

TEXT CUT AND
TIGHT BINDING
BOOK

UNIVERSAL
LIBRARY

OU_148327

UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No. 517.2
H71B

Acc.No. ^G21485

Hoffman Stephen. P
Basic Analysis.

OSMANIA UNIVERSITY LIBRARY

Call No. 517.2 Accession No. G 21485

H 71 B
Author Hoffman Stephen. P

Title Basic Analysis.

This book should be returned on or before the date
last marked below.

BASIC ANALYSIS

Stephen P. Hoffman

Trinity College

Holt, Rinehart and Winston

New York

Copyright, ©, 1961 by

Holt, Rinehart and Winston, Inc.

Library of Congress Catalog Card Number: 61-7858

23688-0111

Printed in the United States of America

All Rights Reserved

to
R.V.C.



FOREWORD

It is my firm conviction that the main business of mathematics is the making of mathematics. By this I mean proving theorems to construct the theory of that particular branch of the subject which is under consideration. This book represents an attempt to do that for the theory of differential and integral calculus of functions of one real variable.

Experience has convinced me that many students falter in their study of calculus because of their shortcomings in algebraic manipulation. In an attempt to minimize the effect of these shortcomings, I have arranged this book so that the usual differentiation and integration theory is presented first for polynomial functions only. The theory is then extended to rational functions, algebraic functions, trigonometric and inverse-trigonometric functions, and logarithmic and exponential functions. Finally, a chapter on infinite series attempts to show how the subject may be considered as one involving only polynomials and power series.

The first seven chapters constitute a leisurely one-year (3 hours per week) course. The first nine chapters cover the material usual in a first-year course in calculus. The entire book is designed for three (3 hours per week) half-year courses. An appendix on the conic sections is intended to round out the study of plane analytic geometry. This appendix may be used at any point after Chapter Three, with its insertion following Chapter Three or Chapter Six being probably the most feasible.

Differentials have been avoided on the ground that they are functions of several variables and thus cannot be handled properly within the scope of this book. Newton's method and a section on rates of change are included for the purpose of approximations.

Many people have given me aid and comfort in the preparation of this book. I should like particularly to cite Prof. Burton W. Jones, Prof. Mario J. Poliferno, my colleagues at Trinity College, Mrs. Wanda Pineault, and Mr. Richard A. Stroud.

S. H.

Hartford, Connecticut
February, 1961



CONTENTS

1. The Number System

1.1	Counting: Straight Lines	1
1.2	The Number Line	4
1.3	Arithmetic and the Natural Numbers	5
1.4	Zero: the Integers	11
1.5	The Rational Numbers	16
1.6	Algebraic Properties of the Rational Numbers	19
1.7	Reciprocals	20
1.8	Ordering Properties of the Rational Numbers	23
1.9	Some Facts about the Natural Numbers	25
1.10	The Real Numbers	26
1.11	Linear Coordinate Systems	30

2. Equations and Inequalities

2.1	Variables and Constants	31
2.2	Equations	34
2.3	Inequalities	41
2.4	Quadratic Inequalities	45
2.5	Other Inequalities	52
2.6	Absolute Values	55

3. Plane Analytic Geometry

3.1	Plane Coordinate Systems	60
3.2	Distance	63
3.3	Graphs of Equations and Inequalities	65
3.4	Straight Lines	69
3.5	Algebraic Aspects of Geometric Properties of Lines	78
3.6	Distance between a Point and a Line	84
3.7	The Graph of the Equation $y = x^2$	87
3.8	Further Investigation of the Graph of $y = x^2$	93

X CONTENTS

4. Functions

4.1 Functions	100
4.2 Graphs of Functions	105
4.3 Constant Functions	111
4.4 The Sum of Two Functions	112
4.5 Products and Quotients of Functions	120
4.6 The Identity Function: Polynomials	125

5. Limits and Derivatives

5.1 Approximations on Graphs	128
5.2 Limits of Functions	132
5.3 Some Theorems of Limits	137
5.4 Limits for Polynomial Functions	140
5.5 Derivatives	143
5.6 The Derivative as a Rate of Change	150
5.7 Tangent Lines	156
5.8 Increasing and Decreasing Functions	159
5.9 Higher Derivatives and Anti-derivatives	166

6. Applications of the Derivative

6.1 Continuity and Differentiability	170
6.2 Maximum and Minimum Points on Graphs	174
6.3 Theorems on Relative Maxima and Minima	178
6.4 Further Theorems on Extremes	181
6.5 Applications of Maxima and Minima	185
6.6 Concavity	188
6.7 Points of Inflection	191
6.8 Newton's Method	193

7. Integration

7.1 Some Facts about the Natural Numbers	197
7.2 Area	201
7.3 Sequences	207
7.4 Upper and Lower Integrals	210
7.5 The Definite Integral	213
7.6 Functions without Integrals	215
7.7 Some Theorems about Integrals	216
7.8 Area	222
7.9 The Fundamental Theorem of Calculus	223
7.10 Application of the Fundamental Theorem to the Evaluation of Integrals	230

8. Algebraic Functions

8.1 Integer Exponents	234
-----------------------	-----

8.2	The Quotient of Two Functions	235
8.3	Derivative and Anti-derivative for $\{I^p\}()$, p a Negative Integer	239
8.4	The Derivative of a Quotient	242
8.5	Derivative of the Product of Two Functions	245
8.6	Rational Functions	248
8.7	"Infinite" Limits	249
8.8	Other Asymptotes	254
8.9	Vertical Tangents	257
8.10	Rational Number Exponents	258

9. Composites of Functions

9.1	The Composite of Two Functions	263
9.2	The Limit Theorem for the Composite of Two Functions	267
9.3	The Derivative of a Composite ("Chain Rule")	269
9.4	Functions Given by an Equation	274
9.5	Related Rates	279
9.6	The Inverse of a Function	283
9.7	Review of Differentiation	287
9.8	Anti-differentiation	288
9.9	Anti-differentiation and Composites	292
9.10	Anti-differentiation by Parts	297
9.11	The Calculus of Algebraic Functions	299

10. Trigonometric Functions

10.1	Sines and Cosines of Numbers	300
10.2	The Cosine and Sine Functions	303
10.3	Other Trigonometric Functions	305
10.4	Some Useful Relationships	308
10.5	Some Limit Theorems	313
10.6	Derivatives of the Trigonometric Functions	316
10.7	Anti-derivatives	318
10.8	Inverse Trigonometric Functions	321
10.9	The Derivatives of the Inverse Trigonometric Functions	323
10.10	Some Anti-differentiation	327
10.11	A Method of Redoing the Trigonometric and Inverse Trigonometric Functions	330

11. Logarithmic and Exponential Functions

11.1	The Function Log ($\cdot$)	332
11.2	Logarithmic Properties of Log ($\cdot$)	335
11.3	The Exponential Function	338
11.4	Properties of the Exponential Function	339
11.5	The Irrational Number e	341

XII CONTENTS

11.6	The Extended Exponential Functions	342
11.7	The Hyperbolic Functions	345
11.8	The Inverse Hyperbolic Functions	350

12. Applications of Integration

12.1	Volume	355
12.2	Right Circular Cylinders	356
12.3	Spheres	363
12.4	Volume of a Solid of Revolution	365
12.5	Solids with Regular Cross-sections	370
12.6	Shell Method for Volumes of Revolution	373
12.7	Algebraic Simplicity and Geometric Complexity	378
12.8	Arc Length	382

13. Infinite Series

13.1	Sequence of Numbers	387
13.2	Test for Convergence and Divergence	392
13.3	Series of Functions	397
13.4	Differentiation, Anti-differentiation, and Integration	402
13.5	Taylor's Theorem and Taylor's Series	406
13.6	Rational Functions	409
13.7	Algebraic Functions	413
13.8	Trigonometric Functions	415
13.9	Logarithmic and Exponential Functions	418
13.10	Conclusion	422

Appendix

CLASSIFICATION OF THE GRAPHS OF SECOND-DEGREE EQUATIONS; CONIC SECTIONS		423
A.1	Definitions of Circle, Parabola, Ellipse, and Hyperbola	423
A.2	Rotation Principle	427
A.3	Justification of Rotation Principle	429
A.4	Translation Principle	431
A.5	Reduction of the Classification Problem	434
A.6	Conclusion and Summary	440

Answers to Odd-Numbered Exercises	443
--	------------

Index	457
--------------	------------



THE NUMBER SYSTEM

1.1. COUNTING: STRAIGHT LINES

Simple questions often have nonsimple answers. It is easy to ask the question: "What is a number?" but not easy to answer it. In this chapter we will attempt a partial answer.

Everyone learns to count at an early age. There are two distinct acts involving counting: counting things and "just counting." The act of counting—as distinct from the act of counting things—consists in presenting the numbers 1, 2, 3, 4, etc., in their "natural order." The act of counting a set of objects consists in using these numbers in the following manner:

Given a collection of objects, we choose any one of these objects and label that one with the number 1; we then choose any one of the remaining, unlabeled objects, and label that one with the number 2; we then choose one of the remaining, unlabeled objects, and label that one with the number 3; we continue this process of choosing a remaining, unlabeled object and giving that object a label, until we have exhausted the set of objects—that is, until no more objects remain unlabeled. We have used for the labeling some of the numbers: 1, 2, 3, 4, etc. Of those numbers which we actually have used for the labeling, we consider the last (and largest) as the "number of" objects in the collection.

There are two basic mathematical ideas involved in this process of counting objects. First, we have the concept of the numbers: 1, 2, 3, 4, etc.—those numbers which we used as labels. Second, we have the concept of labeling objects with numbers, or, more broadly, the concept of setting up a correspondence between the members of two collections—in this particular instance, between a given collection of objects and a collection of numbers, the latter being part of the set of numbers: 1, 2, 3, 4, etc. Whenever a correspondence between the members of two sets is set up in such a manner that each single thing in one set corresponds to one and only one thing in the second set, and vice versa, then such a correspondence is called a *one-to-one correspondence*.

As an example of a one-to-one correspondence, let us consider what occurs when a business firm sends out a day's mail. Normally, each letter involves two separate pieces of paper: the letter and the envelope. The one-to-one correspondence here is governed by the rule that each letter must be inserted in the envelope addressed to the person who is to receive that letter. One can imagine the confusion which reigns when a one-to-one correspondence is not set up between the letters and the envelopes, so that at the end of the day there is one envelope left over, or one letter without an envelope.

Even though the establishment of a one-to-one correspondence is part of the process of counting objects, it is possible to use a one-to-one correspondence to avoid counting. Let us suppose that a hardware dealer has one large box filled with bolts and another large box filled with nuts of a size to match the bolts. If he wanted to know whether or not he has just as many bolts as he has nuts, he could certainly set out to count the number of bolts, and then, separately, count the number of nuts. If the two numbers were the same, then he could validly use the phrase "just as many" to describe the number of nuts and bolts in the boxes. However, the counting process may be avoided completely by merely pairing the nuts and bolts in such a manner that each nut is matched with exactly one bolt and each bolt is matched with exactly one nut. In fact, a most natural way to accomplish this pairing would be to thread one nut lightly on the end of each bolt. If there were no bolts and no nuts remaining after this pairing operation, then the dealer would validly use the phrase "just as many" to describe the number of nuts and bolts in the boxes, even though he would not know that number.

A question which begins with the words "how many" is usually answered by using a one-to-one correspondence. When the answer to such a question is a specific counting number (for example, 5, or 57, or 3,001,204), the one-to-one correspondence occurred in counting. When the answer to the "how many" question is not a specific counting number, but involves the phrase "just as many," the one-to-one correspondence occurred in a pairing of the elements of two sets of objects.

Some "how many" questions cannot be answered by using a specific counting number. For example, the question "How many counting numbers are there?" cannot be answered by using one of the counting numbers. But it is possible to give a reasonable definition to the phrase "just as many," which will tell us about some aspect of this set of numbers: 1, 2, 3, 4, etc.

To say that two sets of objects contain "just as many" objects means that it is possible to establish a one-to-one correspondence between the objects in the two sets, which correspondence does apply to all the objects in both sets. I now assert that, under this meaning of the phrase "just as many," the set of all natural numbers: 1, 2, 3, 4, etc., and the set of all even natural numbers: 2, 4, 6, 8, etc., contain just as many numbers.

In order to show that this assertion is true, it is necessary to establish a one-to-one correspondence between the elements of the set of all natural numbers and the elements of the set of all even natural numbers. The most

natural such one-to-one correspondence is governed by the following rule: To each number chosen from the set of all natural numbers, we correspond the even natural number which is the double of the chosen one. Thus, 1 in the set of all natural numbers is paired with 2 in the set of even natural numbers; 2 in the set of all natural numbers is paired with 4 in the set of even natural numbers; 3 in the set of all natural numbers is paired with 6 in the set of even natural numbers; and so on. Clearly, this rule does establish a one-to-one correspondence between the elements of the set of all natural numbers and the elements of the set of all even natural numbers, which correspondence does apply to all the numbers in both sets. Thus, there are "just as many" natural numbers as there are even natural numbers.

The set of even natural numbers is a "proper subset" of the set of all natural numbers. Any set which has the property that a one-to-one correspondence can be established between the entire set and one of its proper subsets is called an infinite set. A finite set then is defined to be not an infinite one, so that we cannot put a finite set into a one-to-one correspondence with any of its proper subsets.

Completely separate and distinct from the idea of counting, and from the idea of a one-to-one correspondence, is the idea one learns at an early age that there "exists" such a thing as a *straight line*. Of course, this "existence" is a purely mental one—we conceive of, or imagine, a straight line. For the study of mathematics, it is important to realize that we consider a straight line not as a succession of points, but as a flow of points. That is, given a straight line, and given one particular point on that line, there is no point "next to" that given one. Another way of saying the same thing is to say that if we are given any two distinct points of a straight line, then between these two points there are other points of that line. In fact, we can go on and say that between any two given points of a straight line there are an "infinite number" of points of this line. However, the phrase "infinite number," although appealing, is rather a subtle one, so that we will avoid the use of this phrase, or any phrase containing the word "infinite" or "infinity," whenever we possibly can.

An important elementary concept associated with the idea of a straight line is that of distance. Usually distance is measured (that is, given in terms of a number) by counting. The objects which are counted are "units of distance" (inches, miles, kilometers). Somehow, then, we have a primitive concept of one particular distance's serving as a standard, and we give all other distances in terms of that one particular *unit distance*.

We take then, as a starting point, these four basic concepts:

- (1) the set of natural numbers: 1, 2, 3, 4, etc.;
- (2) a one-to-one correspondence;
- (3) a straight line;
- (4) a unit of distance.

Exercises

1. When the set of natural numbers is illustrated by:

1, 2, 3, 4, etc.,

what is the significance of the "etc."?

2. Which of the following numbers are natural numbers:

$\frac{1}{2}$, 3, 79, -4, 5102, 2,880,903, 0, ?

3. How can a one-to-one correspondence be used to discover whether or not a theatre has a capacity audience?
4. Why does "counting noses" tell you how many people are present?
5. Describe the one-to-one correspondences used in the children's game of "musical chairs."

Use a one-to-one correspondence to show:

4. There are just as many states as state capitals.
5. There are seven words in this sentence.
6. There are just as many natural numbers which end in the digit 5 (5, 15, 25, 35, etc.) as there are natural numbers.
7. There are just as many natural numbers which end in the digit 7 as there are natural numbers which end in the digit 2.
8. There are just as many natural numbers which do not end in the digit 7 as there are natural numbers which do not end in the digit 4.
9. There are not just as many states as there are Senators.
10. There are just as many natural numbers which do not end in the digit 6 as there are natural numbers which do end in the digit 9.

1.2. THE NUMBER LINE

In this section, and for the rest of this chapter, we will keep our attention fixed on one particular straight line, placed horizontally, and extending indefinitely both to the left and to the right. This line will be called *the number line*. Our goal is to attempt to use some of the points of this line to illustrate the set of all natural numbers, and, in order to attain this objective, we will set up a one-to-one correspondence between these points and the natural numbers.

To begin with, we choose two distinct points on this line, and say that the distance between those two will serve us throughout as *the* unit of distance. Our one-to-one correspondence will be established by labeling points on the number line. Of the two points chosen, we attach to the left-hand point the label "zero" and to the right-hand point the label "1."

In order to give the label "zero" to the left-hand point, we had to have the word "zero" in our vocabulary. Let us postpone, temporarily, a discussion of the number 0, and merely consider the word "zero" as a word whose meaning has not yet been given.

We now proceed to label more points on the number line. Technically, we call a number which is used as the label of a point *the coordinate* of that point. That point which is one unit to the right of the point with coordinate

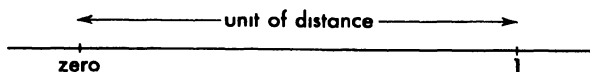


Fig. 1.1

"1" will be given the label "2." That point one unit to the right of the point with coordinate "2" will be given the label "3." That point one unit to the right of the point with coordinate "3" will be given the label "4." If we envisage this labeling process as continuing indefinitely, we will have labeled certain points on the number line with all the natural numbers, and this labeling process establishes a one-to-one correspondence between these points and the elements of the set of all natural numbers. This one-to-one correspondence is so arranged that to each natural number, n , there corresponds one and only one point, namely that point with coordinate " n ." The point with coordinate " n "—where n is some natural number—is precisely n units to the right of the point with label "zero." Clearly there are many points on the number line which we have not labeled.

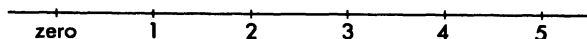


Fig. 1.2

1.3. ARITHMETIC AND THE NATURAL NUMBERS

In this section we will examine certain facts about the operations of arithmetic—insofar as they apply to the set of natural numbers, and examine the connections between these algebraic facts and certain geometric facts about the number line.

The first algebraic operation usually learned is that of addition. We state the basic properties of addition as "laws," using an indexing system for these laws which indicates what kind of numbers we are discussing, what kind of operation, and which law is involved. First, we have:

N-A-1: If m and n are natural numbers, then there is a unique natural number, written $m + n$, which is the result of adding n to m , and which is called the sum of m and n .

The index, N-A-1, indicates that this law is the first addition law for the natural numbers. We may interpret N-A-1 on the number line if we have a method of interpreting the operation of adding one natural number to another. If m is a natural number, and n is a natural number, the $m + n$ will be the natural number used for the coordinate of that one point which is precisely n units to the right of the point with coordinate " m ." We thus interpret "adding the natural number n " as "moving precisely n units to the right."

The law N-A-1 says that the new point reached by moving n units to the right of the point with coordinate " m " has already been labeled with a natural number, so that the result of adding any two natural numbers (which need not be distinct ones) is always another natural number.

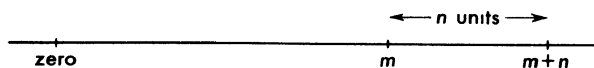


Fig. 1.3

N-A-2: If m and n are natural numbers, then $m + n$ and $n + m$ are the same natural number. (Symbolically, $m + n = n + m$.)

This second addition law for the natural numbers may be interpreted by saying that the point which is n units to the right of the point with coordinate " m " is also m units to the right of the point with coordinate " n ."

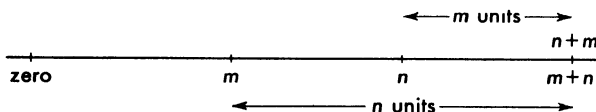


Fig. 1.4

For example, $3 + 5$ and $5 + 3$ are both the natural number 8. Geometrically, the point 5 units to the right of the point with coordinate "3" is also 3 units to the right of the point with coordinate "5."

N-A-3: If m , n , and p are natural numbers, then the sum of m and $n + p$ is the same natural number as the sum of $m + n$ and p . (Symbolically, $m + (n + p) = (m + n) + p$.)

This third addition law for the natural number is used whenever we have to find the sum of three or more natural numbers. For a specific case, we find the sum of the numbers 3, 4, and 6 by first finding the sum: $3 + 4$ —that is, the natural number 7—and then, to this result, adding 6, so that we have

$$(3 + 4) + 6 = 7 + 6 = 13.$$

The law N-A-3 says that the natural number 13 will also result from adding the sum $4 + 6$ to the natural number 3:

$$3 + (4 + 6) = 3 + 10 = 13.$$

An interpretation of N-A-3 on the number line is: that point which is $n + p$ units to the right of the point with coordinate " m " is also p units to the right of the point with coordinate " $m + n$."

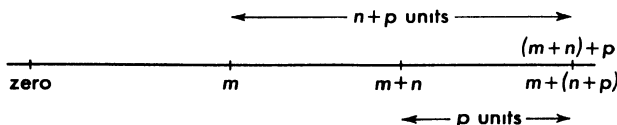


Fig. 1.5

There are two ways of giving geometric meaning on the number line to the multiplication of natural numbers, one in terms of distances and areas, and the other in terms of distances alone. The determination of the product, mn , of two natural numbers, m and n , in terms of distances and areas, requires that one count the number of squares whose sides have unit length, which squares are contained in a rectangle whose sides have lengths m and n units of distance. Since our purpose is to interpret algebraic properties in terms of the number line, we will not use the concept of area, but will give an interpretation of multiplication which involves distances only. The first three laws concerning multiplication of natural numbers are similar to those concerning addition:

- N-M-1:** If m and n are natural numbers, then there is a unique natural number, written mn , which is the result of multiplying m by n , and is called the product of m and n .
- N-M-2:** If m and n are natural numbers, then mn and nm are the same natural number. (Symbolically, $mn = nm$.)
- N-M-3:** If m , n , and p are natural numbers, then the product of m and np is the same natural number as the product of mn and p . (Symbolically, $m(np) = (mn)p$.)

As with addition, we need to begin our interpretation of these laws by giving an interpretation of the process of multiplying one natural number by another natural number. This number-line interpretation of the arithmetic process of multiplication is: the point with coordinate " mn " is reached by measuring, to the right of the point with coordinate "zero," n units of distance, and by doing so m times.

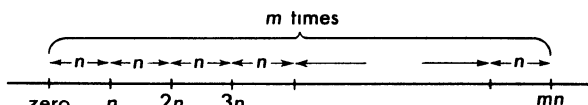


Fig. 1.6

The law N-M-1 says that the point reached by measuring to the right of the point with coordinate "zero," n units of distance, m times, will be a point already labeled with a natural number, and that label will be the natural number mn .

The second multiplication law for the natural numbers is interpreted on the number line by saying that the point reached by measuring to the right of the point with label "zero" m units of distance, n times, is the same point as the one reached by measuring to the right of the point with label "zero" n units of distance, m times. The construction of a diagram to illustrate this fact is left as an exercise for the reader.

The law N-M-3 may be interpreted on the number line by saying that the point reached by measuring to the right of the point with label "zero," np units of distance, m times, is also reached by measuring to the right of

the point with label "zero," p units of distance, mn times. For example, in a specific case, N-M-3 requires that

$$[(2)(5)] 7 = 2[(5)(7)];$$

that is,

$$(10) 7 = 2(35).$$

On the number line, the two statements indicate that the point with coordinate "70" may be reached either by measuring to the right of the point with label "zero," 7 units, 10 times, or 35 units, twice.

By this time we have used the phrase "the point with label 'zero'" fairly often. It behooves us, then, to adopt a shorthand for this phrase. Our shorthand consists of a formal definition of a new word:

Definition 1.1: That point on the number line whose label is "zero" is called *the origin*.

There is a law for the multiplication of natural numbers which has no counterpart in the addition of natural numbers. This law shows the special role of the natural number 1.

N-M-4: 1 is a natural number. For each and every natural number, m , $m(1)$ and m are the same natural number. (Symbolically, $m(1) = m$.)

The number-line interpretation of this law is extremely simple: that point which is m units to the right of the origin has coordinate " m ."

Finally, there is a law which connects the two operations: addition and multiplication, which is called the distributive law for the natural numbers:

N-D: If m , n , and p are natural numbers, then the product of m and $n + p$ is the same natural number as the sum of mn and mp . (Symbolically, $m(n + p) = mn + mp$.)

On the number line, this distributive law for the natural numbers says that the point reached by measuring to the right of the origin $n + p$ units of distance, m times, is also the point mp units to the right of the point with coordinate " mn ." For example, N-D requires

$$2(3 + 5) = (2)(3) + (2)(5);$$

that is

$$2(8) = 6 + 10.$$

From these eight basic facts of life about the natural numbers, we can deduce other facts, which we call theorems. The deduction process is carried out in the proofs of these theorems. We will construct our proofs in a form similar to that normally used in the study of plane geometry, namely as a succession of steps and reasons, the last step being the conclusion of the theorem. As an illustration of this process and form, we state and prove a theorem the truth of which no one would doubt.

Theorem 1.1:

If m is a natural number
then $(1)m = m$.

In this theorem the statement " m is a natural number" is called the *hypothesis* and the statement " $(1)m = m$ " is called the *conclusion*.

Proof of 1.1:

- | | |
|----------------------------|------------------------------|
| 1. m is a natural number | by hypothesis. |
| 2. $m(1) = m$ | by step 1 and N-M-4. |
| 3. $m(1) = (1)m$ | by step 1, N-M-4, and N-M-2. |
| 4. $(1)m = m$ | by steps 2 and 3. |

In this proof, we have used a basic law concerning the relation of equality. We pause to point out that the relation of equality has three basic properties:

- (1) For any thing, that thing = that thing.
- (2) For any two things, if first thing = second thing
then second thing = first thing.
- (3) For any three things, if first thing = second thing,
and if second thing = third thing,
then first thing = third thing.

These statements about equality are true no matter what things are being discussed, and, in particular, are true if the things are natural numbers. For the natural numbers, we also have:

If m , n , and p are natural numbers, and
if $m = n$, then $m + p = n + p$ and $mp = np$.

The one-to-one correspondence which we established between the natural numbers and certain points on the number line depended on our understanding of the terms "left" and "right." Since any pair of distinct points on the number line consists of a definite left-hand point and a definite right hand point, we are led to relating the sizes of natural numbers by their relative positions on the number line. We will say that the natural number m is less than the natural number n provided that the point with coordinate " m " is to the left of the point with coordinate " n ."

The laws governing the ordering of the natural numbers are:

N-O-1: If m and n are two natural numbers, then one and only one of the following three statements is true:

$$m = n; \quad m < n; \quad m > n.$$

(Here, the symbol $<$ is read "is less than" and the symbol $>$ is read "is greater than.")

N-O-2: If m and n are natural numbers, then $m < n$ if and only if there is some natural number, p , such that $m + p = n$.

N-O-3: If m , n , and p are natural numbers, and if $m < n$, then $m + p < n + p$.

N-O-4: If m , n , and p are natural numbers, and if $m < n$, then $mp < np$.

In terms of the number line, the first ordering law for the natural numbers says that given any two different natural numbers as coordinates of two points, one of these points will lie to the left of the other.

The second ordering law says that $m < n$ if and only if the point with coordinate " n " is actually to the right of the point with coordinate " m ," and further, some natural number of units of distance to the right. For example, this law requires $3 < 12$ to be true, since there does exist a natural number, namely 9, such that $3 + 9 = 12$.

The third ordering law for natural numbers says that if the point with coordinate " n " is to the right of the point with coordinate " m ," then the point p units to the right of the point with coordinate " n " will be to the right of the point p units to the right of the point with coordinate " m ." For example, this law requires that, since $2 < 7$, $2 + 3 < 7 + 3$, that is that $5 < 10$.

The fourth ordering law for natural numbers says that the point arrived at by measuring to the right of the origin p units of distance, n times, will be further to the right than the point arrived at by measuring to the right of the origin p units of distance, m times, under the hypothesis $m < n$. For example, $4 < 15$ implies that $4(3) < 15(3)$, that is $12 < 45$.

Exercises

1. Define the word "unique."
2. Construct a number-line illustration of N-A-2 for $m = 2$, $n = 7$.
3. Construct a number-line illustration of N-A-3 for $m = 1$, $n = 4$, and $p = 3$.
4. List all possible ways in which the sum of the natural numbers 3, 4, and 6 can be found, and show each way on the number line.
5. Construct a number-line illustration of N-M-2 for $m = 3$, $n = 4$.
6. Construct a number-line illustration of N-M-3 for $m = 4$, $n = 2$, and $p = 3$.
7. List all possible ways in which the product of the natural numbers 3, 4, and 6 can be found, and show each way on the number line.
8. Construct a number-line illustration of N-D for the particular cases:
 - a) $m = 2$, $n = 1$, $p = 3$.
 - b) $m = 2$, $n = 3$, $p = 3$.
 - c) $m = 2$, $n = 2$, $p = 2$.
9. In the proof of Theorem 1.1, what part of N-M-4 is used in the reason for step 2? in the reason for step 3?

10. Prove the theorem: If m , n , and p are natural numbers then $m(n + p)$ and $pm + nm$ are the same natural number.
11. Give five specific examples of N-O-2.
12. Give a number-line interpretation of N-O-3 for $m = 2$, $n = 4$, and $p = 3$.
13. Give a number-line interpretation of N-O-4 for $m = 3$, $n = 5$, and $p = 4$.
14. Prove the theorem: If m , n , and p are natural numbers, then $m(n + p) > mn$.
15. If, in N-A-1, the word "adding" is changed to "subtracting," and other changes are made to get a "law" for subtraction, will the resulting "law" be true? Why or why not?
16. If, in N-M-1, the word "multiplying" is changed to "dividing" and other changes are made to get a "law" for division, will the resulting "law" be true? Why or why not?

1.4. ZERO: THE INTEGERS

In section 1.2, we labeled a point on the number line, now called the origin, with the word "zero." Since we have defined the set of natural numbers to be: 1, 2, 3, etc., zero is not a natural number. When learning the basic algebraic operations of addition and multiplication of natural numbers, one also learns to subtract. In the subtraction learned, the result of subtracting one number from another gave a number. We again emphasize the distinction between pure and applied mathematics: when one "subtracts" three objects from a set consisting of three objects, the result of this subtraction is properly described as "nothing." But when one subtracts the *number* 3 from the *number* 3, the result of this subtraction of numbers should also be a *number*, in the case the *number* 0. Even in applied mathematics, there are examples where subtraction naturally always yields a number. If the temperature is 15° in the afternoon, and drops 15 degrees during the night, then the next morning the temperature is 0° , which is not nothing, but merely a mark on the thermometer. It is thus false to say "0 is nothing," since 0 is very definitely something, a number. This number, 0, to repeat, is not one of the natural numbers, so that we must extend our idea of "number" beyond the natural numbers to include 0. When we do, the origin of the number line then has coordinate "0" rather than label "zero."

In learning subtraction in elementary school, one is usually taught that the number 5 cannot be subtracted from the number 3. This statement is incomplete. If the only numbers available are 0 and the natural numbers, then $3 - 5$ is certainly not one of these. But, in order to allow any natural number to be subtracted from any other natural number, and to have the result of this subtraction be a number (although not necessarily a natural

number) we invent a new kind of number—one which is the “negative” of a natural number. We formalize this invention in:

Definition 1.2: The number 0 is defined to be that one number which has the property that the sum of 0 and any given number is that given number. (Symbolically, $a + 0 = a$, for any number, a .)

Definition 1.3: The set of *integers* consists of 0, the natural numbers, and the negatives of the natural numbers. The natural numbers are called *positive integers*. The negatives of the natural numbers are called *negative integers*.

On the number line, we give that point one unit to the left of the origin the coordinate “-1.” We give the point one unit to the left of the point labeled “-1” the coordinate “-2.” We continue labeling in this manner, and, stretching our imagination as before, visualize a one-to-one correspondence set up between certain points of the number line and the negative integers.

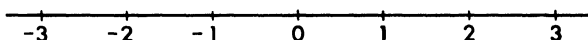


Fig. 1.7

This labeling of certain points to the left of the origin, together with our previous labeling, establishes a one-to-one correspondence between certain points on the number line and the integers. Under this correspondence, the distance between any two of these points is a certain number of units of distance, which number is a natural number found by subtracting the coordinate of the left-hand point from the coordinate of the right-hand point. We have certain algebraic laws for the integers, some of which are completely analogous to the laws for the natural numbers.

I-A-1: If p and q are integers, then there is a unique integer, written $p + q$, which is the sum of p and q .

Since integers are not necessarily natural numbers, the number-line interpretation of the addition of one integer to another is not as easy as was the case with natural numbers. We interpreted addition of a natural number as a move to the right on the number line. Thus it is reasonable to interpret subtraction of a natural number as a move to the left on the number line. But the subtraction of a natural number is equivalent to the addition of the negative of that natural number. Thus we will interpret the addition of a negative integer as a motion to the left on the number line.

If q is a positive integer, then the point with coordinate “ $p + q$ ” will be q units to the right of the point with coordinate “ p .” If q is the integer 0, the point with coordinate “ $p + 0$ ” will be the point with coordinate “ p .” If q is a negative integer, then the negative of q , $-q$, will be a positive integer.

The point with coordinate " $p + q$," in this case, will be $-q$ units to the left of the point with coordinate " p ." This last fact is one instance of the necessity of remembering that, unless we are sure that all numbers concerned must be positive, the presence of a minus sign does not *necessarily* indicate the presence of a negative number.

I-A-2: If p and q are integers, then $p + q$ and $q + p$ are the same integer. (Symbolically, $p + q = q + p$.)

If we ignore the extremely simple cases in which one or both of the integers p and q are 0, the number-line interpretation of this second addition law for the integers falls into four cases:

CASE 1, p AND q ARE BOTH POSITIVE INTEGERS

In this case we are dealing with natural numbers, so that we have precisely the same number-line interpretation as we had for N-A-2, namely: the point q units to the right of the point with coordinate " p " is also p units to the right of the point with coordinate " q ."

CASE 2, p IS A NEGATIVE INTEGER, AND q IS A POSITIVE INTEGER

Since q is positive, the point with coordinate " q " lies to the right of the origin. Since p is negative, the point with coordinate " p " lies to the left of the origin. The point with coordinate " $p + q$ " will lie between these two points. I-A-2 says that the point q units to the right of the point with coordinate " p " is also $-p$ units to the left of the point with coordinate " q ."

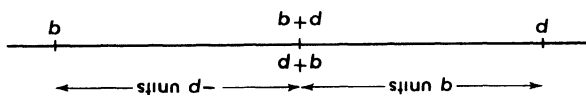


Fig. 1.8

CASE 3, p IS A POSITIVE INTEGER, AND q IS A NEGATIVE INTEGER

The interpretation of this case of I-A-2 is left as an exercise.

CASE 4, p AND q ARE BOTH NEGATIVE INTEGERS

The points with coordinate " p ," " q ," and " $p + q$ " will all be to the left of the origin. I-A-2 says that the point $-q$ units to the left of the point with coordinate " p " is also $-p$ units to the left of the point with coordinate " q ."

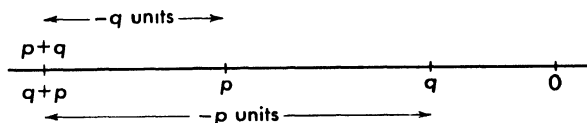


Fig. 1.9

I-A-3: If p , q , and r are integers, then the sum of p and $q + r$ is the same integer as the sum of $p + q$ and r . (Symbolically, $p + (q + r) = (p + q) + r$.)

If we ignore the cases in which one or more of the integers involved is 0, then the number-line interpretation of this third addition law for integers depends on the positive or negative character of the integers involved. By I-A-2, we have

$$p + q = q + p$$

so that

$$(p + q) + r = (q + p) + r;$$

we also have

$$p + (q + r) = (q + r) + p.$$

These two facts, taken together, mean that we need only interpret the statement

$$(q + r) + p = (q + p) + r.$$

For this, we need not consider separately those cases in which q is positive and in which q is negative, so that we need to examine the four cases:

- (a) r and p are both positive
- (b) r is negative and p is positive
- (c) r is positive and p is negative
- (d) r and p are both negative.

For case (a), we have: that point p units to the right of the point with coordinate " $q + r$ " is also r units to the right of the point with coordinate " $q + p$." For case (b), we have: that point which is p units to the right of the point with coordinate " $q + r$ " is also $-r$ units to the left of the point with coordinate " $q + p$." The remaining two interpretations are left as exercises.

Since the number-line interpretation of the various multiplication laws becomes cumbersome in the consideration of the various possible cases, we will omit it completely, and merely state these laws:

I-M-1: If p and q are integers, then there is a unique integer, written pq , which is called the product of p and q .

I-M-2: If p and q are integers, then pq and qp are the same integer. (Symbolically, $pq = qp$.)

I-M-3: If p , q , and r are integers, then the product of p and qr is the same integer as the product of pq and r . (Symbolically, $p(qr) = (pq)r$.)

I-M-4: 1 is an integer. For each and every integer, p , $p(1)$ and p are the same integer. (Symbolically, $p(1) = p$.)

I-D: If p , q , and r are integers, then the product of p and $q + r$ and the sum of pq and pr are the same integer. (Symbolically, $p(q + r) = pq + pr$.)

Examples of these laws for specific integers are:

$$\begin{aligned} -3 + 2 &= 2 + (-3); \\ -4 + (-5 + 1) &= [(-4) + (-5)] + 1; \\ -2[3 + (-1)] &= (-2)(3) + (-2)(-1). \end{aligned}$$

We will take as known fact: the product of two negative integers is a positive integer, and the product of a negative integer and a positive integer is a negative integer.

The number-line interpretation of the ordering of integers is precisely the same as that for the natural numbers: $p < q$ means that the point with coordinate " p " is to the left of the point with coordinate " q ." As a result of this meaning of "less than," we will have to become accustomed to such unfamiliar, but true statements as " $-101 < -5$," " $-3 < -2$," " $-1 < 1$," " $-2000 < 3$." We have, for the integers, the ordering laws:

I-O-1: If p and q are integers, then one and only one of the following three statements is true:

$$p = q; \quad p < q; \quad p > q.$$

I-O-2: If p and q are integers, then $p < q$ if and only if there is some positive integer (that is, some natural number), n , such that $p + n = q$.

I-O-3: If p , q , and r are integers, and if $p < q$, then $p + r < q + r$.

There are two "fourth" ordering laws for integers:

I-O-4a: If p and q are integers; if $p < q$; and if r is a *positive* integer, then $pr < qr$.

I-O-4b: If p and q are integers; if $p < q$; and if s is a *negative* integer, then $ps > qs$.

As an example of I-O-4a: $-4 < -1$; $(-4)(3) = -12$; $(-1)(3) = -3$, and $-12 < -3$. As an example of I-O-4b: $-4 < -1$; $(-4)(-2) = 8$, $(-1)(-2) = 2$, and $8 > 2$.

The creation of the new numbers: 0 and the negative integers, gives us two new addition laws for the integers:

I-A-4: 0 is an integer. For each and every integer, p , $p + 0 = p$.

I-A-5: For each and every integer, p , there exists a unique corresponding integer, $-p$, such that $p + (-p) = 0$.

It is interesting to note the similarity between I-A-4 and I-M-4. This similarity shows that the integer 0 plays the same role in addition which is played in multiplication by the natural number 1.

The fifth addition law for integers allows us to define the operation of subtraction for integers by saying that, if p and q are integers, then $p - q$

is defined to be the sum $p + (-q)$. A number-line interpretation of this fifth addition law for integers is that for each point with an integer coordinate there is exactly one corresponding point with integer coordinate. This one corresponding point lies on the opposite side of the origin and is exactly the same distance from the origin as the given point.

Exercises

1. Why is $1/2$ not an integer?
2. What is the distance between the point with coordinate " -3 " and the origin?
3. If q is a negative integer, what is the distance between the point with coordinate " q " and the origin?
4. What is the distance between the point with coordinate " 5 " and the point with coordinate " $5 + (-3)$ "?
5. What is the distance between the point with coordinate " -2 " and the point with coordinate " $-2 + (-3)$ "?
6. If s is a negative integer, what is the distance between the point with label " p " and the point with label " $p + s$ "?
7. Sketch the number-line interpretation of the missing case of I-A-2.
8. Sketch the number-line interpretation of the two missing cases of I-A-3.
9. Draw an accurate illustration on the number line of the statement $(-3)[2 + (-4)] = (-3)(2) + (-3)(-4)$.
10. Show on the number line that:
 - a) $3 + 5 < 4 + 5$.
 - b) $3 + (-3) < 4 + (-3)$.
 - c) $(-4)(5) < (-2)(5)$.
 - b) $(-4)(-2) > (-2)(-2)$.
11. Prove the theorem: If p is a negative integer, then $-p$ is a positive integer. (Hint: use Def. 1.3, I-A-5, and I-A-2.)
12. Prove the theorem: If p is an integer, and $p + p = p$, then p is the integer 0.
13. Give a number line interpretation of I-O-3.

1.5. THE RATIONAL NUMBERS

Let us now consider a simple problem in the measurement of distances. Suppose we have a triangle, with vertices labeled A , B , and C , with the distance between A and B the unit distance of the number line. (See Fig. 1.10).

Let us further suppose that m and n are two natural numbers (that is, positive integers), that $m < n$, and that the distance between the points A and C is precisely n units of distance. On the line segment joining A and C there will be precisely one point which is exactly m units of distance from A ,

and we will call this point Q . We draw a line through this point, Q , which line is parallel to the line joining B and C . The intersection of this line through Q with the line joining A and B will be called R .

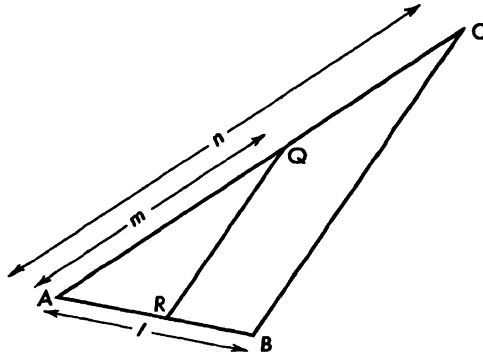


Fig. 1.10

The question to be answered is: "What is the distance between A and R ?" This distance is a "number of" units of distance, and the geometry of similar triangles prompts us to say that this number is m/n . But the quotient, m/n , is not a natural number, nor even an integer, since $m < n$, so that we are once more forced to invent a new kind of number. Rather than use the word "fraction," which has connotations of being somehow "small," we introduce the adjective "rational" (which comes from the word "ratio"), and present:

Definition 1.4a: A *positive rational number* is a number which can be expressed as the quotient of two natural numbers.

We may then extend our labeling system on the number line to points whose coordinates will be numbers such as $1/2$, $5/4$, $47/18$, etc.

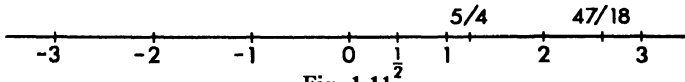


Fig. 1.11

In ordinary arithmetic it is customary to write rational numbers which are greater than 1 (that is, $3/2$, $11/4$) as "mixed numbers" ($1\frac{1}{2}$, $2\frac{3}{4}$). Since any further algebraic manipulations would require that a "mixed number" be converted into an "improper fraction," we will adopt the policy of leaving such quotients as quotients. We thus find the expression $3/2$ more convenient for our purposes than the expression $1\frac{1}{2}$.

Our next step should be to label points on the number line which lie to the left of the origin in a manner similar to that used for the positive rational numbers as coordinates of certain points to the right of the origin. If r is a positive rational number, then we label that point on the number line r units to the left of the origin with the coordinate " $-r$," and we use the definition:

Definition 1.4b: If r is a positive rational number, then the *negative rational number*, $-r$, is that number, which, when added to r gives the sum 0. (Symbolically, $r + (-r) = 0$.)

What we have said, of course, is that a negative rational number is the negative of some positive rational number.

Since any integer, p , may be expressed as the quotient $p/1$, of the integers p and 1, every integer is a positive rational number, or a negative rational number, or 0. We use as a basic definition:

Definition 1.5: A *rational number* is a number which may be expressed as the quotient of two integers, in which the denominator is not the integer 0.

There is a hidden idea in Definition 1.5. This definition gives a name to a certain *kind* of number, but does not give specifically the set of all the numbers of this kind. There is some underlying concept of the set of *all* numbers, and Definition 1.5 says that some (or possibly all) of those numbers are to be called rational numbers.

Where did this idea of the set of all numbers come from? Very simply, the concept of the set of all (real) numbers is a direct outgrowth of the idea of the set of all points on a straight line, and the primitive idea of distance. The one-to-one correspondence between certain numbers and certain points of the number line involved the concept of distance, and the idea of the set of all numbers is really the same as the idea of the set of all possible distances between points on a straight line.

Exercises

- Construct, on the number line, those points with coordinates:
 $2/3$, $4/3$, $-11/5$, $-2/7$, $25/4$.
- Construct, on the number line, the points whose labels are " $-2/3$ " and " $5/4$." Find, on the number line, the point with coordinate " $-2/3 + 5/4$." What is the coordinate of this point in simpler form?
- Construct, on the number line, the points whose coordinates are " -2 " and " $1/2$." Find that point which is midway between these two. What is its coordinate?
- Use the idea of similar triangles to show that the point with coordinate " $2/3$ " and the point with coordinate " $4/6$ " are the same point.
- Construct, on the number line, the point with coordinate " $7/8$ " and the point with coordinate " $9/11$." Does this construction show which of the two numbers is the larger? If so, why? If not, why not? How else could we determine which of these two rational numbers is the larger?
- List the arithmetic steps necessary to add the two rational numbers: $3/4$ and $2/3$.
 - Interpret each step in part (a) on the number line.
 - Compare the addition of $3/4$ and $2/3$ to the addition of 9 and 8.

1.6. ALGEBRAIC PROPERTIES OF THE RATIONAL NUMBERS

The rational numbers obey laws which are similar to those governing the behavior of integers:

- R-A-1:** If r and s are rational numbers, then there is a unique rational number, $r + s$, which is the sum of r and s .
- R-A-2:** If r and s are rational numbers, then $r + s$ and $s + r$ are the same rational numbers. (Symbolically, $r + s = s + r$.)
- R-A-3:** If r , s , and t are rational numbers, then the sum of r and $s + t$ is the same rational number as the sum of $r + s$ and t . (Symbolically, $r + (s + t) = (r + s) + t$.)
- R-A-4:** 0 is a rational number. For each and every rational number, r , $r + 0 = r$.
- R-A-5:** For each rational number, r , there exists a unique rational number, $-r$, such that $r + (-r) = 0$.

The number-line interpretations of these five addition laws for the rational numbers are precisely the same as the interpretations of the similar laws for the integers.

- R-M-1:** If r and s are rational numbers, then there is a unique rational number, rs , which is the product of r and s .
- R-M-2:** If r and s are rational numbers, then rs and sr are the same rational number. (Symbolically, $rs = sr$.)
- R-M-3:** If r , s , and t are rational numbers, then the product of r and st is the same rational number as the product of rs and t . (Symbolically, $r(st) = (rs)t$.)
- R-M-4:** 1 is a rational number. For each and every rational number, r , $r(1) = r$.

There is also a fifth multiplication law for the rational numbers, the discussion of which we defer to the next section.

- R-D:** If r , s , and t are rational numbers, then the product of r and $s + t$ is the same rational number as the sum of rs and rt . (Symbolically, $r(s + t) = rs + rt$.)

Any number-line interpretation of these laws involves an interpretation of the product, rs , of the rational numbers r and s . If r and s were both positive rational numbers, then one would be tempted to say that the point with coordinate " rs " is reached by measuring to the right of the origin r units of distance, s times. The difficulty with such phraseology is that the rational number s need not be a positive integer, so that the words " s times" may be meaningless. That is, while it is perfectly meaningful in English to speak of something as being done three times, or being done not at all, or even as being half-done, it is completely meaningless to speak of something that is done "half a time." Rather than strain unnecessarily for our purposes, we will omit giving number line interpretations for multiplication laws.

Exercises

1. Show that if p , q , m , and n are integers, and if neither q nor n is 0, then the sum $p/q + m/n$, as normally found by the "addition of fractions," actually is expressed as the quotient of two integers.
2. Discuss the number-line interpretations of the addition and multiplication laws for the rational numbers in those cases in which one or more of the rational numbers involved are the rational number 0.

Give number-line interpretations of the laws R-A-2, R-A-3 and R-A-5 for the specific cases in which:

3. $r = 1/2$, $s = 1/3$, $t = 1/4$.
4. $r = -1/2$, $s = -1/3$, $t = 1/4$.
5. $r = 1/2$, $s = -1/3$, $t = -1/4$.
6. $r = -1/2$, $s = -1/3$, $t = -1/4$.
7. Are the numbers 3 and 5 rational numbers? Why or why not?
8. Show how the product $(3)(5)$ can be interpreted geometrically in terms of the area of a rectangle made of 15 squares, each of which squares has sides of length 1.
9. Interpret the product $(1/3)(1/4)$ in terms of the area of a rectangle made up of an appropriate number of squares.
10. Interpret the product $(2/3)(3/4)$ in terms of the area of a rectangle made up of an appropriate number of squares.
11. How is the interpretation of exercise 7 changed if we consider the products $(-2/3)(3/4)$, $(2/3)(-3/4)$ and $(-2/3)(-3/4)$?

1.7. RECIPROCAL

In addition to the multiplication laws already given for the rational numbers, there is a further law which makes use of the concept of the reciprocal of a rational number. However, there is one unusual aspect of the idea of a reciprocal which must be investigated first, namely the special character of the rational number 0. We first state and prove two theorems.

Theorem 1.2:

If r is a rational number, and if $r + r = r$,
then r is the rational number 0.

Remark: This theorem may be stated less formally as: "0 is the only (rational) number which is its own double." In this form the truth of the theorem may seem glaringly obvious, but the reader should realize that the seeming obviousness comes not from the simplicity of the proof, but from the fact that this theorem restates something which was learned as a fact some time ago. Here we will concern ourselves not only with the facts, but with the reason for those facts' being facts.

Proof of 1.2:

1. r is a rational number by hypothesis.
2. $r + r = r$ by hypothesis.
3. $(r + r) + (-r) = r + (-r)$ by step 2, step 1, R-A-5, and R-A-1.
4. $r + [r + (-r)] = r + (-r)$ by step 3, R-A-3.
5. $r + 0 = 0$ by step 4, R-A-5.
6. $r = 0$ by step 5, R-A-4.

Theorem 1.3:

If s is any rational number, then $s(0) = 0$.

Proof:

1. s is a rational number by hypothesis.
2. $0 + 0 = 0$ by R-A-4.
3. $s(0 + 0) = s(0)$ by step 1, step 2, R-M-1.
4. $s(0) + s(0) = s(0)$ by step 3, R-D.
5. $s(0) = 0$ by step 4, Theorem 1.2.

Again, it should be noted that this theorem states the usually known fact that the product of 0 and any number is 0, while the proof of the theorem shows why this fact is a fact.

In order to discuss the reciprocal of a rational number, or of any number for that matter, we need a definition of the word:

Definition 1.6: If r is a rational number, then the rational number s is called a *reciprocal* of r if and only if $rs = 1$.

In actual fact, if a rational number does have a reciprocal, then that reciprocal is unique, so that we could have used the definite article in Definition 1.6, and we could have said that s is *the* reciprocal of r . Further, if s is the reciprocal of r then r is also the reciprocal of s , so that we can say that r and s are reciprocals of one another. Now, if every number were to have a reciprocal, then that reciprocal would have to be a number, since, in Definition 1.6, we really don't need the adjective "rational." Further, the arithmetic laws which we have for rational numbers are applicable to all (real) numbers, so that in the next theorem, we do not really need the word "rational" at any point.

Theorem 1.4:

There is no rational number which is a reciprocal of 0.

Proof:

1. 0 is a rational number by R-A-4.
2. Assume there is some specific rational number, t , which is a reciprocal of 0.
3. $t(0) = 1$ by step 2 and the definition (1.6) of reciprocal.

4. $t(0) = 0$ by step 2 and Theorem 1.3.
5. $0 = 1$ by steps 3 and 4.
6. Step 5 is a false statement, thus the assumption in step 2 is false.
7. There is no rational number which
is the reciprocal of 0 by steps 2 and 6.

Every rational number other than 0 does, in fact, have a reciprocal. More formally, we have a final algebraic law for the rational numbers:

R-M-5: For every rational number, r , which differs from 0, there is a unique rational number, $1/r$, which is the reciprocal of r .

Since we may define the quotient, r/s , the rational number r divided by the rational number s , as the product of r and $1/s$, we may always replace the phrase "divide by" with the phrase "multiply by the reciprocal of." Theorem 1.4 then gives us the first prohibition of algebra:

NEVER DIVIDE BY ZERO

This prohibition occurs not because mathematicians forbid division by 0, but because mathematics does. The proof of Theorem 1.4 shows that allowing division by 0 results in the truth of the statement " $0 = 1$," which statement would be interpreted in applied mathematics as "nothing = something."

Exercises

1. Explain in detail the reasons given in the proof of Theorems 1.2 and 1.3.
2. Prove the theorem: If r is a rational number other than 0, then $(1/r)r = 1$.
3. Prove the theorem: If the product, rs , of two rational numbers, r and s , is 0, then $r = 0$, or $s = 0$, or both. (Hint: Use the fact that $rs + rs = 0$, then R-D.)

Which of the following statements is true, and which false:

4. The negative of the sum of two rational numbers is the sum of their negatives.
5. The negative of the product of two rational numbers is the product of their negatives.
6. The negative of the reciprocal of a rational number is the reciprocal of its negative.
7. The sum of the negative of two rational numbers is the negative of their sum.
8. The product of the negatives of two rational numbers is the negative of their product.
9. The reciprocal of the product of two rational numbers is the product of their reciprocals.
10. The sum of the reciprocals of two rational numbers is the reciprocal of their sum.

11. The reciprocal of the reciprocal of a rational number is that rational number itself.
12. Why is the statement: "If x is any number whatsoever, then $x/x = 1$," a false statement?
13. Find the flaw in the following "reasoning":

If $a = b$ then $a^2 - b^2 = a^2 + b^2 - 2ab$,
 so that $(a - b)(a + b) = (a - b)(a - b)$
 and thus $a + b = a - b$
 which gives $b = -b$.
 Consequently, every number must be 0.
14. Comment on the statement: "The reciprocal of a small number is a large number, and the reciprocal of a large number is a small number."

1.8. ORDERING PROPERTIES OF THE RATIONAL NUMBERS

The ordering laws for the rational numbers are exactly like those for the integers:

- R-O-1:** If r and s are two rational numbers, then one and only one of the following three statements is true: $r = s$; $r < s$; $r > s$.
- R-O-2:** If r and s are rational numbers, then $r < s$ if and only if there is some *positive rational number*, t , such that $r + t = s$.
- R-O-3:** If r , s , and t are rational numbers, and if $r < s$, then $r + t < s + t$.
- R-O-4a:** If r and s are rational numbers, if $r < s$, and if t is a *positive rational number*, then $rt < st$.
- R-O-4b:** If r and s are rational numbers, if $r < s$, and if w is a *negative rational number*, then $rw > sw$.

The number-line interpretation of the first three of these ordering laws for rational numbers is precisely the same as the interpretation of the corresponding laws for the ordering of integers: I-O-1, I-O-2, and I-O-3.

A vitally important consequence of the ordering and algebraic laws for the rational numbers now occurs. This property points up the fact that, for rational numbers, the idea of "next" is meaningless. Given any *integer*, there is exactly one integer immediately larger and exactly one integer immediately smaller. But, given any rational number, there is *no next* rational number. More formally:

Theorem 1.5:

If r and s are two rational numbers, and if $r < s$, then there is at least one rational number, v , such that $r < v < s$ (that is, such that v is between r and s .)

Proof:

1. There is some positive rational number, t , such that $r + t = s$ by the hypothesis and R-O-2.
2. $0 < \frac{1}{2} < 1$ a known fact about specific numbers.
3. $0(t) < \frac{1}{2}(t) < (1)t$ by steps 1 and 2 and R-O-4a.
4. $0 < t/2 < t$ by step 3, Theorem 1.3, the definition of a quotient, and R-M-4.
5. $r < r + t/2 < s$ by step 4, R-A-3, and step 1.
6. There is at least one rational number, v , (namely $v = r + t/2$) such that $r < v < s$ by step 5.

Corollary to Theorem 1.5:

If r is a rational number, then there is no rational number which is nearest to r .

Proof:

1. Assume the rational number, s , is nearest to r , but $s \neq r$.
2. Either $r < s$ or $s < r$ by step 1, R-O-1.
3. Whether $r < s$ or $s < r$, there is a rational number, v , between r and s by Theorem 1.5.
4. s is not nearest to r by step 3.
5. Step 4 contradicts the assumption in step 1, thus there is no rational number nearest to r .

As a special case of the corollary, we have:

There is no number next to 0.

Exercises

1. Find the " t " mentioned in R-O-2 for
 - a) $r = 1/2, s = 7/5$.
 - b) $r = -7/4, s = 1/5$.
 - c) $r = -11/3, s = -4/7$.
2. Sketch a number-line illustration of R-O-3 for
 - a) $r = 1/5, s = 8/3, t = 7$.
 - b) $t = 1/5, s = 8/3, r = -5$.
 - c) $r = -1, s = 5/3, t = 1$.
 - d) $r = -1, s = 5/3, t = -3$.
 - e) $r = -3, s = -1/2, t = 2$.
 - f) $r = -3, s = -1/2, t = -1/2$.
3. Examine R-O-4a for the specific numbers
 - a) $r = 1/4, s = 3/2, t = 3$.
 - b) $r = -3/2, s = 1/2, t = 6$.
 - c) $r = -4, s = -1, t = 1/2$.
4. Examine R-O-4b for the specific numbers
 - a) $r = 7/3, s = 5, w = -1/2$.
 - b) $r = -3/7, s = 1, w = -1/3$.
 - c) $r = -11/2, s = -4/3, w = -6$.
5. Prove that there is no rational number nearest to $3781/1002$.
6. Explain in detail the reasons for steps 3, 4, and 5 of the proof of Theorem 1.5.

Consider the following sets of rational numbers. If any of these sets contains within it a largest rational number, give that number. If any of these sets contains within it a smallest rational number, give that number.

7. The set of all rational numbers.
8. The set of all integers.
9. The set of all natural numbers.
10. The set of all positive rational numbers.
11. The set of all non-negative rational numbers.
12. The set of all rational numbers which are neither less than 2 nor greater than 3.
13. The set of all rational numbers which are either greater than 2 or not greater than 3.
14. The set of all rational numbers which are either not less than 2 or less than 3.
15. The set of all rational numbers which are between 2 and 3.
16. The set of all rational numbers whose squares are less than or equal to 1.
17. The set of all rational numbers whose squares are less than or equal to 2.

1.9. SOME FACTS ABOUT THE NATURAL NUMBERS

It might be hoped that the set of all rational numbers is the same as the set of all numbers. Such, however, is not the case. In order to prove that this fact is a fact, it is sufficient that we find a number which is not a rational number. Before doing this, we will need certain facts about the natural numbers, which we will give in this section.

Definition 1.7: If n and p are natural numbers, then n is *divisible by p* means that there exists a natural number, k , such that $n = pk$.

Special cases of divisibility occur in the use of words such as "even" and "odd." An even natural number is defined to be one which is divisible by 2.

Definition 1.8: A natural number, p , other than 1, is called a *prime* if and only if p is divisible only by 1 and p .

From this definition we see that the first few primes are: 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, etc.

We will use the notation n^2 to indicate the product of the number n with itself. With this manner of writing, we have:

Theorem 1.6:

If n is a natural number, and if n^2 is divisible by the prime number p , then n is divisible by the prime number p .

The proof of this theorem will be omitted, since its length tends to get us off our main track of looking for a number which is not a rational number.

The final theorems of this section will be stated without proof, again because the length of their proofs tends to distract us from our main goal.

Theorem 1.7:

If n is any natural number, then n may be expressed as the product of prime natural numbers in one and only one way, apart from rearrangements of these prime natural numbers.

For example, the natural number 60 can be expressed as the product of two 2's, one 3, and one 5, and can be expressed as the product of no other primes than these.

Theorem 1.8:

If m and n are natural numbers, then there are natural numbers, p and q , which have no common divisors, such that m/n and p/q are the same (positive) rational number.

This theorem merely formalizes the process known as "reducing a fraction to its lowest terms."

1.10. THE REAL NUMBERS

We will now prove that the rational numbers do not constitute the set of all numbers, as we understand that set to be the set of all possible coordinates of points on the number line. In order to achieve this purpose, it is sufficient that we find one number which is not a rational number, that is, that we find one point whose distance from the origin cannot be given as the quotient of two natural numbers.

The Pythagorean theorem of elementary geometry tells us that the hypotenuse of an isosceles right triangle whose sides have length 1 will have length $\sqrt{2}$. (It should be noted here and forevermore that the symbol $\sqrt{}$ is read "the square root of," and, because of the use of the definite article, means "the positive number whose square is.")

Theorem 1.9:

$\sqrt{2}$ is not a rational number.

Proof:

1. Assume $\sqrt{2}$ is a rational number
2. $\sqrt{2}$ is a positive rational number by step 1 and the definition of "the square root."
3. There are natural numbers, m and n , such that $\sqrt{2} = m/n$ by step 2 and Definition 1.4.

4. There are natural numbers, p and q , which have no common divisor, such that $\sqrt{2} = p/q$ by step 3 and Theorem 1.8.
5. $q\sqrt{2} = p$ by step 4 and various R-M laws.
6. $2q^2 = p^2$ by step 5 and various R-M laws.
7. p^2 is an even natural number by step 6 and the definition of even.
8. p is an even natural number by step 7 and Theorem 1.6.
9. There is some natural number, k , such that $p = 2k$ by step 8 and the definition of even.
10. $p^2 = 4k^2$ by step 9 and various R-M laws.
11. $2q^2 = 4k^2$ by steps 6 and 10.
12. $q^2 = 2k^2$ by step 11 and various R-M laws.
13. q^2 is an even natural number by step 12 and the definition of even.
14. q is an even natural number by step 13 and Theorem 1.6.
15. p and q have the common divisor 2 by steps 8 and 14.
16. Step 15 contradicts step 4, so that the assumption in step 1 is false. Thus $\sqrt{2}$ is not a rational number.

We have now found at least one point on the number line which does not have a rational number coordinate, namely that point to the right of the origin whose distance from the origin is $\sqrt{2}$ units of distance. There are, of course, many non-rational numbers, to which we give the name "irrational."

Definition 1.9: A number is *irrational* if that number is not a rational number.

The set of all (real) numbers—that set which can be put into one-to-one correspondence with *all* the points on the number line—is called the set of all real numbers. We will usually drop the adjective "real," and simply refer, as we have been doing, to the set of all numbers. Fortunately, this new, expanded set does not have new algebraic or ordering properties, so that we can adopt the laws used for the rational numbers, merely by omitting the word "rational." In summary, these laws are:

A-1: If x and y are numbers, then $x + y$ exists uniquely.

A-2: If x and y are numbers, then $x + y = y + x$.

A-3: If x , y , and z are numbers, then

$$x + (y + z) = (x + y) + z.$$

A-4: 0 is a number. For every number, x , $x + 0 = x$.

A-5: For every number, x , there is a unique number, $-x$, such that $x + (-x) = 0$.

M-1: If x and y are numbers, then xy exists uniquely.

M-2: If x and y are numbers, then $xy = yx$.

M-3: If x , y , and z are numbers, then $x(yz) = (xy)z$.

- M-4:** 1 is a number. For every number, x , $x(1) = x$.
- M-5:** For every non-zero number, x , there is a unique number, $1/x$, such that $x(1/x) = 1$.
- D:** If x , y , and z are numbers, then $x(y + z) = xy + xz$.
- O-1:** If x and y are numbers, then one and only one of the following three statements is true:
 $x = y$; $x < y$; $x > y$.
- O-2:** If x and y are numbers, then $x < y$ if and only if there is some positive number, p , such that $x + p = y$.
- O-3:** If x , y , and z are numbers, and if $x < y$, then $x + z < y + z$.
- O-4a:** If x and y are numbers, if $x < y$, and if z is a positive number, then $xz < yz$.
- O-4b:** If x and y are numbers, if $x < y$, and if w is a negative number, then $xw > yw$.

From the ideas given in the fifth addition and multiplication laws, we define subtraction by the statement

$$x - y = x + (-y)$$

and we define division by the statement

$$x/y = x(1/y) \quad (\text{where } y \neq 0).$$

From these basic laws we can deduce many theorems. We note two of particular importance:

If x is any number, then $x(0) = 0$.

If $xy = 0$ then $x = 0$, or $y = 0$, or both.

The addition, multiplication, and distributive laws are those used in the algebraic manipulations involved with equations, while the ordering laws are those used in the algebraic manipulations involved with inequalities. Most of the manipulations performed involve very rapid mental use of these basic laws—so rapid, in fact, that one is usually unconscious of the large number of steps involved. In order to aid us in learning not merely the facts, but why those facts are indeed facts, we will consider the algebraic formula:

$$(x + y)^2 = x^2 + 2xy + y^2.$$

In words, this formula becomes the statement: "the square of the sum of two numbers is the sum of their squares and twice their product."

To convince oneself of the truth of the formula, using the basic laws unconsciously, one usually has:

$$\begin{aligned} (x + y)^2 &= (x + y)(x + y) \\ &= x^2 + xy + xy + y^2 \\ &= x^2 + 2xy + y^2, \end{aligned}$$

that is, three steps.

We will now develop this same formula, specifically citing each basic law used, and using only one of them at a time.

- | | | |
|-----|---|--|
| 1. | $(x + y)^2 = (x + y)(x + y)$ | by the definition of the square of a number; |
| 2. | $= (x + y)(x) + (x + y)(y)$ | by D; |
| 3. | $= x(x + y) + y(x + y)$ | by M-2; |
| 4. | $= (xx + xy) + (yx + yy)$ | by D; |
| 5. | $= (x^2 + xy) + (yx + y^2)$ | by the definition of the square of a number; |
| 6. | $= (x^2 + xy) + (xy + y^2)$ | by M-2; |
| 7. | $= [(x^2 + xy) + xy] + y^2$ | by A-3; |
| 8. | $= [x^2 + (xy + xy)] + y^2$ | by A-3; |
| 9. | $= [x^2 + \{(xy)(1) + (xy)(1)\}] + y^2$ | by M-4; |
| 10. | $= [x^2 + (xy)(1 + 1)] + y^2$ | by D; |
| 11. | $= [x^2 + (xy)(2)] + y^2$ | by the fact that $1 + 1 = 2$; |
| 12. | $= [x^2 + 2(xy)] + y^2$ | by M-2; |
| 13. | $= [x^2 + 2xy] + y^2$ | by M-3; |
| 14. | $= x^2 + 2xy + y^2$ | by A-3. |

This gives us 14 separate applications of either basic algebraic laws or of elementary definitions.

Exercises

- Which R-M laws are used in steps 6, 10, and 12 of the proof of Theorem 1.9?
- Prove that $\sqrt{3}$ is not a rational number.
- Show that if a and b are rational numbers, $b \neq 0$, then every number of the form $a + b\sqrt{2}$ is an irrational number.
- Derive the formula

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

in a step-by-step manner, using only the basic algebraic laws. State this formula in words.

- Prove the theorem: If x and y are numbers, then there is at least one number between x and y .
- Derive the following formulas, in a condensed manner, showing, for each step, the algebraic laws which have been used to derive that step from its predecessor.
 - $x^3 - y^3 = (x - y)(x^2 + xy + y^2)$.
 - $x^3 + y^3 = (x + y)(x^2 - xy + y^2)$.
- State in words the facts about numbers given in the formulas of exercise 6.

1.11. LINEAR COORDINATE SYSTEMS

In this section we give the formal mathematical name to the one-to-one correspondence which we have set up between the points of the number line and the (real) numbers.

Definition 1.10: A *linear coordinate system* is a one-to-one correspondence between the points of a straight line and the (real) numbers, such that the distance between two points with coordinates “ x ” and “ y ” is whichever of the two numbers $x - y$ and $y - x$ is not negative.

In Chapter 3 we will present a plane coordinate system, which will be a method of labeling points in the plane.

EQUATIONS AND INEQUALITIES

2.1. VARIABLES AND CONSTANTS

There are many occasions in which it is necessary to talk about some specific set of numbers. When we want to make a statement which will apply to each one of the numbers in that specific set, we use, in speech, the word “each.” For example, we might make the true statement, “Each positive number is greater than -4 ,” or the false statement, “Each positive number is less than -4 .” In these instances, the specific set of numbers under consideration is the set of all positive numbers. Despite the falseness of the statement “Each positive number is less than -4 ,” we are still talking *about* the set of all positive numbers. In order to arrive at a symbolism which will replace the use of words, we introduce:

Definition 2.1: Given a specific set of numbers, a *variable* is a symbol used to represent each of the numbers in that specific set. The specific set of numbers is called the *range* of that variable.

It should be noted that we are *not* using the ordinary English meaning of the word “variable” as an “entity which varies.” We will usually use letters in the latter part of the alphabet as variables, with prejudice for the use of x and/or y . It must be noted that a variable is not present unless we know both the symbol and the specific set of numbers to which that symbol applies. Thus a variable representing each negative number would be written as

$x, x \text{ negative}$

or as

$y, y \text{ negative,}$

or with any other symbolism which gives both the variable and the range of the variable.

Of all possible sets of numbers, the largest is certainly the set of all numbers. If we deal with a variable whose range is the set of all numbers, then it

will be convenient to adopt a convention which will identify the range of such a variable, without having to write a description of this range at all times.

Convention: If a variable appears alone, it will be assumed to have as range the set of all numbers. If a variable appears in an algebraic or other expression, the range of that variable will be assumed to be the set of all numbers for which the given expression will be a number. In all other cases the range of the variable will be specifically stated.

Under this convention, the statement $x < 1$ is a statement about the set of all numbers. The variable x has, as range, the set of all numbers, and the statement $x < 1$ is true if x is any number which is less than 1, and is false if x is any number which is not less than 1. (It should be noted that the truth or falseness of a statement has no bearing on that statement's being a statement!)

The statement $3/x = 2$ is a statement about the set of all numbers other than 0. The variable x appears in the algebraic expression $3/x$. Since this algebraic expression will not be a number if x is the number 0, the range of the variable x , in this case, cannot include the number 0. The statement $3/x = 2$ is true if x is the rational number $3/2$, and is false if x is a number other than $3/2$.

The statement, " $\sqrt{x-4}$ is an algebraic expression," is a statement involving the variable x . The convention which we have adopted says that the range of the variable x is the set of all numbers which are not less than 4. (Since no negative number has a (real) square root, the expression $\sqrt{x-4}$ is a number only if x is a number for which $x-4$ is not negative.)

If, rather than discuss numbers in a given set of numbers, we want to talk about one particular number, but do not know, or care, which number that one is, we use a symbol to represent that number alone. Formally,

Definition 2.2: A *constant* is a symbol used to represent one specific number.

Speaking loosely, a variable is a symbol which represents many numbers, while a constant is a symbol for just one number. Given a variable, a number in the range of that variable is often called a *value* of that variable. Using this terminology, one may say that -3 is a value of the variable x for which the statement $x < 1$ is true, and that 7 is a value of the variable x for which the statement $x < 1$ is false.

It is necessary to distinguish between a value of a variable which makes a statement containing that variable a false statement, and a number which cannot be a value of that variable. For example, 2 is a value of the variable x which makes the statement

$$\frac{5}{x-1} = 10$$

false, while 1 is a number which cannot be a value of the variable, x , in this statement.

☞ In elementary algebra one is concerned with equations which involve one variable. To solve such an equation is to find a number, or numbers, which is, or are, values of the given variable making that equation a true statement. If we have the equation $3x + 5 = 17$, the convention we have adopted tells us that the range of the variable x is the set of all numbers. The number 4 is a solution of this equation—that is, is a value of the variable x for which the equation is a true statement—and every number other than 4 is a value of the variable x for which the equation is a false statement.

Also considered in elementary algebra are “algebraic expressions involving one variable.” Such an expression is, in itself, a symbol for a number, which number results from performing certain indicated algebraic operations. For example, $3x + 5$ is an algebraic expression involving the variable x . According to the convention we have adopted, the range of this variable is the set of all numbers, and the expression $3x + 5$ is a symbol for the number which results from adding 5 to the result of multiplying x by 3. As another example, $\sqrt{x} - 2$ is an algebraic expression involving the variable x . Because of the definition of the square root symbol, the range of the variable x is the set of all numbers not less than 2, since only these numbers make the number $x - 2$ non-negative. The expression $\sqrt{x} - 2$ is then a symbol for the number which results from taking the square root of the result of subtracting 2 from x .

We may then ascertain the range of the variable in a given situation by examining the two operations: division, and taking the square root—which are not always possible. We have already noted that the first prohibition of algebra is: never divide by zero. For the algebra of real numbers, the second is:

NEVER TAKE *THE* SQUARE ROOT OF A NEGATIVE NUMBER

It should be noted that we are using the word “number” as arising from the idea of all possible distances measured on a straight line. That is, the only numbers to be considered in this book are “real” numbers. If one introduces the concept of “imaginary” and “complex” numbers, then it is possible to find a square ~~root~~ of a negative number, but not *the* square root. With our exclusive use of “real” numbers, no negative number will have even a square root.

Further, since all even powers involve squaring, the admonition given should be extended to a prohibition of taking any even root of a negative number.

Exercises

Determine the range of the variable in each of the following expressions or statements. In the case of statements, determine which values of the variable make the statement true.

- | | |
|------------------------|-------------------------|
| 1. x . | 11. $4/(y - 3)$. |
| 2. $2x + 1$. | 12. $(z + 1)/(z + 2)$. |
| 3. $\sqrt{x - 4}$. | 13. $r + 1 < 3$. |
| 4. $\sqrt{r - 4}$. | 14. $s - 2 > 5$. |
| 5. $\sqrt{s - 5}$. | 15. $3t < 9$. |
| 6. $\sqrt{t - (-1)}$. | 16. $3u + 1 < 10$. |
| 7. $\sqrt{u + 3}$. | 17. $-2 < v < 2$. |
| 8. $\sqrt{-v}$. | 18. $1/w < 5$. |
| 9. $\sqrt{1 - w}$. | 19. $x^2 < 4$. |
| 10. $2/x$. | |

2.2. EQUATIONS

We have used in the discussion of the basic laws for the (real) number system, various ideas about equality as such. These ideas are:

- 1) Every number is equal to itself.
(For all a , $a = a$.)
- 2) If the number a is equal to the number b , then the number b is equal to the number a .
(For all a and b , if $a = b$, then $b = a$.)
- 3) If the number a is equal to the number b , and if the number b is equal to the number c , then the number a is equal to the number c .
(For all a , b , and c , if $a = b$ and $b = c$, then $a = c$.)

This third law governing the relation of equality is often expressed by saying "things equal to the same thing are equal to each other." All three of these laws apply to any things whatever, and are not restricted to numbers alone. We do have, as a result of the laws A-1 and M-1, two further basic properties of the relation of equality, as that relation concerns numbers:

- 4) If a , b , and c are numbers, and if $a = b$, then $a + c = b + c$.
- 5) If a , b , and c are numbers, and if $a = b$, then $ac = bc$.

In this section we will consider the solution of equations involving one variable. We have already defined the word "variable," but it is useful to emphasize the meaning of the word "equation" by giving a formal definition.

Definition 2.3: An *equation* is a statement saying that two things are the same thing.

It is noted here that there is a difference between our making a statement about an equation and our making a statement which is an equation. For example, $2 + 3 = 5$ and $7 - 1 = 6$ are both statements which are equations, while "The solution of ' $x + 2 = 1$ ' is the set consisting of the number -1 " is a statement about the equation " $x + 2 = 1$." We could indicate (as shown by these examples) the distinction by putting quotation marks when we are talking about an equation and omitting quotation marks when we are making a statement which is an equation. However, in most cases, the distinction is made very clear by the text immediately preceding each equation. Consequently, it is felt that no confusion will result from omitting the quotation marks, except in those cases where the distinction needs special emphasis.

The usually ignored aspect of this definition is that the words "true," "false," and "number" are not included. Thus the statements: $1 = 1$; $3 = 5$; and $\text{pigs} = \text{pigs}$; are all equations, of which the first and third are true statements, the second is a false statement, and the third is a statement which does not involve numbers. Of course we will concern ourselves particularly with equations which are statements about mathematical entities, particularly statements about numbers. In this chapter, our major concern will be finding solutions of equations. We have:

Definition 2.4: Given an equation involving one variable, a *solution* of that equation is a value of that variable for which the given equation is a true statement. *The solution* of that equation is the set of all values of that variable for which the given equation is a true statement.

As examples: The equation

$$2x + 3 = 9$$

has, as a solution, the number 3. This equation has no solution other than 3, so that the solution of the equation $2x + 3 = 9$ is the set consisting of the single number 3.

The equation $x^2 = 4$ has, as a solution, the number -2 , and also has, as a solution, the number 2. The solution of the equation $x^2 = 4$ is the set consisting of the two numbers: -2 and 2.

The equation $3/x = 0$ has no solution, since no value of the variable x (whose range here is the set of all numbers other than 0) makes this equation a true statement.

The equation $x + 3x = 4x$ has every number as a solution, so that the solution of the equation $x + 3x = 4x$ is the set of all numbers.

The equation $x^2/x = x$ has every non-zero number as a solution, so that the solution of this equation is the set of all numbers other than 0.

The equation $\sqrt{x^2} = x$ has every non-negative number as a solution, but no negative number can be a solution to this equation. The solution of this equation is the set of all non-negative numbers.

In considering how to find the solution of an equation we start with the simplest kind of equation, and then proceed to more complicated kinds.

Our purpose in this chapter is not particularly to develop a technique for solving "first degree" or "second degree" equations—we assume the reader knows these techniques—but it is to put the techniques on a solid foundation by proving theorems which show why these techniques work. We will examine here only two elementary kinds of equations, those which are called "first degree" and those which are called "second degree" or "quadratic."

Definition 2.5: A *first degree equation* involving the variable x is an equation of the form

$$ax + b = c,$$

where a , b , and c are constants, and $a \neq 0$.

Theorem 2.1:

If a , b , and c are constants, and $a \neq 0$, then the solution of $ax + b = c$ is the set consisting of the number $\frac{c - b}{a}$.

Remark: It is necessary that we prove both that the number $\frac{c - b}{a}$ is a solution of $ax + b = c$, and also that no other number is.

Proof of 2.1:

1. The range of the variable x , as contained in the equation

$$ax + b = c,$$

is the set of all numbers

by convention.

2. $a \neq 0$

by hypothesis.

3. $\frac{c - b}{a}$ is a number

by step 2 and algebraic laws.

4. $\frac{c - b}{a}$ is a value of the variable x

by steps 1 and 3.

5. $\frac{c - b}{a}$ is a value of the variable x

which makes the equation $ax + b = c$ into the statement

$$a \frac{c - b}{a} + b = c$$

by step 4.

6. $a \frac{c - b}{a} + b = c$ is true if and only if $c = c$ is true

by various algebraic laws.

7. $c = c$ is true

by the definition of equation.

8. $\frac{c - b}{a}$ is a value of the variable x

which makes the equation $ax + b = c$ a true statement

by steps 5, 6, and 7.

9. $\frac{c-b}{a}$ is a solution of the equation $ax + b = c$ by step 8 and the definition (2.4) of a solution of an equation.
10. Assume that s is a solution of $ax + b = c$.
11. $as + b = c$ by step 10 and the definition (2.4) of a solution.
12. $as = c - b$ by step 11 and algebraic laws.
13. $s = \frac{c-b}{a}$ by steps 2 and 12 and algebraic laws
14. $\frac{c-b}{a}$ is the only number which is a solution of the equation $ax + b = c$ by steps 9, 10, and 14.

It should be noted that the usual algebraic procedure used to *find* the solution of the equation $ax + b = c$ appears in the proof of Theorem 2.1 only to show the uniqueness of the one number which is a solution, and that the procedure usually known as “verifying a suspected solution” is the actual proof that the number $\frac{c-b}{a}$ is a solution of $ax + b = c$.

It will be assumed that the reader is already completely familiar with the techniques of solving such first degree equations involving one variable, so that any specific example of such an equation can be solved mentally without resorting to writing down each and every step. For example, it should be immediately apparent that the only number which is a solution of the equation $3x - 5 = 7$ is the number 4.

Definition 2.6: A *second degree* (or *quadratic*) *equation* involving the variable x is an equation of the form

$$ax^2 + bx + c = 0,$$

where a , b , and c are constants, and $a \neq 0$.

Theorem 2.2:

Consider the equation

$$ax^2 + bx + c = 0$$

where a , b , and c are constants, and $a \neq 0$;

- (i) if $b^2 - 4ac < 0$, this equation has no real numbers as solutions.
- (ii) if $b^2 - 4ac = 0$, the number $-b/2a$ is the only solution of this equation.
- (iii) if $b^2 - 4ac > 0$, the numbers

$$\frac{-b - \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

are the only solutions of this equation.

Remark: The symbol $\sqrt{\quad}$ always means “the *positive* number which, when squared, gives,” so that $\sqrt{4}$ is the number 2, and is not the number -2 .

Proof of 2.2:

1. $a \neq 0$ by hypothesis.
2. $ax^2 + bx + c = 0$ is true if and only if $x^2 + \frac{b}{a}x + \frac{c}{a} = 0$ is true by step 1, multiplication of each side of the first equation by $\frac{1}{a}$, and algebraic laws.
3. $ax^2 + bx + c = 0$ is true if and only if $x^2 + \frac{b}{a}x = -\frac{c}{a}$ is true by step 2, addition of $-\frac{c}{a}$ to each side of the second equation in step 2, and algebraic laws.
4. $ax^2 + bx + c = 0$ is true if and only if $x^2 + 2\frac{b}{2a}x = -\frac{c}{a}$ is true by step 3 and the fact that $\frac{b}{a} = 2\frac{b}{2a}$.
5. $ax^2 + bx + c = 0$ is true if and only if $x^2 + 2\frac{b}{2a}x + \frac{b^2}{4a^2} = \frac{b^2}{4a^2} - \frac{c}{a}$ is true by step 4 and the addition of $b^2/4a^2$ to each side of the second equation in step 4.
6. $ax^2 + bx + c = 0$ is true if and only if $\left[x + \frac{b}{2a}\right]^2 = \frac{b^2 - 4ac}{4a^2}$ is true by step 5, the fact that $(p + q)^2 = p^2 + 2pq + q^2$, and the addition of the “fractions” on the right-hand side of the second equation in step 5.
7. $4a^2 > 0$ by step 1 and the fact that the square of a non-zero number must be positive.
8. If $b^2 - 4ac < 0$, then $ax^2 + bx + c = 0$ has no solutions by step 6 and the fact that the square of a number is never negative.
9. If $b^2 - 4ac = 0$, then $ax^2 + bx + c = 0$ is true if and only if x is the number $-b/2a$ by step 6 and the fact that 0 is the only number whose square is 0.

10. If $b^2 - 4ac > 0$, then $ax^2 + bx + c = 0$ is true if and only if

$$\left(x + \frac{b}{2a}\right)^2 - \left(\frac{\sqrt{b^2 - 4ac}}{2a}\right)^2 = 0 \text{ by step 6 and the fact that}$$

$$\left(\frac{\sqrt{b^2 - 4ac}}{2a}\right)^2 \text{ and } \frac{b^2 - 4ac}{4a^2}$$

are the same number.

11. If $b^2 - 4ac > 0$ then
 $ax^2 + bx + c = 0$
 is true if and only if

$$\left(x + \frac{b}{2a} + \frac{\sqrt{b^2 - 4ac}}{2a}\right)\left(x + \frac{b}{2a} - \frac{\sqrt{b^2 - 4ac}}{2a}\right) = 0$$

by step 10 and the fact that

$$(p + q)(p - q) = p^2 - q^2.$$

12. If $b^2 - 4ac > 0$ then

$$ax^2 + bx + c = 0$$

is true if and only if

$$x \text{ is the number } \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

or x is the number

$$\frac{-b - \sqrt{b^2 - 4ac}}{2a}$$

by step 11 and the fact that the product of two numbers is 0 if and only if one or the other of these numbers is 0.

We will use these theorems in the next section to provide clues for the solution of inequalities involving one variable.

In the process of finding solutions of an equation involving one variable, we are continually changing the given equation to a new one. The new one is related to the given one by having the same solution. We say that the equations are *equivalent* provided that the set of solutions of one is the same as the set of solutions of the other. Under this terminology the following equations are all equivalent:

$$3x^2 + 2x + 5 = 2x^2 + x + 11;$$

$$x^2 + x - 6 = 0;$$

$$(x - 2)(x + 3) = 0.$$

Exercises

In the following sentences, first find the verb, then determine whether or not each sentence is an equation, and finally determine, in the case of an equation, whether or not the equation is a true statement:

1. $1 + 1 = 2$.
2. " $1 + 1 = 2$ " is a true statement.
3. $2 + 3 = -4$.
4. " $2 + 3 = -4$ " is the same equation as " $1 + 4 = -2$."
5. Identical twins are the same person.

Write an equation involving the variable t whose solution is the set consisting of:

6. the number 4.
7. the numbers $\sqrt{3}$ and $-\sqrt{3}$.
8. the number $\sqrt{5}$.
9. all numbers other than 0.
10. all numbers other than 1.
11. all numbers other than -1 and 3 .
12. all numbers.
13. no numbers.
14. Exactly which algebraic laws are involved in steps 3, 6, 12, and 13 of the proof of Theorem 2.1?
15. Find the solution of
 - a) $3x + 4 = 5$.
 - b) $-2x + 1 = 4$.
 - c) $\frac{4}{3}x - \frac{1}{2} = -\frac{2}{5}$.
 - b) $\sqrt{5}x + 1 = \sqrt{3}$.
16. Exactly which algebraic laws are involved in steps 2, 3, and 10 of the proof of Theorem 2.2.

Use Theorem 2.2 to find the solution of:

17. $3x^2 - 6x + 3 = 0$.
18. $3x^2 - 6x + 2 = 0$.
19. $3x^2 - 6x + 4 = 0$.
20. $\frac{2}{3}x^2 + \frac{4}{5}x - \frac{6}{7} = 0$.
21. $\frac{1}{3}x^2 - \frac{2}{7}x + \frac{1}{2} = 0$.
22. $-\sqrt{5}x^2 + 2x + \sqrt{3} = 0$.
23. $\sqrt{3}x^2 + x + \sqrt{7} = 0$.

2.3. INEQUALITIES

In this section we will examine methods of solving inequalities involving one variable. As with equations, it is useful to have a formal definition:

Definition 2.7: An *inequality* is a statement which says that one number is less than (or greater than) another number.

There is nothing in this definition which mentions the words "true" and "false," but the definition does specifically require a statement about numbers. Thus $1 < 1$, $3 < 5$, and $5 < 3$ are all inequalities, of which the first and third are false statements, and the second is a true statement, but "dogs are less than cats" and "a boy is less than a man" are not inequalities. We will be concerned with inequalities which involve one variable. Specifically,

Definition 2.8: Given an inequality involving one variable, a *solution* of that inequality is a value of that variable for which the given inequality is a true statement. The *solution* of that inequality is the set of all values of the variable for which the given inequality is a true statement.

As an example, the inequality $3x < 6$, involving the variable x has, as a solution, the number 0. Each of the numbers -5 , $1/2$, and $9/5$ is also a solution of this inequality. Since every number which is less than 2 is a solution of the inequality $3x < 6$, the solution of this inequality is the set of all numbers which are less than 2.

In order to describe conveniently the sets of numbers which will be met in finding the solution of an inequality, we have:

Definition 7.9: An *interval* is one of three kinds of sets of numbers:

- (i) the set of all numbers between two fixed numbers;
- or (ii) the set of all numbers larger than a fixed number;
- or (iii) the set of all numbers less than a fixed number.

Each interval may or may not include the fixed number(s), which fixed number(s) is (are) called the *end point(s)* of the interval.

In this manner of speaking, the solution of the inequality $3x < 6$ is the interval consisting of all numbers less than 2. The number 2 is the right-hand end point of this interval, and this interval has no left-hand end point.

If we start with the two fixed numbers, -1 and 4 , there are four intervals which have these numbers as left and right-hand end points, respectively:

- (i) the set of all numbers between -1 and 4 , excluding -1 and excluding 4 ;
- (ii) the set of all numbers between -1 and 4 , including -1 and excluding 4 ;

- (iii) the set of all numbers between -1 and 4 , excluding -1 and including 4 ;
- (iv) the set of all numbers between -1 and 4 , including -1 and including 4 .

Rather than use so many words to describe these intervals, we will use a variable whose range is specifically given. (See Definition 2.1 and the convention adopted in Section 1 of this chapter.) Then the four intervals above are described by:

- (i) The set of all x , $-1 < x < 4$;
- (ii) The set of all x , $-1 \leq x < 4$;
- (iii) The set of all x , $-1 < x \leq 4$;
- (iv) The set of all x , $-1 \leq x \leq 4$.

(The symbol " $\leq$ " is read, and means, "is less than or is equal to." The symbol " $\geq$ " is read, and means, "is greater than or equal to.") In each of these four intervals, -1 is the left-hand end point and 4 is the right-hand end point.

According to Definition 2.8, the set of all numbers which are greater than 10 is an interval. The number 10 is the left-hand end point of this interval, and the interval is described by:

the set of all x , $x > 10$.

The interval consisting of the set of all numbers which are not less than 10 is described by:

the set of all x , $x \geq 10$.

Also, the set of all numbers less than -1 is an interval whose right-hand end point is the number -1 . This interval is described by:

the set of all x , $x < -1$.

The interval consisting of the set of all numbers which are not greater than -1 is described by:

the set of all x , $x \leq -1$.

On a linear coordinate system, an interval of numbers corresponds to an interval of points, where the first use of "interval" is in the sense of the definition (2.8), and the second use is in the usual geometric meaning. It is thus particularly convenient to refer to a linear coordinate system when finding the solution of an inequality. We will treat the problem of finding the solution of an inequality involving one variable in a manner similar to that used in the preceding section on equations. That is, we will first consider simple inequalities, and then more complicated ones as they arise.

By application of the ordering law, O-3, for the set of all numbers, any inequality is equivalent to a statement that a number is positive, or to a statement that a number is negative. For example, the inequality $A < B$

is true if and only if the inequality $A - B < 0$ or the inequality $B - A > 0$ is true. Thus if A and/or B are expressions involving one variable (the same one in both A and B), the solution of the inequality $A < B$ will also be the solution of the inequalities $A - B < 0$ and $B - A > 0$. As a result of this fact, it is sufficient to consider only inequalities which are statements that a number is positive, or statements that a given number is negative.

To begin with, we have

Definition 2.10: A *first degree inequality* involving the variable x is an inequality of the form

$$ax + b < 0$$

or of the form

$$ax + b > 0,$$

where a and b are constants, and $a \neq 0$.

The solution of a first degree inequality is given in:

Theorem 2.3:

For the first degree inequality involving the variable x :

$$ax + b < 0,$$

where a and b are constants and $a \neq 0$,

(i) if $a > 0$, then the solution of $ax + b < 0$ is the interval

the set of all x , $x < -b/a$;

(ii) if $a < 0$, then the solution of $ax + b < 0$ is the interval:

the set of all x , $x > -b/a$.

Proof:

1. $ax + b < 0$ is true if and only if
 $ax < -b$ is true by O-3 and algebraic laws.
2. If $a > 0$, then $ax + b < 0$ is true
if and only if $x < -b/a$ by step 1, O-4a, and algebraic laws.
3. If $a < 0$, then $ax + b < 0$ is true
if and only if $x > -b/a$ by step 1, O-4b, and algebraic laws.
4. If $a > 0$, then the solution of
 $ax + b < 0$ is the interval:
the set of all x , $x < -b/a$ by step 2 and the definition (2.8).
5. If $a < 0$, then the solution of
 $ax + b < 0$ is the interval:
the set of all x , $x > -b/a$ by step 3 and the definition (2.8).

EXAMPLE:

Find the solution of

$$3x - 4 < 1.$$

 $3x - 4 < 1$ is true if and only if

$3x < 5$ is true

by O-3.

 $3x - 4 < 1$ is true if and only if

$x < 5/3$ is true

by O-4a.

EXAMPLE:

Find the solution of

$$-2x + 3 < -1.$$

 $-2x + 3 < -1$ is true if and only

if $-2x < -4$ is true

by O-3.

 $-2x + 3 < -1$ is true if and only

if $x > 2$ is true

by O-4b.

Theorem 2.3, and the two examples shown, use the ordering laws O-4a and O-4b. It should be noted very definitely that care must be exercised in any manipulations involving multiplication. We must be certain that the laws O-4a and O-4b do apply, and we must know which one of these two laws does apply.

Exercises

1. What is meant by saying that two inequalities are equivalent?

Find the solution of the following inequalities:

2. $2x - 3 > 5.$

12. $-2x - 1 > -3.$

3. $4x + 1 > 6.$

13. $-5x + 2 > -4.$

4. $3x - 1 > -4.$

14. $-3x - 4 < 1.$

5. $2x - 4 > -1.$

15. $-2x + 5 < 2.$

6. $5x - 1 < 3.$

16. $-x - 4 < -3.$

7. $4x + 1 < 2.$

17. $-5x + 2 < -4.$

8. $3x - 3 < -2.$

18. $3x + 5 < 2x - 1.$

9. $2x + 5 < -4.$

19. $2x - 4 > 5 - 3x.$

10. $-4x - 3 > 2.$

20. $\sqrt{5}x + 1 < \sqrt{3} - x.$

11. $-3x + 5 > 1.$

21. $(x - 1)(x - 2) > 0.$

2.4. QUADRATIC INEQUALITIES

Before proving theorems about inequalities which are of the second degree involving one variable (that is, inequalities of the form

$$ax^2 + bx + c < 0, \text{ or of the form } ax^2 + bx + c > 0,$$

where a , b , and c are constants, and $a \neq 0$), let us consider three specific examples:

EXAMPLE I:

The inequality

$$3x^2 + 5x - 2 > 0:$$

If x is any number,

$$3x^2 + 5x - 2 = (3x - 1)(x + 2).$$

Then $3x^2 + 5x - 2 > 0$ is true if and only if

$$(3x - 1)(x + 2) > 0$$

is true.

EXAMPLE II:

The inequality

$$x^2 - 6x - 1 > 0:$$

If x is any number,

$$x^2 - 6x - 1 = (x - 3 + \sqrt{10})(x - 3 - \sqrt{10}).$$

Then $x^2 - 6x - 1 > 0$ is true if and only if

$$(x - 3 + \sqrt{10})(x - 3 - \sqrt{10}) > 0$$

is true.

EXAMPLE III:

The inequality

$$4x^2 + 4x + 7 > 0:$$

If x is any number,

$$\begin{aligned} 4x^2 + 4x + 7 &= 4x^2 + 4x + 1 + 6 \\ &= (2x + 1)^2 + 6. \end{aligned}$$

Then $4x^2 + 4x + 7 > 0$ is true if and only if

$$(2x + 1)^2 + 6 > 0$$

is true.

Example I illustrates the case in which the algebraic expression $ax^2 + bx + c$ may be expressed as (factored into) the product of two first degree algebraic expressions, which have rational numbers as constants. Example II

illustrates the case in which the algebraic expression $ax^2 + bx + c$ may be expressed as the product of two first degree algebraic expressions, which have one or more irrational numbers as constants. Example III illustrates the case in which the algebraic expression $ax^2 + bx + c$ cannot be expressed as the product of two first degree algebraic expressions. In terms of Theorem 2.2, on the solution of a quadratic equation, examples I and II illustrate different-appearing aspects of the case in which the number $b^2 - 4ac$ is positive, and example III illustrates the case in which the number $b^2 - 4ac$ is negative. There is also a fourth case to be considered, namely the case in which the number $b^2 - 4ac$ is 0.

For definiteness, and to avoid tripping over the ordering law O-4b, let us look at an inequality of the form

$$ax^2 + bx + c < 0,$$

in which a , b , and c are constants, with the assumption

$$a > 0.$$

This assumption is not really a restriction on our investigation, since an inequality of the form $sx^2 + wx + r < 0$, in which s is a negative number, is equivalent to the inequality

$$(-s)x^2 + (-w)x + (-r) > 0,$$

in which the first constant is a positive number. We consider then, the inequality:

$$ax^2 + bx + c < 0, \quad \text{with } a > 0.$$

CASE 1, $b^2 - 4ac > 0$

From Theorem 2.2, we know that the equation

$$ax^2 + bx + c = 0$$

has, as its only solutions, the two numbers

$$\frac{-b - \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

The second of these numbers is the larger.

The fact that these two numbers are solutions of the equation $ax^2 + bx + c = 0$ implies that, if x is any number,

$$x^2 + \frac{b}{a}x + \frac{c}{a} = \left(x + \frac{b + \sqrt{b^2 - 4ac}}{2a}\right)\left(x + \frac{b - \sqrt{b^2 - 4ac}}{2a}\right).$$

Since the constant, a , is a positive number, the ordering law O-4a tells us that the inequality

$$ax^2 + bx + c < 0$$

is true if and only if the inequality

$$x^2 + \frac{b}{a}x + \frac{c}{a} < 0$$

is true, and the equation above tells us that we are concerned with those values of the variable x for which the product on the right-hand side of that (true) equation is a negative number.

At this point it is convenient to bring in the concept of a linear coordinate system, and to simplify the problem of solving the given inequality by considering the wider question: "For what values of the variable x , is the algebraic expression

$$ax^2 + bx + c$$

a positive number, or a negative number, or the number 0?" This question is equivalent to the same question concerning the expression

$$x^2 + \frac{b}{a}x + \frac{c}{a},$$

and, finally, we will determine those values of the variable x for which the expression

$$\left(x - \frac{-b - \sqrt{b^2 - 4ac}}{2a}\right)\left(x - \frac{-b + \sqrt{b^2 - 4ac}}{2a}\right)$$

is positive, or negative, or 0.

Since the range of the variable x , in the algebraic expressions, equations, and inequalities considered, is the set of all numbers, it is a natural first step to use what we already know: those two numbers which are the values of the variable x for which this last expression is the number 0, and then to examine the remainder of the set of all numbers.

The two numbers

$$\frac{-b - \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

divide the set of all numbers into three intervals:

$$\text{the set of all } x, \quad x < \frac{-b - \sqrt{b^2 - 4ac}}{2a};$$

$$\text{the set of all } x, \quad \frac{-b - \sqrt{b^2 - 4ac}}{2a} < x < \frac{-b + \sqrt{b^2 - 4ac}}{2a};$$

and

$$\text{the set of all } x, \quad x > \frac{-b + \sqrt{b^2 - 4ac}}{2a}.$$

If x is in the first interval, then

$$x < \frac{-b - \sqrt{b^2 - 4ac}}{2a};$$

$$x - \frac{-b - \sqrt{b^2 - 4ac}}{2a} < 0 \quad \text{by O-3.}$$

$$x - \frac{-b + \sqrt{b^2 - 4ac}}{2a} < -\frac{-b + \sqrt{b^2 - 4ac}}{2a} + \frac{-b - \sqrt{b^2 - 4ac}}{2a}$$

by 0 - 3;

$$x - \frac{-b + \sqrt{b^2 - 4ac}}{2a} < -2 \frac{\sqrt{b^2 - 4ac}}{2a} \quad \text{by algebraic laws;}$$

$$x - \frac{-b + \sqrt{b^2 - 4ac}}{2a} < 0 \quad \text{by the definition of the square root and the fact that } a > 0.$$

and

$$\left(x - \frac{-b - \sqrt{b^2 - 4ac}}{2a}\right)\left(x - \frac{-b + \sqrt{b^2 - 4ac}}{2a}\right) > 0 \quad \text{by the fact}$$

that the product of two negative numbers is positive.

In a similar manner, we find that if x is in the second interval:

$$\text{the set of all } x, \quad \frac{-b - \sqrt{b^2 - 4ac}}{2a} < x < \frac{-b + \sqrt{b^2 - 4ac}}{2a},$$

then the first factor of the product is positive, and the second factor is negative, so that the product itself is a negative number.

For the third interval:

$$\text{the set of all } x, \quad x > \frac{-b + \sqrt{b^2 - 4ac}}{2a},$$

if x is a number larger than both of the solutions of the equation

$$ax^2 + bx + c = 0,$$

then both of the factors in the product under consideration must be positive numbers, so that the product itself is a positive number.

The original problem was to solve the inequality

$$ax^2 + bx + c < 0.$$

We have discovered that the algebraic expression

$$ax^2 + bx + c$$

is a negative number if and only if x is a number in the interval consisting of all the numbers between two solutions of the equation $ax^2 + bx + c = 0$.

Then the solution of this inequality must be the second of our three intervals:

$$\text{the set of all } x, \quad \frac{-b - \sqrt{b^2 - 4ac}}{2a} < x < \frac{-b + \sqrt{b^2 - 4ac}}{2a}.$$

Now let us look back at the first two specific examples, each of which comes under the case: $b^2 - 4ac > 0$.

EXAMPLE I: $3x^2 - 5x - 2 > 0$:

$$3x^2 - 5x - 2 = (3x - 1)(x + 2).$$

$1/3$ and -2 are the solutions of the equation $3x^2 - 5x - 2 = 0$.

The numbers $1/3$ and -2 divide the set of all numbers into three intervals:

- (i) the set of all x , $x < -2$;
- (ii) the set of all x , $-2 < x < 1/3$;
- (iii) the set of all x , $x > 1/3$.

We consider each of these intervals in turn:

(i) If $x < -2$, then

$$3x < -6$$

by O-4a;

$$3x - 1 < -7$$

by O-3;

$$3x - 1 < 0$$

by the fact that $-7 < 0$.

$$x + 2 < 0$$

by O-3.

$$(3x - 1)(x + 2) > 0$$

by the fact that the product of two negative numbers is positive.

(ii) If $-2 < x < 1/3$, then

$$-6 < 3x < 1$$

by O-4a;

$$-7 < 3x - 1 < 0$$

by O-3.

$$0 < x + 2 < 7/3$$

by O-3.

$$(3x - 1)(x + 2) < 0$$

by the fact that the product of a positive number and a negative number is negative.

(iii) If $x > 1/3$, then

$$3x > 1$$

by O-4a;

$$3x - 1 > 0$$

by O-3.

$$x + 2 > 7/3$$

by O-3.

$$x + 2 > 0$$

by the fact that $7/3 > 0$.

$$(3x - 1)(x + 2) > 0$$

by the fact that the product of two positive numbers is positive.

We conclude that the solution of the inequality

$$3x^2 - 5x - 2 > 0$$

is the set consisting of the two intervals:

the set of all x , $x < -2$

and

the set of all x , $x > 1/3$.

EXAMPLE II: $x^2 - 6x - 1 < 0$:

$$x^2 - 6x - 1 = (x - 3 - \sqrt{10})(x - 3 + \sqrt{10}).$$

$3 - \sqrt{10}$ and $3 + \sqrt{10}$ are the solutions of the equation $x^2 - 6x - 1 = 0$.

The numbers $3 - \sqrt{10}$ and $3 + \sqrt{10}$ divide the set of all numbers into three intervals:

- (i) the set of all x , $x < 3 - \sqrt{10}$;
- (ii) the set of all x , $3 - \sqrt{10} < x < 3 + \sqrt{10}$;
- (iii) the set of all x , $x > 3 + \sqrt{10}$.

Rather than use all the words necessary to describe the investigation of the sign of the expression $x^2 - 6x - 1$ for x having a value in each of these three intervals, we will introduce the following chart:

	$x - 3 + \sqrt{10}$	$x - 3 - \sqrt{10}$	product
$x < 3 - \sqrt{10}$	neg.	neg.	pos.
$3 - \sqrt{10} < x < 3 + \sqrt{10}$	pos.	neg.	neg.
$x > 3 + \sqrt{10}$	pos.	pos.	pos.

The vertical columns under the first degree expressions are filled with the abbreviations for "positive" or "negative" according to the sign of the first degree expression for x in the interval described at the left. When the signs of both factors of the product are determined, we write the sign of the product. A glance at the completed chart then tells us that the solution of the inequality

$$x^2 - 6x - 1 < 0$$

is the interval:

the set of all x , $3 - \sqrt{10} < x < 3 + \sqrt{10}$.

CASE 2, $b^2 - 4ac = 0$

From Theorem 2.2, we see that $b^2 - 4ac = 0$ implies that

$$ax^2 + bx + c = a \left(x + \frac{b}{2a} \right)^2.$$

If, as we have assumed, a is positive, then an inequality of the form

$$ax^2 + bx + c < 0$$

has, in this case, no solution, since the square of a number is never negative. An inequality of the form

$$ax^2 + bx + c > 0$$

would, in this case, have the solution consisting of the set of all numbers other than $-b/2a$.

CASE 3, $b^2 - 4ac < 0$

Theorem 2.2 tells us that the equation $ax^2 + bx + c = 0$ is never true in this case. This fact suggests that the expression $ax^2 + bx + c$ has the same sign for all values of the variable x . This suggestion is, in fact, true, but we will omit a general proof. In order to solve the inequality $ax^2 + bx + c < 0$, it is sufficient to note that $a > 0$ implies that $ax^2 + bx + c$ is always positive, and $a < 0$ implies that $ax^2 + bx + c$ is always negative. For our specific example,

EXAMPLE III: $4x^2 + 4x + 7 > 0$:

$$4x^2 + 4x + 7 = (2x + 1)^2 + 6.$$

Since $(2x + 1)^2$ is never negative,

$$(2x + 1)^2 + 6 > 6,$$

so that the solution of the given inequality is the set of all numbers.

Exercises

Solve the inequalities:

1. $x^2 - 2x - 3 < 0$
2. $x^2 - 2x > 3$.
3. $x^2 - 3x - 4 < 0$.
4. $2x^2 - x - 6 > 0$.
5. $2x^2 - 3x - 5 > 0$.
6. $3x^2 + 4x - 1 < 0$.
7. $-2x^2 + 12x - 18 > 0$.
8. $3x^2 > 6x - 3$.
9. $x^2 - 3x + 10 < 0$.
10. $2x^2 - 5x + 19 > 0$.
11. $x^3 < 0$.
12. $x^2(x^2 - 2x - 3) < 0$.
13. $(x^2 + 1)(x^2 - 2x - 3) < 0$.

14. Prove the theorem:

If a , b , and c are constants, if $a > 0$, and if $b^2 - 4ac < 0$, then the solution of the inequality $ax^2 + bx + c > 0$ is the set of all numbers.

15. Prove the theorem:

If a , b , and c are constants, if $a < 0$, and if $b^2 - 4ac < 0$, then the solution of the inequality $ax^2 + bx + c > 0$ is the set consisting of no numbers.

16. What is the difference between saying that an inequality has no solutions and saying that the solution of an inequality is the set consisting of no numbers.
17. Prove the theorem:
If a , b , and c are constants, if $a > 0$, and if $b^2 - 4ac = 0$, then the solution of the inequality $ax^2 + bx + c > 0$ is the set of all numbers other than $-b/2a$.

2.5. OTHER INEQUALITIES

The techniques which may be used to find the solution of an inequality involving one variable other than one of the first or second degree depend on the type of algebraic expression involved in such an inequality. In this section we will consider one technique which is applicable in many situations.

The basic facts about the set of all numbers which are used are that the product of an odd (natural) number of negative numbers is a negative number, and that the product of an even (natural) number of negative numbers is a positive number. We will concern ourselves with algebraic expressions which can be expressed as the product or quotient of (that is, factored into) expressions which are of first and/or second degree. For the purposes of dealing with such algebraic expressions, we have

Definition 2.11: In an algebraic expression involving one variable, a *critical number* for the expression is either a number which is not in the range of the variable or a number which makes the algebraic expression 0.

For example, the algebraic expression

$$\frac{(x + 1)(x - 4)}{x - 7}$$

has the critical numbers -1 , 4 , and 7 .

Whenever we can find all the critical numbers of an algebraic expression, which expression occurs in an inequality of the form "algebraic expression < 0 " or the form "algebraic expression > 0 ," we find the solution of this inequality by considering, in turn, each of the intervals formed from the set of all numbers by removing from this set the critical numbers of the expression. Our procedure consists of considering the sign of each factor of the algebraic expression for each of the intervals, and then examining the sign of the expression itself for each of the intervals. We thus make use of the fact that the given algebraic expression—as well as each of its first or second degree factors, will have the same sign for every number *within* one of these intervals. We will omit the proof of this fact, as well as a precise definition of the phrase "algebraic expression," but will accept the fact as fact for the purposes of manipulation.

In those cases in which an algebraic expression can be expressed as the product or quotient of first degree algebraic expressions alone, we can immediately determine the critical numbers, and proceed to examine each interval formed by them. In those cases in which some of the factors of an algebraic expression are second degree algebraic expressions, then we may use Theorem 2.2 to discover the critical numbers, and proceed as before. In those cases in which we cannot express an algebraic expression as the product or quotient of first and second degree expressions only, then we cannot use the technique prescribed here.

For example, let us consider the inequality:

$$x^3 - 2x^2 - 5x + 6 > 0.$$

If x is any number,

$$x^3 - 2x^2 - 5x + 6 = (x - 1)(x + 2)(x - 3).$$

(This "factoring" is admittedly pulled out of thin air. The techniques of "factoring" a third degree expression are not in the scope of this book, but properly precede the study of calculus.) The numbers -2 , 1 , and 3 are the critical numbers of the given algebraic expression of the left-hand side of our inequality. These three numbers divide the set of all numbers into four intervals:

- (i) the set of all x , $x < -2$;
- (ii) the set of all x , $-2 < x < 1$;
- (iii) the set of all x , $1 < x < 3$;
- (iv) the set of all x , $x > 3$.

Examination of each of these intervals, in turn, yields the following chart:

	$x - 1$	$x + 2$	$x - 3$	product
$x < -2$	neg.	neg.	neg.	neg.
$-2 < x < 1$	neg.	pos.	neg.	pos.
$1 < x < 3$	pos.	pos.	neg.	neg.
$x > 3$	pos.	pos.	pos.	pos.

The solution of the inequality $x^3 - 2x^2 - 5x + 6 > 0$ is the set consisting of the two intervals:

the set of all x , $-2 < x < 1$
and the set of all x , $x > 3$.

As a second example, we consider the inequality:

$$\frac{2}{x-3} < \frac{1}{x+4}.$$

There is a great temptation here to perform the manipulation known as "cross multiplying." Such a manipulation would involve the multiplication of both sides of the inequality by, among other things, the expression $x - 3$. The ordering laws O-4a and O-4b tell us how to perform such a multiplication in the case of a *known* (that is, constant) number, but do not tell us anything about multiplication by a number whose sign is not known. Therefore no multiplication of both sides of an inequality is allowable when the multiplier involves the variable itself, in a form in which the sign of the multiplier is not always the same for all possible values of the variable.

However, the ordering law O-3 does allow the addition of any number, positive or negative, or unknown, to each side of an inequality so that we may say

$$\frac{2}{x-3} < \frac{1}{x+4}$$

is true if and only if

$$\frac{2}{x-3} - \frac{1}{x+4} < 0$$

is true. If x is any number other than -4 or 3 ,

$$\frac{2}{x-3} - \frac{1}{x+4} = \frac{x+11}{(x-3)(x+4)},$$

we conclude that the original inequality is true if and only if

$$\frac{x+11}{(x-3)(x+4)} < 0$$

is true. The critical numbers of the algebraic expression on the left side of this inequality are -11 , -4 , and 3 . These three numbers divide the set of all numbers into four intervals. We have the chart:

	$x + 11$	$x + 4$	$x - 3$	expression
$x < -11$	neg.	neg.	neg.	neg.
$-11 < x < -4$	pos.	neg.	neg.	pos.
$-4 < x < 3$	pos.	pos.	neg.	neg.
$x > 3$	pos.	pos.	pos.	pos.

So that the solution of our original inequality is the set consisting of the two intervals:

the set of all x , $x < -11$
and the set of all x , $-4 < x < 3$.

Exercises

Solve the inequalities:

1. $(2x + 1)(x - 3)(x + 1) < 0$.
2. $(3x - 4)(2 - x)(3x + 1) > 0$.
3. $2x^3 - x^2 - 3 > 0$.
4. $x^3 - 5x^2 + 6x < 0$.
5. $x^4 + 6x^3 + 9x^2 > 0$.
6. $x^3 - 4x^2 + 5x < 0$.
7. $(x^2 - 4)(x^2 - 9) < 0$.
8. $(x - 1)(x + 2)(x - 3)(x + 4)(x - 5) > 0$.
9. $\frac{3}{x - 5} < 0$.
10. $\frac{2x + 1}{x - 4} > 0$.
11. $\frac{1}{x} + \frac{2}{x - 1} < 0$.
12. $\frac{3}{x^2 - 1} > \frac{1}{x + 1}$.
13. Prove the theorem:

If a , b , and c are constants, and if $a < b < c$, then the solution of the inequality

$$(x - a)(x - b)(x - c) < 0$$

is the set consisting of the two intervals:

the set of all x , $x < a$
and the set of all x , $b < x < c$.

14. Prove the theorem:

If a and b are constants, and $a < b$, then the solution of the inequality

$$\frac{1}{x - a} > \frac{1}{x - b}$$

is the interval:

the set of all x , $a < x < b$.

2.6. ABSOLUTE VALUES

Suppose we attempt to answer the question: "On a linear coordinate system, what is the distance between the point with coordinate 2 and the point with coordinate 7?" This question can be answered rapidly, by saying that the point with label "7" is 5 units away from the point with label "2." Now let us slow the speed of giving an answer. We have the two numbers, 2 and 7, to work with. There are two ways of subtracting one of these

/ 0 . - / - -

numbers from the other, namely $2 - 7$ and $7 - 2$. Of these two possible subtractions, we choose $7 - 2$ in order to get the correct answer to the original question. But why would $2 - 7$ be an incorrect answer? The reason, of course, is that *distance* is never a negative number. In order to have a convenient manner of writing about distances, we give

Definition 2.12: If a is a number, the *absolute value* of a , written $|a|$, is the distance, on a straight line with a linear coordinate system, between the point with coordinate a and the origin.

Since distance is never negative, this definition says that the absolute value of a number is never negative. Thus we have a new operation which can be performed, namely, taking the absolute value of a number. For a number which is positive, the absolute value is the same as the number. For a number which is negative, the absolute value is the negative of that negative number. In symbols, this last statement says:

$$\text{if } a < 0, \text{ then } |a| = -a.$$

For example, $|3| = 3$, and $|-5| = -(-5) = 5$. Directly from the definition, the absolute value of the number 0 is 0.

An inequality which mentions absolute values provides a particularly convenient short-hand way of making statements about distances on a line with a linear coordinate system, or about differences between numbers. The statement

$$|x| < 1$$

means "The distance between the point with coordinate x and the origin is less than 1 unit of distance." This may be rephrased as "The point with coordinate x is within 1 unit of the origin." But this last statement involving the linear coordinate system may be translated into a statement about numbers: "The number x is within 1 unit of the number 0." This statement about numbers may also be made by stating the inequality: $-1 < x < 1$.

In like manner, the statement

$$|x - a| < b, \text{ where } b > 0$$

is equivalent to " $x - a$ is a number within b units of 0," which, in turn, is equivalent to the inequality

$$a - b < x < a + b.$$

This inequality may be restated as "The number x is within b units of the number a ."

We will be particularly concerned with proving the truth of statements of the form

"If $|x - a| < d$ then $|(\text{expression involving } x) - L| < e$." For example, consider the assertion:

$$\text{"If } |x - 1| < 1/4 \text{ then } |(3x + 5) - 8| < 3/4."$$

This assertion is equivalent to the statement:

“If x is a number within $1/4$ unit of 1, then $3x + 5$ will be a number within $3/4$ unit of 8.”

This statement, in turn, may be translated to read:

“If x is a number between $3/4$ and $5/4$, then $3x + 5$ will be a number between $29/4$ and $35/4$.”

Expressed in terms of inequalities, we are concerned with the assertion:

“If $3/4 < x < 5/4$ then $29/4 < 3x + 5 < 35/4$.”

This last statement is clearly true, since,

$$\begin{aligned} &\text{if } 3/4 < x < 5/4, \\ &\text{then } 9/4 < 3x < 15/4 \text{ by the law O-4a;} \\ &(9/4) + 5 < 3x + 5 < (15/4) + 5 \text{ by the law O-3;} \\ &29/4 < 3x + 5 < 35/4 \text{ by addition.} \end{aligned}$$

The use of the absolute value symbol to form a statement saying that a number is within a certain “distance” of another number allows us to describe intervals by reference to their centers and lengths, rather than by means of the end points.

The interval:

the set of all x , $a < x < b$,

may also be described by

$$\text{the set of all } x, \left| x - \frac{a+b}{2} \right| < \frac{b-a}{2}$$

since this interval does consist of all the numbers within $\frac{b-a}{2}$ units of the number $\frac{a+b}{2}$.

Conversely, the interval:

the set of all x , $|x - p| < q$

may also be described by

the set of all x , $p - q < x < p + q$.

As another use for the absolute value notation, we are now able to simplify an algebraic expression of the form

$$\sqrt{x^2}.$$

There is a great temptation to reason that finding *the* square root of a number undoes the process of squaring, and to write

$$\sqrt{x^2} = x,$$

as an equation which is true for all numbers, x . But the definition of the square root symbol requires that the lefthand side of this equation never

be a negative number. Thus the equation $\sqrt{x^2} = x$ is false for any value of the variable x which is a negative number. We may, however, write the equation

$$\sqrt{x^2} = |x|,$$

which is true for all numbers.

The relation of the operation of taking absolute value and the algebraic operations of addition and multiplication are given in the following two theorems:

Theorem 2.4:

If a and b are numbers,
then $|a + b| \leq |a| + |b|$.

Proof:

- | | |
|---|--|
| 1. $- a \leq a \leq a $ | by the definition (2.12) of absolute value. |
| 2. $- b \leq b \leq b $ | by the definition (2.12) of absolute value. |
| 3. $-(a + b) \leq a + b \leq a + b $ | by steps 1 and 2 and the law O-3. |
| 4. $ a + b \leq a + b $ | by step 4 and the definition (2.12) of absolute value. |

(The inequality in Theorem 2.4 is often called the "triangle inequality.")

Theorem 2.5:

If a and b are numbers,
then $|ab| = |a||b|$.

Proof:

- If $a > 0$ and $b > 0$, then $|ab| = ab$ and $|a||b| = ab$.
- If $a < 0$ and $b > 0$, then $|ab| = -ab$ and $|a||b| = (-a)b$.
- If $a > 0$ and $b < 0$, then $|ab| = -ab$ and $|a||b| = a(-b)$.
- If $a < 0$ and $b < 0$, then $|ab| = ab$ and $|a||b| = (-a)(-b)$.

Exercises

- Sketch, on a line with a linear coordinate system, the intervals described by:
 - the set of all x , $|x| < 3$.
 - the set of all x , $|x - 2| < 4$.
 - the set of all x , $|x + 1| < 5$.
 - the set of all x , $|x - 4| < 1$.
 - the set of all x , $|x - 4| < 0.1$.
 - the set of all x , $|x - 4| < 0.01$.

2. Prove the theorem:

If a is a number, then $-|a| \leq a \leq |a|$.

3. Give specific numerical examples to illustrate the four steps in the proof of Theorem 2.5.

4. Prove the theorem:

If $|x - a| < d$ then $|(mx + b) - (ma + b)| < |m|d$.

5. If x is a number in the interval:

the set of all x , $|x + 1| < 3$,

find the largest interval inside of which x^2 may be found.

Solve the following inequalities:

6. $|3x - 12| < 9$.

7. $|2x + 4| < 4$.

8. $|7x - 2| < 1$.

9. $|5x - 20| < 1$.

10. $|5x - 20| < 0.1$.

11. $|5x - 20| < 0.01$.

12. $|5x - 20| < 0.001$.

13. $|5x - 20| < e$, where e is some positive number.



PLANE ANALYTIC GEOMETRY

3.1. PLANE COORDINATE SYSTEMS

In order to relate geometric facts about a straight line and algebraic facts about the set of all numbers, we established a one-to-one correspondence between the points on a straight line and the numbers. In this section we will develop a method of relating geometric facts about a plane and algebraic facts about numbers. The results of this development are known, collectively, as plane analytic geometry, and are used to answer two kinds of questions: First, given certain geometric statements about a plane, how can we express those statements algebraically? Second, given certain algebraic statements about numbers, how can we express these statements geometrically in a plane?

The basic concept used in plane analytic geometry is that of an ordered pair of numbers. The adjective “ordered” is used to mean that the order of presentation, as well as the numbers themselves, is of significance. We will write an ordered pair of numbers by using parentheses and a comma: (a, b) , where a and b are numbers. Thus $(3, 5)$ and $(5, 3)$ are two different ordered pairs. The most significant fact about ordered pairs of numbers is that two ordered pairs are equal if and only if the respective first numbers are equal *and* the respective second numbers are equal. Formally, we have

Definition 3.1: An *ordered pair of numbers* is a pair of numbers, such as a and b , written in the form (a, b) . The number a is called the *first element* of the ordered pair (a, b) , and b is called the *second element* of this ordered pair. Given two such ordered pairs, (a, b) and (c, d) , these two ordered pairs are equal if and only if $a = c$ and $b = d$.

We will now proceed to set up a one-to-one correspondence between the set of all points in a plane and the set of all possible ordered pairs of numbers. Let us start with a plane and the primitive concept of unit distance. On this plane we place two lines which are perpendicular to each other. For defini-

teness, we will consider the plane in such a position that one of the lines will appear horizontal and the other vertical. In this way, there is meaning attached to saying that a point is to the right of (or to the left of) the vertical line, and to saying that a point is above (or below) the horizontal line.

Since our concept of the set of all numbers is the same as the concept of the set of all possible points on a straight line, we can consider all points of the plane by considering those points which are at all possible distances, horizontally to the left or right, and vertically upward or downward, from the two given lines.

We will label each point in the plane with an ordered pair of numbers, which ordered pair of numbers will be called the *coordinates* of that point. The first element of the ordered pair will always refer to a distance measured horizontally from the vertical line (although that number may not be the distance itself, but instead may be the negative of the distance). The second element of the ordered pair will always refer to a distance measured vertically from the horizontal line (although that number may not be the distance itself).

The horizontal line on the plane is called the *x-axis*, or *horizontal axis*. (The symbol " x " is used merely as a symbol here, not as a variable.) The vertical line on the plane is called the *y-axis*, or *vertical axis*. The first number of an ordered pair which labels a given point is called the *x-coordinate*, or *horizontal coordinate*, or *abscissa*, of that point. The second number of an ordered pair which labels a point is called the *y-coordinate*, or *vertical coordinate*, or *ordinate* of that point. It should be noted that the horizontal and vertical coordinates of a point are numbers, not line segments, nor anything else but numbers.

We are now ready to establish the one-to-one correspondence which gives us the method of labeling points in a plane.

Definition 3.2: A *plane coordinate system* is a one-to-one correspondence between the point of a plane and all possible ordered pairs of numbers, in which:

- 1) the horizontal coordinate of any point which is to the right of, or on, the vertical axis is the distance between that point and the vertical axis.
- 2) the horizontal coordinate of any point which is to the left of the vertical axis is the negative of the distance between that point and the vertical axis.
- 3) the vertical coordinate of any point which is above, or on, the horizontal axis is the distance between that point and the horizontal axis.
- 4) the vertical coordinate of any point which is below the horizontal axis is the negative of the distance between that point and the horizontal axis.

According to this definition, all points which are both above the *x-axis* and to the right of the *y-axis* have both of their coordinates positive numbers.

All points which are to the left of the vertical axis and above the horizontal axis have a negative number for an x -coordinate and a positive number for a y -coordinate. All points which are below the x -axis and to the left of the y -axis have both of their coordinates negative numbers. All points which are to the right of the vertical axis and below the horizontal axis have positive numbers for their x -coordinates and negative numbers for their y -coordinates. The horizontal coordinate of any point on the y -axis is 0 and the y -coordinate of any point on the horizontal axis is 0. The point with coordinates $(0, 0)$ —the intersection of the two axes—is called the *origin* of the coordinate system.

For convenience, we call the parts of the plane formed by the two axes “quadrants,” and number these quadrants, arbitrarily choosing as the “first quadrant” that portion of the plane in which both coordinates are positive, and continuing the numbering in a counter-clockwise manner.

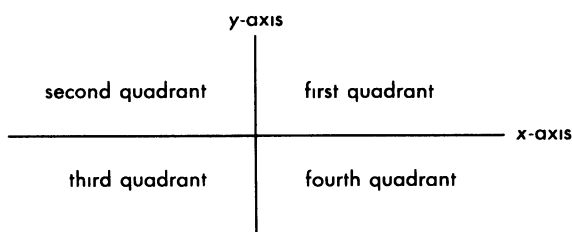


Fig. 3.1

The definition of a plane coordinate system allows us to find the coordinates of any given point, starting from a geometric description of the location of that point, and also allows us to locate, geometrically, any point whose coordinates are given. For example, that point which is 3 units of distance to the left of the vertical axis and 5 units of distance above the horizontal axis has coordinates $(-3, 5)$; that point with coordinates $(1, -4)$ is located one unit of distance to the right of the y -axis and 4 units of distance below the x -axis.

Exercises

1. Sketch a plane with a plane coordinate system, and indicate the location of the points with coordinates:
 $(1, 1)$, $(-1, 1)$, $(1, -1)$, $(-1, -1)$, $(-3, 7)$, $(-4, 1/2)$, $(3, -2)$, $(-1/3, 5)$, $(-2, 7)$.
2. Let (a, b) be the coordinates of a point in the second quadrant of a plane. Describe the location of the points with coordinates $(a, -b)$, $(-a, b)$, $(-a, -b)$, and (b, a) .
3. Describe the location of all the points in the plane whose x -coordinate is the number -4 .
4. Describe the location of all the points in the plane whose vertical coordinate is the number 7.

5. Describe the location of all the points in the plane whose horizontal and vertical coordinates are the same.
6. Describe the location of all the points in the plane whose x - and y -coordinates are negatives of each other.
7. How can one find the distance, in terms of the coordinates, between two points which lie on a vertical line?
8. How can one find the distance, in terms of the coordinates, between two points which lie on a horizontal line?
9. Find a one-to-one correspondence between the set of all numbers and the points in some set on the plane.

3.2. DISTANCE

Most geometric properties of interest involve the concept of distance. It behooves us, then, to discover how we can find the distance between two points in a plane (a purely geometric property) if we known their coordinates (an algebraic property). Since a plane coordinate system is defined by means of distances measured along horizontal and vertical lines, there is no great difficulty in finding the distance between points which are on a vertical line, or which are on a horizontal line.

Theorem 3.1:

If the point A has coordinates (a, b) and the point P has coordinates (a, q) , then the distance between P and A is $|b - q|$.

Proof:

1. The y -coordinates of the points on the vertical line through A and P form a linear coordinate system by the definition (1.10) of a linear coordinate system.
2. The distance between A and P is $|b - q|$ by the definition (1.10) and the definition (2.12) of absolute value.

For the measure of distances along horizontal lines we have

Theorem 3.2:

If the point A has coordinates (a, b) and the Point P has coordinates (p, b) , then the distance between P and A is $|a - p|$.

The proof of this theorem is left as an exercise.

In order to find the distance between two points which are not directly above each other, nor on one horizontal line, we need to use the Pythagorean theorem of elementary plane geometry in proving:

Theorem 3.3:

If the point A has coordinates (a, b) , if the point P has coordinates (p, q) , and if $a \neq p$ and $b \neq q$, then the distance between A and P is

$$\sqrt{(a - p)^2 + (b - q)^2}.$$

In words, this theorem says that the distance between two points on a plane with a plane coordinate system is the square root of the sum of the squares of the differences between the respective coordinates of those points. In the proof which follows, no specific diagram is referred to, but the argument applies with equal validity to each of the four diagrams in Fig. 3.2. These four diagrams illustrate all possible relative positions of the two points in question.

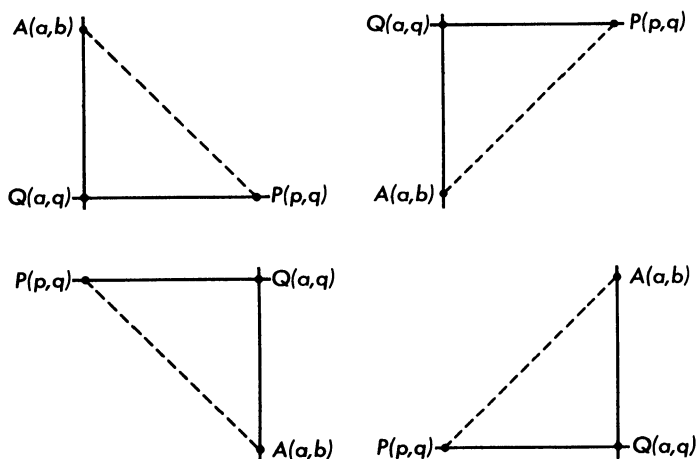


Fig. 3.2

Proof of 3.3:

1. A has coordinates (a, b) by hypothesis.
2. P has coordinates (p, q) by hypothesis.
3. Let Q be the intersection of a horizontal line through P with a vertical line through A .
4. Q has x -coordinate a by steps 1 and 3.
5. Q has y -coordinate q by steps 2 and 3.
6. The distance between A and Q is $|b - q|$ by steps 1, 4, and 5, and Theorem 2.1.
7. The distance between P and Q is $|a - p|$ by steps 2, 4, and 5, and Theorem 3.2.
8. The line segments AP , QP , and AQ form a right triangle by step 3.

9. The distance between A and P is

$$\sqrt{|a - p|^2 + |b - q|^2}$$

by steps 6, 7, and 8, and the Pythagorean theorem.

10. $|b - q|^2 = (b - q)^2$

by the definition (2.12) of absolute value and the fact that the square of a number is never negative.

11. $|a - p|^2 = (a - p)^2$

as in step 10.

12. The distance between A and P is

$$\sqrt{(a - p)^2 + (b - q)^2}$$

by steps 9, 10, and 11.

Exercises

1. Prove Theorem 3.2.

2. How does the conclusion of Theorem 3.3 include the conclusions of Theorems 3.1 and 3.2?

Find the distance between the points with coordinates:

3. $(0, 0)$ and $(-4, 1)$.
4. $(-1, -3)$ and $(2, 1/2)$.
5. $(1/2, -4)$ and $(-3/2, 1)$.
6. $(1/3, 1/2)$ and $(1/2, -1/3)$.
7. $(-2, 1)$ and $(-3, 4)$.
8. $(-2, -1)$ and $(-3, -1)$.
9. $(-1/2, 0)$ and $(4, -2)$.
10. $(1, -1)$ and $(-3, -4)$.
11. $(1, -1)$ and $(1, -4)$.
12. $(-\sqrt{2}, \sqrt{5})$ and $(\sqrt{3}, -\sqrt{7})$.

3.3. GRAPHS OF EQUATIONS AND INEQUALITIES

In this section we will concern ourselves with statements which contain at most two variables, and specifically the variables will be x and y . We use these specific variables for definiteness, and because we have defined a plane coordinate system in terms of the symbols x and y . We have, first

Definition 3.3: Given a statement involving at most the variables x and y , a point in a plane with a plane coordinate system is said to *satisfy* that statement if the x -coordinate of that point is a value of the variable x , and the y -coordinate is a value of the variable y , which values together make the given statement true.

For example, the point with coordinates $(1, 2)$ satisfies the statements:

$$\begin{aligned}x &\text{ and } y \text{ have the same sign;} \\y &= 2x; \\y &> x;\end{aligned}$$

while the point with coordinates $(0, -1)$ does not satisfy any of them.

The point with coordinates $(-1, 4)$ satisfies the statement

$$2x + y = 2,$$

since the statement

$$2(-1) + 4 = 2$$

is true, and the point with coordinates $(4, -1)$ does not satisfy the statement

$$2x + y = 2,$$

since the statement

$$2(4) + (-1) = 2$$

is false.

We will be primarily concerned with statements which are either equations or inequalities, so that the collection of all those points in the plane which satisfy a given equation or a given inequality is of particular interest.

Definition 3.4: Given a statement involving at most the variables x and y (in particular, an equation or an inequality), the *graph of that statement* is the collection of all those points in a plane with a plane coordinate system, which points satisfy that statement (equation, inequality).

Thus, the graph of the statement,

$$x \text{ and } y \text{ have the same sign,}$$

is the collection of all those points whose coordinates have the same sign. This set of points consists of all the points whose coordinates are both negative, together with all those points whose coordinates are both positive. This set of points is the set of all points which are in either the first or third quadrants of the plane, with the exception of those points on the coordinate axes.

The graph of the equation

$$y = 2x$$

is the set of all those points whose y -coordinates are twice their x -coordinates. As it happens, this graph consists of the straight line (extended indefinitely) passing through the origin and the point $(1, 2)$.

The graph of the inequality

$$y > x$$

consists of all those points whose y -coordinates are bigger than their x -coordinates. This set consists of all those points to the left of a line through the origin and the point $(1, 1)$.

In these examples, the ranges of the variables x and y were the set of all numbers. If we were to consider a statement involving the variables x and y , in which statement the ranges of these variables were some sets smaller than the set of all numbers, we would either mention the ranges explicitly,

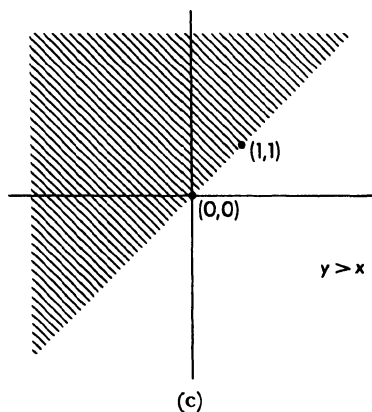
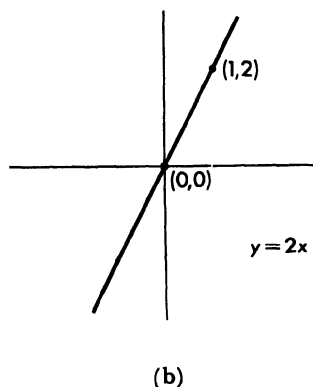
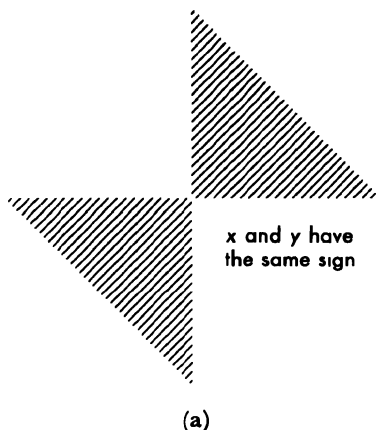


Fig. 3.3

or use the convention which we have adopted and which warns us off the forbidden operations of division by zero and taking the square root of a negative number. For example, the modified equation

$$x = y, \quad x > 0$$

has, as its graph, those points in the first quadrant lying on the line through $(1, 1)$ and $(2, 2)$. Of course those *points* on the graph of this modified equation have only positive y -coordinates, but the characteristics of the graph of a statement have no necessary connection with the *ranges* of the variables involved. Thus, in this modified equation, the variable x has, for its range, the set of all positive numbers, and the x -coordinates of the points on the

graph of this modified equation are only positive numbers; but the range of the variable y in this modified equation is the set of all numbers, even though the set of y -coordinates of the points on the graph consists of positive numbers only.

This distinction between the ranges of the two variables involved and the values of those variables which make a statement true can be noticed particularly in the modified equation

$$y^2 = x - 1, \quad y \geq 0.$$

The range of the variable x , in this statement, is the set of all numbers, but the statement is true only for values of the variable x which are not less than 1. The range of the variable y is, as indicated, the set of all non-negative numbers, which is also the set of values of this variable for which the equation could be true. This graph is the same as the graph of the equation

$$y = \sqrt{x - 1},$$

but the variable x , in this second equation, has as its range, the set of all numbers which are not less than 1, and the variable y has, as its range, the set of all numbers. We note also that we have a situation in which two different equations have the same graph.

The two basic problems of analytic geometry are:

- (1) Given an equation, or inequality, or statement, involving at most the variables x and y , what are the geometric properties of the graph of that equation, or inequality, or statement?
- (2) Given a collection of points in a plane with certain geometric properties, is that collection of points the graph of some equation, or inequality, or statement, involving at most the variables x and y ?

The second of these questions gives rise to

Definition 3.5: An *equation of a collection of points* is an equation involving at most the variables x and y whose graph is that collection of points.

While we could also define “an inequality of a collection of points” or “a statement of a collection of points,” we will not do so, and will confine our attention mainly to those collections of points which are graphs of equations.

Exercises

Describe and sketch the graphs of:

- | | |
|---------------|-------------------|
| 1. $x = 4$. | 6. $y = 0$. |
| 2. $y = 3$. | 7. $xy = 0$. |
| 3. $x = -7$. | 8. $x + 1 = 0$. |
| 4. $y = -1$. | 9. $y + 2 = 0$. |
| 5. $x = 0$. | 10. $x + y = 0$. |

11. $y - x = 0$. 15. $x + y = 1$.
 12. $y = 3x$. 16. $x + y = 2$.
 13. $y = 3x + 1$. 17. $x + y = -1$.
 14. $y = 3x + 2$.

Find an equation of:

18. All points 3 units to the right of the y -axis.
 19. All points 2 units above the x -axis.
 20. All points 1 unit to the left of the y -axis.
 21. All points 4 units below the x -axis.
 22. All points on the x -axis.
 23. All points on the y -axis.
 24. All points whose x - or y -coordinates are 0.
 25. All points on the line through the origin and $(1, 1)$.
 26. All points whose x - and y -coordinates are negatives of each other.

3.4. STRAIGHT LINES

In this section we will examine the graphs of simple equations involving at most the variables x and y . The simplest kind of an equation involving "at most" the variables x and y is one which does not actually mention either of these variables. The graph of such an equation is obviously either the entire plane or the set of no points at all, depending on whether the given equation not involving x and y is false or true.

The first non-trivial kind of equation involving at most the variables x and y which we will consider is an equation which mentions one of these variables, but not the other. The equation

$$x = a,$$

where a is a specific number, is an algebraic statement about numbers. Translated into geometric terminology, we have

"The x -coordinate of a point is the number a ." The graph of this statement is obviously a vertical line, $|a|$ units from the y -axis, to the right or left of the y -axis according to the number a 's being positive or negative. In like manner, the graph of the equation

$$y = b,$$

where b is a specific number, is a horizontal line, $|b|$ units from the x -axis, above or below the x -axis according to the number b 's being positive or negative.

We will examine an equation of the form

$$Ax + By + C = 0,$$

in which A , B , and C are constants.

If $A = 0$, and $B \neq 0$, then the graph of

$$(0)x + By + C = 0$$

is also the graph of

$$y = -C/B,$$

which graph is a horizontal straight line.

If $A \neq 0$ and $B = 0$, then the graph of

$$Ax + (0)y + C = 0$$

is also the graph of

$$x = -C/A,$$

which graph is a vertical straight line.

Real interest occurs when both the variables x and y are in fact involved in the equation under consideration, so that, as the simplest case, we consider the equation

$$Ax + By + C = 0,$$

in which A , B , and C are constants, $A \neq 0$, and $B \neq 0$. This equation will then be true if and only if

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

is true, since we have assumed that B is not 0. The number $-A/B$ may be either positive or negative, but is not 0, since we have also assumed that A is not 0. For convenience, we will let

$$m = -A/B, \quad b = -C/B,$$

so that we are examining the graph of the equation

$$y = mx + b,$$

in which b and m are constants, and $m \neq 0$. We consider the two possibilities:

CASE 1, $m > 0$

Let p and q be two values of the variable x , (whose range is the set of all numbers) such that $p < q$.

Let r be the number

$$mp + b,$$

so that the point (p, r) is on the graph of $Ax + By + C = 0$.

Let s be the number

$$mq + b,$$

so that the point (q, s) is on the graph of $Ax + By + C = 0$. Applying the algebraic laws, we have

$$\begin{aligned} s - r &= (mq + b) - (mp + b) \\ &= m(q - p). \end{aligned}$$

Since $p < q$, we have $q - p > 0$. Since $m > 0$, we have $s - r > 0$. Since $s - r > 0$, we have $r < s$.

Then, if we examine the graph of $Ax + By + C = 0$ for the case in which $m (= -A/B)$ is positive, we find that, loosely speaking, larger x -coordinates mean larger y -coordinates, or that the graph may be described as "rising."

CASE 2, $m < 0$

Let p , q , r , and s be as described in Case 1. We have

$$s - r = m(q - p).$$

Since $p < q$, we have $q - p > 0$. Since $m < 0$, we have $s - r < 0$. Since $s - r < 0$, we have $r > s$.

Then, if we examine the graph of $Ax + By + C = 0$ for the case in which $m (= -A/B)$ is negative, we find that, loosely speaking, larger x -coordinates mean smaller y -coordinates, or that the graph may be described as "falling."

Now let us look at any three distinct points on the graph of

$$Ax + By + C = 0 \quad (y = mx + b),$$

whose coordinates are (a, d) , (b, e) , and (c, f) , so arranged that $a < b < c$. That is, we consider the left-to-right arrangement of these three points and assume that (a, d) are the coordinates of the first point, (b, e) the coordinates of a point to the right of the first point, and that (c, f) are the coordinates of a point to the right of the second. Since these points are on the graph of $Ax + By + C = 0$, the definition (3.4) of the graph of an equation tells us that

$$\begin{aligned} Aa + Bd + C &= 0; \\ Ab + Be + C &= 0; \\ \text{and } Ac + Bf + C &= 0. \end{aligned}$$

From our previous investigation, we have discovered that the graph of " $Ax + By + C = 0$ " is either "falling" (that is, $m = -A/B < 0$) or "rising" (that is, $m = -A/B > 0$). Insofar as these three points are concerned, this choice between a "rising" or "falling" graph tells us that we have either $d < e < f$ or $d > e > f$.

CASE 1, $m = -A/B > 0$, "RISING GRAPH," $d < e < f$

The true statements $Aa + Bd + C = 0$

and $Ab + Be + C = 0$

yield the fact

$$A(b - a) + B(e - d) = 0.$$

The true statements $Ab + Be + C = 0$

and $Ac + Bf + C = 0$

yield the fact

$$A(c - b) + B(f - e) = 0.$$

Since $A \neq 0$, $a \neq b$, and $b \neq c$, there is no possibility of an attempted division by 0, so we have from the two (true) statements given which do not mention C ,

$$\frac{e-d}{b-a} = -A/B \quad \text{and} \quad \frac{f-e}{c-b} = -A/B.$$

Since A and B are fixed constants, these last two statements together imply that

$$\frac{e-d}{b-a} = \frac{f-e}{c-b}.$$

We have chosen a , b , and c so that $a < b < c$, which means that the numbers $b-a$ and $c-b$ are positive. In the case under consideration (a "rising" graph), we have $d < e < f$, so that the numbers $e-d$ and $f-e$ are positive. Finally, looking at the plane itself, we see that the numbers $e-d$, $f-e$, $b-a$, and $c-b$ are all the lengths of various line segments. (See Fig. 3.4.)

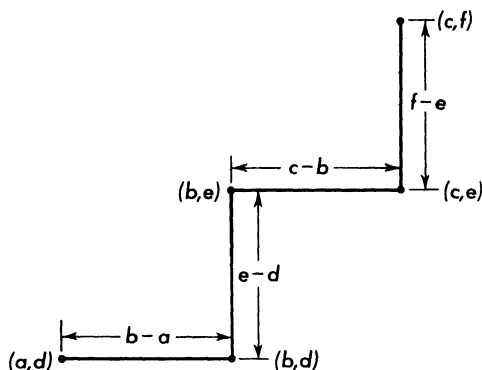


Fig. 3.4

Let us consider the points with coordinates (b, d) and (c, e) . The point with coordinates (b, d) is the intersection of a horizontal line from (a, d) with a vertical line from (b, e) . The point with coordinates (c, e) is the intersection of a horizontal line from (b, e) with a vertical line from (c, f) .

The number, $b-a$, is the length of the line segment joining (a, d) and (b, d) . The number, $c-b$, is the length of the line segment joining (b, e) and (c, e) . The number, $f-e$, is the length of the line segment joining (c, e) and (c, f) . The number, $e-d$, is the length of the line segment joining (b, d) and (b, e) .

Since

$$\frac{e-d}{b-a} = \frac{f-e}{c-b},$$

the right triangles formed by connecting (a, d) , (b, d) , and (b, e) and by connecting (b, e) , (c, e) , and (c, f) are similar to each other. Since the horizontal sides of these triangles are parallel to each other, and the vertical

sides of these triangles are parallel to each other, the hypotenuses—the line segment joining (a, d) to (b, e) and the line segment joining (b, e) to (c, f) —are also parallel to each other. But both these line segments pass through the point (b, e) . Consequently, the points (a, d) , (b, e) , and (c, f) are all on one straight line.

CASE 2, $m < 0$

For the proof that the points (a, d) , (b, e) , and (c, f) are all on one straight line, see the exercises.

Both cases yield the same result: Any three points on the graph of $Ax + By + C = 0$, where A , B , and C are constants, $A \neq 0$, and $B \neq 0$, lie on one straight line. Because of this fact, the graph itself must be either a straight line or a segment thereof. But since the ranges of the variables x and y are both the set of all numbers, we will not have merely a line segment, but an entire straight line which is neither vertical nor horizontal. The graph of $By + C = 0$ is a horizontal line and the graph of $Ax + C = 0$ is a vertical line, so that what we have done is a proof of

Theorem 3.4:

If A , B , and C are constants, and A and B are not both 0, the graph of the equation

$$Ax + By + C = 0$$

is a straight line.

In case $B \neq 0$, the number $m = -A/B$, whose positive or negative character determines the “rising” or “falling” character of the straight line which is the graph of $Ax + By + C = 0$, is of sufficient importance to be given a specific name. Since this number is essentially a geometric feature of the straight line, we will give a definition which starts with the geometry involved:

Definition 3.6: Given a straight line in a plane with a plane coordinate system, the *slope* of this line is defined as follows:

- a) if the given line is not vertical, and (a, b) and (p, q) are the coordinates of any two distinct points on the given line, the slope of the given line is the number

$$\frac{q - b}{p - a}$$

- b) if the given line is vertical, the line does not have a slope.

It should be noted that part b) of Definition 3.6 does *not* say that a vertical line has a slope, which slope is 0. That is, it should be realized that the number 0 is not nothing. A horizontal line, in fact, has for its slope the number 0.

The details making up the proof of Theorem 3.4 show that the number $-A/B$ is the slope of the line which is the graph of the equation $Ax + By + C = 0$, where $B \neq 0$.

Theorem 3.4 tells us that any equation of the form

$$Ax + By + C = 0$$

must have a straight line for its graph. If the constant $B = 0$, then the graph of this equation is a vertical line, which line does not have a slope. If the constant $B \neq 0$, then the graph of this equation is a non-vertical line, which line has for its slope the number $-A/B$ (which number may be 0).

For the reverse problem, that is, starting with a geometrical description of a straight line in a plane with a plane coordinate system, and attempting to find an equation of that line, we need to know how a straight line can be geometrically described. A vertical line presents no problems, since we know that such a line is the graph of an equation of the form

$$x = a.$$

A particular non-vertical line may be specified by giving either

- (1) the coordinates of two points on that line;

or

- (2) the slope, and the coordinates of one point on that line.

In each of these situations we will find an equation whose graph is the given straight line.

Theorem 3.5:

An equation of the non-vertical line through the points with coordinates (p, q) and (r, s) is

$$y - q = \frac{s - q}{r - p} (x - p).$$

Proof:

1. The slope of the given non-vertical line is the number

$$\frac{s - q}{r - p}$$

by the definition (3.6) of slope.

2. If (x, y) are the coordinates of a point, other than (p, q) , on the given line, then the slope of the given line is the number

$$\frac{y - q}{x - p}$$

by the definition (3.6) of slope.

3. If (x, y) are the coordinates of a point, other than (p, q) , on the given line, then

$$\frac{y - q}{x - p} = \frac{s - q}{r - p}$$

by steps 1 and 2.

4. If (x, y) are the coordinates of a point, other than (p, q) , on the given line, then

$$y - q = \frac{s - q}{r - p} (x - p) \quad \text{by step 3 and algebraic laws.}$$

5. If (x, y) are the coordinates of the point (p, q) , then

$$y - q = \frac{s - q}{r - p} (x - p) \quad \text{since } 0 = 0 \text{ is a true statement.}$$

6. If (x, y) are the coordinates of any point on the given line then

$$y - q = \frac{s - q}{r - p} (x - p) \quad \text{by steps 4 and 5.}$$

7. Every point on the given line satisfies the equation

$$y - q = \frac{s - q}{r - p} (x - p) \quad \text{by step 6 and the definition (3.3) of a point satisfying an equation.}$$

8. The graph of the equation

$$y - q = \frac{s - q}{r - p} (x - p)$$

is a straight line

by Theorem 3.4.

9. The graph of the equation

$$y - q = \frac{s - q}{r - p} (x - p)$$

is the given line

by steps 7 and 8.

It should be noted that the graph of the equation

$$\frac{y - q}{x - p} = \frac{s - q}{r - p}$$

is *not* the given line through (p, q) and (r, s) , since the point (p, q) itself cannot be on the graph of this last equation.

Theorem 3.6:

An equation of the line through the point (p, q) , which line has slope m is

$$y - q = m(x - p).$$

Proof:

1. If (a, b) are the coordinates of a point, other than (p, q) , on the line through (p, q) , which line has slope m , then

$$m = \frac{b - q}{a - p} \quad \text{by the definition of slope.}$$

2. If (a, b) are the coordinates of a point, other than (p, q) on the line through (p, q) , which line has slope m , then

$$a - p = m(b - q) \quad \text{by step 1 and algebraic laws.}$$

3. If (a, b) are the coordinates of a point, other than (p, q) , on the line through (p, q) , which line has slope m , then that point satisfies the equation

$$y - q = m(x - p) \quad \text{by step 2 and the definition of a point's satisfying an equation.}$$

4. The point (p, q) satisfies the equation

$$y - q = m(x - p) \quad \text{since } 0 = 0 \text{ is a true statement.}$$

5. Every point on the straight line through (p, q) , which line has slope m satisfies the equation

$$y - q = m(x - p) \quad \text{by steps 3 and 4.}$$

6. The graph of the equation

$$y - q = m(x - p)$$

is a straight line by Theorem 3.4.

7. The graph of the equation

$$y - q = m(x - p)$$

is the line through the point (p, q) ,
which line has slope m by steps 5 and 6.

As a special case of Theorem 3.4, we have the very simple equation whose graph is the straight line identified by its slope and the point of intersection with the y -axis.

Corollary to Theorem 3.6:

An equation of the non-vertical line through the point $(0, b)$, which line has slope m is

$$y = mx + b.$$

Proof:

Theorem 3.4 with $p = 0$, $q = b$.

The equation given in Theorem 3.5 is called “the two-point form of an equation of a line”; the equation in Theorem 3.6 is called “the point-slope form of an equation of a line”; and the equation given in the corollary to Theorem 3.6 is called “the slope-intercept form of an equation of a line.” This last name—“slope-intercept”—is used since the number which is the y -coordinate of the point of intersection of a line with the y -axis is called the y -intercept of that line. Similarly, the number which is the x -coordinate of the point of intersection of a line with the horizontal axis is called the x -intercept of that line. Obviously, vertical lines have no y -intercept and horizontal lines have no x -intercept.

We have, in Theorems 3.5 and 3.6, found an equation for any given non-vertical line. To formalize and complete our set of theorems, we have

Theorem 3.7:

An equation of the vertical line through the point with coordinates (p, q) is

$$x = p$$

Proof:

All points on a vertical line have the same x -coordinate.

The theorems which we have proved in this section give us full and complete information about all equations involving the variables x and y of the form

$$Ax + By + C = 0,$$

where A , B , and C are constants, and A and B are not both 0. Conversely, these theorems tell us how to find an equation for any straight line. Thus, as a first step toward our double goal of classifying all equations from the point of view of the shapes of their graphs, and classifying all geometric entities from the point of view of the algebraic forms of their equations, we have tied together equations involving at most the variables x and y in which one or both of these variables actually appear to the first degree, with the set of all straight lines in a plane with a plane coordinate system. (This tying together is not a one-to-one correspondence, since one line may be the graph of many equations.)

Exercises

1. What is the graph of each of the equations:

a) $3 = 4$.

b) $-5 = 3$.

c) $x = 1$.

d) $y = -2$.

e) $x + 3 = 0$.

f) $y - 4 = 0$.

g) $2x + 5 = 0$.

h) $3y - 4 = 0$.

2. If r is the number $mp + q$, why is the point with coordinates (p, r) on the graph of $y = mx + b$?
3. Find the y -coordinates of the points on the graph of

$$2x - 3y + 4 = 0,$$
 whose x -coordinates are: $-100, -50, -10, -1, 0, 1, 2, 3, 10, 100, 1000$.
4. Find the y -coordinates of the points on the graph of

$$3x + 2y - 4 = 0,$$
 whose x -coordinates are: $-100, -50, -10, -1, 2, 100, 10000$.
5. Complete part II of the proof of Theorem 3.4.
6. Write out a formal, step-by-step proof of part I of Theorem 3.4.
7. Find the slope of the line joining

a) $(2, 1)$ and $(-4, 5)$.	c) $(3, 5)$ and $(-4, 5)$.
b) $(-1, 0)$ and $(3, -2)$.	d) $(2, 4)$ and $(2, 17)$.
8. Find an equation for each line in exercise 7.
9. What is an equation of the line through $(-1, 3)$ which line has slope:

a) -4 .	b) $3/4$.	c) 0 .	d) $-2/3$.	e) $3/2$.	f) $\sqrt{3}$.
-----------	------------	----------	-------------	------------	-----------------
10. What is an equation of the line with slope 4 which passes through:

a) $(0, 0)$.	d) $(1, 2)$.
b) $(-4, -2)$.	e) $(\sqrt{2}, -\sqrt{3})$.
c) $(-3, 10)$.	
11. Find the slope of the lines which are the graphs of the following equations, and sketch those graphs.

a) $y - 2 = \frac{1}{3}(x + 1)$.	h) $3x - 4y = 1$.
b) $y = 4x - 5$.	i) $3x - 4y + 5 = 0$.
c) $-y = 4x - 5$.	j) $6x - 8y + 10 = 0$.
d) $2x + y = 7$.	k) $\frac{2}{5}x - \frac{4}{5}y + 1 = 0$.
e) $y - 2x = 7$.	l) $\sqrt{2}x - y + 5 = 0$.
f) $3x - 4y = 720$.	m) $x + \sqrt{3}y - 3 = 0$.
g) $3x - 4y = 1$.	n) $\sqrt{5}x + \sqrt{7}y + \sqrt{11} = 0$.

3.5. ALGEBRAIC ASPECTS OF GEOMETRIC PROPERTIES OF LINES

From a purely geometric point of view, a question of interest is the determination of the angle between two lines. We will look only at the extreme cases, namely, the case in which the two lines do not intersect (which lines are parallel), and the case in which the two lines intersect at right angles (which lines are perpendicular).

Theorem 3.8:

Two non-vertical lines are parallel if and only if their slopes are equal.

Proof:

Let l_1 and l_2 be the two lines, with corresponding slopes m_1 and m_2 .

1. An equation of l_1 is $y = m_1x + b_1$,
where b_1 is some constant by Theorem 3.6.
2. An equation of l_2 is $y = m_2x + b_2$,
where b_2 is some constant by Theorem 3.6.
3. l_1 and l_2 will be parallel if and only
if they do not intersect by the definition of parallel.
4. l_1 and l_2 will be parallel if and only
if no point of the plane is on both
lines by step 3.
5. l_1 and l_2 will be parallel if and only
if no point of the plane satisfies
both of the equations
 $y = m_1x + b_1$ and $y = m_2x + b_2$ by step 4 and the definition of the
graph of an equation.
6. The equations $y = m_1x + b_1$ and
 $y = m_2x + b_2$ have no common
solution if and only if the equation
 $m_1x + b_1 = m_2x + b_2$ has no solu-
tion by step 5 and various algebraic laws.
7. The equation $m_1x + b_1 = m_2x$
 $+ b_2$ has no solution if, and only
if $m_1 = m_2$ by the definition of a solution of an
equation.
8. l_1 and l_2 are parallel if and only
if $m_1 = m_2$ by steps 3–7.

This theorem enables us to tell, almost at a glance, whether or not the graphs of two first-degree equations in the variables x and y (which graphs must be straight lines) are parallel to each other. If the equations are in one of the forms in which the slope is easily found (such as the point-slope or slope-intercept forms), then there is no computation involved, and we may merely compare the two slopes. If the equations are not in such a nice form, but require some manipulation in order to find the slopes of the lines, it is possible to discover that the lines are parallel (or not) without actually finding their slopes. Suppose we are presented with an equation of the form

$$Ax + By + C = 0.$$

From Theorem 3.2, we know that the graph of this equation is a straight line. If $B \neq 0$, then this line will be non-vertical. If $A \neq 0$, then this line will be non-horizontal. If A and B are both not 0, a particularly convenient way of ascertaining the slope of the line which is the graph of $Ax + By + C = 0$ is to note that this graph is also the graph of the equation

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

This last equation is the slope-intercept form of an equation of the line through $(0, -C/B)$, which line has slope $-A/B$. Thus the numbers A and B , appearing in the equation $Ax + By + C = 0$, determine the slope of the line which is the graph of this equation. Then any line which is the graph of an equation of the form

$$Ax + By + D = 0,$$

where D is any number different from C , will have the same slope as the line which is the graph of $Ax + By + C = 0$, and will thus be parallel to this first line. Further, if k is any number other than 0, the graph of

$$kAx + kBy + E = 0,$$

where E is any number, is the same as the graph of

$$Ax + By + \frac{E}{k} = 0,$$

so that the graph of $kAx + kBy + E = 0$ is also a line parallel to the graph of $Ax + By + C = 0$. For example, the graphs of

$$2x - 3y - 5 = 0;$$

$$2x - 3y + 4 = 0;$$

$$6x - 9y + 1 = 0;$$

are three mutually parallel lines.

A particularly convenient method exists for determining an equation of a straight line when the slope of the line, and the coordinates of one of its points, are known, and when the slope is a rational number. If the line has slope m/n , where m and n are integers, and if the line passes through the point (p, q) , then the point-slope form of an equation of this line will be (by Theorem 3.4)

$$y - q = \frac{m}{n} (x - p).$$

This equation is equivalent to—and thus has the same graph as—the equation

$$mx - ny = mp - nq.$$

We arrive at this second equation by saying that any line with slope m/n will have an equation whose left-hand side is $mx - ny$. Since the point with coordinates (p, q) must satisfy this equation, the right-hand side must be the result of substituting p and q for x and y , respectively, in the left-hand side; that is, the right-hand side must be the number $mp - nq$. For example, an equation of the line with slope $2/3$, through the point $(2, 7)$, will have left-hand side

$$2x - 3y$$

and right-hand side

$$2(2) - 3(7),$$

so that an equation of this line is

$$2x - 3y = -17,$$

or, if we prefer,

$$2x - 3y + 17 = 0.$$

For the case of perpendicular lines, we have:

Theorem 3.9:

Two lines, neither of which is vertical, are perpendicular to each other if and only if the product of their slopes is -1 .

Proof:

Let l_1 and l_2 be the two lines, with slopes m_1 and m_2 , respectively. Let the point of intersection, P , have coordinates (a, b) . (See Fig. 3.5.)

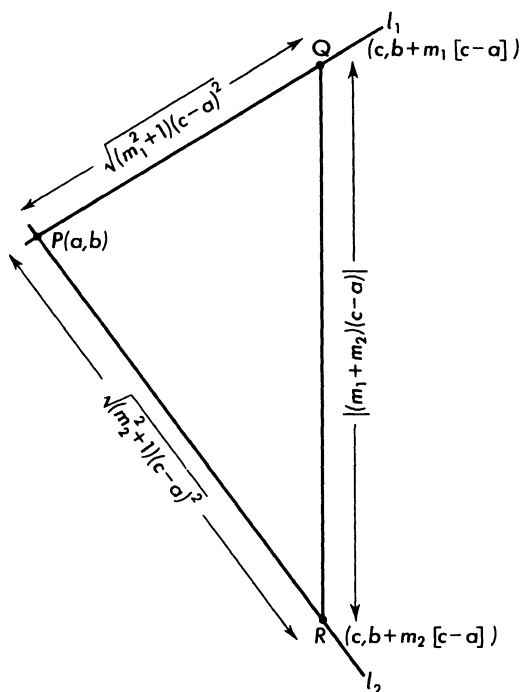


Fig. 3.5

1. If l_1 and l_2 are perpendicular to each other, their slopes have opposite signs.
2. If $m_1 m_2 = -1$, then m_1 and m_2 have opposite signs.
3. Without loss of generality, we may assume m_1 is negative and m_2 is positive by steps 1 and 2.

4. An equation of the line l_1 is

$$y - b = m_1(x - a)$$
 by Theorem 3.6.
5. An equation of the line l_2 is

$$y - b = m_2(x - a)$$
 by Theorem 3.6.
6. Let Q be a point on the line l_1 , to the right of P , whose x -coordinate is the number c , so that $c > a$.
7. Let R be the point on line l_2 whose x -coordinate is c .
8. The y -coordinate of Q is the number

$$b + m_1(c - a)$$
 by steps 4 and 6.
9. The y -coordinate of R is the number

$$b + m_2(c - a)$$
 by steps 5 and 7.
10. The length of the line segment joining P and R is

$$\sqrt{(m_1^2 + 1)(c - a)^2}$$
 by Theorem 3.3.
11. The length of the line segment PQ is

$$\sqrt{(m_1^2 + 1)(c - a)^2}$$
 by Theorem 3.3.
12. PQR is a right triangle if and only if

$$\begin{aligned} &(\text{length of } QR)^2 \\ &= (\text{length of } PR)^2 \\ &\quad + (\text{length of } PQ)^2, \end{aligned}$$
 by the Pythagorean theorem.
13. The length of the line segment joining Q and R is

$$|(m_1 - m_2)(c - a)|$$
 by steps 6–9 and Theorem 3.1.
14. PQR is a right triangle if and only if

$$\begin{aligned} &(m_1^2 + m_2^2 - 2m_1m_2)(c - a)^2 \\ &= (m_1^2 + m_2^2 + 2)(c - a)^2 \end{aligned}$$
 by steps 10–13.
15. PQR is a right triangle if and only if

$$m_1m_2 = -1$$
 by step 14.
16. The lines l_1 and l_2 are perpendicular to each other if and only if

$$m_1m_2 = -1$$
 by step 15 and the construction.

In words, this theorem says that two lines which have slopes are perpendicular if and only if their slopes are negative reciprocals of each other.

The theorems thus far presented provide the tools with which one may analyze many problems of plane geometry, as expressed in terms of a plane coordinate system.

Exercises

- Draw diagrams to illustrate the two cases (positive and negative slopes) of the proof of Theorem 3.8.
- Are the lines which are the graphs of the following pairs of equations parallel, or perpendicular, or neither:

a) $x + y - 1 = 0,$ $2x + 2y + 3 = 0.$ b) $3x - 4y + 5 = 0,$ $4x + 3y - 2 = 0.$ c) $2x - y + 3 = 0,$ $4x - 2y + 6 = 0.$	d) $x - y + 7 = 0,$ $x + y - 2 = 0.$ e) $3x - 5y + 4 = 0,$ $3x + 5y - 1 = 0.$ f) $2x - 3y + 4 = 0,$ $4x + 5y - 1 = 0.$
---	--
- Make a diagram to illustrate the proof of Theorem 3.7.
- Find an equation of the line through the point $(2, -1)$ which is parallel to the line with equation:

a) $3x - 4y + 7 = 0.$ b) $2x + y - 4 = 0.$ c) $x + 4 = 0.$	d) $y - 2 = 0.$ e) $x + y = 1.$
--	---
- Find an equation of the line through $(-3, 1)$ which is perpendicular to each of the lines in exercise 4.
- Show that all lines with equation

$$5x - 4y + k = 0,$$
 where k is any constant, are parallel to each other. Find the particular ones of those lines which pass through

a) $(0, 0).$ b) $(-1, 4).$	c) $(1, 5).$ d) $(3, -3).$
--	--
- Show that the points $(1, 2)$, $(6, 4)$, $(9, 9)$, and $(4, 7)$ are vertices of a parallelogram.
- Show that $Ax + By = Aa + Bb$ is an equation of the line through the point with coordinates (a, b) , which line is parallel to the line with equation $Ax + By + C = 0$.
- Show that $Bx - Ay = Ba - Ab$ is an equation of the line through the point with coordinates (a, b) , which line is perpendicular to the line with equation $Ax + By + C = 0$.
- What is the difference between “ A and B are not both 0” and “ A and B are both not 0”?

3.6. DISTANCE BETWEEN A POINT AND A LINE

One of the interesting problems of geometry is to find the distance between a given point and a given line. This distance is the length of a line segment drawn from the given point perpendicular to and ending at the given line.

For example, let us try to find the distance between the point $(2, 3)$ and the line joining the points $(1, -4)$ and $(9, 2)$. Using Theorem 3.6, an equation of the line through $(1, -4)$ and $(9, 2)$ is

$$y - 2 = \frac{6}{8}(x - 9).$$

Algebraic manipulation yields, as an equation of this line,

$$3x - 4y = 19.$$

The distance between $(2, 3)$ and the line with equation $3x - 4y = 19$ is the length of a line segment drawn from $(2, 3)$ perpendicular to, and ending at, the line with equation $3x - 4y = 19$. If we know the coordinates of the point at which this perpendicular meets the line $3x - 4y = 19$, then we can easily compute the distance between $(2, 3)$ and that point of intersection.

Since the line with equation $3x - 4y = 19$ has slope $3/4$, Theorem 3.9 tells us that the new line, through $(2, 3)$, will have slope $-4/3$. An equation of this new line will then be

$$4x + 3y = 17.$$

The point of intersection of the lines with equations

$$\begin{aligned} 3x - 4y &= 19 \\ \text{and } 4x + 3y &= 17 \end{aligned}$$

will have coordinates which satisfy both equations. We may find these numbers by solving the pair of equations simultaneously:

$3x - 4y = 19$ is true if and only if $9x - 12y = 57$ is true.

$4x + 3y = 17$ is true if and only if $16x + 12y = 68$ is true.

If $9x - 12y = 57$ and $16x + 12y = 68$ are both true, then $25x = 125$ is true, so x is the number 5.

If x is the number 5 and $3x - 4y = 19$ is true,

then $15 - 4y = 19$ is true, so that y is the number -1 .

If $x = 5$ and $y = -1$, then both equations are true.

We have thus found that the point $(5, -1)$ is the intersection of the line through $(2, 3)$, which line is perpendicular to the line through $(1, -4)$ and $(9, 2)$. The distance between $(2, 3)$ and the line through $(1, -4)$ and $(9, 2)$ is thus the distance between $(2, 3)$ and $(5, -1)$. This distance is

$$\sqrt{(5-2)^2 + (3+1)^2} = 5$$

units of distance.

In the process of solving this problem of finding the distance between a point and a line, we have also come across a connection between the purely algebraic process of finding the solution of a pair of first-degree simultaneous equations in two variables and the purely geometric process of determining the point of intersection of two straight lines: namely, the coordinates of the point of intersection will be the values of the variables x and y which make both of the equations simultaneously true.

In general, to find the distance between a point and a line, we have:

Theorem 3.10:

If A and B are both not 0, then the distance between the point with coordinates (p, q) and the line with equation

$$Ax + By + C = 0$$

is

$$\frac{|Ap + Bq + C|}{\sqrt{A^2 + B^2}} \text{ units of distance.}$$

Proof:

1. $Ax + By + C = 0$ is true if and only if

$$y = -\frac{A}{B}x - \frac{C}{B} \text{ is true} \quad \text{by the hypothesis that } B \neq 0.$$

2. The graph of $Ax + By + C = 0$ is a straight line

by Theorem 3.4.

3. The slope of the line which is the graph of $Ax + By + C = 0$ is $-A/B$

by steps 1 and 2 and Theorem 3.6.

4. Any line perpendicular to the line which is the graph of $Ax + By + C = 0$ will have slope B/A

by step 3, Theorem 3.9, and the hypothesis that $A \neq 0$.

5. The line through (p, q) which is perpendicular to the line which is the graph $Ax + By + C = 0$ has equation:

$$y - q = \frac{B}{A}(x - p) \quad \text{by step 4 and Theorem 3.6.}$$

6. The line through (p, q) which is perpendicular to the line which is the graph of $Ax + By + C = 0$ has equation:

$$Bx - Ay = Bp - Aq \quad \text{by step 5 and various algebraic laws.}$$

7. The point of intersection of the line which is the graph of

$$Ax + By + C = 0$$

and the line which is the graph of

$$Bx - Ay = Bp - Aq$$

has coordinates

$$\frac{B^2p - ABq - CA}{A^2 + B^2}, \quad \frac{A^2q - ABp - BC}{A^2 + B^2}$$

since these numbers satisfy both equations.

8. The distance between (p, q) and the line which is the graph of $Ax + By + C = 0$ is the distance between (p, q) and the point of intersection of the two lines in step 7

since these lines are perpendicular to each other.

9. The distance between (p, q) and the line which is the graph of $Ax + By + C = 0$ is

$$\sqrt{\left(p - \frac{B^2p - ABq - CA}{A^2 + B^2}\right)^2 + \left(q - \frac{A^2q - ABp - BC}{A^2 + B^2}\right)^2}$$

by steps 7 and 8.

10. The distance between (p, q) and the line which is the graph of $Ax + By + C = 0$ is

$$\sqrt{\left(\frac{A^2p + B^2p - B^2p + ABq + CA}{A^2 + B^2}\right)^2 + \left(\frac{A^2q + B^2q - A^2q + ABp + BC}{A^2 + B^2}\right)^2}$$

by step 9 and algebraic manipulation.

11. The distance between (p, q) and the line which is the graph of $Ax + By + C = 0$ is

$$\sqrt{\left(\frac{A(Ap + Bq + C)}{A^2 + B^2}\right)^2 + \left(\frac{B(Ap + Bq + C)}{A^2 + B^2}\right)^2}$$

by step 10 and algebraic manipulation.

12. The distance between (p, q) and the line which is the graph of $Ax + By + C = 0$ is

$$\sqrt{(A^2 + B^2) \frac{(Ap + Bq + C)^2}{(A^2 + B^2)^2}}$$

by step 11 and algebraic manipulation.

13. The distance between (p, q) and the line which is the graph of the equation $Ax + By + X = 0$ is

$$\frac{|Ap + Bq + C|}{\sqrt{A^2 + B^2}}$$

by step 12, algebraic manipulation, and the definitions of the square root and absolute value.

Exercises

1. Draw a diagram to illustrate the proof of Theorem 3.10.
2. Perform the algebraic manipulations mentioned in steps 10-13 of the proof of Theorem 3.10.
3. How does the conclusion of Theorem 3.10 emphasize the definition of a point's satisfying an equation of a straight line?
4. Find the distance between the origin and the line which is the graph of:

a) $x + y = 1$.	c) $y = 4x + 5$.
b) $2x - 5y = 4$.	d) $x = 2x - 3$.
5. Find the distance between the line which is the graph of $12x + 5y = 60$ and the point:

a) $(17, 5)$.	c) $(0, 0)$.
b) $(-2, -17)$.	d) $(5, 0)$.

3.7. THE GRAPH OF THE EQUATION $y = x^2$

In this section and the next, we consider the graph of the equation

$$y = x^2,$$

which graph is the collection of all the points in the plane whose y -coordinates are the squares of their x -coordinates. Since the y -coordinates are all the squares of some numbers, none of them will be negative, and thus none of the points of the graph of $y = x^2$ will lie below the horizontal axis.

But, in one glance at the equation, we have been able to determine a geometric property of the graph from an algebraic property of the set of all numbers. We will continue to examine the graph of this equation from an algebraic point of view, and then interpret our algebraic findings geometrically.

If the point with coordinates (p, q) is on the graph of $y = x^2$, then $q = p^2$. Since p^2 and $(-p)^2$ are the same number, the point with coordinates $(-p, q)$ would also be on the graph. Thus the point with coordinates (p, q) is on the graph of $y = x^2$ if and only if the point with coordinates $(-p, q)$ is on the graph of $y = x^2$. Speaking loosely, the points on this graph can be considered in pairs, which pairs of points have the same y -coordinate, and opposite x -coordinates. Since any one such pair of points consists of two points on opposite sides of the y -axis, equally far from the y -axis, and the same distance above the x -axis, the graph of $y = x^2$ can be considered as being made up of two parts, which parts are "mirror images" of each other in a mirror held perpendicular to the plane, running along the y -axis. The technical word for this situation is "symmetric," and the graph of $y = x^2$ is said to be symmetric with respect to the y -axis.

Insofar as our discovery of the geometric description of the graph of $y = x^2$ is concerned, the symmetry with respect to the y -axis means that we need only concern ourselves with the behavior of the graph in that part of the plane which lies to the right of the y -axis, since such behavior must always be matched by a similar behavior in that part of the plane to the left of the y -axis. Since we have already discovered that the graph of $y = x^2$ does not lie below the x -axis, it is sufficient to concentrate our attention on the first quadrant of the plane. This is particularly convenient for any manipulations with inequalities, since the coordinates of any point in the first quadrant are both positive numbers, and only negative numbers are particularly troublesome in handling inequalities.

Algebraically,

$$\text{if } 0 < a < 1, \text{ then } 0 < a^2 < 1.$$

Geometrically, any point on the graph of $y = x^2$ whose x -coordinate is between 0 and 1 also has y -coordinate between 0 and 1. Thus, all points on the graph of $y = x^2$ which lie in the vertical strip between the lines with equations $x = 0$ and $x = 1$ are contained within the square formed by these two vertical lines and the horizontal lines with equations $y = 0$ and $y = 1$. (See Fig. 3.6.)

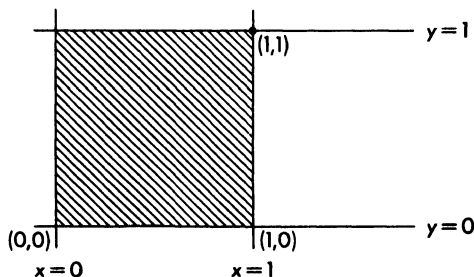


Fig. 3.6

Algebraically,

if $a > 1$, then $a^2 > 1$.

Geometrically, any point on the graph of $y = x^2$ which is to the right of the vertical line with equation $x = 1$ is also above the horizontal line with equation $y = 1$.

These last two paragraphs show us not only where in the first quadrant of the plane the graph of $y = x^2$ may lie, but also where this graph cannot be. That is, there are no points of the graph lying in that part of the vertical strip between the lines with equations $x = 0$ and $x = 1$ which part is above the horizontal line with equation $y = 1$, and there are no points of the graph lying to the right of the vertical line with equation $x = 1$ which are also below the horizontal line with equation $y = 1$.

Almost incidentally, we note that the points $(0, 0)$ and $(1, 1)$ are on the graph of $y = x^2$. Since it is almost always easy to do arithmetic with the numbers 0 and 1, little effort is involved in discovering those points on this graph having x -coordinates 0 or 1, and we have done this arithmetic. There is a temptation to find a few more specific points on this graph (such as the points with coordinates $(2, 4)$, $(3, 9)$, $(4, 16)$, and so on.) and then to connect these points with what appears to be an attractive smooth "curve." Extending such a point-plotting process to any graph would require that one plot "a sufficient number of points to bring out the essential features of the graph." The only thing wrong with such a process is that we do not know what features of any graph *are* essential until after we have the graph, and that we do not know how many points are a "sufficient number" to bring out these unknown features. That is, the assumption that the graph of a given equation "behaves nicely" between two specific points is only an assumption and can be justified only by investigating algebraic facts about the set of all numbers.

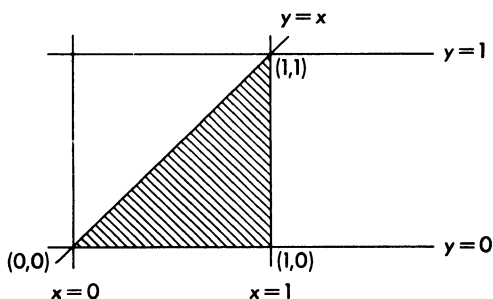


Fig. 3.7

The algebraic fact that every number has a square (which fact is embodied in the multiplication law M-1) implies the geometric fact that there will be a point on the graph of $y = x^2$ for every possible x -coordinate. In other words, every vertical line in the plane will intersect the graph of $y = x^2$.

If $0 < a < 1$, then application of the ordering law O-4a yields $0 < a^2 < a$. This algebraic fact implies the geometric fact that, inside the square bounded

by the lines with equations $x = 0$, $x = 1$, $y = 0$, and $y = 1$, a point on the graph of $y = x^2$ with coordinates (a, a^2) is below the point with coordinates (a, a) . In other words, inside this square the graph of $y = x^2$ lies below the diagonal of the square, which diagonal is part of "the 45-degree line," an equation of which is $y = x$. Conversely, no point on the graph of $y = x^2$ which is within this square lies above or on that diagonal. (See Fig. 3.7.)

If $b > 1$, then the ordering law O-4a yields $b^2 > b$. This algebraic fact gives us the geometric fact that, in that part of the plane which is to the right of the vertical line with equation $x = 1$, the graph of $y = x^2$ lies above the line with equation $y = x$.

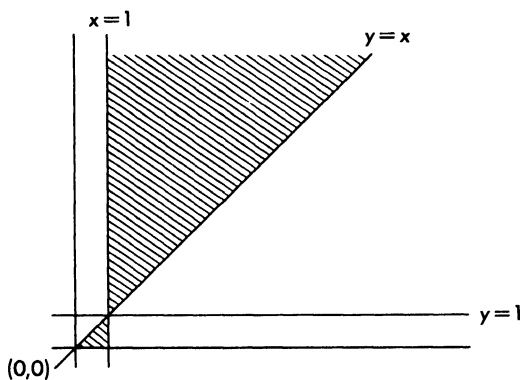


Fig. 3.8

We have discovered, through the use of inequalities:

- (1) In the first quadrant of the plane, the graph of $y = x^2$ includes the points $(0, 0)$ and $(1, 1)$.
- (2) Between $(0, 0)$ and $(1, 1)$ the graph of $y = x^2$ is below the graph of $y = x$.
- (3) To the right of $(1, 1)$ the graph of $y = x^2$ is above the graph of $y = x$.

In order to justify the assumption that the graph of $y = x^2$ is actually a collection of points which can be called a "curve" we must show that points on this graph which have almost the same x -coordinates do not have too different y -coordinates. This last statement is vague and imprecise, even though it may be somewhat meaningful. We are, to a certain extent, dependent on the idea of a "continuous curve with no holes or gaps." In the next section we will try to make precise just what is meant by a hole or a gap.

In this section, we have used some tools for translating algebraic facts about an equation involving the variables x and y into geometric facts

about the graph of that equation. The first of these tools brought up the idea of symmetry. Formally, we have

Definition 3.7: A graph is *symmetric with respect to the y -axis* (horizontally symmetric), provided that a point with coordinates (p, q) is on the graph, if and only if the point with coordinates $(-p, q)$ is on the graph.

Clearly, any equation in which the variable x occurs to powers which are even integers only will have a graph symmetric with respect to the y -axis.

Definition 3.8: A graph is *symmetric with respect to the x -axis* (vertically symmetric), provided that a point with coordinates (p, q) is on the graph, if and only if the point with coordinates $(p, -q)$ is on the graph.

From this definition, any equation in which the variable y appears to powers which are even integers only will have a graph which is symmetric with respect to the x -axis.

Each of these concepts of symmetry is connected with the algebraic fact that the square of a number is the same as the square of the negative of that number. A similar algebraic fact is that the square root of a negative number is not a (real) number.

As an illustration of the geometric use which can be made of the fact that the square of a number is never negative, and the fact that one cannot take the square root of a negative number, let us consider the graph of the equation

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

This graph is the same as the graph of the equation

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

and as the graph of the equation

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

Consequently, no point on the graph of $x^2 + y^2 = 1$ can have an x -coordinate whose square is greater than 1, nor a y -coordinate whose square is greater than 1. Thus no point on the graph of $x^2 + y^2 = 1$ is outside the square bounded by the horizontal lines with equations $y = -1$ and $y = 1$, and the vertical lines with equations $x = -1$ and $x = 1$.

The discovery of which numbers cannot be x - or y -coordinates of points on the graph of an equation is known as determining the horizontal or vertical extent of that graph. The *horizontal extent* of the graph of an equation is the set of numbers which are x -coordinates of all the points on that graph, and the *vertical extent* of the graph of an equation is the set of numbers which are y -coordinates of the points on that graph. In order to find either extent, it is sufficient to find those numbers which cannot be coordinates of points

on the graph, which means finding those numbers which would require that one perform either of the two forbidden algebraic operations: taking the square root of a negative number and division by zero.

For example, the graph of the equation

$$x^2 = y + 2$$

cannot have any point whose y -coordinates are less than -2 , but can have any number as x -coordinate of a point. The vertical extent of this graph is the interval:

$$\text{the set of all } y, y > -2,$$

and the horizontal extent is the set of all numbers.

As another example, we consider the graph of

$$4x^2 + 9y^2 = 36.$$

If (p, q) is on this graph, then $4p^2 + 9q^2 = 36$. However, if $4p^2 + 9q^2$ is the number 36, then $4(-p)^2 + 9q^2$, $4p^2 + 9(-q)^2$, and $4(-p)^2 + 9(-q)^2$ are also 36, so that the points $(-p, q)$, $(p, -q)$, and $(-p, -q)$ are also on the graph of $4x^2 + 9y^2 = 36$. This graph is then symmetric with respect to the origin, with respect to the x -axis and with respect to the y -axis.

Further, $4x^2 + 9y^2 = 36$ will be true if and only if the following equations are true:

$$\begin{aligned} 4x^2 &= 36 - 9y^2; & 9y^2 &= 36 - 4x^2; \\ 4x^2 &= 9(4 - y^2); & 9y^2 &= 4(9 - x^2); \\ \frac{4}{9}x^2 &= 4 - y^2; & \frac{9}{4}y^2 &= 9 - x^2. \end{aligned}$$

Since the squares of (real) numbers are never negative, $\frac{4}{9}x^2 = 4 - y^2$ cannot be true if $4 - y^2 < 0$, that is, if $y^2 > 4$. Thus no point on the graph of $4x^2 + 9y^2 = 36$ can have a y -coordinate greater than 2 or less than -2 . Similarly, $\frac{9}{4}y^2 = 9 - x^2$ cannot be true if $9 - x^2 < 0$, that is, if $x^2 > 9$. Thus no point on the graph of $4x^2 + 9y^2 = 36$ can have an x -coordinate greater than 3 or less than -3 .

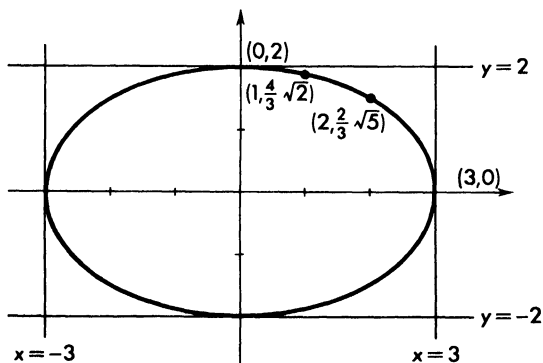


Fig. 3.9

The graph of $4x^2 + 9y^2 = 36$ is consequently found to lie entirely within the rectangle bounded by the (straight lines which are the) graphs of the equations $x = -3$, $x = 3$, $y = -2$, and $y = 2$.

The computation of the coordinates of a few points in the first quadrant and use of the symmetries discovered gives the graph of Fig. 3.9.

Exercises

1. In which quadrants do the graphs of the following equations lie:

- | | |
|------------------|---------------------|
| a) $y = x^3$. | d) $x^2 = y^2$. |
| b) $y^2 = x$. | e) $y^3 = x^4$. |
| c) $y^2 = x^3$. | f) $y = \sqrt{x}$. |

2. For what vertical strips in the first quadrant are the graphs of the following equations above or below the graph of $y = 0$; above or below the graph of $y = x$:

- | | |
|---------------------|---------------------|
| a) $y = x^3$. | d) $y = \sqrt{x}$. |
| b) $y = x^2 - x$. | e) $y = 3x + 5$. |
| c) $y = x^3 - 4x$. | f) $y = 4 - x^2$. |

3. Are the graphs of the following equations symmetric with respect to the x - or y -axes?

- | | |
|----------------|----------------------|
| a) $y = x^3$. | d) $x^2 = y^3$. |
| b) $y^2 = x$. | e) $x^2 + y^3 = 7$. |
| c) $y = x^4$. | f) $x^4 + y^2 = 1$. |

g) $y^2 = x^6 - 4x^4 + 7x^2 - 2x + 1$.

4. Find the horizontal and vertical extent of the graphs of:

- | | |
|----------------------|-----------------------|
| a) $x^2 + y^2 = 4$. | d) $y = x + 1/x$. |
| b) $y = x^2 + 1$. | e) $y^2 = x^2 - 2x$. |
| c) $y^2 = x - 2$. | f) $y = 9 - x^2$. |

5. Prove the theorem: If $0 < a < 1$, then $0 < a^2 < a$.

6. Assuming that the graphs are "smooth curves", sketch on the same plane the graphs of:

$$y = x^2, \quad y = x^3, \quad y = x^4, \quad \text{and} \quad y = x^5,$$

giving particular care to that part of those graphs in the vertical strip between the lines with equations $x = -2$ and $x = 2$.

3.8. FURTHER INVESTIGATION OF THE GRAPH OF $y = x^2$

In this section we will look closely at the idea of a "hole" in a "continuous curve." Suppose (a, a^2) are the coordinates of a point on the graph of the equation

$$y = x^2.$$

How can we investigate the location of points whose coordinates are $(a + h, (a + h)^2)$, where h is some number "close to" 0—either positive or negative? The first stumbling-block to the investigation is the phrase

"close to." Substitution of the word "small" for "close to 0" leaves us in no better position, since the definition of "small" is purely subjective. After all, one dollar is "small" in relation to the national debt, but "large" to a child who must save pennies in order to obtain a dollar.

What we are really concerned with is the closeness of the points with coordinates $(a + h, (a + h)^2)$, for different values of the variable h , to the point (a, a^2) . Because of our convention the variable h has, as range, the set of all numbers, but our major interest lies in the part of the range of this variable which is near 0. The x -coordinates of the two points (a, a^2) and $(a + h, (a + h)^2)$ differ by the number $|h|$ and the y -coordinates of these two points differ by the number $|(a + h)^2 - a^2|$, which is the number $|2ah + h^2|$.

Then the imprecise geometric idea that, for small values of the variable h , the point with coordinates $(a + h, (a + h)^2)$ is close to the point (a, a^2) may be translated into the equally imprecise, but purely algebraic idea that small values of the variable h make the number $|2ah + h^2|$ small. This algebraic statement certainly seems reasonable, but we are still behind the barrier of the word "small."

We avoid the vagueness of the words "small" or "close to" in the usual manner of mathematicians: we do not use these words! What we will do is consider the following question: What could the difference between the number a and the number $a + h$ be in order that the number $(a + h)^2$ be within some desired closeness to the number a^2 ? In terms of inequalities, this question may be restated:

If e is the desired closeness of $(a + h)^2$ to a^2 , what is a solution of the inequality

$$|2ah + h^2| < e.$$

in which a is a constant and h is a variable?

It should be noted that what we are after is merely a solution to this inequality, rather than *the* solution.

The inequality

$$|2ah + h^2| < e$$

is true if and only if the inequality

$$|h| \cdot |2a + h| < e \text{ is true}$$

by Theorem 2.4, which states that the absolute value of the product of two numbers is the product of their absolute values. Since we are doing all our investigating of the graph of the equation $y = x^2$ in the first quadrant of the plane, the number a , in the inequality which we are trying to solve, will be positive. We may begin our partial solution of this inequality by considering only those values of the variable h , for which the number $2a + h$ is positive. But, $2a + h$ will be positive if and only if $h > -2a$. Since we find intervals convenient to deal with and since our concern lies

with part of the range of the variable h which includes the number 0, we will seek a solution of the given inequality inside the interval:

$$\text{the set of all } h, \quad -2a < h < 2a$$

(which may also be described as:

$$\text{the set of all } h, \quad |h| < 2a).$$

For those values of h in this interval, application of the ordering law O-3 to the inequality

$$-2a < h < 2a$$

yields the inequality

$$0 < 2a + h < 4a.$$

Thus all values of h in the interval:

$$\text{the set of all } h, \quad |h| < 2a$$

satisfy the two statements:

$$|2a + h| = 2a + h$$

from the definition of absolute value

and

$$|2a + h| < 4a$$

from the fact that, for these values of the variable h , $2a + h < 4a$.

Consequently, for all values of the variable h in the interval:

$$\text{the set of all } h, \quad |h| < 2a,$$

we will have

$$|h| \cdot |2a + h| < 4a|h|.$$

Because of this last statement, any value of the variable h , in the interval:

$$\text{the set of all } h, \quad |h| < 2a,$$

which value is a solution of the inequality

$$4a|h| < e,$$

will also be a solution of the inequality

$$|h| \cdot |2a + h| < e.$$

The solution of the new inequality $4a|h| < e$ is the interval:

$$\text{the set of all } h, \quad |h| < e/4a$$

(which may also be described as:

$$\text{the set of all } h, \quad -e/4a < h < e/4a).$$

We may then conclude that a solution of the inequality

$$|h| \cdot |2a + h| < e$$

is the smaller of the two intervals:

$$\begin{array}{l} \text{the set of all } h, \quad |h| < 2a \\ \text{and the set of all } h, \quad |h| < e/4a. \end{array}$$

For example, if a is the number 3, and e is the number $1/2$, then our original question is: How close to the number 3 should $3 + h$ be taken in order that the number $(3 + h)^2$ be within $1/2$ unit of 9? In terms of inequalities, we have to find a solution to:

$$|6h + h^2| < 1/2.$$

This inequality will be true if and only if the inequality

$$|h| \cdot |6 + h| < 1/2$$

is true.

If we consider only those values of the variable h in the interval:

$$\text{the set of all } h, \quad |h| < 6,$$

then, for these values of h , we have

$$|h| \cdot |6 + h| < 12|h|.$$

Consequently, any value of h in the interval:

$$\text{the set of all } h, \quad |h| < 6,$$

which value is a solution of the inequality

$$12|h| < 1/2$$

will also be a solution of the original inequality

$$|h| \cdot |6 + h| < 1/2.$$

The solution of the inequality $12|h| < 1/2$ is the interval:

$$\text{the set of all } h, \quad |h| < 1/24$$

(which may also be expressed as:

$$\text{the set of all } h, \quad -1/24 < h < 1/24),$$

which interval is contained entirely within the interval:

$$\text{the set of all } h, \quad -6 < h < 6.$$

We conclude, then, that a solution of the inequality

$$|h| \cdot |6 + h| < 1/2$$

is the interval:

$$\text{the set of all } h, \quad -1/24 < h < 1/24.$$

This algebraic result may then be interpreted geometrically by saying that if we do take the number $3 + h$ within $1/24$ unit of the number 3, then the number $(3 + h)^2$ will certainly be within $1/2$ unit of the number 9.

In general, for any fixed positive number a , if the number ϵ is given, there *will* be values of the variable h which achieve any desired closeness of $(a + h)^2$ to a^2 , since there *are* values of h which are sufficiently close to 0 to achieve this goal, namely, those values of h which are in both the intervals:

$$\begin{aligned} &\text{the set of all } h, \quad -2a < h < 2a \\ \text{and} \quad &\text{the set of all } h, \quad -\epsilon/4a < h < \epsilon/4a. \end{aligned}$$

These manipulations with inequalities have far-reaching consequences for the graph of $y = x^2$. We can now see that there are no points of this graph which are "off by themselves." Each point of the graph has other points of the graph close by, regardless of how small a distance we take to indicate closeness. Thus, it is justifiable to call the graph a curve and to draw an approximation to the graph by drawing some curve from the point $(0, 0)$ through the point with coordinates $(1, 1)$, which curve lies below the line with equation $y = x$ between $(0, 0)$ and $(1, 1)$ and lies above that line to the right of $(1, 1)$.

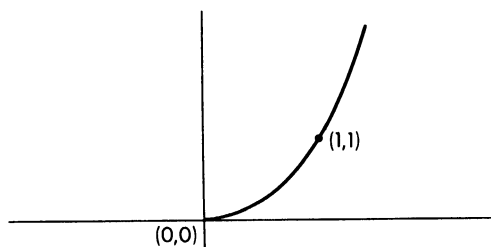


Fig. 3.10

It should be realized that the curve drawn is only an approximation to the actual graph, but that at least the shape of the graph is suggested. The approximation may be improved by finding more specific points on the graph, which is purely an arithmetic process. One may conveniently construct a table of corresponding values of the variables x and y for the equation $y = x^2$, and use this table to plot some of the points which lie on the graph. For example,

x	y
2	4
3	9
4	16
5	25

is such a table. This table shows that the points with coordinates $(2, 4)$, $(3, 9)$, $(4, 16)$, and $(5, 25)$ are on the graph of $y = x^2$. We may plot these

points, together with $(0, 0)$ and $(1, 1)$, and provide ourselves with a better guide than we had in drawing an approximation to the graph of $y = x^2$, but it should be realized that whether we have 6 points or 600 points, the fact which allows us to connect the plotted points with what we hope is a continuous curve, and to use this curve as an approximation to the actual graph, is the algebraic fact embodied in the manipulations with inequalities which we have performed.

Finally, having found an approximation to that part of the graph of $y = x^2$, which part lies in the first quadrant, we use the symmetry with respect to the y -axis to complete our approximation by drawing, in the second quadrant, a curve which is symmetric to the one which we have drawn in the first quadrant.

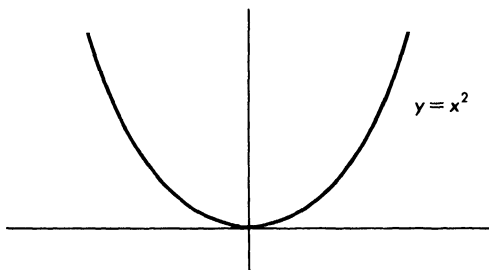


Fig. 3.11

Exercises

1. Draw the graph of $y = 3x + 5$.

2. Solve the inequality

$$8 - 1/2 < 3x + 5 < 8 + 1/2.$$

3. Draw the graph of

$$y = \frac{(3x + 5)(x - 1)}{x - 1}.$$

4. How close to 1 should the x -coordinates of the points on the graph of the equation in exercise 3 be taken in order to be sure that those points will have y -coordinates within $1/2$ unit of 8?

5. On the same plane draw the graphs of the inequalities

$$8 - 1/2 < y < 8 + 1/2$$

and

$$1 - 1/6 < x < 1 + 1/6.$$

Then draw the graph of $y = 3x + 5$.

What conclusions can be drawn from the result of exercise 4, and these three graphs?

6. Find a solution of the inequality

$$15 < x^2 < 17.$$

7. Find a solution of the inequality

$$| (4 + h)^2 - 16 | < 1.$$

8. How close to 4 should the x -coordinates of points on the graph of $y = x^2$ be taken to be sure that the y -coordinates of those points will be within 1 unit of 16?

Sketch approximations to the graphs of:

- 9. $y = x^3$.
- 10. $y = x^4$.
- 11. $y = x^k$, where k is an even natural number.
- 12. $y = x^p$, where p is an odd natural number.

chapter four



FUNCTIONS

4.1. FUNCTIONS

In the preceding chapter, we examined relationships between sets of points in a plane with a coordinate system and equations involving the variables x and y . We found that any one equation involving these two variables had a graph, which graph consisted of some points, which points were identified by their coordinates, which coordinates were ordered pairs of numbers. This relating a statement about two variables with a collection of ordered pairs of numbers leads into almost all of what is ordinarily called “the calculus.”

From the definition (2.1), a variable is a symbol standing for a number taken from a specific set of numbers, which set of numbers is called the range of that variable. If we are dealing with some statement (namely, an equation) involving two variables, then we are necessarily dealing with two sets of numbers—the ranges of the two variables. If we consider the fact that such a statement relates the two variables to each other, then we are necessarily considering some sort of relationship between two sets of numbers.

Variables are certainly more tangible things, in *written appearance*, than the idea of a correspondence between two sets of numbers, but too heavy a dependence on algebraic symbolism frequently obscures the real meaning of much of mathematics. Let us look at something we have already seen, but do our looking from a slightly different vantage point.

A set of ordered pairs of numbers—in appearance a set of things of the form (p, q) , where p and q are numbers, such as the set consisting of the three ordered pairs:

$$(1, 1), (2, 1), \text{ and } (1, 2),$$

or such as the set of all ordered pairs which are the coordinates of the points on the 45-degree line of a plane—is a set of ordered pairs of numbers, and, so far, nothing more. If we choose to interpret the ordered pairs as coordinates of points in a plane with a plane coordinate system, then *we* are

looking at the given set of ordered pairs of numbers from a geometric standpoint, and *we* are identifying the set of ordered pairs of numbers with some set of points in the plane. But, if we look at the set of ordered pairs of numbers from the standpoint of the numbers involved—that is, algebraically—then that set of ordered pairs of numbers is merely a set of things offered for our contemplation and consideration.

In the beginning of such contemplation of a set of ordered pairs of numbers, there are three basic questions:

- 1) Where did the numbers which are the first elements of the ordered pairs come from?
- 2) Where did the numbers which are the second elements of the ordered pairs come from?
- 3) How did the paired numbers become paired?

It should be recalled, at this point, that an ordered pair of numbers is a pair, written in the form (a, b) , for which the number a is called the first element of the ordered pair and the number b is called the second element of the ordered pair.

There are many different kinds of sets of ordered pairs of numbers, but we will restrict our attention to one specific kind. For any set of ordered pairs of numbers, we have a definition of convenience:

Definition 4.1: A *relation* is a set of ordered pairs of numbers.

For the specific kinds of sets of ordered pairs of numbers which we will consider, we have

Definition 4.2: A *function* is a set of ordered pairs of numbers, in which set there are no two ordered pairs with the same first element and different second elements.

(We have defined what is often called a “single-valued” function.)

The definitions given are definitions of the mathematical entities known as “a relation” and “a function,” and have no *necessary* connection with the meanings of the English words “relation” and “function.”

We can now look at the three basic questions asked above for the kinds of sets of ordered pairs of numbers which we will consider. For the first question, we have

Definition 4.3: Given a function, the *domain* of that function is the set of all numbers which are the first elements of the ordered pairs which make up the function.

For the second question:

Definition 4.4: Given a function, the *range* of that function is the set of all numbers which are the second elements of the ordered pairs which make up the function.

In this second definition, we are making a new use of the word "range," distinct from, but similar to, the concept of the range of a variable. The difficulty of using the same word in two different senses may be lessened by inserting the words "of values," so that we speak of the range of values of a function, but of the range of a variable.

For an example, let us consider the extremely simple function which is the set of the following five ordered pairs:

$$(0, 3), (1, -1), (2, 4), (3, 3), \text{ and } (4, 5).$$

This collection of ordered pairs of numbers is a function, since no two of the ordered pairs have the same first element. The domain of this function is the set of numbers: 0, 1, 2, 3, and 4. The range of values of this function is the set of numbers: 3, -1, 4, and 5.

As a less simple example, let us consider the set of ordered pairs which are the coordinates of all the points on the graph of the equation $y = x^2$. This set of ordered pairs of numbers is a function because no two of the ordered pairs have the same first element (since no two points on the graph of $y = x^2$ have the same x -coordinate). The domain of this function is the set of all numbers. The range of values of this function is the set of all non-negative numbers.

As an example of a relation which is not a function, consider the set of ordered pairs which are the coordinates of all points on the graph of the equation $y^2 = x$. This set of ordered pairs of numbers is not a function since there are, in this set, ordered pairs with the same first element but different second elements—for example, the ordered pairs:

$$(4, -2) \text{ and } (4, 2).$$

For the third of the basic questions (How did the paired numbers become paired?), we have

Definition 4.5: Given a function, a *rule* of that function is a set of instructions which, when applied to any one number in the domain, yields that one number in the range of values which is paired with the chosen number in the domain to constitute an ordered pair of the function.

It should be noted that the use of the indefinite article indicates that a given function may have more than one rule. For example, the extremely simple function which is the set of ordered pairs:

$$(0, 0), (-1, 1), \text{ and } (1, 1),$$

has, as one of its rules, the instruction: "square," and, as another of its rules: "raise to the power 4." if this function is listed by means of a table:

first element	second element
-1	1
0	0
1	1

then a rule would also be "look at the table."

One should also note that we are not answering these three basic questions, but merely providing names for the characteristics involved, so that we may apply these names in specific cases.

Those functions which consist of a finite number of ordered pairs may be given merely by listing the ordered pairs, but those functions which consist of infinitely many ordered pairs of numbers can only be given by some scheme which tells us which ordered pairs are present. By far the most usual method of presentation is telling the domain and a rule. Once the domain and a rule of a function are known, then the range of values is completely determined, although it may involve some labor to find that set of numbers. Thus, in order to know completely a function, it is sufficient to know the domain of this function and a rule. For the simple example above, with the domain the set of numbers: $-1, 0$, and 1 , and a rule: "square," the range of values must consist of the set of numbers: 0 and 1 , since only these two numbers result from applying the given rule separately to each number in the domain.

Since the domain and range of values of a function are sets of numbers, we may well have two variables representing numbers in those sets. These variables are deserving of special names, which are given them in

Definition 4.6: Given a function, an *independent variable* of that function is a variable whose range is the domain of the given function.

Definition 4.7: Given a function, a *dependent variable* of that function is a variable whose range is the range of values of the given function.

It is usual, although by no means necessary, to use the symbol x as an independent variable of functions, and the symbol y as a dependent variable of functions. With this usage, it is then reasonable to say that a dependent variable is a *function of* an independent variable, where we are using the phrase "function of" in an English rather than a mathematical sense. This manner of speaking indicates that to each value of an independent variable there will correspond (be paired) one and only one value of a dependent variable of that function.

From the definition of the word "function," we are able to see a function in the graph of an equation involving the variables x and y if we can see that the set of coordinates of the points on the graph actually forms a set of ordered pairs of numbers, which set of ordered pairs of numbers is a function. In this sense, the equation itself is said to give a function, but, of course, the equation we start with is an equation, and only that, and it is really we who find the function.

In order to have some way of writing down various properties of functions, we introduce the following notation: A function will be denoted by a letter or combination of letters, followed by blank parentheses:

$f()$, $g()$, $h()$, $F()$, $G()$, $H()$, $\{f + g\}()$, and so on.

If p is a number in the domain of the function $f()$, then the number paired with p in the function $f()$ —that is, that number in the range of values of $f()$ which results from applying a rule of $f()$ to the number p —will be denoted by the symbol $f(p)$ (read “ f of p ”). If we are using the variable x to represent any number in the domain of the function $f()$ —that is, as an independent variable of $f()$ —then the expression $f(x)$ will represent any number in the range of values of $f()$. In other words, the symbol $f(x)$ represents that number which is the second element of that ordered pair of the function whose first element is the number x . That is, we may say that the function $f()$ consists of all ordered pairs of the form

$$(x, f(x)),$$

where x represents any number in the domain of $f()$.

Exercises

Which of the following collections of ordered pairs of numbers are functions:

1. $(1, 1), (2, 1), (3, 1), (4, 1)$
2. $(1, 1), (2, 2), (3, 3), (4, 4)$.
3. $(1, 1), (1, 2), (2, 1), (3, 2)$.
4. The set of coordinates of all the points on the graph of $y = 5$.
5. The set of coordinates of all the points on the graph of $x = 3$.
6. The set of coordinates of all the points on the graph of $x = 2y$.
7. The set of coordinates of all the points on the graph of $x^2 + y^2 = 1$.
8. The set of coordinates of all the points on the graph of $x^2 + y^2 = 1$; $y \geq 0$.
9. The set of coordinates of all the points on the graph of $x^2 + y^2 = 1$; $x \geq 0$.
10. Of the functions identified in exercises 1-9, what are their domains and ranges. Find a rule for each.
11. If the domain of a function is the interval:

$$\text{the set of all } x, \quad 1 < x < 2,$$
 and a rule of that function is “square,” what is the range of values?
12. Let r and w be variables, whose ranges are the set of all numbers, and for which

$$2r + 3w = 6.$$

- a) If r is an independent variable of the function $g()$ and w is a dependent variable of $g()$, find the domain and range of $g()$, and a rule of $g()$. Find an equation involving the variables x and y whose graph consists of a set of points whose coordinates are $g()$.
- b) If w is an independent variable of the function $h()$ and r is a dependent variable of $h()$, find the domain and range of $h()$, and a rule of $h()$. Find an equation involving the variables x and y whose graph consists of a set of points whose coordinates are $h()$.

13. If $f()$ is a function whose domain is the interval:

the set of all t , $-1 \leq t \leq 1$,

and a rule of $f()$ is given by the equation

$$f(v) = v^2,$$

what set of ordered pairs is $f()$?

14. The domain of the function $g()$ is the interval:

the set of all z , $0 \leq z \leq 2$.

An independent variable of $g()$ is x and a dependent variable of $g()$ is y . A rule of $g()$ is given by the equation

$$g(w) = 2w + 1.$$

What is the range of the variable y ?

15. The functions $f()$, $g()$, and $h()$ all have a rule which says "square."
The domain of $f()$ is the interval:

the set of all r , $-2 \leq r \leq 0$;

the domain of $g()$ is the interval:

the set of all s , $0 \leq s \leq 1$;

and the domain of $h()$ is the interval:

the set of all t , $-1 \leq t \leq 2$.

What ordered pairs do

- a) $f()$ and $g()$ have in common?
- b) $f()$ and $h()$ have in common?
- c) $g()$ and $h()$ have in common?

16. The domain of the functions $F()$, $G()$ and $H()$ is the interval:

the set of all x , $1 < x < 2$.

A rule of $F()$ is given by

$$F(r) = 2r + 5;$$

a rule of $G()$ is given by

$$G(s) = s^2;$$

and a rule of $H()$ is given by

$$H(t) = 4t^2 + 2t + 7.$$

What are the ranges of values of these three functions?

4.2. GRAPHS OF FUNCTIONS

Since we came to the concept of function by looking at collections of points in a coordinate plane, let us now reverse the procedure.

Definition 4.8: The *graph of a function* is the collection of those points in a coordinate plane whose coordinates are the ordered pairs comprising the function.

This definition says that the collection of those points, and only those points, whose coordinates make up the set of ordered pairs of numbers,

which set is the function $f()$, is the graph of $f()$. We will use the same examples as in the preceding section. The graph of the function which is the set:

$$(0, 3), (1, -1), (2, 4), (3, 3), \text{ and } (4, 5),$$

consists of precisely the five points which have these ordered pairs of numbers as their coordinates. The graph of the function, which is the set of ordered pairs, which are the coordinates of the points on the graph of the equation $y = x^2$, is the collection of points which is the graph of the equation $y = x^2$. This last example illustrates the principle: When a function is given by an equation involving the variables x and y , in which x is an independent variable and y is a dependent variable, and in which the range of the variable x includes the domain of the function, and the range of the variable y includes the range of values of the function, and in which the horizontal extent of the graph of the equation is the domain of the function and the vertical extent of the graph of the equation is the range of values of the function, then the graph of the given equation and the graph of the given function are the same. However, if all of these conditions are not met, then the graph of an equation may contain more points than the graph of a function which that equation gives. For example, consider the function $g()$, which is the collection of those ordered pairs of numbers, which are the coordinates of those points on the graph of $y = 3x$ which are in the first quadrant of the plane. The domain of $g()$ is the set of all positive numbers and the range of values of $g()$ is also the set of all positive numbers. The graph of $g()$ is the set of those points on the graph of $y = 3x$ which lie in the first quadrant, but is not all of the graph of the equation $y = 3x$.

From the definition of function, no two distinct points on the graph of a function can have the same x -coordinate. Geometrically, any vertical line in the plane can intersect the graph of a function at most once. Speaking loosely, this idea means that the graph of a function never "doubles back" on itself. For example, the curve shown in Fig. 4.1a can be the graph of a function (and, in fact, this curve is the graph of precisely that function made up of the ordered pairs which are the coordinates of the points on this curve), while the curve shown in Fig. 4.1b cannot be the graph of a function.

Since we can discover, by geometric and algebraic means, much information about the graph of an equation, we should look at the possible interpretations of such information when that graph is also the graph of a function.

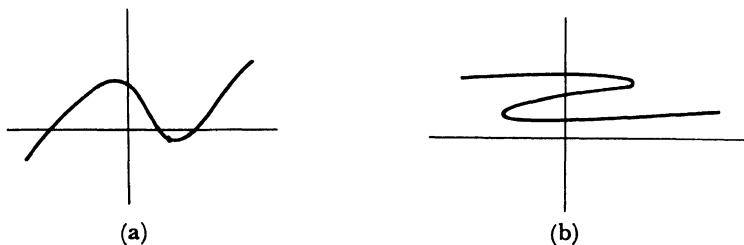


Fig. 4.1

The horizontal extent of the graph of a function is the set of numbers which are the x -coordinates of the points on this graph. But this set of numbers is the set of numbers, which are the first elements of the ordered pairs which make up the function itself. Consequently, the horizontal extent of the graph of a function is the domain of that function.

In like manner, the vertical extent of the graph of a function is the set of numbers which are the y -coordinates of the points on this graph. This set of numbers is the set of numbers, which are the second elements of the ordered pairs, which make up the function. Thus the vertical extent of the graph of a function is the range of values of that function.

If we are given the domain of a function and a rule of this function, then these alone determine what the range of values must be. These two alone completely determine the function. But a rule, by itself, does not determine a function unless we have some convention in operation which tells us what the domain of that function will be. The same set of instructions may be a rule for many different functions, each one of which has a different domain from any other. Geometrically, given the graph of one specific function, any collection of points which is part of this graph will be the graph of a function, which has a rule the same as the rule of the given function.

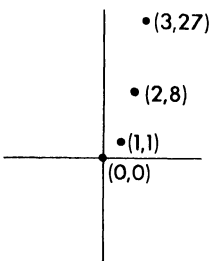
For example, let us consider the rule: cube. (Refer to Fig. 4.2a.) This rule, together with the domain: 0, 1, 2, and 3, gives the function consisting of the four ordered pairs:

$$(0, 0), (1, 1), (2, 8), \text{ and } (3, 27).$$

The same rule, with domain the interval:

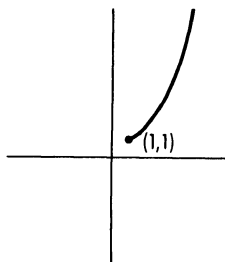
$$\text{the set of all } x, \quad x \geq 1,$$

gives the function consisting of those ordered pairs, which are the coordinates of all those points on the graph of the equation $y = x^3$, which points are to the right of and above the point with coordinates (1, 1), together with the ordered pair (1, 1) itself. (See Fig. 4.2b.)



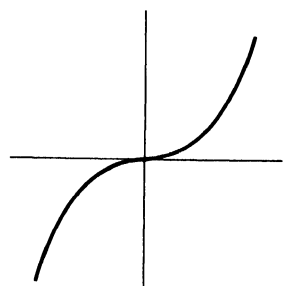
domain 0, 1, 2, 3
rule cube

(a)



domain the set of
all $x, x \geq 1$
rule cube

(b)



domain all numbers
rule cube

(c)

Fig. 4.2

The same rule, with the set of all numbers as domain, gives the function consisting of the ordered pairs, which are the coordinates of *all* the points on the graph of $y = x^3$. (See Fig. 4.2c.)

It is thus not valid to say that the graph of a function having x as an independent variable and y as a dependent variable, which variables are related by the equation $y = x^3$, is the graph of the equation $y = x^3$, since the graph of this equation may contain many more points than does the graph of the function. That is, any equation relating two variables does not give one specific function unless we adopt a convention which will allow us to specify a function by means of an equation alone.

In order to achieve an economy of language and of the physical labor involved in *writing*, we will adopt the

Convention: When only a rule of a function is prescribed, the domain of that function will consist of the set of all the numbers to which one can do that which this rule says to do. This function will be called *the* function defined by that rule.

Since the only algebraic operations forbidden us are division by 0 and taking the square root of a negative number, any rule which involves algebraic operations other than division or taking square roots will specify a function whose domain, unless specifically stated otherwise, is the set of all numbers.

Further, we will often give a function by means of an equation involving two variables or by an algebraic expression involving one variable. Unless otherwise stated, when the variables x and y are used in an equation specifying a function, x will be an independent variable and y will be a dependent variable. The equation will be of the form

$$y = f(x),$$

and the domain will be the range of the variable x . In this case we will use the words "the function given by the equation $y = f(x)$." Thus the function given by the equation $y = 2x + 5$ is the function consisting of all the ordered pairs which are the coordinates of the points on the graph of $y = 2x + 5$. If variables other than x are used, they will always be specifically labeled as to which is an independent variable and which is a dependent variable. Thus the function defined by $y = 2x + 5$ is also the function specified by $r = 2s + 5$, in which s is an independent variable and r is a dependent variable, but is not the function defined by $w = 2t + 5$ in which w is an independent variable and t is a dependent variable.

Algebraic expression involving one variable will be assumed to specify a function which has that variable as an independent variable. If we want a dependent variable for such a function, we may use any variable other than the one already used. Thus the function given by the algebraic expression $2x + 5$, the function given by $2t + 5$ and the function given by $2z + 5$ are all the same function, which function may have, as a dependent variable, any variable other than x , t , or z , respectively.

However, economy in the physical labor of writing should not mean, necessarily, economy in speech, so that—for the beginner—it is advisable to *say* all of the words and to avoid such short-cuts as “the function $2x + 5$ ” or “the function $y = 2x + 5$.”

If we want to speak of a function whose domain is not the largest possible domain for a given rule—that is, whose domain is somehow restricted—then we need to attach a qualifying statement to the equation or algebraic expression which we write. Thus the function $g()$, defined above as the set of all ordered pairs of numbers, which are the coordinates of those points on the graph of $y = x^3$ which lie in the first quadrant, may be specified by the modified equation

$$y = x^3, \quad x > 0$$

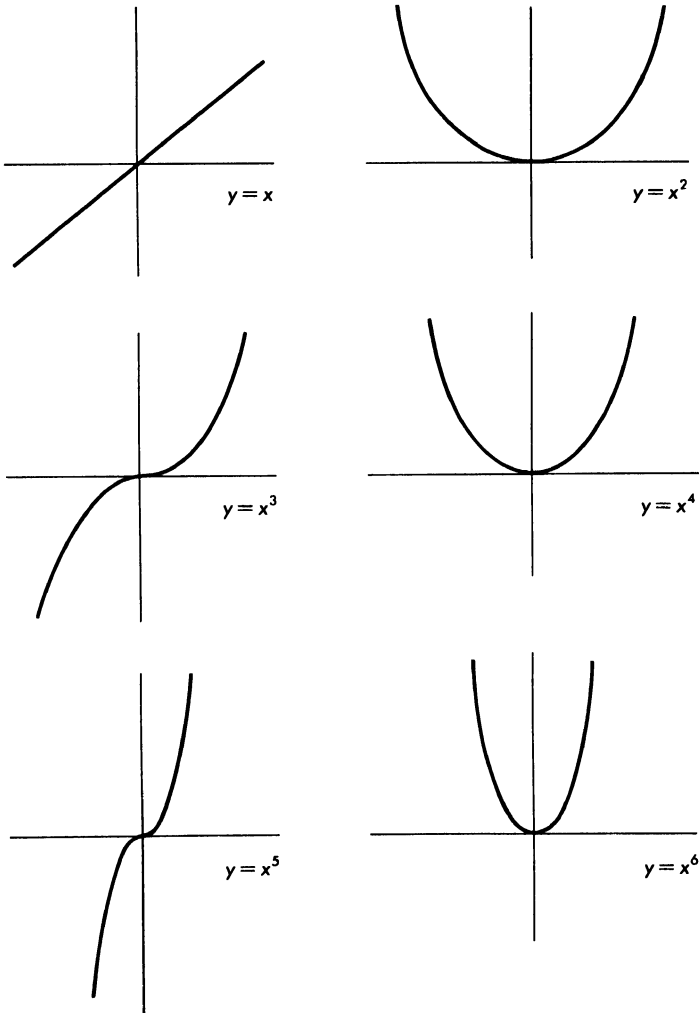


Fig. 4.3

or by any one of the modified algebraic expressions

$$\begin{array}{ll} x^3, & x > 0; \\ t^3, & t > 0; \\ r^3, & r > 0. \end{array}$$

Fig. 4.3 shows the graphs of the functions given by the equation $y = x^n$, where $n = 1, 2, 3, 4, 5, 6$.

Exercises

- Sketch the graph of the functions:
 - $(1, 1), (2, 1), (3, 1), (4, 1)$.
 - $(0, 0), (1, 2), (3, -1)$.
 - the set of coordinates of the points on the graph of $y = 3x + 1$.
 - the set of all ordered pairs whose first and second elements are the same.
 - the set of coordinates of those points on the graph of $x^2 + y^2 = 1$ which are above or on the x -axis.
- What is the domain of each of the functions in exercise 1?
- What is the range of values of each of the functions in exercise 1?
- Which of the following collections of points are the graphs of functions:
 - the points with coordinates $(2, 1), (-3, 4), (5, 7)$, and $(2, -4)$.
 - the graph of $x = -4$.
 - the graph of $x^2 + y = 0$.
 - the graph of $x + y^2 = 0$.
 - the first quadrant.
- Find three distinct functions, each of which has rule: square, but not one of which is the function defined by the equation $y = x^2$.
- State in words a rule of the function defined by the equation $2r + s = 1$, in which
 - r is an independent variable and s is a dependent variable.
 - s is an independent variable and r is a dependent variable.

Find the domain, and state a rule of the functions given by:

- $r^2 + 2r - 4$.
- $(s^3 + 3)/s$.
- $(t - 1)/(t + 1)$.
- $u\sqrt{u - 4}$.
- $\sqrt{v^2 + 1}$.
- $\sqrt{1 - w^2}$.
- $1/\sqrt{1 - w^2}$.
- $1 + 1/x, \quad x < 1$.
- $1 + 1/y, \quad y > 3$.
- $1 + 1/z$.

4.3. CONSTANT FUNCTIONS

We begin the endless task of answering the question: "What functions are there?" Since we have already defined the graph of a function and shown how functions can be specified by equations involving at most two variables, let us look at what we already know about graphs of equations and see how we may interpret this knowledge in terms of functions.

The simplest equations—those in which the variables x and y appear to at most the first degree—have, as their graphs, straight lines. Of all possible straight lines in a coordinate plane, a vertical line cannot be the graph of a function, since each point on a vertical line has the same x -coordinate. Any non-vertical line, however, is the graph of some function. In a sense, the geometrical opposite of a vertical line is a horizontal line. The coordinates of the points on a horizontal line are a collection of ordered pairs of numbers, no two of which ordered pairs have the same first element. Thus a horizontal line can be the graph of a function. Such a function has the peculiar property that its range of values consists of exactly one number. Then a dependent variable of such a function is a symbol standing for a number from a set consisting of just one number. Such a symbol has already been encountered in the word "constant." As we have defined it, a constant is a symbol standing for a specific number, but we will stretch the noun "constant" to act as an adjective in:

Definition 4.9: A constant function is a function whose range of values consists of just one number.

If we are to specify a constant function by means of an equation in which x is an independent variable and y is a dependent variable, the equation must then have the form $y = k$, where k is the one number in the range of values. Such a function has, as a result of the convention adopted the set of all numbers as domain. If we are to specify a constant function whose domain is not the set of all numbers, then we could use a modified equation, such as

$$y = k, \quad x > 0.$$

In this case, however, the situation is clearer if we use the modified equation

$$f(x) = k, \quad x > 0,$$

so as to emphasize that we are dealing with a function, an independent variable of which is x and whose domain is the set of all positive numbers.

We will adopt a particular means of writing constant functions whose domains are the set of all numbers. If the one number in the range of values of such a constant function is the number k , then we will denote that constant function by $k()$. Thus the constant function defined by the equation $y = 3$, that is the function whose domain is the set of all numbers and whose range of values is the set consisting of the one number 3, will be denoted by $3()$. The constant function whose range of values consists of the number 0

and the constant function whose range of values consists of the number 1 are of sufficient importance to be given special names. This we do in:

Definition 4.10: $\theta()$ is the constant function whose domain is the set of all numbers and whose range of values is the set consisting of the single number 0.

Definition 4.11: $U()$ is the constant function whose domain is the set of all numbers and whose range of values is the set consisting of the single number 1.

4.4. THE SUM OF TWO FUNCTIONS

As we have seen, an equation involving two variables and an algebraic expression containing one variable may be used to specify a function. Since algebraic expressions may also be considered as symbols representing numbers, this new interpretation of an algebraic expression as a symbol specifying a function gives rise to the following question: "Can functions be combined algebraically in a manner similar to that used in combining numbers?" That is, we are looking for meanings of the phrases "the sum of two functions," "the product of two functions," "the cube of a function," and so on.

The fact that it is necessary to seek a definition for the phrase "the sum of two functions" comes from the definition of the word function. A function is a certain kind of set of ordered pairs, and nothing else but this: In particular, a function is not an equation; a function is not an algebraic expression; and a function is not a number. Thus, while we would certainly expect to use the idea of the addition of two numbers in the process of defining what is to be meant by the sum of two functions, we must actually do some mental labor.

Suppose that $f(x)$ is an algebraic expression involving the variable x and is used to specify the function $f()$, as the function whose domain is the set of all values of the variable x for which the expression $f(x)$ is a number—or perhaps some smaller set.

In any case, $f()$ will have some specific set of numbers as domain and a rule for $f()$ will be given by stating what must be done to the number x in order to arrive at the number $f(x)$. There is certainly nothing special, insofar as the function $f()$ is concerned, about using the particular variable x as an independent variable. The algebraic expression $f(s)$, which results from replacing the variable x by the variable s in the original expression $f(x)$, certainly will give the same function $f()$, provided only that we attach the same domain to the rule inspired by the expression $f(s)$ as we attached to the rule inspired by the expression $f(x)$.

Suppose then, that we have two algebraic expressions, each involving one variable: $f(x)$ and $g(z)$. Each of these expressions will specify a function, once the ranges of the variables x and z are specified. For the sake of sim-

plicity, let us suppose that, in each case, we use the convention we have adopted and attach to each rule the largest possible domain. Thus, the expression $f(x)$ gives the function $f()$, whose domain is the set of all values of the variable x for which $f(x)$ is a number, and the expression $g(z)$ gives the function $g()$, whose domain is the set of all values of the variable z for which the expression $g(z)$ is a number. Of course, the function $f()$ will also result from the new algebraic expression which results from replacing x by any variable whatsoever, and the function $g()$ will also result from replacing the variable z by any variable whatsoever.

If we were merely to consider the expressions $f(x)$ and $g(z)$ as two symbols for numbers, then the new expression

$$f(x) + g(z)$$

is certainly an algebraic expression, but is one which involves two separate variables. Since we are allowing a function to be specified by an algebraic expression involving one variable, the new expression involving two variables is of no use at all in determining any new function.

If, on the other hand, we were to use the same variable in each of the expressions specifying $f()$ and $g()$, so that we had $f(x)$ and $g(x)$, then the new expression

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

would be an algebraic expression involving only one variable, and could then be used to prescribe a new function. Under the convention we have adopted, this new algebraic expression $f(x) + g(x)$ will prescribe the function whose domain must be the largest possible domain for the indicated rule, that is, the set of all values of the variable x for which $f(x) + g(x)$ is a number. Since the first addition law for numbers, A-1, says that the sum of any two numbers is always a number, the only restrictions on the values of the variable x which make $f(x) + g(x)$ a number are those imposed by the functions $f()$ and $g()$ separately. But these restrictions are precisely the determinations of the domains of the function $f()$ and $g()$. We are thus led to:

Definition 4.12: If $f()$ and $g()$ are two functions, then the new function $h()$ will be called the *sum* of the functions $f()$ and $g()$, provided that

- 1) the domain of $h()$ is the set of all numbers which are simultaneously in the domains of $f()$ and of $g()$.
- 2) a rule of $h()$ is given by $h(x) = f(x) + g(x)$.

In words, a rule of the function which is the sum of the functions $f()$ and $g()$ would say: For a number in the domain of this new function, first apply a rule of the function $f()$; then, to the number chosen in the domain, apply a rule of the function $g()$; finally, add the two numbers which resulted from the preceding two steps.

Speaking loosely, the sum of the functions $f()$ and $g()$ is the collection of ordered pairs of numbers of the form

$$(x, f(x) + g(x)),$$

where the range of the variable x is the set of those numbers which are both in the domain of $f()$ and in the domain of $g()$.

It should be noted that we have avoided using the symbol “+” as an addition symbol for functions, and have used this symbol only when we meant the addition of numbers. Since the symbol “+” is used to indicate the addition of two *numbers*, it might be confusing to use this same symbol to indicate the operation which we have called the addition of two *functions*. However, the symbol “+” is by now so ingrained in us as an addition symbol that we will adopt a manner of indicating the addition of two functions which will not dispense with the symbol “+,” but yet is sufficiently different from the manner of writing the sum of two numbers that the point is made that the addition of functions is a different thing from the addition of numbers. For the function $h()$ to be the sum of the functions $f()$ and $g()$, we will write

$$h() = \{f + g\}(),$$

where the brace is used to indicate the combining of the functions $f()$ and $g()$, and the empty parentheses to the right of the brace are used to indicate that this sum of two functions is another function.

The defining of the addition of two functions, as an operation distinct from the addition of two numbers, leads to the question of the properties this operation might have, in comparison with the properties which we know the addition of numbers has. If we can show that the addition of functions satisfies the same sort of laws as the addition of numbers, then we will be able to perform the same additive operations and manipulations which we can perform with numbers.

A function is a certain kind of set of ordered pairs of numbers. If we do not have a set of ordered pairs then we do not have a function. The sum of two functions has been defined to be a new function, found in a specified manner. In order that the sum of two given functions, $f()$ and $g()$, exist at all, we must be able to use Definition 4.12 to construct this new function. But this definition says, among other things, that the domain of the function $\{f + g\}()$ consists of the set of those numbers which are simultaneously in the domains of $f()$ and $g()$. Consequently, if the domains of $f()$ and $g()$ have *no* numbers in common, then the sum of the functions $f()$ and $g()$ does not exist at all. We have as a first addition law for functions:

If $f()$ and $g()$ are functions, and if the domains of $f()$ and $g()$ have at least one number in common, then $\{f + g\}()$ is a function.

It is possible to state formally further addition laws for functions, using the same terminology as the addition laws for numbers, but we will note

only those corresponding to the laws A-4 and A-5 for the addition of numbers. First we have:

Theorem 4.1:

$\theta()$ is a function. For each and every function $f()$, $\{f + \theta\}()$ and $f()$ are the same function.

Proof:

1. $\theta()$ is a function whose domain is the set of all numbers by the definition (4.10) of $\theta()$.
2. The domain of $\{f + \theta\}()$ is the domain of $f()$ by step 1 and the definition (4.12) of the sum of two functions.
3. If x is any number in the domain of $f()$, then

$$f(x) + \theta(x) = f(x) + 0$$
 by the definition (4.10) of $\theta()$.
4. If x is any number in the domain of $\{f + \theta\}()$, then

$$\{f + \theta\}(x) = f(x) + \theta(x)$$
 by the definition (4.12) of the sum of two functions.
5. If x is any number in the domain of $\{f + \theta\}()$, then

$$\{f + \theta\}(x) = f(x) + 0$$
 by steps 2, 3, and 4.
6. If x is any number in the domain of $\{f + \theta\}()$, then

$$\{f + \theta\}(x) = f(x)$$
 by step 5 and law A-4 for numbers.
7. $f()$ and $\{f + \theta\}()$ are the same function by steps 2 and 6.

It should be noted in this proof that, in order to prove the two functions $f()$ and $\{f + \theta\}()$ were the same, it was necessary to show that they had the same domains and also that their two rules, when applied to a single number, always gave the same result.

For a property of functions corresponding to the addition law A-5, for numbers, we have:

Definition 4.13: If $f()$ is a function, then the function

$$\{-f\}()$$

is defined as that function with the same domain as $f()$ given by the rule

$$\{-f\}(x) = -f(x).$$

It might now be expected that we would have a theorem about functions, the statement of which theorem would correspond to the addition law A-5 for numbers. This law for numbers states that the sum of a number x and

the negative of that number $-x$ is the number 0. The constant function $\theta()$ seems to play the role in the addition of functions which was played by the number 0 in the addition of numbers. One might expect, then, that the sum of a function $f()$ and the negative of that function $\{-f\}()$ will be the constant function $\theta()$. However, this function, $\theta()$ has a property not shared with *all* functions: The domain of the function $\theta()$ is the set of all numbers. Thus, if $f()$ is some function whose domain is smaller than the set of all numbers, then $\{-f\}()$ will have the same domain and the sum $\{f + \{-f\}\}()$ will also have that same set of numbers as domain. We must then conclude that there is no general law governing the addition of functions corresponding to the law A-5 governing the addition of numbers. The most we can say is that the sum $\{f + \{-f\}\}()$ is a constant function whose range of values is the set consisting of the single number 0.

As with numbers, we can define subtraction of functions by:

Definition 4.14: If $f()$ and $g()$ are functions, then the function $\{f - g\}()$ is defined to be $\{f + \{-g\}\}()$.

Finally, we consider three examples:

EXAMPLE 1:

$f()$ is the function specified by the algebraic expression

$$2x^3;$$

$g()$ is the function specified by the algebraic expression

$$x^2 - 4.$$

In both functions, the convention which we have adopted on the use of the definite article "the" tells us that the domain of each of these functions is the set of all numbers. By the definition of the sum of two functions, the domain of the function $\{f + g\}()$ will also be the set of all numbers.

Since a rule of the function $\{f + g\}()$ is given by the algebraic expression

$$2x^3 + x^2 - 4,$$

we see that the sum of the functions $f()$ and $g()$ is the function specified by the algebraic expression which is the sum of the algebraic expressions defining $f()$ and $g()$, where we have used the same variable as an independent variable of each function. (See Fig. 4 4.)

EXAMPLE 2:

$f()$ is the function given by the algebraic expression

$$\sqrt{x - 4};$$

$g()$ is the function given by the algebraic expression

$$\sqrt{5 - x}.$$

Our convention says that the domain of $f()$ is the interval:

the set of all x , $x \geq 4$,

and the domain of $g()$ is the interval:

the set of all x , $x \leq 5$.

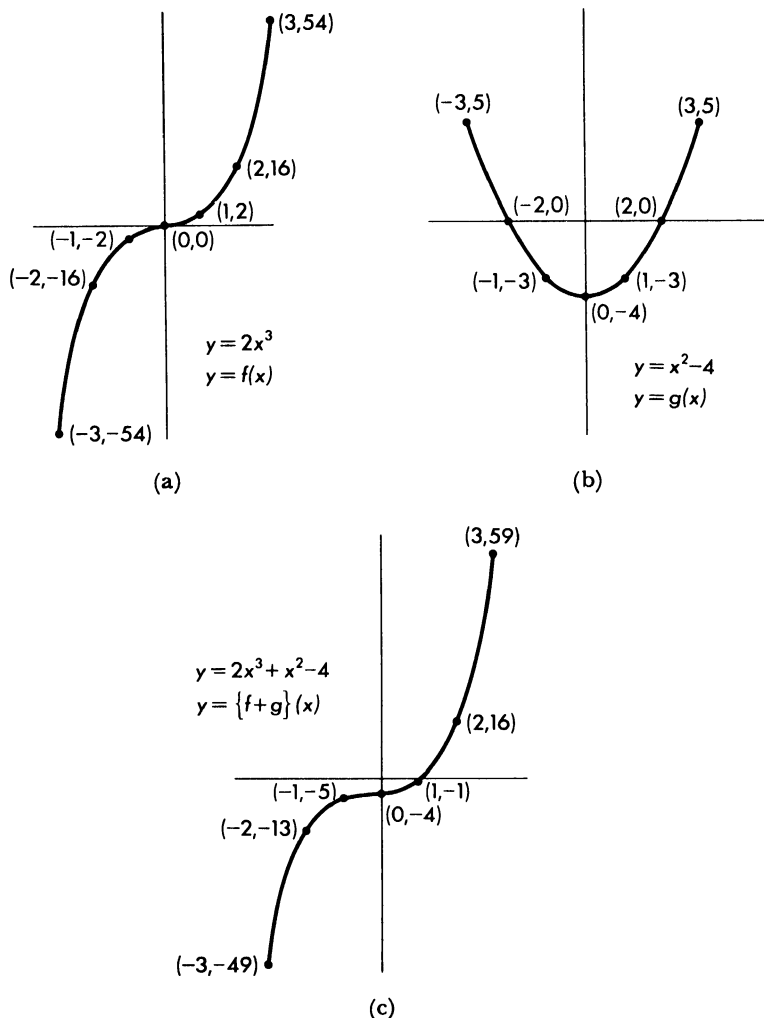


Fig. 4.4

The definition of the sum of two functions then tells us that the domain of the function $\{f + g\}()$ is the interval:

the set of all x , $4 \leq x \leq 5$.

A rule of $\{f + g\}()$ will be given by the algebraic expression

$$\sqrt{x-4} + \sqrt{5-x},$$

and the function $\{f + g\}()$ will be *the* function given by this algebraic expression. (See Fig. 4.5.)

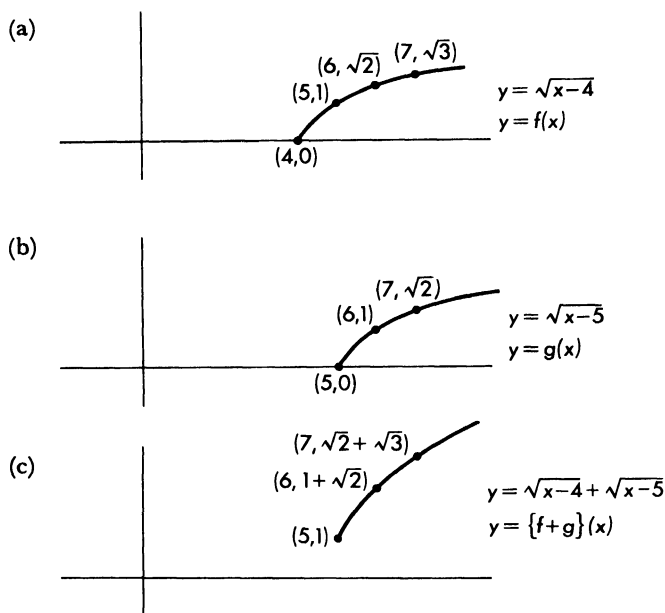


Fig. 4.5

EXAMPLE 3:

$f()$ is the function consisting of the ordered pairs:

$$(-3, 1), (-2, 2), (-1, 0), (0, 5), \text{ and } (1, -4);$$

$g()$ is the function consisting of the ordered pairs:

$$(-1, 5), (0, 3), (1, -1), (2, 4), \text{ and } (3, 17).$$

The domain of $f()$ is the set of numbers:

$$-3, -2, -1, 0, \text{ and } 1.$$

The domain of $g()$ is the set of numbers:

$$-1, 0, 1, 2, \text{ and } 3.$$

By the definition of the sum of two functions, the domain of $\{f + g\}()$ is the set of numbers:

$$-1, 0, \text{ and } 1.$$

Following this definition, we have:

$$\begin{aligned}\{f+g\}(-1) &= f(-1) + g(-1) = 0 + 5 = 5; \\ \{f+g\}(0) &= f(0) + g(0) = 5 + 3 = 8; \\ \{f+g\}(1) &= f(1) + g(1) = -4 + (-1) = -5.\end{aligned}$$

The function $\{f+g\}()$ thus consists of the three ordered pairs:

$$(-1, 5), (0, 8) \text{ and } (1, -5).$$

(See Fig. 4.6.)

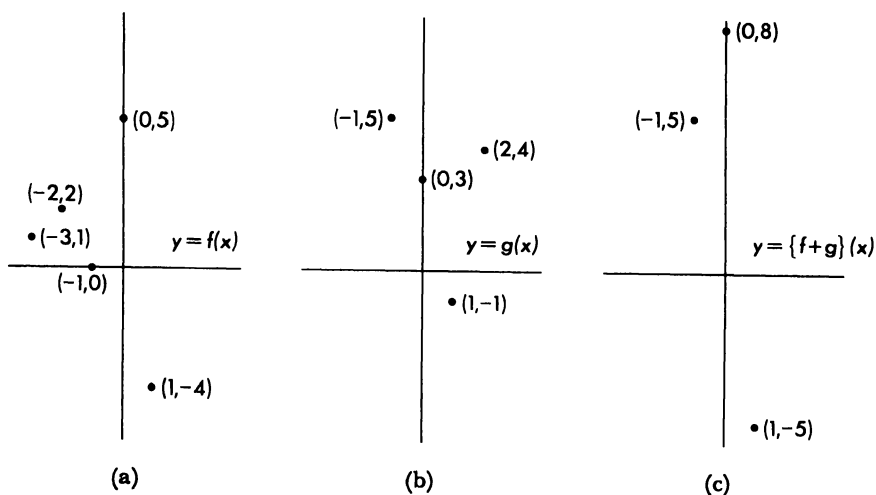


Fig. 4.6

Exercises

In each of the following, $f()$ and $g()$ are the functions defined by the given equations. Find an algebraic expression which specifies $\{f+g\}()$:

- | | |
|-----------------------------|-------------------------|
| 1. $f(r) = 3r$, | $g(s) = s^2$. |
| 2. $f(t) = t$, | $g(u) = u$. |
| 3. $f(v) = 3v$, | $g(w) = 4w$. |
| 4. $f(x) = x^2 + 3x$, | $g(y) = 5y - 2$. |
| 5. $f(z) = 3z^2 - 4z + 1$, | $g(r) = 2r^3 + r - 3$. |

What is the domain of $\{f+g\}()$ in each of the following:

- | | |
|----------------------|------------------|
| 6. $f(s) = 3s^2$, | $g(t) = 4 - t$. |
| 7. $f(u) = u$, | $g(v) = v - 1$. |
| 8. $f(w) = w$, | $g(x) = -x$. |
| 9. $f(y) = y - 2$, | $g(z) = 5 - z$. |
| 10. $f(r) = r - 7$, | $g(s) = 4 - s$. |

In each of the following, find the set of ordered pairs which is $\{f + g\}()$; and which is $\{f - g\}()$:

11. $f()$ is: $(0, 2)$, $(1, 1)$, and $2, -4)$,
 $g()$ is: $(0, 1)$, $(1, 1)$, and $(2, 2)$.
12. $f()$ is: $(-1, 2)$, $(0, 3)$, $(1, 5)$, and $(2, 6)$,
 $g()$ is: $(-2, 5)$, $(-1, 4)$, $(0, -2)$, and $(1, 3)$.
13. $f()$ is: $(-3, 1)$, $(-2, 5)$, $(1, 4)$, and $(3, 6)$,
 $g()$ is: $(-2, -3)$, $(-1, 7)$, $(0, 5)$, and $(2, 1)$.
14. $f()$ is: $(-10, 3)$, $(-4, 8)$, $(0, 1)$, and $(15, 3)$,
 $g()$ is: $(-5, 1)$, $(1, 3)$, $(2, 2)$, and $(10, 10)$.

4.5. PRODUCTS AND QUOTIENTS OF FUNCTIONS

If we recall the distinction made between *the* function defined by an algebraic expression involving one variable and *a* function defined by an algebraic expression involving one variable, we can see that the product of two functions, each of which is the function specified by an algebraic expression involving one variable, should be defined to be the function specified by the new algebraic expression in one variable formed by multiplying the given algebraic expressions. If we are dealing with functions with smaller domains, then we must again take as the domain of the product of these functions the set of all numbers which are in the domains of both the factors. We then have

Definition 4.15: If $f()$ and $g()$ are functions, then the new function $\{fg\}()$ will be called the *product of the functions* $f()$ and $g()$ provided that

- 1) the domain of $\{fg\}()$ is the set of all numbers which are simultaneously in the domains of $f()$ and $g()$;
- 2) A rule of $\{fg\}()$ is given by $\{fg\}(x) = f(x)g(x)$.

In the particular case that one of the functions making up a product is a constant function whose domain is the set of all numbers, we will suppress the notation for constant functions and merely use the number which constitutes the entire range of values of that constant function. Thus, the product of a function $f()$ and the constant function $2()$ will be written as $\{2f\}()$. With this manner of writing, we may then say that $\{2f\}()$ and $\{f + f\}()$ are the same function, since the domain of the constant function $2()$ is the set of all numbers.

As a multiplication law for functions we have:

If $f()$ and $g()$ are functions, and the domains of $f()$ and $g()$ have at least one number in common, then $\{fg\}()$ is a function.

Further, we have, corresponding to the fourth multiplication law for numbers M-4:

Theorem 4.2:

$U()$ is a function. For every function $f()$, $\{Uf\}()$ and $f()$ are the same function.

Proof:

1. $U()$ is a function whose domain is the set of all numbers by the definition (4.11) of $U()$.
2. The domain of $\{Uf\}()$ and the domain of $f()$ are the same set of numbers by step 1 and the definition (4.15) of the product of two functions.
3. For every number x in the domain of $f()$

$$\{Uf\}(x) = U(x)f(x)$$
 by the definition (4.15) of the product of two functions.
4. For every number x in the domain of $f()$,

$$\{Uf\}(x) = (1)f(x)$$
 by step 3 and the definition (4.11) of $U()$.
5. For every number x in the domain of $f()$,

$$\{Uf\}(x) = f(x)$$
 by step 4 and the law M-4.
6. $\{Uf\}()$ and $f()$ are the same function by steps 2 and 5.

In this proof, as in the proof of Theorem 4.1, we had to prove that two functions were the same. This can only be done by proving both that the domains are the same and that the two functions have a rule in common.

In the first chapter, we treated the question of division by first considering the reciprocals of numbers. The multiplication law M-5 for numbers tells us that every non-zero number has a reciprocal. Naturally, then, any constant function whose range of values consists of the single number 0 (that is, the constant function $\theta()$ or any function consisting only of some of the ordered pairs of the function $\theta()$) cannot be expected to have a reciprocal. For other functions, however, which have numbers other than 0 in their ranges of values, we can define the reciprocal function by using the idea of the reciprocal of a number, making sure that we take into account the fact that the number 0 does not have a reciprocal. Thus we have:

Definition 4.16: If $f()$ is a function whose range of values contains at least one number other than 0, then the *reciprocal of the function* $f()$ is that function, written $\{1/f\}()$, such that

- 1) the domain of $\{1/f\}()$ is the set of all those numbers in the domain of $f()$ which are not the first elements of ordered pairs of $f()$ whose second elements are the number 0.
- 2) a rule of $\{1/f\}()$ is given by $\{1/f\}(x) = 1/f(x)$.

With this definition of the reciprocal of a function, we would expect to find that the product of any given function and its reciprocal would be the constant function $U()$, whose range of values consists of the single number 1. However, as with the negative of a function, we must look carefully at the domains of the functions involved and, in doing so, we can see that the only way in which this could occur would be for both $f()$ and $\{1/f\}()$ to have, as their domains, the set of all numbers. Since there are many perfectly simple functions for which this does not occur, we can only say that the product of $f()$ and $\{1/f\}()$, if such a product exists at all, is a constant function whose range of values consists of the single number 1, but whose domain consists of only those numbers which are in the domain of $\{1/f\}()$.

We may now define the quotient of two functions, $f()$ and $g()$, as the product of $f()$ and $\{1/g\}()$. However, for the first part of this text, we will not concern ourselves with the quotients of non-constant functions, so that we will not now pursue the question of quotients and reciprocals further.

EXAMPLE 1:

$f()$ is the function given by the algebraic expression

$$x^2 - 1;$$

$g()$ is the function given by the algebraic expression

$$2x + 3.$$

From the convention adopted, the domains of $f()$ and $g()$ are, in each case, the set of all numbers. The definition of the product of two functions then says the domain of $\{fg\}()$ is the set of all numbers, so that $\{fg\}()$ is the function given by the algebraic expression

$$(x^2 - 1)(2x + 3).$$

EXAMPLE 2:

$f()$ is the function given by the algebraic expression

$$\sqrt{x} - 4;$$

$g()$ is the function given by the algebraic expression

$$\sqrt{5 - x}.$$

The domain of $f()$ is the interval:

$$\text{the set of all } x, \quad x \geq 4.$$

The domain of the function $g()$ is the interval:

$$\text{the set of all } x, \quad x \leq 5.$$

From the definition of the product of two functions, the domain of $\{fg\}()$ is the interval:

$$\text{the set of all } x, \quad 4 \leq x \leq 5.$$

This product is then the function given by the algebraic expression

$$\sqrt{x} - 4\sqrt{5-x},$$

or by the expression

$$\sqrt{-20 + 9x - x^2}.$$

EXAMPLE 3:

$f()$ is the function specified by the algebraic expression

$$\sqrt{x-1},$$

and $g()$ is the same function.

The domain of $f()$ is the interval:

the set of all x , $x \geq 1$.

The product, $\{ff\}()$ —which will be written as $\{f^2\}()$ —is the function specified by the algebraic expression

$$(\sqrt{x-1})^2.$$

But this function is not *the* function specified by the algebraic expression $x-1$, since this latter function has for its domain the set of all numbers. If we want a simpler appearance, we must use the modified algebraic expression

$$x-1, \quad x \geq 1.$$

EXAMPLE 4:

$f()$ is the function consisting of the ordered pairs:

$(-3, 1), (-2, 0), (-1, 5), (0, 3),$ and $(1, 0)$;

$g()$ is the function consisting of the ordered pairs:

$(-1, 2), (0, 3), (1, 10), (2, 1),$ and $(3, 4)$.

The domain of $f()$ consists of the numbers:

$-3, -2, -1, 0,$ and 1 .

The domain of $g()$ consists of the numbers:

$-1, 0, 1, 2,$ and 3 .

The domain of $\{fg\}()$ consists of the three numbers:

$-1, 0,$ and 1 .

For these three numbers, the definition of the product of functions tells us

$$\begin{aligned} \{fg\}(-1) &= f(-1)g(-1) = 5(2) = 10; \\ \{fg\}(0) &= f(0)g(0) = 3(3) = 9; \\ \{fg\}(1) &= f(1)g(1) = 0(10) = 0. \end{aligned}$$

The product function $\{fg\}()$ thus consists of the three ordered pairs: $(-1, 10), (0, 9),$ and $(1, 0)$.

Exercises

- Find the products $\{fg\}()$ of the 14 pairs of functions given in the exercises following section 4.4.
- Find the reciprocals $\{1/g\}()$ of the functions labeled $g()$ in the four examples of this section.
- Construct a pair of functions, each of which consists of three ordered pairs, whose product consists of
 - three ordered pairs.
 - two ordered pairs.
 - one ordered pair.
 - no ordered pairs.
- Construct a function, consisting of three ordered pairs, whose reciprocal consists of
 - three ordered pairs.
 - two ordered pairs.
 - one ordered pair.
 - no ordered pairs.
- Find two distinct constant functions, neither of which has a reciprocal.
- Let the function $f()$ be specified by the modified expression

$$f(t) = t^2, \quad t > -2.$$

- What is the function

$$g() = \left\{ \frac{1}{f} - f + f^2 - \frac{f^3 + 1}{f} + \frac{2f^4}{f^3} \right\} ()?$$

- If the domain of $f()$ is changed to

$$\text{the set of all } t, \quad -2 < t < 2,$$

what is the domain of $g()$?

- If the domain of $f()$ is changed to

$$\text{the set of all } t, \quad t > 0,$$

what is the domain of $g()$?

- If the domain of $f()$ is changed to

$$\text{the set of all } t, \quad t \geq 0,$$

what is the domain of $g()$?

For the functions $f()$ and $g()$ as specified, find the domains of $\{fg\}()$, $\{f/g\}()$, and $\{g/f\}()$.

$$7. f(x) = \sqrt{x-3}; \quad g(x) = \sqrt{4-x}.$$

$$8. f(y) = \sqrt{y+1}; \quad g(z) = \sqrt{z-1}.$$

$$9. f(r) = \sqrt{r+4}; \quad g(s) = \sqrt{7-s}.$$

$$10. f(t) = \sqrt{t-5}; \quad g(u) = \sqrt{4-u}.$$

$$11. f(v) = \sqrt{v}; \quad g(w) = \sqrt{-w}.$$

4.6. THE IDENTITY FUNCTION: POLYNOMIALS

Among the most used of all functions is the extremely simple one which consists of all ordered pairs which are the coordinates of the points on the 45-degree line—the graph of the equation $y = x$. This function may also be described as the collection of all possible ordered pairs of numbers in which each ordered pair contains the same first and second element, or as the function specified by the equation $y = x$. This set of all ordered pairs of the form (a, a) , where a is any number, is given a special name in

Definition 4.17: The function $I()$, which consists of all ordered pairs of numbers with the first and second elements of each pair equal, is called *the identity function*.

The domain of the function $I()$ is, of course, the set of all numbers, and the range of values of $I()$ is also the set of all numbers. A rule of $I()$ is given by the equation $I(x) = x$.

The particular advantage of the identity function is that we may express any function which is *the* function specified by an algebraic expression as a function-algebraic (not number-algebraic) combination of the function $I()$ and various constant functions. For example, the function given by the algebraic expression

$$x^2 - 2x + 5$$

is the function

$$\{I^2 - 2I + 5\}(),$$

where the “2” and “5” refer to the constant functions $2()$ and $5()$, rather than to the numbers 2 and 5.

Since we have defined the function $I()$ to have, as domain, the set of all numbers, we cannot use this function to represent a function whose domain is any smaller set, except in the special cases in which the algebraic laws for combinations of functions themselves restrict the domain of the combination. Thus the function given by the modified algebraic expression

$$2x^3 - 4x + 7, \quad 0 < x < 1,$$

whose domain is the interval:

$$\text{the set of all } x, \quad 0 < x < 1,$$

cannot be represented as a function-algebraic combination of $I()$ with the constant functions $2()$, $4()$, and $7()$ (except in a very complicated, tortured way). But the function specified by the expression

$$\frac{x + 4}{x - 2},$$

whose domain is the set of all numbers other than 2, can be and is expressed as

$$\left\{ \frac{I + 4}{I - 2} \right\}().$$

Among the many facts about the identity function, one of the most significant is the following: If ϵ is a positive number, then, for each and every number x , the statement

$$\text{"if } 0 < |x - a| < \epsilon \text{ then } |I(x) - a| < \epsilon\text{"}$$

is true for each and every number a . In less exact words, this statement says that $I(x)$ is a number close to a for those value of the variable x which are close to a , regardless of what number a may be.

For the next few chapters, we will be concerned almost exclusively with combinations of the identity function and constant functions which are called polynomial functions:

Definition 4.18: A *polynomial function* is a function which is the sum of functions which are either the product of a positive integer power of the identity function and a constant function, or are themselves constant functions.

This definition implies that every polynomial function has the form

$$\{a_n I^n + a_{n-1} I^{n-1} + a_{n-2} I^{n-2} + \dots + a_2 I^2 + a_1 I + a_0\}(),$$

where n is some positive integer and the $n + 1$ symbols $a_0, a_1, a_2, \dots, a_{n-2}, a_{n-1}$, and a_n designate constant functions. From this definition of a polynomial function, we may define an algebraic expression in one variable as a polynomial algebraic expression in that variable, or, more simply, as a polynomial in that variable, by saying that a polynomial (algebraic expression) in the variable x is the sum of terms which are either the product of a positive integer power of x and a constant, or are themselves constants. Thus the polynomial function above is the function specified by the polynomial expression

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_2 x^2 + a_1 x + a_0.$$

The *degree* of a polynomial function (or a polynomial expression) is the largest positive integer exponent contained in the function (or expression). We have already done some analysis of polynomial functions of degree 1 and 2. In particular, we have proved that the graph of every polynomial function of degree 1 is a straight line. We have also looked rather closely at the graph of the specific polynomial function of degree 2, $\{I^2\}()$. In Section 7 of Chapter 3 we discovered that if we are given the positive number ϵ , as a desired closeness of the numbers $\{I^2\}(a + h)$ and $\{I^2\}(a)$, then this desired closeness may be achieved by choosing the number h so that both the inequalities $|h| < 2a$ and $|h| < \epsilon/4a$ are true.

Since the graph of every polynomial function of degree 1 is a straight line (and, in fact, a non-vertical straight line), we call such functions *linear*. To conform with usual usage, we will call polynomial functions of degree 2 quadratic functions, and polynomial functions of degree 3 cubic functions. The constant functions involved in a polynomial function are called the *coefficients* of the polynomial function. It should be remembered, however, that these coefficients of a polynomial function are themselves functions.

In elementary algebra one learns to solve equations involving polynomial algebraic expressions. A solution of an equation of the form $P(x) = 0$, where $P()$ is a polynomial function, will be called a *zero* of the polynomial function $P()$. The various algebraic facts about polynomial algebraic expressions remain facts, so that we know that a polynomial function of degree n has at most n zeros. We also know that a polynomial algebraic expression which can be expressed as the product of first degree algebraic expressions defines a polynomial function which can be expressed as the product of linear functions. Theorems 2.2 and 2.3 tell us how to find the zeros of polynomial functions of degree 1 and 2.

Exercises

If $P()$ is each of the following polynomial functions, find $P(x)$, $P(t)$, $P(2)$, $P(1)$, $P(0)$ and $P(-1)$:

1. $\{I^3 - 4I^2 + 7I - 2\}()$.
2. $\{3I^2 + 5I - 6\}()$.
3. $\{-I^3 + 2I^2 - 3I\}()$.
4. $\{I^4 + I\}()$.

Express the functions defined by each of the following algebraic expressions as a function-algebraic combination of the identity function and constant functions:

5. $r^3 - 3r^2 + 2r + 1$.
6. $3s^3 + 4s^2 - s + 5$.
7. $t^2 - 3t - 5$.
8. $u^5 - 17u^2 + 1$.
9. $v - 3v^2 + 7v^5$.
10. $w + 1/w$.
11. $(x + 1)/x$.
12. Draw the graph of the identity function.
13. On one plane, draw the graphs of:
 - a) $\{-3\}()$.
 - b) $\{-4I\}()$.
 - c) $\{I^2\}()$.
 - d) $\{-4I + 3\}()$.
 - e) $\{I^2 - 4I + 3\}()$.



LIMITS AND DERIVATIVES

5.1. APPROXIMATIONS ON GRAPHS

In Chapter 2, we looked closely at the graph of the equation $y = x^2$. In Chapter 4, we saw that the collection of ordered pairs of numbers which make up the coordinates of the points on this graph are the function $\{I^2\}()$, which function has, as domain, the set of all numbers, and, as range of values, the set of all non-negative numbers. In the last section of Chapter 4, we constructed a method of describing this function as the square of the identity function. We may then describe the function $\{I^2\}()$ as the function specified by the equation $f(x) = x^2$ or by the algebraic expression x^2 , or by the equation $y = x^2$.

In this section, we will examine the behavior of the graph of $\{I^2\}()$ “near” the point with coordinates $(2, 4)$. Such an examination may be conducted in one of two ways:

- (1) We can look at the numbers in the domain of $\{I^2\}()$ which are “near” 2 and see what numbers in the range of values of $\{I^2\}()$ correspond to (that is, get paired with) them.
- (2) We can look at the numbers in the range of values of $\{I^2\}()$ which are “near” 4 and see what numbers in the domain of $\{I^2\}()$ they correspond to (that is, are paired with).

In geometric terms, these two lines of inquiry are:

- (1) We examine the points on the graph of $y = x^2$ whose x -coordinates are “near” 2, and see what y -coordinates these points have.
- (2) We examine the points on the graph of $y = x^2$ whose y -coordinates are “near” 4, and see what x -coordinates these points have.

Whichever of these two lines of investigation is chosen, our concern is NOT with the point $(2, 4)$ itself, since we know that this point is on the graph of $\{I^2\}()$. Our concern lies with those points on the graph of this function which are—or ought to be—“near” the point with coordinates $(2, 4)$, and, in order to allow us to generalize this investigation to any

point involved with the graph of any function, a major task will be either to give precise meaning to the word “near” or to avoid using such an imprecise word.

Pursuing the first line of investigation, it seems most reasonable to guess that points on the graph of the function $\{I^2\}()$ whose x -coordinates are “near” 2 will have y -coordinates which are “near” 4. But this guess is merely a sketchy application of the laws governing inequalities. To be exact and to avoid the use of the vague word “near,” we may use the ordering laws for numbers to show that each of the following five statements is true:

If $1 < x < 3$,	then $1 < x^2 < 9$.
If $1.5 < x < 2.5$,	then $2.25 < x^2 < 6.25$.
If $1.9 < x < 2.1$,	then $3.61 < x^2 < 4.41$.
If $1.99 < x < 2.01$,	then $3.9601 < x^2 < 4.0401$.

For d , a number between 0 and 2,

$$\text{if } 2 - d < x < 2 + d, \text{ then } 4 - d(4 - d) < x^2 < 4 + d(4 + d).$$

In each of the true statements above, we now replace the second inequality by an inequality which is less restrictive on the values of the variable x which satisfy this inequality. That is, we replace the interval in which we found the number x^2 by an interval which is larger, and which larger interval has the number 4 as mid-point. The following five statements are true:

If $1 < x^2 < 9$,	then $-1 < x^2 < 9$.
If $2.25 < x^2 < 6.25$,	then $1 < x^2 < 7$.
If $3.61 < x^2 < 4.41$,	then $3.5 < x^2 < 4.5$.
If $3.9601 < x^2 < 4.0401$,	then $3.95 < x^2 < 4.05$.

For d , a number between 0 and 2,

$$\text{if } 4 - d(4 - d) < x^2 < 4 + d(4 + d), \text{ then } 4 - 6d < x^2 < 4 + 6d.$$

Since these five statements are true, we may combine these with the original five statements to have the following five true statements:

If $1 < x < 3$,	then $-1 < x^2 < 9$.
If $1.5 < x < 2.5$,	then $1 < x^2 < 7$.
If $1.9 < x < 2.1$,	then $3.5 < x^2 < 4.5$.
If $1.99 < x < 2.01$,	then $3.95 < x^2 < 4.05$.

For d , a number between 0 and 2,

$$\text{if } 2 - d < x < 2 + d, \text{ then } 4 - 6d < x^2 < 4 + 6d.$$

We may use the absolute value symbol to restate these five true statements:

If $ x - 2 < 1$,	then $ x^2 - 4 < 5$.
If $ x - 2 < 0.5$,	then $ x^2 - 4 < 3$.
If $ x - 2 < 0.1$,	then $ x^2 - 4 < 0.5$.
If $ x - 2 < 0.01$,	then $ x^2 - 4 < 0.05$.

For d , a number between 0 and 2,

$$\text{if } |x - 2| < d, \text{ then } |x^2 - 4| < 6d.$$

These five statements may be immediately interpreted in terms of the graph of the function $\{I^2\}()$:

Any point on the graph of $\{I^2\}()$ whose x -coordinate is within 1 unit of 2 will have y -coordinate within 5 units of 4.

Any point on the graph of $\{I^2\}()$ whose x -coordinate is within 0.5 units of 2 will have y -coordinate within 3 units of 4.

Any point on the graph of $\{I^2\}()$ whose x -coordinate is within 0.1 units of 2 will have y -coordinate within 0.5 units of 4.

Any point on the graph of $\{I^2\}()$ whose x -coordinate is within 0.01 units of 2 will have y -coordinate within 0.05 units of 4.

For d , a number between 0 and 2, any point on the graph of $\{I^2\}()$ whose x -coordinate is within d units of 2 will have y -coordinate within $6d$ units of 4.

These five statements, which were developed in the pursuit of the first line of investigation, actually allow us to pursue the second line of investigation. We can now answer the following questions:

How close to 2 could x be in order that x^2 be at least within 5 units of 4?

How close to 2 could x be in order that x^2 be at least within 3 units of 4?

How close to 2 could x be in order that x^2 be at least within 0.5 units of 4?

How close to 2 could x be in order that x^2 be at least within 0.05 units of 4?

For ϵ , some preassigned positive number, how close to 2 could x be in order that x^2 be within ϵ units of 4.

There is no *one* correct answer to any of these questions, but a correct answer to the first of them is 1; a correct answer to the second is 0.5; a correct answer to the third is 0.1; and a correct answer to the fourth is 0.01. Of course, the number 0.01 is a correct answer to all four questions, since if the number x^2 is within 0.05 units of 4 then that number is also within 0.5, 3, 5, or 1,000,000 units of 4.

An answer to the fifth question is partially given by the last of the five statements, namely, for ϵ , a number less than 12, x should be within $\epsilon/6$ units of 2 in order that x^2 be at least within ϵ units of 4. But if ϵ is a number not less than 12, the first statement—which says that if x is within 1 unit of 2, then x^2 will be within 5 units of 4—allows us to say: Whatever positive number ϵ may be, we may make the number x^2 be within at least ϵ units of 4 merely by taking the number x to be within either $\epsilon/6$ units or within 1 unit of 2, whichever of the numbers $\epsilon/6$ and 1 is the smaller.

We have now proved:

For any $\epsilon > 0$, there does exist a corresponding $d > 0$, namely, d the smaller of the two numbers $\epsilon/6$ and 1, such that

$$\text{if } |x - 2| < d, \text{ then } |x^2 - 4| < \epsilon.$$

The first inequality in this last statement says that the number x is in an interval whose center is the number 2 and whose width is $2d$. We said before that we are not in the least concerned with the fact that the point $(2, 4)$ is on the graph of the equation $y = x^2$, but are concerned with points "near" $(2, 4)$. In order to emphasize this total unconcern with the fact that $2^2 = 4$, we will state our conclusion as:

For any $\epsilon > 0$, there does exist a $d > 0$, namely d the smaller of the two numbers $\epsilon/6$ and 1, such that

$$\text{if } 0 < |x - 2| < d, \text{ then } |x^2 - 4| < \epsilon.$$

The implication contained in this statement (that is, the if-then statement contained in our conclusion) may then be expressed in words as: If x is a number other than 2, but within d units of 2, then x^2 will be a number within ϵ units of 4.

A less precise, but perhaps more descriptive, rephrasing of our conclusion is: With no concern for the number 2 itself, we can make the number x^2 as close as we please to the number 4, merely by taking the number x sufficiently close to the number 2. In the exact statement we have made, the positive number ϵ is the closeness which pleases, and the positive number d is the closeness to 2 which accomplishes that which we say we can do.

Exercises

- How close to 3 should the number r be taken to be sure that the number r^2 is at least within
 - 1 unit;
 - 0.1 units;
 - 0.01 units;
 - ϵ units (where $\epsilon > 0$)
 of the number 9?

- Show that, given any positive number ϵ , there does in fact exist a positive number d —whose values may depend on ϵ —such that if

$$0 < |t - 3| < d, \text{ then } |t^2 - 9| < \epsilon.$$

- Find an interval, whose center is the number 3, which interval is a solution of the inequality

$$9 - \epsilon < s^2 < 9 + \epsilon,$$

where s is a variable, and ϵ is a positive constant.

- Let $f()$ be the function defined by the expression

$$4t - 2.$$

For a given positive number ϵ , how close to the number p should we take numbers in the domain of $f()$ to be sure that the corresponding numbers in the range of values of $f()$ will be at least within ϵ units of the number $f(p)$?

- What is the answer to the question in exercise 4 if $f()$ is the function defined by $f(x) = mx + b$?

6. If the graph of the equation $y = x^3$ is cut by the two horizontal lines with equations $y = 8.1$ and $y = 7.9$,
- what are the points of intersection of these two lines with the graph?
 - of the two points of intersection, which of them has x -coordinate closer to 2?
 - Find an interval, centered at 2, such that if t is a number in this interval, then $|t^3 - 8| < 0.1$.

Repeat the procedure in exercise 6 for the intersection of the graph of the given equation with the given horizontal lines:

- $y = x^3$; $y = 8.01$; $y = 7.99$.
- $y = x^3$; $y = 8.001$; $y = 7.999$.
- $y = x^3$; $y = 8 + e$; $y = 8 - e$, (where $e > 0$).
- $y = 3x^3$; $y = 8 + e$; $y = 8 - e$, (where $e > 0$).

5.2. LIMITS OF FUNCTIONS

In the preceding section, we examined the graph of the function $\{I^2\}()$ near the point with coordinates $(2, 4)$. If we remove the ordered pair $(2, 4)$ from the collection of ordered pairs which is the function $\{I^2\}()$, then we will have a new function. We will call this new function $g()$, so that $g()$ is specified by the equation

$$g(x) = x^2 \frac{x-2}{x-2},$$

or by the modified equation

$$g(x) = x^2, \quad x \neq 2.$$

The domain of the function $g()$ is the set of all numbers other than 2, and the range of values of $g()$ is the set of all non-negative numbers. For example, the number 4 is in the range of values of $g()$ since $(-2, 4)$ is one of the ordered pairs of $g()$. Every ordered pair of the function $g()$ is also an ordered pair of the function $\{I^2\}()$, and the ordered pair $(2, 4)$ is in the function $\{I^2\}()$ but not in the function $g()$. Geometrically, the graph of $g()$ consists of every point on the graph of $\{I^2\}()$ except the point with coordinates $(2, 4)$.

Since our concern with the function $\{I^2\}()$ was not with the point with coordinates $(2, 4)$ itself, but with points "near" this one, we find that—in inexact words—any point on the graph of $g()$ whose x -coordinate is near 2 has y -coordinate near 4. In exact words, if we are given any positive number ϵ , we can find a corresponding positive number δ , namely δ the smaller of the two numbers $\epsilon/6$ and 1, such that if x is a number other than 2, but within δ units of 2, then x will be in the domain of $g()$ and the number $g(x)$ will be within ϵ units of 4.

In general, if $f()$ is any function, and a is any fixed number, we ask the following question: "As we look at the set of all numbers, squeezing our view down on the number a , do we eventually get all of the numbers in our view in the domain of $g()$, and do the corresponding numbers in the range of values of $f()$ squeeze down on one fixed number?" If the answer to this question is: "Yes, and that one fixed number is L ," then we may say:

"Given any $\epsilon > 0$, there does exist a corresponding $\delta > 0$ such that, if $0 < |t - a| < \delta$, then t is in the domain of $f()$ and $|f(t) - L| < \epsilon$."

It is convenient to have a simple, short label for such a situation for a function $f()$, a number a , and a number L . This label is given in

Definition 5.1: If $f()$ is a function, a is a number, and L is a number, then

$$\lim_{t \rightarrow a} f(t) \text{ is the number } L$$

(read "the limit of f of t as t approximates a .") means:

For any $\epsilon > 0$, there exists a corresponding $\delta > 0$ such that if $0 < |t - a| < \delta$, then t is in the domain of $f()$ and

$$|f(t) - L| < \epsilon.$$

Remark: It is my opinion that this definition (5.1) is the most important single definition in the study of mathematics.

To use this new notation, we have discovered in Section 1, that

$$\lim_{x \rightarrow 2} \{I^2\}x = 4,$$

or

$$\lim_{t \rightarrow 2} \{I^2\}(t) = 4,$$

or

$$\lim_{() \rightarrow 2} \{I^2\}() = 4,$$

where any variable is inserted in the blank parentheses. Since the manner of writing adopted for the limit of a function considered an expression of the form $\lim_{t \rightarrow a} f(t)$ as a single entity—in fact, as a number—we may in this instance use directly an algebraic expression which defines a function, so that

$$\lim_{x \rightarrow 2} \{I^2\}(x) = \lim_{x \rightarrow 2} x^2 = 4$$

and

$$\lim_{x \rightarrow 2} g(x) = \lim_{x \rightarrow 2} \left[x^2 \frac{x - 2}{x - 2} \right] = 4.$$

If, for a given function $h()$, we want to find the number $\lim_{z \rightarrow b} h(z)$, we must answer the question: "What (if any) is the number to which numbers in the range of values of $h()$ squeeze down, if we squeeze down toward b in our view of the set of all numbers?" It should be noted that we are NOT asking: "What is the number in the range of values of $h()$ which corresponds

to the number b in the domain of $h()$?" In fact, from the definition of limit, it is not at all necessary, in order that $\lim_{t \rightarrow c} h(t)$ be the number P , that c be in the domain of $h()$, or that P be in the range of values of $h()$, or that, even if c were in the domain of $h()$ and P were in the range of values of $h()$, that $h(c) = P$. For example, 2 is not in the domain of $g()$, 4 is in the range of values of $g()$ (since $g(-2) = 4$), but we still have

$$\lim_{t \rightarrow 2} g(t) = 4.$$

It is possible that the limit of a given function, as we approximate a given number, may not exist at all. We will defer the examination of such cases until we have found a little more about limits which do exist, and have seen some of their uses.

In order to find limits for polynomial functions, which are our first variety of interest, we will not examine each function by itself, but merely consider a polynomial function as an algebraic combination of constant functions and the identity function. Thus, it behooves us to investigate the limit properties of constant functions, of the identity function, and then of algebraic combinations of functions.

Theorem 5.1:

If $c()$ is a constant function, whose domain is the set of all numbers and whose range of values consists of the single number c , and if a is a number, then

$$\lim_{t \rightarrow a} c(t) = c.$$

Proof:

1. Let ϵ be a positive number.

2. $|c - c| = |0| = 0$

by the hypothesis that c is a number and the definition of absolute value.

3. For every value of the variable t ,

$$|c(t) - c| = 0$$

by step 2, the hypothesis that $c()$ is a constant function, and the definition of a constant function.

4. For every value of the variable t ,

$$|c(t) - c| < \epsilon$$

by steps 1 and 3.

5. If d is any positive number, then

$0 < |t - a| < d$ implies that t is in the domain of $c()$

by the hypothesis that the domain of $c()$ is the set of all numbers.

6. If d is any positive number, then

$0 < |t - a| < d$ implies that $|c(t) - c| < \epsilon$

by step 4.

7. $\lim_{t \rightarrow a} c(t) = c$

by steps 1, 5, and 6, and the definition of limit.

Of course, the following interpretation in words of this theorem does not sound very deep: "As t squeezes down on the number a , then the constant c squeezes down on the constant c ." Yet, no one would doubt the truth of this non-profound statement.

For the identity function, we have another non-profound, but true statement: "As t squeezes down on the number a , then t squeezes down on the number a ." This is given formally in

Theorem 5.2:

If a is a number, then

$$\lim_{t \rightarrow a} I(t) = a.$$

Proof:

1. For every number t ,

$$I(t) = t$$

by the definition of the identity function.

2. Let ϵ be a positive number.

3. For every value of the variable t ,

$$\begin{aligned} &\text{if } 0 < |t - a| < \epsilon, \\ &\text{then } |t - a| < \epsilon \end{aligned}$$

since the truth of a more restrictive statement implies the truth of the less restrictive statement.

4. For every value of the variable t ,

$$\begin{aligned} &\text{if } 0 < |t - a| < \epsilon, \\ &\text{then } |I(t) - a| < \epsilon \end{aligned}$$

by steps 1 and 3.

5. The domain of $I()$ is the set of all numbers

by the definition of the identity function.

6. For every value of the variable t ,

$$\begin{aligned} &\text{if } 0 < |t - a| < \epsilon, \\ &\text{then } t \text{ is in the domain of } I(). \end{aligned}$$

by step 5.

7. For any $\epsilon > 0$, there does exist a corresponding $d > 0$ —namely $d = \epsilon$ itself—such that if

$$\begin{aligned} &0 < |t - a| < d, \\ &\text{then } t \text{ is in the domain of } I() \text{ and} \\ &|I(t) - a| < \epsilon \end{aligned}$$

by steps 4 and 6.

8. $\lim_{t \rightarrow a} I(t) = a$

by step 7 and the definition (5.1).

Exercises

1. Find $\lim_{z \rightarrow 3} f(z)$ for $f()$ the function specified by

a) $f(r) = r^2$.

b) $f(s) = s^2 \frac{s-3}{s-3}$.

$$\text{c) } f(t) = t^2 \frac{t-2}{t-2}.$$

$$\text{d) } f(u) = u^2, \quad 0 \leq u \leq 5.$$

$$\text{e) } f(v) = v^2, \quad 2 < v < 3.$$

$$\text{f) } f(w) = w^2, \quad 3 \leq w < 4.$$

$$\text{g) } f(x) = x^2, \quad x < 1.$$

2. Illustrate the proof of Theorem 5.1 on a graph.

3. Consider the point with coordinates $(5, 5)$ on the graph of the equation $y = x$. If the y -coordinates of points on this graph are in the interval:

$$\text{the set of all } y, \quad 4 < y < 6,$$

in which interval are the x -coordinates of these points? If e is a positive number and the interval for the y -coordinates is:

$$\text{the set of all } y, \quad 5 - e < y < 5 + e,$$

in which interval are the x -coordinates?

4. Illustrate the proof of Theorem 5.2 on a graph.

5. Let the function $h()$ consist of the following ordered pairs: all ordered pairs of the form $(p, 1)$, where p is positive; all ordered pairs of the form $(n, -1)$, where n is negative; and the ordered pair $(0, 0)$.

a) Draw the graph of $h()$.

b) Find $h(-1)$, $h(0)$, and $h(1)$.

c) Find $\lim_{r \rightarrow -1} h(r)$; $\lim_{s \rightarrow 0} h(s)$; $\lim_{t \rightarrow 1} h(t)$.

6. Prove that a function which consists of only a finite number of ordered pairs has no limit as we approximate any number.

7. If the domain of the function $f()$ is the interval:

$$\text{the set of all } t, \quad a \leq t \leq b,$$

show that neither $\lim_{x \rightarrow a} f(x)$ nor $\lim_{x \rightarrow b} f(x)$ exists.

8. Find $\lim_{r \rightarrow 2} g(r)$ for $g()$ the function specified by

$$\text{a) } g(r) = r^3.$$

$$\text{b) } g(r) = r^3 \frac{r-2}{r-2}.$$

$$\text{c) } g(s) = s^3, \quad s \neq 2.$$

$$\text{d) } g(t) = t^3, \quad t \neq 0.$$

$$\text{e) } g(u) = u^3, \quad 0 \leq u \leq 4.$$

$$\text{f) } g(v) = v^3, \quad 0 \leq v < 2.$$

$$\text{g) } g(w) = w^3, \quad 2 \leq w < 5.$$

$$\text{h) } g(x) = x^3, \quad -3 < x < 1.$$

$$\text{i) } g(y) = y^3, \quad y > 0.$$

$$\text{j) } g(z) = z^3, \quad z \leq 0.$$

5.3. SOME THEOREMS ON LIMITS

In order to provide ourselves with the tools for evaluating the limits of polynomial functions, we will need some theorems relating the limits of given functions to the limit of a function which is an algebraic combination (sum, difference, product) of these given functions. From the definition of a polynomial function, we see that we need to know about positive integer powers of functions, about the product of a function and a constant function, and about the sum of functions.

The reasoning used in the proofs of some of these theorems is complicated and somewhat lengthy, so that only one proof will actually be given. The reader's understanding of this one proof will provide a test of his understanding of the concept of the limit of a function.

Theorem 5.3:

If $f()$ and $g()$ are functions; if a is a number; if $\lim_{t \rightarrow a} f(t)$ is the number M ; and if $\lim_{t \rightarrow a} g(t)$ is the number N ,

then $\lim_{t \rightarrow a} \{f + g\}(t)$ is the number, $M + N$.

Proof:

1. Let ϵ be a positive number.

2. $\epsilon/2$ is a positive number by step 1, law O-4a.

3. $\lim_{t \rightarrow a} f(t) = M$ by hypothesis.

4. Corresponding to the positive number $\epsilon/2$, there is a positive number d_1 , such that if

$$0 < |t - a| < d_1,$$

then t is in the domain of $f()$, and

$$|f(t) - M| < \epsilon/2$$

by step 3 and the definition (5.1.)

5. $\lim_{t \rightarrow a} g(t) = N$ by hypothesis.

6. Corresponding to the positive number $\epsilon/2$, there is a positive number d_2 , such that if

$$0 < |t - a| < d_2,$$

then t is in the domain of $g()$, and

$$|g(t) - N| < \epsilon/2$$

by step 5 and the definition (5.1.)

7. For any number t in the domain of $\{f + g\}()$,

$$\{f + g\}(t) - (M + N) =$$

$$f(t) + g(t) - (M + N)$$

by the definition of the sum of two functions.

8. For any number t in the domain of $\{f + g\}()$,

$$\{f + g\}(t) - (M + N) = (f(t) - M) + (g(t) - N)$$

by step 7 and various algebraic laws for numbers.

9. For any number t in the domain of $\{f + g\}()$,

$$|\{f + g\}(t) - (M + N)| \leq |f(t) - M| + |g(t) - N|$$

by step 8 and the theorem (2.3) which says that

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

10. t is in the domain of $\{f + g\}()$ if and only if t is both in the domain of $f()$ and the domain of $g()$

by the definition of the sum of two functions.

11. Let d be the smaller of the two positive numbers, d_1 and d_2 .

12. If $0 < |t - a| < d$, then t is in the domain of $\{f + g\}()$ and

$$|\{f + g\}(t) - (M + N)| < e/2 + e/2$$

by steps 11, 4, 6, 9, and 10.

13. For every $e > 0$, there does exist a corresponding $d > 0$ —namely d the smaller of the numbers, d_1 and d_2 , such that if

$$0 < |t - a| < d,$$

then t is in the domain of $\{f + g\}()$, and

$$|\{f + g\}(t) - (M + N)| < e \quad \text{by steps 1 and 12.}$$

14. $\lim_{t \rightarrow a} \{f + g\}(t)$ is the number, $M + N$

by step 13 and the definition (5.1).

We now state two theorems whose proofs will be omitted.

Theorem 5.4:

If $f()$ is a function; if a is a number; if $c()$ is a constant function; and if $\lim_{t \rightarrow a} f(t)$ is the number M ,

then $\lim_{t \rightarrow a} \{cf\}(t)$ is the number cM .

Theorem 5.5:

If $f()$ and $g()$ are functions; if a is a number; if $\lim_{t \rightarrow a} f(t)$ is the number M ; and if $\lim_{t \rightarrow a} g(t)$ is the number N ,

then $\lim_{t \rightarrow a} \{fg\}(t)$ is the number MN .

This second theorem on the limit of the product of two functions allows us to find the limit of any positive integer power of a function. For the purposes of polynomial functions we need

Theorem 5.6:

If n is a positive integer and a is a number, then

$$\lim_{t \rightarrow a} \{I^n\}(t) = a^n.$$

Proof:

1. $\{I^n\}() = \{I() I() I() \dots I()\}()$ by the definition of the product of functions, where there are n factors on the right.
2. $\lim_{t \rightarrow a} I(t) = a$ by Theorem 5.2.
3. $\lim_{t \rightarrow a} \{I^n\}(t) = \{\lim_{t \rightarrow a} I(t)\}^n$ by step 2 and n applications of Theorem 5.5.
4. $\lim_{t \rightarrow a} \{I^n\}(t) = a^n$ by steps 2 and 3.

Exercises

Let $f()$ be specified by $f(x) = 3x$ and $g()$ be specified by $g(x) = 4x$

1. Find $\lim_{r \rightarrow 2} f(r)$.
2. Find $\lim_{s \rightarrow 2} g(s)$.
3. Find $\lim_{t \rightarrow 2} \{f + g\}(t)$.
4. Find $\lim_{u \rightarrow 2} \{fg\}(u)$.
5. Solve the inequality $|3v - 6| < 1/2$.
6. Solve the inequality $|4w - 8| < 1/2$.
7. Which of the intervals found in exercises 5 and 6 is a solution of the inequality $|7x - 14| < 1$.
8. For ϵ a fixed positive number, find a value for the number d_1 , in terms of ϵ , which makes the statement:
 “If $0 < |y - 2| < d_1$, then $|3y - 6| < \epsilon$ ” true.
9. For ϵ a fixed positive number, find a value for the number d_2 , in terms of ϵ , which makes the statement:
 “If $0 < |z - 2| < d_2$, then $|4z - 8| < \epsilon$ ” true.

10. Show that the smaller of the two numbers, d_1 and d_2 , found in exercises 8 and 9 is a value of the number d which makes the statement

"If $0 < |t - 2| < d$, then $|7t - 14| < \epsilon$ " true.

11. Prove Theorem 5.4.

5.4. LIMITS FOR POLYNOMIAL FUNCTIONS

We have now the machinery which enables us to prove the following theorem, which theorem allows us to find the limit of any polynomial function, as we approximate any number.

Theorem 5.7:

If $P()$ is a polynomial function and b is a number, then $\lim_{t \rightarrow b} P(t)$ is the number $P(b)$.

1. $P()$ is a polynomial function by hypothesis.
2. There is a positive integer n , and there are $n + 1$ constants: $a_0, a_1, a_2, \dots, a_{n-1}, a_n$, such that $P()$ is the function specified by the equation

$$"P(t) = a_n t^n + a_{n-1} t^{n-1} + \dots + a_2 t^2 + a_1 t + a_0"$$
by step 1 and the definition of a polynomial function.
3. $\lim_{t \rightarrow b} t = b$; $\lim_{t \rightarrow b} t^2 = b^2$;
 $\lim_{t \rightarrow b} t^{n-1} = b^{n-1}$; and $\lim_{t \rightarrow b} t^n = b^n$ by n applications of Theorem 5.6.
4. $\lim_{t \rightarrow b} a_1 t = a_1 b$; $\lim_{t \rightarrow b} a_2 t^2 = a_2 b^2$;
 $\dots$; $\lim_{t \rightarrow b} a_{n-1} t^{n-1} = a_{n-1} b^{n-1}$;
and $\lim_{t \rightarrow b} a_n t^n = a_n b^n$ by step 3 and n applications of Theorem 5.4.
5. $\lim_{t \rightarrow b} a_0 = a_0$ by Theorem 5.1.
6. $\lim_{t \rightarrow b} P(t) = a_n b^n + a_{n-1} b^{n-1} + \dots + a_2 b^2 + a_1 b + a_0$ by steps 4 and 5 and n applications of Theorem 5.3.
7. $P(b) = a_n b^n + a_{n-1} b^{n-1} + \dots + a_2 b^2 + a_1 b + a_0$ by step 2.
8. $\lim_{t \rightarrow b} P(t) = P(b)$ by steps 6 and 7.

This theorem tells us that, for polynomial functions, seeking $\lim_{r \rightarrow c} P(r)$ is the same as seeking the value of the function $P()$ for the number c in the domain of $P()$. For a function which is not a polynomial function, however, this situation need not occur.

As an example of a function for which $f(a)$ and $\lim_{s \rightarrow a} f(s)$ are not the same, let us examine the function formed by multiplying a polynomial function $P()$ and the constant function $\{I/I\}()$, which consists of all ordered pairs with second element 1, except for the ordered pair $(0, 1)$. We will let the function $f()$ be specified by the equation

$$f(h) = P(h) \frac{h}{h}.$$

We have just proved that $\lim_{h \rightarrow 0} P(h)$ is the number $P(0)$. One can easily see that, for the function $\{I/I\}()$, we have

$$\lim_{h \rightarrow 0} \frac{h}{h} = 1,$$

since the function $\{I/I\}()$ is a constant function whose domain is the set of all numbers other than 0. (That is, $\{I/I\}()$ consists of all the ordered pairs of the constant function $U()$ except for the ordered pair $(0, 1)$.) Since both of the factors of $f()$ have limits as we approximate 0, Theorem 5.5 applies and we have

$$\begin{aligned} \lim_{h \rightarrow 0} \left[P(h) \frac{h}{h} \right] &= \left[\lim_{h \rightarrow 0} P(h) \right] \left[\lim_{h \rightarrow 0} \frac{h}{h} \right] \\ &= P(0) \cdot 1 \\ &= P(0). \end{aligned}$$

We have thus shown that $\lim_{h \rightarrow 0} f(h)$ is the number $P(0)$, even though 0 is not in the domain of the function $f()$.

As a specific example, let us use the function $\{I^2\}()$ to construct a new function $S_a()$. We will specify the new function by the equation:

$$S_a(h) = \frac{\{I^2\}(a+h) - \{I^2\}(a)}{h},$$

in which h is the independent variable used for $S_a()$, and a is some constant.

The number 0 is obviously not in the domain of the function $S_a()$. We have, from the definition of the identity function,

$$\begin{aligned} \{I^2\}(a+h) - \{I^2\}(a) &= (a+h)^2 - a^2 \\ &= a^2 + 2ah + h^2 - a^2 \\ &= 2ah + h^2 \\ &= (2a+h)h. \end{aligned}$$

The function $S_a()$ may then be seen to be the function specified by the equation

$$S_a(h) = \frac{(2a+h)h}{h},$$

or by the equation

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

We then have

$$\begin{aligned}
 \lim_{h \rightarrow 0} S_a(h) &= \lim_{h \rightarrow 0} \left[(2a + h) \frac{h}{h} \right] \\
 &= \left[\lim_{h \rightarrow 0} (2a + h) \right] \left[\lim_{h \rightarrow 0} \frac{h}{h} \right] . \\
 &= \left[\lim_{h \rightarrow 0} 2a + \lim_{h \rightarrow 0} h \right] \left[\lim_{h \rightarrow 0} \frac{h}{h} \right] \\
 &= (2a + 0) (1) = 2a.
 \end{aligned}$$

We will make use later on of the following theorem, which we state without proof, and without a number:

Theorem (the “pinching theorem”):

If $f()$, $g()$, and $h()$ are functions; if the interval:

$$\text{the set of all } t, \quad a \leq t \leq b$$

is in the domains of $f()$, $g()$, and $h()$; if the number c is between a and b (that is, $a < c < b$); if, for all values of the variable t in the given interval,

$$f(t) \leq g(t) \leq h(t);$$

and if

$$\lim_{x \rightarrow c} f(t) \quad \text{and} \quad \lim_{x \rightarrow c} h(t)$$

both exist and are the same number L , then

$$\lim_{x \rightarrow c} g(t)$$

also exists and is the number L .

Exercises

Evaluate the following limits, using the theorems of this and the preceding section, stating which theorem is used where.

1. $\lim_{r \rightarrow 3} (2r + 5) (r^2 - 5r + 1).$

2. $\lim_{s \rightarrow 1} \frac{(s - 1) (s^3 + s - 1)}{s - 1}.$

3. $\lim_{t \rightarrow 2} \frac{t^2 - 4}{t - 2}.$

4. $\lim_{u \rightarrow 0} (u + 1)^2.$

5. $\lim_{v \rightarrow 1} \frac{v + 1}{v + 1}.$

6. $\lim_{w \rightarrow 0} (|w| + 3)^2.$

7. $\lim_{x \rightarrow 1} (x^3 - 5x^2 + 7x + 3).$

8. $\lim_{y \rightarrow 0} \frac{y^2 + 3y}{y}.$
9. $\lim_{z \rightarrow 0} \frac{|z|}{z}.$
10. $\lim_{h \rightarrow 0} \frac{(r+h)^3 - r^3}{h}.$
11. $\lim_{r \rightarrow -2} \frac{\frac{1}{r+1} + 1}{r+2}.$
12. $\lim_{s \rightarrow 1} \frac{\sqrt{s+3} - 2}{s-1}.$
13. $\lim_{t \rightarrow 0} \frac{\sqrt{a+t} - \sqrt{a}}{t}.$
14. $\lim_{u \rightarrow 2} \frac{u^2 - 5u + 6}{u^2 - 6u + 8}.$
15. $\lim_{v \rightarrow -1} \frac{v^2 - 1}{v^2 - 2v - 3}.$
16. $\lim_{w \rightarrow 1} \frac{(w^2 + 2w - 3)^2}{(w-1)(w+5)^2}.$

5.5. DERIVATIVES

We come now to a concept most useful and used. The uses are so varied that consideration of one particular use (namely, the geometric aspect) is really as much of a hindrance later as a help now. We will then present this concept in a purely analytic manner, concerning ourselves only with functions and numbers.

Definition 5.2: If $f()$ is a function, then the *derivative* of the function $f()$ is the new function, $f'()$, which new function is specified by

$$1) f'(t) = \lim_{h \rightarrow 0} \frac{f(t+h) - f(t)}{h}.$$

2) the domain of $f'()$ consists of all the numbers in the domain of $f()$ for which the above limit exists.

It should be noted in this definition that, in 1), t is an independent variable of the function $f'()$, while h is an independent variable of the function whose limit is being found, and that, in the process of evaluation of this limit, t is not a variable. Further, it should be emphasized that the derivative of a function is always another function, and that only functions have derivatives.

From the preceding section, we see that, for $f()$ the function $\{I^2\}()$, we have $f'() = \{2I\}()$. Using the prime symbol, we have

$$\{I^2\}'() = \{2I\}().$$

In general, we will use the prime symbol to denote derivatives, putting this symbol outside of the brace used to indicate an algebraic combination of functions.

Our first concern will be with polynomial functions. In order to find the derivative of (we say, to differentiate) a polynomial function, we need to investigate constant functions, the identity function, and positive integer powers of the identity function.

Theorem 5.8:

If $f()$ is a constant function whose domain is the set of all numbers and whose range of values consists of the single number c (that is, if $f() = c()$), then $f'()$ is the function $\theta()$, where $\theta()$ is the constant function whose domain is the set of all numbers and whose range of values consists of the number 0.

Proof:

1. $f() = c()$ by hypothesis.
2. The domain of $f()$ is the set of all numbers by step 1 and the hypothesis.
3. $f'()$ is the function specified by

$$f'(t) = \lim_{h \rightarrow 0} \frac{f(t+h) - f(t)}{h}$$
by the definition of derivative.
4. For every value of the variable t , and for every value, other than 0, of the variable h ,

$$\frac{f(t+h) - f(t)}{h} = \frac{c - c}{h} = 0$$
by steps 1, 2, and 3, and various algebraic laws.
5. For every value of the variable t ,

$$\lim_{h \rightarrow 0} \frac{f(t+h) - f(t)}{h} = 0$$
by step 4 and Theorem 5.1.
6. $f'()$ is the function specified by

$$f'() = 0$$
by steps 3 and 5.
7. $f'() = \theta()$ by steps 5 and 6 and the definition of $\theta()$.

This theorem may be stated imprecisely as “the derivative of a constant is 0.” The imprecision occurs because of the words omitted. To say “the derivative of a constant is—” suggests that constants have derivatives, which is false since only functions have derivatives. To say “the derivative of something is 0” suggests that the derivative of something is a number, which is false since derivatives are always functions. The complete, precise statement of the theorem is “the derivative of a constant function is the ‘zero function,’ $\theta()$.”

Theorem 5.9:

$I'() = U()$, where $U()$ is the constant function whose domain is the set of all numbers and whose range of values consists of the number 1.

Proof:

1. $I()$ is the function specified by the equation

$$I(x) = x,$$

whose domain is the set of all numbers

by the definition of $I()$.

2. $I'()$ is the function specified by the equation

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

by the definition of derivative.

3. $I'()$ is the function specified by the equation

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

by steps 1 and 2.

4. For every value of the variable x , and for every value, other than 0, of the variable h ,

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

by various algebraic laws.

5. For every value of the variable x ,

$$I'(x) = 1$$

by steps 3 and 4 and Theorem 5.1.

6. $I'() = U()$

by step 5 and the definition of $U()$.

We may rephrase Theorem 5.9 by saying that the derivative of the identity function $I()$ is the unit function $U()$. Finally, we have

Theorem 5.10:

If n is a positive integer,
then $\{I^n\}'() = \{n I^{n-1}\}()$.

Proof:

1. $\{I^n\}()$ is the function specified by

$$\{I^n\}(x) = x^n$$

by the definition of $I()$ and the definition of the product of functions.

2. $\{I^n\}'()$ is the function specified by

$$\{I^n\}'(x) = \lim_{h \rightarrow 0} \frac{(x+h)^n - x^n}{h}$$

by step 1 and the definition of derivative.

3. For every value of the variable x ,
and every value of the variable h ,

$$(x+h)^n - x^n = (x+h-x) [(x+h)^{n-1} + (x+h)^{n-2}x + (x+h)^{n-3}x^2 + \dots + (x+h)^2x^{n-3} + (x+h)x^{n-2} + x^{n-1}]$$

by various algebraic laws.

4. For every value of the variable x ,
and every value, other than 0,
of the variable h ,

$$\frac{(x+h)^n - x^n}{h} = (x+h)^{n-1} + (x+h)^{n-2}x + (x+h)^{n-3}x^2 + \dots + (x+h)^2x^{n-3} + (x+h)x^{n-2} + x^{n-1}$$

by step 3 and various algebraic laws.

5. $\{I^n\}'()$ is the function specified by

$$\{I^n\}'(x) = \lim_{h \rightarrow 0} [(x+h)^{n-1} + (x+h)^{n-2}x + (x+h)^{n-3}x^2 + \dots + (x+h)^2x^{n-3} + (x+h)x^{n-2} + x^{n-1}]$$

by steps 2 and 4.

6. $\{I^n\}'()$ is the function specified by

$$\{I^n\}'(x) = \lim_{h \rightarrow 0} (x+h)^{n-1} + \lim_{h \rightarrow 0} (x+h)^{n-2}x + \dots + \lim_{h \rightarrow 0} (x+h)^2x^{n-3} + \lim_{h \rightarrow 0} (x+h)x^{n-2} + \lim_{h \rightarrow 0} x^{n-1}$$

by step 5 and n applications of the theorem (5.3) on the limit of the sum of functions.

7. $\lim_{h \rightarrow 0} [(x+h)^{n-2}x] = x [\lim_{h \rightarrow 0} (x+h)^{n-2}]$;

$$\lim_{h \rightarrow 0} [(x+h)^{n-3}x^2] = x^2 [\lim_{h \rightarrow 0} (x+h)^{n-3}]$$

.....

$$\lim_{h \rightarrow 0} [(x+h)^2x^{n-3}] = x^{n-3} [\lim_{h \rightarrow 0} (x+h)^2]$$

$$\lim_{h \rightarrow 0} [(x+h)x^{n-2}] = x^{n-2} [\lim_{h \rightarrow 0} (x+h)]$$

by $n-2$ applications of the theorem (5.4) on the limit of the product of a function and a constant function.

8. $\lim_{h \rightarrow 0} x^{n-1} = x^{n-1}$

by the theorem (5.1) on the limit of a constant function.

9. $\lim_{h \rightarrow 0} (x+h)^{n-1} = [\lim_{h \rightarrow 0} (x+h)]^{n-1}$;

$$\lim_{h \rightarrow 0} (x+h)^{n-2} = [\lim_{h \rightarrow 0} (x+h)]^{n-2}$$

$$\lim_{h \rightarrow 0} (x+h)^{n-3} = [\lim_{h \rightarrow 0} (x+h)]^{n-3}$$

...

$$\lim_{h \rightarrow 0} (x+h)^2 = [\lim_{h \rightarrow 0} (x+h)]^2$$

by $n-2$ applications of the theorem (5.5) on the limit of the product of two functions.

$$0. \lim_{h \rightarrow 0} (x + h) = x$$

by the theorem (5.7) on the limit of a polynomial function.

1. $\{I^n\}'()$ is the function specified by

$$\{I^n\}'(x) = n x^{n-1}$$

by steps 6, 7, 8, 9, and 10.

2. $\{I^n\}'() = \{n I^{n-1}\}()$

by step 11.

This theorem tells us that the derivative of the function specified by the expression x^n , where n is a fixed positive integer and x is an independent variable, is the function defined by the expression $n x^{n-1}$. Of course, it is equally true that the derivative of the function specified by t^n , where n is a fixed positive integer is the function specified by the expression $n t^{n-1}$, where t is an independent variable of the original function and r is an independent variable of the derivative. Thus, the derivative of the function specified by x^3 is the function specified by $3x^2$; the derivative of the function specified by y^4 is the function specified by $4y^3$; and so on. It should be noted that we have proved nothing about a function defined by x^p where p is any number other than a positive integer, so that in order to discuss functions other than polynomials, we will have to prove further theorems about derivatives.

We have still not completed the full list of theorems needed to differentiate (meaning to find the derivative of) any polynomial function. A glance at the definition of polynomial function shows that we need a theorem about the product of a given function and a constant function, and another theorem about the derivative of the sum of two functions.

Theorem 5.11:

If $f()$ is a function and $c()$ is a constant function, then

$$\{cf\}'() = \{cf'\}().$$

Proof:

1. $\{cf\}'()$ is the function specified by

$$\{cf\}'(x)$$

$$= \lim_{h \rightarrow 0} \frac{\{cf\}(x+h) - \{cf\}(x)}{h}$$

by the definition of derivative.

2. $\{cf\}'()$ is the function specified by

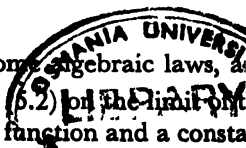
$$\{cf\}'(x) = \lim_{h \rightarrow 0} \frac{cf(x+h) - cf(x)}{h}$$

by step 1, the definition of the product of two functions, and the hypothesis that $c()$ is a constant function.

3. $\{cf\}'()$ is the function specified by

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

by step 2, some algebraic laws, the theorem (5.2) on the limit of the product of a function and a constant function.



4. $\{cf\}'()$ is the function specified by

$$\{cf\}'(x) = cf'(x)$$
by step 3 and the definition of derivative.
5. $\{cf\}'()$ is the function specified by

$$\{cf\}'(x) = \{cf'\}(x)$$
by step 4 and the definition of the product of two functions.
6. $\{cf\}'() = \{cf'\}()$
by step 5.

In words, this theorem says that the derivative of the product of a function and a constant function is the product of that constant function and the derivative of the original function. Thus, the derivative of the function specified by $6x^3$ is the function defined by $18x^2$, as the product of 6 and $3x^2$.

Theorem 5.12:

If $f()$ and $g()$ are functions, then

$$\{f + g\}'() = \{f' + g'\}().$$

Proof:

1. Let a be a number in the domains of both $f'()$ and $g'()$.
2. a is in the domain of $f()$ and $g()$
by step 1 and the definition of derivative.
3. a is in the domain of $\{f' + g'\}()$
by step 1 and the definition of the sum of two functions.
4. a is in the domain of $\{f + g\}()$
by step 2 and the definition of the sum of two functions.
5. $\lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}$ exists and is
the number $f'(a)$
by step 1 and the definition of derivative.
6. $\lim_{h \rightarrow 0} \frac{g(a+h) - g(a)}{h}$ exists and is
the number $g'(a)$
by step 1 and the definition of derivative.
7. $\lim_{h \rightarrow 0} \left[\frac{f(a+h) - f(a)}{h} + \frac{g(a+h) - g(a)}{h} \right]$
exists and is the number
 $f'(a) + g'(a)$
by steps 5 and 6 and the theorem (5.3) on the limit of the sum of two functions.
8. $\lim_{h \rightarrow 0} \frac{[f(a+h) + g(a+h)] - [f(a) + g(a)]}{h}$
exists and is the number
 $f'(a) + g'(a)$
by step 7 and various algebraic laws.

9. $\lim_{h \rightarrow 0} \frac{\{f+g\}(a+h) - \{f+g\}(a)}{h}$
 exists and is the number $f'(a) + g'(a)$ by step 8 and the definition of the sum of two functions.
10. $\{f+g\}'(a)$ exists and is the number $f'(a) + g'(a)$ by step 9 and the definition of derivative.
11. a is in the domain of $\{f+g\}'()$ by step 10 and the definition of derivative.
12. $\{f+g\}'(a)$ exists and is the number $\{f' + g'\}(a)$ by steps 3 and 11 and the definition of the sum of two functions.
13. $\{f+g\}'() = \{f' + g'\}()$ by steps 1 and 12.

This theorem, which says that the derivative of the sum of two functions is the sum of their derivatives, together with the previous theorems, allows us to find the derivative of any polynomial function. For example, if

$$P() = \{I^5 - 3I^4 + 7I^3 + 2I^2 + 4I - 2\}(),$$

then

$$P'() = \{\{I^5\}' + \{-3I^4\}' + \{7I^3\}' + \{2I^2\}' + \{4I\}' + \{-2\}'\}()$$

by Theorem 5.12,

$$P'() = \{\{I^5\}' + \{-3I^4\}' + \{7I^3\}' + \{2I^2\}' + \{4I\}'\}()$$

by Theorem 5.8,

$$P'() = \{\{I^5\}' - 3\{I^4\}' + 7\{I^3\}' + 2\{I^2\}' + 4I'\}()$$

by Theorem 5.12,

$$P'() = \{5I^4 - 12I^3 + 21I^2 + 4I + 4\}()$$

by Theorem 5.10.

In terms of algebraic expressions, the derivative of the polynomial function given by the expression

$$t^5 - 3t^4 + 7t^3 + 2t^2 + 4t - 2$$

is the function given by the expression

$$5t^4 - 12t^3 + 21t^2 + 4t + 4.$$

Exercises

1. For the function $f()$, specified by $f(x) = x^2$, find

a) $f(3)$.

b) $f(3+h)$.

c) $\frac{f(3+h) - f(3)}{h}$.

d) $f'(3)$.

2. For the function $g()$, specified by $g(t) = -3$, find

- a) $g(1)$.
- b) $g(1 + h)$.
- c) $\frac{g(1 + h) - g(1)}{h}$.
- d) $g'(1)$.

3. For the function $f()$, specified by $f(x) = mx + b$, where m and b are constants and $m \neq 0$, find

- a) $f(a)$.
- b) $f(a + h)$.
- c) $\frac{f(a + h) - f(a)}{h}$.
- d) $f'(a)$.

Find the derivatives of the polynomial function specified by

- 4. $r^3 - 4r + 7$.
- 5. $5s^{10} - 12s^7 + 3s^4 - 2s^3 + 15s^2 + 6s - 72,196$.
- 6. $2t^5 - \sqrt{3}t^4 + 4\sqrt{5}t^2 - 15$.
- 7. $4u^5 + 7u^4 - 2u^3 + 17$.
- 8. $-3v^4 + 12v^2 - v + 1/2$.
- 9. $8w - 15w^5 + 5w^{12} - 3$.
- 10. $x^7 + 3x^5 - (1/8)x^4 + (2/3)x^3 - \sqrt{5}x^2 + 10$.
- 11. $(1/3)y^3 - (1/2)y^2 + y - 2$.
- 12. $(1/7)z^7 + (1/6)z^6$.

5.6. THE DERIVATIVE AS A RATE OF CHANGE

We look again at the now well-known function $\{I^2\}()$ —the function specified by the equation $y = x^2$, where x is an independent variable and y is a dependent variable. Given a value of the variable x , there will be a corresponding value of the variable y . If we change the value of the variable x , then the corresponding value of the variable y is necessarily changed. The question: “How much does the variable y change for a given change in the variable x ?” is a loaded one. Let us examine some specific examples for this function, so that we may answer this loaded question.

If 2 is a value of the variable x , then 4 is the corresponding value of the variable y . If the value of the variable x is changed by 0.1 units, then the new value of the variable x will be either 1.9 ($= 2 - 0.1$) or 2.1 ($= 2 + 0.1$), depending on whether the change in the value of x was negative or positive. Thus, a change of 0.1 units in this value of the variable x induces a change of either 0.39 or 0.41 units in the corresponding value of the variable y .

If the value of the variable x is changed from 2 by 0.01 units, then the new value of x will be either 1.99 ($= 2 - 0.01$) or 2.01 ($= 2 + 0.01$), and the new value of the variable y will be either 3.9601 or 4.0401, again depending on whether the change in the value of x was negative or positive. Then a change of 0.01 in the value of the variable x induces a change of either 0.0399 or 0.0401 units in the value of y .

Speaking roughly, a change in the value of x of 0.1 units induces a change in the value of y of 0.4 units, and a change in the value of x of 0.01 units induces a change in the value of y of approximately 0.04 units. Further investigation of small changes in the value of the variable x would reveal that the induced change in the value of y would be approximately 4 times as large as the inducing change in the value of x .

If 3 is a value of the variable x , then 9 is the corresponding value of the variable y . If the value of x is changed by 0.1 units, the new value of this variable will be either 2.9 ($= 3 - 0.1$) or 3.1 ($= 3 + 0.1$), and the new value of y will be either 8.41 or 9.61, depending on whether the change in the value of x was negative or positive. Then a change of 0.1 units in the value of x induces a change of either 0.59 or 0.61 units in the value of y .

If the value of x is changed from 3 by 0.01 units, then the new value of this variable will be either 2.99 ($= 3 - 0.01$) or 3.01 ($= 3 + 0.01$), and the new value of y will be either 8.9391 or 9.0601, again depending on whether the change in x was negative or positive. Then a change of 0.01 units in the value of x induces a change of either 0.0599 or 0.0601 units in the value of y .

Speaking roughly, a change, from 3, in the value of the variable x of 0.1 units induces a change in the value of y of approximate 0.6 units, and a change in the value of x of 0.01 units induces a change in the value of the variable y of approximately 0.06 units. Further investigation of small changes in the value of the variable x would reveal that the induced change in the value of y would be approximately 6 times as large as the inducing change in the value of x .

Why did the variable y change 4 times as fast as the variable x in one case, and 6 times as fast in the other? Evidently, the manner in which a change in the value of x induces a change in the value of y is not always the same, but depends somehow on which value of x we start with. We have found the ratio of these changes were given by the quotients

$$\frac{(2.1)^2 - (2)^2}{0.1} = 4.1;$$

$$\frac{(1.9)^2 - (2)^2}{-0.1} = 3.9;$$

$$\frac{(2.01)^2 - (2)^2}{0.01} = 4.01;$$

$$\frac{(1.99)^2 - (2)^2}{-0.01} = 3.99;$$

and

$$\frac{(3.1)^2 - (3)^2}{0.1} = 6.1;$$

$$\frac{(2.9)^2 - (3)^2}{-0.1} = 5.9;$$

$$\frac{(3.01)^2 - (3)^2}{0.01} = 6.01;$$

$$\frac{(2.99)^2 - (3)^2}{-0.01} = 5.99.$$

These quotients are all of the form

$$\frac{f(a+h) - f(a)}{h},$$

where $f() = \{t^2\}()$, and for the first four, $a = 2$, $h = 0.1, -0.1, 0.01, -0.01$; and for the second four, $a = 3$, $h = 0.1, -0.1, 0.01, -0.01$. This quotient is precisely the one encountered in the definition of the derivative of the function, namely,

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

We may then answer our loaded question by saying that the ratio of an induced change in the value of y to an inducing change in the value of x is approximately the value of the derivative of the function having y for a dependent variable and x an independent variable. For this reason we have

Definition 5.3: Given the function $f()$, which has an independent variable r and a dependent variable z , the (*instantaneous*) *rate of change* of z with respect to r is a dependent variable of the function $f'()$.

From this definition, the phrase “rate of change” is meaningful only when followed by the phrase “of one variable with respect to another,” and this rate of change is itself a variable, and thus has values. The adjective “instantaneous” is often used to emphasize certain physical aspects of a situation.

EXAMPLE 1:

In physics, a “formula” which relates the distance which a freely falling body travels and the time which has elapsed since the beginning of the fall is

$$s = 16t^2.$$

This “formula” is an equation, which equation will give a function, an independent variable of which function is t and a dependent variable of which function is s . A positive value of the variable t is interpreted as the number of seconds which have elapsed since the beginning of the fall and

a (necessarily positive) corresponding value of the variable s is interpreted as the number of feet which the body has fallen. Because of this physical interpretation, we restrict the range of the variable t to the set of all non-negative numbers. Thus, we are dealing with the function specified by the modified equation

$$s = 16t^2, \quad t \geq 0.$$

By what has already been discovered about derivatives, the derivative of the function we are handling is the function specified by the modified algebraic expression

$$32t, \quad t > 0,$$

where t is an independent variable. If we introduce a dependent variable—say the variable v —then the derivative of our function is now specified by the modified equation

$$v = 32t, \quad t > 0.$$

By the definition (5.3), the variable v is the (instantaneous) rate of change of s with respect to t . In the physical interpretation of the mathematics, the variable v is the instantaneous rate of change of the distance (in feet) traveled with respect to the time (in seconds) elapsed. It is but a small step to insert the word “velocity” for this rate of change, and this step is precisely the definition of the word “velocity.” It should be realized, however, that this insertion of the word “velocity” is a deliberate act in theoretical physics, and the physicist is actually making a formal definition of the word itself. In our terminology, when an independent variable of a function has values which are interpreted as the distance from a fixed point of an object moving along a straight line, and a dependent variable of that function has values which are interpreted as the time elapsed since the moving object began moving, then, and only then, are the values of a dependent variable of the derivative of that function interpreted as the velocity of that moving object.

In our specified “formula,” s had values expressed in feet, and t had values expressed in second. The velocity v thus has values expressed in feet per second, and we may say that a falling body moves at a velocity of 32 feet per second after 1 second has elapsed, at 64 feet per second after 2 seconds have elapsed, and so on. (As incidental information: 88 feet per second = 60 miles per hour.)

EXAMPLE 2:

In geometry, a “formula” which relates the area of a circle to its radius is

$$A = \pi r^2.$$

Because of the geometrical interpretation of the values of these variables, as lengths of lines or areas of circles, we cannot have negative values of these variables, so we will take the function specified by the modified equation

$$A = \pi r^2, \quad r > 0,$$

where r is an independent variable and A is a dependent variable. The rate of change of the variable A with respect to the variable r will be a dependent variable of the function specified by the modified expression

$$2\pi r, \quad r > 0.$$

If we like, we may interpret this result geometrically by saying that the rate of change of (the number of square units of) area with respect to the (number of units of length of the) radius of a circle is (the number of units of length of) the circumference. Thus, changing the radius of a circle from 10 feet to 10.1 feet will change the area of that circle by 2π ($= 20\pi$ (0.1)) square feet, approximately—that is, approximately 6.28 square feet.

EXAMPLE 3:

In economics, if the total cost of a product and the number of units produced are related by the “formula”

$$M = f(n),$$

where the variable n has, as its range, the set of non-negative integers, the variable M has, as its range, a set of positive numbers, and the function $f()$ has a derivative, then the cost of making n objects is taken to be M dollars. The economist then defines the phrase “marginal cost” to be a dependent variable of the derivative $f'()$, of the function $f()$.

EXAMPLE 4:

If a hemispherical bowl is filled with liquid, then one can measure the rate at which the liquid enters (the rate of change of the volume occupied) and one can also measure the depth of the liquid. There is a “formula”

$$V = \frac{\pi h^2}{3} (3r - h)$$

which gives the volume V of that portion of the sphere of radius r which is occupied by a liquid of depth h . If we have a hemispherical bowl of radius 10 feet, then the equation

$$V = \frac{\pi h^2}{3} (30 - h)$$

will specify a function with independent variable h and dependent variable V . Since we are interpreting a value of the variable h as a depth of liquid in this hemispherical bowl, we restrict the range of this variable to be the interval:

$$\text{the set of all } h, \quad 0 \leq h \leq 10.$$

A simple exercise in manipulation with inequalities shows that the dependent variable will then have values in the interval:

$$\text{the set of all } V, \quad 0 \leq V \leq \frac{2000}{3} \pi.$$

If we now consider the function $V()$, given by

$$V(h) = \frac{\pi h^2}{3} (30 - h), \quad 0 \leq h \leq 10,$$

we have

$$V(h) = 10\pi h^2 - \frac{\pi}{3} h^3, \quad 0 \leq h \leq 10.$$

$$V'(h) = 20\pi h - \pi h^2, \quad 0 < h < 10.$$

$$= \pi h(20 - h), \quad 0 < h < 10.$$

The rate of change of volume with respect to depth then depends on depth, and also then the rate of change of depth with respect to volume depends on depth. For a specific instance, consider the tank already holding liquid to the depth of 5 feet. Then, since $V'(5) = 75\pi (= 5\pi(20-5))$, changing the depth by 1 inch ($= 1/12$ feet) will change the volume by approximately $\frac{75}{12}\pi$ (approximately 19.65) cubic feet. Speaking roughly, then, when the tank has 5 feet of water in it, it will take just a little less than 20 cubic feet (approximately 147 gallons) of liquid to raise the level one inch. For another instance, suppose there is 6 feet of liquid in the tank. $V'(6) = 84\pi$. Changing the depth by 1 inch will change the volume approximately $\frac{84}{12}\pi (= 7\pi)$ cubic feet. Speaking roughly, then, when the tank has 6 feet of water in it, it will take about 22 cubic feet (approximately 165 gallons) of liquid to raise the level one inch.

Exercises

Find the following rates of change:

1. The volume of a sphere with respect to its radius.
2. The surface area of a sphere with respect to its radius.
3. The area of a square with respect to the length of a side.
4. The perimeter of a square with respect to the length of a side.
5. The volume of a cube with respect to the length of a side.
6. The surface area of a cube with respect to the length of a side.
7. A projectile shot upward will have height given by

$$s = -16t^2 + 3200t, \quad 0 \leq t \leq 200,$$

where t represents the time (in seconds) elapsed from firing, and s the height (in feet) reached.

- a) Find an expression whose values give the velocity of the projectile.
 - b) What is the range of values of the function given above with dependent variable s ?
 - c) What is the range of values of the velocity?
 - d) How high will the projectile go?
8. The radius of a sphere is changed from 10 to 10.2 feet. What is an approximate change in the volume and surface area of that sphere?

9. The circumference of the earth at the equator is 25,000 miles. If a satellite were to orbit at a height of 50 miles, remaining always above the equator, approximately how far would the satellite travel in each circuit of its orbit?
10. Why is the derivative of the function specified by $16t^2$, $t \geq 0$, the function specified by $32t$, $t > 0$, and not the function specified by $32t$, $t \geq 0$?

5.7. TANGENT LINES

In this section, we will seek a connection between the value of the derivative of a function and the graph of that function. The quotient

$$\frac{f(a+h) - f(a)}{h},$$

which appears in the definition of the derivative of a function may be interpreted geometrically as follows.

The number $f(a+h)$ is the y -coordinate of that point on the graph of the function $f()$ whose x -coordinate is the number $a+h$. The number $f(a)$ is the y -coordinate of that point on the graph of the function $f()$ whose x -coordinate is the number a . If h is positive, the point with coordinates $(a+h, f(a+h))$ will be to the right of the point with coordinates $(a, f(a))$. If h is negative, the point with coordinates $(a+h, f(a+h))$ will be to the left of the point with coordinates $(a, f(a))$.

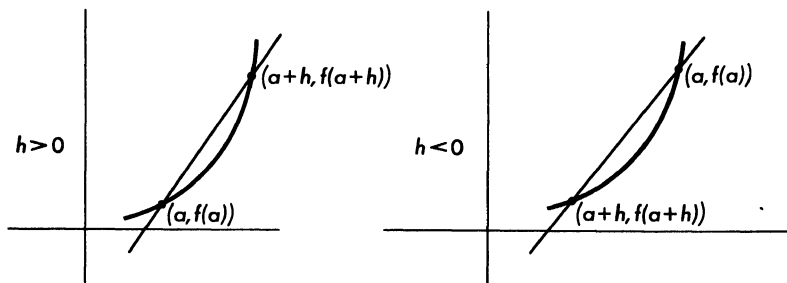


Fig. 5.1

Regardless of whether h is negative or positive, the number

$$\frac{f(a+h) - f(a)}{(a+h) - a}$$

is the slope of the line through the points $(a+h, f(a+h))$ and $(a, f(a))$. Then, for each value of the variable h (other than 0, of course), there is a straight line through the point $(a, f(a))$, which line has slope

$$\frac{f(a+h) - f(a)}{h}.$$

Asking if the numbers which are this quotient squeeze down on one fixed number as we squeeze our view of the values of the variable h down on the number 0 (that is, asking for $f'(a)$) is also asking if these straight lines come closer and closer to one fixed line. Because of this geometric situation we have

Definition 5.4: Given a function $f()$ and a number a in the domain of $f'()$, the *tangent line* to the graph of $f()$, at the point with coordinates $(a, f(a))$ is that line through that point whose slope is the number $f'(a)$.

It should be noted that this definition agrees with our intuitive concept of what a tangent line should be, but that heretofore the only precisely defined tangent lines in geometry were tangents to circles. While this definition agrees with an intuitive idea of tangent line, the expressing of this intuitive idea of the line's "touching the curve only once," or "touching, but not crossing" is too restrictive.

For example, the tangent line to the graph of the polynomial function $\{I^3 - 3I\}()$, the graph of the equation

$$y = x^3 - 3x,$$

at the point with coordinates $(-1, 2)$ is that line through $(-1, 2)$ which has slope $\{I^3 - 3I\}'(-1)$. But $\{I^3 - 3I\}'()$ is the function $\{3I^2 - 3\}()$, so that the slope of this tangent line is the number 0. Thus, a horizontal line through the point $(-1, 2)$ is tangent to the graph of the equation $y = x^3 - 3x$.

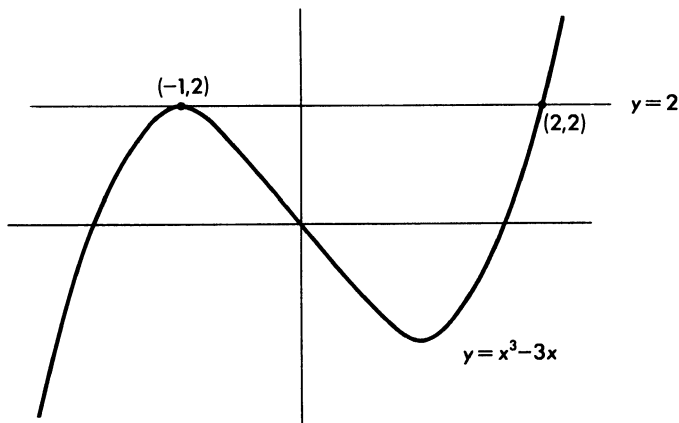


Fig. 5.2

This horizontal line also intersects the graph of $y = x^3 - 3x$ at the point $(2, 2)$, so that it is not really valid to insist that a tangent line "touch the graph only once."

As another example, the tangent line to the graph of the function $\{I^3\}()$ at the point with coordinates $(0, 0)$ has slope $\{I^3\}'(0) = \{3I^2\}() = 0$.

This horizontal line is the x -axis itself, and the graph of the equation $y = x^3$ is above the x -axis to the right of the origin and below the x -axis to the left of the origin. Thus, we have a tangent line which actually crosses the curve at the point of tangency, so that it is short-sighted to insist that a tangent line "touch, but not cross" the curve.

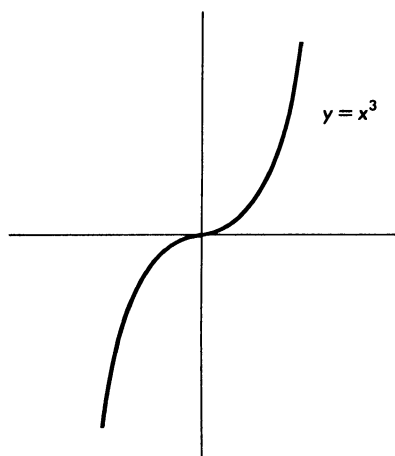


Fig. 5.3

We can write an equation of a straight line if we know the coordinates of one of its points and its slope. Hence, if we are given a function and the coordinates of a point on the graph of that function, we can immediately write an equation of the tangent line to the graph at that point (if such a tangent line exists). For example, an equation of the tangent line to the graph of the equation

$$y = 4x^5 - 8x^2 - 2x,$$

at the point $(1, -6)$ will be

$$y + 6 = 2(x - 1)$$

or

$$2x - y = 8.$$

In general, an equation of the tangent line to the graph of the function $f()$ at the point $(a, f(a))$ will be

$$y - f(a) = f'(a)(x - a).$$

It should be noted that the number $f'(a)$, appearing in this equation is, and must be, a *number*, and not any expression involving any variable, particularly not the variable x .

Exercises

Find an equation of the tangent line to the graph of the given equation at the indicated point or points.

1. $y = x^2$; $(2, 4)$.
2. $y = x^2$; $(-2, 4)$.
3. $y = 3x + 5$; $(1, 8)$.
4. $y = x^3 - 3x^2$; $(1, -2)$, $(0, 0)$, $(2, 2)$.
5. $y = x^4 + 5x^3 + 3x^2 + 2x - 10$; $(0, -10)$, $(1, 1)$.
6. Find an equation of the tangent line to the graph of the equation $y = x^2$, which tangent line has slope 10.

7. Find an equation of a tangent line to the graph of $y = x^3$, which tangent line has slope 27.
8. Why does exercise 6 say "the tangent line" and exercise 7 says "a tangent line"?
9. Find an equation of the tangent line to the graph of $y = x^3$, which tangent line is perpendicular to the line with equation $x + 12y = 15$.
10. Two lines are drawn from the point $(7/3, -2)$ tangent to the graph of $y = x^3 - 3x$. Find the points of tangency of the two lines, and an equation of each.

5.8. INCREASING AND DECREASING FUNCTIONS

It is customary, in reading and writing English, to do so from left to right. In like manner, it is customary, when looking at the domain of a function, to look "from left to right." That is, one normally looks at the numbers in the domain of a function (or at any set of numbers) by looking from the smaller numbers to the larger. This manner of examining the domain is actually equivalent to looking at the graph of the function from left to right on the plane.

The intuitive idea of the word "increasing," as applied to the graph of a function, is synonymous with the idea of a graph's "rising." In such a case, as we look from left to right at the points on the graph, our glance travels upward. Similarly, "decreasing" would mean that our glance at the graph moves downward, as we sweep our eyes from left to right. In this section, we will attempt to make the intuitive ideas precise.

First of all, we will look at the graph of a function, not in its entirety, but in a point-by-point manner. That is, we will look at one specific point—say the point with coordinates $(a, f(a))$ —and see what we mean by saying that the graph is increasing at this one point. Intuitively, such a statement should mean that nearby points of the graph which are to the right of $(a, f(a))$ should have larger y -coordinates than the number $f(a)$, and that nearby points of the graph which are to the left of $(a, f(a))$ should have smaller y -coordinates than the number $f(a)$. A precise statement must avoid the vague word "nearby," and, for purely technical reasons, we will only ask that this intuitive idea be not violated, rather than strictly met.

Definition 5.5: Given a function $f()$ and a number a in the domain of $f()$, $f()$ is *increasing at a* provided that there is an interval:

$$\text{the set of all } t, \quad |t - a| < d,$$

where d is some positive number, which interval is entirely contained in the domain of $f()$, for which:

$$\begin{aligned} \text{if } a - d < t < a, \quad \text{then } f(t) &\leq f(a); \\ \text{if } a < t < a + d, \quad \text{then } f(t) &\geq f(a). \end{aligned}$$

Definition 5.6: Given a function $f()$ and a number b in the domain of $f()$, $f()$ is *decreasing at b* provided that there is an interval:

$$\text{the set of all } t, \quad |t - b| < d,$$

where d is some positive number, which interval is entirely contained in the domain of $f()$, for which:

$$\text{if } b - d < t < b, \quad \text{then } f(t) \geq f(b);$$

$$\text{if } b < t < b + d, \quad \text{then } f(t) \leq f(b).$$

It is worthy of note that these definitions, though inspired by intuitive geometric ideas, are purely analytic and contain no reference to the geometry of the graph of a function. The geometry of the graph does, however, provide us with a useful tool by suggesting relationships. For instance, let us combine this analytic definition of a function's increasing at a number with the analytic definition of the derivative of that function. In geometric terms, we are trying to find some relation between a "rising" graph and a tangent line to that graph. Since the value of the derivative of a function is the slope of the line which is tangent to the graph of that function—the number in the domain of the derivative being the same as the x -coordinate of the point of tangency—we would expect that a positive slope for the tangent line would mean a rising graph, or that the function would be increasing at this point. In order to prove that such is actually the case, we need first a theorem on limits.

Theorem 5.13:

If $g()$ is a function, p is a number, $\lim_{t \rightarrow p} g(t)$ is the number B , $B > 0$, and $g(p) > 0$;
then there is an interval:

$$\text{the set of all } t, \quad |t - p| < d,$$

such that if $|t - p| < d$, then $g(t) > 0$.

This theorem says that if $\lim_{t \rightarrow p} g(t)$ is the positive number B and if $g(p)$ (which may well be the number B) is also positive, then, not only are the points (p, B) and $(p, g(p))$ (which may be one and the same point) above the x -axis, but there is some piece of the graph of $g()$ which surrounds the point (p, B) and is entirely above the x -axis.

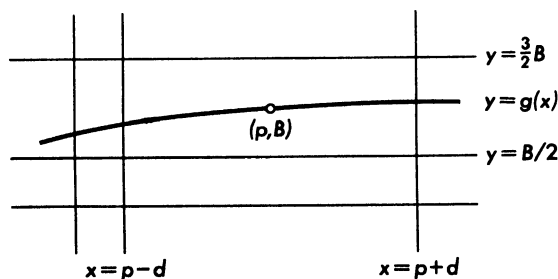


Fig. 5.4

Proof of 5.13:

1. $\lim_{t \rightarrow p} g(t) = B$ by hypothesis.
2. $B > 0$ by hypothesis.
3. $B/2 > 0$ by step 2.
4. Corresponding to the positive number $B/2$, there exists a positive number d , such that
 if $0 < |t - p| < d$,
 then t is in the domain of $g()$ and
 $|g(t) - B| < B/2$ by steps 1 and 3 and the definition of limit.
5. There exists a positive number d , such that
 if $0 < |t - p| < d$,
 then t is in the domain of $g()$ and
 $B/2 < g(t) < \frac{3}{2}B$ by step 4 and the definition of absolute value.
6. There exists a positive number d , such that
 if $0 < |t - p| < d$,
 then t is in the domain of $g()$ and
 $g(t) > B/2$ by step 5.
7. There exists a positive number d , such that
 if $0 < |t - p| < d$,
 then t is in the domain of $g()$ and
 $g(t) > 0$ by steps 3 and 6.
8. $g(p) > 0$ by hypothesis.
9. There exists a positive number d , such that
 if $|t - p| < d$,
 then t is in the domain of $g()$ and
 $g(t) > 0$ by steps 7 and 8.
10. There exists an interval:
 the set of all t , $|t - p| < d$,
 such that if $|t - p| < d$, then
 $g(t) > 0$ by step 9.

We can also prove a similar theorem:

Theorem 5.14:

If $g()$ is a function, q is a number, $\lim_{s \rightarrow q} g(s)$ is the number C , $C < 0$, and $g(q) < 0$; then there is an interval:

$$\text{the set of all } s, \quad |s - q| < d,$$

such that if $|s - q| < d$, then $g(s) < 0$.

These two theorems say that knowledge of the sign the number $\lim_{t \rightarrow a} g(t)$ and $g(a)$ (which, as we saw, are the same number in the case of a polynomial function, but need not be the same in the case of a function not a polynomial function) gives us knowledge of that part of the range of values of $g()$ corresponding to some (small) interval containing the number a . Since the derivative of a function is defined by means of the limit of a new function formed from the original one, we will use these two theorems to prove:

Theorem 5.15:

If $f()$ is a function, if a is a number in the domain of $f'()$, and if $f'(a) > 0$; then $f()$ is increasing at a .

Proof:

1. $f'(a) > 0$

by hypothesis.

2. $f'(a)$ is the number

$$\lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}$$

by the definition of derivative and the hypothesis that a is in the domain of $f'()$.

3. Let the function $g()$ be specified by

$$g(h) = \frac{f(a+h) - f(a)}{h}, \quad \text{for } h \neq 0;$$

$$g(0) = f'(a).$$

4. $g()$ satisfies the hypotheses of Theorem 5.13

by steps 1-3.

5. There is an interval:

$$\text{the set of all } h, \quad |h| < d$$

for which, if $|h| < d$, then

$$g(h) > 0$$

by step 4 and Theorem 5.13.

6. There is an interval:

the set of all h , $|h| < d$,
for which: if $-d < h < 0$, then

$$\frac{f(a+h) - f(a)}{h} > 0$$

and if $0 < h < d$, then

$$\frac{f(a+h) - f(a)}{h} > 0$$

by steps 3 and 5.

7. There is an interval:

the set of all h , $|h| < d$,
for which: if $-d < h < 0$, then

$$f(a+h) - f(a) < 0$$

and if $0 < h < d$, then

$$f(a+h) - f(a) > 0$$

by step 6 and the fact that a quotient is positive if and only if both numerator and denominator have the same sign.

8. There is an interval:

the set of all h , $|h| < d$,
such that: if $-d < h < 0$, then

$$f(a+h) < f(a)$$

and if $0 < h < d$, then

$$f(a+h) > f(a)$$

by step 7 and O-3.

9. There is an interval:

the set of all t , $|t - a| < d$,
such that: if $a - d < t < a$, then

$$f(t) < f(a)$$

and if $a < t < a + d$ then

$$f(t) > f(a)$$

by step 8 and relating the variables t and h by $t = a + h$.

10. There is an interval:

the set of all t , $|t - a| < d$,
such that: if $a - d < t < a$, then

$$f(t) \leq f(a)$$

and if $a < t < a + d$, then

$$f(t) \geq f(a)$$

by step 9 and the fact that $r < s$ implies $r \leq s$.

11. $f()$ is increasing at a

by step 10 and the definition of a function which increases at a number.

We may also prove the analogous theorem:

Theorem 5.16:

If $f()$ is a function, if b is a number in the domain of $f'()$, and if $f'(b) < 0$; then $f()$ is decreasing at b .

These theorems give a sufficient condition that a function be increasing or decreasing at a number in its domain. For a polynomial function, every number is in the domain of both the function and its derivative. We may then find the set of all the numbers for which a polynomial function is increasing (or decreasing) by finding the solution of an inequality of the form $P'(x) > 0$ (or $P'(x) < 0$).

For example, let $P()$ be the function specified by

$$P(x) = x^3 - 3x^2 - 9x + 4.$$

By the theorems on differentiation, $P'()$ will be the function specified by

$$P'(x) = 3x^2 - 6x - 9.$$

The solution of the inequality

$$3x^2 - 6x - 9 > 0$$

is also the solution of

$$x^2 - 2x - 3 > 0$$

and of

$$(x - 3)(x + 1) > 0.$$

The solution of this last inequality consists of the two intervals:

the set of all x , $x < -1$

and the set of all x , $x > 3$.

Thus, the original polynomial function $P()$ is increasing at every number in either of these two intervals. We may use this fact to sketch the graph of $P()$ —the graph of the equation $y = x^3 - 3x^2 - 9x + 4$ —by drawing a curve which (coming from the left) rises to the point $(-1, 9)$ and which (going to the right) rises from the point $(3, -23)$.

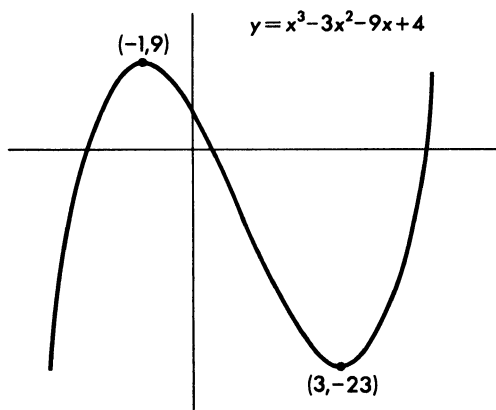


Fig. 5.5

As a further example, let $f()$ be the function given by

$$f(t) = 4t^2 - 8t + 5.$$

The derivative $f'()$ is specified by

$$f'(t) = 8t - 8.$$

The solution of $8t - 8 < 0$ is the interval:

the set of all t , $t > 1$.

The graph of $f()$ (that is, the graph of $y = 4x^2 - 4x + 5$), when scanned from left to right, decreases to the point $(1, 1)$ and increases to the right of this point.

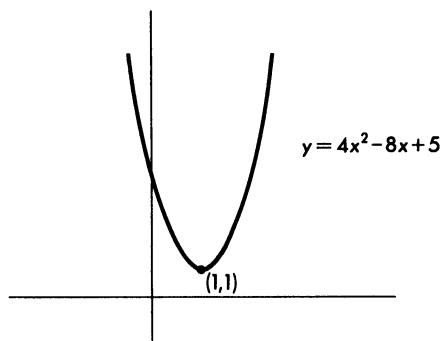


Fig. 5.6

In general, any function whose derivative has the set of all numbers as domain will be increasing at any number which is a solution of the inequality $f'(x) > 0$ and will be decreasing at any number which is a solution of the inequality $f'(x) < 0$.

In Chapter 2, we noticed that an inequality may have, as the solution, the set of all numbers. In the terminology of Chapter 4, this amounts to saying that a function may have, as range of values, a set consisting of positive numbers only. Such a function may well be the derivative of another function, which other function will then be increasing at every number in its domain. To cover this situation, we have:

Definition 5.7: A function is *increasing* if that function is increasing at every number in its domain.

Definition 5.8: A function is *decreasing* if that function is decreasing at every number in its domain.

Exercises

1. Is a constant function increasing at a number in its domain? Why or why not?
2. Find an interval containing the number $1/2$ in which the function given by $f(t) = t^3$ has positive values only.

3. Explain in detail the reasons for steps 6 and 9 of the proof of Theorem 5.13.
4. Prove Theorem 5.14.
5. For the function specified by $f(z) = z^2$:
 - a) Find $f'(2)$.
 - b) Find $\frac{f(2+h) - f(2)}{h}$, for $h = 0.1$ and $h = -0.1$.
 - c) What is the ordering of the numbers $f(1.9)$, $f(2)$, $f(2.1)$?
 - d) Repeat a) through c) for the number -2 rather than the number 2 .
6. Prove Theorem 5.16.
7. For what numbers is the function $f()$, specified by each of the following equations, increasing, and decreasing. Use this information to sketch the graph of each function.

a) $f(r) = r^3$.	c) $f(t) = 3t^2 - t + 5$.
b) $f(s) = s^3 - 3s$.	d) $f(u) = 2u^2 + 3u - 17$.
8. Which of the functions defined by the following expressions is increasing? Which is decreasing?

a) v^2 .	c) $4 - x^2$.
b) w^3 .	d) $5 - y^3$.

5.9. HIGHER DERIVATIVES AND ANTI-DERIVATIVES

By definition, the derivative of a function is another function. The new function may well have a derivative of its own, which is related to the original function by

Definition 5.9: If $f()$ is a function, the *second derivative* of $f()$ —written $f''()$ —is the derivative of $f'()$; the *third derivative* of $f()$ —written $f'''()$ —is the derivative of $f''()$; and so on.

Usually, problems in elementary calculus do not involve derivatives of order higher than the fourth, so that we will not really need, at this time, any notation for indicating derivatives other than the prime, double prime, triple prime, and quadruple prime. For other purposes, however, it is convenient to consider other ways of indicating derivatives.

A particularly important application of differentiation occurs in that part of the study of physics which involves motion along a straight line. We have already noted that a function whose independent variable has values which are interpreted as time elapsed and whose dependent variable has values which are interpreted as distance travelled will then have a derivative, the values of the dependent variable of which will be interpreted as the (instantaneous) velocity of the moving object. Specifically, if s is a dependent variable of the function $f()$, and t is an independent variable of $f()$, then the equation

$$s = f(t)$$

is interpreted physically by saying that the moving body will be s units from its starting point at a time t seconds after the body began moving. In case $f()$ has both positive and negative numbers in its range of values, then the physical interpretation involves direction as well as distance. Because of the physical considerations involved, the physicist then *defines* acceleration to be a dependent variable of the function $f''()$, the second derivative of $f()$. In this sense, one says that velocity is the derivative of position, and that acceleration is the derivative of velocity.

It often happens, in physics, that one may find the function which gives acceleration, and from this information want to find the functions which give velocity and position. Mathematically, this is equivalent to starting with a function known to be the second derivative of some other function, and to ask for that other function. There are two aspects to this mathematical problem: what to call that which we seek and how to write it.

Since that which we are trying to do is reverse the process of differentiation, we have

Definition 5.10: Given a function $f()$, any function $F()$, for which $F'() = f()$, is called an *anti-derivative* of $f()$.

(It should be noted that the indefinite article "an" is used here, so that there is no such thing as "the anti-derivative.")

One of the ways in which derivatives may be indicated is by use of the symbol " D ," not as a number, but as an indication of differentiation. In this manner of writing $f'()$ is written as $\{Df\}()$; $f''()$ is written as $\{D^2f\}()$; $f'''()$ is written as $\{D^3f\}()$; and so on. Of course, the positive integers which appear to be exponents on the symbol " D " are not exponents in the number sense, since the symbol " D " is not a number, nor a symbol for a number. Because of the similarity of appearance to exponents, however, this manner of indicating derivatives is particularly convenient to denote anti-derivatives. We write an anti-derivative of the function $f()$ as $\{D^{-1}f\}()$ and read this symbol as " D inverse of f ."

Now that we have discussed what we are to call that which we seek and how to write that which we seek, it might be useful to find that which we seek. Unfortunately for the general case, this is more easily said than done. For polynomial functions, however, we may use the following theorems:

Theorem 5.17:

If $c()$ is a constant function whose domain is the set of all numbers, then there is a constant k , such that an anti-derivative of $c()$ is the function

$$\{cI + k\}().$$

Proof:

$$\{cI + k\}'() = c() \quad \text{by the theorems on differentiation.}$$

This theorem may be re-stated as

$$\{D^{-1}c\}() = \{cI + k\}().$$

Theorem 5.18:

If a is a constant, and n is a positive integer, there is a constant k , such that an anti-derivative of the function $\{a I^n\}()$ is the function

$$\left\{ \frac{a}{n+1} I^{n+1} + k \right\} ().$$

Proof:

$$\left\{ \frac{a}{n+1} I^{n+1} + k \right\}' () = \{a I^n\} ().$$

Theorem 5.19:

If $F()$ is an anti-derivative of $f()$ and $G()$ is an anti-derivative of $g()$, then $\{F + G\}()$ is an anti-derivative of $\{f + g\}()$.

Proof:

Theorem 5.12.

As an example of the process of anti-differentiation, let us consider the physical problem of a falling body, which starts to fall with velocity zero (that is, no downward push) from a platform 100 feet high and whose acceleration is always 32 feet per second. In mathematical terms, we are dealing with a function $f()$, specified by some equation

$$s = f(t),$$

whose derivative $f'()$, is given by the equation

$$v = f'(t),$$

and whose second derivative is given by the equation

$$a = f''(t).$$

The information we start with is, in terms of these variables, when $t = 0$, $v = 0$; when $t = 0$, $s = 100$; and $a = -32$ for all values of t . In terms of the functions involved, we have

$$f'(0) = 0; f(0) = 100; a() = -32().$$

Since $f''()$ is the derivative of $f'()$, $f'()$ is an anti-derivative of the constant function $-32()$. By theorem 5.17, $f'()$ is specified by some equation

$$f'(t) = -32t + k,$$

where k is some constant. But the information $f'(0) = 0$ tells us that

$$0 = (-32)(0) + k$$

must be a true statement, so that k must be the number 0. Thus, $f'()$ is specified by the equation

$$f'(t) = -32t,$$

or, in terms of the variables, by

$$v = -32t.$$

Since $f'()$ is the derivative of $f()$, $f()$ is an anti-derivative of $f'()$, and Theorem 5.18 tells us that there is some constant b , such that $f()$ is specified by the equation

$$f(t) = -16t^2 + b.$$

But the information $f(0) = 100$ tells us that the equation

$$100 = (-16)(0) + b$$

must be a true statement, so that b must be the number 100. Then $f()$ is specified by the equation

$$f(t) = -16t^2 + 100,$$

and consideration of the physical situation tells us that we are dealing with a function whose domain is not the set of all numbers, but is the interval:

$$\text{the set of all } t, \quad 0 \leq t \leq 25/8.$$

Exercises

Find two distinct anti-derivatives of the function specified by the following equations:

1. $f(r) = 2r + 3.$
2. $f(s) = 3s^2 - 4s + 5.$
3. $f(t) = t^3 - t + 1.$
4. $f(u) = u^4 + 3u - 2.$
5. $f(v) = 7v^4 - 3v^3 + 4v^2 + v - 1.$
6. $f(w) = 3.$
7. $f(x) = 0.$
8. $f(y) = 2y, \quad y > 0.$
9. $f(z) = z^3, \quad z \neq 1.$

Find the derivative, second derivative, and third derivative of each of the functions in the preceding 9 exercises.

10. Prove the theorem: If n is a positive integer and $P()$ is a polynomial function of degree n , then $\{D^{n+1}P\}() = \theta()$.
11. If an object falls from a height of 200 feet with an initial downward velocity of 10 feet per second, find the function which describes the position of this object at any time.
12. If an object is thrown *upward* from a height of 300 feet with an initial velocity of 200 feet per second, when will that object hit the ground?



APPLICATIONS OF THE DERIVATIVE

6.1. CONTINUITY AND DIFFERENTIABILITY

In the investigation of limits, it was stated that the numbers, $\lim_{t \rightarrow a} f(t)$ and $f(a)$, need not be the same, but it was proven that for a polynomial function $P()$, $\lim_{t \rightarrow a} P(t)$ and $P(a)$ were the same number. In general, we have

Definition 6.1: The function $f()$ is *continuous* at the number a provided that

- 1) a is in the domain of $f()$,
- 2) $\lim_{t \rightarrow a} f(t)$ exists,
- 3) $\lim_{t \rightarrow a} f(t) = f(a)$.

In this section, we will prove a theorem relating the continuity of a function at a particular number in the domain with the differentiability of that function at that same number. This theorem is important because of its being used in the proofs of many other theorems. Throughout this section, it should be realized that we are considering continuity and differentiability as properties which a function may or may not have at one particular number.

The definition of the derivative of a function shows that the existence of the number

$$\lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}$$

means that the function $f()$ is differentiable at the number a .

Further, $\lim_{t \rightarrow a} f(t)$ and $\lim_{h \rightarrow 0} f(a+h)$ will be the same number, whenever both of these limits actually are numbers. Thus, any attempt to prove that

$$\lim_{t \rightarrow a} f(t) = f(a)$$

or that

$$\lim_{h \rightarrow 0} f(a+h) = f(a)$$

is actually an attempt to prove that

$$\lim_{h \rightarrow 0} [f(a+h) - f(a)] = 0.$$

Consequently, an alternate way of stating that the function $f()$ is continuous at the number a is to state that

$$\lim_{h \rightarrow 0} [f(a+h) - f(a)] = 0.$$

Our theorem is:

Theorem 6.1:

If $f()$ is a function, and a is a number in the domain of $f'()$, then $f()$ is continuous at a .

(In shorter form, this theorem says that differentiability implies continuity.)

Proof of 6.1:

1. a is in the domain of $f'()$ by hypothesis.
2. $\lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}$ exists and is the number $f'(a)$ by step 1 and the definition of derivative.
3. For any value of the variable h , other than 0, for which $a+h$ is in the domain of $f()$,

$$\frac{f(a+h) - f(a)}{h} h = f(a+h) - f(a)$$
by algebraic laws
4. $\lim_{h \rightarrow 0} h = 0$ by the theorem on the limit of a polynomial function.
5. $\lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h} h$ exists and is the product

$$\left[\lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h} \right] \left[\lim_{h \rightarrow 0} h \right]$$
by steps 2 and 4 and the theorem on the limit of the product of two functions.
6. $\lim_{h \rightarrow 0} [f(a+h) - f(a)]$ exists and is the product

$$[f'(a)] [0]$$
by steps 2, 4, and 5.
7. $\lim_{h \rightarrow 0} [f(a+h) - f(a)]$ exists and is the number 0 by step 6 and the fact that the product of 0 and any number is 0.
8. $f()$ is continuous at the number a by step 6 and the alternate definition of continuity.

This theorem is one whose converse is a false statement. That is, while the existence of the number $f'(a)$ guarantees the continuity of the function $f()$ at the number a , it is possible to have a function $g()$ continuous at the number b , but have the domain of the derivative $g'()$ exclude the number b . For example, consider the function specified by the equation

$$g(x) = |x - 1|.$$

The number 1 is in the domain of $g()$, and, since

$$\lim_{t \rightarrow 1} g(t) = \lim_{t \rightarrow 1} |t - 1| = 0,$$

$g()$ is continuous at the number 1. In order to discover whether or not 1 is in the domain of $g'()$, we must investigate the existence of

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

If $h \neq 0$, then

$$\frac{g(1+h) - g(1)}{h} = \frac{|1+h-1| - |1-1|}{h} = \frac{|h|}{h}.$$

The definition of limit requires that we “squeeze down” our view of the domain of the function specified by $\frac{|h|}{h}$, always surrounding the number 0.

We can see that for any negative value of the variable h , $\frac{|h|}{h}$ will be the number -1 , and for any positive value of the variable h , $\frac{|h|}{h}$ will be the number 1. Thus, no matter how small an interval surrounding the number 0 we may consider, we have for this small interval some values of the function whose limit we seek being the number -1 and other values being the number 1. Then the values of this function do not “squeeze down” on one fixed number, and $\lim_{h \rightarrow 0} \frac{|h|}{h}$ is not a number. Consequently, 1 is not on the domain of $g'()$.

A look at the graph of $g()$ (Fig. 6.1) shows that the point (1,0) is the location of what might well be described as a “sharp corner” of the graph. While

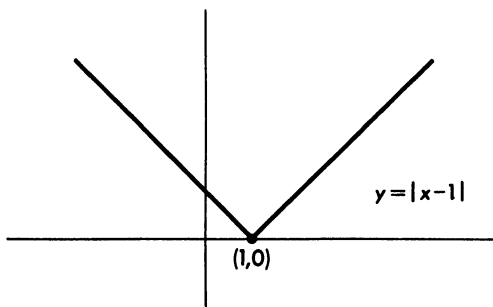


Fig. 6.1

there is certainly no “missing point” or “hole” in the graph at this point, there is equally certainly no one line in the plane, through $(1, 0)$, which can be reasonably called the tangent line to the graph of $g()$, at the point $(1, 0)$.

Exercises

1. $1/2$ is a positive number. Show that, for every positive number d , there is a value of the variable h , such that

a) $|h| < d$.

b) h is in the domain of the function given by $\frac{|h|}{h}$.

c) $\left| \frac{|h|}{h} - 1 \right| > 1/2$.

2. Explain why exercise 1 proves that 1 is not $\lim_{h \rightarrow 0} \frac{|h|}{h}$.

3. Show that, if a is any number other than 1, then for every positive number d , there is a value of the variable h such that

a) $|h| < d$.

b) h is in the domain of the function given by $\frac{|h|}{h}$.

c) $\left| \frac{|h|}{h} - a \right| > \frac{|a - 1|}{2}$.

4. Explain why exercise 2 proves that no number other than 1 could be

$$\lim_{h \rightarrow 0} \frac{|h|}{h}.$$

Which of the functions specified by the following equations is continuous at the number 2? Why or why not?

5. $f(r) = r^2 - 3r + 1$.

6. $f(s) = s - 2$.

7. $f(t) = t^3 - 3t + 5, \quad t \geq 0$.

8. $f(u) = u^2 + u - 7, \quad u > 5$.

9. $f(v) = 2 - 4/v$.

10. $f(w) = \frac{w + 1}{w - 2}$.

11. $f(x) = \frac{x^2 - 4}{x - 2}$.

12. $f(y) = y + 2, \quad y \neq 2$.

13. $f(z) = z + 2$.

6.2. MAXIMUM AND MINIMUM POINTS ON GRAPHS

In this section, we will examine the graph of the function specified by the modified algebraic expression

$$t^3 - 3t, \quad -3 \leq t \leq 3.$$

The graph of this function consists of those points on the graph of the equation $y = x^3 - 3x$ whose x -coordinates are numbers between -3 and 3 , inclusive. The graph appears in Fig. 6.2.

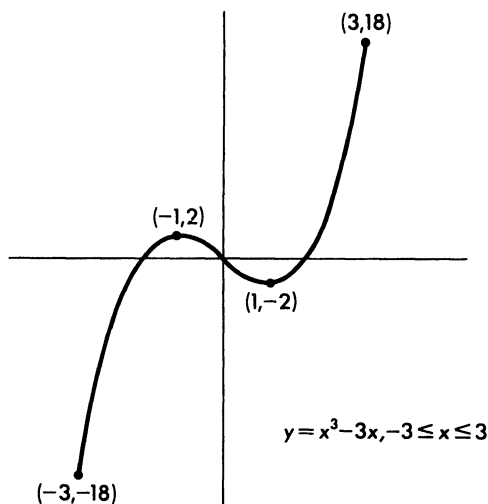


Fig. 6.2

From what has been developed on the differentiation of polynomial functions, the derivative of this function is the new function specified by the modified algebraic expression

$$3s^2 - 3, \quad -3 < s < 3.$$

The numbers -3 and 3 are in the domain of the original function, but are not in the domain of the derivative, since, for this original function $f()$, which is part of the function $\{I^3 - 3I\}()$, the domain is the interval:

the set of all r , $-3 \leq r \leq 3$,

and the expressions

$$\lim_{h \rightarrow 0} \frac{f(3+h) - f(3)}{h} \quad \text{and} \quad \lim_{h \rightarrow 0} \frac{f(-3+h) - f(-3)}{h}$$

are not numbers. Consequently, -3 and 3 are not in the domain of $f'()$, the derivative of $f()$. This instance of the failure of a limit to exist occurs because the definition of limit requires that there is an interval—of values of the variable h —which interval contains the number 0 , such that, among other things, $3+h$ and $-3+h$ are in the domain of $f()$. But,

for any negative value of h , $-3 + h$ will not be in the domain of $f()$, and for any positive value of h , the number $3 + h$ will not be in the domain of $f()$. Thus, neither of the limits indicated above can exist, and the domain of $f'()$ consists of the interval:

$$\text{the set of all } x, \quad -3 < x < 3.$$

As we have seen before, a tangent line to the graph of $f()$ at the point $(-1, 2)$, and a tangent line to the graph of $f()$, at the point $(1, -2)$ will both be horizontal lines, since the numbers -1 and 1 are numbers in the domain of $f'()$ and $f'(-1) = f'(1) = 0$. What we will examine are the properties which the points $(-1, 2)$, $(1, -2)$, and the points $(-3, -18)$ and $(3, 18)$ seem to have, in relation to the graph of $f()$, and we will hope to do our examining, as much as possible, by using that which we know or can discover about the behavior of the derivative $f'()$.

The point $(-3, -18)$ certainly appears to be the lowest point on the graph of $f()$. Moreover, the algebraic laws governing inequalities allow us to prove that, in fact, $(-3, -18)$ is the lowest point of this graph. The proof consists of two observations: first, $(-3, -18)$ is actually on the graph of $f()$, since $f(-3) = -18$; second, since $x^3 - 3x = x(x^2 - 3)$, for any value of the variable x , if we restrict the range of this variable to the interval:

$$\text{the set of all } x, \quad -3 \leq x \leq 3,$$

then the factor $x^2 - 3$ is governed by the inequality

$$-3 \leq x^2 - 3 \leq 6.$$

The product $x(x^2 - 3)$ then certainly cannot be more negative than -18 . Thus, -18 is indeed the smallest number in the range of values of $f()$.

Similarly, 18 is a number in the range of values of $f()$, since $f(3) = 18$, and the product $x(x^2 - 3)$ certainly cannot exceed 18 , so that 18 is the largest number in the range of values of $f()$.

The graph of $f()$ however, indicated that, somehow, the points $(-1, 2)$ and $(1, -2)$, while not the lowest or highest points on the graph, still are the lowest and highest points for at least part of that graph.

If we now look at the derivative $f'()$, we see that the highest and lowest points— $(3, 18)$ and $(-3, -18)$ —are points whose x -coordinates are not in the domain of $f'()$. At the other two points of interest $(-1, 2)$ and $(1, 2)$, the x -coordinates are numbers which give the derivative the value 0 . We may then expect that consideration of maximum and minimum values of a function will have to involve both values of $f()$ which are actually the largest and smallest numbers in the range of values, and values of $f()$ which are somehow largest and smallest for some restricted part of the domain. We may also expect that consideration of the role played by the derivative of the function will involve consideration of those numbers in the domain of the function which are not in the domain of the derivative and those numbers in the domain of the derivative for which the derivative has value 0 . In geometric terms, we will be concerned with points which are end points

of a graph or we will be concerned with points at which the tangent line is horizontal.

With these ideas in mind, we make the following definitions:

Definition 6.2: A number M is the *absolute maximum value* of the function $f()$ provided that

- 1) there is a number a in the domain of $f()$, such that $f(a) = M$;
- 2) for every number x , in the domain of $f()$, $f(x) \leq M$.

Definition 6.3: A number m is the *absolute minimum value* of the function $f()$ provided that

- 1) there is a number b in the domain of $f()$, such that $f(b) = m$;
- 2) for every number x in the domain of $f()$, $f(x) \geq m$.

We also say, in these definitions, that the number a gives the absolute value of M to the function $f()$ and that the number b gives the absolute minimum value m to the function $f()$. For the specific function examined above, 18 is the absolute maximum value and -18 is the absolute minimum value. The number -3 gives this function its absolute minimum value and the number 3 gives this function its absolute maximum value.

Definition 6.4: The number K is a *relative maximum value* of the function $f()$ provided that

- 1) there is a number c in the domain of $f()$ for which $f(c) = K$;
- 2) there is an interval, containing this number c in its interior, such that, for every number x in this interval, which is in the domain of $f()$, we have $f(x) \leq K$.

Definition 6.5: The number k is a *relative minimum value* of the function $f()$ provided that

- 1) there is a number p in the domain of $f()$ for which $f(p) = k$;
- 2) there is an interval, containing this number p in its interior, such that, for every number x in this interval, which is in the domain of $f()$, we have $f(x) \geq k$.

Comparison of the definition (6.2) with the definition (6.4), and of the definition (6.3) with the definition (6.5) shows that every absolute maximum value is also a relative maximum value and every absolute minimum value is also a relative minimum value.

In the use of these two definitions, we say that the number c gives $f()$ a relative maximum value of K and that the number p gives $f()$ a relative minimum value of k . In the specific function examined above, -1 gives $f()$ a relative maximum value of 2, and 1 gives $f()$ a relative minimum value of -2 .

Because of the use of “ $\leq$ ” and “ $\geq$ ” in the four definitions, we are including cases which are not ordinarily thought of as usual in using the words “maximum” and “minimum.” For example, according to these four definitions, a constant function has an absolute maximum, absolute minimum, relative maximum, and relative minimum value at every number in its domain. The reason for stretching the definitions to include such a usage is basically technical. This usage will make the statements of certain of our following theorems simpler than would otherwise be the case.

Exercises

1. Prove that, if $-3 \leq x \leq 0$, then $0 \leq x^3 - 3x \leq 2$.
2. Prove that, if $0 \leq x \leq 3$, then $-2 \leq x^3 - 3x \leq 0$.
3. Prove that, if $-3 \leq x \leq -\sqrt{3}$, then $-18 \leq x^3 - 3x \leq 0$.
4. Prove that, if $\sqrt{3} \leq x \leq 3$, then $0 \leq x^3 - 3x \leq 18$.
5. Why is every absolute maximum value of a function also a relative maximum value?
6. Let the function $g()$ consist of the three ordered pairs:

$$(-1, 5), (0, -4), \text{ and } (1, 11).$$

Sketch the graph of $g()$. Find the absolute maximum value, the absolute minimum value, and all relative maximum and relative minimum values of $g()$.

7. Let the function $F()$ be specified by

$$F(x) = 2x, \quad -1 < x < 2.$$

Show that $F()$ has no absolute or relative maximum or minimum values. (Hint: there is no number next to 0.)

Find the absolute maximum and minimum values (if any) for the function specified by the following:

8. $f(r) = r^2, \quad -1 \leq r \leq 2.$
9. $f(s) = s^2, \quad -1 \leq s < 2.$
10. $f(t) = t^2, \quad -1 < t < 2.$
11. $f(u) = u^2, \quad -1 < u \leq 2.$
12. $f(v) = (v - 1)^2 + 1, \quad 0 \leq v \leq 2.$
13. $f(w) = (w - 1)^2 + 1, \quad 0 \leq w \leq 1.$
14. $f(x) = (x - 1)^2 + 1, \quad 0 \leq x \leq 3.$
15. $f(y) = (y - 1)^2 + 1, \quad 1 \leq y \leq 2.$
16. $f(z) = (z - 1)^2 + 1, \quad 1 \leq z \leq 2.$
17. For each extreme point of the graphs of the functions in exercises 6-16, find the value of $f'()$.

6.3. THEOREMS ON RELATIVE MAXIMA AND MINIMA

If we are able to draw the graph of a function with any reasonable degree of accuracy, then we will be able to determine—with the same degree of accuracy—the location of those points which give relative or absolute maximum or minimum values of that function. But, if we had a method of determining exactly what these relative extreme values are, without recourse to the graph, then we would be spared the inaccuracy inherent in using an approximation to a graph, and we would be able to improve the accuracy of the graph itself.

We will content ourselves with examining relative extremes, since it is only for these that we have available some general theorems. Since every absolute extreme is also a relative extreme, we are not restricting ourselves too much. We expect to find some connection between the location of a relative extreme point and the location of a point of horizontal tangency. Of course, this connection can exist only when there is actually a tangent line, so that we have

Theorem 6.2:

If $f()$ is a function and if a is a number in the domain of $f'()$ for which $f()$ has a relative maximum or relative minimum value, then $f'(a) = 0$.

Proof:

CASE 1:

$f(a)$ is a relative maximum value for $f()$:

1. a is in the domain of $f'()$ by hypothesis.
2. $\lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}$ exists and
is the number $f'(a)$ by step 1 and the definition of derivative.
3. $f(a)$ is a relative maximum value for $f()$ by hypothesis.
4. There is an interval:
the set of all x ,
 $a - d_1 < x < a + d_1$
(where $d_1 > 0$), such that,
if $a - d_1 < x < a + d_1$,
then $f(x) \leq f(a)$ by step 3 and the definition of relative maximum.
5. There is an interval:
the set of all h , $-d_1 < h < d_1$,
such that, if $d_1 < h < d_1$, then
 $f(a+h) \leq f(a)$ by step 4, and relating the variables x and h by $x = a + h$.

6. There is an interval:
the set of all h , $-d_1 < h < d_1$,
such that, if $-d_1 < h < d_1$, then
 $f(a+h) - f(a) \leq 0$ by step 5 and O-3.
7. Assume $f'(a) > 0$.
8. There is an interval:
 $-d_2 < h < d_2$ (where $d_2 > 0$),
such that, if $0 < |h| < d_2$, then
 $\frac{f(a+h) - f(a)}{h}$ is positive by step 7 and Theorem 5.13.
9. Let d_3 be the smaller of the two
positive numbers d_1 and d_2 .
10. If $0 < h < d_3$, then
 $\frac{f(a+h) - f(a)}{h}$ is positive by step 8 and 9.
11. If $-d_3 < h < 0$, then
 $\frac{f(a+h) - f(a)}{h}$ is positive by steps 8 and 9.
12. If $0 < h < d_3$, then
 $f(a+h) - f(a)$ is positive by step 10 and the fact that a
quotient is positive if and only if
both numerator and denominator
have the same sign.
13. If $-d_3 < h < 0$, then
 $f(a+h) - f(a) \leq 0$ by step 6 and 9.
14. Steps 12 and 13 contradict each other, thus the assumption in step 7
is false.
15. Assume $f'(a) < 0$.
16. There is an interval:
the set of all h , $-d_4 < h < d_4$
(where $d_4 > 0$), such that, if
 $0 < |h| < d_4$, then
 $\frac{f(a+h) - f(a)}{h}$ is negative by step 15 and Theorem 5.14.
17. Let d_5 be the smaller of the two
positive numbers d_1 and d_4 .
18. If $0 < h < d_5$, then
 $\frac{f(a+h) - f(a)}{h}$ is negative by steps 16 and 17.
19. If $0 < h < d_5$, then
 $f(a+h) - f(a) \leq 0$ by steps 6 and 17.

20. If $-d_5 < h < 0$, then

$f(a+h) - f(a)$ is positive

by step 18 and the fact that a quotient is positive if and only if both numerator and denominator have the same sign.

21. Steps 19 and 20 contradict each other, thus the assumption in step 15 is false.

22. $f'(a) = 0$ by steps 14 and 20 and O-1.

CASE 2: see the exercises.

It should be remarked that this theorem says only what it says and nothing more. That is, if we start with the two pieces of information: a is in the domain of the derivative $f'()$ and a gives a relative extreme value for $f()$, then we may conclude that $f'(a) = 0$. But the converse may not be true, since we may have a number in the domain of a function, but not in the domain of the derivative of that function, which number gives a relative extreme value to the function, but for which the derivative does not exist at all. Further, we may have a function $f()$ and a number p in the domain of $f'()$, for which $f'(p) = 0$, and yet not have $f(p)$ be a relative extreme value of $f()$. For example 0 is a number in the domain of the function $\{I^3\}()$, for which $\{I^3\}'() = \{3I^2\}()$ has the value 0, but this number does not give a relative extreme value for $\{I^3\}()$. Geometrically, the graph of the equation $y = x^3$ does indeed have a horizontal tangent at the point $(0, 0)$, but this point is not a relative extreme point.

Exercises

Find all the points which might be relative extreme points of the graphs of the function specified by the following expressions:

1. r^2 .
2. $s^2 - 4$.
3. $4 - t^2$.
4. $2u^2 - 3u + 4$.
5. $-2v^2 + 3v - 4$.
6. $w^2 + w + 1$.
7. $\frac{1}{3}x^2 + \frac{1}{4}x - \frac{1}{5}$.
8. $\sqrt{2}y^2 - \sqrt{3}y + \sqrt{7}$.
9. $2z^3 - 9z^2 + 12z - 5$.
10. $r^3 - 12r + 3$.
11. $2s^3 + 3s^2 - 72s + 17$.
12. $t^3 - 4t^2 - 3t + 2$.
13. $4u^3 + u^2 - 4u + 375$.
14. $v^3 - 3v^2 + 3v - 1$.
15. Complete the proof of Theorem 6.1 for the case in which $f(a)$ is a relative minimum value for $f()$.

6.4. FURTHER THEOREMS ON EXTREMES

If we are presented with a function, and can find the derivative of that function, then the discovery of those numbers which give the derivative the value 0 is an algebraic process. In particular, given a polynomial function, the derivative will be another polynomial function, and, if all goes well, we can find the zeros of this derivative. Since we know that a number for which the derivative of a function has value 0, might give a relative extreme value for the function—but also might not—we need some way of testing a suspected relative extreme to determine the true character of such a zero of the derivative.

Let us consider a simple example, the function $\{4 - I^2\}()$. Geometrically, we are looking at the graph of the equation $y = 4 - x^2$. Purely algebraic considerations will tell us that any point on this graph whose x -coordinate is in either of the intervals:

$$\begin{aligned} &\text{the set of all } x, \quad x > 2 \\ \text{or} \quad &\text{the set of all } x, \quad x < -2, \end{aligned}$$

will have a negative y -coordinate. Further, since $x^2 \geq 0$ no matter what number x is, $-x^2 \leq 0$ and $4 - x^2 \leq 4$, no matter what number x is. The point $(0, 4)$ is on the graph of this function, so that 0 is in the domain of $\{4 - I^2\}()$ (which is hardly surprising, since the domain of a polynomial function is the set of all numbers). Since, for every value of the variable x ,

$$\{4 - I^2\}(x) \leq \{4 - I^2\}(0),$$

then 0 gives this function the absolute maximum value of 4. Without recourse to differentiation, we then know that 4 is the absolute maximum value of $\{4 - I^2\}()$, that this absolute maximum value is attained for the number 0 in the domain, and sketching the graph of this function makes it appear a most reasonable guess that there are no other relative extremes.

Now that we have determined the only relative extreme for the function $\{4 - I^2\}()$ by purely algebraic means, let us look at the derivative of the function, and see what is to be seen.

$$\begin{aligned} \{4 - I^2\}'() &= \{-2I\}(). \\ \{-2I\}(0) &= 0. \end{aligned}$$

The number 0 is the only zero of the derivative of $\{4 - I^2\}()$. Since the domain of this derivative is the set of all numbers, if there were any number other than 0 which gave a relative extreme value to the function, the value of the derivative would have to be 0 for this number. Thus, Theorem 6.1 tells us that 0 is the only number giving a relative extreme to the function $\{4 - I^2\}()$, since 0 is the only number making $\{-2I\}()$ take the value 0, and the domain of $\{-2I\}()$ is the set of all numbers.

Since relative extremes involve intervals in the domain of a function,

which intervals contain a number at which the relative extreme occurs, let us examine the interval:

$$\text{the set of all } x, \quad -1 < x < 1,$$

which interval contains the number 0.

If $-1 < x < 0$, then $0 < \{-2I\}(x) < 2$, so that, by Theorem 5.15, $\{4 - I^2\}()$ is increasing at every number between -1 and 0 . If $0 < x < 1$, then $-2 < \{-2I\}(x) < 0$, so that, by Theorem 5.16, $\{4 - I^2\}()$ is decreasing at every number between 0 and 1 . These two facts suggest that the graph of $\{4 - I^2\}()$ increases from $(-1, 3)$ to $(0, 4)$ and decreases from $(0, 4)$ to $(1, 3)$, which would make $(0, 4)$ a relative maximum point. The procedure followed to determine whether a suspected relative extreme is an actual one, by examining the sign of the values of the derivative in an interval surrounding the suspected number in the domain, suggests the following theorems:

Theorem 6.3:

If $f()$ is a function and a is a number in the domain of $f'()$, for which $f'(a) = 0$;

if there is an interval:

$$\text{the set of all } x, \quad a - d < x < a + d$$

(where $d > 0$), such that,

$$\text{if } a - d < x < a, \text{ then } f'(x) \geq 0,$$

$$\text{if } a < x < a + d, \text{ then } f'(x) \leq 0;$$

then $f(a)$ is a relative maximum value for $f()$.

Theorem 6.4:

If $f()$ is a function and a is a number in the domain of $f'()$, for which $f'(a) = 0$;

if there is an interval:

$$\text{the set of all } x, \quad a - d < x < a + d$$

(where $d > 0$), such that,

$$\text{if } a - d < x < a, \text{ then } f'(x) \leq 0,$$

$$\text{if } a < x < a + d, \text{ then } f'(x) \geq 0;$$

then $f(a)$ is a relative minimum value for $f()$.

We will omit the proofs of these two theorems, leaving them as exercises.

Looking again at the function $\{4 - I^2\}()$, we noted that the derivative of this function has positive values to the left of the number 0 and negative values to the right of the number 0. Since this derivative has value 0 at the number 0, it would appear that the derivative itself is decreasing at the number 0, and, by the relation between a decreasing function and a

negative derivative, that the second derivative (as the derivative of the derivative) should have a negative value for the number 0. This suggests:

Theorem 6.5:

If $f()$ is a function and a is a number in the domain of $f''()$, for which $f'(a) = 0$ and $f''(a) < 0$, then $f(a)$ is a relative maximum value for $f()$.

Proof:

1. $f''(a) > 0$ by hypothesis.
2. $f''() = \{f'\}'()$ by the definition of second derivative.
3. $f'()$ is decreasing at a by steps 1 and 2 and Theorem 5.16.
4. There is an interval:
the set of all x ,
$$a - d < x < a + d$$

(where $d > 0$), such that, if
$$a - d < x < a,$$

then $f'(x) \geq f'(a)$, and, if
$$a < x < a + d,$$

then $f'(x) \leq f'(a)$ by step 3 and the definition of a function which decreases at a number.
5. $f'(a) = 0$ by hypothesis.
6. There is an interval:
the set of all x ,
$$a - d < x < a + d$$

(where $d > 0$), such that, if
$$a - d < x < a,$$

then $f'(x) \geq 0$,
and, if $a < x < a + d$,
then $f'(x) \leq 0$ by steps 4 and 5.
7. $f(a)$ is a relative maximum value for $f()$ by step 6 and Theorem 6.3.

In like manner, we have

Theorem 6.6:

If $f()$ is a function, and a is a number in the domain of $f''()$, for which $f'(a) = 0$ and $f''(a) > 0$, then $f(a)$ is a relative minimum value for $f()$.

(The proof of this theorem is left as an exercise.)

We may now use these theorems to discover all relative extreme values of those functions for which we can find all the zeros of the derivative. The procedure consists of the following: Given $f()$,

1. Find $f'()$ and $f''()$.
2. Find the solution of the equation $f'(x) = 0$.
3. For each zero of $f'()$ —say the number a —find $f''(a)$. If $f''(a) > 0$, then $f(a)$ is a relative minimum value. If $f''(a) < 0$, then $f(a)$ is a relative maximum value. If $f''(a) = 0$, go on to the next step.
4. If $f'(a) = 0$ and $f''(a) = 0$, examine the sign of $f'()$ for a small interval containing a , but not containing any other zero of $f'()$. Use Theorems 6.2 and 6.3 to determine whether or not $f(a)$ is a relative extreme value, and, if so, which kind of an extreme.

EXAMPLE 1:

$$\begin{aligned} & \{I^3 - 3I\}(). \\ & \{I^3 - 3I\}'() = \{3I^2 - 3\}(). \\ & \{I^3 - 3I\}''() = \{6I\}(). \end{aligned}$$

The solution of $\{3I^2 - 3\}(x) = 0$ is the solution of $3x^2 - 3 = 0$, which is the solution of $x^2 = 1$, which is the set consisting of the numbers -1 and 1 .

$\{6I\}(-1) = -6$, so that $\{I^3 - 3I\}(-1)$ (which is the number 2) is a relative maximum value for this function.

$\{6I\}(1) = 6$, so that $\{I^3 - 3I\}(1)$ (which is the number -2) is a relative minimum value for this function.

The domain of $\{I^3 - 3I\}'()$ is the set of all numbers, so there are no further relative extremes.

EXAMPLE 2:

$$\begin{aligned} & \{I^4\}(). \\ & \{I^4\}'() = \{4I^3\}(). \\ & \{I^4\}''() = \{12I^2\}(). \end{aligned}$$

The only zero of $\{4I^3\}()$ is the number 0 . $\{I^4\}''(0) = 0$, so that we need to examine the sign of the values of $\{I^4\}'()$ for an interval containing 0 . If $x < 0$, then $4x^3 < 0$, so that $\{I^4\}'()$ is decreasing to the left of 0 . If $x > 0$, then $4x^3 > 0$, so that $\{I^4\}'()$ is increasing to the right of 0 . Thus, by Theorem 6.2, $\{I^4\}(0)$ is a relative minimum value for the function $\{I^4\}()$.

Since the domain of $\{I^4\}'()$ is the set of all numbers, there are no further extremes.

Exercises

Find all the relative extremes of the following functions:

1. $\{I^3\}()$.
2. $\{I^3 - 4\}()$.
3. $\{4 - I^3\}()$.

4. $\{2I^2 - 3I + 4\}()$.
5. $\{I^2 + I + 1\}()$.
6. $\{2I^3 - 9I^2 + 12I - 5\}()$.
7. $\{I^3 - 12I + 3\}()$.
8. $\{2I^3 - 3I^2 - 72I + 17\}()$.
9. $\{I^3 - 4I^2 - 3I + 2\}()$.
10. $\{4I^3 + I^2 - 4I + 375\}()$.
11. $\{I^3 - 3I^2 + 3I - 1\}()$.
12. Complete the proof of Theorem 6.2.
13. Use the information on critical points found in exercises 1-11 to sketch the graph of each of these functions.

6.5. APPLICATIONS OF MAXIMA AND MINIMA

In this section, we will examine some functions which were formed to describe some specific physical, geometric or other consideration, and try to apply the theorems and techniques of the preceding sections.

In economics, if the function $C()$, given by some equation " $C(n) = \text{expression involving } n$," describes the cost, in dollars, of manufacturing n items (where n may be any non-negative number, not necessarily an integer), then the economist defines the derivative $C'()$ to be a function whose values are called the "marginal cost." Similarly, if the function $R()$, given by some equation " $R(n) = \text{expression involving } n$," describes the income, in dollars, from selling n items, the economist defines the derivative $R'()$ to be a function whose values are called the "marginal revenue." Since the profit made will be given by the function $\{R - C\}()$, the statement of the economist that profit is maximized when marginal cost and marginal revenue are equal is merely a restatement of the mathematical principle that the function $\{R - C\}()$ will have a maximum value whenever $R'()$ and $C'()$ have equal value. Of course, there is a hidden assumption that $R''()$ has only negative values and $C''()$ has only positive values, so that $\{R - C\}''()$ will have only negative values.

In physics, if a projectile is shot upward with an initial velocity of 320 feet per second, the height of the projectile will be given by the equation

$$h = 320t - 16t^2,$$

where the range of the variable t is the interval:

$$\text{the set of all } t, \quad 0 \leq t \leq 20.$$

We then have the function $h()$, which is a function given by this equation for which t is an independent variable and h a dependent variable. The

domain of $h()$ consists of all the numbers between 0 and 20 inclusive. The derivative $h'()$ is the function specified by the equation

$$h'(t) = 320 - 32t, \quad 0 < t < 20$$

and the second derivative $h''()$ is specified by

$$h''(t) = -32, \quad 0 < t < 20.$$

Clearly, the number 10 is a value of the variable t for which $h'(t) = 0$. For this number, $h''(10) = -32$, since $h''()$ is a constant function. Thus, by Theorem 6.5, $h(10)$ is a relative maximum value for $h()$. Since 0 and 20 are the only numbers in the domain of $h()$ which are not in the domain of $h'()$, and since $h(0) = h(20) = 0$, neither $h(0)$ nor $h(20)$ is a relative maximum value for $h()$. Further, 10 is the only number for which $h'()$ has the value 0, so that there is no other relative extreme *inside* the domain of $h()$. Thus, $h(10)$ is not only a relative maximum value for $h()$, but is also the absolute maximum value of $h()$. Since $h(10) = 1600$, we may translate the mathematical fact that 1600 is the absolute maximum value of $h()$ into the physical fact that a projectile shot upward with an initial velocity of 320 feet per second will reach a maximum height of 1600 feet.

In making open cardboard boxes, a manufacturing process involves cutting squares from each of the four corners of a rectangular sheet of cardboard, folding up the four resulting projections to form the sides of the box, and then somehow fastening the edges of these projections to each other. In order to determine the size of the cut-out squares in relation to the size of the rectangular sheet one begins with, we will take as our highly desired objective the making of a box holding the largest possible volume. Suppose then that we are presented with a sheet of cardboard which is in the shape of a rectangle 6 feet by 10 feet. If we cut squares from each of the four corners of this rectangular sheet of cardboard, which squares are x feet on a side, then x is a variable whose range is the interval:

$$\text{the set of all } x, \quad 0 < x < 3.$$

(If we attempt to cut *four* squares whose sides are 3 feet or more we annihilate the cardboard.) Folding up the projections to form an open box gives us a box whose base is a rectangle measuring $(6 - 2x)$ feet by $(10 - 2x)$ feet, and whose height is x feet. Then $x(6 - 2x)(10 - 2x)$ cubic feet will be the volume of this box. The largest possible box will then be determined by that value or values of the variable x which will give the absolute maximum value to the function $V()$, given by

$$V(x) = x(6 - 2x)(10 - 2x), \quad 0 < x < 3.$$

We also have, by multiplication,

$$V(x) = 4x^3 - 32x^2 + 60x, \quad 0 < x < 3.$$

The derivative is given by

$$V'(x) = 12x^2 - 64x + 60, \quad 0 < x < 3,$$

and the second derivative by

$$V''(x) = 24x - 64, \quad 0 < x < 3.$$

We find the zeros of the function $V'()$ by solving the equation

$$12x^2 - 64x + 60 = 0.$$

This equation is a polynomial equation of degree two, so that, by Theorem 2.2, there are two solutions, the numbers

$$\frac{8 - \sqrt{19}}{3} \quad \text{and} \quad \frac{8 + \sqrt{19}}{3}.$$

In order to get an idea of the size of these two numbers, we have:

$$16 < \sqrt{19} < 25.$$

$$4 < \sqrt{19} < 5. \qquad -5 < -\sqrt{19} < -4.$$

$$12 < 8 + \sqrt{19} < 13. \qquad 3 < 8 - \sqrt{19} < 4.$$

$$4 < \frac{8 + \sqrt{19}}{3} < \frac{13}{3}. \qquad 1 < \frac{8 - \sqrt{19}}{3} < \frac{4}{3}.$$

We can then see that $\frac{8 - \sqrt{19}}{3}$ is the only one of these two numbers which is in the domain of $V()$. Since this number is less than 2, the value of $V''()$ for this number will be negative, and Theorem 6.5 tells us that this number gives $V()$ a relative maximum value. Since there are no other zeros of $V'()$ in the domain of $V()$, and, since the endpoints give $V(0) = V(3) = 0$, we conclude that $V\left(\frac{8 - \sqrt{19}}{3}\right)$ is the absolute maximum value of $V()$.

Returning to the concrete (or rather, cardboard) problem which started us, we are thus certain that cutting out squares which measure $\frac{8 - \sqrt{19}}{3}$ feet on a side will give us the largest possible box which can be constructed from a rectangular sheet of cardboard 6 feet by 10 feet.

Exercises

1. A manufacturer has a plant which can produce at most 150 typewriters per day. If he produces n typewriters, and prices each at a price of $(1000 - 5n)$ dollars, then he will sell them all. He finds that the cost to him of producing n typewriters is $(300,000 - 200n + n^2)$ dollars. How many typewriters should he produce in order to make the most money?
2. A used helicopter dealer must spend $(200 + 1000n + 50n^2)$ dollars in order to acquire and sell n helicopters. In order actually to sell n helicopters, his price per helicopter must be $\{100(70 - \frac{3}{2}n)\}$ dollars. How many should he buy and sell in order to make the most money?

3. If a grocer offers a large can of applesauce at a price of $(40 - 2n)$ cents per can, then he will sell n cans. In order to buy and distribute n cans, it costs him $(10 + 2n + n^2)$ dollars. How many cans should he plan to sell, and at what price, in order to make the most money?
4. A price of $(60 - n)$ dollars will sell precisely n horses. The cost of raising n horses is $(100 + 30n + n^2)$ dollars. How many horses should be raised and sold to make the most money (Hint: no one will buy half a horse.)
5. A projectile fired directly upward at an initial velocity of 15,000 feet per second will be at a height of $(15000t - 16t^2)$ feet t seconds after firing. How high will it go?
6. A box with open top is made by cutting equal squares from the corners of a 20 inch by 50 inch sheet of steel, and folding up the sides. Find the size of the largest box which can be so made.
7. A piece of wire 5 feet long is bent to form a rectangle. Find the dimensions of the largest rectangle which can be formed in this manner.
8. A fence 10 feet high is 3 feet away from a wall. Find the length of the shortest ladder which will reach from the ground to the top of the fence, rest on the top of the fence, and touch the bottom of the wall.
9. The graphs of $y = 15 - x^2$ and $y = x^2 - 12$ enclose a piece of the plane in which we are to inscribe a rectangle with horizontal and vertical sides. Find the size of the largest such rectangle.
10. Of all the pairs of numbers whose sum is 6, one such pair has the largest product. Find that pair of numbers.

6.6. CONCAVITY

We turn now to a geometric aspect of the graph of a function. It has already been seen that at a relative maximum point on the graph of the function the second derivative often has a negative value, and never has a positive value. That is, if we exclude the awkward cases in which the second derivative has the value 0, then there is a connection between the value of the second derivative and the relative positions of the graph and the tangent line at a relative extreme point. We will make this connection precise in this section.

We start with a purely geometric idea:

Definition 6.6: If a is a number in the domain of the derivative $f'()$ of the function $f()$, then the graph of $f()$ is said to be *concave down at the point* $(a, f(a))$ provided that there is an interval:

$$\text{the set of all } x, \quad a - h < x < a + h$$

(where $h > 0$), such that, if $0 < |a - t| < h$, then the point with x -coordinate t on the line tangent to the graph of $f()$ at $(a, f(a))$ is not below the point $(t, f(t))$.

This definition says that a graph is concave down at a point $(a, f(a))$ provided that the tangent line to the graph at this point lies above, or at least not below, a small piece of the curve through this point. The point-slope form of an equation of a straight line tells us that an equation of this tangent line will be

$$y - f(a) = f'(a)(x - a),$$

or, equivalently,

$$y = f(a) + (x - a)f'(a).$$

The definition says that if t is a number in an appropriately small interval containing the number a , then the curve's being concave down at $(a, f(a))$ is equivalent to the statement that, for $t \neq a$,

$$f(a) + (t - a)f'(a) \geq f(t),$$

or, equivalently, for $t \neq a$,

$$(t - a)f'(a) \geq f(t) - f(a).$$

It might seem, from what we have seen above relative to maximum points, that a graph's being concave down at $(a, f(a))$ might be equivalent to the condition $f''(a) < 0$. There is, however, no implication starting from the fact of concavity, unless we impose further conditions (such as continuity) on the function $f''()$. That is, there exist functions which are concave down at a particular point, but which functions do not have a negative value for their second derivatives at that point.

For example, the function $\{-I^4\}()$ is certainly concave down at $(0, 0)$, since the tangent line to the graph of $y = -x^4$ at the point $(0, 0)$, is the x -axis itself, and the entire graph of $\{-I^4\}()$ lies below the x -axis. But, if $f() = \{-I^4\}()$, then $f''() = \{-12I^2\}()$, and $f''(0) = 0$, which is certainly not a negative number.

However, our purpose is not to discover the implications of geometric facts about graphs of functions in terms of the functions and their derivatives, but to use the geometric implications of analytic facts. We have then

Theorem 6.7:

If $f()$ is a function and a is a number in the domain of $f''()$, if $f''()$ is continuous at a , and $f''(a) < 0$; then the graph of $f()$ is concave down at the point $(a, f(a))$.

Proof:

1. Let the function $g()$ be specified by

$$g(t) = (t - a)f'(a) - f(t) + f(a).$$

2. $g'()$ is specified by

$$g'(t) = f'(a) - f'(t) \quad \text{by step 1 and differentiation.}$$

3. $g''()$ is specified by

$$g''(t) = -f''(t) \quad \text{by step 2 and differentiation.}$$

4. $g''(a) = -f''(a)$ by step 3.
5. $f''(a) < 0$ by hypothesis.
6. $g''(a) > 0$ by steps 4 and 5.
7. $g'(a) = 0$ by step 2.
8. $g(a)$ is a relative minimum value for $g()$ by steps 6 and 7 and Theorem 6.6.
9. There is an interval:
the set of all t ,
$$a - d < t < a + d$$

(where $d > 0$), such that, if
$$a - d < t < a + d$$

then $g(t) \geq g(a)$ by step 8 and the definition of relative minimum value.
10. There is an interval:
the set of all t ,
$$a - d < t < a + d$$

(where $d > 0$), such that, for any t in this interval
$$f(a) + (t - a)f'(a) \geq f(t)$$
 by steps 1 and 9.
11. An equation of the tangent line at $(a, f(a))$ to the graph of $f()$ is
$$y = f(a) + (x - a)f'(a)$$
 by the definition of tangent line and the theorem on the point-slope form of an equation of a line.
12. There is an interval:
the set of all x ,
$$a - d < x < a + d$$

(where $d > 0$), such that, for t in this interval, the point with x -coordinate t on the line tangent to the graph of $f()$ at $(a, f(a))$ is not below the point $(t, f(t))$ by steps 10 and 11.
13. The graph of $f()$ is concave down at $(a, f(a))$ by step 12 and the definition (6.6).

In this proof, the function-of-convenience $g()$ was constructed by noting that the inequality

$$(t - a)f'(a) \geq f(t) - f(a)$$

which was necessary for downward concavity, is equivalent to

$$(t - a)f'(a) - f(t) + f(a) \geq 0,$$

the left side of which inequality was used to construct $g()$.

We have a similar definition, which will be numbered 6.7, for the graph of a function being concave up at a point. There is a theorem similar to 6.7, which says that a sufficient condition for upward concavity of that graph of $f()$ at $(b, f(b))$ is that $f''()$ be continuous at b and that $f''(b) > 0$.

Exercises

1. State the definition (6.7) of upward concavity at a point.
2. Prove Theorem 6.8: If $f()$ is a function, $f''()$ is continuous at b , and $f''(b) > 0$; then the graph of $f()$ is concave up at the point $(b, f(b))$.
3. If the point (p, q) is not below the point (p, r) , which of the following is true: $q = r$, $q < r$, $q > r$, $q \leq r$, $q \geq r$?
4. If the point (s, t) is not above the point (s, u) , what is the ordering relation between t and u ?
5. Find the slope-intercept form of an equation of the tangent line to the graph of the function $G()$, at the point whose x -coordinate is p .
6. Why is the inequality

$$f(a) + (t - a)f'(a) \geq f(t)$$

equivalent to the inequality

$$(t - a)f'(a) \geq f(t) - f(a)?$$

Are these two inequalities equivalent to

$$f'(a) \leq \frac{f(t) - f(a)}{t - a}?$$

7. If $g()$ is specified as in step 1 of the proof of Theorem 6.7, what is $g(a)$?
8. If $f()$ is a function, and we specify a new function by the expression $(t - a)f'(a)$, what expression will give the derivative of this new function, and why?

Determine the points at which the graphs of the following functions are concave up; concave down.

- | | |
|-------------------------------------|--------------------------------------|
| 9. $\{3 - I^2\}()$. | 10. $\{I^2 + 4\}()$. |
| 11. $\{2 + 3I - 5I^2\}()$. | 12. $\{4I^2 - 3I + 10\}()$. |
| 13. $\{2I^3 + 3I^2 - 12I + 5\}()$. | 14. $\{2I^3 - 15I^2 + 36I - 1\}()$. |

6.7. POINTS OF INFLECTION

As we saw in the preceding section, “nice” functions—which in this context mean functions with continuous second derivatives—reveal the concavity of their graphs by the value of the second derivative. It is reasonable, then, to examine the graph of such a function at a point where the

second derivative changes its sign, which point will necessarily be a point at which the graph changes its concavity. Such a point is given a name in

Definition 6.8: A point $(a, f(a))$ on the graph of the function $f()$ is called a *point of inflection* provided that there exists an interval:

the set of all x , $a - d < x < a + d$ (where $d > 0$), such that,

either, if $a - d < t < a$, then the graph is concave down

at $(t, f(t))$ and, if $a < t < a + d$,

then the graph is concave up at $(t, f(t))$

or, if $a - d < t < a$, then the graph is concave up at

$(t, f(t))$ and, if $a < t < a + d$,

then the graph is concave down at $(t, f(t))$.

If the graph has a tangent line at the point of inflection, then, from the definitions of concavity, this tangent line must cross the graph. The theorem which enables us to find all points of inflection, which theorem we state without proof, is

Theorem 6.9:

If $f()$ is a function and a is a number, if $f''()$ is continuous at a , $f''(a) = 0$, $f'''(a) \neq 0$; then $(a, f(a))$ is a point of inflection of the graph of $f()$.

The requirement $f'''(a) \neq 0$ insures that the function $f''()$ actually changes sign as we pass through the number a , so that $f''()$ will have values of opposite sign for numbers on opposite sides of a .

(Remark: Because of the definition of concavity, a straight line is both concave up and concave down at every point.)

Exercises

Find the points of inflection, relative maximum points, and relative minimum points on the graphs of the following equations. Sketch the graphs.

1. $y = x^3$.

2. $y = x^3$.

3. $y = x^3 - 3x$.

4. $y = x^3 - 3x^2$.

5. $y = x^3 - 3x + 3x - 1$.

6. $y = x^3 - 3x^2 + 4x - 5$.

7. $y = x^3 + 2x^2 + 3x + 4$.

8. $y = x^3 + x^2 + x + 2$.

9. $y = x^3 - 2x^2 + 2$.

10. $y = x^3 - 4x^2 + 3x + 4$.

11. $y = x^3 - \frac{5}{4}x^4 + 3$.

6.8. NEWTON'S METHOD

One useful application of the theorems relating geometric properties of the graph of a function with algebraic properties of the function and its derivatives is known as Newton's method for solving equations.

Suppose we have a function $f()$ and an interval:

the set of all t , $a \leq t \leq b$,

in the domain of $f()$. Further, this function has the property that the given interval is also in the domain of both $f'()$ and $f''()$ —the first and second derivatives of $f()$. If, in addition,

$$f(a) < 0$$

$$f(b) > 0$$

for every t in the given interval,

$$f'(t) > 0$$

$$f''(t) > 0,$$

then the graph must increase from $(a, f(a))$ to $(b, f(b))$, must be a continuous curve, and must be concave up at each point. There is thus one and only one point at which the graph crosses the x -axis.

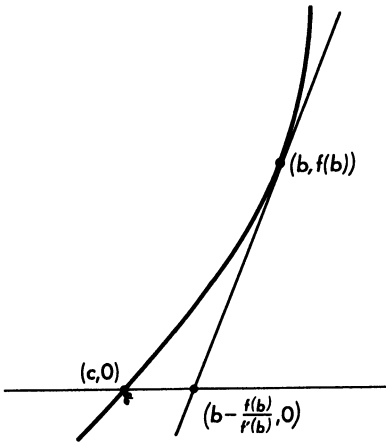


Fig. 6.3

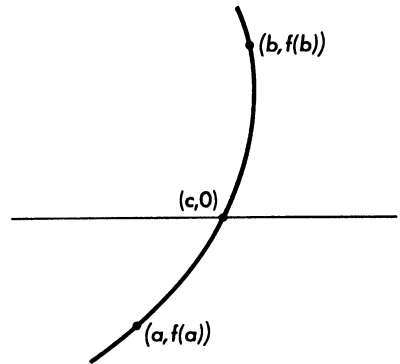


Fig. 6.4

Algebraically, all of this implies that there is, between a and b , exactly one number c , which is a solution of the equation $f(x) = 0$.

The tangent line to the graph of $f()$ at the point $(b, f(b))$ will be below the graph (from the concavity) and will have, as an equation

$$y - f(b) = f'(b) (x - b)$$

or

$$f'(b) x - y = f'(b) b - f(b).$$

This line will cross the x -axis at the point

$$\left(\frac{bf'(b) - f(b)}{f'(b)}, 0 \right),$$

which is

$$\left(b - \frac{f(b)}{f'(b)}, 0 \right).$$

We will use the number $b - \frac{f(b)}{f'(b)}$, or, for convenience in arithmetic, a slightly larger number, as a first approximation to the number c .

We find a second approximation to c by using the first approximation—or a slightly larger number—as the x -coordinate of a new point on the graph, find the tangent line at that new point, and then find the x -coordinate of the point where this new tangent line crosses the x -axis. We do the same to this second approximation to find a third approximation, and continue until we have any desired degree of accuracy for our approximation to the number c .

It is a fact that these successive approximations are increasingly accurate, but the proof is too tedious for our purpose here.

As examples we will find a solution to two equations:

$$x^2 - 2 = 0 \quad \text{and} \quad x^3 - 3x^2 - 9x + 15 = 0.$$

EXAMPLE 1:

$$x^2 - 2 = 0.$$

a) If $f(x)$ is the function given by

$$f(x) = x^2 - 2,$$

then $f'(x)$ is the function given by

$$f'(x) = 2x,$$

and $f''(x)$ is the function given by

$$f''(x) = 2.$$

Since $f(1) = -1$ and $f(2) = 2$, there is some number—in fact $\sqrt{2}$ —between 1 and 2 such that $f(\sqrt{2}) = 0$. The tangent line to the graph of $y = x^2 - 2$ at $(2, 2)$ has equation $4x - y = 6$. This line intersects the x -axis at $(1.5, 0)$, so that our first approximation to $\sqrt{2}$ is 1.5.

b) $f(1.5) = 0.25$ and $f'(1.5) = 3$. The tangent line to the graph of $y = x^2 - 2$ at $(1.5, 0.25)$ has equation

$$3x - y = 4.25,$$

and crosses the x -axis at $(4.25/3, 0)$. For convenience in arithmetic, we take our second approximation to $\sqrt{2}$ as 1.42, a number slightly larger than $4.25/3$.

c) $f(1.42) = 0.0164$ and $f'(1.42) = 2.84$. The tangent line to the graph of $y = x^2 - 2$ at $(1.42, 0.0164)$ has equation $2.84x - y = 4.0264$ and crosses the x -axis at $(4.0264/2.84, 0)$. For convenience in arithmetic, we take our third approximation to $\sqrt{2}$ as 1.418, a number slightly larger than $4.0264/2.84$.

d) $f(1.418) = 0.010724$ and $f'(1.418) = 2.836$. The tangent line to the graph of $y = x^2 - 2$ at $(1.418, 0.010724)$ has equation

$$2.836x - y = 4.010724$$

and crosses the x -axis at $(4.010724/2.836, 0)$. For convenience in arithmetic, we take our fourth approximation to $\sqrt{2}$ as 1.4143, a number slightly larger than $4.010724/2.836$. This approximation is, in fact, accurate to three decimal places.

EXAMPLE 2:

$$x^3 - 3x^2 - 9x + 15 = 0.$$

a) $f(x) = x^3 - 3x^2 - 9x + 15,$

$$f'(x) = 3x^2 - 6x - 9 = 3(x - 3)(x + 1),$$

$$f''(x) = 6x - 6 = 6(x - 1).$$

The graph of $f()$ has a relative maximum point at $(-1, 20)$ (since $f'(-1) = 0$, $f''(-1) = -12 < 0$); it has a relative minimum point at $(3, -12)$ (since $f'(3) = 0$, $f''(3) = 12 > 0$); and an inflection point at $(1, 4)$ (since $f'(1) = -12 \neq 0$, $f''(1) = 0$). These are the only critical points. If $x < 1$, the graph is concave down at $(x, f(x))$; if $x > 1$, the graph is concave up at $(x, f(x))$. The graph appears as

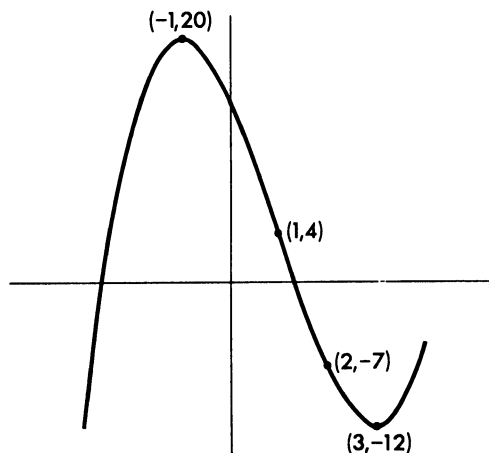


Fig. 6.5

There are three points at which this graph crosses the x -axis. We will try to find the x -coordinate of the middle one. That is, we will attempt to find a number c between 1 and 3, which is a solution of

$$x^3 - 3x^2 - 9x + 15 = 0.$$

Since $f(2) = -7$, we know that $1 < c < 2$. Between $(1, 4)$ and $(2, -7)$ the graph is concave up at each point.

b) $f(1) = 4$ and $f'(1) = -12$. The tangent line at $(1, 4)$ has equation $12x + y = 16$. Our first approximation to c would be $4/3$, and would be an underestimate, since the tangent line is below the graph. For convenience, we will use 1.3 , a number slightly less than $4/3$.

c) $f(1.3) = 0.427$ and $f'(1.3) = -11.73$.

$$11.73x + y = 15.676,$$

$$15.676/11.73 = 1.33.$$

d) $f(1.33) = 0.075937$ and $f'(1.33) = -11.6733$.

$$11.6733x + y = 15.601426,$$

$$15.601426/11.6733 = 1.336.$$

e) $f(1.336) = 0.005933056$ and $f'(1.336) = -11.661312$.

$$1.661312x + y = 15.58544588,$$

$$15.58544588/11.661312 = 1.3365.$$

So that we take 1.3365 as an approximation to the one solution of

$$x^3 - 3x^2 - 9x + 15 = 0,$$

which is between 1 and 3 .

Exercises

1. Draw diagrams to illustrate parts a), b), c) and d) of example 1.

Use Newton's method to

2. Find $\sqrt{3}$ to 3 decimal place accuracy.

3. Find the other two solutions of $x^3 - 3x^2 - 9x + 15 = 0$ to one decimal place accuracy.

4. Solve $x^3 - 3x^2 - 9x - 18 = 0$ to one decimal place accuracy.

5. How can Newton's method be applied to situations in which the desired intercept is also an inflection point? A relative maximum point?



INTEGRATION

7.1. SOME FACTS ABOUT THE NATURAL NUMBERS

A cornerstone of all the mathematics we have examined thus far is the set of natural numbers: 1, 2, 3, 4, and so on. Since these numbers are used for counting, there are situations in which we use the natural numbers in describing sets of natural numbers. For example, the set consisting of all the natural numbers which do not exceed 3,004 contains exactly 3,004 natural numbers; the set consisting of the first 84 natural numbers is the set of all natural numbers not exceeding 84; the set of the first 15 even natural numbers consists of the numbers: 2, 4, 6, 8, and so on, up to and including 30.

In this section, we will examine a set of natural numbers, and from the description of that set, ascertain which natural number results from the addition of all the numbers in the described set. Our first such set will be the set of all the natural numbers which do not exceed a given natural number. Thus, if the given natural number is n , then we are looking at the set consisting of the first n natural numbers, and we are trying to find, in terms of this given natural number n , the natural number which is the sum of the n natural numbers in this set. We are again faced with the problem of deciding how we will write things about our new situation, and we solve this problem by using the device of inserting a string of dots to indicate the words: "and continuing in the indicated manner until we come to."

With this notation, the set of all natural numbers not exceeding the natural number n may be written as:

$$1, 2, 3, 4, \dots, n - 2, n - 1, n.$$

The sum of these first n natural numbers will then be indicated by:

$$1 + 2 + 3 + 4 + \dots + (n - 2) + (n - 1) + n.$$

It is customary in mathematics to adopt a shorthand form of this notation, but we will content ourselves with this "writing out" of a sum.

We have the following theorem:

Theorem 7.1:

If n is a natural number, then

$$1 + 2 + 3 + 4 + \dots + (n - 2) + (n - 1) + n$$

is the natural number $n(n + 1)/2$.

Remark: If n is a natural number, then either n or $n + 1$ is an even natural number, so that $n(n + 1)/2$ will always be a natural number.

Proof of 7.1:

1. Let S be the sum of the first n natural numbers,

$$S = 1 + 2 + 3 + 4 + \dots + (n - 2) + (n - 1) + n.$$
2. S is the sum of precisely n natural numbers by step 1.
3. $S = n + (n - 1) + (n - 2) + (n - 3) + \dots + 3 + 2 + 1$ by step 1 and the laws N-A-2 and N-A-3.
4. $2S = 1 + 2 + 3 + 4 + \dots + (n - 2) + (n - 1) + n + n + (n - 1) + (n - 2) + (n - 3) + \dots + 3 + 2 + 1$ by steps 1 and 3.
5. $2S = [1 + n] + [2 + (n - 1)] + [3 + (n - 2)] + [4 + (n - 3)] + \dots + [(n - 2) + 3] + [(n - 1) + 2] + [n + 1]$ by step 4 and laws N-A-2 and N-A-3.
6. The sum in step 4 contains $2n$ natural numbers by steps 2 and 4.
7. The sum in step 5 contains n terms, which are enclosed in brackets by steps 4, 5, and 6.
8. Each term in brackets in the sum in step 5 is the natural number $n + 1$ by step 5 and various algebraic laws.
9. $2S = n(n + 1)$ by steps 7 and 8.
10. $S = n(n + 1)/2$ by step 9.
11. $1 + 2 + 3 + 4 + \dots + (n - 2) + (n - 1) + n = n(n + 1)/2$ by steps 1 and 10.

Since we have found that the sum of the first n natural numbers is the natural number $n(n + 1)/2$, we will use this fact to determine the sum of the squares of the first n natural numbers:

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

We will need to use another fact about natural numbers, which fact we have already encountered in the differentiation of the function $\{I^k\}()$, where k is a natural number, namely the factoring

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

In particular, for $k = 3$,

$$a^3 - b^3 = (a - b) (a^2 + ab + b^2).$$

Let us consider the set of all natural numbers not exceeding the natural number n . We want to examine the expressions

$$\begin{aligned} n^3 - (n-1)^3, \\ (n-1)^3 - (n-2)^3, \\ (n-2)^3 - (n-3)^3, \\ \cdot \\ \cdot \\ \cdot \\ 4^3 - 3^3, \\ 3^3 - 2^3, \\ 2^3 - 1^3. \end{aligned}$$

We can examine each and every one of these $n - 1$ expressions by realizing that, if k is a natural number, then

$$\begin{aligned} k^3 - (k-1)^3 &= [k - (k-1)] [k^2 + k(k-1) + (k-1)^2] \\ &= (1) (k^2 + k^2 - k + k^2 - 2k + 1) \\ &= 3k^2 - 3k + 1. \end{aligned}$$

Our interest lies in the $n - 1$ true equations formed by letting $k = 2, 3, 4, \dots, n-2, n-1$, and n . We have

$$\begin{aligned} n^3 - (n-1)^3 &= 3n^2 - 3n + 1; \\ (n-1)^3 - (n-2)^3 &= 3(n-1)^2 - 3(n-1) + 1; \\ (n-2)^3 - (n-3)^3 &= 3(n-2)^2 - 3(n-2) + 1; \\ \cdot \\ \cdot \\ \cdot \\ 4^3 - 3^3 &= 3(4)^2 - 3(4) + 1; \\ 3^3 - 2^3 &= 3(3)^2 - 3(3) + 1; \\ 2^3 - 1^3 &= 3(2)^2 - 3(2) + 1. \end{aligned}$$

It is also true that

$$1^3 - 0^3 = 3(1)^2 - 3(1) + 1,$$

so that we can append this equation to the $n - 1$ equations above to give us a set of n true equations.

If we add the numbers which are the left-hand sides of these n equations we have, as the sum,

$$\begin{aligned} n^3 - (n-1)^3 + (n-1)^3 - (n-2)^3 + (n-2)^3 - (n-3)^3 + \dots \\ \dots + 4^3 - 3^3 + 3^3 - 2^3 + 2^3 - 1^3 + 1^3 - 0^3, \end{aligned}$$

which is certainly the number n^3 , since all except the first term add to 0.

If we add the numbers which are the right-hand sides of these n equations we have, as the sum,

$$\begin{aligned} & 3 [n^2 + (n-1)^2 + (n-2)^2 + \dots + 4^2 + 3^2 + 2^2 + 1^2] \\ & + 3 [n + (n-1) + (n-2) + \dots + 4 + 3 + 2 + 1] \\ & + n. \end{aligned}$$

The second of the terms in this sum contains the sum of the first n natural numbers, which, as we saw, is the natural number $n(n+1)/2$. We conclude

$$\begin{aligned} n^3 = & 3 [n^2 + (n-1)^2 + (n-2)^2 + \dots + 4^2 + 3^2 + 2^2 + 1^2] \\ & + 3n(n+1)/2 + n. \end{aligned}$$

This equation says that the sum of the squares of the first n natural numbers,

$$1^2 + 2^2 + 3^2 + 4^2 + \dots + (n-2)^2 + (n-1)^2 + n^2,$$

is the natural number

$$1/3 [n^3 - 3n(n+1)/2 - n],$$

which simplifies to

$$n(n+1)(2n+1)/6.$$

We have thus proved

Theorem 7.2:

The sum of the squares of the first n natural numbers is the natural number

$$n(n+1)(2n+1)/6.$$

Exercises

1. Write a formal proof of Theorem 7.2.
2. If n is a natural number, why can one be certain that $n(n+1)(2n+1)/6$ is a natural number?
3. Prove that the sum of the cubes of the first n natural numbers is the natural number $n^2(n+1)^2/4$.

Derive "formulas" for the following sums:

4. $1^4 + 2^4 + 3^4 + \dots + (n-2)^4 + (n-1)^4 + n^4$.
5. $1 + 3 + 5 + 7 + \dots + (2n-3) + (2n-1) + (2n+1)$.
6. $2 + 4 + 6 + 8 + \dots + (2n-4) + (2n-2) + 2n$.
7. $1^3 + 3^3 + 5^3 + \dots + (2n-3)^3 + (2n-1)^3 + (2n+1)^3$.
8. $2^3 + 4^3 + 6^3 + \dots + (2n-4)^3 + (2n-2)^3 + (2n)^3$.

7.2. AREA

In elementary geometry, one learns to find the area of a triangle by using the "formula"

$$A = \frac{1}{2}bh,$$

where b is the number which gives the length of the base of the triangle, h is the number which gives the length of the altitude of the triangle, and A is the number which gives the area of the triangle. One also learns to find the area of a sector of a circle by the "formula"

$$A = \frac{\theta}{360} \pi r^2,$$

where r is the number which gives the length of the radius of the circle and θ is the number of degrees contained in the angle forming the sector.

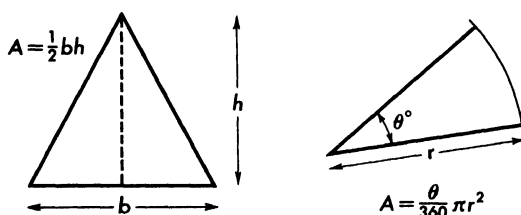


Fig. 7.1

Any part of the plane which is made up entirely of triangles or sectors of circles then has an area, which area can be found by adding the areas of the component parts. On the other hand, a part of the plane which is *not* made up of triangles or sectors of circles, does not, as yet, have any area at all, until we find some way of defining such a thing.

In the process of attempting to discover a definition of area, we will actually invent a new mathematical entity, whose uses are manifold, the definition of area being only one use. It is convenient, however, to approach the problem by discussing the geometric aspect of area. The reader should keep in mind that a discussion of area is merely the device we are using to introduce our subject, and that, when we have finished, we will have a new mathematical entity, which stands alone, with no necessary connection with area, or with any other particular use of this entity. Such a situation has already been seen in the derivative of a function, the *definition* of which involves nothing about rates of change, or tangent lines, or relative extremes.

If we are to attempt to define the area of parts of the plane which are not made up of triangles or sectors of circles, it would be wise to begin the investigation with a part of the plane which does not differ too radically from such simple areas. We turn to our well-known function $\{I^2\}()$ and look at part of the graph of this function. Suppose we let the function $f()$ be specified by the modified equation

$$f(x) = x^2, \quad 0 \leq x \leq 1.$$

The graph of $f()$ consists of that part of the graph of $\{I^2\}()$ which lies between the points $(0, 0)$ and $(1, 1)$, and includes these two points. (Fig. 7.2.)

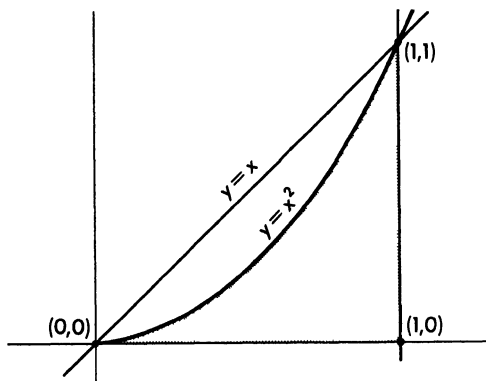


Fig. 7.2

We will investigate that part of the plane which is bounded by (that is, contained within) the graph of $f()$, the portion of the x -axis between $(0, 0)$ and $(1, 0)$, and the portion of the vertical line with equation $x = 1$ between $(1, 0)$ and $(1, 1)$. This part of the plane looks somewhat like a right triangle whose hypotenuse has ceased to be a straight line. Suppose we have, somehow, defined area for such an almost-triangle, and we discover that the area of this almost-triangle is given by the number (of units of area) A . How could we, not knowing actually which number A is, find some reasonable approximation to this number? First of all, whatever the definition of area may be, every reasonable definition must require A to be a positive number. Second, we know that the graph of $f()$ always lies below the graph of the equation $y = x$, so that A should be less than the area of the actual triangle formed by the graph of $y = x$ and the two vertical and horizontal line segments. Thus, whatever number A may be, we know that

$$0 < A < \frac{1}{2}.$$

Any further attempt at a better approximation of the number A must somehow involve the function $f()$, and also involve parts of the plane whose area has already been defined. Of all possible parts of a plane which are made up of triangles or sectors of circles, that one whose area is most easily found is a rectangle. Let us then look at the point $(\frac{1}{2}, \frac{1}{4})$ on the graph of $f()$ —which seems to be a reasonably midway point—and see if we can find any rectangles which might have something to do with the number A , which we have assumed will give the area of our almost-triangle. Since we are dealing with a rectangular coordinate system, life will be somewhat easier if we look for rectangles whose sides are vertical and horizontal lines.

Let us then draw a vertical line through the point $(\frac{1}{2}, \frac{1}{4})$, and start looking for some horizontal lines to form rectangles. We have basically two choices: 1) Consider the horizontal line segment from $(0, 0)$ to $(\frac{1}{2}, 0)$ and the horizontal line segment from $(\frac{1}{2}, \frac{1}{4})$ to $(1, \frac{1}{4})$ (Fig. 7.3)

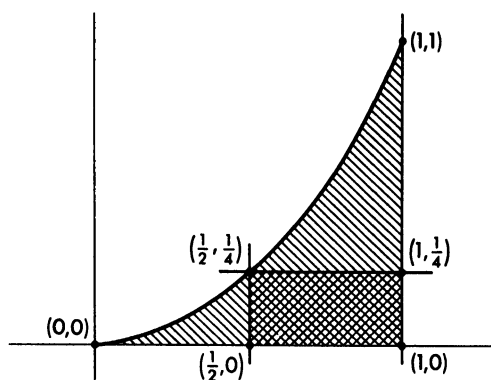


Fig. 7.3

The line segment along the x -axis forms, with the vertical line $x = \frac{1}{2}$ and the y -axis, a “rectangle” of 0 area. The line segment along the line $y = \frac{1}{4}$ forms, with the x -axis and the vertical lines $x = \frac{1}{2}$ and $x = 1$, a rectangle of area $\frac{1}{2} \cdot \frac{1}{4} (= \frac{1}{8})$. In terms of the function $f()$, we are looking at the sum of the areas of two rectangles, which sum is the number

$$f(0) \frac{1}{2} + f\left(\frac{1}{2}\right) \frac{1}{2},$$

and, whatever the definition of area for our almost-triangle may be, every reasonable definition must have the area of the almost-triangle larger than the number $\frac{1}{8}$. It should be noted that $f(0)$ is the absolute minimum value of $f()$ for the interval:

$$\text{the set of all } x, \quad 0 \leq x \leq \frac{1}{2}$$

and $f(\frac{1}{2})$ is the absolute maximum value of $f()$ for the interval:

$$\text{the set of all } x, \quad \frac{1}{2} \leq x \leq 1.$$

2) Consider the horizontal line segment from $(\frac{1}{2}, 1)$ to $(1, 1)$ and the horizontal line segment from $(0, \frac{1}{4})$ to $(\frac{1}{2}, \frac{1}{4})$. (Fig. 7.4).

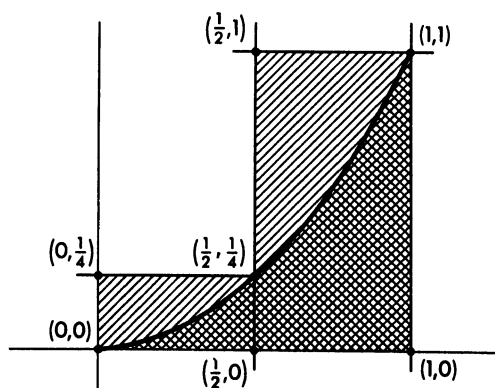


Fig. 7.4

The left-hand line segment gives us a rectangle whose area is $\frac{1}{4} \cdot \frac{1}{2} (= \frac{1}{8})$ and the right-hand line segment gives us a rectangle whose area is $1 \cdot \frac{1}{2} (= \frac{1}{2})$. The sum of these two numbers is $\frac{5}{8}$, and, in terms of the function $f()$,

$$f(\frac{1}{2}) \frac{1}{2} + f(1) \frac{1}{2} = \frac{5}{8}.$$

Again, whatever the definition of area may be, every reasonable definition must have the area of our almost-triangle less than the number $\frac{5}{8}$. It should be noted here that $f(\frac{1}{2})$ is the absolute maximum value of $f()$ for the interval:

$$\text{the set of all } x, \quad 0 \leq x \leq \frac{1}{2},$$

and $f(1)$ is the absolute maximum value of $f()$ for the interval:

$$\text{the set of all } x, \quad \frac{1}{2} \leq x \leq 1.$$

These two lines of investigation together give us

$$\frac{1}{8} < A < \frac{5}{8}.$$

Of course, $A < \frac{5}{8}$ is no improvement over $A < \frac{1}{2}$, but let us see what occurs if we consider two points on the graph of $f()$ rather than the one point $(\frac{1}{2}, \frac{1}{4})$. For convenience, let us take the x -coordinates evenly spaced, so that we will split the domain of $f()$ into three intervals:

$$\text{the set of all } x, \quad 0 \leq x \leq \frac{1}{3};$$

$$\text{the set of all } x, \quad \frac{1}{3} \leq x \leq \frac{2}{3};$$

$$\text{and the set of all } x, \quad \frac{2}{3} \leq x \leq 1.$$

The absolute maximum value of $f()$ on the first interval is the number $f(\frac{1}{3}) = \frac{1}{9}$. The absolute maximum value of $f()$ on the second interval is the number $f(\frac{2}{3}) = \frac{4}{9}$. The absolute maximum value of $f()$ on the third interval is the number $f(1) = 1$. The three rectangles formed by drawing horizontal lines from the points $(\frac{1}{3}, f(\frac{1}{3}))$, $(\frac{2}{3}, f(\frac{2}{3}))$, and $(1, f(1))$, leftward until one reaches the vertical lines $x = 0$, $x = \frac{1}{3}$, and $x = \frac{2}{3}$, respectively, will have total area

$$\begin{aligned} & f(\frac{1}{3}) \frac{1}{3} + f(\frac{2}{3}) \frac{1}{3} + f(1) \frac{1}{3} \\ &= (\frac{1}{9} \cdot \frac{1}{3}) + (\frac{4}{9} \cdot \frac{1}{3}) + (1 \cdot \frac{1}{3}) \\ &= \frac{1}{27}. \end{aligned}$$

As can be seen from Fig. 7.5, every reasonable definition of the area of our almost-triangle will give $A < \frac{1}{27}$.

The absolute minimum value of $f()$ on the first interval is the number $f(0) = 0$. The absolute minimum value of $f()$ on the second interval is the number $f(\frac{1}{3}) = \frac{1}{9}$. The absolute minimum value of $f()$ on the third interval is the number $f(\frac{2}{3}) = \frac{4}{9}$. The three rectangles formed by drawing horizontal line segments from the points $(0, f(0))$, $(\frac{1}{3}, f(\frac{1}{3}))$, and $(\frac{2}{3}, f(\frac{2}{3}))$ to the right until one reaches the vertical lines $x = \frac{1}{3}$, $x = \frac{2}{3}$, and $x = 1$, respectively, will have total area

$$\begin{aligned} & f(0) \frac{1}{3} + f(\frac{1}{3}) \frac{1}{3} + f(\frac{2}{3}) \frac{1}{3} \\ &= (0 \cdot \frac{1}{3}) + (\frac{1}{9} \cdot \frac{1}{3}) + (\frac{4}{9} \cdot \frac{1}{3}) \\ &= \frac{5}{27}. \end{aligned}$$

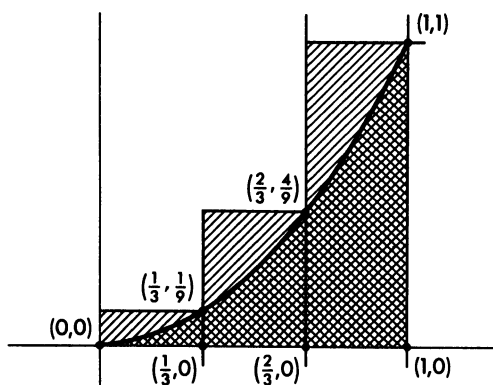


Fig. 7.5

As can be seen from Fig. 7.6, every reasonable definition of the area of our almost-triangle will give $A > \frac{5}{27}$.

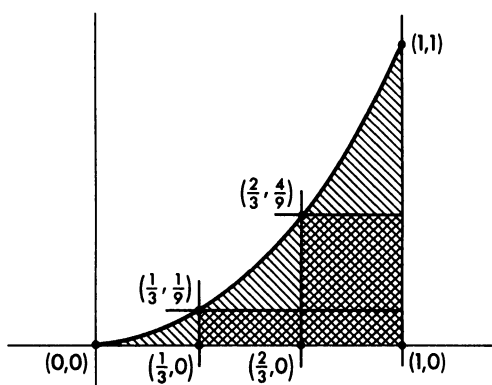


Fig. 7.6

These two lines of investigation give

$$\frac{5}{27} < A < \frac{14}{27}.$$

Since dividing the domain of $f()$ into three small intervals gave us a better approximation to the number A than dividing the domain of $f()$ into two smaller intervals, it is reasonable to hope that dividing the domain of $f()$ into four small intervals will give an approximation to A which will be even better. But then, dividing the domain into five small intervals, six small intervals, and so on, could well give successively better approximations to the number A . Then, it will be of great value to us to consider what approximation to A we have when we divide the domain of $f()$ into n intervals, where n is a natural number, and our computation will be eased if we make each of those intervals have the same length.

We then have, serving as end points of these n intervals, the $n + 1$ numbers

$$0, \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \frac{4}{n}, \dots, \frac{n-2}{n}, \frac{n-1}{n}, 1.$$

For each small interval, the absolute minimum value of $f()$ occurs at the left-hand end point and the absolute maximum value of $f()$ occurs at the right-hand end point. We then have an upper bound to the number A by considering the sum

$$\begin{aligned} & f\left(\frac{1}{n}\right) \frac{1}{n} + f\left(\frac{2}{n}\right) \frac{1}{n} + f\left(\frac{3}{n}\right) \frac{1}{n} + \dots + f\left(\frac{n-1}{n}\right) \frac{1}{n} + f(1) \frac{1}{n} \\ &= \frac{1^2}{n^2} \frac{1}{n} + \frac{2^2}{n^2} \frac{1}{n} + \frac{3^2}{n^2} \frac{1}{n} + \dots + \frac{(n-1)^2}{n^2} \frac{1}{n} + \frac{1}{n} \\ &= \frac{1}{n^3} [1^2 + 2^2 + 3^2 + \dots + (n-1)^2 + n^2] \\ &= \frac{1}{n^3} \frac{n(n+1)(2n+1)}{6} \quad (\text{by Theorem 7.2}) \\ &= \frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2}. \end{aligned}$$

We have a lower bound to the number A by considering the sum

$$\begin{aligned} & f(0) \frac{1}{n} + f\left(\frac{1}{n}\right) \frac{1}{n} + f\left(\frac{2}{n}\right) \left(\frac{1}{n}\right) + \dots + f\left(\frac{n-2}{n}\right) \frac{1}{n} + f\left(\frac{n-1}{n}\right) \frac{1}{n} \\ &= 0 + \frac{1^2}{n^2} \frac{1}{n} + \frac{2^2}{n^2} \frac{1}{n} + \dots + \frac{(n-2)^2}{n^2} \frac{1}{n} + \frac{(n-1)^2}{n^2} \frac{1}{n} \\ &= \frac{1}{n^3} [1^2 + 2^2 + \dots + (n-2)^2 + (n-1)^2] \\ &= \frac{1}{n^3} \frac{(n-1)(n)(2n-1)}{6} \quad (\text{by Theorem 7.2}) \\ &= \frac{1}{3} - \frac{1}{2n} + \frac{1}{6n^2} \end{aligned}$$

We then conclude that, for every reasonable definition of the area of our almost-triangle, and for every natural number, n , we have

$$\frac{1}{3} - \frac{1}{2n} + \frac{1}{6n^2} < A < \frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2}.$$

But the only manner in which the fixed number A can satisfy the above inequality for *every* natural number n , is for the number A actually to be $\frac{1}{3}$. Consequently, every reasonable definition for the area of the almost-triangle bounded by the graph of $f()$, the horizontal line segment joining $(0, 0)$ to $(1, 0)$, and the vertical line segment joining $(1, 0)$ to $(1, 1)$ must yield that area to be the number $\frac{1}{3}$.

Exercises

Let $g() = I()$.

1. Find the values of $g()$ for the numbers

$$0, \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \frac{4}{n}, \dots, \frac{n-2}{n}, \frac{n-1}{n}, 1,$$

where n is some natural number.

2. Find

$$\frac{1}{n} \left[g(0) + g\left(\frac{1}{n}\right) + g\left(\frac{2}{n}\right) + g\left(\frac{3}{n}\right) + \dots + g\left(\frac{n-2}{n}\right) + g\left(\frac{n-1}{n}\right) \right].$$

3. Find

$$\frac{1}{n} \left[g\left(\frac{1}{n}\right) + g\left(\frac{2}{n}\right) + g\left(\frac{3}{n}\right) + g\left(\frac{4}{n}\right) + \dots + g\left(\frac{n-1}{n}\right) + g(1) \right].$$

4. Draw the graph of $g()$ and the graph of $x = 1$.

5. What is the area of that part of the plane bounded by the piece of the graph of $g()$ between $(0, 0)$ and $(1, 1)$, the line $x = 1$, and the x -axis. Why?

6. Why are the numbers found in exercises 2 and 3 estimates of the area found in exercise 5?

7. Repeat exercises 1–6 for $g() = \{I^3\}()$.

8. Find successive estimates to the area of that part of the plane bounded by the graphs $y = x^2$, $y = 0$, and $x = 1$, by subdividing, and then drawing secant line segments joining the points

$$(0, 0), \left(\frac{1}{n}, \frac{1}{n^2}\right), \left(\frac{2}{n}, \frac{4}{n^2}\right), \dots, \left(\frac{n-1}{n}, \frac{(n-1)^2}{n^2}\right), (1, 1).$$

7.3. SEQUENCES

In the preceding section, we found that we could get approximations to an area by successively dividing an interval into more and more pieces. We kept track of these successive subdivisions—and their resulting approximations to the area—by indicating a natural number n , which natural number was the number of small intervals into which we divided the domain of a function. If we were to use the symbol $\overline{A}_n$ to indicate the number which resulted from dividing the domain of $f()$ into n small intervals (of equal length) and considering the rectangles whose height was the absolute maximum value of $f()$ on each small interval, then to each natural number n , there would correspond some number $\overline{A}_n$. We may then consider the expression $\overline{A}_n$ as giving a function whose domain is the set of natural numbers, and whose range of values is a certain set of numbers.

In like manner, we may denote by $\underline{A}_n$ the number which is a lower bound to the area of the almost-triangle we dealt with, and this expression will give another function whose domain is the set of natural numbers, and whose range of values is some set of numbers.

Generally, without concern as to the manner in which we arrive at a particular expression involving natural numbers, we have

Definition 7.1: A *sequence* is a function whose domain is the set of natural numbers.

We are again faced with the usual double problem found in mathematics on encountering a new entity: what is this thing and how do we write the thing? We have defined what a sequence is, and, to emphasize that we are dealing with that special kind of function called a sequence, we will not use the usual notation for functions, but will adopt the following:

If a function $B()$ is a sequence, then the range of values of this sequence will be denoted by the symbol

$$\{B_n \mid n = 1, 2, \dots\}.$$

and particular numbers in this range of values by the symbols

$$B_1, B_2, B_3, \text{ and so on.}$$

It may happen that a sequence has the property that, at least beyond a certain place in the indexing, the numbers in the range of values somehow are "getting close to" one fixed number. That is, as we look through the natural numbers, which constitute the domain of a sequence, we see that the corresponding numbers in the range of values seem to be getting closer and closer to some one fixed number. For example, the sequence specified by

$$\left\{\frac{1}{n} \mid n = 1, 2, \dots\right\},$$

is so arranged that numbers in the range of values which are "far out"—that is, which correspond to "large" natural numbers—are "close to" the number 0. More exactly, however close to the number 0 we may want to be, we can find a natural number such that all numbers in the range of values of this sequence which correspond to that natural number, or any larger natural number, will be within the desired closeness of the number 0. This situation is similar to that encountered in defining the limit of a function, as we approximate a fixed number, so that we use the same word, but write things in a slightly different manner. Thus, we have:

Definition 7.2: If $\{B_n \mid n = 1, 2, \dots\}$ is the range of values of a sequence, then

$$\lim_n B_n$$

(which is read: "the limit, on n , of B_n ") exists and is the number L means:

For every $\epsilon > 0$, there exists a natural number N , such that,

$$\text{if } n \geq N, \text{ then } |B_n - L| < \epsilon.$$

According to this definition, we have

$$\lim_n \frac{1}{n} = 0,$$

since whatever positive number ϵ may be, we need only let N be an integer larger than $\frac{1}{\epsilon} + 1$ in order to be assured that if $n \geq N$, then $|\frac{1}{n} - 0| < \epsilon$. As with ordinary limits of functions, we are able to prove theorems about the sum, difference, or product of two sequences, where those algebraic combinations are already defined, since sequences are functions. Our main purpose here is to use the idea of a sequence in what is to follow, so that we will merely state these theorems without proof.

Theorem 7.3:

If $\{A_n \mid n = 1, 2, \dots\}$ and $\{B_n \mid n = 1, 2, \dots\}$ are the ranges of two sequences; if $\lim_n A_n$ is the number L ; and if $\lim_n B_n$ is the number M ; then

- a) $\lim_n (A_n + B_n)$ is the number $A + B$.
- b) $\lim_n (A_n B_n)$ is the number AB .
- c) $\lim_n (kA_n)$ is the number kA , where k is a constant.

This theorem may be used to investigate the two sequences constructed in the preceding section. We have

$$\begin{aligned} \lim_n \overline{A_n} &= \lim_n \left[\frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2} \right] \\ &= \lim_n \frac{1}{3} + \lim_n \frac{1}{2n} + \lim_n \frac{1}{6n^2} \\ &= \lim_n \frac{1}{3} + \frac{1}{2} \lim_n \frac{1}{n} + \frac{1}{6} \left[\lim_n \frac{1}{n} \right]^2 \\ &= \frac{1}{3}, \end{aligned}$$

since the first term gives a constant sequence, and $\lim_n \frac{1}{n} = 0$, as we have proved above. We also have

$$\begin{aligned} \lim_n \underline{A_n} &= \lim_n \left[\frac{1}{3} - \frac{1}{2n} + \frac{1}{6n^2} \right] \\ &= \lim_n \frac{1}{3} - \frac{1}{2} \lim_n \frac{1}{n} + \frac{1}{6} \left[\lim_n \frac{1}{n} \right]^2 \\ &= \frac{1}{3}. \end{aligned}$$

Exercises

Find the limits, if any, of the sequences specified by:

1. $A_n = n, n = 1, 2, \dots$
2. $A_n = 1, n = 1, 2, \dots$

3. $A_n = \frac{2}{n}, n = 1, 2, \dots$
4. $A_n = \frac{n+1}{n+2}, n = 1, 2, \dots$
5. $A_n = \frac{n+2}{2n+1}, n = 1, 2, \dots$
6. $A_n = \frac{3n-4}{5n+1}, n = 1, 2, \dots$
7. $A_n = \frac{3n-4}{5n^2+n}, n = 1, 2, \dots$
8. $A_n = \frac{3n-4}{5n^2+1}, n = 1, 2, \dots$
9. $A_n = \frac{2n^3+3n-1}{4n^2+2n+7}, n = 1, 2, \dots$
10. $A_n = \frac{2n^2+3n-1}{4n^3-2n+1}, n = 1, 2, \dots$
11. $A_n = \frac{2n^2+3n-1}{4n^2-2n+3}, n = 1, 2, \dots$
12. If a sequence is specified by

$$A_n = \frac{1}{n^3} \quad \text{for } n = 1, 3, 5, 7, \dots$$

and

$$A_n = -\frac{3}{n} \quad \text{for } n = 2, 4, 6, 8, \dots,$$

does this sequence have a limit, and if so, what number is its limit?

7.4. UPPER AND LOWER INTEGRALS

If we combine the ideas in the preceding three sections, we can see that the problem of defining what is meant by the area of a part of the plane bounded by the graph of a function and some vertical and horizontal lines involves dividing the domain of this function into small intervals, constructing two sequences by means of finding the absolute maximum and absolute minimum of the function on each of these small intervals, and investigating the limits of the sequences formed. The end result of this process is useful and used, and finding area is only one of the many uses to which we put this process. Thus, we will suppress the geometric aspect and give our definition purely in terms of the functions involved.

Definition 7.3: If $f()$ is a function, and the interval:

the set of all $t, \quad a \leq t \leq b$

is in the domain of $f()$, then

$$\int_a^b f() \, d()$$

(which is read "the upper integral, from a to b , of f ") is defined to be the limit of the sequence

$$\{\overline{S}_n \mid n = 1, 2, \dots\},$$

provided this limit exists, and where this sequence is constructed as follows:

For the natural number n , we divide the interval:

the set of all t , $a \leq t \leq b$,

into n subintervals:

the set of all t , $a \leq t \leq a + \frac{b-a}{n}$,

the set of all t , $a + \frac{b-a}{n} \leq t \leq a + 2 \frac{b-a}{n}$, ,

the set of all t , $a + 2 \frac{b-a}{n} \leq t \leq a + 3 \frac{b-a}{n}$,

$\vdots$

the set of all t , $a + (n-2) \frac{b-a}{n} \leq t \leq a + (n-1) \frac{b-a}{n}$,

and the set of all t , $a + (n-1) \frac{b-a}{n} \leq t \leq b$.

For each of these n subintervals, we find the absolute maximum value of $f()$. We let $\overline{x}_1$ be a number in the first interval which gives $f()$ an absolute maximum value for that first interval: we let $\overline{x}_2$ be a number in the second interval which gives $f()$ an absolute maximum value for that second interval; we let $\overline{x}_3$ be a number in the third interval which gives $f()$ an absolute maximum value for that third interval; ...; we let $\overline{x}_{n-1}$ be a number in the $(n-1)$ st (next to the last) interval which gives $f()$ an absolute maximum value for that $(n-1)$ st interval; and we let $\overline{x}_n$ be a number in the n -th interval which gives $f()$ an absolute maximum value for that n -th interval.

$\overline{S}_n$ is the number

$$[f(\overline{x}_1) + f(\overline{x}_2) + f(\overline{x}_3) + \dots + f(\overline{x}_{n-1}) + f(\overline{x}_n)] \frac{b-a}{n}.$$

According to this definition, if $g()$ is specified by

$$g(x) = x^2, \quad 0 \leq x \leq 1,$$

then

$$\begin{aligned} \int_0^1 g() d() &= \lim_n \left[\frac{1^2}{n^2} + \frac{2^2}{n^2} + \frac{3^2}{n^2} + \dots + \frac{(n-1)^2}{n^2} + \frac{n^2}{n^2} \right] \frac{1}{n} \\ &= \lim_n \left[\frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2} \right] \\ &= \frac{1}{3}. \end{aligned}$$

In a similar manner, we define the lower integral of a function, over a particular interval, by using, instead of the absolute maximum values of the function on each subinterval, the absolute minimum values. We have:

Definition 7.4: If $f()$ is a function, and if the interval:

$$\text{the set of all } t, \quad a \leq t \leq b$$

is in the domain of $f()$, then

$$\int_a^b f() \, d()$$

(which is read “the lower integral, from a to b , of f ”) is defined to be the limit of the sequence

$$\{S_n \mid n = 1, 2, \dots\},$$

provided this limit exists, and where this sequence is constructed in a manner similar to that contained in the definition (7.3), except that we find the numbers

$$\underline{x}_1, \underline{x}_2, \underline{x}_3, \dots, \underline{x}_{n-1}, \underline{x}_n,$$

which give values of $f()$:

$$f(\underline{x}_1), f(\underline{x}_2), f(\underline{x}_3), \dots, f(\underline{x}_{n-1}), f(\underline{x}_n),$$

which are the absolute minimum values of $f()$ for the corresponding subintervals. Thus,

$$S_n = [f(\underline{x}_1) + f(\underline{x}_2) + \dots + f(\underline{x}_n)] \frac{b-a}{n}.$$

For the function $g()$ specified by

$$g(x) = x^2, \quad 0 \leq x \leq 1,$$

we have

$$\begin{aligned} \int_0^1 g() \, d() &= \lim_n \left[\frac{0^2}{n^2} + \frac{1^2}{n^2} + \frac{2^2}{n^2} + \dots + \frac{(n-2)^2}{n^2} + \frac{(n-1)^2}{n^2} \right] \frac{1}{n} \\ &= \lim_n \left[\frac{1}{3} - \frac{1}{2n} + \frac{1}{6n^2} \right] \\ &= \frac{1}{3}. \end{aligned}$$

It should be noted that *all* of the steps called for in the definitions must be possible, so that any function for which we cannot always find absolute extremes for all subintervals will simply not have an upper or lower integral.

Exercises

Find the sequences whose limits are $\overline{\int_a^b f() d()}$ and $\underline{\int_a^b f() d()}$, where $f()$ is specified by the following, and a and b have the values given.

1. $f(x) = x$; $a = 0$, $b = 1$.
2. $f(x) = 3x$; $a = 0$, $b = 1$.
3. $f(x) = -2x$; $a = 0$, $b = 1$.
4. $f(x) = x$; $a = 1$, $b = 2$.
5. $f(x) = x^2$; $a = 0$, $b = 2$.
6. $f(x) = x^2$; $a = -1$, $b = 0$.
7. $f(x) = x^2$; $a = -1$, $b = 1$.
8. $f(x) = 2x^2$; $a = 0$, $b = 1$.
9. $f(x) = 2x^2$; $a = 0$, $b = 2$.
10. $f(x) = -3x^2$; $a = 0$, $b = 1$.
11. $f(x) = -x^2$; $a = -1$, $b = 0$.

7.5. THE DEFINITE INTEGRAL

In the example considered, we saw that the upper integral of $\{I^2\}()$, over the interval:

the set of all t , $0 \leq t \leq 1$,

and the lower integral of $\{I^2\}()$, over the same interval, were the same number. In general, if such should occur, we have

Definition 7.5: If $f()$ is a function and

$$\overline{\int_a^b f() d()} \quad \text{and} \quad \underline{\int_a^b f() d()}$$

both exist and are the same number, then this number is defined to be

$$\int_a^b f() d()$$

(read “the integral, from a to b , of $f()$ ”).

We have discovered, then, that for $f()$ given by

$$f(x) = x^2, \quad 0 \leq x \leq 1,$$

we have

$$\int_0^1 f() d() = \frac{1}{3}.$$

Now that we have defined what is meant by the integral of a function, over an interval in the domain, we need a way of writing which will be

more concise than that which we have been using. We will adopt the idea of a "variable of integration," using the blank parentheses already present in our symbolism. Thus, if $g()$ is specified by

$$g(s) = s^2 + s,$$

that is, if $g() = \{I^2 + I\}()$, we will adopt the policy of using, as a variable of integration, any convenient letter (preferably one not already used somewhere else in the same expression), and denote the integral of $\{I^2 + I\}()$, over the interval:

$$\text{the set of all } x, \quad 3 \leq x \leq 4,$$

by any of the following:

$$\int_3^4 \{I^2 + I\}() d(), \quad \int_3^4 (r^2 + r) dr, \quad \int_3^4 (t^2 + t) dt, \\ \int_3^4 (x^2 + x) dx, \quad \text{or even} \quad \int_3^4 [()^2 + ()] d().$$

In this manner of writing, we use $d()$, dr , dt , dx , etc., purely as a device to indicate the end of the integral, and to point out which variable is being used.

At this point, we have defined what an upper integral, lower integral, and integral are, and have given in the definitions a fairly complicated scheme for finding these numbers. Our next task will be first to see just what kind of functions have integrals, and over which intervals, and then to examine at least a few functions for which some or all of these integrals do not exist.

In order to allow us room for maneuvering, we will state without proof:

Theorem 7.4:

If $f()$ is a function; if the interval:

$$\text{the set of all } t, \quad a \leq t \leq b$$

is in the domain of $f()$; and if $f()$ is continuous at every number in this interval; then $\int_a^b f() d()$ exists.

(We also state the conclusion of this theorem as " $f()$ is integrable over the interval a, b .")

The gist of the proof of this theorem consists in observing that, as we look at larger and larger natural numbers, the number of subdivisions of the original interval is so large that the lengths of the individual subintervals become extremely small. Since a continuous function does not change its values too rapidly on a very small subinterval, the absolute maximum and absolute minimum values of this continuous function, on such a small interval, do not differ greatly. Thus, if the limits of the two sequences—one for the upper integral and the other for the lower integral—are to

exist, they will not differ too much, if any, since the terms in one sequence which "far out" are more or less the same numbers as the terms in the other sequence which are "far out." This plausibility argument is not a proof.

Exercises

Find the integrals for the functions and intervals in the exercises of the preceding section.

7.6. FUNCTIONS WITHOUT INTEGRALS

In this section, we give an example of a function for which the upper integral does not exist, and thus one whose integral cannot exist; and also an example of a function whose upper and lower integrals exist, but are different numbers, so that the integral of such a function does not exist.

Let $F()$ be the function whose domain is the interval:

$$\text{the set of all } x, \quad 0 \leq x \leq 1,$$

given by:

$$\text{a) } F(0) = 0;$$

$$\text{b) } F(t) = \frac{1}{t^2}, \quad 0 < t \leq 1.$$

Since all the numbers in the range of values of this function are non-negative, all the terms in each "upper sum," $\overline{S}_n$, will certainly be non-negative numbers. Then, each upper sum $\overline{S}_n$ will not be less than any one of its own terms. If we divide the interval which is the domain of the function $F()$ into n subintervals of equal length (where n is a natural number), then the second of these subintervals is:

$$\text{the set of all } t, \quad \frac{1}{n} \leq t \leq \frac{2}{n}.$$

The function $F()$ is decreasing at each point in this small interval, so that the absolute maximum value of $F()$ for this interval will be attained at the left-hand end point, $1/n$, and this absolute maximum value will be the number

$$F(1/n) = n^2.$$

Since the length of the small interval is $1/n$, the second term of the upper sum $\overline{S}_n$ will be the number n . We then have

$$\overline{S}_n > n.$$

The sequence $\{\overline{S}_n \mid n = 1, 2, \dots\}$ then cannot have a limit, since the numbers in the range of values increase indefinitely. Thus, $\int_0^1 F() d()$ does not exist.

Let $G()$ be the function whose domain is the interval:

$$\text{the set of all } x, \quad 0 \leq x \leq 1,$$

given by:

- a) If x is a rational number, $G(x) = 1$;
- b) If x is an irrational number, $G(x) = 0$.

Every sub-interval of the domain, no matter how small, will contain both irrational and rational numbers. Thus, on every subinterval, the absolute maximum value of $G()$ for that subinterval will be 1 and the absolute minimum value of $G()$ for that subinterval will be 0.

Each number in the range of values of the sequence $\{\overline{S_n} \mid n = 1, 2, \dots\}$ will then be 1, and each number in the range of values of the sequence $\{S_n \mid n = 1, 2, \dots\}$ will be 0. Since the limits of constant sequences are particularly easy to find, we have

$$\int_0^1 G() d() = 1$$

and

$$\int_0^1 G() d() = 0.$$

Since the upper and lower integrals are different numbers, $\int_0^1 G() d()$ does not exist.

Exercises

1. Show that the upper integral, from 0 to 1, of the function given by

$$f(z) = \frac{1}{z^3}, \quad z \neq 0; \quad f(0) = 0$$

does not exist.

2. Find the upper and lower integrals for the function specified by:

$$\begin{aligned} H(t) &= t, & 0 \leq t \leq 1, & \quad t \text{ rational;} \\ H(t) &= 1, & 0 \leq t \leq 1, & \quad t \text{ irrational.} \end{aligned}$$

Find the upper and lower integrals for the specified functions over the indicated intervals:

- 3. $f(t) = t + 1, \quad 0 \leq t \leq 1, \quad t \text{ rational};$
 $f(t) = 2 - t, \quad 0 \leq t \leq 1, \quad t \text{ irrational.}$
- 4. $g(r) = r^2, \quad 0 \leq r \leq 1, \quad r \text{ rational};$
 $g(r) = 0, \quad 0 \leq r \leq 1, \quad r \text{ irrational.}$

7.7. SOME THEOREMS ABOUT INTEGRALS

The integral of a function over an interval in the domain of that function has a number of important properties, which we will give in this section.

Theorem 7.5:

If the interval:

the set of all t , $a \leq t \leq b$,

is in the domain of the function $f()$, and c is a number inside this interval, then

$$\int_a^b f() d() = \int_a^c f() d() + \int_c^b f() d(),$$

provided only that all three integrals exist.

The proof of this theorem depends on the subdivision of the original interval into so many subintervals that we can distinguish between those subintervals lying between a and c and those lying between c and b . The manner of subdividing into subintervals all of whose lengths are the same introduces an awkwardness into the proof of this theorem in the case that none of the points of subdivision falls exactly on the number c . This is purely a technical awkwardness—at least when the function concerned is continuous—but it does complicate the proof, so that we will take the easy way out and omit the proof of this theorem.

Theorem 7.5 is particularly applicable when integrating a function which has a rule saying different things for different parts of the domain. For example, if $H()$ were to have domain:

the set of all r , $0 \leq r \leq 2$

and a rule which says:

a) If $0 \leq r \leq 1$, then $H(r) = r$;

b) If $1 \leq r \leq 2$, then $H(r) = 1$,

the graph of $H()$ would consist of the line segment joining $(0, 0)$ to $(1, 1)$ together with the line segment joining $(1, 1)$ to $(2, 1)$. Applying this theorem, we have

$$\begin{aligned} \int_0^2 H() d() &= \int_0^1 H() d() + \int_1^2 H() d() \\ &= \int_0^1 r dr + \int_1^2 1 dr. \end{aligned}$$

We have seen that the first integral on the right is the number $\frac{1}{2}$, and the second, as we will see, is the number 1, so that

$$\int_0^2 H() d() = \frac{1}{2} + 1 = \frac{3}{2}.$$

Theorem 7.6:

If $k()$ is a constant function, and $a \leq b$, then

$$\int_a^b k() d() = k(b - a).$$

Proof:

1. $k()$ is a constant function by hypothesis.
2. On every subinterval of every subdivision of the interval of integration, the absolute maximum value of $k()$ is the number k by step 1 and the definition of absolute maximum.
3. For the subdivision of the interval of integration into n subintervals (where n is natural number), $\overline{S}_n$ is the number $k(b - a)$ since the sum of the lengths of the subintervals is the length of the entire interval.
4. $\overline{S}_n, n = 1, 2, \dots$ is a constant sequence by step 3.
5. $\int_a^b k() d()$ is the number $k(b - a)$ by steps 3 and 4 and the definition of upper integral.
6. On every subinterval, of every subdivision of the interval of integration, the absolute minimum value of $k()$ is the number k by step 1 and the definition of absolute minimum.
7. For the subdivision of the interval of integration into n subintervals (where n is a natural number), $\underline{S}_n$ is the number $k(b - a)$ since the sum of the lengths of the subintervals is the length of the entire interval.
8. $\underline{S}_n, n = 1, 2, \dots$ is a constant sequence by step 7.
9. $\int_a^b k() d()$ is the number $k(b - a)$ by steps 6 and 7 and the definition of lower integral.
10. $\int_a^b k() d()$ is the number $k(b - a)$ by steps 5 and 9 and the definition of integral.

This theorem means that, as with differentiation, constant functions are the easiest to handle.

Theorem 7.7:

If $k()$ is a constant function; $f()$ is a function; and $\int_a^b f() d()$ exists; then

$$\int_a^b \{kf\}() d() \text{ exists and is } k \int_a^b f() d().$$

Proof:

CASE 1, $k = 0$:

Here all terms in the upper and lower sums are 0, so that $\int_a^b kf() d()$ and $k \int_a^b f() d()$ are both 0.

CASE 2, $k > 0$:

Here all terms in the upper and lower sums merely contain the constant factor k , so that the theorem on limits of sequences applies and we have

$$\overline{\int_a^b \{kf\}() d()} = k \overline{\int_a^b f() d()}$$

and

$$\underline{\int_a^b \{kf\}() d()} = k \underline{\int_a^b f() d()}.$$

CASE 3, $k < 0$:

Here all terms in the upper sums for $f()$ become terms in the lower sums for $\{kf\}()$, and vice-versa, with the constant factor k appearing everywhere. Again, the theorem on limits of sequences applies and we have

$$\overline{\int_a^b \{kf\}() d()} = k \underline{\int_a^b f() d()}$$

and

$$\underline{\int_a^b \{kf\}() d()} = k \overline{\int_a^b f() d()}.$$

As a result of this theorem, our labors in the vineyard of the function $\{I^2\}()$ enable us to evaluate

$$\int_0^1 \{kI^2\}() d(),$$

no matter what constant k might be. Thus,

$$\int_0^1 3t^2 dt, \quad \int_0^1 \left[-\frac{4}{3}u^2\right] du, \quad \int_0^1 \sqrt{5}z^2 dz$$

are the numbers

$$3\left(\frac{1}{3}\right) = 1, \quad -\frac{4}{3}\left(\frac{1}{3}\right) = -4/9, \quad \text{and} \quad \sqrt{5}\left(\frac{1}{3}\right) = \sqrt{5}/3,$$

respectively.

Finally, we have

Theorem 7.8:

If $\int_a^b f() d()$ and $\int_a^b g() d()$ both exist, then $\int_a^b \{f + g\}() d()$ exists and is the number $\int_a^b f() d() + \int_a^b g() d()$.

The proof of this theorem follows almost immediately from the definitions of upper integral, lower integral, integral, and the sum of two functions.

Corollary to Theorem 7.8:

Under the same hypothesis as in Theorem 7.8,

$$\int_a^b \{f - g\}() d() = \int_a^b f() d() - \int_a^b g() d().$$

Proof: $\{f - g\}() = \{f + \{-g\}\}()$, Theorems 7.7 and 7.8.

Since the numbers that were the end points of the interval of the integral of a function appear in the symbolism for the integral, the definition of integral requires that the lower "limit of integration" be a number less than the upper "limit of integration." Then, if we want to write

$$\int_p^q f() d(),$$

where $f()$ is a (hopefully, continuous) function and $p \geq q$, we must define precisely what we mean by this symbol.

Definition 7.6: If $p > q$, then

$$\int_p^q f() d() = - \int_q^p f() d().$$

If $p = q$, then

$$\int_p^p f() d() = 0.$$

This definition implies that reversing the limits of integration is equivalent to taking the negative of the integral. For example,

$$\int_1^0 t^2 dt = - \int_0^1 t^2 dt = -\frac{1}{3}.$$

As a result of this definition, we have

Corollary to Theorem 7.5:

If a , b , and c are any three numbers, then

$$\int_a^b f() d() = \int_a^c f() d() + \int_c^b f() d(),$$

provided only that all three integrals exist.

Proof:

Part I.

If any two of the three numbers: a , b , and c are the same, then one of the numbers on the right is 0 and the theorem is true by law A-4.

Part II.

CASE 1, $c < a < b$:

By Theorem 7.5,

$$\int_a^b f() d() = \int_a^c f() d() + \int_c^b f() d().$$

Then,

$$-\int_c^a f() d() + \int_c^b f() d() = \int_a^b f() d().$$

Since $a < c$,

$$\int_a^b f() d() = \int_a^c f() d() + \int_c^b f() d().$$

CASE 2, $a < c < b$:

This case is the same as Theorem 7.5.

CASE 3, $a < b < c$;

CASE 4, $c < b < a$;

CASE 5, $b < c < a$;

CASE 6, $b < a < c$.

Exercises

1. Complete the proof of the Corollary to Theorem 7.5 in cases 3, 4, 5, and 6.

Show that:

$$\int_0^b r dr = \frac{b^2}{2} \quad \text{and} \quad \int_0^c t^2 dt = \frac{c^3}{3}.$$

Use the theorems of this section to find:

- | | |
|------------------------------------|-------------------------------------|
| 2. $\int_0^2 r dr.$ | 11. $\int_0^3 r^2 dr.$ |
| 3. $\int_0^4 s ds.$ | 12. $\int_0^4 s^2 ds.$ |
| 4. $\int_2^4 t dt.$ | 13. $\int_3^4 t^2 dt.$ |
| 5. $\int_0^{-3} u du.$ | 14. $\int_0^{-2} u^2 du.$ |
| 6. $\int_{-3}^0 v dv.$ | 15. $\int_{-2}^0 v^2 dv.$ |
| 7. $\int_{-3}^2 w dw.$ | 16. $\int_{-2}^3 w^2 dw.$ |
| 8. $\int_{-3}^4 x dx.$ | 17. $\int_0^1 [x + x^2] dx.$ |
| 9. $\int_{-6}^7 y dy.$ | 18. $\int_a^b [y + y^2] dy.$ |
| 10. $\int_a^b z dz.$ | 19. $\int_{-1}^3 [1 + z + z^2] dz.$ |
| 20. $\int_p^a [ar^2 + br + c] dr.$ | |

7.8. AREA

Now that we have looked at the question of defining area for parts of the plane other than triangles, rectangles, or sectors of circles, and have defined a new mathematical entity: the integral of a function over an interval in the domain of that function, we are able actually to give a definition of area.

Definition 7.7: If $f()$ and $g()$ are functions; if the interval:

the set of all t , $a \leq t \leq b$,

is in the domain of both $f()$ and $g()$; and if, for all t in this interval, $f(t) \leq g(t)$, then the *area* of that portion of the plane bounded by the piece of the graph of $f()$ between $(a, f(a))$ and $(b, f(b))$, the piece of the graph of $g()$ between $(a, g(a))$ and $(b, g(b))$, the vertical line with equation $x = a$, and the vertical line with equation $x = b$, is the number

$$\int_a^b \{g - f\}() d().$$

In this definition, it is vital that the piece of the graph of $f()$ between $(a, f(a))$ and $(b, f(b))$ never crosses that piece of the graph of $g()$ which lies between $(a, g(a))$ and $(b, g(b))$. We have then defined the word “area” only for such parts of the plane as are bounded above and below by pieces of the graphs of two functions (either or both of which may be constant), and to the right and left by vertical line segments (either or both of which may have zero length. A variety of such situations is shown in Fig. 7.7.

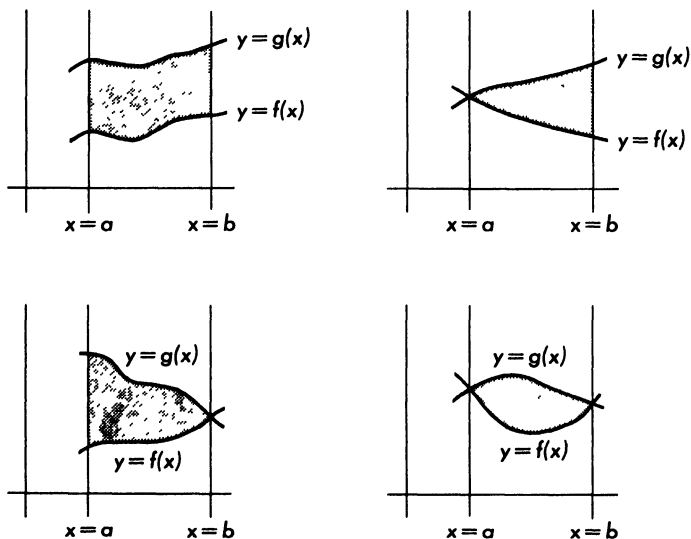


Fig. 7.7

In Fig. 7.7, the area of the shaded portion of the plane is, in each case, the number

$$\int_a^b \{g - f\} (x) dx.$$

As a direct result of this definition of area, the area of the almost-triangle encountered in section 7.1 is the number

$$\int_0^1 r^2 dr = \frac{1}{3}$$

Exercises

For any numbers, a and b ,

$$\begin{aligned}\int_a^b s ds &= \frac{b^2 - a^2}{2}, \\ \int_a^b t^2 dt &= \frac{b^3 - a^3}{3}, \\ \int_a^b u^3 du &= \frac{b^4 - a^4}{4}.\end{aligned}$$

Use these facts, together with the theorems on integrals, to find the area of that part of the plane bounded by the graphs of:

1. $y = x, x = 3, y = 0$.
2. $y = x, x = -2, y = 0$.
3. $y = x, x = -4, y = 3$.
4. $y = x^2, x = 1, x = 2, y = 0$.
5. $y = x^2, x = 1, x = 2, y = -1$.
6. $y = x^2, x = 1, x = 2, y = 1$.
7. $y = x^3, x = 0, x = 1, y = 0$.
8. $y = x^3, x = -1, x = 0, y = 0$.
9. $y = x^3, x = -1, x = 1, y = -1$.
10. $y = x^3 - 3x, x = -3, x = 3, y = 0$.

7.9. THE FUNDAMENTAL THEOREM OF CALCULUS

If we were to find the number which is the integral of a given function, over a given interval in the domain of that function, only by using the definitions of upper and/or lower integral, the process would be laborious, and, in fact, would be impossible for most functions other than polynomials. We need, then, to investigate an alternate method of evaluating an integral, which method does not directly make use of the definition of integral. Because of the importance of the theorem of this section, we will prove the theorem both developmentally (that is, before stating it) and formally (after stating it).

Let us assume that a given function $f()$ is continuous at each number in some interval:

the set of all t , $a \leq t \leq b$.

Then, if x is some fixed number of this interval, the function $f()$ is also continuous at every number of the smaller interval:

the set of all t , $a \leq t \leq x$.

By Theorem 7.4, for any fixed number x in the original interval,

$$\int_a^x f() d()$$

will exist and be a number. Since, to each number x in the original interval, there corresponds (gets paired) the number

$$\int_a^x f() d(),$$

we have actually specified a new function $F()$ by the modified equation

$$F(x) = \int_a^x f() d(), \quad a \leq x \leq b.$$

By this rule, we have

$$F(a) = \int_a^a f() d() = 0,$$

and

$$F(b) = \int_a^b f() d().$$

If we could find another rule for the function $F()$, then applying that rule to the number b would yield the integral we are seeking. Let us see if we can differentiate the new function $F()$.

For x , a fixed number between a and b , $F'(x)$ will be the number

$$\lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h},$$

provided that this limit exists.

By the rule specifying $F()$,

$$F(x+h) = \int_a^{x+h} f() d();$$

$$F(x) = \int_a^x f() d().$$

For any value of h for which the number $x+h$ is in the domain of $F()$, we have

$$\begin{aligned} F(x+h) - F(x) &= \int_a^{x+h} f() d() - \int_a^x f() d() \\ &= \int_x^{x+h} f() d() \end{aligned}$$

by the Corollary to Theorem 7.5.

We then need to evaluate

$$\lim_{h \rightarrow 0} \left[\frac{1}{h} \int_x^{x+h} f(t) dt \right].$$

We started with the assumption that $f(t)$ was continuous at every number in the original interval. Thus, $f(t)$ is continuous at the number x . By the definition of continuity, this means

$$\lim_{t \rightarrow x} f(t) = f(x).$$

By the definition of limit, if ϵ is any positive number, there is an interval:

$$\text{the set of all } t, \quad x - d < t < x + d \quad (\text{where } d > 0),$$

such that, if t is in this (small) interval around x ,

$$f(x) - \epsilon < f(t) < f(x) + \epsilon$$

(that is, $f(t)$ will be within ϵ units of the number $f(x)$). Considering t as a variable, the expressions $f(x) - \epsilon$ and $f(x) + \epsilon$, specify constant functions. Considering the integrals of the three functions specified by the terms in the above string of inequalities, the definition of integral tells us that

$$\int_x^{x+h} [f(x) - \epsilon] dt < \int_x^{x+h} f(t) dt < \int_x^{x+h} [f(x) + \epsilon] dt$$

whenever $|h| < d$.

This last string of inequalities implies that, for $h > 0$,

$$h[f(x) - \epsilon] < \int_x^{x+h} f(t) dt < h[f(x) + \epsilon]$$

and, if $h < 0$,

$$h[f(x) + \epsilon] < \int_x^{x+h} f(t) dt < h[f(x) - \epsilon],$$

by examining the similar inequalities for the sums whose limits are the middle integrals.

Then, regardless of whether h is positive or negative

$$\frac{1}{h} \int_x^{x+h} f(t) dt$$

will be between the numbers $f(x) - \epsilon$ and $f(x) + \epsilon$, whenever $0 < |h| < d$.

This means that, given any $\epsilon > 0$, we can find a positive number d , such that,

$$\text{if } 0 < |h| < d, \quad \text{then} \quad \left| \frac{1}{h} \int_x^{x+h} f(t) dt - f(x) \right| < \epsilon.$$

This statement appears in the definition of limit, so we conclude

$$F'(x) = \lim_{h \rightarrow 0} \left[\frac{1}{h} \int_x^{x+h} f(t) dt \right] = f(x).$$

Since there was nothing special about which number in the given interval x was, we find that $F'(x)$ is specified by

$$F'(x) = f(x), \quad a \leq x \leq b.$$

This equation says that the function $F()$ is an antiderivative of $f()$. Since it happens that any two antiderivatives of a given function can differ only by a constant function, if $G()$ is an anti-derivative of $f()$ which we can find, then we know that

$$F() = \{G + k\}(),$$

where $k()$ is some constant function,

We can now, using a rule of $G()$, find

$$\begin{aligned} \int_a^b f() d() &= F(b) \\ &= F(b) - 0 \\ &= F(b) - F(a) \\ &= G(b) + k - G(a) + k \\ &= G(b) - G(a). \end{aligned}$$

Our theorem is:

Theorem 7.9:

(The Fundamental Theorem of Calculus)

If the function $f()$ is continuous at each number in the interval:

the set of all t , $a \leq t \leq b$;

and the function $G()$ is an anti-derivative of $f()$; then

$$\int_a^b f() d() = G(b) - G(a).$$

Proof:

1. Let r be a number in the interval:

the set of all t ,

$$a \leq t \leq b.$$

2. $f()$ is continuous at r

$$3. \lim_{t \rightarrow r} f(t) = f(r)$$

by step 1 and the hypothesis.

by step 2 and the definition of continuity.

4. Pick $\epsilon > 0$.

5. There exists an interval:

the set of all t ,

$$r - d < t < r + d \text{ (where } d > 0),$$

such that, if t is in this interval,
then

$$t \text{ is in the domain of } f()$$

and

$$f(r) - \epsilon < f(t) < f(r) + \epsilon$$

by steps 3 and 4 and the definition of limit.

6. Let the function $F()$ be specified by

$$F(s) = \int_a^s f() d(), \quad a \leq s \leq b.$$

7. r is in the domain of $F'()$ if and only if

$$\lim_{h \rightarrow 0} \frac{F(r+h) - F(r)}{h}$$

exists

by the definition of derivative.

8. If $0 < |h| < d$, then $r+h$ is in the domain of $f()$ by step 5.
 9. If $0 < |h| < d$, then $r+h$ is in the domain of $F()$ by step 6.
 10. If $0 < |h| < d$, then

$$F(r+h) - F(r) = \int_a^{r+h} f() d() - \int_a^r f() d()$$

by steps 6, 8, and 9.

11. If $0 < h < d$, then

$$F(r+h) - F(r) = \int_r^{r+h} f() d()$$

by step 10 and Theorem 7.5.

12. If $-d < h < 0$, then

$$F(r+h) - F(r) = - \int_{r+h}^r f() d()$$

by step 10 and Theorem 7.5.

13. If $0 < h < d$, then

$$\int_r^{r+h} [f(r) - e] dt < \int_r^{r+h} f(t) dt < \int_r^{r+h} [f(r) + e] dt$$

by step 5 and the definition of integral.

14. If $-d < h < 0$, then

$$\int_{r+h}^r [f(r) - e] dt < \int_{r+h}^r f(t) dt < \int_{r+h}^r [f(r) + e] dt$$

by step 5 and the definition of integral.

15. If $0 < h < d$, then

$$\int_r^{r+h} [f(r) - e] dt < F(r+h) - F(r) < \int_r^{r+h} [f(r) + e] dt$$

by steps 11 and 13.

16. If $-d < h < 0$, then

$$-\int_r^{r+h} [f(r) - e] dt < -\int_r^{r+h} f(t) dt < -\int_r^{r+h} [f(r) + e] dt$$

by step 14 and the definition (7.6).

17. If $-d < h < 0$, then

$$\int_r^{r+h} [f(r) - e] dt > F(r+h) - F(r) > \int_r^{r+h} [f(r) + e] dt$$

by steps 16 and 12 and O-4b.

18. If $0 < h < d$, then

$$h[f(r) - e] < F(r+h) - F(r) < h[f(r) + e]$$

by step 15 and Theorem 7.6 on the integral of a constant function.

19. If $-d < h < 0$, then

$$(-h)[f(r) - e] < [F(r) - F(r+h)] < (-h)[f(r) + e]$$

by step 17 and Theorem 7.6 and O-4b.

20. If $0 < h < d$, then

$$f(r) - e < \frac{F(r+h) - F(r)}{h} < f(r) + e$$

by step 18 and O-4a.

21. If $-d < h < 0$, then

$$f(r) - e < \frac{F(r+h) - F(r)}{h} < f(r) + e$$

by step 19 and O-4b.

22. If $0 < |h| < d$, then

$$f(r) - e < \frac{F(r+h) - F(r)}{h} < f(r) + e$$

by steps 20 and 21.

23. For every $e > 0$, there is a $d > 0$, such that,

$$\text{if } 0 < |h| < d,$$

then h is in the domain of the function given by

$$\frac{F(r+h) - F(r)}{h}$$

and

$$\left| \frac{F(r+h) - F(r)}{h} - f(r) \right| < e$$

by steps 4, 5, and 22.

24. $\lim_{h \rightarrow 0} \frac{F(r+h) - F(r)}{h}$ exists and is the number $f(r)$ by step 23 and the definition of limit.
25. r is in the domain of $F'()$ and $F'(r) = f(r)$ by step 24 and the definition of derivative.
26. $F()$ is an anti-derivative of $f()$ by step 25 and the definition of anti-derivative.
27. $G()$ is an anti-derivative of $f()$ by hypothesis.
28. $\{F - G\}'()$ is a constant function whose range of values is the number 0 by steps 26 and 27.
29. $\{F - G\}()$ is some constant $k()$ by step 28 and the theorem that any anti-derivative of $\theta()$ is a constant function.
30. $F()$ is specified by
$$F(s) = G(s) + k, \quad a \leq s \leq b$$
 by step 29.
31. $\int_a^b f() d() = F(b)$ by step 6.
32. $F(a) = 0$ by step 6 and the definition of $\int_a^a f() d()$.
33. $\int_a^b f() d() = F(b) - F(a)$ by steps 28 and 29 and A-4.
34. $\int_a^b f() d() = [G(b) + k] - [G(a) + k]$ by steps 30 and 33.
35. $\int_a^b f() d() = G(b) - G(a)$ by step 34 and A-5.

Exercises

Use the Fundamental Theorem to show that

1. $\int_a^b r dr = \frac{b^2 - a^2}{2}.$
2. $\int_a^b s^2 ds = \frac{b^3 - a^3}{3}.$
3. $\int_a^b t^3 dt = \frac{b^4 - a^4}{4}.$
4. Draw that part of the graph of $y = x^3$ between $(0, 0)$ and $(3, 9)$.

5. Is the interval:

the set of all t , $2 - 0.1 < t < 2 + 0.1$
 composed of numbers, each of which is a solution of the inequality
 $4 - 0.5 < v^2 < 4 + 0.5$?

How does this relate to steps 4 and 5 of the proof of the Fundamental Theorem?

6. Draw that part of the plane bounded by the graphs of $y = 0$, $x = 1.9$, $x = 2.1$, and $y = x^2$.
7. Find the area of the rectangle bounded by the graphs of $y = 0$, $x = 1.9$, $x = 2$, and $y = 3.5$.
8. Find an integral whose value is the area of that part of the plane bounded by the graphs of $y = 0$, $x = 1.9$, $x = 2$, and $y = x^2$.
9. Find the area of the rectangle bounded by the graphs of $y = 0$, $x = 1.9$, $x = 2$, and $y = 4.5$.
10. By examining the graphs, how do the numbers found in exercises 7, 8, and 9 compare in size with each other? How does this comparison relate to steps 13 and 14 of the proof of the Fundamental Theorem?
11. Find the area of the rectangle bounded by the graphs of $y = 0$, $x = 2$, $x = 2.1$, and $y = 3.5$.
12. Find an integral whose value is the area of that part of the plane bounded by the graphs of $y = 0$, $x = 2$, $x = 2.1$, and $y = x^2$.
13. Find the area of the rectangle bounded by the graphs of $y = 0$, $x = 2$, $x = 2.1$, and $y = 4.5$.
14. By examining the graphs, how do the numbers found in exercises 11, 12, and 13 compare in size with each other? How does this comparison relate to steps 13 and 14 of the proof of the Fundamental Theorem?
15. Find an anti-derivative of $\{I^2\}()$.
16. If the anti-derivative found in exercise 15 is called $F()$, find $F(1.9) - F(2)$. How does this number relate to the result of exercise 10?
17. With $F()$ as in the preceding exercise, find $F(2.1) - F(2)$. How does this number relate to the result of exercise 14?
18. With $F()$ as above, find

$$\frac{F(1.9) - F(2)}{1.9 - 2}, \quad \frac{F(2.1) - F(2)}{2.1 - 2}, \quad F'(2)$$

What is another method for finding $F'(2)$?

7.10. APPLICATION OF THE FUNDAMENTAL THEOREM TO THE EVALUATION OF INTEGRALS

Now that we have established a connection between the integral of a function, over an interval in its domain, and an anti-derivative of that function—which integral and anti-derivative are two vastly different things

—the work of evaluating integrals becomes the work of finding anti-derivatives.

For polynomial functions, which are the sums of natural number powers of the identity function $I()$, multiplied by constants, we will find all that we need to use by proving:

Theorem 7.10:

If n is a natural number, then

$\left\{ \frac{1}{n+1} I^{n+1} \right\} ()$ is an anti-derivative of $\{I^n\} ()$.

Proof:

$$\begin{aligned} \left\{ \frac{1}{n+1} I^{n+1} \right\}' () &= \left\{ \frac{1}{n+1} (n+1) I^n \right\} () \\ &= \{I^n\} (). \end{aligned}$$

This theorem, together with the theorems on the integral of the sum of functions and on the integral of a constant multiple of a function, enables us to use the Fundamental Theorem in the evaluation of the integral of any polynomial function over any interval of numbers. For example,

$$\begin{aligned} &\int_2^3 [5t^3 - 2t^2 + 7t - 4] dt \\ &= \int_2^3 (5u^3) du + \int_2^3 (-2v^2) dv + \int_2^3 (7w) dw + \int_2^3 (-4) dx \\ &= 5 \int_2^3 y^3 dy - 2 \int_2^3 z^2 dz + 7 \int_2^3 r dr - 4 \int_2^3 ds. \end{aligned}$$

An anti-derivative of the function given by y^3 is $\left\{ \frac{1}{4} I^4 \right\} ()$, so

$$\int_2^3 y^3 dy = \frac{1}{4} (3)^4 - \frac{1}{4} (2)^4.$$

An anti-derivative of the function given by z^2 is $\left\{ \frac{1}{3} I^3 \right\} ()$, so

$$\int_2^3 z^2 dz = \frac{1}{3} (3)^3 - \frac{1}{3} (2)^3.$$

An anti-derivative of the function given by r is $\left\{ \frac{1}{2} I^2 \right\} ()$, so

$$\int_2^3 r dr = \frac{1}{2} (3)^2 - \frac{1}{2} (2)^2.$$

The last integral is the integral of a constant function, and Theorem 7.6 tells us

$$\int_2^3 ds = 1.$$

Combining these results, we have

$$\begin{aligned} & \int_2^3 [5t^3 - 2t^2 - 7t - 4] dt \\ &= 5 \left(\frac{1}{4}\right) (81 - 16) - 2 \left(\frac{1}{3}\right) (27 - 8) + 7 \left(\frac{1}{2}\right) (9 - 4) - 4(1) \\ &= 685/12. \end{aligned}$$

Since $2 \leq t \leq 3$ implies that

$$5t^3 - 2t^2 + 7t - 4 \geq 0,$$

the area of that part of the plane bounded by the graphs of

$$y = 5x^3 - 2x^2 + 7x - 4, y = 0, x = 2, \text{ and } x = 3$$

will be 685/12.

We will adopt a notational device which enables us to write down the steps in the evaluation of an integral by using the Fundamental Theorem. If $f()$ is being integrated over the interval:

the set of all s , $a \leq s \leq b$,

and $F()$ is an anti-derivative of $f()$, we will write

$$F(t) \Big|_{t=a}^{t=b}$$

as shorthand for $F(b) - F(a)$.

For the example shown, this shorthand gives

$$\begin{aligned} & 5 \int_2^3 y^3 dy - 2 \int_2^3 z^2 dz + 7 \int_2^3 r dr - 4 \int_2^3 ds \\ &= 5 \left[\frac{y^4}{4} \right]_{y=2}^{y=3} - 2 \left[\frac{z^3}{3} \right]_{z=2}^{z=3} + 7 \left[\frac{r^2}{2} \right]_{r=2}^{r=3} - 4 \left[s \right]_{s=2}^{s=3} \\ &= \frac{5}{4} [(3)^4 - (2)^4] - \frac{2}{3} [(3)^3 - (2)^3] + \frac{7}{2} [(3)^2 - (2)^2] - 4 [3 - 2] \\ &= \frac{325}{4} - \frac{38}{3} + \frac{35}{2} - 4 \\ &= 685/12. \end{aligned}$$

Exercises

Evaluate the following:

1. $\int_0^3 [r^3 - 9r] dr.$

4. $\int_0^4 u^3 du.$

2. $\int_{-3}^0 [s^3 - 9s] ds.$

5. $\int_{-4}^0 v^3 dv.$

3. $\int_{-2}^3 [t^3 - 9t] dt.$

6. $\int_{-4}^4 w^3 dw.$

Find the area of the portion of the plane bounded by the graphs of:

7. $y = x^2 + x, y = 0, x = 2.$

8. $y = x^3 + 2x, y = 0, x = 3.$

9. $y = x^3 + 2x, y = -1, x = 0, x = 3.$

10. $y = x^3, x = -1, x = 1, y = 1.$

11. $y = x^3 - 3x, y = 0.$

Evaluate:

12. $\int_1^2 [y + 1] dy.$

13. $\int_2^3 [3z^2 - 4z + 1] dz.$

14. $\int_{-1}^1 [2r^3 - r^2 + 5r - 3] dr.$

15. $\int_{-2}^1 [5s^4 + 3s^2 - 2] ds.$

16. $\int_{-1}^3 [t^4 - 3t^3 + 6t^2 + 8t - 1] dt.$

17. $\int_{-2}^0 [6u^5 + u^3 - 7u] du.$

18. $\int_4^5 [v^5 - 16v^2] dv.$

19. $\int_{1/2}^1 [w^3 + 7w - 1] dw.$

20. $\int_{-1/2}^{1/2} [x^2 - 7x + 2] dx.$

21. $\int_0^3 [y^3 + 15y] dy.$

22. $\int_2^3 [z^2 - 1] dz.$

ALGEBRAIC FUNCTIONS

8.1. INTEGER EXPONENTS

A symbol such as b^a , where a and b are numbers, has been met before only in the case that the number a is a positive integer. Thus, in order that this symbol have meaning where the number a is anything other than a positive integer, it is necessary that we give a formal definition, which definition will depend entirely on what kind of a number the “exponent” a actually is. In this section, we restrict ourselves to integer exponents, recalling that the integers consist of 0, the natural numbers, and the negatives of the natural numbers.

Definition 8.1: If a is a positive integer and b is any number, then b^a is the product of a factors, each one of which is b .

It should be noted that the phrase “ a factors” or “ a times” is meaningless if a is anything other than a positive integer. That is, one can do something “3 times,” or “7 times,” or “1,813,204 times,” but one cannot do something “– 4 times” or “– 10 times,” or even “0 times.”

Definition 8.2: If q is a negative integer and b is any number other than 0, then b^q is the number

$$\frac{1}{b^{-q}}.$$

Here one should note that q 's being a negative integer implies that $-q$ is a positive integer, so that b^q is the reciprocal of the product: b multiplied by itself $-q$ times.

For example

$$(2)^{-3} = \frac{1}{2^3} = \frac{1}{8};$$

$$\left(\frac{1}{3}\right)^{-2} = \frac{1}{\left(\frac{1}{3}\right)^2} = \frac{1}{\frac{1}{9}} = 9.$$

Definition 8.3: If b is any number, $b^0 = 1$.

This third definition does not say that the product of the number b multiplied by itself "0 times" is the number 1, because there is no meaning to the phrase "0 times."

The algebraic laws for multiplication and division then give us the following rules for manipulation with integer exponents:

Whenever p and q are integers,

$$b^p b^q = b^{p+q};$$

$$\frac{b^p}{b^q} = b^{p-q};$$

$$(b^p)^q = b^{pq};$$

for every number b , except where a division by 0 is to be avoided. We have constructed the definitions (8.2) and (8.3) so that these familiar rules for exponents which are positive integers remain true if we allow any integer exponents.

Exercises

Find the number indicated by:

1. 3^{-2} , 2^{-3} , $(-2)^3$, $(-3)^{-2}$.

2. $\frac{4^{-1}}{2^{-3}}$, $\frac{3^4}{7^0}$, $\frac{2^4}{5^{-1}}$, $\frac{5^{-2}}{5^{-5}}$.

3. $(2^3)(2^{-2})$, $(3^{-2})(3^{-1})$, $(2^{-2})(2^3)$, $(4^{-1})(4^4)$.

4. $((-2)^2)^{-2}$, $((3^{-1})^3)^{-2}$, $(4^5)^0$, $(5^{-1})^2$.

5. $(2^3)^2$, 2^{3^2} .

Solve the equations:

6. $2^p = 1/16$.

7. $3^p = 27$.

8. $4^p = 1/64$.

9. $5^p = 1$.

10. $(-2)^p = 16$.

8.2. THE QUOTIENT OF TWO FUNCTIONS

Since negative integer exponents indicate division, if we specify a function by an algebraic expression involving one variable, in which expression the variable has a negative integer exponent, then we are led to the idea of dividing functions. We have the obvious definition:

Definition 8.4:

If $f()$ and $g()$ are functions, then the *quotient* $\left\{\frac{f}{g}\right\}()$, of $f()$ divided by $g()$ is defined by:

a) The domain of $\left\{\frac{f}{g}\right\}()$ is the set of all numbers which are simultaneously in the domains of $f()$ and $g()$, except for those numbers in the domain of $g()$ which are solutions of the equation $g(x) = 0$;

b) a rule of $\left\{\frac{f}{g}\right\}()$ is given by

$$\left\{\frac{f}{g}\right\}(x) = \frac{f(x)}{g(x)}.$$

As a result of this definition, the quotient of two functions, each of which is *the* function specified by some algebraic expression, will be the function specified by the quotient of the algebraic expressions. Because of the prohibition on dividing by 0, even in the simple case in which both the "numerator function" and the "denominator function" are polynomials, the domain of the quotient is not necessarily the set of all numbers. For example, if $f()$ is the function specified by

$$f(r) = 3r^2 - 5r + 4$$

and $g()$ is the function specified by

$$g(s) = s^2 - 3s + 2,$$

then $\left\{\frac{f}{g}\right\}()$ will be the function specified by

$$\left\{\frac{f}{g}\right\}(t) = \frac{3t^2 - 5t + 4}{t^2 - 3t + 2},$$

whose domain is the set of all numbers other than 1 and 2.

In what follows, it will be necessary for us to find a limit of the form

$$\lim_{h \rightarrow 0} \frac{N(h)}{D(h)},$$

where $N()$ and $D()$ are functions, such that $\lim_{h \rightarrow 0} N(h)$ exists, and $\lim_{h \rightarrow 0} D(h)$ exists, and is not 0. In general, we have

Theorem 8.1:

If $f()$ and $g()$ are functions;

if $\lim_{t \rightarrow a} f(t)$ exists, and is the number A ;

if $\lim_{t \rightarrow a} g(t)$ exists, and is the number B ; and if $B \neq 0$,

then $\lim_{t \rightarrow a} \left\{\frac{f}{g}\right\}(t)$ exists and is the number A/B .

(A more concise, but less precise statement is given by: The limit of the quotient of two functions is the quotient of their limits, provided we do not divide by 0.)

Proof of 8.1:

1. Pick $\epsilon > 0$.

2. $B \neq 0$
so $|B| > 0$

by hypothesis,
by the definition of absolute value.

3. There exists an interval:

the set of all t ,

$$|t - a| < d_1 \text{ (where } d_1 > 0),$$

such that, if t is in this interval,
then t is in the domain of $g()$ and

$$|g(t) - B| < \frac{1}{2} |B|$$

by steps 1 and 2 and the definition
of limit.

4. If $0 < |t - a| < d_1$, then

$$|g(t)| > |B| - \frac{1}{2} |B|$$

$$= \frac{1}{2} |B| > 0$$

by steps 2 and 3.

5. If $0 < |t - a| < d_1$,

$$\text{then } \frac{1}{|g(t)|} < \frac{2}{|B|}$$

by step 4 and the laws governing
inequalities.

6. $\lim_{t \rightarrow a} f(t) = A$

by hypothesis.

7. There exists an interval:

the set of all t ,

$$0 < |t - a| < d_2$$

(where $d_2 > 0$), such that, if t is
in the interval,

then t is in the domain of $f()$

$$\text{and } |f(t) - A| < \frac{|B|}{4} \epsilon$$

by steps 2 and 6 and the definition
of limit.

8. There exists an interval:

the set of all t ,

$$0 < |t - a| < d_3$$

(where $d_3 > 0$), such that, if t is
in this interval, then t is in the
domain of $g()$

$$\text{and } |g(t) - B| < \frac{|B|^2}{4|A|} \epsilon$$

by steps 1 and 2 and the definition
of limit.

9. Let d be the smallest of the three numbers d_1 , d_2 , d_3 .

10. If $0 < |t - a| < d$, then t is

in the domain of $\left\{ \frac{f}{g} \right\}()$;

$$\frac{1}{|g(t)|} < \frac{2}{|B|};$$

$$|f(t) - A| < \frac{|B|}{4} \epsilon;$$

$$\text{and } |g(t) - B| < \frac{|B|^2}{4|A|} \epsilon \quad \text{by steps 5, 7, and 8, and the definition of the quotient of two functions.}$$

11. If t is in the domain of $\left\{\frac{f}{g}\right\}()$,
then

$$\begin{aligned} \left\{\frac{f}{g}\right\}(t) - \frac{A}{B} &= \frac{f(t)}{g(t)} - \frac{A}{B} \\ &= \frac{Bf(t) - Ag(t)}{Bg(t)} \\ &= \frac{Bf(t) - BA + AB - Ag(t)}{Bg(t)} \\ &= \frac{1}{g(t)} ([f(t) - A] + \frac{A}{B} [B - g(t)]) \end{aligned}$$

by the definition of the quotient of two functions and various algebraic laws.

12. If t is in the domain of $\left\{\frac{f}{g}\right\}()$,
then

$$\left| \left\{\frac{f}{g}\right\}(t) - \frac{A}{B} \right| \leq \frac{1}{|g(t)|} \left(|f(t) - A| + \left| \frac{A}{B} \right| |g(t) - B| \right)$$

by step 11 and the theorems on the absolute values of the sum and product of numbers.

13. If $0 < |t - a| < d$, then t is in
the domain of $\left\{\frac{f}{g}\right\}()$ and

$$\left| \left\{\frac{f}{g}\right\}(t) - \frac{A}{B} \right| \leq \frac{2}{|B|} \left(\frac{|B|}{4} \epsilon + \left| \frac{A}{B} \right| \frac{|B|^2}{4|A|} \epsilon \right)$$

by steps 10 and 12.

14. If $0 < |t - a| < d$, then t is in
the domain of $\left\{\frac{f}{g}\right\}()$ and

$$\left| \left\{\frac{f}{g}\right\}(t) - \frac{A}{B} \right| < \epsilon$$

by step 13 and various algebraic laws

15. $\lim_{t \rightarrow a} \left\{\frac{f}{g}\right\}(t) = \frac{A}{B}$

by steps 1, 9, 14, and the definition of limit.

Exercises

1. Find an interval, containing the number 2 in its interior, such that, for every number t in this interval, $g(t) > 2$, where $g()$ is specified by

a) $g(r) = 3r + 5$.

b) $g(s) = s^2$.

2. If $f()$ is specified by

$$f(t) = t + 1,$$

how close to 2 must t be in order that

$$|f(t) - 3| < e$$

be true?

3. If t is a number in the interval found in exercise 2, find an upper bound to

a) $\left| \left\{ \frac{f}{g} \right\} (t) - \frac{3}{8} \right|,$

where $f()$ is as in exercise 2 and $g()$ is specified by $g(u) = 3u + 5$.

b) $\left| \left\{ \frac{f}{g} \right\} (t) - \frac{3}{4} \right|,$

where $g()$ is specified by $g(v) = v^2$.

Find the following limits:

4. $\lim_{w \rightarrow 1} \frac{3w - 4}{w^2 + 2w - 6}.$

5. $\lim_{x \rightarrow 2} \frac{x^3 - 4x^2 + 7x - 1}{3x^2 - 16}.$

6. $\lim_{y \rightarrow 0} \frac{y^3 + 2y - 4}{y^2 + y - 1}.$

7. $\lim_{z \rightarrow -1} (z + 2)(z + 1)^{-1}.$

8. $\lim_{t \rightarrow 1} (t + 2)(t + 1)^{-1}.$

9. $\lim_{u \rightarrow 2} (u^2 - 3u + 1 - 4u^{-1} + u^{-2}).$

10. $\lim_{v \rightarrow 3} \frac{v^2 - 9}{v^3 - 27}.$

8.3. DERIVATIVE AND ANTI-DERIVATIVE FOR $\{I^p\}()$, p A NEGATIVE INTEGER

For the function $\{I^p\}()$, where p is an integer, we have found the derivative in the case that p is a positive integer in Theorem 5.10, and in the case that $p = 0$, which case gives the constant function $U()$, in Theorem 5.8. In this section, we will find the derivative of $\{I^p\}()$, where p is a negative integer.

Theorem 8.2:

If p is a negative integer, then

$$\{I^p\}'() = \{p I^{p-1}\}().$$

Proof:

1. p is a negative integer by hypothesis.
2. Let $k = -p$.
3. k is a positive integer by steps 2 and 3.
4. $\{I^p\}(t) = \frac{1}{t^k}$ by step 2 and the definition (8.2).
5. $\{I^p\}(x+h) - \{I^p\}(x) = \frac{1}{(x+h)^k} - \frac{1}{x^k}$, whenever $x \neq 0$, and for every value of the variable h except $h = -x$.
6. $\{I^p\}(x+h) - \{I^p\}(x) = \frac{x^k - (x+h)^k}{x^k(x+h)^k}$ by step 5 and various algebraic laws.
7. $\frac{\{I^p\}(x+h) - \{I^p\}(x)}{h} = (-1) \frac{(x+h)^k - x^k}{h} \frac{1}{x^k(x+h)^k}$ by step 6 and various algebraic laws.
8. $\lim_{h \rightarrow 0} -x^k(x+h)^k$ exists and is the number $-x^{2k}$ by the fact that $-x^k(x+h)^k$, $h \neq -x$, specifies a continuous function.
9. $\lim_{h \rightarrow 0} \frac{(x+h)^k - x^k}{h}$ exists and is the number kx^{k-1} since this limit gives the derivative of $\{I^k\}()$, and k is a positive integer by step 3.
10. $\lim_{h \rightarrow 0} \left[(-1) \frac{(x+h)^k - x^k}{h} \frac{1}{x^k(x+h)^k} \right]$ exists and is the number $\frac{kx^{k-1}}{-x^{2k}}$ by steps 8 and 9 and Theorem 8.1.
11. $\{I^p\}'(x)$ is the number $\frac{-kx^{k-1}}{x^{2k}}$ by steps 7 and 10 and the definition of derivative.
12. $\{I^p\}'(x) = -k x^{-k-1}$ by step 11 and the algebraic laws for integer exponents.
13. $\{I^p\}'(x) = p x^{p-1}$ by steps 12 and 2.
14. $\{I^p\}'() = \{p I^{p-1}\}()$ by steps 5 and 13.

Corollary to Theorem 8.2:

If p is any integer,

$$\{I^p\}'() = \{p I^{p-1}\}().$$

Proof:

Theorems 5.8, 5.10, and 8.2.

For anti-derivatives, we have:

Theorem 8.3:

If p is an integer other than -1 , then

$$\left\{ \frac{1}{p+1} I^{p+1} \right\}()$$

is an anti-derivative of $\{I^p\}()$.

Proof:

The Corollary to Theorem 8.2.

Exercises

Find the derivative of

1. $\{I^3 - 2I^2 + 3 - 4I^{-2} - 5I^{-4}\}()$.
2. $\{(I^2 + 3I) I^{-2}\}()$.
3. $\{\frac{1}{2}I^3 - I^2 - \frac{1}{3}I + \frac{3}{7} + \frac{1}{4}I^{-1} - 2I^{-2}\}()$.
4. $\{(I^{-2} - \frac{1}{6}I^{-1})(\frac{1}{2}I^3)\}()$.
5. $\{\sqrt{2}I^{-4} + \sqrt{5}I^{-1} - I\}()$.
6. $\left\{ \frac{I^4 + 2I^3 - 3I^2 + 7I - 4}{5I^6} \right\}()$.

Find an anti-derivative of the function specified by:

7. $r^2 - 3r^{-4} + 2r^3$.
8. $s^{-4} + 2s^{-3} - \frac{1}{3}s^{-2} + 1$.
9. $-t + \sqrt{3}t^{-5} + \sqrt{5}t^2 - \frac{1}{2}t^{-8}$.
10. $(u^2 - 3u + 1)(2u^{-4})$.
11. $(v^3 + v^{-1})^2$.

Find the area of that portion of the plane bounded by the graphs of:

12. $y = x^{-2}$, $x = 1$, $x = 2$, $y = 0$.
13. $y = x^{-3}$, $y = x$, $x = 2$.

Sketch the graphs of the following equations, indicating relative extreme points and points of inflection:

14. $y = x + x^{-1}$.

15. $y = x + x^{-2}$.

16. $y = x^2 + x^{-2}$.

17. Find an expression specifying the n -th derivative of the function specified by $f(t) = t^{-1}$, where n is a natural number.

8.4. THE DERIVATIVE OF A QUOTIENT

In order to differentiate the quotient of two functions we have:

Theorem 8.4:

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

Proof:

1. Let a be a number in the domain of $f'(\cdot)$, and in the domain of $g'(\cdot)$, for which $g(a) \neq 0$.

2. a is in the domain of

$$\left\{ \frac{gf' - fg'}{g^2} \right\}(\cdot)$$

by step 1 and the definitions of the product, difference, and quotient of functions.

$$3. \left\{ \frac{f}{g} \right\}'(a) = \lim_{h \rightarrow 0} \frac{\left\{ \frac{f}{g} \right\}(a+h) - \left\{ \frac{f}{g} \right\}(a)}{h}$$

by the definition of derivative.

4. For any value of the variable h , other than 0 and those values for which $g(a+h) \neq 0$,

$$\frac{\left\{ \frac{f}{g} \right\}(a+h) - \left\{ \frac{f}{g} \right\}(a)}{h} = \frac{1}{h} \left[\frac{f(a+h)}{g(a+h)} - \frac{f(a)}{g(a)} \right]$$

by the definition of the quotient of two functions.

5. For any value of the variable h , other than 0 and those values of h for which $g(a+h) \neq 0$,

$$\frac{\left\{ \frac{f}{g} \right\}(a+h) - \left\{ \frac{f}{g} \right\}(a)}{h} = \frac{1}{h} \left[\frac{f(a+h)g(a) - f(a)g(a+h)}{g(a+h)g(a)} \right]$$

by step 4 and algebraic laws.

6. For any value of the variable h , other than 0 and those values of h for which $g(a + h) \neq 0$,

$$\frac{\left\{\frac{f}{g}\right\}(a + h) - \left\{\frac{f}{g}\right\}(a)}{h} = \frac{1}{h} \left[\frac{f(a + h)g(a) - f(a)g(a) + f(a)g(a) - f(a)g(a + h)}{g(a + h)g(a)} \right]$$

by step 5 and the addition of 0 in convenient form.

7. For any value of the variable h , other than 0 and those values for which $g(a + h) \neq 0$,

$$\frac{\left\{\frac{f}{g}\right\}(a + h) - \left\{\frac{f}{g}\right\}(a)}{h} = \left[g(a) \frac{f(a + h) - f(a)}{h} - f(a) \frac{g(a + h) - g(a)}{h} \right] \frac{1}{g(a + h)g(a)}$$

by step 6 and various algebraic laws.

8. $g()$ is continuous at a

by step 1 and the theorem (6.1) that differentiability implies continuity.

9. $\lim_{h \rightarrow 0} g(a + h)$ exists and is the number $g(a)$

by step 8 and the definition of continuity.

10. $\lim_{h \rightarrow 0} [g(a + h)g(a)]$ exists and is the number $[g(a)]^2$

by step 9 and the theorem (5.4) on the limit of a constant multiple of a function.

11. $\lim_{h \rightarrow 0} \frac{f(a + h) - f(a)}{h}$ exists and is the number $f'(a)$

by step 1 and the definition of derivative.

12. $\lim_{h \rightarrow 0} \left[g(a) \frac{f(a + h) - f(a)}{h} \right]$ exists and is the number $g(a)f'(a)$

by step 11 and the theorem (5.4) on the limit of a constant multiple of a function.

13. $\lim_{h \rightarrow 0} \frac{g(a + h) - g(a)}{h}$ exists and is the number $g'(a)$

by step 2 and the definition of derivative.

$$14. \lim_{h \rightarrow 0} \left[f(a) \frac{g(a+h) - f(a)}{h} \right] \text{ exists}$$

and is the number $f(a) g'(a)$

by step 13 and the theorem (5.4) on the limit of a constant multiple of a function.

$$15. \lim_{h \rightarrow 0} \left[g(a) \frac{f(a+h) - f(a)}{h} - f(a) \frac{g(a+h) - g(a)}{h} \right]$$

exists and is the number

$$g(a) f'(a) - f(a) g'(a)$$

by steps 12 and 14 and the theorems (5.4 and 5.5) on the limit of a constant multiple of a function and on the limit of the sum of functions.

$$16. \left\{ \frac{f}{g} \right\}'(a) \text{ exists and is the number}$$

$$\frac{g(a) f'(a) - f(a) g'(a)}{[g(a)]^2}$$

by steps 15, 10, 3, and 7, and the theorem (8.1) on the limit of the quotient of functions.

$$17. \left\{ \frac{f}{g} \right\}'() = \left\{ \frac{gf' - fg'}{g^2} \right\}()$$

by step 1 and the definitions of the product, difference, and quotient of functions.

As for anti-derivatives, life is simple: there is no anti-derivative theorem of any general significance for the quotient of two functions.

Exercises

Differentiate the functions given by:

$$1. f(r) = \frac{r+1}{r+2}.$$

$$2. f(s) = \frac{s^2 + 3s - 1}{2s - 3}.$$

$$3. g(t) = \frac{3t^3 - 3}{2t + 5}.$$

$$4. g(u) = \frac{u^3 - 2u^2 + 5u - 1}{u^2 + 4u - 7}.$$

$$5. h(v) = \frac{v^{-1} + v^2}{2v - 5}.$$

$$6. h(w) = \frac{w^{-2} + w^{-3}}{w^{-1} + w^{-4}}.$$

$$7. F(x) = (x^3 - 3x^2 + 1)(3x + 5)^{-1}.$$

$$8. F(y) = (y + 3)^2 (2y - 1)^{-2}.$$

9. $G(z) = \frac{\sqrt{3} z^{-1} + \frac{1}{2} z^{-3}}{z^5 - \sqrt{2} z^4}.$

10. Find any relative maximum or relative minimum points on the graph of the equation

$$y = \frac{2x + 5}{x^2 - 4}.$$

8.5. DERIVATIVE OF THE PRODUCT OF TWO FUNCTIONS

Now that we have proved theorems concerning the derivative of the sum, difference, and quotient of two functions, we will tie up the loose end of the differentiation of the product of two functions:

Theorem 8.5:

$$\{fg\}'() = \{fg' + f'g\}().$$

Proof:

1. Let a be a number in the domain of $f'()$ and in the domain of $g'()$.

2. a is in the domain of

$$\{fg' + f'g\}()$$

by step 1 and the definitions of derivative and of the product and sum of functions.

$$3. \{fg\}'(a) = \lim_{h \rightarrow 0} \frac{\{fg\}(a+h) - \{fg\}(a)}{h}$$

by the definition of derivative.

4. For any value of h , other than 0, for which $(a+h)$ is in the domain of $\{fg\}()$,

$$\frac{\{fg\}(a+h) - \{fg\}(a)}{h} = \frac{f(a+h)g(a+h) - f(a)g(a)}{h}$$

by the definition of the product of two functions.

5. For any value of h , other than 0, for which $(a+h)$ is in the domain of $\{fg\}()$,

$$\begin{aligned} & \frac{\{fg\}(a+h) - \{fg\}(a)}{h} \\ &= \frac{f(a+h)g(a+h) - f(a)g(a+h) + f(a)g(a+h) - f(a)g(a)}{h} \end{aligned}$$

by step 4 and the addition of 0 in convenient form.

6. For any value of h , other than 0,
for which $(a + h)$ is in the
domain of $\{fg\}()$,

$$\frac{\{fg\}(a + h) - \{fg\}(a)}{h} = g(a + h) \frac{f(a + h) - f(a)}{h} + f(a) \frac{g(a + h) - g(a)}{h}$$
by step 5 and various algebraic laws.
7. $\lim_{h \rightarrow 0} \frac{f(a + h) - f(a)}{h}$ exists and is
the number $f'(a)$
by step 1 and the definition of
derivative.
8. $g()$ is continuous at a
by step 1 and the theorem (6.1)
that differentiability implies con-
tinuity.
9. $\lim_{h \rightarrow 0} g(a + h)$ exists and is the
the number $g(a)$
by step 8 and the definition of con-
tinuity.
10. $\lim_{h \rightarrow 0} \left[g(a + h) \frac{f(a + h) - f(a)}{h} \right]$
exists and is the number $g(a) f'(a)$
by steps 7 and 9 and the theorem
on the limit of the product of
functions.
11. $\lim_{h \rightarrow 0} \frac{g(a + h) - g(a)}{h}$ exists and is
the number $g'(a)$
by step 1 and the definition of
derivative.
12. $\lim_{h \rightarrow 0} f(a) \frac{g(a + h) - g(a)}{h}$ exists
and is the number $f(a) g'(a)$
by step 11 and the theorem on the
limit of a constant multiple of a
function.
13. $\{fg\}'(a)$ exists and is the number
 $f(a) g'(a) + f'(a) g(a)$
by steps 3, 6, 10, and 12 and the
theorem on the limit of the sum of
functions.
14. $\{fg\}'() = \{fg' + f'g\}()$
by steps 1 and 13 and the definition
of the product and sum of functions.

This theorem provides us with an alternate method of differentiating various polynomials, and will later on provide us with a method of differentiating functions other than polynomials.

As an example, consider the function

$$\left\{ \frac{I^3 + I^2}{I^6} \right\} (I).$$

Considered as the quotient of $\{I^3 + I^2\} (I)$ by $\{I^6\} (I)$, we have

$$\begin{aligned} \left\{ \frac{I^3 + I^2}{I^6} \right\}' (I) &= \left\{ \frac{I^6 (3I^2 + 2I) - (I^3 + I^2) 5I^4}{I^{10}} \right\} (I) \\ &= \left\{ \frac{3I^7 + 2I^6 - 5I^7 - 5I^6}{I^{10}} \right\} (I) \\ &= \left\{ \frac{-2I^7 - 3I^6}{I^{10}} \right\} (I) \\ &= \{-2I^{-3} - 3I^{-4}\} (I). \end{aligned}$$

Considered as the product of $\{I^3 + I^2\} (I)$ and $\{I^{-6}\} (I)$, we have

$$\begin{aligned} \{(I^3 + I^2) (I^{-6})\}' (I) &= \{(I^3 + I^2) (-5I^{-6}) + (3I^2 + 2I) (I^{-6})\} (I) \\ &= (-5I^{-3} - 5I^{-4} + 3I^{-3} + 2I^{-4}) (I) \\ &= \{-2I^{-3} - 3I^{-4}\} (I). \end{aligned}$$

Exercises

Find the derivative of the functions specified by:

1. $(2r + 3) (5r - 1)$.
2. $(s^2 - 3s) (2s + 1)$.
3. $(t^3 + 5t^2 - 7t + 1) t$.
4. $(u^{-3} + 2u^3) (3u^2 - u^{-4})$.
5. $\frac{(2v - 3) (4v + 5)}{6v - 7}$.

Find all relative maximum and relative minimum values of the functions specified by

6. $f(w) = (3w - 1) (w + 2)$.
7. $f(x) = (x - 3) (x^2 + 2x + 1)$.
8. $f(y) = y(y^2 - 12)$.
9. $f(z) = (z^2 - 6) (2z + 3)$.
10. Use the theorem about the derivative of the product of two functions to derive directly the fact that

$$\{f^n\}' (I) = \{n f^{n-1} f'\} (I),$$

where n is a natural number.

8.6. RATIONAL FUNCTIONS

In the discussion of the various kinds of numbers contained in Chapter 1, rational numbers were defined to be those numbers which could be expressed as the quotient of two integers. Integers are the simplest numbers with which to do arithmetic, and polynomial functions are the easiest functions with which to do calculus—that is, to find the derivative, anti-derivative, and integral. Just as we called the numbers next in complication to integers rational numbers, so will we call the functions next in complication to polynomial functions rational functions:

Definition 8.5: A *rational function* is a function which may be expressed as quotient of two polynomial functions.

Thus, we are concerning ourselves with functions specified by expressions of the form

$$\frac{N(x)}{D(x)},$$

where $N()$ and $D()$ are polynomial functions. Since the domain of a polynomial function is the set of all numbers, the definition of the quotient of two functions tells us that the domain of a rational function consists of all numbers except the zeros of the denominator polynomial—in the illustration, those values of the variable x for which $D(x)$ is the number 0.

The arithmetic process known as long division is actually a theorem about natural numbers and rational numbers which says that if n and d are natural numbers, then there is a non-negative integer q , and a natural number r , with $1 \leq r < d$, such that

$$\frac{n}{d} = q + \frac{r}{d}.$$

This theorem is usually first seen in a form which says that the result of dividing the natural number n by the natural number d is a “quotient” of q with a “remainder” of r . The analogous theorem for rational functions is:

Theorem 8.6:

If $N()$ and $D()$ are polynomial functions, then there exist polynomial functions, $Q()$ and $R()$ (either of which may be constant), such that

$$\left\{ \frac{N}{D} \right\} () = \left\{ Q + \frac{R}{D} \right\} (),$$

and the degree of $R()$ is less than the degree of $D()$.

For example,

$$\left\{ \frac{I^4 + I^3 - 2I^2 + 3I - 7}{I^2 + 3I - 2} \right\} () = \left\{ I^2 - 2I + 6 + \left(\frac{-19I + 19}{I^2 + 3I - 2} \right) \right\} ().$$

The proof of this theorem consists of establishing the validity of the process of long division for polynomials. It is not necessary for the purposes of calculus for one to go through the details of this proof. It is necessary for the purposes of calculus for one to know how to use this theorem.

Exercises

Perform the indicated divisions:

$$1. \frac{2x^5 - 3x^4 + 7x^3 - x^2 + 4x - 1}{x^3 + 2x^2 - 3x + 4}.$$

$$2. \frac{x^5 - 3x^4 + 7x^3 + x - 1}{x^2 + x - 1}.$$

$$3. \frac{7x^6 - 5x^5 + x^3 - 3}{x + 1}.$$

4. If a polynomial of degree 6 is divided by a polynomial of degree 4, what can be said of the degree of the remainder polynomial?
5. If $N()$ is a polynomial of degree n , and $D()$ is a polynomial of degree m , and

$$\left\{ \frac{N}{D} \right\} () = \left\{ Q + \frac{R}{D} \right\} (),$$

what can be said of the degrees of $Q()$ and $R()$ in the cases: $m > n$, $m = n$, $m < n$.

$$6. \text{ a) Divide } \frac{2y^4 - 4y^3 + 7y^2 - 8y + 1}{y - 2}.$$

b) If $f()$ is the function given by the expression in the numerator of part a), find $f(2)$.

7. Prove the following theorem: If $P()$ is a polynomial function, and a is a number, and

$$\frac{P(t)}{t - a} = Q(t) + \frac{R}{t - a},$$

then R and $F(a)$ are the same number.

8. Suppose that $F()$ and $G()$ are polynomial functions, and that a is a number such that $F(a) = G(a) = 0$. If $F()$ and $G()$ are of the same degree, show that

$$\lim_{t \rightarrow a} \frac{F(t)}{G(t)}$$

exists and is not 0 only in the case that $F(t)$ and $G(t)$ contain the same number of first degree factors of the form $(t - a)$.

8.7. "INFINITE" LIMITS

The domain of a rational function must exclude those numbers which give the denominator polynomial a zero value. Consequently, the graph of such a rational function must not contain any points whose x -coordinates are the numbers excluded from the domain. We are here faced with

a situation in which an attempted division by 0 leads to interesting results, as opposed to the situation encountered with derivatives, where an apparent division by zero in

$$\lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}$$

led to a number.

If we examine the graph of the simple rational function specified by the equation

$$y = \frac{1}{x^2},$$

we note that 0 is not in the domain of this function, so that no point on the graph of this equation will have an x -coordinate of 0. Geometrically, this last statement means that the y -axis never crosses or touches the graph of $y = \frac{1}{x^2}$. If we were to plot points on the graph of this rational function, the x -coordinates of which points were numbers "near" 0, we would find that the y -coordinates of these points were "large," and that the closer to 0 we took the x -coordinates, the larger the y -coordinates would be. That is, however large we might want the y -coordinates to be, we could find points on this graph whose y -coordinates were at least that large, merely by taking the x -coordinates of those points sufficiently close to 0. Since we can do this, then there is a closeness to 0 which is sufficiently close to accomplish our desires. We will abbreviate this situation in:

Definition 8.6: If $f()$ is a function and a is a number,

$$\lim_{t \rightarrow a} f(t) = +\infty$$

means:

For every $B > 0$, there exists an interval:

the set of all t , $|t - a| < d$ (where $d > 0$),

such that,

$$\text{if } 0 < |t - a| < d,$$

then t is in the domain of $f()$ and $f(t) > B$.

(This notation is read: "As t approximates a , f of t becomes infinite.")

It should be most strongly emphasized that we are NOT defining the symbol " $+\infty$ " by itself, but are defining the entire shorthand sentence " $\lim_{t \rightarrow a} f(t) = +\infty$." Roughly speaking, the definition says that, if t is a number sufficiently close to a , then t will be in the domain of $f()$ and the corresponding value $f(t)$ will be large.

Similarly we have:

Definiton 8.7: If $f()$ is a function and b is a number,

$$\lim_{t \rightarrow a} f(t) = -\infty$$

means:

For every $N < 0$, there exists an interval:

the set of all t , $|t - a| < d$ (where $d > 0$),

such that,

$$\text{if } 0 < |t - a| < d,$$

then t is in the domain of $f()$ and $f(t) < N$.

Roughly speaking, the definition (8.7) describes the situation in which values of t close to a yield values of $f(t)$ which are very negative.

One can easily see that

$$\lim_{t \rightarrow 0} \frac{1}{t^2} = +\infty.$$

$$\lim_{t \rightarrow 0} -\frac{1}{t^2} = -\infty.$$

If we examine the graph of the very simple rational function $\{\frac{1}{t}\}()$, we see that points whose x -coordinates are near 0, and positive, have very large positive y -coordinates, while points with x -coordinates near 0, but negative, have very large negative y -coordinates. Then, we can say that x -coordinates near 0 yield y -coordinates which are very far from 0, some of which y -coordinates are negative and some positive. Combining our newly introduced notation with the absolute value symbol, we can see that

$$\lim_{x \rightarrow 0} \left| \frac{1}{x} \right| = +\infty.$$

Once more, it should be noted that we have not defined the symbols “ $+\infty$ ” and “ $-\infty$ ” have themselves, nor have we defined the word “infinity,” but have defined entire sentences written in a shorthand which includes these symbols. Just as the symbol ‘, used to indicate differentiation, or the symbol $\int$ used to indicate integration, have no meaning by themselves, so the symbols “ $+\infty$ ” and “ $-\infty$ ” have no meaning by themselves. What we have done is to produce a shorthand notation for saying that the graph of a function rises or falls very rapidly as we come close to a fixed vertical line.

Geometrically, if $\lim_{t \rightarrow a} f(t) = +\infty$, or $\lim_{t \rightarrow a} f(t) = -\infty$, or $\lim_{t \rightarrow a} |f(t)| = +\infty$, then the line with equation $x = a$ has the property that the graph of $f()$ never touches this line, but does come arbitrarily close to the line as we look farther (up or down) from the horizontal axis. The geometric terminology is given in:

Definition 8.8: The vertical line with equation $x = a$ is a *vertical asymptote* of the graph of the function $f()$ provided that

$$\lim_{t \rightarrow a} |f(t)| = +\infty.$$

For rational functions, in which the numerator and denominator polynomials have no common polynomial factors, vertical asymptotes occur for every number which is a zero of the denominator.

For example, let the rational function $f()$ be specified by

$$f(x) = \frac{x^2 + 1}{x^2 - 3x + 2}.$$

There are no numbers which make $x^2 + 1$ have the value 0, so that we need not concern ourselves with any algebraic factoring. The denominator function $D()$ is given by

$$D(x) = x^2 - 3x + 2,$$

so that $D(1) = D(2) = 0$. We have

$$f(x) = \frac{x^2 + 1}{(x - 1)(x - 2)}.$$

$$\lim_{t \rightarrow 1} |f(t)| = \lim_{t \rightarrow 1} \frac{t^2 + 1}{|t - 1| \cdot |t - 2|} = +\infty,$$

since, for t near 1, $(t^2 + 1)$ is near 1, $|t - 2|$ is near 1, and $|t - 1|$ is a positive number near 0. The quotient is then positive and large.

Similarly,

$$\lim_{t \rightarrow 2} \frac{t^2 + 1}{|t - 1| \cdot |t - 2|} = +\infty.$$

Thus, the vertical lines with equations $x = 1$ and $x = 2$ are vertical asymptotes of the graph of $f()$.

Furthermore, for this function, given by

$$f(x) = \frac{x^2 + 1}{(x - 1)(x - 2)},$$

$$f(0.99) = \frac{1.9801}{(-0.01)(-1.01)} = 200;$$

$$f(1.01) = \frac{2.0201}{(0.01)(-0.99)} = -200;$$

$$f(1.99) = \frac{4.9601}{(0.99)(-0.01)} = -500;$$

$$f(2.01) = \frac{5.0401}{(1.01)(0.01)} = 500;$$

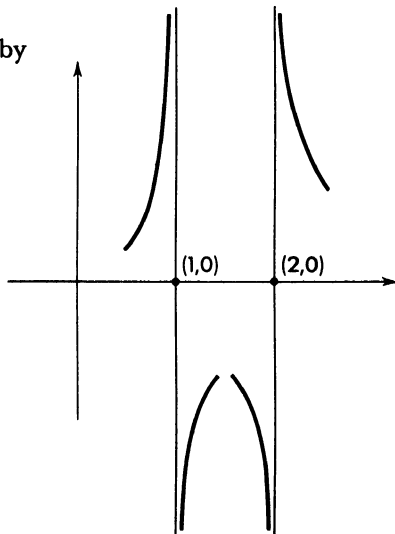


Fig. 8.1

so that we may sketch an approximation to part of the graph of $f()$ as shown in Fig. 8.1.

Exercises

1. Solve the inequalities:

a) $\frac{1}{(x-1)^2} > 100.$

b) $\frac{1}{(x-1)^2} > 10,000.$

c) $\frac{1}{(x-1)^2} > B$, where $B > 0.$

2. Solve the inequalities:

$$1 - \frac{1}{x^2} < N,$$

where: a) $N = -99$; b) $N = -9999$; c) N is some negative number.

3. Solve the inequalities:

$$\left| \frac{1}{x-3} \right| > B,$$

where: a) $B = 100$; b) $B = 1,000,000$; c) B is some positive number.

4. Find the vertical asymptotes of the graphs of the equations:

a) $xy = 4.$

b) $(x-1)(y-2) = 1.$

c) $(x-2)(y+3) = 2.$

d) $(x-4)(y-2) = 3.$

e) $y = \frac{x-2}{x^2-5x+4}.$

f) $y = \frac{x-3}{x^2-4x+3}.$

g) $y = \frac{1}{(x+1)(x-2)(x-4)}.$

5. Sketch the graphs of the equations in exercise 4.

6. Does the graph of $y = \frac{x^2-4}{x-2}$

have a vertical asymptote? Why or why not?

7. Why does $\lim_{t \rightarrow a} f(t) = -\infty$ imply that

$$\lim_{t \rightarrow a} |f(t)| = +\infty?$$

8. Does $\lim_{t \rightarrow a} |f(t)| = +\infty$ imply that either

$$\lim_{t \rightarrow a} f(t) = +\infty \quad \text{or} \quad \lim_{t \rightarrow a} f(t) = -\infty? \text{ Why or why not?}$$

Find the following limits:

9. $\lim_{t \rightarrow 3} \frac{t^3+1}{t^4-81}.$

10. $\lim_{u \rightarrow 0} \frac{1}{u^p}$, where p is a natural number.
11. $\lim_{v \rightarrow a} \frac{F(v)}{G(v)}$, where $F()$ and $G()$ are polynomials,
 $F(a) \neq 0$, $G(v) = (v - a)^2 H(v)$, and $H(v) \neq 0$.

8.8. OTHER ASYMPTOTES

The graph in the preceding section is incomplete, since we have not indicated how this graph behaves for x -coordinates very far from 0. It appears, however, that if x is a very large positive number, both factors $(x - 2)$ and $(x - 1)$ in the denominator will be large and positive, and the factor $(x^2 + 1)$ in the denominator will also be large. For the function specified by

$$f(x) = \frac{x^2 + 1}{(x - 1)(x - 2)},$$

we have

$$f(1,000) = \frac{1,000,001}{(999)(998)} \approx 1;$$

$$f(1,000,000) = \frac{1,000,000,000,001}{(999,999)(999,998)} \approx 1;$$

so that it appears that the quotient will be approximately 1 for large positive values of x .

In general, we have

Definition 8.9: If $f()$ is a function and b is a number, " $\lim_{t \rightarrow +\infty} f(t)$ is the number b " means that, for any $\epsilon > 0$, there exists a number $B > 0$, such that, if $t > B$, then t is in the domain of $f()$ and $|f(t) - b| < \epsilon$.

Speaking loosely, we say "the limit of $f(t)$, as t becomes very large" is the number b whenever we can make $f(t)$ as close as we please to b , merely by taking t large enough.

Similarly, we have

Definition 8.10: If $f()$ is a function and c is a number " $\lim_{t \rightarrow -\infty} f(t)$ is the number c " means that, for every $\epsilon > 0$, there will exist a number $N < 0$, such that, if $t < N$, then t is in the domain of $f()$ and $|f(t) - c| < \epsilon$.

Roughly, this definition says that "the limit of $f(t)$, as t becomes very large negatively" is the number c whenever the negative values of t , vary far from 0, give corresponding values of $f(t)$ closer and closer to c .

For the example of the preceding section,

$$\lim_{t \rightarrow +\infty} \frac{t^2 + 1}{(t - 1)(t - 2)} = 1$$

and

$$\lim_{t \rightarrow -\infty} \frac{t^2 + 1}{(t-1)(t-2)} = 1$$

Geometrically, we have

Definition 8.11: The horizontal line with equation $y = b$ is a *horizontal asymptote* of the graph of the function $f()$ provided that either or both of

$$\lim_{t \rightarrow -\infty} f(t) \quad \text{or} \quad \lim_{t \rightarrow +\infty} f(t)$$

is the number b .

We will not concern ourselves here with lines other than horizontal or vertical ones as asymptotes of graphs.

A particularly simple way to discover horizontal and vertical asymptotes of the graph of a function $f()$ is afforded by considering solutions of an equation of the form

$$r = f(t),$$

for t in terms of r . If, of several possible solutions, one is in the form

$$t = g(r),$$

where $g()$ is a function, then any vertical asymptote of the graph of $g()$ will be a horizontal asymptote of the graph of $f()$. That is, if $x = c$ is a vertical asymptote of the graph of $y = g(x)$, then $y = c$ will be a horizontal asymptote of the graph of $y = f(x)$.

For example, if $f()$ is the function

$$\left\{ \frac{1}{t^2 - 3t + 2} \right\} (),$$

we have

$$r = \frac{1}{t^2 - 3t + 2};$$

$$t^2 - 3t + 2 - \frac{1}{r} = 0;$$

$$t = \frac{9 \pm \sqrt{9 - 4\left(2 - \frac{1}{r}\right)}}{2}.$$

If, in this third equation, we choose either “+” or “−” for the numerator, we cannot have $r = 0$. Thus, the vertical line with equation $x = 0$ is a horizontal asymptote of the graph of the given function. This fact may be confirmed by noting that

$$\lim_{x \rightarrow +\infty} \frac{1}{x^2 - 3x + 2} = 0.$$

Exercises

1. Solve the inequalities:

a) $\left| \frac{3x+2}{x-1} - 3 \right| < 1, \quad x > 0.$

b) $\left| \frac{3x+2}{x-1} - 3 \right| < 0.01, \quad x > 0.$

c) $\left| \frac{3x+2}{x-1} - 3 \right| < e, \quad x > 0 \text{ (where } e > 0\text{)}.$

2. Find negative values of x which are solutions to the inequalities in exercise 1.

3. Find: a) $\lim_{t \rightarrow +\infty} \frac{3t+2}{t-1}.$

b) $\lim_{t \rightarrow -\infty} \frac{3t+2}{t-1}.$

4. How does exercise 3 relate to exercises 1 and 2?

5. Find the horizontal asymptotes of the graphs of the equations in exercise 4 of the preceding section.

6. Sketch the graphs of the equations in exercise 4 of the preceding section.

Find the following limits:

7. $\lim_{r \rightarrow -\infty} \frac{2r^2 + 5r - 1}{4r^2 - 3r + 7}.$

8. $\lim_{s \rightarrow +\infty} \frac{4s - 1}{3s^2 - 2s + 5}.$

9. $\lim_{t \rightarrow -\infty} \frac{2t^2 + 4t - 1}{t + 6}.$

10. $\lim_{u \rightarrow +\infty} \frac{1}{\sqrt{u}}.$

11. $\lim_{v \rightarrow -\infty} \frac{1}{\sqrt{v}}.$

12. $\lim_{w \rightarrow +\infty} \sqrt{\frac{w-2}{w+3}}.$

13. Show that, if $f()$ is specified by

$$f(x) = x + 1/x,$$

then

$$\lim_{x \rightarrow +\infty} f'(x) = \lim_{x \rightarrow -\infty} f'(x) = 1.$$

How does this indicate that the graph of $f()$ has, as a non-vertical and non-horizontal asymptote, the line with equation $y = x$.

8.9. VERTICAL TANGENTS

Before using the concept of the derivative of a function to define what is meant by the tangent line to the graph of a function, we should recall that it was possible to define a tangent line to a circle without making use of derivatives at all. The definition, learned in elementary geometry, is that a line through a point on the circumference of a circle, which line is perpendicular to the radius from the center to that point, is tangent to the circle.

If we look at the graph of the equation

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

we can see that any point on the graph of this equation must be precisely one unit from the origin, and thus that this graph is a circle, centered at the origin, with radius 1. (See Fig. 8.2.) Since a vertical line through the point with coordinate $(1, 0)$ certainly should be tangent to the circle, our definition of the tangent line to the graph of a function should not prevent this vertical line from being tangent to the graph of this circle.

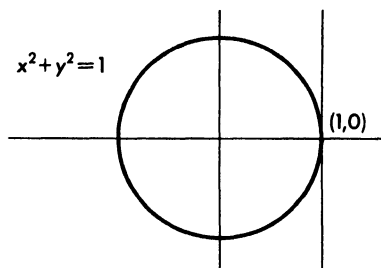


Fig. 8.2

The definition given of tangent line to the graph of a function says that the line through the point $(a, f(a))$ whose slope is the number $f'(a)$ will be the tangent line. According to this definition, no vertical line could be the tangent line to the graph of a function since vertical lines do not have any slope.

But we have no contradiction here, since the circle of radius 1, centered at the origin is not the graph of a function, but is merely the graph of an equation. For the present, then we will consider only graphs of functions, and we will make the following definition:

Definition 8.12: The line with equation $x = a$ is tangent to the graph of the function $f()$ at the point $(a, f(a))$ provided that a is in the domain of $f()$, but is not in the domain of $f'()$, and that

$$\lim_{t \rightarrow a} |f'(t)| = +\infty.$$

Since the situation described in this definition cannot occur for any rational function, we will have no vertical tangents to the graph of any of the functions thus far encountered. Then we must keep in mind that we have two things to do: 1) find some functions other than rational ones, and see if this definition applies to them; and 2) find some way of determining tangent lines to graphs of equations which graphs are not graphs of functions.

Exercise

Show that the graph of a rational function cannot have a vertical tangent anywhere.

8.10. RATIONAL NUMBER EXPONENTS

In Chapter 1, we defined a rational number to be any number which could be expressed as the quotient of two integers. Our concern here will be with rational numbers which are not integers, so that we are looking at numbers of the form p/q , where p and q are integers. However, since a positive rational number can be expressed as the quotient of two positive integers, and a negative rational number can be expressed as the quotient of a negative integer divided by a positive integer, it is sufficient for us to look at numbers of the form p/q , where p is any non-zero integer (positive or negative), but q is definitely a positive integer.

If b is any non-zero number, and p is any non-zero integer, then the number b^p has already been defined by saying:

if p is a positive integer, b^p is the product:

b multiplied by itself p times;

if p is a negative integer, b^p is $\frac{1}{b^{-p}}$.

If we succeed in defining the number

$$(b)^{1/q},$$

where q is a positive integer, then we can define

$$(b)^{p/q}$$

as the p -th power of $b^{1/q}$.

In order to avoid difficulties with non-real numbers, we will restrict the values of b to positive ones, giving

Definition 8.13: If $a > 0$ and q is a positive integer,

$$a^{1/q}$$

is that positive number which, when raised to the power q , yields a . $0^{1/q} = 0$.

According to this definition

$$(a^{1/q})^q$$

is the number a , while the requirement that the result of raising a number to the power $1/q$ (where q is a positive integer) means that

$$(a^q)^{1/q} = |a|.$$

That is, $a^{1/q}$ is the non-negative number which is a solution of the equation $x^q = a$.

For $q = 2$, we already have a symbolism for $a^{1/2}$, namely $\sqrt{a}$. For positive integers other than 2, a usual symbol is given by

$$a^{1/q} = \sqrt[q]{a},$$

which is read as "the q -th root of a ." The use of the definite article, "the," emphasizes the non-negative character of this " q -th root."

Finally, to allow any rational number to be an exponent of a non-negative number, we have

Definition 8.14: If p is an integer, and q is a positive integer, and $a > 0$, then

$$a^{p/q} = (a^{1/q})^p.$$

If p and q are positive integers, $0^{p/q} = 0$.

We may now use rational numbers as exponents in combining functions. For example, the function

$$\{I^{3/2}\}()$$

is the function specified by the equation

$$y = x^{3/2},$$

which is equivalent to the equation

$$y = (\sqrt{x})^3.$$

The domain of this function consists of the set of all non-negative numbers.

The rational number $1/3$ and the rational number $2/6$ are certainly the same rational number, and if $a > 0$, then $a^{1/3}$ and $a^{2/6}$ will be the same number. However, if we are to allow negative numbers to be raised to powers which powers are rational numbers, then we must be very careful to make sure that every symbol we write is actually a symbol for a number. If the definition of rational number exponents is to remain as above, namely $a^{p/q} = (a^{1/q})^p$, then we have the following strange situation:

$(-8)^{1/3}$ should be a solution of the equation $x^3 = -8$. The only number which is a solution of this equation is -2 , so that it would seem to offer no difficulty if we were to say that $(-8)^{1/3} = -2$.

$(-8)^{2/6}$ should $[(-8)^{1/3}]^2$, so that we must first find $(-8)^{1/3}$. $(-8)^{1/6}$ should be a solution of $x^6 = -8$. But there is no solution of this equation, since the sixth power of any number is not negative. Thus, $(-8)^{2/6}$ is not a number.

We are then forced to conclude that, as exponents, $1/3$ and $2/6$ are not the same number, or at least do not have the same meaning, and we make the following definition:

Definition 8.15: If p is an integer, if q is an odd positive integer, and $a < 0$, then

$$a^{p/q} \text{ is } (a^{1/q})^p,$$

where $a^{1/q}$ is the one solution of $x^q = a$.

According to this definition, $(-8)^{1/3}$ is -2 , and $(-8)^{2/6}$ is meaningless. Then the function $\{I^{2/3}\}()$ is the function specified by the equation

$$y = x^{2/3},$$

which is equivalent to the equation

$$y = (\sqrt[p]{x})^2,$$

and the domain of this function consists of the set of all numbers. But the domain of $\{I^{4/3}\}()$ is the set of all non-negative numbers.

Thus $\{I^{p/q}\}()$ is a well defined function, whose domain sometime is the set of all numbers, sometime the set of all non-negative numbers, sometime the set of all numbers except 0, and sometime the set of all positive numbers, depending on which integers p and q are.

To find the derivative of $\{I^{p/q}\}()$, where p is any non-zero integer, and q is a positive integer, we need to investigate

$$\lim_{h \rightarrow 0} \left[\frac{(x+h)^{p/q} - x^{p/q}}{h} \right].$$

For convenience in this investigation, we let

$$\begin{aligned} a &= (x+h)^{p/q} \\ \text{and } b &= x^{p/q} \end{aligned}$$

We then have

$$\begin{aligned} a^q &= (x+h)^p \\ \text{and } b^q &= x^p. \end{aligned}$$

Further, since q is a positive integer,

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

Then the numerator of the quotient whose limit will give us the derivative becomes

$$a - b = \frac{a^q - b^q}{a^{q-1} + a^{q-2}b + a^{q-3}b^2 + \dots + a b^{q-2} + b^{q-1}},$$

where there are precisely q terms in the denominator. But

$$a^q - b^q = (x+h)^p - x^p,$$

so that

$$\begin{aligned} & \frac{(x+h)^{p/q} - x^{p/q}}{h} \\ &= \left[\frac{1}{\begin{aligned} & (x+h)^{p/q(q-1)} + (x+h)^{p/q(q-2)} x^{p/q} + (x+h)^{p/q(q-3)} x^{(p/q)(2)} + \dots \\ & + (x+h)^{p/q} x^{p/q(q-2)} + x^{p/q(q-1)} \end{aligned}} \right] \\ & \quad \cdot \left[\frac{(x+h)^p - x^p}{h} \right] \end{aligned}$$

Since p is an integer,

$$\lim_{h \rightarrow 0} \frac{(x+h)^p - x^p}{h}$$

is precisely the limit encountered in finding the derivative of $\{I^p\}()$, which limit exists and is $p x^{p-1}$.

Each term in the denominator of the first factor has, as $h \rightarrow 0$, the limit $x^{p/q(q-1)}$, since each of these terms gives a rational number exponent function, with independent variable h . Since there are q terms in this denominator, we have

$$\lim_{h \rightarrow 0} \frac{(x+h)^{p/q} - x^{p/q}}{h} = \frac{1}{q x^{p/q(q-1)}} p x^{p-1}$$

by the theorems of the limit of sums, product, and quotients of functions.

Furthermore,

$$\begin{aligned} \frac{p}{q} (q-1) &= p - \frac{p}{q}, \\ \frac{1}{q x^{p/q(q-1)}} p x^{p-1} &= \frac{p}{q} x^{p-1-(p-p/q)} \\ &= \frac{p}{q} x^{p/q-1}, \end{aligned}$$

and we have proved

Theorem 8.7:

If r is a rational number,
then

$$\{I^r\}'() = \{r I^{r-1}\}().$$

An immediate consequence is the anti-derivative theorem:

Theorem 8.8:

If s is a rational number other than -1 ,

$$\left\{ \frac{1}{s+1} I^{s+1} \right\} () \text{ is an anti-derivative of } \{I^s\} ().$$

Proof:

$$\left\{ \frac{1}{s+1} I^{s+1} \right\}'() = \{I^s\} ().$$

Exercises

1. Let $f()$ be the function $\{I^{1/3}\}()$.
 - a) Find $\lim_{t \rightarrow 0} f(t)$.
 - b) Find $f'()$.
 - c) Find $\lim_{t \rightarrow 0} |f'(y)|$.
 - d) Sketch the graph of $f()$.
2. Repeat exercise 1 with $f() = \{I^{1/2}\}()$.

Simplify:

3. $4^{1/2}, 8^{1/3}, 9^{1/2}, 16^{1/4}$.
4. $4^{3/2}, 8^{2/3}, 9^{5/2}, 16^{5/4}$.
5. $4^{-1/2}, 8^{-1/3}, 9^{-1/2}, 16^{-1/4}$.
6. $4^{-3/2}, 8^{-2/3}, 9^{-5/2}, 16^{-5/4}$.
7. $(-8)^{1/3}, (-8)^{2/3}, (-8)^{-1/3}, (-8)^{-2/3}$.

Differentiate:

8. $\{I^{1/2}\}()$.
9. $\{I^{1/3} - 2I^{3/5} + I\}()$.
10. $\{3I^{1/8} + 2I^{-4/5} - 8I^{3/2}\}()$.
11. $\{\frac{1}{4}I^{3/2} - \frac{1}{3}I^{-3/2} + \frac{1}{5}I^{-1/10}\}()$.
12. $\{\sqrt{2}I^{2/3} - \sqrt{5}I^{-1/5} + \sqrt{3}I^4\}()$.
13. Write out a formal proof of Theorem 8.6.
14. Find the area of that part of the plane bounded by the graphs of $y = x^{1/2}$, $x = 1$, $x = 2$, and $y = 0$.
15. Find all relative maximum and minimum values of the function specified by

$$F(x) = x^{1/2} + x^{-1/2}.$$

COMPOSITES OF FUNCTIONS

9.1. THE COMPOSITE OF TWO FUNCTIONS

We now have defined the operations of addition, subtraction, multiplication, division, and raising to rational number powers for functions. If we combine the new operation of raising to rational number powers with some or all of the previous operations, we will have an entirely new kind of function. For example, if t is a variable, then $t^2 + 1$ will be a symbol for a number, and, in fact, a positive number, so that $(t^2 + 1)^{1/2}$ will be a number. The algebraic expression

$$(t^2 + 1)^{1/2}$$

specifies a function whose domain is the set of all numbers, but one whose derivative we cannot find by the use of any of the theorems thus far encountered.

Let us look more slowly at the construction of this new function. Suppose we have the equations

$$r = s^{1/2} \quad \text{and} \quad s = t^2 + 1.$$

Each of these equations will specify a function, and the particular symbols we have used for the variables involved in these functions are completely immaterial insofar as the two sets of ordered pairs are concerned.

Thus, the equation

$$r = s^{1/2},$$

will specify the function $\{I^{1/2}\}()$ when s is an independent variable and r is a dependent variable. The domain of $\{I^{1/2}\}()$ is the set of non-negative numbers, and the range of values of $\{I^{1/2}\}()$ is also the set of non-negative numbers. For convenience, we will denote $\{I^{1/2}\}()$ by $f()$.

The equation

$$s = t^2 + 1$$

specifies the function $\{I^2 + 1\}()$ when t is an independent variable and s is a dependent variable. The domain of the polynomial function $\{I^2 + 1\}()$ is the set of all numbers, and the range of values of $\{I^2 + 1\}()$ is the set of all numbers not less than 1. For convenience, we will call $\{I^2 + 1\}()$ the function $g()$.

We have now:

$f() = \{I^{1/2}\}()$. s is an independent variable of $f()$. r is a dependent variable of $f()$. Thus, $r = f(s)$.

$g() = \{I^2 + 1\}()$. t is an independent variable of $g()$. s is a dependent variable of $g()$. Thus, $s = g(t)$.

The use of the same symbol s for an independent variable of one function and a dependent variable of another at the same time is motivated by the following observation: if, in the equation $r = s^{1/2}$, we substitute for s that which the equation $s = t^2 + 1$ tells us that s is, we will have a new equation

$$r = (t^2 + 1)^{1/2},$$

which new equation specifies a function with dependent variable r and independent variable t .

In terms of the variables, the two equations involving three variables may be interpreted as saying that r is a function of s , while s , in turn, is a function of t , so that the value of r does indeed depend on the value of t , with the dependence determined by the double role played by s as simultaneously a dependent and independent variable. This variable s is some kind of intermediate variable of some new kind of function, which new function is a new kind of combination of the original two functions. There should be no concern over the phrase " s is both a dependent and an independent variable," since this phrase, by itself, means nothing. There is no such thing as a "dependent variable," but only a variable's being a dependent variable *of a function*, and we have merely said that we are using the same symbol s as a dependent variable of the function $g()$ and as an independent variable of the function $f()$.

We have a situation in which the variable r is a function of, a function of t . We will now drop the comma and use the terminology that r is a "function-of-a-function" of t , and see what this means in terms of the two component functions, $f()$ and $g()$.

If we consider one number a in the domain of $g()$ —that is, a is one value of the variable t —then the number $g(a)$ will be in the range of values of $g()$ —that is, a value of the variable s , and the ordered pair $(a, g(a))$ will be one of the ordered pairs of the function $g()$.

Now $g(a)$ is a number, and, for the specific example we have,

$$g(a) = a^2 + 1.$$

Thus, $g(a)$ is not a negative number, so that $g(a)$ is in the domain of the function $f()$ —meaning that it is again a value of the variable s . Since the number $g(a)$ is in the domain of $f()$, then the number $f(g(a))$ will be the

corresponding number in the range of values of $f()$ —namely, a value of the variable r . This means that the ordered pair $(g(a), f(g(a)))$ is one of the ordered pairs of the function $f()$.

But the number $f(g(a))$ is the value of the variable r which corresponds to the number a as a value of the variable t . That is, the ordered pair

$$(a, f(g(a)))$$

is one of the ordered pairs of the function specified by the equation

$$r = (t^2 + 1)^{1/2},$$

in which r is a dependent variable and t is an independent variable. This function then must consist of all ordered pairs of the form

$$(a, (a^2 + 1)^{1/2}).$$

We have determined a rule of this new function by combining rules for $f()$ and $g()$ so that, to each number in the domain of our new function, we first apply a rule of $g()$, and then, to the result of that applying of the rule of $g()$, we apply a rule of $f()$. Evidently, the only restriction which must be observed is that the result of applying a rule of $g()$ must be a number actually in the domain of $f()$. In general,

Definition 9.1: If $f()$ and $g()$ are functions, then the *composite* function $\{f(g)\}()$ is defined by

- a) The domain of $\{f(g)\}()$ is the set of all numbers, a , in the domain of $g()$, for which $g(a)$ is in the domain of $f()$;
- b) A rule of $\{f(g)\}$ is given by

$$\{f(g)\}(a) = f(g(a)).$$

EXAMPLE 1:

$g()$ is the function consisting of the ordered pairs:

$$(1, 2), (2, 0), (3, 4), (4, -2), \text{ and } (5, 3).$$

$f()$ is the function consisting of the ordered pairs:

$$(1, 0), (2, 2), (3, -1), (4, -2), \text{ and } (5, 1).$$

The domain of $g()$ consists of the numbers: 1, 2, 3, 4, and 5. For these,

$$g(1) = 2, g(2) = 0, g(3) = 4, g(4) = -2, \text{ and } g(5) = 3.$$

Thus the range of values of $g()$ consists of the numbers: 2, 0, 4, -2, and 3. Of these five numbers, only $g(1)$, $g(3)$, and $g(5)$ are in the domain of $f()$, so that the domain of $\{f(g)\}()$ consists of the three numbers: 1, 3, and 5.

$$\begin{aligned} f(g(1)) &= f(2) = 3; f(g(3)) = f(4) = -2; \\ f(g(5)) &= f(1) = 2. \end{aligned}$$

Thus, $\{f(g)\}()$ consists of the three ordered pairs:

$$(1, 2), (3, -2), \text{ and } (5, 2).$$

(The composite, $\{g(f)\}()$, of these functions, taken in the opposite order consists of the two ordered pairs: $(2, 0)$ and $(5, 2)$.)

EXAMPLE 2:

$g()$ is the function specified by the equation

$$y = 3x - 6,$$

in which x is an independent variable and y a dependent variable.

$f()$ is the function specified by the equation

$$y = x^{1/2},$$

in which x is an independent variable and y a dependent variable.

The domain of $g()$ is the set of all numbers and the domain of $f()$ is the set of all non-negative numbers. In order to discover which part of the range of values of $g()$ is actually in the domain of $f()$, we must solve the inequality

$$3x - 6 \geq 0.$$

The solution of this inequality is the interval:

$$\text{the set of all } x, \quad x \geq 2.$$

Consequently, this interval is the domain of the composite function $\{f(g)\}()$. This function is specified by the equation:

$$y = (3x - 6)^{1/2}.$$

It should be noted that we could have used any variables in giving the function $f()$, so that it would have been particularly convenient for us to use, as an independent variable of this "outside" function, the same symbol as we used for a dependent variable of $g()$, the "inside" function. Thus, specifying $g()$ by

$$y = g(x) = 3x - 6$$

and $f()$ by

$$z = f(y) = y^{1/2}$$

would allow simple substitution to give

$$z = \{f(g)\}(x) = (3x - 6)^{1/2},$$

which equation does indeed specify $\{f(g)\}()$.

EXAMPLE 3:

$g()$ is the function $\{I^2 - 3I + 4\}()$.

$f()$ is the function $\{I^2\}()$.

Here the domain of $f()$ consists of the set of all numbers, so that the domain of the composite $\{f(g)\}()$ is all of the domain of $g()$. If we use the equations

$$\begin{aligned}s &= g(t) = t^2 - 3t + 4 \\ r &= f(s) = s^2,\end{aligned}$$

then the equation

$$r = \{f(g)\}(t) = (t^2 - 3t + 4)^2$$

will specify $\{f(g)\}()$.

Exercises

For the pairs of functions given, find the domain and a rule of the two composites: $\{f(g)\}()$ and $\{g(f)\}()$.

1. $f() = \{I^2\}(); g() = \{I^3\}()$.
2. $f() = \{2I + 1\}(); g() = \{I^3\}()$.
3. $f() = \{3I - 6\}(); g() = \{I^{1/4}\}()$.
4. $f() = \{3I - 6\}(); g() = \{I^{1/3}\}()$.
5. $f() = \{I^{1/2}\}(); g() = \{I^{1/3}\}()$.
6. $f() = \{I^{1/2}\}(); g() = \{I^2\}()$.
7. $f() = \{I^{1/3}\}(); g() = \{I^3\}()$.
8. $f() = \{4I + 5\}(); g() = \{3I - 2\}()$.
9. $f() = \{3 - I\}(); g() = f()$.
10. $f() = \{I^2 + I\}(); g() = \{I + 1\}()$.
11. If $f() = \{I + 1\}()$, find a rule for

$$\{f(f(f(f(f))))\}().$$

12. If $f()$ and $g()$ are rational functions, find the domain of $\{f(g)\}()$.

9.2. THE LIMIT THEOREM FOR THE COMPOSITE OF TWO FUNCTIONS

In investigating the limit, as we approximate a given number, of the composite $\{f(g)\}()$, we would expect that, unless the individual functions misbehave, if we have t a number near a in the domain of $g()$, with $g(t)$ a number near b , and if $f(s)$ is a number near c for s any number near b , then $\{f(g)\}(t)$ —which is $f(g(t))$ —ought to be a number near c . We have

Theorem 9.1:

$$\text{If } \lim_{t \rightarrow a} g(t) = B, \lim_{s \rightarrow B} f(s) = C,$$

$$\text{and } f(B) = C,$$

$$\text{then } \lim_{t \rightarrow a} \{f(g)\}(t) \text{ exists and is } C.$$

Proof:

1. Pick $\epsilon > 0$.
2. $\lim_{t \rightarrow B} f(s) = C$ by hypothesis.
3. There is an interval:
 the set of all s ,
 $|s - B| < d$ (where $d > 0$),
 such that, if s is in this interval,
 then s is in the domain of $f()$ and
 $|f(s) - C| < \epsilon$ by step 2, the definition of limit,
and the hypothesis that $f(B) = C$.
4. $\lim_{t \rightarrow a} g(t) = B$ by hypothesis.
5. There is an interval:
 the set of all t ,
 $|t - a| < d'$ (where $d' > 0$),
 such that, if $0 < |t - a| < d'$,
 then t is in the domain of $g()$ and
 $|g(t) - B| < d$ by step 3, step 4, and the definition
of limit.
6. If $0 < |t - a| < d'$, then $g(t)$
 is in the domain of $f()$ by steps 3 and 5 and the hypothesis
that B is in the domain of $f()$.
7. If $0 < |t - a| < d'$, then t is
 in the domain of $\{f(g)\}()$ by step 6 and the definition of the
composite.
8. If $0 < |t - a| < d'$, then
 $|f(g(t)) - C| < \epsilon$ by steps 3 and 5.
9. If $0 < |t - a| < d'$, then t is
 in the domain of $\{f(g)\}()$ and
 $|\{f(g)\}(t) - C| < \epsilon$ by steps 6 and 8 and the definition
of the composite.
10. $\lim_{t \rightarrow a} \{f(g)\}(t) = C$ by steps 1, 5, 7, and 9, and the
definition of limit

An immediate result of this theorem is

Corollary to Theorem 9.1:

If $g()$ is continuous at a and $f()$ is continuous at $g(a)$, then $\{f(g)\}()$ is continuous at a .

This corollary is often stated more concisely as “a continuous function of a continuous function is continuous.”

Exercises

Let $f() = \{I^2\}()$ and $g() = \{3I - 1\}()$.

1. Find $\{f(g)\}()$.
2. Find $\{f(g)\}(2)$ and $\lim_{r \rightarrow 2} \{f(g)\}(r)$.
3. Show that the interval:
the set of all s , $2 - 0.1 < s < 2 + 0.1$
is a solution of the inequality:
$$4 - 0.5 < t^2 < 4 + 0.5.$$
4. Find the solution of the inequality:
$$2 - 0.1 < 3u - 1 < 2 + 0.1.$$
5. Show that the interval found in exercise 4 is a solution of the inequality:
$$4 - 0.5 < \{f(g)\}(v) < 4 + 0.5.$$
6. Show that if $f()$ and $g()$ are polynomial functions, then $\{f(g)\}()$ is also a polynomial function.
7. Show that if $f()$ and $g()$ are rational functions, then $\{f(g)\}()$ is also a rational function.
8. Give an example of pairs of functions for which
 - a) $\{f(g)\}() = f()$.
 - b) $\{f(g)\}() = g()$.
 - c) $\{f(g)\}()$ exists, but $\{g(f)\}()$ does not.
9. Let $f() = \{I^2\}()$ and $g() = \{I^3\}()$.
 - a) Find $\{f(g)\}(1)$ and $\{f(g)\}(1 + h)$.
 - b) Find $\{f(g)\}(1 + h) - \{f(g)\}(1)$.
 - c) Find $\lim_{h \rightarrow 0} \frac{\{f(g)\}(1 + h) - \{f(g)\}(1)}{h}$.
 - d) Find $f'()$ and the composite $\{f'(g)\}()$.
 - e) Evaluate $\{f'(g)\}(1)$ and $g'(1)$.
 - f) Compare the results of parts c) and e).

9.3. THE DERIVATIVE OF A COMPOSITE ("CHAIN RULE")

We come now to the differentiation of the composite of two functions. Once we have found such a derivative, we have the complete machinery to enable us to differentiate any algebraic function. In order to find the derivative of the composite $\{f(g)\}()$, we must investigate

$$\lim_{h \rightarrow 0} \frac{\{f(g)\}(a + h) - \{f(g)\}(a)}{h}.$$

The quotient in this limit specifies a function whose independent variable is h , given by the expression

$$\frac{f(g(a+h)) - f(g(a))}{h}.$$

If $g()$ is a continuous function, the number $g(a+h)$ will not differ too much from the number $g(a)$. In fact, another way of saying that $g()$ is continuous at a is to say

$$g(a+h) = g(a) + k(h),$$

$$\lim_{h \rightarrow 0} k(h) = 0.$$

With the new manner of expressing $g(a+h)$, the quotient involved in the derivative becomes

$$\frac{f(g(a) + k(h)) - f(g(a))}{h}.$$

Since $g(a+h) - g(a) = k(h)$, we may rewrite this quotient by noting that

$$\begin{aligned} & \frac{f(g(a) + k(h)) - f(g(a))}{h} \\ &= \left[\frac{f(g(a) + k(h)) - f(g(a))}{k(h)} \right] \left[\frac{k(h)}{h} \right] \\ &= \left[\frac{f(g(a) + k(h)) - f(g(a))}{k(h)} \right] \left[\frac{g(a+h) - g(a)}{h} \right], \end{aligned}$$

recognizing, of course, that this equation is true only for values of the variable h for which $h \neq 0$ and $k(h) \neq 0$.

The limit of the product of two functions is the product of their limits, so we need to look separately at the factors:

$$\lim_{h \rightarrow 0} \frac{f(g(a) + k(h)) - f(g(a))}{k(h)}$$

and

$$\lim_{h \rightarrow 0} \frac{g(a+h) - g(a)}{h}.$$

The second of these limits is, by the definition of derivative, the number $g'(a)$, provided only that a is in the domain of $g'()$. But if a is in the domain of $g'()$, then $g()$ is continuous at a , by the theorem that differentiability implies continuity, so that

$$\lim_{h \rightarrow 0} k(h) = \lim_{h \rightarrow 0} [g(a+h) - g(a)] = 0$$

It seems plausible, then, that we should be able to suppress the variable h in the first limit, and evaluate

$$\lim_{k \rightarrow 0} \frac{f(g(a) + k) - f(g(a))}{k}$$

which is merely the number $f'(g(a))$, provided that the number $g(a)$ is in the domain of $f'()$. If such a procedure is valid, then we have

$$\{f(g)\}'(a) = f'(g(a)) g'(a).$$

This conclusion is correct, and we have

Theorem 9.2:

("The Chain Rule")

If $f()$ and $g()$ are functions, if a is in the domain of $g'()$ and $g(a)$ is in the domain of $f'()$, then a is in the domain of $\{f(g)\}'()$ and

$$\{f(g)\}'(a) = f'(g(a)) g'(a).$$

(An alternate statement, in terms of the functions only, is:

$$\{f(g)\}'() = \{f'(g)\} \cdot g'().)$$

Proof:

$$1. \frac{\{f(g)\}(a+h) - \{f(g)\}(a)}{h} = \frac{f(g(a+h)) - f(g(a))}{h},$$

for every value of the variable h ,
other than 0, for which $a+h$ is
in the domain of $\{f(g)\}()$ by the definition of the composite.

2. Let the function $k()$ be specified by

$$k(h) = g(a+h) - g(a).$$

3. $f(g(a+h)) = f(g(a) + k(h))$ by step 2.

$$4. \frac{\{f(g)\}(a+h) - \{f(g)\}(a)}{h} = \frac{f(g(a) + k(h)) - f(g(a))}{h}$$

for every value of the variable h ,
other than 0, for which $a+h$ is
in the domain of $\{f(g)\}()$ by steps 1 and 3.

$$5. \frac{\{f(g)\}(a+h) - \{f(g)\}(a)}{h} = \left[\frac{f(g(a) + k(h)) - f(g(a))}{k(h)} \right] \left[\frac{k(h)}{h} \right],$$

for those values of the variable h
for which $g(a) + k(h)$ is in the
domain of $f()$, except for 0 and
those values of h for which
 $k(h) = 0$ by step 4 and M-4.

$$6. \frac{k(h)}{h} = \frac{g(a+h) - g(a)}{h}$$

for those values, other than 0,
of the variable h for which $a+h$
is in the domain of $g()$ by step 2.

7. a is in the domain of $g'()$ by hypothesis.

8. $\lim_{h \rightarrow 0} \frac{k(h)}{h}$ exists and is the number $g'(a)$

by steps 6 and 7 and the definition of derivative.

9. Pick $\epsilon > 0$.

10. $g(a)$ is in the domain of $f'()$

by hypothesis.

11. There is an interval:

the set of all t ,

$|t| < d_1$ (where $d_1 > 0$),

such that, if $0 < |t| < d_1$, then

t is in the domain of the function

specified by

$$\frac{f(g(a) + t) - f(g(a))}{t}$$

and

$$\left| \frac{f(g(a) + t) - f(g(a))}{t} - f'(g(a)) \right| < \epsilon$$

by steps 10, 11, and the definitions of derivative and limit.

12. $g()$ is continuous at a

by step 7 and the theorem which states that differentiability implies continuity.

13. $\lim_{h \rightarrow 0} k(h) = 0$

by step 2, 13, and the definition of continuity.

14. There exists an interval:

the set of all h ,

$|h| < d_2$ (where $d_2 > 0$),

such that, if $|h| < d_2$, then h

is in the domain of $k()$ and

$|k(h)| < d_1$

by step 13 and the definition of limit.

15. If $0 < |h| < d_1$,

then $k(h)$ is in the domain of the function specified by

$$\frac{f(g(a) + t) - f(g(a))}{t}$$

and

$$\left| \frac{f(g(a) + k(h)) - f(g(a))}{k(h)} - f'(g(a)) \right| < \epsilon$$

by steps 14 and 11.

16. $\lim_{h \rightarrow 0} \frac{f(g(a) + k(h)) - f(g(a))}{k(h)} = f'(g(a))$

by steps 9, 14 and 15, and the definition of limit.

$$17. \lim_{h \rightarrow 0} \frac{\{f(g)\}(a+h) - \{f(g)\}(a)}{h}$$

exists and is the number

$$f'(g(a)) g'(a)$$

by steps 5, 8, and 16, and the theorem on the limit of the product of functions.

18. a is in the domain of $f(g)'()$ and

$$\{f(g)\}'(a) = f'(g(a)) g'(a)$$

by step 17 and the definition of derivative.

This theorem allows us to differentiate some new kinds of functions, and also allows us to simplify the differentiation of some of the functions already encountered.

For example, consider the function

$$\{(I^3 + 3)^2\}().$$

By squaring, we have

$$\{(I^3 + 3)^2\}() = \{I^6 + 6I^3 + 9\}(),$$

so that

$$\{(I^3 + 3)^2\}'() = \{6I^5 + 18I^2\}().$$

However, $\{(I^3 + 3)^2\}()$ is also the composite

$$\{I^2(I^3 + 3)\}().$$

The derivative of this composite is given by

$$\begin{aligned} \{I^2(I^3 + 3)\}'() &= \{(2I\{I^3 + 3\}) \cdot (3I^2)\}() \\ &= \{(2I^3 + 6) \cdot (3I^2)\}() \\ &= \{6I^5 + 18I^2\}(). \end{aligned}$$

In terms of variables, this example says for

$$\begin{aligned} z &= f(y) = y^2 \\ \text{and } y &= g(x) = x^3 + 3, \\ z &= \{f(g)\}(x) = (x^3 + 3)^2 \\ \text{and } \{f(g)\}'(x) &= f'(g(x)) g'(x). \end{aligned}$$

Since $f'(y) = 2y$, we have

$$\begin{aligned} f'(g(x)) &= 2g(x) \\ &= 2x^3 + 6, \end{aligned}$$

and since $g'(x) = 3x^2$, we have

$$\begin{aligned} \{f(g)\}'(x) &= (2x^3 + 6)(3x^2) \\ &= 6x^5 + 18x^2. \end{aligned}$$

As a second example, let

$$F() = \{(I^2 + 1)^{1/2}\}().$$

If $f()$ is specified by

$$f(r) = r^{1/2}$$

and $g()$ is specified by

$$g(x) = x^2 + 1,$$

then $F() = \{f(g)\}()$, and we have

$$f'(r) = \frac{1}{2} r^{-1/2},$$

$$g'(x) = 2x,$$

so that

$$F'(x) = \frac{1}{2} (x^2 + 1)^{-1/2} (2x),$$

or, in other words, $F'()$ is specified by

$$F'(x) = \frac{x}{\sqrt{x^2 + 1}}.$$

Exercises

1. Find the derivatives of the composites in exercises 1-10 of Section 9.1.

Find the derivatives of the functions specified by:

2. $(r^2 + 3)^{10}$.

3. $(2s + 1)^3 - (2s + 1)^2 + 5$.

4. $(t^{-1} + t)^{-1}$.

5. $(u^{1/3} + u^{1/2})^{1/5}$.

6. $(v^3 + 3v^2 - 5v + 1)^{30}$.

7. $(w^{-1} - w^2)^{-15}$.

8. $(x + 1)^{333}$.

9. $(2y + 3)^{471}$.

10. $\frac{(z^2 - 5z)^4}{(3z + 1)^3}$.

Sketch the graphs of the following equations, showing relative extreme points and inflection points:

11. $y = (x^2 + 1)^{1/2}$.

12. $y = (1 - x^{2/3})^{3/2}$.

13. $y = (1 + 2x)^{1/3}$.

9.4. FUNCTIONS GIVEN BY AN EQUATION

If we have an equation involving two variables, we may be able to solve this equation, in a purely algebraic manner, so that one of the variables is expressed in terms of the other. For example if

$$3x + 4y = 5$$

is one such equation, then we may consider the equation

$$y = -\frac{3}{4}x + \frac{5}{4}$$

as an equivalent equation, in the sense that both these equations will be either simultaneously true or simultaneously false for any specific values of the variables involved.

If

$$y^2 = x$$

is another such equation involving two variables, we have the unfortunate problem of having to decide between the equations

$$y = x^{1/2} \quad \text{and} \quad y = -x^{1/2}.$$

If

$$x^5 + 3xy + y^6 = 1$$

is still another such equation, we can find no algebraic way to express y in terms of x or x in terms of y . However, if we assume that such a "solution" exists, then we are able to look at the original equation in a new light.

To be specific, we have

Definition 9.2: A function $f()$ is called a *function given by an equation involving two variables* provided that the substitution of an independent variable of $f()$ for one of the variables and of a dependent variable of $f()$ for the other variable in the equation makes the equation an identity.

It should be recalled that the range of an independent variable of a function is the domain of that function, that the range of a dependent variable of a function is the range of values of the function, and that an identity is an equation involving variables which equation is true for every allowable value of these variables.

In this manner,

$$\left\{-\frac{3}{4}I + \frac{5}{4}\right\}()$$

is a function given by the equation

$$3x + 4y = 5,$$

since, in this equation, if x is an independent variable of this function and y is a dependent variable of this function, then we have

$$y = -\frac{3}{4}x + \frac{5}{4}$$

and the equation

$$3x + 4\left\{-\frac{3}{4}x + \frac{5}{4}\right\} = 5$$

is an identity, in which the range of the variable x is the set of all numbers.

It is also true that

$$\left\{-\frac{4}{3}I + \frac{5}{3}\right\}()$$

is a function given by the equation

$$3x + 4y = 5,$$

since, in this equation, if y is an independent variable of the function and x is a dependent variable of the function, then we have

$$x = -\frac{4}{3}y + \frac{5}{3}$$

and the equation

$$3\left\{-\frac{4}{3}y + \frac{5}{3}\right\} + 4y = 5$$

is an identity, since the range of the variable y is the set of all numbers.

These two are not the only functions given by the equation $3x + 4y = 5$. Among others, we have the functions specified by

$$-\frac{4}{3}r + \frac{5}{3}, \quad 0 \leq r \leq 5;$$

and

$$-\frac{3}{4}s + \frac{5}{4}, \quad s > 0.$$

Let us consider the first of the functions above which are given by the equation $3x + 4y = 5$, namely

$$\left\{-\frac{3}{4}I + \frac{5}{4}\right\}(),$$

and, for convenience, call this function $f()$. By our definition, the equation

$$3x + 4f(x) = 5$$

is an identity.

The left-hand side of this identity specifies a function and the right-hand side of this identity specifies another function (which is a constant function). Thus, under the assumption that y is a dependent variable of the function $f()$, we may interpret the equation $3x + 4y = 5$ as a statement saying that two functions are the same. That is, if $H()$ is specified by

$$H(x) = 3x + 4f(x)$$

and $G()$ is specified by

$$G(x) = 5,$$

the assumption that $f()$ is a function given by the equation $3x + 4y = 5$ is equivalent to the assumption that $G()$ and $H()$ are the same function.

But, if $G() = H()$, then $G'() = H'()$. $G'()$ is specified by the equation

$$G'(x) = 3 + 4f'(x),$$

and $H'()$ is specified by

$$H'(x) = 0,$$

so that we have a new identity

$$3 + 4f'(x) = 0.$$

This last identity implies that $f'()$ is specified by

$$f'(x) = -\frac{3}{4},$$

and we have managed to find the derivative of $f()$ without actually finding $f()$ itself.

For the second example, the equation

$$y^2 = x,$$

let us assume that $f()$ is some function given by this equation with x an independent variable of $f()$ and y a dependent variable of $f()$.

There are many possible functions given by the equation $y^2 = x$. If we look at the graph of this equation (which is the collection of points resulting from interchanging x and y -coordinates of the points on the graph of $y = x^2$), we see that the graph is symmetric with respect to the x -axis, and lies in the first and fourth quadrants of the plane. (See Fig. 9.1.)

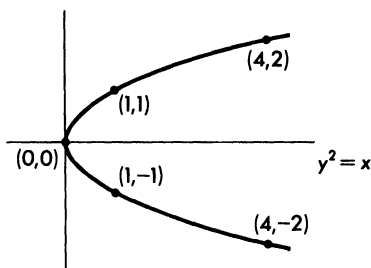


Fig. 9.1

Any collection of points on this graph, of which no two have the same x -coordinate, will have coordinates making up a function given by this equation. Examples of such functions are:

- the set of ordered pairs: $(0, 1)$, $(1, 1)$, $(4, -2)$, and $(8, -\sqrt{8})$.
- the set of ordered pairs of the form $(t, \sqrt{t})$, where $1 \leq t \leq 2$.
- the set of ordered pairs of the form $(r, -\sqrt{r})$, where $0 \leq r \leq 1$.
- the function given by $y = \sqrt{x}$.
- the function given by $y = -\sqrt{x}$.
- the set of ordered pairs of the form $(s, \sqrt{s})$, where s is a rational number between 4 and 9, together with the set of all ordered pairs of the form $(v, -\sqrt{v})$, where v is an irrational number between 4 and 9.

Here, we do not care which of the many possible such functions $f()$ actually is, but we will assume that $f()$ has a derivative, $f'()$. Of the functions above, $a)$ and $f)$ have no derivative, while the remaining ones do have a derivative. The equation $y^2 = x$ thus becomes the identity

$$[f(x)]^2 = x.$$

Since we have assumed that $f()$ has a derivative, the left-hand side of this identity specifies a function whose derivative is given by

$$2f(x)f'(x),$$

as can be seen by applying Theorem 9.2.

The right-hand side of the identity specifies the function $I()$, or a restriction of $I()$ to have as domain the range of the variable x . That is, the right-hand side gives the function specified by

$$x, x \text{ in the domain of } f().$$

This function has a derivative which will be a constant function whose range of values consists of the number 1. We thus have a new identity in

$$2f(x)f'(x) = 1.$$

This identity is equivalent to

$$f'(x) = \frac{1}{2f(x)}.$$

In this, we have found the derivative $f'()$ of a differentiable function $f()$, given by the equation $y^2 = x$, without actually knowing the form of $f()$.

If $f()$ happened to be $\{I^{1/2}\}()$, then

$$f'() = \left\{\frac{1}{2} I^{-1/2}\right\}().$$

If $f()$ happened to be $\{-I^{1/2}\}()$, then

$$f'() = \left\{-\frac{1}{2} I^{-1/2}\right\}().$$

In either case, $f'() = \left\{\frac{1}{2} I^{-1/2}\right\}()$.

For the third equation

$$x^5 + 3xy + y^6 = 5,$$

we cannot find any explicit form of any function given by this equation, but if we do assume that $f()$ is such a function, and also that $f()$ is differentiable, then the equation can be considered as an identity, and the differentiation of the two functions specified by each side of this identity yields the new identity:

$$5x^4 + 3xf'(x) + 3f(x) + 6[f(x)]^5 f'(x) = 0,$$

so that $f'()$ is specified by

$$f'(x) = -\frac{5x^4 + 3f(x)}{3x + 6[f(x)]^5}.$$

Exercises

In the following equations, x is an independent variable of the function $f()$, and y is a dependent variable of the function $f()$. The function $f()$ has a derivative.

Find an expression giving $f'()$:

1. $x^2 + y^2 = 25.$

2. $xy = 5.$

3. $x^{2/3} + y^{2/3} = 4.$

4. $x^3 + 2x^2y - 3xy^2 + 4y^3 = 2$.
5. $y^2 = x^3$.
6. Find an equation of the line tangent to the graph of the equation in exercise 1 at the points:
 $(3, 4)$, $(-4, -3)$, $(0, 5)$, and $(-5, 0)$.
7. Find an equation of the line tangent to the graph of the equation in exercise 2 at the point $(1, 5)$.
8. Find an equation of the line tangent to the graph of the equation in exercise 3 at the points:
 $(8, 0)$, $(\sqrt{8}, \sqrt{8})$, $(0, 8)$.
9. Sketch the graphs of the equations in exercises 1-3.
10. Prove that the graphs of the equations
 $x^2 + 3y^2 = 24$ and $3x^2 - y^2 = 12$
 intersect at right angles.
11. Find three distinct differentiable functions given by the equation
 $y^2 = 9x$.
12. The function $f()$, given by the equation
 $y + \sqrt{xy} - \frac{1}{4}x^3 = 0$,
 is twice differentiable at the number 1. Find $f''(1)$.
13. The graph of the equation
 $b^2x + a^2y = a^2b^2$
 (where a and b are constants) is an ellipse. Find an equation of the tangent line to this ellipse at any point (p, q) of the ellipse.

9.5. RELATED RATES

As a particularly useful aspect of the differentiation of functions whose existence is implied by an equation—which process is called implicit differentiation—we consider a situation in which an equation contains many variables, each of these being a dependent variable of some function, and each of these functions having the same variable (which may or may not appear) as an independent variable.

For example, suppose the variables x , y , and z are related by the equation

$$3x^2 + 4y^3 + 5z^4 = 1,$$

and we have the further information that x is a dependent variable of the function $F()$, y is a dependent variable of the function $G()$, z is a dependent variable of the function $H()$, and that the variable t —not mentioned in this equation—is an independent variable of the functions $F()$, $G()$, and $H()$. Further, we will suppose that one set of numbers acts as the domain of each function and that the range of the variable t is that set of numbers.

The equation, together with these facts, implies that

$$3[F(t)]^2 + 4[G(t)]^3 + 5[H(t)]^4 = 1$$

is an identity. Both sides of this equation specify functions, so that differentiation of each of these two functions yields the new identity

$$6F(t) F'(t) + 12[G(t)]^2 G'(t) + 20[H(t)]^3 H'(t) = 0.$$

If we have the further information that, for $t = a$, the numbers $F(a)$, $G(a)$, $H(a)$, $G'(a)$, and $H'(a)$ are known, then we can find easily the number $F'(a)$.

As an actual physical example, consider the situation in which two automobiles leave an intersection at the same time, one traveling north at 30 miles per hour, and the other traveling east at 40 miles per hour.

We let t be an independent variable of all the functions we are about to create, and interpret a value of t as the time in hours, which has elapsed since the departure of the autos. From this physical situation, the range of the variable t contains only non-negative numbers.

We let N be a dependent variable of the function $f()$, such that, for a given value of the variable t , $f(t)$ is the distance in miles traveled by the north-moving car, t hours after departure.

We let E be a dependent variable of the function $g()$, such that, for a given value of the variable t , $g(t)$ is the distance in miles traveled by the east-moving car, t hours after departure.

The information given about the velocities of these cars, together with the meaning of "rate of change," tells us that $f'()$ and $g'()$ are the constant functions given by

$$\begin{aligned} f'(t) &= 30, \quad t \geq 0 \\ \text{and } g'(t) &= 40, \quad t \geq 0. \end{aligned}$$

We let D be a dependent variable of the function $h()$, such that, for a given value of the variable t , $h(t)$ is the distance in miles between the two cars, t hours after departure. Then, for any value of the variable t , $h'(t)$ will be the rate at which the cars are separating.

The variables, E , N , and D are related by the equation

$$E^2 + N^2 = D^2,$$

from the Pythagorean theorem of elementary plane geometry. Then the equation

$$[f(t)]^2 + [g(t)]^2 = [h(t)]^2$$

is an identity. Differentiation yields the new identity

$$2f(t) f'(t) + 2g(t) g'(t) = 2h(t) h'(t),$$

which is equivalent to

$$f(t) f'(t) + g(t) g'(t) = h(t) h'(t).$$

When $t = 1$ (that is, one hour after departure), we have

$$\begin{aligned}f(1) &= 30, \\g(1) &= 40,\end{aligned}$$

so that, by the relation given between these three functions

$$h(1) = 50.$$

Substitution of 1 for t in the identity yields the true statement

$$f(1)f'(1) + g(1)g'(1) = h(1)h'(1),$$

i.e.

$$(30)(30) + (40)(40) = 50h'(1),$$

or, that $h'(1)$ is the number 50. This mathematical fact implies the physical fact that, at the end of one hour, the cars are separating at the rate of 50 miles per hour.

As a geometric example, consider the situation in which the length and width of a rectangle are increasing, so that the area of that rectangle is also increasing. The equation

$$A = b h$$

is a "formula" which gives the area of the rectangle in terms of the length and width. To say that these three quantities are changing may be interpreted as saying that the symbols A , b , and h are dependent variables of three functions, say $u()$, $v()$, and $w()$, respectively, each of which functions has independent variable t , interpreted as the number of units of time elapsed since some starting time.

The equation relating these variables then says that

$$u(t) = v(t) w(t)$$

is an identity. This identity is equivalent to

$$u(t) = \{v w\}(t),$$

and implicit differentiation yields the new identity

$$u'(t) = \{v w' + v' w\}(t),$$

i.e.

$$u'(t) = v(t) w'(t) + v'(t) w(t).$$

If we have also the information that the length, b , of the rectangle is increasing at the rate of 4 units per minute, then this geometric information translates into the statement that $v'()$ is the constant function $4()$. The information that the height, h , of the rectangle is increasing at the rate of 5 units per minute would translate into the statement that the function $w'()$ is the constant function $5()$. We then have the identity

$$u'(t) = 5 v(t) + 4 w(t).$$

Since a number in the range of values of $u'()$ may be interpreted as a rate of change of the variable A , we can see that the area of the rectangle under consideration is changing at a non-constant rate, which rate depends on what the length, b , and width, h , of the rectangle are at the particular instant under consideration.

Exercises

1. A ship leaves Nantucket at noon, heading south at 10 miles per hour. A second ship leaves the same pier at one p.m., heading east at 15 miles per hour. How fast are the ships separating at two p.m.?
2. Albany is 150 miles north of New York. A plane leaves Albany at 3 p.m., headed due west at 200 mph. A second plane leaves New York, headed due north at 300 mph. How fast are the planes separating when the second plane passes over Albany?
3. The base of a triangle is increasing at a rate of 4 units per hour. The altitude of this triangle is decreasing at a rate of 3 units per hour. At the instant when the altitude is 8 feet and the base is 10 feet, what is the rate of change of the area?
4. The radius of a sphere is increasing at a rate of 3 units per second. At the instant that the radius is 10 inches, how fast are the surface area and the volume changing?
5. A part of the plane bounded by the graphs of $y = x^2$, $y = 0$ and $x = a$ is increased by sliding the vertical line with equation $x = a$ to the right at the rate of 3 units per minute. At the instant that the area is 9 square units, how fast is the area increasing?
6. A canal running north-south is crossed by a road running east-west. The bridge carrying the road is 30 feet above the water level. At 6 a.m., a barge passes under the bridge and a truck crosses the bridge. The barge is traveling at 5 mph and the truck at 40 mph. How fast are they separating at 8 a.m.?
7. The volume of a cylinder increases at a constant rate of 6 cubic inches per second. The diameter of the base of the cylinder is increasing at a constant rate of 2 inches per second. How fast is the height of the cylinder changing at the instant that the volume is 120π cubic feet and the height is 30 feet?
8. A conical tank is positioned point downward. The tank is 10 feet deep and the top is a circle whose radius is 5 feet. If the water level drops at a constant rate of 1 foot per hour, at what rate is the water leaving the tank at the instant that the tank runs dry?
9. A light is placed atop a pole 20 feet high. A man walks away from the pole at a pace of 4 miles per hour. How fast is his shadow moving? (60 miles per hour = 88 feet per second.)

10. A bridge crosses a canal. The canal runs due east and west and the bridge carries a road running north and south. A wagon, going 10 feet per second, crosses the bridge at the instant a ship, going 5 feet per second, passes underneath. How fast are the wagon and boat separating 10 seconds after this instant, if the bridge is 20 feet above the water?

9.6. THE INVERSE OF A FUNCTION

As should be known by now, a function is a set of ordered pairs of numbers, no two of which pairs have the same first element. If that particular collection of ordered pairs has the property that no two of the pairs have the same second element, then the new collection of ordered pairs obtained by reversing the order of the ordered pairs will also be a function.

For example, if $f()$ is the function consisting of the set or ordered pairs:

$$(1, 2), (2, 3), (3, 8), \text{ and } (4, 0),$$

then the collection of ordered pairs:

$$(2, 1), (3, 2), (8, 3), \text{ and } (0, 4)$$

will also be a function.

However, if $g()$ is the function consisting of the set of ordered pairs:

$$(1, 3), (2, 3), (3, 7), \text{ and } (4, 0),$$

then the collection of ordered pairs:

$$(0, 4), (3, 1), (3, 2), \text{ and } (7, 3)$$

is not a function.

Geometrically, if $f()$ is a function, then the graph of $f()$ is the graph of the equation

$$y = f(x).$$

To say that $f()$ consists of ordered pairs, no two of which have the same first element is equivalent to saying that the graph of $f()$ is a collection of points, no two of which have the same x -coordinate. In turn, this statement is equivalent to saying that the graph of a function has the property that any vertical line will intersect this graph at most once.

The requirement that a collection of ordered pairs also have no two second elements the same is equivalent to the geometric requirement that no points on the graph of that collection have the same y -coordinate, or that any horizontal line cuts this graph at most once.

The graph of the function $\{I^3\}()$ does have this additional property that any horizontal line cuts at most once. The graph of the function $\{I^2\}()$ does not have this additional property, since any horizontal line which lies above the x -axis will actually cut the graph of $y = x^2$ twice. (For example, the horizontal line with equation $y = 4$ cuts the graph of $y = x^2$ at $(-2, 4)$ and $(2, 4)$.)

We are then examining a property which some functions have and some do not. To give this property an explicit name we have:

Definition 9.3: If $f()$ is a function, no two of the ordered pairs of which have the same second element, then the new function

$$\{f^{-1}\}(),$$

called the *inverse* of $f()$ consists of the collection of ordered pairs which results from reversing the order of the ordered pairs of $f()$.

(The symbol $\{f^{-1}\}()$ is read “ f -inverse.”)

Some immediate consequences of this definition follow.

Theorem 9.3:

If the function $f()$ has an inverse, $\{f^{-1}\}()$, and t is in the domain of $f()$, then

$$\{f^{-1}\}(f(t)) = t.$$

Remark: This theorem says that a rule of $\{f^{-1}\}()$ undoes that which a rule of $f()$ does.

Proof of 9.3:

- | | |
|--|--|
| 1. t is in the domain of $f()$ | by hypothesis. |
| 2. $(t, f(t))$ is an ordered pair of $f()$ | by step 1 and the definition of function. |
| 3. $(f(t), t)$ is an ordered pair of $\{f^{-1}\}()$ | by step 2 and the definition of inverse. |
| 4. $(f(t), f^{-1}(f(t)))$ is an ordered pair of $\{f^{-1}\}()$ | by the definition of function. |
| 5. $\{f^{-1}\}(f(t)) = t$ | by steps 3 and 4 and the fact that no two ordered pairs of a function can have the same first element. |

Theorem 9.4:

If the function $f()$ has an inverse, $\{f^{-1}\}()$, and r is in the domain of $\{f^{-1}\}()$, then

$$f(f^{-1}(r)) = r.$$

Remark: This theorem says that a rule of $f()$ undoes that which a rule of $\{f^{-1}\}()$ does.

Proof of 9.4:

- | | |
|--|---|
| 1. r is in the domain of $\{f^{-1}\}()$ | by hypothesis. |
| 2. $(r, f^{-1}(r))$ is an ordered pair of $\{f^{-1}\}()$ | by step 1 and the definition of function. |

3. $(f^{-1}(r), r)$ is an ordered pair of $f()$ by step 2 and the definition of inverse.
4. $(f^{-1}(r), f(f^{-1}(r)))$ is an ordered pair of $f()$ by the definition of function.
5. $f(f^{-1}(r)) = r$ by steps 3 and 4 and the fact that no two ordered pairs of a function have the same first element.

These theorems show that the composites

$$\{f(f^{-1})\}() \quad \text{and} \quad \{f^{-1}(f)\}()$$

are part or all of the identity function $I()$. It should also be noted that the domain of $f()$ is the range of values of $\{f^{-1}\}()$ and that the range of values of $f()$ is the domain of $\{f^{-1}\}()$.

As an example, we let $f() = \{3I + 2\}()$. $f()$ is then the function specified by the equation

$$f(t) = 3t + 2,$$

and a rule of $f()$ is: "Multiply by 3, then add 2."

A rule undoing that which the given rule of $f()$ does would then be: "Subtract 2, then divide by 3." This is a rule of the function $\left(\frac{I-2}{3}\right)()$, which is the same function as $\left\{\frac{1}{3}I - \frac{2}{3}\right\}()$. If we call this new function $g()$, then $g()$ is the function specified by the equation

$$g(r) = \frac{1}{3}r - \frac{2}{3}.$$

Examining the composites, we have

$$\begin{aligned} f(g(r)) &= f\left(\frac{1}{3}r - \frac{2}{3}\right) \\ &= 3\left(\frac{1}{3}r - \frac{2}{3}\right) + 2 \\ &= r - 2 + 2 \\ &= r. \\ g(f(t)) &= g(3t + 2) \\ &= \frac{1}{3}(3t + 2) - \frac{2}{3} \\ &= t + \frac{2}{3} - \frac{2}{3} \\ &= t. \end{aligned}$$

Then $\{f(g)\}()$ and $\{g(f)\}()$ are both the identity function $I()$, so that $f()$ and $g()$ are inverses of each other.

Pursuing this example geometrically, the graph of $f()$ is the graph of the equation

$$y = 3x + 2.$$

This straight line is also the graph of

$$3x - y = -2.$$

The graph of the function $g()$ is the graph of the equation:

$$y = -\frac{1}{3}x - \frac{1}{3},$$

which is also the graph of

$$3y - x = 2.$$

But the second equation can be produced from the first merely by interchanging the variables x and y , so that the equation

$$3x - y = -2$$

gives two functions, each of which is the inverse of the other.

Because of the relation between a function $f()$ and its inverse $g()$, the definition of composite, and the definition of an identity, the equation

$$\{f(g)\}(x) = x$$

must be an identity.

The situation in which an equation is considered as an identity is precisely that encountered in the study of implicit differentiation. Then, provided that the functions involved have derivatives, if $f()$ and $g()$ are inverses of each other, implicit differentiation of the identity

$$\{f(g)\}(x) = x$$

yields the new identity

$$f'(g(x)) g'(x) = 1$$

by the "chain rule."

This new identity also says

$$g'(x) = \frac{1}{f'(g(x))},$$

and we need not worry about a possible division by 0 since the identity

$$f'(g(x)) g'(x) = 1$$

says that neither of the numbers $f'(g(x))$ and $g'(x)$ can be 0.

We are led to the conclusion

Theorem 9.5:

If the function $f()$ has an inverse $g()$, if a is a number in the domain of $g'()$ such that $g'(a) \neq 0$, $g(a)$ is in the domain of $f'()$, and $f'(g(a)) \neq 0$, then

$$g'(a) = \frac{1}{f'(g(a))}.$$

Exercises

Find the inverses, where they exist, of the following functions:

1. $\{I^3\}()$.
2. $\{I^4\}()$.
3. $\{[I^3 - 2]^{1/3}\}()$.

4. $\{3 - I\}()$.
5. $\{I^{-3}\}()$.
6. $\{1/I\}()$.
7. The function specified by $f(x) = x^2, x \leq 0$.
8. The function specified by $f(x) = x^2, x > 0$.
9. The function specified by $f(x) = x^2, -1 < x < 1$.
10. Find the derivatives of each of the inverses found in the preceding nine exercises.
11. What was left unsaid in step 4 of the proofs of Theorems 9.3 and 9.4?

9.7. REVIEW OF DIFFERENTIATION

We have thus far proved the following differentiation theorems:

1. $\{I^n\}'() = \{n I^{n-1}\}()$, where n is any rational number.
2. $\{f + g\}'() = \{f' + g'\}()$.
3. $\{kf\}'() = \{kf'\}()$, where k is any constant.
4. $\{fg\}'() = \{f g' + f' g\}()$.
5. $\left\{\frac{f}{g}\right\}'() = \left\{\frac{g f' - f g'}{g^2}\right\}()$.
6. $\{f(g)\}'() = \{f'(g) g'\}()$.
7. $\{f^{-1}\}'(a) = \left\{\frac{1}{f'(f^{-1})}\right\}(a)$, whenever there are no 0 values.

As a special case of the sixth theorem, if $f() = \{I^n\}()$, where n is a rational number,

$$\{g^n\}'() = \{n g^{n-1} g'\}().$$

These theorems allow us to differentiate any function whose rule involves any of the algebraic operations of addition, subtraction, multiplication, division, raising to rational number powers, or extracting positive integer roots.

For example, the function specified by the equation

$$f(t) = \left[\frac{t^5 - 3t^{1/3} + 1}{(4t + 7)^{1/3}} \right]^{1/2}$$

is a composite of $\{I^{1/2}\}()$ with a function which is the quotient of two functions. The numerator of this quotient is the function specified by

$$t^5 - 3t^{1/3} + 1$$

and the denominator of this quotient is the composite of $\{I^{1/3}\}()$ with the function specified by

$$4t + 7.$$

Using the various theorems, we have

$$\begin{aligned}
 & \left\{ \left[\frac{I^5 - 3I^{1/3} + 1}{(4I + 7)^{1/3}} \right]^{1/2} \right\}' () \\
 = & \left\{ \frac{1}{2} \left[\frac{I^5 - 3I^{1/3} + 1}{(4I + 7)^{1/3}} \right]^{-1/2} \left[\frac{I^5 - 3I^{1/3} + 1}{(4I + 7)^{1/3}} \right]' \right\} () \\
 = & \left\{ \frac{1}{2} \left[\frac{I^5 - 3I^{1/3} + 1}{(4I + 7)^{1/3}} \right]^{-1/2} \right. \\
 & \left. \frac{(4I + 7)^{1/3} [I^5 - 3I^{1/3} + 1]' - [I^5 - 3I^{1/3} + 1] [(4I + 7)^{1/3}]'}{(4I + 7)^{2/3}} \right\} () \\
 = & \left\{ \frac{1}{2} \left[\frac{I^5 - 3I^{1/3} + 1}{(4I + 7)^{1/3}} \right]^{-1/2} \right. \\
 & \left. \frac{(4I + 7)^{1/2} (5I^4 - I^{-2/3}) - (I^5 - 3I^{1/3} + 1) \frac{1}{3} (4I + 7)^{-2/3} (4)}{(4I + 7)^{2/3}} \right\} () \\
 = & \left\{ \frac{1}{6} \left[\frac{56I^5 + 105I^4 - 4 - 21I^{-2/3}}{(I^5 - 3I^{1/3} + 1)^{1/2} (4I + 7)^{7/6}} \right] \right\} ()
 \end{aligned}$$

Exercises

Find the first derivative of the function specified by:

1. $((2t + 10)^{1/3} + 2)^3 + 56t$.
2. $[s^2(3s + 5)^{1/3}]^6$.
3. $(r^3 + 3r - 1)^4 (2r - 1)^6$.
4. $[u^{1/5} + u^{-1/5}]^{3/4}$.
5. $\{(v^3 + 1)^3 + 2\}^4 - 4\}^5$.
6. $\left(\frac{3w - 2}{4w - 5}\right)^{-1/4}$.
7. $\left(\frac{x^7 + 3x^4 - 2x + 5x^{-2}}{4x + 1}\right)^3$.
8. $(2y + 1)^{1/2} (3y + 2)^{1/3} (4y + 5)^{1/3}$.
9. $\{(z^{1/5} - 1)^{1/3} + 1\}^5$.
10. Find the second derivative of the functions given in exercises 1-9.

9.8. ANTI-DIFFERENTIATION

With the notation $\{Df\}()$ rather than $f'()$ to indicate the derivative of the function $f()$, we can write theorems on anti-differentiation which result immediately from the theorems on differentiation.

Of these possible anti-differentiation theorems, only five will be of concern to us. Before we state these theorems, we must look more closely at the notation used for anti-derivatives. If $f()$ is a function, we write $\{Df\}()$ for that one function which is the derivative of $f()$, and we have written $\{D^{-1}f\}()$ for any function whose derivative is $f()$.

That is, if $g()$ is a function such that $g'() = f()$, then we write $g() = \{D^{-1}f\}()$. But, to take a simple example, there are many functions whose derivative is the function $\{2I\}()$. Among these are $\{I^2\}()$, $\{I^2 + 3\}()$, $\{I^2 - 7/2\}()$, and $\{I^2 + k\}()$, where $k()$ is any constant function. As a result of this situation, it is not correct to speak of "*the* anti-derivative," and one can only say "*an* anti-derivative." If we use the notation $\{D^{-1}f\}()$ to represent *an* anti-derivative of the function $f()$, then an equation such as

$$\{D^{-1}f\}() = \{D^{-1}g\}()$$

does not say "the anti-derivative of $f()$ is the same function as the anti-derivative of $g()$," since one cannot use definite articles with anti-derivatives. This equation merely says that one of the many anti-derivatives of $f()$ is also one of the many anti-derivatives of $g()$, or, in simpler words, that $\{D^{-1}f\}()$ and $\{D^{-1}g\}()$ are functions which may not themselves be the same, but which have the same derivative.

If two functions, $u()$ and $v()$ have the same derivative—that is, $u'()$ and $v'()$ are the same function—then the function $\{u - v\}()$ must have, as its derivative, $\{u' - v'\}()$, which is a constant function whose range of values consists of the number 0. But the only functions whose derivatives are the zero function are constant functions, so that we can conclude that $\{D^{-1}u\}()$ and $\{D^{-1}v\}()$, if they do differ, differ only by a constant function.

But then the equation

$$\{D^{-1}f\}() = \{D^{-1}g\}()$$

is really the equation

$$\{D^{-1}f\}() = \{D^{-1}g + C\}(),$$

where $C()$ is some constant function.

We have then a wierd situation in which an equation involving anti-derivatives does not seem to say what it says. We got into such a situation by using a notation for anti-derivatives—writing $\{D^{-1}f\}()$ —which seems to say that *an* anti-derivative is *the* anti-derivative. We will get ourselves out of this situation by never using the symbolism $\{D^{-1}f\}()$ without also using the words "an anti-derivative of $f()$."

In this manner, we have, corresponding to our first differentiation theorem of the preceding section, the statement

$$\left\{ \frac{1}{p+1} I^{p+1} \right\}() \text{ is an anti-derivative of } \{I^p\}(),$$

whenever p is a rational number."

In order to avoid an attempted division by 0, we must also make the provision that the exponent, p , is a rational number different from -1 . This gives us:

Theorem 9.6:

(replacing Theorem 5.18)

If p is a rational number other than -1 , and $\{D^{-1}I^p\}()$ is an anti-derivative of $\{I^p\}()$, then there is a constant, C , such that

$$\{D^{-1}I^p\}() = \left\{ \frac{1}{p+1} I^{p+1} + C \right\}().$$

The next theorem on differentiation says that the derivative of the sum of two functions is the sum of their derivatives. The corresponding anti-derivative theorem is

Theorem 9.7:

If $\{D^{-1}f\}()$ is an anti-derivative of $f()$, $\{D^{-1}g\}()$ is an anti-derivative of $g()$, and $\{D^{-1}(f+g)\}()$ is an anti-derivative of $\{f+g\}()$, then there is a constant, C , such that

$$\{D^{-1}(f+g)\}() = \{D^{-1}f + D^{-1}g + C\}().$$

The third theorem on differentiation says that the derivative of a constant multiple of a function is the product of that constant and the derivative of the function. We have

Theorem 9.8:

If $\{D^{-1}f\}()$ is an anti-derivative of $f()$, k is a constant, and $\{D^{-1}(kf)\}()$ is an anti-derivative of $\{kf\}()$, then there is a constant, C , such that

$$\{D^{-1}(kf)\}() = \{k(D^{-1}f) + C\}().$$

We will now apply these anti-derivative theorems to the anti-differentiation of polynomial functions. Suppose we consider the function specified by

$$3t^5 - 5t^3 + 4t^2 - 8t + 1$$

It is certainly a reasonable guess that an anti-derivative of this polynomial function will be another polynomial function. From Theorem 9.7, we see that we need only look at each term separately in the above expression.

Our first chore is then to find an anti-derivative of the function specified by the expression

$$3t^5.$$

Without trying to remember the explicit statement on the anti-differentiation of $\{3I^5\}()$, we should know that the differentiation of some function specified by an expression involving t^6 will give us a function specified by another expression involving t^5 . Let us make a first guess by saying that $G_1()$, given by

$$G_1(t) = t^6,$$

is an anti-derivative of the function given by $3t^5$.

But $G_1'()$ is specified by

$$G_1'(t) = 6t^5,$$

not by the expression $3t^5$, so that our guess, $G_1()$, was wrong. But just how was this guess wrong? We found that $G_1'()$ gave us a 6 where we wanted a 3. Then, if we make a second guess, $G_2()$, specified by

$$G_2(t) = \frac{1}{2} G_1(t) = \frac{1}{2} t^6,$$

we will have

$$G_2'(t) = 3t^5,$$

and $G_2()$ will be an anti-derivative of $\{3t^5\}()$.

For the second term, $-5t^3$, a reasonable first guess for an anti-derivative would be

$$G_1(t) = t^4,$$

which gives

$$G_1'(t) = 4t^3.$$

We want the constant in this derivative to be -5 rather than 4 , so that our second guess, $G_2()$, would be specified by

$$G_2(t) = -\frac{5}{4} G_1(t) = -\frac{5}{4} t^4,$$

which gives

$$G_2'(t) = -5t^3.$$

In like manner we find that we need to differentiate the function specified by

$$\frac{4}{3} t^3,$$

in order to have the function specified by $4t^2$ as a derivative; we need to differentiate the function specified by

$$-4t^2,$$

in order to have the function specified by $-8t$ as a derivative; and we need to differentiate the function specified by

$$t,$$

in order to have the constant function specified by 1 as a derivative.

Consequently, an anti-derivative of the function specified by

$$3t^5 - 5t^3 + 4t^2 - 8t + 1$$

will be the function specified by

$$\frac{1}{2} t^6 - \frac{5}{4} t^4 + \frac{4}{3} t^3 - 4 t^2 + t$$

and for any anti-derivative, there will be some constant, C , for which this anti-derivative will be the function specified by

$$\frac{1}{2} t^6 - \frac{5}{4} t^4 + \frac{4}{3} t^3 - 4 t^2 + t + C.$$

Exercises

Find an anti-derivative of the function specified by:

1. $2r - 1$.
2. $s^2 + 3s - 4$.
3. $5t^3 - 4t^2 + 3t - 2$.
4. $-2u^4 - 7u^2 + 3u + 1$.
5. $14v^5 + 19v^4 - 3v^3 + 3v^2 - 11v + 43$.
6. $(w^2 - 3w^3)^2$.
7. $4x^7 - 5x^5 + x^3 - 2x^2 + 7x - 4$.
8. $y^8 + 7y^2 - 3y$.
9. $z^{15} + 2z^3$.
10. $3r^{1/2} - 2$.
11. $s^{3/4} - 3s + 4s^{-2}$.
12. $t^{-3/5} + 2t - 1$.
13. $-4u^{-5} + 7u^{-4} + 3u^{-3}$.
14. $\frac{v^3 + 2v^2 - 4}{v^5}$.
15. $2 \mid x \mid$.

9.9. ANTI-DIFFERENTIATION AND COMPOSITES

Corresponding to the differentiation theorem for the composite of two functions—the “chain rule,” we have the following anti-differentiation theorem:

Theorem 9.9:

If $\{D^{-1}f\}$ is an anti-derivative of $f()$ and $\{D^{-1}\{f(g)\}\}g'()$ is an anti-derivative of $\{f(g)g'\}()$, then there is a constant, C , such that

$$\{D^{-1}\{f(g)g'\}\}() = \{\{D^{-1}f\}(g) + C\}().$$

In other notation, this theorem says that if $F()$ is an anti-derivative of $f()$, then $\{F(g)\}()$ will be an anti-derivative of $\{f(g) \cdot g'\}()$.

To use this theorem in finding an anti-derivative of a given function, one must regard an expression specifying the given function as an expression of the form

$$f(g(t))g'(t),$$

decide which parts of the given expression are the two factors, and then find an anti-derivative of the “outside function” $f()$. For example, in order to find an anti-derivative of the function specified by the expression

$$(t^2 + 1)^{1/2}t,$$

it is reasonable to consider the more complicated factor as being an expression for the composite $f(g(t))$. If

$$f(g(t)) = (t^2 + 1)^{1/2},$$

then $f()$ and $g()$ might be a number of functions, but again, it is plausible to consider $g()$ as the function specified by

$$g(t) = t^2 + 1,$$

in which case we will have

$$g'(t) = 2t$$

as an equation specifying $g'()$.

With the "inside function" so specified, then the "outside function," $f()$, of the composite must be given by

$$f(z) = z^{1/2}.$$

If we had, for our original expression

$$(t^2 + 1)^{1/2} (2t),$$

then this would be exactly of the form

$$f(g(t)) g'(t),$$

and, for $F()$ any anti-derivative of $f()$, the desired anti-derivative would be the composite $\{F(g)\}()$.

An anti-derivative of $\{t^{1/2}\}()$ is $\{\frac{2}{3} t^{3/2}\}()$, by Theorem 9.6, so that we know that an anti-derivative of the function specified by

$$(t^2 + 1)^{1/2} (2t)$$

will be

$$\frac{2}{3} (t^2 + 1)^{3/2}.$$

We are however seeking an anti-derivative of the function specified by an expression which is 1/2 of the one we have found. But then Theorem 9.7 tells us that an anti-derivative will be 1/2 of the anti-derivative we found, so that

$$\frac{1}{3} (t^2 + 1)^{3/2}$$

is an expression specifying an anti-derivative of the function given by

$$(t^2 + 1)^{1/2} t.$$

As a second method of finding an anti-derivative of the function specified by

$$(t^2 + 1)^{1/2} t,$$

we consider making a succession of reasonable guesses, improving each guess over the preceding until we arrive at a correct anti-derivative.

Examining the given expression, the first thing to be noted (usually unconsciously) is that this expression does not specify a simple polynomial function, an anti-derivative of which could be found immediately. If we ask ourselves the question: "Is there one simple-appearing thing in the expression which prevents our having a polynomial?," the answer is immediate: "If the exponent $1/2$ were not in the expression, we would have a polynomial." Ignoring everything except the exponent $1/2$, we would have a function for which something with exponent $3/2$ would be an anti-derivative. Then, a not unreasonable first guess would be given by

$$G_1(t) = (t^2 + 1)^{3/2}.$$

Use of the "chain rule" for differentiation gives

$$\begin{aligned} G_1'(t) &= \frac{3}{2} (t^2 + 1)^{1/2} (2t) \\ &= 3 (t^2 + 1)^{1/2} t. \end{aligned}$$

The derivative of our first guess is 3 multiplied by the given function. Thus, if the second guess is given by

$$G_2(t) = \frac{1}{3} G_1(t) = \frac{1}{3} (t^2 + 1)^{3/2},$$

we will have

$$\begin{aligned} G_2'(t) &= \frac{1}{3} (3) (t^2 + 1)^{1/2} t \\ &= (t^2 + 1)^{1/2} t, \end{aligned}$$

and we have found an anti-derivative of the given function.

Both of the methods of anti-differentiation presented here involve a certain amount of intelligent guessing. It is always a problem to decide just what to guess in such a situation. Let us re-examine the same function, the one given by

$$(t^2 + 1)^{1/2} t,$$

and see if we can trace through the general ideas of the differentiation of composites, in order to see both where the guess should be made and what the guess should be.

We want to have

$$f(g(t)) g'(t) = (t^2 + 1)^{1/2} t.$$

We have started with an expression (on the right-hand side of this equation) which is to give the derivative of some function, and it is this function which we seek. That is, there is a function, $F()$, unknown to us, for which

$$F'(t) = (t^2 + 1)^{1/2} t.$$

Now if $g()$ is any function whatsoever, the derivative of the composite, $\{F(g)\}()$, is given by

$$\begin{aligned} \{F(g)\}'(t) &= F'(g(t)) g'(t) \\ &= [(g(t))^2 + 1]^{1/2} g(t) g'(t). \end{aligned}$$

Since this last equation is true whatever function $g()$ may be, it is most reasonable to choose $g()$ to be a function which simplifies the expression we have. But the complicating factor in the expression is the exponent $1/2$. If then, we choose $g()$ so that the expression

$$[g(t)]^2 + 1$$

is a "perfect square," then we will remove this square root. The simplest such "perfect square" would be the expression t^2 , and if we assume that all the numbers involved are positive, we have

$$\begin{aligned} [g(t)]^2 + 1 &= t^2 \\ g(t) &= [t^2 - 1]^{1/2} \\ g'(t) &= \frac{1}{2} (t^2 - 1)^{-1/2} (2t) \\ &= \frac{t}{(t^2 - 1)^{1/2}}. \end{aligned}$$

$$\begin{aligned} \{F(g)\}'(t) &= [(t^2 - 1) + 1]^{1/2} (t^2 - 1)^{1/2} \left[\frac{t}{(t^2 - 1)^{1/2}} \right] \\ &= t^2. \end{aligned}$$

But then it is easy to find an anti-derivative of $\{F(g)\}()$, namely that given by

$$\{F(g)\}(t) = \frac{1}{3} t^3.$$

By the definition of composite, we then have

$$F(g(t)) = \frac{1}{3} t^3.$$

Using the equation specifying $g()$, we have

$$\begin{aligned} g(t) &= (t^2 - 1)^{1/2} \\ [g(t)]^2 &= t^2 - 1 \\ t &= \{[g(t)]^2 + 1\}^{1/2}, \end{aligned}$$

so that substitution in the expression for $F(g(t))$ gives

$$F(g(t)) = \frac{1}{3} \{[g(t)]^2 + 1\}^{3/2}.$$

But merely substituting a new variable for $g()$ gives the equation

$$F(z) = \frac{1}{3} (z^2 + 1)^{3/2},$$

and the function $F()$ is an anti-derivative of the function specified by

$$(t^2 + 1)^{1/2} t.$$

Each of these three methods of attacking an anti-differentiation problem involves some measure of uncertainty along the way, but this uncertainty can always be dispelled by simply differentiating the function which we claim is an anti-derivative.

For algebraic functions, the method of making a guess which involves the most complicated part of the function yields fairly rapid results when such results are there to be obtained. As an example, to find an anti-derivative of the function specified by

$$f(t) = \frac{4t^2}{(t^3 + 1)^3},$$

we rewrite this equation in the form

$$f(t) = (t^3 + 1)^{-3} (4t^2).$$

A first guess is given by

$$G_1(t) = (t^3 + 1)^{-2},$$

for which we have

$$\begin{aligned} G_1'(t) &= -2 (t^3 + 1)^{-3} (3t^2) \\ &= (t^3 + 1)^{-3} (-6t^2). \end{aligned}$$

A second, and final guess then must be

$$G_2(t) = -\frac{2}{3} G_1(t) = -\frac{2}{3} (t^3 + 1)^{-2}.$$

Exercises

Find an anti-derivative of the function specified by:

1. $\sqrt{3 - 4r}$.
2. $(1 + 5s)^{-1/2}$.
3. $t(t^2 + 4)^{-3}$.
4. $u^2(u^3 - 1)^{1/3}$.
5. $(v + 1)(v^2 + 2v + 3)^{-2}$.
6. $(w^2 + 1)^{1/4} w$.
7. $x^{1/2} (x^{3/2} + 1)^{-3}$.
8. $y^{-2} (2y^{-3} + 41)^{1/3}$.
9. $z^{-5} (3z^{-4} + 1)^{15}$.

Evaluate the following integrals:

$$10. \int_{-1}^1 r(r^2 + 3)^{1/2} dr.$$

$$11. \int_0^2 s^2(s^3 - 8)^{1/3} ds.$$

$$12. \int_{-1}^1 \frac{t + 3}{(t^2 + 6t + 10)^2} dt.$$

13. Find the area bounded by the graphs of $y = x^2 (x^3 + 1)^{-3}$, $y = 0$, $x = 1$, and $x = 2$.

9.10. ANTI-DIFFERENTIATION BY PARTS

Corresponding to the theorem on the derivative of the product of two functions, we have the following anti-derivative theorem:

Theorem 9.10:

If $f()$ and $g()$ are functions, with corresponding derivatives $f'()$ and $g'()$, if $\{D^{-1}\{fg'\}\}()$ is an anti-derivative of $\{fg'\}()$ and $\{D^{-1}\{f'g\}\}()$ is an anti-derivative of $\{f'g\}()$, then there is a constant, C , such that

$$\{D^{-1}\{fg'\}\}() = \{fg - D^{-1}\{f'g\} + C\}().$$

In effect, this theorem says that if $g()$ is an anti-derivative of $g'()$, then an anti-derivative of $\{fg'\}()$ will be $\{fg - h\}()$, where $h()$ is an anti-derivative of $\{f'g\}()$. This theorem is applied to find anti-derivatives of the product of two functions, where we cannot use the results of the preceding section, but when we suspect that this product may be one of the two terms involved in the differentiation of the product of some two functions.

For example, if we want to find an anti-derivative of the function given by the expression

$$t(t+1)^{1/2},$$

we would expect that this expression is of the form

$$f(t)g'(t)$$

and we must decide which one of the factors in the given expression we will take as giving $f(t)$ and which we will take as giving $g'(t)$. Whichever choice we make, we must then find $f'(t)$ and $g(t)$, and seek an anti-derivative of the function given by the product $f'(t)g(t)$.

If we choose

$$f(t) = t \quad \text{and} \quad g'(t) = (t+1)^{1/2},$$

we then must have

$$f'(t) = 1$$

and we may have

$$g(t) = \frac{1}{2}(t+1)^{3/2},$$

so that, for this choice of the "parts," we need to seek an anti-derivative of the function specified by

$$f'(t)g(t) = \frac{2}{3}(t+1)^{3/2}.$$

If we choose

$$f(t) = (t+1)^{1/2} \quad \text{and} \quad g'(t) = t,$$

then we must have

$$f'(t) = \frac{1}{2}(t+1)^{-1/2}$$

and we may have

$$g(t) = \frac{1}{2}t^2,$$

so that, for this choice of the "parts," we need to seek an anti-derivative of the function specified by

$$f'(t) g(t) = \frac{1}{4} t^2 (t + 1)^{-1/2}.$$

It should be clear that the first choice of "parts" is preferable to the second, since an anti-derivative of the function specified by $\frac{2}{3} (t + 1)^{3/2}$ is the function given by

$$\frac{4}{15} (t + 1)^{5/2}.$$

The theorem then says that an anti-derivative of the function, given by

$$F(t) = f(t) g'(t) = t (t + 1)^{1/2},$$

will be given by

$$f(t) g(t) - \frac{4}{15} (t + 1)^{5/2} = \frac{2}{3} t (t + 1)^{3/2} - \frac{4}{15} (t + 1)^{5/2} + C,$$

where C is some constant.

In general, the efficacy of anti-differentiation by parts is dependent on the choice of the "parts." We must choose two parts of our given product, one of these to be differentiated, and the other to be anti-differentiated. We then have to find an anti-derivative of a new product. If all these steps can be done, then this method can be used, but if one of them (particularly the last) cannot be done readily, then this method cannot be used. As a "rule of thumb," however, it should be noted that many very fine mathematicians say "when in doubt, use anti-differentiation by parts."

When the Fundamental Theorem is used to evaluate integrals, our theorem on anti-differentiation by parts yields

$$\int_a^b f(t) g'(t) dt = f(t) g(t) \Big|_{t=a}^{t=b} - \int_a^b f'(t) g(t) dt.$$

Exercises

Evaluate:

1. $\int_{-1}^1 r^3(1 + r^2)^{-1/2} dr$. (Hint: $r^2 = rr$.)
2. $\int_0^1 s^5(2 + s^3)^{-3} ds$.
3. $\int_0^1 t^7(1 + t^4)^{1/3} dt$.
4. $\int_0^1 u^9(a + u^5)^{-4} du$, where a is a positive constant.
5. $\int_0^1 v^{2n-1}(a + v^n)^p dv$, where a is a positive constant and p is a positive rational number.
6. $\int_0^1 w^3(1 + w^2)^{1/2} dw$.

7. If n is a natural number, how many anti-differentiations by parts must be performed in order to find an anti-derivative of the function specified by

$$x^n(a + x)^p,$$

where p is some rational number other than 0?

9.11. THE CALCULUS OF ALGEBRAIC FUNCTIONS

Differential calculus involves the differentiation of functions. Integral calculus involves the integration of functions. By the Fundamental Theorem of Calculus, we may find integrals by first finding anti-derivatives. All of these things we can now do for every polynomial function.

An algebraic function is one which has a rule which involves only the algebraic operations of addition, subtraction, multiplication, division, and raising to rational number powers. With the theorems we have available, we can differentiate any algebraic function.

Since integration also involves anti-differentiation, we can find the integral of any function for which we can find an anti-derivative. But we cannot find an anti-derivative for every algebraic function—at least we cannot yet find such an anti-derivative among the algebraic functions. Thus, a continuation of the study of calculus will involve the investigation of new, non-algebraic functions.

TRIGONOMETRIC FUNCTIONS

10.1. SINES AND COSINES OF NUMBERS

In order to create a new class of functions, which will be called trigonometric functions, it is necessary to change one's point of view of such operations as "taking the cosine of an angle." We will adopt as undefined, yet instinctively meaningful words, the two words "angle" and "arc-length-along-a-circle." Since we want to define a function whose rule will be "take cosine," and since we want that which we have defined to be a function, it is necessary, from the definition of the word function, that we know how to take the cosine of a number.

In the elementary trigonometry of right triangles, we have the concept of cosine (and sine) *of angles* (but not of *numbers*) presented in the following manner: if ABC is a right triangle (Fig. 10.1) with the right angle at the point C , and if θ is the angle between the line segments AB and AC , then $\sin \theta$ (which is read as "the sine of θ ") is the number

$$\frac{\text{length of } BC}{\text{length of } AB}$$

and $\cos \theta$ (read as "the cosine of θ ") is the number

$$\frac{\text{length of } AC}{\text{length of } AB}$$

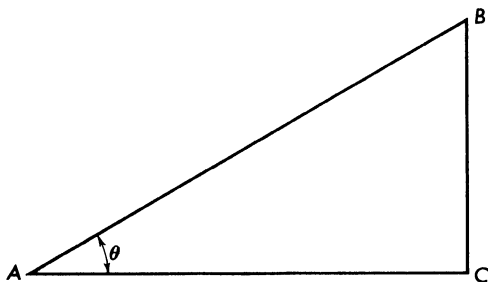


Fig. 10.1

These two quotients would certainly appear simpler if the denominators, the length of the hypotenuse of the triangle, were the number 1. But if we were to have a number of right triangles such as is shown in Fig. 10.1, and their hypotenuses were all one unit long, then we could arrange them

so that one vertex is always at the same point and one side is always along a horizontal line. (Fig. 10.2)

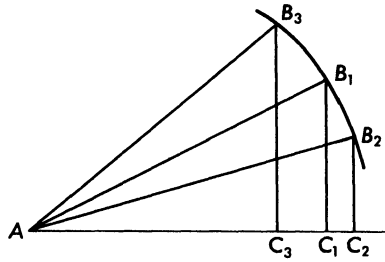


Fig. 10.2

For such an arrangement the second vertex of each triangle (the points B_1 , B_2 , B_3 , etc.) would all be one unit from the single point A , and would thus lie on a circle, centered at A , of radius 1.

Now, let us look at this same situation, reversing the order of our steps. Suppose we *start* with a circle of radius 1, with a horizontal line running through the center of that circle. Since we have a horizontal line, we might as well have a vertical line also, and it seems plausible to have these two lines intersect at the center of the circle. Further, since we are somewhat at home with plane coordinate systems, we will let these vertical and horizontal lines be the axes of the coordinate system, and our circle will be the graph of the equation

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

(See Fig. 10.3.)

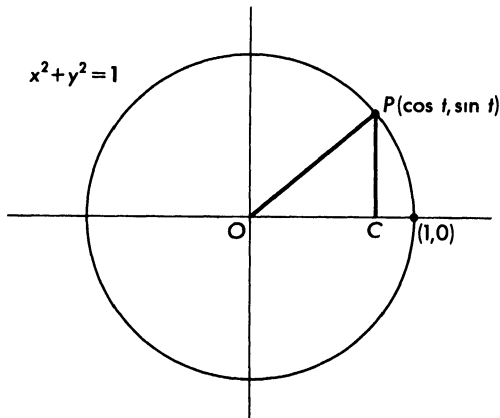


Fig. 10.3

On Fig. 10.3, the line segment OP has length 1, and the angle between this line segment and the right-hand part of the horizontal axis has, for its cosine, the length of the line segment CO and, for its sine, the length of

the line segment PC . But these two lengths are, respectively, the x and y -coordinates of the point P .

We also note that, since the radius of this circle is 1, the length of its circumference is 2π , so that

$$\frac{\text{length of arc from } (1, 0) \text{ to } P}{2\pi}$$

and

$$\frac{\text{number of degrees in the angle between } OP \text{ and } OC}{360}$$

are the same number. We may then set up a one-to-one correspondence between all possible angles from the line segment joining the origin to $(1, 0)$ to a line joining the origin to some point on the circle, and the interval of numbers: all t , $0 \leq t < 2\pi$. Under this correspondence, the number t corresponds to the angle formed at the origin by the right-hand half of the x -axis and the line from the origin to that point whose distance from $(1, 0)$, measured in a counter-clockwise manner along the circumference of this circle, is the number t . Thus the number π corresponds to an angle of 180° , the number $\pi/2$ corresponds to an angle of 90° , etc.

We may now extend the idea of this correspondence in the following manner: given any non-negative number t , there is precisely one point on this "unit circle" whose distance from $(1, 0)$, measured in a counter-clockwise manner along the circumference, is that number t . We then make the following definition:

Definition 10.1: If t is a non-negative number, then $\cos t$ is the x -coordinate of that point on the graph of $x^2 + y^2 = 1$ whose distance from $(1, 0)$, measured in a counter-clockwise manner along the circumference of that circle, is t ; $\sin t$ is the y -coordinate of that same point. If t is a negative number, then

$$\begin{aligned}\sin t &= -\sin(-t), \\ \cos t &= \cos(-t).\end{aligned}$$

Since the point $(1, 0)$ is at a distance of 0 units from the point $(1, 0)$, $\sin 0 = 0$ and $\cos 0 = 1$. The point $(0, 1)$ is at a distance—along the circumference, in a counter-clockwise manner—of $\pi/2$ units from the point $(1, 0)$, so that $\cos(\pi/2) = 0$ and $\sin(\pi/2) = 1$. The reader should verify that the following table of corresponding values of t , $\cos t$, and $\sin t$, are all valid:

t	$\cos t$	$\sin t$
$\pi/6$	$1/2$	$1/2\sqrt{3}$
$\pi/4$	$1/\sqrt{2}$	$1/\sqrt{2}$
$\pi/3$	$\sqrt{3}/2$	$1/2$
π	-1	0
$3\pi/2$	0	-1

Exercises

1. Use the definition of cosine and sine of a number to find $\cos t$ and $\sin t$, when t is the number:
 $2\pi, -\pi, 3\pi, 7\pi, n\pi$ (where n is an integer),
 $2\pi/3, 3\pi/4, 5\pi/6, 7\pi/6, 5\pi/4, 4\pi/3, 5\pi/3, 7\pi/4, 11\pi/6$.
2. Construct a table showing the relation between the numbers t , $\cos t$, and $\sin t$, when t has the values:
 $-4\pi, -3\pi, -2\pi, -3\pi/2, -5\pi/4, -\pi, -5\pi/6, -3\pi/4, -2\pi/3, -\pi/2$.
3. On coordinate paper, construct, as accurately as possible, a circle. Using the definition of cosine and sine of a number, find the cosine and sine of the following numbers: 1, 2, 3, 4, 5, 6, 7, 30, 45, 90.

10.2. THE COSINE AND SINE FUNCTIONS

We now have new operations which can be performed on a number, namely, taking the cosine of that number and taking the sine of that number. Since we have defined these operations so that each of them may be performed on any number whatsoever, we may use the operations as rules of functions whose domains will be the set of all numbers. We then will define:

Definition 10.2: The function $\cos()$ is the set of ordered pairs of the form $(t, \cos t)$, where t is any number.

Definition 10.3: The function $\sin()$ is the set of ordered pairs of the form $(t, \sin t)$, where t is any number.

From these definitions, we see that the domain of both of these new functions is the set of all numbers. Since the numbers $\cos t$ and $\sin t$ are x or y coordinates of points on the circle, centered at the origin, of radius 1, we must have $-1 \leq \cos t \leq 1$ and $-1 \leq \sin t \leq 1$, so that the ranges of values of both these new functions are the interval:

$$\text{The set of all } z, \quad -1 \leq z \leq 1.$$

There are many relations between these two functions, but the most used one is given in:

Theorem 10.1:

$$\{\sin^2() + \cos^2()\}() = U().$$

(The reader is reminded that the constant function $U()$ has, as its range of values, the set consisting of the single number, 1.)

Proof of 10.1:

If t is any number whatsoever, by the definition (10.1), $(\cos t, \sin t)$ are the coordinates of a point on the graph of $x^2 + y^2 = 1$.

By definition of the graph of an equation, we must have

$$\cos^2 t + \sin^2 t = 1.$$

Since $\cos()$ and $\sin()$ are functions, they have graphs, and, by the definition of graph of a function, the graph of the cosine function must be the graph of the equation

$$y = \cos x.$$

In like manner, the graph of the sine function is the graph of the equation

$$y = \sin x.$$

The form of these graphs is shown in Fig. 10.4. It will be noted that the shape of these graphs is repeated as we look from left to right. This repetitive feature of the graph is given analytic form in

Theorem 10.2:

If t is any number,

$$\cos(t + 2\pi) = \cos t$$

and

$$\sin(t + 2\pi) = \sin t.$$

Proof:

If t is non-negative, the point on the graph of $x^2 + y^2 = 1$ which is t units from $(1, 0)$, measured along the circumference in a *counter-clockwise* manner, is also $t + 2\pi$ units from $(1, 0)$, measured in the same manner, since the length of the entire circumference is 2π units.

If t is negative, $\cos(-t)$ and $\sin(-t)$ are the x and y -coordinates of that point on the graph of $x^2 + y^2 = 1$ whose distance from $(1, 0)$, measured along the circumference in a *clockwise* manner, is the (positive) number $-t$. But if $t + 2\pi$ is positive, then this same point is also $2\pi - (-t)$ units from $(1, 0)$ measured in a *counter-clockwise* manner, and if $t + 2\pi$ is negative, then this same point is also $-t - 2\pi$ units from $(1, 0)$ measured in a *counter-clockwise* manner.

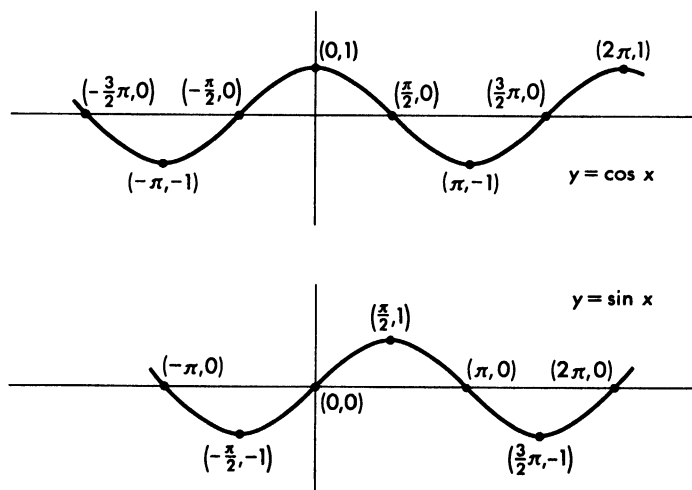


Fig. 10.4

Exercises

1. Rewrite the proof of Theorem 10.1 in a formal manner, citing the reason for each statement written.

2. Prove that if t is any number,

$$\begin{aligned}\cos(t - 2\pi) &= \cos t \\ \text{and } \sin(t - 2\pi) &= \sin t.\end{aligned}$$

3. Prove that if t is any number, and n is any integer (positive or negative), then

$$\begin{aligned}\cos(t + 2n\pi) &= \cos t \\ \text{and } \sin(t + 2n\pi) &= \sin t.\end{aligned}$$

4. If the function $f()$ has the set of all numbers as domain and there is a fixed number p , such that

$$f(x + p) = f(x), \text{ for all } x,$$

prove that, for any integer n ,

$$f(x + np) = f(x).$$

(In this case, the number p is called a *period* of $f()$. The smallest such positive number p is called the period of $f()$. Thus, $\cos()$ is said to have period 2π .)

On the same coordinate plane, sketch the graphs of

5. $y = \cos x, y = \cos 2x, y = \cos 3x.$

6. $y = \sin x, y = \sin 2x, y = \sin 3x.$

7. $y = \sin x, y = \sin^2 x.$

8. $y = \cos x, y = \cos^2 x.$

9. $y = 1 - \cos x, y = 1 - \cos 2x, y = \frac{1}{2}(1 - \cos 2x).$

10. $y = 1 + \cos x, y = 1 + \cos 2x, y = \frac{1}{2}(1 + \cos 2x).$

10.3. OTHER TRIGONOMETRIC FUNCTIONS

In addition to the cosine and sine functions, we also define:

Definition 10.4: The functions $\tan()$, $\sec()$, $\csc()$, and $\cot()$ are specified by

$$\tan() = \left\{ \frac{\sin}{\cos} \right\}(),$$

$$\sec() = \left\{ \frac{1}{\cos} \right\}(),$$

$$\csc() = \left\{ \frac{1}{\sin} \right\}(),$$

$$\cot() = \left\{ \frac{\cos}{\sin} \right\}().$$

These functions are named as follows: $\tan()$ is the tangent function; $\sec()$ is the secant function; $\csc()$ is the cosecant function; and $\cot()$ is the cotangent function.

It should be noted that each of these new functions is defined as the quotient of two other functions—for example, $\sec()$ is the quotient of the constant function $U()$ by the function $\cos()$. Consequently, these new functions cannot have in their domains any number which would require a division by zero. Thus, the domain of the tangent function consists of all the numbers except those which are odd integer multiples of $\pi/2$, since the cosine of any of these numbers is 0. The determination of the numbers excluded from the domains of $\sec()$, $\csc()$ and $\cot()$ is left as an exercise for the reader (see exercise 1).

Since the two basic trigonometric functions are $\cos()$ and $\sin()$, any expression involving any of the four new functions can always be rewritten in terms involving $\cos()$ and $\sin()$ only. For example, the expression:

$$\sec^2 t - \tan^2 t$$

is a symbol for a number whenever t is in both the domain of the secant function and the domain of the tangent function.

From the definition of these functions,

$$\sec^2 t - \tan^2 t = \frac{1}{\cos^2 t} - \frac{\sin^2 t}{\cos^2 t}.$$

But this last expression may be simplified to

$$\frac{1 - \sin^2 t}{\cos^2 t}.$$

Theorem 10.1 tells us that $1 - \sin^2 t$ and $\cos^2 t$ are the same number. Thus, excluding those values of t which would require a division by 0, we have

$$\sec^2 t - \tan^2 t = 1.$$

Formally, this result may be stated as

Theorem 10.3:

If t is not an odd-integer multiple of $\pi/2$, then

$$\tan^2 t + 1 = \sec^2 t.$$

In a similar manner, we have

Theorem 10.4:

If t is not an integer multiple of π , then

$$\cot^2 t + 1 = \csc^2 t.$$

The proof of this theorem is left as an exercise (see exercise 4).

A particularly simple way of remembering the results of Theorems 10.1, 10.3, and 10.4, is to look only at those values of the variable t which are between 0 and $\pi/2$. If t is such a number, then we may illustrate the theorems with a number of right triangles. (Fig. 10.5).

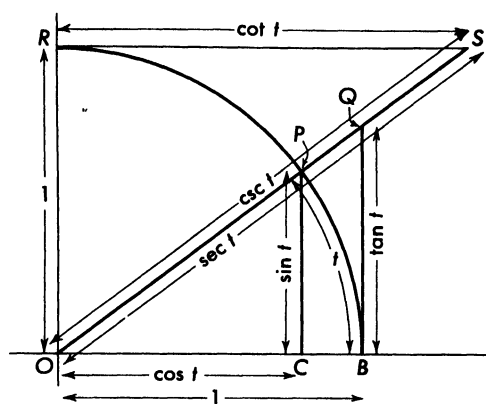


Fig. 10.5

In this diagram, we have three right triangles, $OC P$, $OB Q$, and $OR S$. The angle at S is the same as the angle between OP and OB . Then, from the definitions of cosine and sine of a number, if t is the length of the arc BP , we have

$$\begin{aligned}\sin t &= \text{length of } CP \\ \cos t &= \text{length of } OC.\end{aligned}$$

Since the triangles $OC P$ and $OB Q$ are similar, $\frac{CP}{OC} = \frac{BQ}{OB}$, so that

$$\tan t = \text{length of } BQ.$$

Also, $\frac{OQ}{OB} = \frac{OP}{OC}$, so that

$$\sec t = \text{length of } OQ.$$

Since the triangles $OR S$ and $OC P$ are similar, $\frac{RS}{OR} = \frac{OC}{CP}$, so that

$$\cot t = \text{length of } RS.$$

Also, $\frac{OS}{OR} = \frac{CP}{OP}$, so that

$$\csc t = \text{length of } OS.$$

But these three triangles are right triangles, and the Pythagorean theorem of elementary plane geometry is true, so we have, from the triangle $OC P$:

$$\sin^2 t + \cos^2 t = 1;$$

from the triangle OBQ :

$$\tan^2 t + 1 = \sec^2 t;$$

and from the triangle ORS :

$$\cot^2 t + 1 = \csc^2 t.$$

Exercises

1. Find the domain of the functions: $\sec()$, $\csc()$, and $\cot()$.
2. Find the range of values of the functions: $\tan()$, $\sec()$, $\csc()$, and $\cot()$.
3. Sketch the graphs of the functions: $\tan()$, $\sec()$, $\csc()$, and $\cot()$.
4. Prove Theorem 10.4.
5. Write out a formal proof of Theorem 10.3.
6. Fill in the spaces in the following table:

t	$\cos t$	$\sin t$	$\tan t$	$\sec t$	$\csc t$	$\cot t$
0	1					
$\pi/6$	$\sqrt{3}/2$					
$\pi/4$	$\sqrt{2}/2$					
$\pi/3$	$1/2$					
$\pi/2$	0					

7. Prove that if t is any number in the domain of $\tan()$, then

$$\tan(t + \pi) = \tan t.$$

10.4. SOME USEFUL RELATIONSHIPS

The definition of cosine and sine of a number requires that for any positive number, we find that point on the unit circle whose distance, measured *counter-clockwise*, from $(1, 0)$ is that positive number. This definition also implies that for any negative number, we find that point on the unit circle whose distance, measured *clockwise*, from $(1, 0)$ is the negative of (and thus is the absolute value of) that negative number.

If r is a positive number, the point on the unit circle, r units from $(1, 0)$, measured in a *counter-clockwise* manner, has coordinates $(\cos r, \sin r)$. The actual (straight-line) distance between this point and $(1, 0)$ can be found simply by using the theorem which gives the distance between two points whose coordinates are known. The distance between the points $(\cos r, \sin r)$ and $(1, 0)$ then is

$$\sqrt{(\cos r - 1)^2 + \sin^2 r}.$$

Because of the relations between $\cos ()$ and $\sin ()$ discovered in the preceding section,

$$\begin{aligned} (\cos r - 1)^2 + \sin^2 r &= \cos^2 r - 2 \cos r + 1 + \sin^2 r, \\ &= \cos^2 r + \sin^2 r + 1 - 2 \cos r, \\ &= 1 + 1 - 2 \cos r, \\ &= 2 - 2 \cos r, \end{aligned}$$

so that the distance between the points $(\cos r, \sin r)$ and $(1, 0)$ may be re-expressed as

$$\sqrt{2 - 2 \cos r}.$$

Fig. 10.6 illustrates this property for r a number between π and $3\pi/2$.

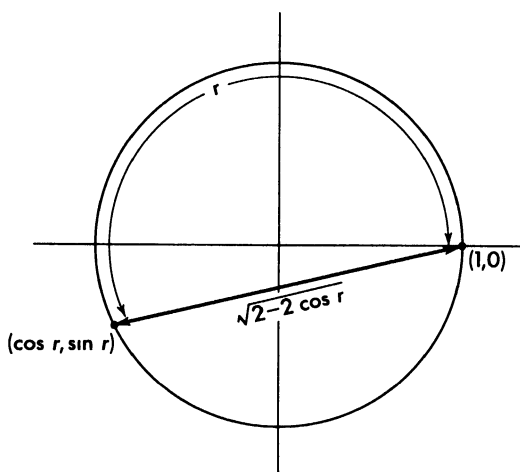


Fig. 10.6

This preliminary result allows us to prove the following theorem, which is basic to all the following ones of this section.

Theorem 10.5:

If s and t are any two numbers,
 $\cos (t - s) = \cos s \cos t + \sin s \sin t.$

Proof:

1. If $t = s$, $\cos(t - s) = \cos 0 = 1$,
 $\cos s \cos t + \sin s \sin t = \cos^2 t + \sin^2 t = 1$;
 and the theorem is true.
2. If $t > s$, let $r = t - s$. Then r is positive.
3. If $s > 0$, then $t > 0$ and the point $(\cos s, \sin s)$ is s units from $(1, 0)$, measured counter-clockwise; the point $(\cos t, \sin t)$ is t units from $(1, 0)$, measured counter-clockwise; and these two points are $(t - s)$ units from each other, measured along the circumference.
4. If $t < 0$, then $s < 0$, and the point $(\cos s, \sin s)$ is $-s$ units from $(1, 0)$, measured clockwise; the point $(\cos t, \sin t)$ is $-t$ units from $(1, 0)$, measured clockwise; and these two points are $-s - (-t)$ $[= t - s]$ units from each other, measured along the circumference.
5. If $t > 0$ and $s < 0$, then the point $(\cos s, \sin s)$ is $-s$ units from $(1, 0)$, measured clockwise; the point $(\cos t, \sin t)$ is t units from $(1, 0)$; and these two points are $-s + t$ $[= t - s]$ units from each other, measured along the circumference.
6. By steps 2-5, if $t > s$, then the distance, measured along the circumference, from $(\cos s, \sin s)$ to $(\cos t, \sin t)$ is $t - s$ ($= r$).
7. By the discussion preceding the theorem, the actual (straight-line) distance between the points $(\cos r, \sin r)$ and $(1, 0)$, is

$$\sqrt{2 - 2 \cos r}.$$

8. The distance from $(1, 0)$ to $(\cos r, \sin r)$, measured along the circumference, is r .
9. By steps 6, 7, and 8, the distance between the points $(\cos s, \sin s)$ and $(\cos t, \sin t)$ is

$$\sqrt{2 - 2 \cos r}.$$

10. By the theorem on finding the actual (straight-line) distance between two points, the distance between the points $(\cos s, \sin s)$ and $(\cos t, \sin t)$ is

$$\sqrt{(\cos s - \cos t)^2 + (\sin s - \sin t)^2}.$$

11. By Theorem 10.1, and algebraic manipulations,

$$\begin{aligned} & (\cos s - \cos t)^2 + (\sin s - \sin t)^2 \\ &= \cos^2 s - 2 \cos s \cos t + \cos^2 t + \sin^2 s - 2 \sin s \sin t + \sin^2 t \\ &= \cos^2 s + \sin^2 s + \cos^2 t + \sin^2 t - 2 [\cos s \cos t + \sin s \sin t] \\ &= 1 + 1 - 2 [\cos s \cos t + \sin s \sin t] \\ &= 2 - 2 [\cos s \cos t + \sin s \sin t]. \end{aligned}$$

12. By steps 9, 10, and 11,

$$\cos r = \cos(t - s) = \cos s \cos t + \sin s \sin t.$$

13. If $t < s$, $\cos(t - s) = \cos(s - t)$, and letting $r = s - t$ allows us to repeat steps 2 - 12.

Fig. 10.7 illustrates this theorem for s and t positive numbers less than 2π .

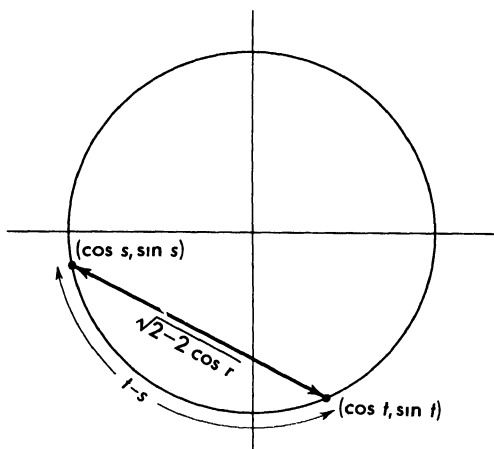


Fig. 10.7

For the particular value of $t = \pi/2$, we have

$$\begin{aligned}\cos\left(\frac{\pi}{2} - s\right) &= \cos\frac{\pi}{2}\cos s + \sin\frac{\pi}{2}\sin s \\ &= (0)\cos s + (1)\sin s \\ &= \sin s,\end{aligned}$$

so that we have

Corollary A to Theorem 10.5:

For any number s , $\cos\left(\frac{\pi}{2} - s\right) = \sin s$.

If, in Theorem 10.5, we replace s by $-x$, we have

Corollary B to Theorem 10.5:

For any numbers, t and x ,

$$\cos(t + x) = \cos t \cos x - \sin t \sin x.$$

Proof:

$$\begin{aligned}\cos(t + x) &= \cos[t - (-x)] \\ &= \cos t \cos(-x) + \sin t \sin(-x) \\ &= \cos t \cos x + \sin t(-\sin x) \\ &= \cos t \cos x - \sin t \sin x.\end{aligned}$$

Using Corollary A to Theorem 10.5, we find that

$$\sin\left(\frac{\pi}{2} - t\right) = \cos\left[\frac{\pi}{2} - \left(\frac{\pi}{2} - t\right)\right] = \cos t.$$

Theorem 10.5 and its corollaries may be used to prove

Theorem 10.6:

If s and t are any numbers,

$$\sin(s + t) = \sin s \cos t + \cos s \sin t.$$

The proof of this theorem is left as an exercise. (See exercise 1.)

Among the most used results of Theorem 10.6 are the following two theorems.

Theorem 10.7:

If x is any number,

$$\cos^2 x = \frac{1}{2} + \frac{1}{2} \cos 2x.$$

Theorem 10.8:

If x is any number,

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

Proof of Theorem 10.7:

- | | |
|---|---------------------------------|
| 1. $\cos 2x = \cos(x + x)$ | by the fact that $1 + 1 = 2$, |
| $= \cos x \cos x - \sin x \sin x$ | by Corollary B to Theorem 10.5, |
| $= \cos^2 x - \sin^2 x$ | by the definition of "square," |
| $= \cos^2 x - (1 - \cos^2 x)$ | by Theorem 10.1, |
| $= 2 \cos^2 x - 1$ | by algebraic manipulation. |
| 2. $2 \cos^2 x = 1 + \cos 2x$ | by step 1. |
| 3. $\cos^2 x = \frac{1}{2} + \frac{1}{2} \cos 2x$ | by step 2. |

Exercises

1. Prove Theorem 10.5.
2. Prove Theorem 10.8.
3. Draw a diagram of the unit circle to illustrate the proof of Theorem 10.6 for:
 - a) $0 < s < \frac{\pi}{2}$, $-\frac{\pi}{2} < t < 0$.
 - b) $3\pi < s < \frac{7}{2}\pi$, $5\pi < t < \frac{11}{2}\pi$.

Use the theorems thus far proven to show that the following equations are identities:

4. $(\sin r - \cos r)^2 = 1 - \sin 2r$.
5. $\cos 3s = \cos^3 s - 3 \cos s \sin^2 s$.
6. $\tan(t - u) = \frac{\tan t - \tan u}{1 + \tan t \tan u}$.
7. $\sin 3v = 4 \sin v \cos^2 v - \sin v$.
8. $\cos 3w \sin 2w - \cos 4w \sin w = \cos 2w \sin w$.
9. $\frac{2 \sin 2x - \sin 4x}{2 \sin 2x + \sin 4x} = \tan^2 x$.
10. $\sin y + \sin z = 2 \sin \frac{1}{2}(y + z) \cos \frac{1}{2}(y - z)$.

10.5. SOME LIMIT THEOREMS

In order to discover facts about derivatives, anti-derivatives, or integrals of any function, it is necessary to use the definitions of derivative, anti-derivative, and integral. All of these definitions involve the limit concept, so that it becomes necessary that we conduct an investigation into the limit properties of the trigonometric functions. Since the trigonometric functions $\cos()$ and $\sin()$ are basic, we may restrict our detailed investigation to these two functions.

Theorem 10.9:

$$\lim_{h \rightarrow 0} \cos h = 1.$$

Proof:

1. From the definition of the cosine and sine functions, if h is any number, the point $(\cos h, \sin h)$ is at a distance, measured along the circumference of the unit circle, of $|h|$ units from the point $(1, 0)$.
2. Since the straight-line distance between any two points is always the shortest, we have, for any number h ,

$$0 \leq \sqrt{(\cos h - 1)^2 + \sin^2 h} \leq |h|.$$

3. Step 3 implies that, for any h ,

$$0 \leq (\cos h - 1)^2 + \sin^2 h \leq h^2.$$

4. Algebraic manipulation and the fact that $\cos^2 h + \sin^2 h = 1$, shows that, for any h ,

$$0 \leq 2[1 - \cos h] \leq h^2.$$

5. $\lim_{h \rightarrow 0} h^2 = 0$ since h^2 specifies a continuous function.
6. $\lim_{h \rightarrow 0} 0 = 0$ since $\theta()$ is a constant function.
7. $\lim_{h \rightarrow 0} 2(1 - \cos h) = 0$ by steps 4, 5, and 6, and the "pinching theorem" on limits.
8. $\lim_{h \rightarrow 0} (1 - \cos h) = 0$ by step 7 and division by 2.
9. $\lim_{h \rightarrow 0} 1 - \lim_{h \rightarrow 0} (\cos h) = 0$ by step 8 and the fact that the limit of the difference of two functions is the difference of their limits.
10. $\lim_{h \rightarrow 0} (\cos h) = 1$ by step 9.

Corollary to Theorem 10.9:

$$\lim_{h \rightarrow 0} \sin h = 0.$$

This corollary is proved by using the fact that

$$\sin^2 h + \cos^2 h = 1.$$

The details of the proof are left as an exercise. (See exercise 1.)

Theorem 10.10:

The function $\cos ()$ is continuous.

Proof:

1. For any a and any h ,

$$\cos (a + h) - \cos a = \cos a \cos h - \sin a \sin h - \cos a.$$

2. $\lim_{h \rightarrow 0} [\cos (a + h) - \cos a]$

$$= \lim_{h \rightarrow 0} [\cos a \cos h] - \lim_{h \rightarrow 0} [\sin a \sin h] - \lim_{h \rightarrow 0} \cos a$$

by step 1 and the theorem about the limit of the sum or difference of two functions.

3. $\lim_{h \rightarrow 0} [\cos (a + h) - \cos a] = \cos a \lim_{h \rightarrow 0} \cos h - \sin a \lim_{h \rightarrow 0} \sin h - \cos a$

since $\cos a$ and $\sin a$ are constants.

4. $\lim_{h \rightarrow 0} [\cos (a + h) - \cos a] = 0$

by step 3, Theorem 10.8 and its corollary.

5. $\cos ()$ is continuous at any number

by step 4 and the definition of continuous.

In a similar manner, we can prove

Theorem 10.11:

The function $\sin ()$ is continuous.

The proof of this theorem is left as an exercise. (See exercise 2.)

We now proceed to the proof of two theorems which will be needed in order to find the derivatives of the functions $\cos ()$ and $\sin ()$.

Theorem 10.12:

$$\lim_{h \rightarrow 0} \frac{1 - \cos h}{h} = 0.$$

Proof:

1. Repeating steps 1-4 of the proof of Theorem 10.9, we have, for any number h ,

$$0 \leq 2(1 - \cos h) \leq h^2.$$

2. From step 1,

$$0 \leq 1 - \cos h \leq \frac{h^2}{2}.$$

3. Step 2 yields

$$0 \leq \frac{1 - \cos h}{|h|} \leq \frac{|h|}{2}, \text{ since } h^2 = |h|^2.$$

4. $\lim_{h \rightarrow 0} \frac{|h|}{2} = \lim_{h \rightarrow 0} 0 = 0.$

5. By steps 3 and 4, and the "pinching theorem,"

$$\lim_{h \rightarrow 0} \frac{1 - \cos h}{h} = 0.$$

Theorem 10.13:

$$\lim_{h \rightarrow 0} \frac{\sin h}{h} = 1.$$

Proof:

1. For any number h , the distance, measured along the circumference of the unit circle, between $(\cos h, \sin h)$ and $(1, 0)$ is not less than the length of the line segment joining $(\cos h, \sin h)$ and $(\cos h, 0)$, since this line segment is perpendicular to the x -axis, and a perpendicular line gives the shortest distance from a given point to a given line. (See Fig. 10.8.)
2. The distance, measured along the circumference of the unit circle, between $(\cos h, \sin h)$ and $(1, 0)$ is not more than the sum of the lengths of the line segments joining $(\cos h, \sin h)$ to $(1, \sin h)$ and $(1, \sin h)$ to $(1, 0)$. (See Fig. 10.8.)

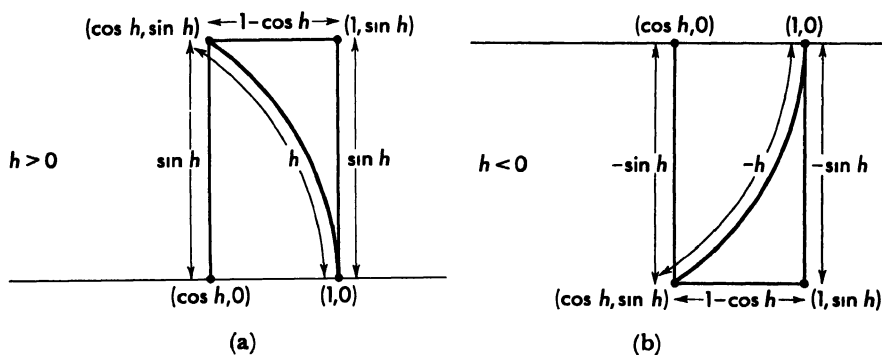


Fig. 10.8

3. If $h > 0$, steps 1 and 2 imply

$$h < 1 - \cos h + \sin h < 1 - \cos h + h.$$

4. From step 3, if $h > 0$,

$$1 < \frac{1 - \cos h}{h} + \frac{\sin h}{h} < \frac{1 - \cos h}{h} + 1.$$

5. If $h < 0$, steps 1 and 2 imply

$$-h < 1 - \cos h - \sin h < 1 - \cos h - h.$$

6. From step 5, if $h < 0$,

$$-1 > \frac{1 - \cos h}{h} - \frac{\sin h}{h} > \frac{1 - \cos h}{h} - 1$$

7. From step 6, if $h < 0$,

$$1 < -\frac{1 - \cos h}{h} + \frac{\sin h}{h} < 1 - \frac{1 - \cos h}{h}.$$

8. By Theorem 10.12, $\lim_{h \rightarrow 0} \frac{1 - \cos h}{h} = 0$.

9. By steps 4, 7, and 8, and the "pinching theorem" on limits,

$$\lim_{h \rightarrow 0} \frac{\sin h}{h} = 1.$$

Exercises

1. Prove the Corollary to Theorem 10.9.
2. Prove Theorem 10.11.

Find the following limits:

3. $\lim_{r \rightarrow 0} \frac{\sin 2r}{r}.$
4. $\lim_{s \rightarrow 0} \frac{\sin^2 s}{s}.$
5. $\lim_{t \rightarrow 0} \frac{\sin^2 t}{t^2}.$
6. $\lim_{u \rightarrow 0} \tan u.$
7. $\lim_{h \rightarrow 0} \frac{\tan h}{h}.$
8. $\lim_{w \rightarrow 0} \frac{\tan^2 w}{w \sin w}.$
9. $\lim_{x \rightarrow 0} x \sec x.$
10. $\lim_{y \rightarrow 0} y \csc y.$
11. $\lim_{z \rightarrow 0} z \cot z.$

10.6. DERIVATIVES OF THE TRIGONOMETRIC FUNCTIONS

Now that we have the necessary limit theorems available, we can proceed to finding the derivatives of the $\cos()$ and $\sin()$ functions.

Theorem 10.14:

$$\{D \cos\}() = \{-\sin\}().$$

Proof:

$$\begin{aligned} & \lim_{h \rightarrow 0} \left[\frac{\cos(a+h) - \cos a}{h} \right] \\ = & \lim_{h \rightarrow 0} \left[\frac{\cos a \cos h - \sin a \sin h - \cos a}{h} \right] \end{aligned}$$

by the theorem on the cosine of the sum of two numbers,

$$= \lim_{h \rightarrow 0} \left[\cos a \frac{\cos h - 1}{h} - \sin a \frac{\sin h}{h} \right]$$

by algebraic manipulation,

$$= \cos a \lim_{h \rightarrow 0} \frac{\cos h - 1}{h} - \sin a \lim_{h \rightarrow 0} \frac{\sin h}{h}$$

by the theorems on limits of differences of functions,

$$= (\cos a) (0) - (\sin a) (1)$$

by Theorems 10.12 and 10.13,

$$= -\sin a.$$

Theorem 10.15:

$$\{D \sin\}() = \cos().$$

Proof:

$$\sin t = \cos\left(\frac{\pi}{2} - t\right)$$

by Corollary A to Theorem 10.5.

$$\{D \sin\}(t) = -\left[\sin\left(\frac{\pi}{2} - t\right)\right](-1) \text{ by the "chain rule" for differentiation.}$$

$$= \cos t.$$

Theorem 10.16:

$$\{D \tan\}() = \{\sec^2\}().$$

Proof:

$$\{D \tan\}() = \left\{D \left\{\frac{\sin}{\cos}\right\}\right\}()$$

$$= \left\{\frac{(\cos)(D \sin) - (\sin)(D \cos)}{\cos^2}\right\}()$$

by the theorem on the differentiation of the quotient of two functions,

$$= \left\{\frac{\cos^2 + \sin^2}{\cos^2}\right\}()$$

by Theorems 10.14 and 10.15,

$$= \left\{\frac{1}{\cos^2}\right\}()$$

by the theorem stating that $\{\sin^2 + \cos^2\}() = U()$,

$$= \{\sec^2\}()$$

by the definition of $\sec()$.

Theorem 10.17:

$$\{D \sec\}() = \{\sec \tan\}().$$

Theorem 10.18:

$$\{D \csc\}() = \{-\csc \cot\}().$$

Theorem 10.19:

$$\{D \cot\}() = \{-\csc^2\}().$$

The proofs of the last three theorems are left as exercises.

Exercises

1. Prove Theorem 10.17.
2. Prove Theorem 10.18.
3. Prove Theorem 10.19.

Find the derivatives of the functions specified by:

- | | |
|---------------------------|-----------------------|
| 4. $2 \sin r \cos r.$ | 10. $v^3 \sec v^3.$ |
| 5. $\sin^3 s.$ | 11. $\csc 3w^3.$ |
| 6. $\cos^3 t + \sin^3 t.$ | 12. $\cos (1 + 2x).$ |
| 7. $\cos^3 u.$ | 13. $\cot^3 (x - 1).$ |
| 8. $t \sin t + \cos t.$ | 14. $\sin (\cos x).$ |
| 9. $\cot u^{1/2}.$ | |

Find all relative maximum and relative minimum points on the graphs of:

15. $y = \cos x.$
16. $y = \sin x.$
17. $y = \cos x + \sin x.$
18. $y = x + \sin x.$
19. Show that if n is a natural number, then the $(n + 4)$ -th derivative of $\cos()$ is the same as the n -th derivative.

10.7. ANTI-DERIVATIVES

Corresponding to the two basic differentiation theorems,

$$\{D \cos\}() = \{-\sin\}() \quad \text{and} \quad \{D \sin\}() = \cos(),$$

we have the obvious anti-derivative theorems:

Theorem 10.20:

If $f()$ is an anti-derivative of $\cos()$, then there exists a constant K , such that

$$f() = \{\sin + K\}().$$

Theorem 10.21:

If $f()$ is an anti-derivative of $\sin()$, then there exists a constant K such that

$$f() = \{-\cos + K\}().$$

These theorems have consequences far beyond their appearance. Let us consider the following example:

EXAMPLE 1:

Find an anti-derivative of $\{\sin^3\}()$.

We have $\sin^3 t = \frac{1}{2} - \frac{1}{2} \cos 2t$. If we make a first guess of

$$g_1(t) = \frac{1}{2}t - \sin 2t,$$

we find

$$g_1'(t) = \frac{1}{2} - 2 \cos 2t,$$

and note that we have 2 where we want $1/2$. We then amend our first guess to

$$g_2(t) = \frac{1}{2}t - \frac{1}{4}\sin 2t,$$

and we find

$$g_2'(t) = \frac{1}{2} - \frac{1}{2}\cos 2t = \sin^2 t,$$

so that we may conclude that any anti-derivative of $\{\sin^2\}()$ will be given by

$$\frac{1}{2}t - \frac{1}{4}\sin 2t + K,$$

where K is some constant.

EXAMPLE 2:

Find an anti-derivative of $\{\sin^3\}()$.

We have

$$\begin{aligned}\sin^3 t &= \sin^2 t \sin t \\ &= (1 - \cos^2 t) \sin t \\ &= \sin t - \cos^2 t \sin t.\end{aligned}$$

If we make a first guess of

$$g_1(t) = \cos t + \cos^3 t,$$

we find

$$g_1'(t) = -\sin t - 3\cos^2 t \sin t,$$

and we note that the first term has a “ $-$,” where we want “ $+$,” and the second term has 3 where we want 1. We then amend our first guess to

$$g_2(t) = -\cos t + \frac{1}{3}\cos^3 t,$$

and then find

$$g_2'(t) = \sin t - \cos^2 t \sin t,$$

so that we may conclude that any anti-derivative of $\{\sin^3\}()$ will be given by

$$-\cos t + \frac{1}{3}\cos^3 t + K,$$

where K is some constant.

EXAMPLE 3:

Find an anti-derivative of

$$\{\sqrt{1 + \sin} \cos\}()$$

The form of this function, as the product of a somewhat complicated function, involving the sine function, with the cosine function, suggests that we might have the derivative of a composite $f(h)$, in which $h() = \sin()$, and $f'() = \{\sqrt{1 + I}\}()$.

Then a reasonable first guess would be given by

$$g_1(t) = (1 + \sin t)^{3/2}.$$

For this,

$$g_1'(t) = \frac{3}{2} (1 + \sin t)^{1/2} \cos t,$$

and this function differs from the original only in the constant factor $3/2$. We may then amend our guess to

$$g_2(t) = \frac{2}{3} (1 + \sin t)^{3/2}$$

to get

$$g_2'(t) = \sqrt{1 + \sin t} \cos t,$$

so that any anti-derivative of

$$\{\sqrt{1 + \sin t} \cos t\}()$$

is given by

$$\frac{2}{3} (1 + \sin t)^{3/2} + K,$$

where K is some constant.

Exercises

Find anti-derivatives of the functions specified by:

- | | |
|-------------------------|-------------------------|
| 1. $\sin r \cos r.$ | 6. $\tan w \sec^3 w.$ |
| 2. $\cos s \csc^3 s.$ | 7. $x \sin x^2.$ |
| 3. $\tan t \sec^2 t.$ | 8. $\cos (3x + 5).$ |
| 4. $\tan^2 u \sec^2 u.$ | 9. $(1 + \sin y)^{-1}.$ |
| 5. $\sin^2 v \sec^4 v.$ | 10. $\sec^3 z.$ |

11. Find the area bounded by the graphs of

$$y = \sin x, \quad y = 0, \quad \text{and} \quad x = \pi/2.$$

12. Find the area bounded by the graphs of

$$y = \sin x, \quad y = 0, \quad \text{and} \quad x = t,$$

where $0 < t < \pi/2$.

13. Find the value of t which makes the area found in problem 12 exactly half that found in problem 11.

Use anti-differentiation by parts to find an anti-derivative of the functions specified by:

14. $r \sin r.$
15. $s^2 \sin s.$
16. $t \cos t.$
17. $u^2 \cos u.$
18. $\cot^2 v.$

10.8. INVERSE TRIGONOMETRIC FUNCTIONS

In order for any function to have an inverse, it is necessary that that function be a collection of ordered pairs, no two of which pairs have either the same first element or the same second element. As we noted in Sec. 6 of Chapter 9, this requirement is equivalent to the geometric statement that a horizontal line cut the graph of the function at most once. A glance at the graphs of the six trigonometric functions we have defined shows that any horizontal line cuts either not at all or "infinitely often"—that is, certainly more than once. Therefore, not one of the functions $\cos()$, $\sin()$, $\tan()$, $\sec()$, $\csc()$, and $\cot()$ has an inverse.

It is possible, however, to create new trigonometric functions from these six, which new trigonometric functions have the same rules as these six, but have much smaller domains. In order to simplify matters, we will admit a prejudice in favor of positive numbers, and make our definitions of these new functions so that their ranges of values will not contain negative numbers. The definitions are given in:

Definition 10.5: $\text{Cos}()$ is defined by

$$\text{Cos } t = \cos t, \quad \text{where } 0 \leq t \leq \pi/2.$$

$\text{Sin}()$ is defined by

$$\text{Sin } t = \sin t, \quad \text{where } 0 \leq t \leq \pi/2.$$

$\text{Tan}()$ is defined by

$$\text{Tan } t = \tan t, \quad \text{where } 0 \leq t < \pi/2.$$

$\text{Sec}()$ is defined by

$$\text{Sec } t = \sec t, \quad \text{where } 0 \leq t < \pi/2.$$

$\text{Csc}()$ is defined by

$$\text{Csc } t = \csc t, \quad \text{where } 0 < t \leq \pi/2.$$

$\text{Cot}()$ is defined by

$$\text{Cot } t = \cot t, \quad \text{where } 0 < t \leq \pi/2.$$

We will thus adopt the notation of using capital letters to indicate the restriction of the regular trigonometric function to domains which are the interval with end points 0 and $\pi/2$, including or excluding the end points as becomes necessary in order to avoid division by 0. The graphs of the first four of these new functions are shown in Fig. 10.9.

Since the ranges of values of all these new functions consist only of non-negative numbers, Theorems 10.1, 10.3, and 10.4 give us

Theorem 10.22:

$$\text{Sin}() = \{\sqrt{1 - \text{Cos}^2}\}().$$

$$\text{Cos}() = \{\sqrt{1 - \text{Sin}^2}\}().$$

$$\tan(\theta) = \sqrt{\sec^2(\theta) - 1}.$$

$$\sec(\theta) = \sqrt{1 + \tan^2(\theta)}.$$

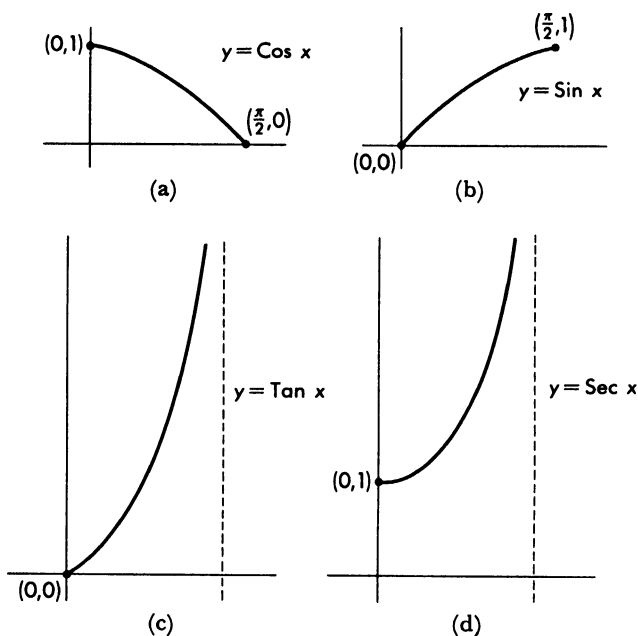


Fig. 10.9

Although all six of our new functions have inverses, it is usual only to consider the first four. Since these are restrictions of (that is, have fewer ordered pairs than) the corresponding non-capitalized functions, we have, immediately from the corresponding derivative theorems for the ordinary trigonometric functions:

Theorem 10.23:

$$\{D \sin\}(t) = \cos(t)$$

$$\{D \cos\}(t) = \{-\sin\}(t) \quad \text{for } 0 < t < \pi/2.$$

$$\{D \tan\}(t) = \{\sec^2\}(t)$$

$$\{D \sec\}(t) = \{\sec \tan\}(t)$$

We now give names to the inverses of the four restricted trigonometric functions.

Definition 10.6:

$$\arcsin(\theta) = \{\sin\}^{-1}(\theta)$$

$$\arccos(\theta) = \{\cos\}^{-1}(\theta)$$

$$\arctan(\theta) = \{\tan\}^{-1}(\theta)$$

$$\operatorname{arcsec}(\theta) = \{\sec\}^{-1}(\theta).$$

In further work, it is usual to consider inverses to many different portions of given trigonometric functions. Certain of these inverses are then designated as “principal values,” and the entire collections are called “multiple-valued functions.” We merely note here that we have given the four functions specified in the definition (10.6) the interval:

$$\text{the set of all } t, \quad 0 \leq t \leq \pi/2$$

(with 0 excluded if necessary) as ranges of values. We could have extended these functions to have somewhat larger domains and ranges, but chose not to do so in order to simplify the analysis.

Exercises

1. What is the domain of each of the functions: $\arcsin()$, $\arccos()$, $\arctan()$, and $\operatorname{arcsec}()$?
2. Find: $\arcsin 1$; $\arccos 1$; $\arctan 1$; $\operatorname{arcsec} 1$.

Evaluate the following:

3. $\arcsin 1/2$.
4. $\arccos \sqrt{3}/2$.
5. $\arctan 1$.
6. $\operatorname{arcsec} 2$.
7. $\arctan (\sin \pi/4)$.
8. $\arcsin (\tan 3\pi)$.
9. $\tan (\arcsin \sqrt{3}/2)$.

Simplify the expressions:

10. $\sin (\arccos r)$.
11. $\sin (\arccos s)$.
12. $\tan (\operatorname{arcsec} t)$.
13. $\tan (\operatorname{arcsec} u)$.
14. $\sec (\arctan \sqrt{v^2 - 1})$.
15. $\sin (\arccos \sqrt{1 - w^2})$.
16. $\cos (\frac{1}{2} \arcsin x)$.
17. $\arccos (\sin y)$.
18. $\arcsin [\cos (z + 3\pi)]$.

10.9. THE DERIVATIVES OF THE INVERSE TRIGONOMETRIC FUNCTIONS

Theorem 9.5 states that if the function $g()$ is the inverse of the function $f()$ then

$$g'() = \left\{ \frac{1}{f'(g)} \right\} ().$$

Applying this theorem successively to the functions $\sin ()$, $\cos ()$, $\tan ()$, and $\sec ()$, we get the following theorems:

Theorem 10.24:

$$\{D \arcsin\}(t) = \left\{ \frac{1}{\sqrt{1-t^2}} \right\}(t), \quad 0 < t < 1$$

Proof:

If t is any number in the domain of $\arcsin ()$,

$$\begin{aligned} \{D \arcsin\}(t) &= \frac{1}{\{D \sin\} \arcsin t} \quad \text{by Theorem 9.5,} \\ &= \frac{1}{\{\cos\}(\arcsin t)} \quad \text{by Theorem 10.23,} \\ &= \frac{1}{\{\sqrt{1-\sin^2}\}(\arcsin t)} \\ &\quad \text{by Theorem 10.22,} \\ &= \frac{1}{\sqrt{1-\{\sin^2\}(\arcsin t)}} \\ &\quad \text{by applying the rule of the denominator function,} \\ &= \frac{1}{\sqrt{1-[\sin(\arcsin t)]^2}} \\ &\quad \text{by the definition of the square of a function,} \\ &= \frac{1}{\sqrt{1-t^2}} \quad \text{by the definition (10.6) and the definition of the inverse of a function,} \\ &= \left\{ \frac{1}{\sqrt{1-t^2}} \right\}(t) \quad \text{by the definition of the identity function, } I(). \end{aligned}$$

The result of this theorem is really worthy of note. We have defined a function, $\sin ()$, in terms of coordinates of points along the unit circle—a most non-algebraic process. Next we defined a restriction, $\sin ()$, of this function. We then found the inverse of this restriction, $\arcsin ()$. All of these operations are well removed from the algebraic ones of addition, subtraction, multiplication, division, and raising to rational number powers. Yet we find that the derivative, $\{D \arcsin\}()$, is an algebraic function.

For the other three inverse trigonometric functions to be considered, we have:

Theorem 10.25:

$$\{D \arccos\}(t) = \left\{ \frac{-1}{\sqrt{1-t^2}} \right\}(t), \quad 0 < t < 1.$$

Proof:

For any number t in the domain of $\arccos$ (),

$$\arccos t = \frac{\pi}{2} - \arcsin t,$$

since

$$\cos\left(\frac{\pi}{2} - \arcsin t\right) = \sin(\arcsin t) = t.$$

Thus,

$$\{D \arccos\}(t) = -\{D \arcsin\}(t) = \frac{-1}{\sqrt{1-t^2}}$$

Theorem 10.26:

$$\{D \arctan\}(t) = \left\{ \frac{1}{1+t^2} \right\}(t), \quad t \geq 0.$$

Proof:

For any number t in the domain of $\arctan$ (),

$$\begin{aligned} \{D \arctan\}(t) &= \frac{1}{\{D \tan\}(\arctan t)} \\ &= \frac{1}{\{\sec^2\}(\arctan t)} \\ &= \frac{1}{\{1 + \tan^2\}(\arctan t)} \\ &= \frac{1}{1 + \{\tan^2\}(\arctan t)} \\ &= \frac{1}{1 + [\tan(\arctan t)]^2} \\ &= \frac{1}{1 + t^2} \\ &= \left\{ \frac{1}{1+t^2} \right\}(t), \quad t \geq 0. \end{aligned}$$

The reasons for each step follows in a manner similar to that in the preceding proof.

Theorem 10.27:

$$\{D \operatorname{arcsec}\}(t) = \left\{ \frac{1}{t\sqrt{t^2-1}} \right\}(t), \quad t > 1.$$

Proof:

If t is any number in the domain of arcsec (),

$$\{D \operatorname{arcsec}\}(t) = \frac{1}{\{D \sec\}(\operatorname{arcsec} t)}$$

$$\begin{aligned}
&= \frac{1}{\{\sec \tan\} \{\sec t\}} \\
&= \frac{1}{[\sec (\sec t)] \cdot [\tan (\sec t)]} \\
&= \frac{1}{t \cdot \tan (\sec t)} \\
&= \frac{1}{t \cdot \{\sqrt{\sec^2 - 1}\} (\sec t)} \\
&= \frac{1}{t \cdot \sqrt{\sec^2} (\sec t) - 1} \\
&= \frac{1}{t \cdot \sqrt{[\sec (\sec t)]^2 - 1}} \\
&= \frac{1}{t \sqrt{t^2 - 1}} \\
&= \left\{ \frac{1}{t \sqrt{t^2 - 1}} \right\} (t), \quad t > 1.
\end{aligned}$$

The reasons for the steps in the proofs of Theorems 10.26 and 10.27 are left as exercises.

Exercises

1. Supply the reasons in the proofs of Theorems 10.26 and 10.27.

Differentiate the functions specified by:

2. $2 \arctan (\sqrt{r-1})$.
3. $\arccos (1-x)$.
4. $\arccos \left(\frac{a-t}{t} \right)$.
5. $u \arcsin u + \sqrt{1-u^2}$.
6. $\arcsin v^2$.
7. $\arctan w^2$.
8. $\arctan \frac{1+x}{1-x}$.
9. $\tan \left(2 \arctan \frac{y}{2} \right)$
10. $\sin (\arcsin z)$.
11. If $f()$ is specified by

$$f(t) = \arcsin t + \arccos t,$$

show that $f()$ is a constant function. (Hint: find $f'()$.)

12. Prove that the graph of the function

$$\{\arctan + \arccos\}()$$

has no relative maximum or relative minimum points. Does this graph have any inflection points?

10.10. SOME ANTI-DIFFERENTIATION

Since the derivatives of the inverse trigonometric functions turn out to be algebraic functions, we may now extend our method of finding anti-derivatives. Previously, in the process of making a first guess at an anti-derivative of an algebraic function, we were restricted to guessing other algebraic functions. Now, we are able to try the inverse trigonometric functions. Of course, we have immediate anti-derivatives for functions given by expressions of the form

$$\frac{1}{\sqrt{1-t^2}}, \quad \frac{1}{1+t^2}, \quad \frac{1}{t\sqrt{t^2-1}},$$

since these functions are directly the derivatives of the functions $\text{arc Sin } ()$, $\text{arc Tan } ()$, and $\text{arc Sec } ()$.

Beyond these simple cases, though, we can apply further techniques.

EXAMPLE 1:

Find an anti-derivative of the function given by

$$\sqrt{1-t^2}, \quad 0 < t < 1.$$

As a first guess, let us try the function given by

$$g_1(t) = \text{arc Sin } t.$$

Then

$$g_1'(t) = \frac{1}{\sqrt{1-t^2}}.$$

We would like to have arrived at $\sqrt{1-t^2}$. It is certainly true that, for our first guess,

$$\begin{aligned} g_1'(t) &= \frac{1}{\sqrt{1-t^2}} + \sqrt{1-t^2} - \sqrt{1-t^2} \\ &= \sqrt{1-t^2} + \frac{1}{\sqrt{1-t^2}} - \sqrt{1-t^2} \\ &= \sqrt{1-t^2} + \frac{t^2}{\sqrt{1-t^2}}. \end{aligned}$$

If we then subtract from $g_1(t)$ some expression giving a function whose derivative is given by the negative of the undesired, extra term in $g_1'(t)$, then we would have the desired anti-derivative. Our problem then is to determine an anti-derivative of the function given by

$$\frac{t^2}{\sqrt{1-t^2}}.$$

We can write this last expression as

$$t^2 (1 - t^2)^{-1/2},$$

which suggests that this might be one of the two terms resulting from the differentiation of the product of two functions. In order to have $(1 - t^2)^{-1/2}$ as an expression involved in the derivative of some function, it would seem most reasonable that the function being differentiated contain the expression $(1 - t^2)^{1/2}$. But, if $f(t) = (1 - t^2)^{1/2}$, then $f'(t) = -t(1 - t^2)^{-1/2}$, by the chain rule, so that it is more reasonable to write our unwanted term in the form

$$t^2 (1 - t^2)^{-1/2}.$$

Considering this new expression as one of the terms involved in differentiating a product, we can make as a reasonable second guess:

$$g_2(t) = \arcsin t + t \sqrt{1 - t^2},$$

for which

$$\begin{aligned} g_2'(t) &= \sqrt{1 - t^2} + \frac{t^2}{\sqrt{1 - t^2}} + \sqrt{1 - t^2} - \frac{t^2}{\sqrt{1 - t^2}} \\ &= 2\sqrt{1 - t^2}. \end{aligned}$$

This new expression is precisely twice the original one for the derivative of the guessed-for function, so that an anti-derivative of the function given by the expression

$$\sqrt{1 - t^2}, \quad 0 < t < 1$$

is the function given by

$$\frac{1}{2} \arcsin t + \frac{1}{2} t \sqrt{1 - t^2}, \quad 0 < t < 1.$$

Needless to say, this process is much involved in making the right guess at the right time. A simpler way is to consider that the differentiation of the composite of two functions is involved. We have, as a beginning,

$$f'(t) = \sqrt{1 - t^2}.$$

For any function $g()$,

$$\begin{aligned} \{Df(g)\}(t) &= f'(g(t)) g'(t) \\ &= \sqrt{1 - [g(t)]^2} g'(t). \end{aligned}$$

The algebraic expression involving the square root is similar to that relating the functions $\sin()$ and $\cos()$, so that, for $g()$ the function $\sin()$, we have

$$\begin{aligned} \{D\{f(\sin)\}\}(t) &= \sqrt{1 - \sin^2 t} \cos t \\ &= \cos^2 t \\ &= \frac{1}{2} + \frac{1}{2} \cos 2t \end{aligned}$$

Then, taking an anti-derivative,

$$f(\sin t) = \frac{1}{2}t + \frac{1}{4}\sin 2t.$$

Since we want to find a rule for the function $f()$, we need to express the right side of this equation entirely in terms of $\sin t$. Thus,

$$\begin{aligned} f(\sin t) &= \frac{1}{2}t + \frac{1}{2}\sin t \cos t \\ &= \frac{1}{2}\arcsin(\sin t) + \frac{1}{2}\sin t \sqrt{1 - \sin^2 t}, \end{aligned}$$

so that $f()$ is defined by

$$f(z) = \frac{1}{2}\arcsin z + \frac{1}{2}z\sqrt{1 - z^2}.$$

EXAMPLE 2:

Find an anti-derivative of $\frac{t^2}{4 + t^2}$.

As a beginning,

$$f'(t) = \frac{t^2}{4 + t^2}.$$

For any function $g()$,

$$\begin{aligned} \{D\{f(g)\}\}(t) &= f'(g(t))g'(t) \\ &= \frac{[g(t)]^2}{1 + [g(t)]^2}g'(t). \end{aligned}$$

The denominator of this quotient suggests the relation between $\tan()$ and $\sec()$, so that, for $g()$ the function $\tan()$, we have

$$\begin{aligned} \{D\{f(\tan)\}\}(t) &= \frac{\tan^2 t}{1 + \tan^2 t} \sec^2 t \\ &= \tan^2 t \\ &= 1 - \sec^2 t. \end{aligned}$$

Then, taking an anti-derivative,

$$\begin{aligned} f(\tan t) &= t - \tan t \\ &= \arcsin(\tan t) - \tan t, \end{aligned}$$

so that $f()$ is defined by

$$f(z) = \arcsin z - z.$$

Finally, whether we attempt anti-differentiation by guessing or by the substitution involved in composite functions, we are always able to determine whether we have in fact found an anti-derivative by differentiating the function which is our result.

Exercises

Find anti-derivatives of the functions specified by:

1. $\frac{1}{4 + r^2}$.
2. $(1 - 9s^2)^{-1/2}$.
3. $(t^2 + 2t + 2)^{-1}$.
4. $(-2u - u^2)^{-1/2}$.
5. $(v^2 + 2v)^{-1} (v + 1)^{-1}$.
6. $w(w^4 - 1)^{-1/2}$.
7. $(b^2 - a^2x^2)^{-1/2}$, where $a > 0$ and $b > 0$.
8. $(b^2 + a^2y^2)^{-1}$, where $a > 0$ and $b > 0$.
9. $\frac{1 - z^2}{1 + z^2}$.

10. Find the area bounded by the graphs of

$$y = \sqrt{r^2 - x^2}, \quad y = 0, \quad x = 0, \quad \text{and} \quad x = a,$$

where $0 < a < r$.

11. Find the area of that part of a circle of radius r , cut off by a line perpendicular to a radius, drawn at a point a units from the center.
12. The graph of $b^2x^2 + a^2y^2 = a^2b^2$ is an ellipse. Find the area enclosed by this graph.

10.11. A METHOD OF REDOING THE TRIGONOMETRIC AND INVERSE TRIGONOMETRIC FUNCTIONS

One result of the Fundamental Theorem of Calculus (Theorem 7.9) is that any function specified by means of an integral, say $F()$, given by

$$F(x) = \int_0^x f,$$

is an anti-derivative of the function being integrated (that is, $F'() = f()$).

We may then *specify* the function $\arcsin ()$ by:

$$\arcsin x = \int_0^x \frac{1}{\sqrt{1 - t^2}} dt, \quad 0 \leq x < 1.$$

So that we have immediately,

$$\{D \arcsin\}(x) = \frac{1}{\sqrt{1 - x^2}}, \quad 0 \leq x < 1.$$

If we specify that the domain of this function, $\arcsin ()$, is the interval: all x , $0 \leq x \leq 1$ (making the specification that $\arcsin 1 = \pi/2$), then we may define the function $\sin ()$ as the inverse of $\arcsin ()$. The domain of the $\sin ()$ function is then the interval: all t , $0 \leq t \leq \pi/2$.

We may now create a new function, called "sin ()" (no capital), by extending the domain of the function Sin (). We simply say,

If $-\pi/2 \leq t < 0$, then $\sin t = -\text{Sin}(-t)$.

If $\pi/2 < t \leq \pi$, then $\sin t = \text{Sin}(\pi - t)$.

If $-\pi \leq t < -\pi/2$, then $\sin t = \text{Sin}(-t - \pi)$.

If $(n - 1)\pi \leq t \leq (n + 1)\pi$, where n is any integer,

$$\sin t = \text{Sin}(t - 2n\pi).$$

We will then have created the sine function, without any recourse to circles or to the really rather deep concept of the length of an arc along the circumference of a circle. We may then define the remaining trigonometric functions in terms of the sine function. In the next chapter, we will adopt this use of the Fundamental Theorem of Calculus to introduce a new kind of function.

The definition of the remaining trigonometric functions is accomplished from $\sin()$ by:

$$\cos t = \sin\left(\frac{\pi}{2} - t\right)$$

$$\tan t = \frac{\sin t}{\cos t}$$

$$\text{cosec } t = \frac{1}{\sin t}$$

$$\sec t = \frac{1}{\cos t}$$

$$\cotan t = \frac{\cos t}{\sin t},$$

where the domains include all numbers except those which would call for a division by 0.

LOGARITHMIC AND EXPONENTIAL FUNCTIONS

11.1. THE FUNCTION $\text{Log}()$

In the investigation of anti-derivatives of algebraic functions, there was one rather simple function, none of whose anti-derivatives could be found: the function $\left\{\frac{1}{I}\right\}()$. The number 0 is not in the domain of this function, but all non-zero numbers are in the domain. Further, the derivative of $\left\{\frac{1}{I}\right\}()$ is, by the theorem on the differentiation of $\{I^p\}()$, where p is any integer, the function $\left\{-\frac{1}{I^2}\right\}()$, whose domain also consists of the set of all non-zero numbers. Since differentiability implies continuity (Theorem 6.1), the function $\left\{\frac{1}{I}\right\}()$ is continuous. If we have any interval which does not include the number 0 either in its interior or as an end point, by Theorem 7.4 the integral of $\left\{\frac{1}{I}\right\}()$ over this interval will exist—that is, will be a number.

As a result, if x is any positive number greater than 1, $\int_1^x \left\{\frac{1}{I}\right\}()$ will exist, and be some positive number. If x is any number between 0 and 1, then $\int_1^x \left\{\frac{1}{I}\right\}()$ will exist and be a negative number, since interchanging the “limits of integration” is equivalent to a multiplication by -1 (from Definition 7.6). Obviously $\int_1^1 \left\{\frac{1}{I}\right\}()$ will be the number 0.

We now make the following definition:

Definition 11.1: The function $\text{Log}()$ is specified by:

- 1) the domain of $\text{Log}()$ is the set of all positive numbers.
- 2) For any positive number x ,

$$\text{Log } x = \int_1^x \left\{\frac{1}{I}\right\}() d().$$

As an immediate consequence of the Fundamental Theorem of Calculus, we have

Theorem 11.1:

$\{D \text{Log}\}()$ is the function specified by

$$\frac{1}{x}, \quad x > 0.$$

It should be noted that the derivative of the function $\text{Log}()$ is *not* the function $\{\frac{1}{t}\}()$, since the domain of the derivative of the $\text{Log}()$ function contains no negative numbers.

As a result of the fact that differentiability implies continuity, we have

Corollary to Theorem 11.1:

The function $\text{Log}()$ is continuous.

In attempting to sketch the graph of the $\text{Log}()$ function, we can make use of Theorem 11.1 to tell us that this function must be increasing at every number in its domain (since the value of the derivative is always positive). Further, the second derivative—given by $-\frac{1}{x^2}$, $x > 0$ —always has a negative value, so that the graph of the $\text{Log}()$ function is always concave down. Examination of the values of $\text{Log}()$ and $\{D \text{Log}\}()$ for numbers which are positive, but near 0, suggests that the y -axis is a vertical asymptote for this graph. This is in fact true, although we will omit a detailed proof. The graph of $\text{Log}()$, which is the graph of the equation $y = \text{Log } x$ is shown in Fig. 11.1.

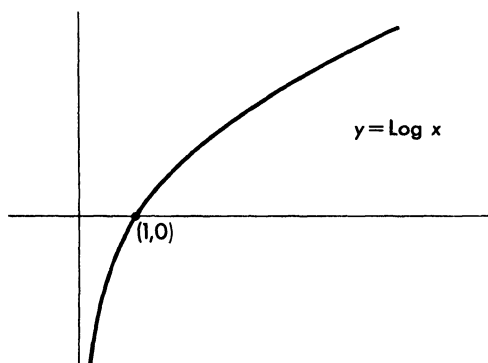


Fig. 11.1

Insofar as finding values of $\text{Log}()$ for various positive numbers, the best we can do at present is use the definition of integral to find approximations. For example, $\text{Log } 2$ is, by the definition (11.1), the integral

$$\int_1^2 \frac{1}{t} dt.$$

From the definition of area, this integral is the number of units of area of that part of the plane bounded by the graphs of $y = 1/x$, $y = 0$, $x = 1$, and $x = 2$. (See Fig. 11.2.)

This area is certainly greater than $1/2$ and less than $3/4$, since that part of the plane is larger than the rectangle with corners at $(1, 0)$, $(2, 0)$, $(2, 1/2)$, and $(1, 1/2)$ and smaller than the trapezoid with corners at $(1, 0)$,

$(2, 0)$, $(2, 1/2)$, and $(1, 1)$. Then we have $0.5 < \text{Log } 2 < 0.75$ in decimal form. We may narrow this estimate further by considering other triangular areas underneath the graph of $y = 1/x$.

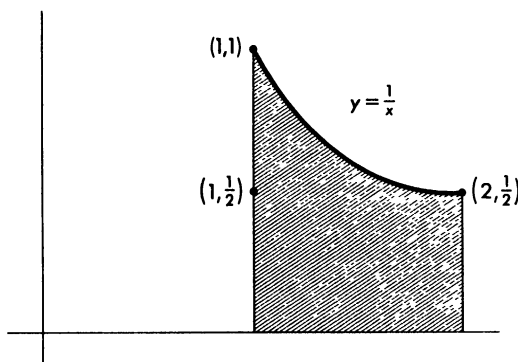


Fig. 11.2

Exercises

1. Let $f()$ be the function specified by $f(t) = 1/t$. Divide the interval: the set of all t , $1 \leq t \leq 2$ into 10 sub-intervals of equal length by the numbers 1.1, 1.2, ..., 1.9. Find the "upper sum" corresponding to this subdivision, which upper sum is one of the numbers in the range of values of the sequence whose limit is the upper integral of this function over this interval.
2. Subdivide the interval: the set of all t , $1 \leq t \leq 2$, as in exercise 1, and find the "lower sum" corresponding to this subdivision.
3. Why are the numbers found in exercises 2 and 3 reasonable approximations to the number $\text{Log } 2$?
4. Repeat exercises 1 and 2 for the interval: the set of all t , $2 \leq t \leq 3$, thus finding an approximation to the number $\int_2^3 \frac{1}{t} dt$.
5. Use the results of the first four exercises to find a reasonable approximation to $\text{Log } 3$.
6. Use the definitions of area and of $\text{Log } ()$ to show that, if $B > 1$, then $\text{Log } B > 1 - \frac{1}{B}$.
7. Improve the lower bound to $\text{Log } B$ found in exercise 6 by considering the point $(B/2, 2/B)$ in addition to the point $(B, 1/B)$, thus getting $\text{Log } B > 3/2 - 2/B$.
8. Does the graph of the equation $y = \text{Log } x$ have any relative maximum or relative minimum points? Why or why not?
9. Does the graph of the equation $y = \text{Log } x$ have any points of inflection? Why or why not?
10. If n is a natural number, find an expression giving the n -th derivative of $\text{Log } ()$.
11. Use anti-differentiation by parts to find an anti-derivative of $\text{Log } ()$.

Find an anti-derivative of the function specified by:

12. $\frac{1}{r}, \quad r < 0.$

13. $\frac{1}{s+1}, \quad s > -1.$

14. $\frac{1}{2t+1}, \quad t > -1/2.$

15. $\frac{u}{u^2+1}.$

16. $\frac{\sin v}{\cos v}, \quad -\pi/2 < v < \pi/2.$

17. $\tan w, \quad \pi/2 < w < \pi.$

18. $\sec x, \quad 0 < x < \pi/2. \quad \left(\text{Hint: } \sec x = \frac{\sec^2 x + \sec x \tan x}{\sec x + \tan x} \right).$

19. $\frac{3y^2+2y-4}{y-5}, \quad y > 5.$

20. $\frac{z+1}{z-1}, \quad z > 1.$

11.2. LOGARITHMIC PROPERTIES OF LOG ()

The new function, Log (), has properties in common with the operation of taking logarithms of a number (which was the reason for giving the name Log ()). We have:

Theorem 11.2:

If $a > 0$ and $x > 0$, then

$$\text{Log}(ax) = \text{Log } a + \text{Log } x.$$

Proof:

1. Let the function $F()$ be given by

$$F(x) = \text{Log}(ax) - \text{Log } x, \quad x > 0.$$

2. $F'(x) = \frac{1}{ax} a - \frac{1}{x}, \quad x > 0$

by Theorem 11.1 and the chain rule.

3. $F'(x) = 0, \quad x > 0$

by step 2 and algebraic manipulation.

4. $F()$ is a constant function

by the theorem that any anti-derivative of the function $\theta()$ is a constant function.

5. $F(1) = \text{Log } a - \text{Log } 1$

by step 1.

6. $F(1) = \text{Log } a$

by the fact that $\text{Log } 1 = 0$.

7. $F()$ is a constant function whose range of values consists of the single number, $\text{Log } a$ by steps 4 and 6.
8. $\text{Log } a = \text{Log } (ax) - \text{Log } x$ by steps 1 and 7.
9. $\text{Log } (ax) = \text{Log } a + \text{Log } x$ by step 8 and algebraic manipulation.

Theorem 11.3:

If $x > 0$, then $\text{Log } (1/x) = -\text{Log } x$.

Proof:

1. Let the function $G()$ be given by

$$G(x) = \text{Log } (1/x) + \text{Log } x, x > 0.$$
2. $G'(x) = \frac{1}{1/x} \left(-\frac{1}{x^2} \right) + \frac{1}{x}, x > 0$ by Theorem 11.1 and the chain rule.
3. $G'(x) = 0, x > 0$ by step 2 and algebraic manipulation.
4. $G()$ is a constant function by step 3.
5. $G(1) = \text{Log } 1 + \text{Log } 1$ by step 1.
6. $G(1) = 0$ by step 5 and the fact that $\text{Log } 1 = 0$.
7. $G()$ is a constant function whose range of values consists of the single number 0.
8. $0 = \text{Log } (1/x) + \text{Log } x$ by steps 1 and 7.
9. $\text{Log } (1/x) = -\text{Log } x$ by step 8 and algebraic manipulation.

Theorem 11.4:

If $a > 0$ and $b > 0$,

$\text{Log } (a/b) = \text{Log } a - \text{Log } b$.

Proof:

1. $a/b = a(1/b)$ by the definition of quotient.
2. $\text{Log } (a/b) = \text{Log } (a(1/b))$ by step 1.
3. $\text{Log } (a/b) = \text{Log } a + \text{Log } (1/b)$ by step 2 and Theorem 11.2.
4. $\text{Log } (a/b) = \text{Log } a - \text{Log } b$ by step 3 and Theorem 11.3.

Theorem 11.5:

If $x > 0$ and p is any rational number,

$\text{Log } (x^p) = p \text{Log } x$.

Proof:

1. Let the function $H()$ be given by

$$H(x) = \text{Log } (x^p) - p \text{Log } x, x > 0.$$

2. $H'(x) = \frac{1}{x^p} p x^{p-1} - p \frac{1}{x}, x > 0$ by Theorem 11.1 and the chain rule.
3. $H'(x) = 0, x > 0$ by step 2 and algebraic manipulation.
4. $H()$ is a constant function by step 3.
5. $H(1) = \text{Log } 1 - p \text{Log } 1$ by step 1.
6. $H(1) = 0$ by step 5 and the fact that $\text{Log } 1 = 0$.
7. $H()$ is a constant function whose range of values consists of the single number 0 by steps 4 and 6.
8. $0 = \text{Log } (x^p) - p \text{Log } x$ by steps 1 and 7.
9. $\text{Log } (x^p) = p \text{Log } x$ by step 8 and algebraic manipulation.

An approximation, accurate to four decimal places of two values of the Log () function is given by:

$$\text{Log } 2 \approx 0.6931;$$

$$\text{Log } 3 \approx 1.0986.$$

Using these values, we can find

$$\begin{aligned} \text{Log } 4 &= \text{Log } (2^2) \\ &= 2 \text{Log } 2 \text{ (by Theorem 11.5)} \\ &\approx 1.3862. \end{aligned}$$

$$\begin{aligned} \text{Log } 6 &= \text{Log } ((2) (3)) \\ &= \text{Log } 2 + \text{Log } 3 \text{ (by Theorem 11.2)} \\ &\approx 1.7917. \end{aligned}$$

$$\begin{aligned} \text{Log } (1.5) &= \text{Log } (3/2) \\ &= \text{Log } 3 - \text{Log } 2 \text{ (by Theorem 11.4)} \\ &\approx 0.4045. \end{aligned}$$

Exercises

Using the approximations given above, together with

$$\text{Log } 5 \approx 1.6094, \quad \text{Log } 7 \approx 1.9459,$$

find approximations to:

- | | | |
|---------------------|---------------------------|-----------------|
| 1. Log 8. | 2. Log 9. | 3. Log 10. |
| 4. Log 12. | 5. Log 14. | 6. Log 15. |
| 7. Log 20. | 8. Log 30. | 9. Log 40. |
| 10. Log 50. | 11. Log 60. | 12. Log 100. |
| 13. Log (1/2). | 14. Log (1/3). | 15. Log (2/3). |
| 16. Log (1/4). | 17. Log (3/4). | 18. Log (1/5). |
| 19. Log (0.1). | 20. Log (0.2), | 21. Log (0.01). |
| 22. Log (0.000001). | 23. Log (0.000000000003). | |

11.3. THE EXPONENTIAL FUNCTION

We now use the function $\text{Log } ()$ to define another new function, which will be the inverse of $\text{Log } ()$. We have, then

Definition 11.2: The function $\exp ()$ is defined as the inverse of the function $\text{Log } ()$.

Since the domain of $\text{Log } ()$ is the set of all positive numbers, the range of values of $\exp ()$ must be the set of all positive numbers. Since the range of values of $\text{Log } ()$ is the set of all numbers, the domain of $\exp ()$ must be the set of all numbers. Because of the definition of inverse of a function, we must have

$$\begin{aligned}\text{Log } (\exp t) &= t, \quad \text{for any number } t, \\ \exp (\text{Log } x) &= x, \quad \text{for any positive number } x.\end{aligned}$$

The inverse of any function is the collection of ordered pairs which results from reversing the order of each ordered pair of the original function. The graph of $\exp ()$ then must be the set of points which results from interchanging the x and y -coordinates of the points on the graph of $\text{Log } ()$. (See Fig. 11.3.)

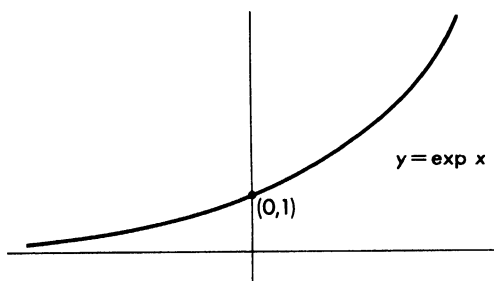


Fig. 11.3

Because of the theorem on the relation between the derivative of a function and the derivative of the inverse of that function, we have:

Theorem 11.6:

$$\{D \{\exp\}\} () = \{\exp\} ().$$

Proof:

$$\begin{aligned}\{D \exp\} () &= \{D \{\text{Log}^{-1}\}\} () && \text{by the definition (11.2),} \\ &= \left\{ \frac{1}{\{D \text{Log}\} (\exp)} \right\} () && \text{by the definition of inverse functions,} \\ &= \left\{ \frac{1}{\frac{1}{\exp}} \right\} () && \text{by Theorem 11.1,} \\ &= \exp () && \text{by algebra.}\end{aligned}$$

This gives us the first non-constant function we have encountered (and, in fact, the only one other than the constant function, $\theta()$, in existence), which is its own derivative.

Exercises

Find the derivatives of the functions specified by:

- | | |
|----------------------|------------------------------|
| 1. $\exp(3r)$. | 6. $\arctan w \exp(w + 1)$. |
| 2. $s \exp s$. | 7. $(\exp x)^3$. |
| 3. $t^2 \exp(2t)$. | 8. $\exp(\sin y)$. |
| 4. $\log u \exp u$. | 9. $\exp(-z)$. |
| 5. $\cos v \exp v$. | |

Find anti-derivatives of the functions specified by:

- | | |
|-----------------------------|------------------------------------|
| 10. $\exp(2r)$. | 15. $\exp(-w)$. |
| 11. $s \exp(s^2)$. | 16. $\exp(5x - 4)$. |
| 12. $t^2 \exp(t^2)$. | 17. $y^{-1} \exp(\log y), y > 0$. |
| 13. $u \exp u$. | 18. $\log(\exp z)$. |
| 14. $\cos v \exp(\sin v)$. | |

Sketch the graphs of the following equations, labeling all relative extremes and inflection points:

19. $y = x \exp x$.
20. $y = x^2 \exp x$.
21. $y = x^3 \exp x$.

11.4. PROPERTIES OF THE EXPONENTIAL FUNCTION

The new function, $\exp()$, has properties similar to that of the operation of raising to powers, as we investigated those properties in Chapter 8, limiting our investigation to powers which were rational numbers. Since the domain of $\exp()$ consists of the set of all numbers, we need not make any restriction on our new "powers" or "exponents." We have

Theorem 11.7:

If a and b are any numbers,

$$\exp(a + b) = \exp a \exp b.$$

Proof:

$$\begin{aligned} \log(\exp a \exp b) &= \log(\exp a) + \log(\exp b) && \text{by Theorem 11.2,} \\ &= a + b && \text{by the definition (11.2).} \\ \exp(a + b) &= \exp(\log(\exp a \exp b)) \\ &= \exp a \exp b && \text{by the definition (11.2).} \end{aligned}$$

Theorem 11.8:

If a is any number,

$$\exp(-a) = \frac{1}{\exp a}.$$

Proof:

$$\begin{aligned}\operatorname{Log}\left(\frac{1}{\exp a}\right) &= -\operatorname{Log}(\exp a) \\ &= -a\end{aligned}$$

by Theorem 11.3,

by the definition (11.2).

$$\begin{aligned}\exp(-a) &= \exp\left[\operatorname{Log}\left(\frac{1}{\exp a}\right)\right] \\ &= \frac{1}{\exp a}\end{aligned}$$

by the definition (11.2).

Theorem 11.9:

$$\exp(a - b) = \frac{\exp a}{\exp b}.$$

Proof:

Left as an exercise.

Theorem 11.10:

$\exp(ap) = (\exp a)^p$, where p is any rational number.

Proof:

$$\operatorname{Log}(\exp(ap)) = ap$$

by the definition (11.2).

$$\begin{aligned}\operatorname{Log}((\exp a)^p) &= p \operatorname{Log}(\exp a) \\ &= ap\end{aligned}$$

by Theorem 11.6,

by the definition (11.2).

$$\begin{aligned}\exp(ap) &= \exp[\operatorname{Log}((\exp a)^p)] \\ &= (\exp a)^p.\end{aligned}$$

We are now able to define an expression of the form a^b , where a is any positive number, and b is any number whatsoever, rational or irrational. If a is positive, then a is in the range of values of $\exp()$. Thus, there does exist some number, c , such that $\exp c = a$. We then define a^b as $\exp(bc)$.

Exercises

1. Prove Theorem 11.9.
2. Give the reasons for the latter part of the proof of Theorem 11.10.

The value of $\exp 1$ is approximately 2.718. Find approximations to:

3. $\exp 2$.
4. $\exp 3$.
5. $\exp 4$.
6. $\exp(-1)$.
7. $\exp(-2)$.

8. $\exp(0)$.

9. $\exp(64)$.

10. Show that, if $a > 0$, then $a^b = \exp(b \operatorname{Log} a)$.

11.5. THE IRRATIONAL NUMBER e

The easiest number, of all numbers, with which to do arithmetic—and thus algebra—is the number 0. The next easiest is 1. For the new, non-algebraic functions, which we have created, we note that $\operatorname{Log} 1$ is the number 0, and $\operatorname{Log} 0$ is meaningless. For the function $\exp()$, we have $\exp 0 = 1$, but, as yet, have no way of getting at $\exp 1$. We make the definition:

Definition 11.3: $e = \exp 1$.

This number e is a fixed, definite number (that is, a constant), and turns out to be vitally important for large branches of mathematics. From our extension of the idea of raising to powers, following Theorem 11.10, we may write

$$\exp x = e^x.$$

In attempting to find approximations to this number e , we have

Theorem 11.11:

$$e = \lim_n \left[\left(1 + \frac{1}{n} \right)^n \right]$$

Proof:

$$\{D \operatorname{Log}\}(x) = \frac{1}{x}, \quad x > 0 \quad \text{by Theorem 11.1.}$$

Thus,

$$1 = \{D \operatorname{Log}\}(1) \quad \text{since 1 is in the domain of both functions.}$$

$$1 = \lim_{h \rightarrow 0} \frac{\operatorname{Log}(1+h) - \operatorname{Log} 1}{h} \quad \text{by the definition of derivative.}$$

$$1 = \lim_{h \rightarrow 0} \frac{1}{h} [\operatorname{Log}(1+h)] \quad \text{by the fact that } \operatorname{Log} 1 = 0.$$

$$1 = \lim_{h \rightarrow 0} [\operatorname{Log}((1+h)^{1/h})] \quad \text{by Theorem 11.5.}$$

$$1 = \operatorname{Log} \left[\lim_{h \rightarrow 0} (1+h)^{1/h} \right] \quad \text{by the fact that } \operatorname{Log}() \text{ is continuous.}$$

$$\exp(1) = \exp(\operatorname{Log}[\lim_{h \rightarrow 0} (1+h)^{1/h}]) \quad \text{by taking } \exp() \text{ of each side.}$$

$$e = \lim_{h \rightarrow 0} (1+h)^{1/h} \quad \text{by the definitions (11.2) and (11.3).}$$

$$e = \lim_{t \rightarrow +\infty} \left(1 + \frac{1}{t} \right)^t \quad \text{by the substitution } t = 1/h.$$

$$e = \lim_n \left[\left(1 + \frac{1}{n} \right)^n \right] \quad \text{by restricting the range of the variable } t \text{ to the natural numbers.}$$

Taking approximations we find:

$$(1 + 1)^1 = 2$$

$$(1 + \frac{1}{2})^2 = 2.25$$

$$(1 + \frac{1}{3})^3 \approx 2.37037$$

$$(1 + \frac{1}{5})^5 \approx 2.48832$$

$$(1 + \frac{1}{10})^{10} \approx 2.59374$$

$$(1 + \frac{1}{100})^{100} \approx 2.70493$$

$$(1 + \frac{1}{1000})^{1000} \approx 2.71807$$

$$(1 + \frac{1}{1000000})^{1000000} \approx 2.71828$$

The number e is irrational, but the proof of this fact is not simple. There exist extensive tables of approximations to values of e^x , where x is some number. A small portion of such a table is

x	e^x
- 10	0.00005
- 5	0.00674
- 3	0.94979
- 2	0.1353
- 1	0.3679
0	1
1	2.718
2	7.389
3	20.90
4	54.60
5	148.4
10	22,026.

11.6. THE EXTENDED EXPONENTIAL FUNCTIONS

Since we have defined e^x as $\exp x$, where x is any number whatsoever (irrational as well as rational), we have a new method for finding any "power" of a *positive* number. If a is a particular, fixed, positive number, a must be in the range of values of the function $\exp()$, and thus must be in the domain of the function $\text{Log}()$,

But the functions $\text{Log}()$ and $\exp()$ are inverses of each other; so, for the positive number a in the domain of $\text{Log}()$, we must have

$$a = \exp(\text{Log } a).$$

We can now make the following definition:

Definition 11.4: If $a > 0$ and x is any number,

$$a^x \text{ is the number } \exp(x \text{Log } a).$$

This definition merely extends those already in force for a^x , where $a > 0$ and x is a rational number. If x is the rational number p/q , where p and q are integers, we have, from previous definitions

$$a^{p/q} = (a^{1/q})^p$$

$a^{1/q}$ is the non-negative number such that

$$(a^{1/q})^q = a.$$

Here we have

$$a^{1/q} = \exp\left(\frac{1}{q} \text{Log } a\right), \text{ which is non-negative.}$$

$$\begin{aligned} (a^{1/q})^q &= \left[\exp\left(\frac{1}{q} \text{Log } a\right)\right]^q \\ &= \exp\left(q \frac{1}{q} \text{Log } a\right) \quad (\text{by Theorem 11.10}) \\ &= \exp(\text{Log } a) \\ &= a. \end{aligned}$$

$$\begin{aligned} (a^{1/q})^p &= \left[\exp\left(\frac{1}{q} \text{Log } a\right)\right]^p \\ &= \exp\left(\frac{p}{q} \text{Log } a\right) \quad (\text{by Theorem 11.10}) \\ &= a^{p/q}. \end{aligned}$$

If, in this definition, we consider x as a variable whose range is the set of all numbers, we may construct a function consisting of the set of all possible ordered pairs of the form

$$(x, a^x),$$

where the expression a^x , $\exp(x \text{Log } a)$, and $e^{x \text{Log } a}$ are all symbols for the same (positive) number.

The derivative of the function so constructed is given in

Theorem 11.12:

If $a > 0$ and the function $g()$ is given by

$$g(x) = a^x,$$

then $g'()$ is given by

$$g'(x) = a^x \text{Log } a.$$

Proof:

- | | |
|---|--|
| 1. $g(x) = a^x$ | by hypothesis. |
| 2. $g(x) = \exp(x \text{Log } a)$ | by the definition (11.3). |
| 3. $g'(x) = \exp(x \text{Log } a) D(x \text{Log } a)$ | by the chain rule, and Theorem 11.6. |
| 4. $g'(x) = \exp(x \text{Log } a) \text{Log } a$ | by the fact that $\{D\{kI\}\}() = k()$. |
| 5. $g'(x) = a^x \text{Log } a$ | by step 4 and the definition (11.3). |

This theorem suggests that any function specified by an expression containing a variable written as an exponent can best be handled, for the purposes of differentiation or anti-differentiation, by rewriting the expression so that it contains only exponents (simple or not) of the number e . For example, consider the function $f()$, defined by

$$f(x) = x^x.$$

In order to find the derivative $f'()$, we may rewrite:

$$f(x) = e^{x \operatorname{Log} x}.$$

Using the chain rule, we get

$$\begin{aligned} f'(x) &= e^{x \operatorname{Log} x} D(x \operatorname{Log} x) \\ &= x^x D(x \operatorname{Log} x) \\ &= x^x (\operatorname{Log} x + 1). \end{aligned}$$

As a more complicated example,

$$\begin{aligned} D(\sin x)^{\cos x} &= D \exp(\cos x \operatorname{Log}(\sin x)) \\ &= \exp(\cos x \operatorname{Log}(\sin x)) D(\cos x \operatorname{Log}(\sin x)) \\ &= (\sin x)^{\cos x} \left(\cos x \frac{\cos x}{\sin x} - \sin x \operatorname{Log}(\sin x) \right) \\ &= (\sin x)^{\cos x} (\cos x \cot x - \sin x \operatorname{Log}(\sin x)). \end{aligned}$$

As a further use, we can now show

$$\{I^n\}'() = \{n I^{n-1}\}()$$

for *any* (real) number n , rather than merely for rational exponents. If we have the function $f()$ given by

$$f(x) = x^n,$$

then:

$$\begin{aligned} f(x) &= \exp(n \log x) \quad \text{for } x > 0 \\ f'(x) &= \exp(n \log x) \frac{n}{x}, \quad x > 0 \quad (\text{by the chain rule}) \\ &= x^n \frac{n}{x} \\ &= n x^{n-1} \quad \text{for } x > 0. \end{aligned}$$

Further,

$$\begin{aligned} f(x) &= [(-1)(-x)]^n, \quad x < 0 \\ &= (-1)^n (-x)^n, \quad x < 0 \\ &= (-1)^n \exp(n \log[-x]), \quad x < 0. \end{aligned}$$

$$\begin{aligned}
 f'(x) &= (-1)^n \exp(n \log[-x]) n \frac{-1}{-x}, \quad x < 0 \quad (\text{by the chain rule}) \\
 &= (-1)^n (-x)^n \frac{n}{x}, \quad x < 0 \\
 &= (-1)^n (-1)^n x^n \frac{n}{x}, \quad x < 0 \\
 &= n x^{n-1}, \quad x < 0.
 \end{aligned}$$

At 0, if 0 is in the domain of $f()$ (that is, if $n > 0$),

$$f'(0) = \lim_{h \rightarrow 0} \frac{f(h) - f(0)}{h} = \lim_{h \rightarrow 0} \frac{h^n}{h}.$$

For $n > 1$,

$$f'(0) = \lim_{h \rightarrow 0} \frac{h}{h} h^{n-1} = 0.$$

For $0 < n < 1$, $f'(0)$ does not exist.

Exercises

Find the derivatives of the functions specified by:

- | | |
|---------------------|------------------------|
| 1. 2^x . | 6. $(w^2 - 1)^w$. |
| 2. $(\sqrt{3})^x$. | 7. $(x^2 + 1)^{2x}$. |
| 3. 4^{2t} . | 8. $[(y^2 + 1)^y]^2$. |
| 4. $(16)^u$. | 9. $(z^2 + 1)z^2$. |
| 5. $(v + 2)^v$. | |

A function $f()$ has the property that

$$f'() = kf(),$$

where k is some constant. Find $f()$ when:

10. $f(0) = 1$, $f(1) = e$.
11. $f(0) = 1$, $f(1) = e^2$.
12. $f(0) = 20$, $f(1) = 400$.
13. $f(1) = 10$, $f(2) = 100$.
14. Find the area of that part of the plane bounded by the graphs of $y = \sin x 2^{\cos x}$, $y = 0$, $x = 0$, and $x = \pi$.

11.7. THE HYPERBOLIC FUNCTIONS

The graph of the equation $x^2 - y^2 = 1$ is a hyperbola, with one vertex at the point $(1, 0)$, which hyperbola is asymptotic to the straight lines with equations $y = x$ and $y = -x$. (For a derivation of this fact, see the appendix.) This graph is shown in Fig. 11.4.

From the definition of the exponential function, we have the following relations:

$$(e^t + e^{-t})^2 = e^{2t} + 2 + e^{-2t}.$$

$$(e^t - e^{-t})^2 = e^{2t} - 2 + e^{-2t}.$$

$$(e^t + e^{-t})^2 - (e^t - e^{-t})^2 = 4.$$

$$\frac{(e^t + e^{-t})^2}{4} - \frac{(e^t - e^{-t})^2}{4} = 1.$$

$$\left[\frac{e^t + e^{-t}}{2} \right]^2 - \left[\frac{e^t - e^{-t}}{2} \right]^2 = 1.$$

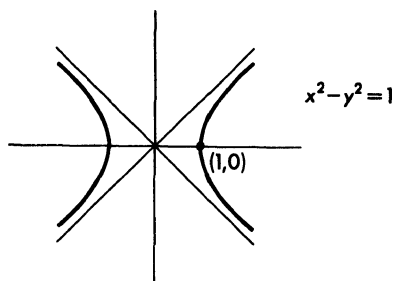


Fig. 11.4

This last relation implies that, if t is any number, the point with coordinates $\left(\frac{e^t + e^{-t}}{2}, \frac{e^t - e^{-t}}{2} \right)$ is on the hyperbola which is the graph of $x^2 - y^2 = 1$.

We defined the cosine and sine functions by using the x and y -coordinates of a point on the circle which is the graph of the equation $x^2 + y^2 = 1$. For this reason, these two functions may be considered as “circular” functions. We now give:

Definition 11.6: The functions $\cosh()$, $\sinh()$, $\tanh()$, and $\coth()$ are specified by:

$$\cosh(t) = \frac{e^t + e^{-t}}{2}, \quad t \text{ any number}$$

$$\sinh(t) = \frac{e^t - e^{-t}}{2}, \quad t \text{ any number}$$

$$\tanh(t) = \frac{\sinh t}{\cosh t}, \quad t \text{ any number}$$

$$\coth(t) = \frac{\cosh t}{\sinh t}, \quad t \neq 0.$$

Then the point with coordinates $(\cosh t, \sinh t)$ will be on the hyperbola which is the graph of the equation $x^2 - y^2 = 1$. For this reason, we call the function $\cosh()$, the hyperbolic cosine; the function $\sinh()$, the hyperbolic sine; the function $\tanh()$, the hyperbolic tangent; and the function $\coth()$, the hyperbolic cotangent. From the construction of these functions we then have

$$\cosh^2 t - \sinh^2 t = 1.$$

The domains of the functions $\cosh()$, $\sinh()$, and $\tanh()$ are, in each case, the set of all numbers, and the domain of the function $\coth()$ consists of the set of all non-zero numbers. In order to find the range of values of these functions, we must examine the solution of certain quadratic equations.

For example, if $\cosh t = x$, we have

$$\frac{e^t + e^{-t}}{2} = x.$$

$$\begin{aligned}
 e^t + e^{-t} &= 2x \\
 e^t(e^t + e^{-t}) &= 2xe^t \\
 e^{2t} + 1 &= 2xe^t \\
 e^{2t} - 2xe^t + 1 &= 0 \\
 (e^t)^2 - 2x(e^t) + 1 &= 0 \\
 e^t &= x + \sqrt{x^2 - 1}.
 \end{aligned}$$

The expression on the right side of this last equation is a number only of $|x| \geq 1$. The expression on the left side is always a positive number. Thus, in order for this equation to be a true statement, x must be a positive number, greater than or equal to 1. This implies that in order for x to be a number in the range of values of the function $\cosh()$, we must have $x \geq 1$. Conversely, if $x \geq 1$, then the equation $\cosh t = x$ does have a solution for t ; namely, $t = \text{Log}(x + \sqrt{x^2 - 1})$, and x is in the range of values of the function $\cosh()$. Thus, the range of values of the hyperbolic cosine function is the interval:

the set of all x , $x \geq 1$.

A similar investigation will show that the range of values of the function $\sinh()$ is the set of all numbers.

For the function $\tanh()$, we have

$$\begin{aligned}
 \tanh t &= x \\
 \frac{e^t - e^{-t}}{e^t + e^{-t}} &= x \\
 e^t - e^{-t} &= xe^t + xe^{-t} \\
 (1 - x)e^t - (1 + x)e^{-t} &= 0 \\
 e^t[(1 - x)e^t - (1 + x)e^{-t}] &= 0 \\
 (1 - x)e^{2t} - (1 + x) &= 0 \\
 e^{2t} &= \frac{x + 1}{1 - x}
 \end{aligned}$$

The expression, e^{2t} , on the left side of this last equation is always positive, so that the range of the variable x must be the solution of the inequality

$$\frac{x + 1}{1 - x} > 0.$$

For this inequality, we have:

$$\begin{aligned}
 \text{if } x < -1, \text{ then } & \frac{x + 1}{1 - x} < 0; \\
 \text{if } x = -1, \text{ then } & \frac{x + 1}{1 - x} = 0; \\
 \text{if } -1 < x < 1, \text{ then } & \frac{x + 1}{1 - x} > 0; \\
 \text{if } x = 1, \text{ then } & \frac{x + 1}{1 - x} \text{ is not a number;} \\
 \text{if } x > 1, \text{ then } & \frac{x + 1}{1 - x} < 0.
 \end{aligned}$$

Thus the range of values of the function $\tanh ()$ is the interval:

$$\text{the set of all } x, \quad -1 < x < 1.$$

In like manner, we can show that the range of values of the function $\coth ()$ is the set consisting of the two intervals:

$$\text{the set of all } x, \quad x < -1 \quad \text{and} \quad \text{the set of all } x, \quad x > 1.$$

Since the hyperbolic functions are algebraic combinations of the exponential functions given by e^t and e^{-t} , these new functions are continuous and differentiable, and their derivatives are given by:

$$\{D \sinh\}() = \cosh ()$$

$$\{D \cosh\}() = \sinh ()$$

$$\{D \tanh\}() = \left\{ \frac{1}{\cosh^2} \right\} ()$$

$$\{D \coth\}() = \left\{ \frac{-1}{\sinh^2} \right\} ()$$

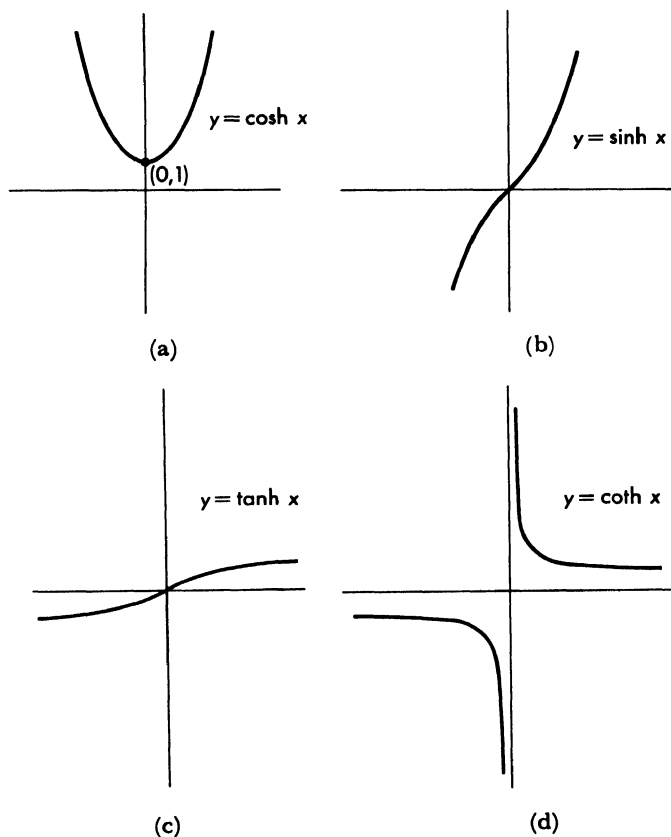


Fig. 11.5

The graphs of the hyperbolic functions are shown in Fig. 11.5. From these graphs it can be seen that the functions $\sinh ()$ and $\tanh ()$ have inverses, while the function $\cosh ()$ has no inverse, and the function $\coth ()$ has an inverse whose domain consists of two disjointed intervals.

Exercises

1. Show that the range of values of the function $\coth ()$ is the set consisting of the two intervals:

the set of all x , $x < -1$

and the set of all x , $x > 1$.

2. Use the fact that $\lim_{z \rightarrow +\infty} e^z = +\infty$ to show that

$$\lim_{r \rightarrow -\infty} \tanh r = 1.$$

3. Use the definition to prove that

$$D \sinh () = \cosh ();$$

$$D \cosh () = \sinh ();$$

$$D \tanh () = \frac{1}{\cosh^2} ();$$

$$D \coth () = \frac{-1}{\sinh^2} ().$$

Find the derivatives of the functions specified by:

4. $\cosh^2 r$.

9. $\text{Log} (\cosh w)$.

5. $\cosh (s^2)$.

10. $\text{Log} (\sinh x)$.

6. $\tanh (\tan t)$.

11. $\tanh^2 y + (\cosh^2 y)^{-1}$.

7. $\sinh (\text{arc Sin } u)$.

12. $\coth^2 z - (\sin^2 z)^{-1}$.

8. $e^v \cosh v$.

Show that

13. $\sinh (a + b) = \sinh a \cosh b + \cosh a \sinh b$.

14. $\cosh (a + b) = \cosh a \cosh b + \sinh a \sinh b$.

15. $\cosh p + \sinh p = e^p$.

16. $(\cosh r + \sinh r)^n = \cosh nr + \sinh nr$.

Find anti-derivatives of the functions specified by:

17. $\tanh r$.

19. $t \sinh t^2$.

18. $s \cosh s$.

20. $u^2 \cosh u^2$.

11.8. THE INVERSE HYPERBOLIC FUNCTIONS

As we noted in the preceding section, the functions $\sinh()$ and $\tanh()$ have inverses. We call these inverse functions $\sinh^{-1}()$ and $\tanh^{-1}()$, respectively. Because of the definition of the hyperbolic functions, it is possible to find explicit rules for these two functions.

Theorem 11.13:

If x is any number,

$$\sinh^{-1} x = \text{Log} (x + \sqrt{x^2 + 1}).$$

Proof:

1. Let x be a number in the domain of $\sinh^{-1}()$.
2. There is a number t in the range of values of $\sinh^{-1}()$, such that

$$\sinh^{-1} x = t$$

$$3. \sinh t = x$$

$$4. x = \frac{e^t - e^{-t}}{2}$$

$$5. e^t - e^{-t} = 2x$$

$$6. e^t (e^t - e^{-t}) = 2x e^t$$

$$7. e^{2t} - 1 = 2x e^t$$

$$8. e^{2t} - 2x e^t - 1 = 0$$

$$9. (e^t)^2 - 2x (e^t) - 1 = 0$$

10. If z is a variable, the equation

$$z^2 - 2xz - 1 = 0$$

has the two solutions given by

$$\text{and } z = x + \sqrt{x^2 + 1}$$

$$z = x - \sqrt{x^2 + 1}$$

11. Either

$$\text{or } e^t = x + \sqrt{x^2 + 1}$$

$$e^t = x - \sqrt{x^2 + 1}$$

12. $e^t > 0$

by step 1 and the definition of function.

by step 2 and the definition of the inverse of a function.

by step 3 and the definition of $\sinh()$.

by step 4 and algebraic laws.

by step 5 and law M-1.

by step 6 and the rules for exponents under multiplication.

by step 7 and algebraic laws.

by step 8 and the rules for exponents under multiplication.

by the theorem (2.2) on the solutions of a quadratic equation.

by steps 9 and 10.

by the fact that the range of values of the exponential function consists only of positive numbers.

13. If $x > 0$, then

$$\sqrt{x^2 + 1} > x$$

by the laws governing inequalities.
14. If $x > 0$, then

$$x - \sqrt{x^2 + 1} < 0$$

by step 13 and the laws governing inequalities.
15. If $x < 0$, then

$$x - \sqrt{x^2 + 1} < 0$$

by the fact that both terms in the expression are negative.
16. If $x = 0$,

$$x - \sqrt{x^2 + 1} = -1 < 0.$$
17. For every value of x ,

$$x - \sqrt{x^2 + 1} < 0$$

by steps 13, 14, and 15.
18. $e^t = x + \sqrt{x^2 + 1}$

by steps 11, 12, and 17.
19. $t = \text{Log } (x + \sqrt{x^2 + 1})$

by step 18 and the fact that the functions $\text{Log } ()$ and $\exp ()$ are inverses of each other.
20. $\sinh^{-1} x = \text{Log } (x + \sqrt{x^2 + 1})$ by steps 2 and 19.

As a result of this determination of a rule for the function $\sinh^{-1} ()$, we can find the derivative of this function.

Theorem 11.14:

$$\{D \sinh^{-1}\}(x) = \frac{1}{\sqrt{1 + x^2}}.$$

Proof:

1. $\sinh^{-1} ()$ is the function specified by

$$\sinh^{-1} x = \text{Log } (x + \sqrt{x^2 + 1}) \quad \text{by Theorem 11.13.}$$

2. $\{D \sinh^{-1}\}(x) = \frac{1}{x + \sqrt{x^2 + 1}} [\{D(I + \sqrt{I^2 + 1})\}(x)]$

by step 1 and the chain rule.

3. $\{D \sinh^{-1}\}(x) = \frac{1}{x + \sqrt{x^2 + 1}} [1 + \{D \sqrt{I^2 + 1}\}(x)]$

by step 2 and the theorem on the derivative of the sum.

4. $\{D \sinh^{-1}\}(x) = \frac{1}{x + \sqrt{x^2 + 1}} [1 + \frac{1}{2} (x^2 + 1)^{-1/2} (2x)]$

by step 3 and the chain rule.

5. $\{D \sinh^{-1}\}(x) = \frac{1}{x + \sqrt{x^2 + 1}} \left[1 + \frac{x}{\sqrt{x^2 + 1}}\right]$

by step 4, the definition of rational number exponents, and algebraic laws.

$$6. \{D \sinh^{-1}\}(x) = \frac{1}{x + \sqrt{x^2 + 1}} \frac{\sqrt{x^2 + 1} + x}{\sqrt{x^2 + 1}}$$

by step 5 and algebraic laws.

$$7. \{D \sinh^{-1}\}(x) = \frac{1}{\sqrt{x^2 + 1}}$$

by step 6 and algebraic laws.

In exactly the same manner, we can prove

Theorem 11.15:

If x is any number for which $-1 < x < 1$, then

$$\tanh^{-1} x = \frac{1}{2} [\text{Log}(1 + x) - \text{Log}(1 - x)].$$

Theorem 11.16:

$D \tanh^{-1}()$ is the function specified by

$$\{D \tanh^{-1}\}(x) = \frac{1}{1 - x^2}, \quad -1 < x < 1.$$

The proofs of these theorems are left as exercises.

Since the derivatives of these two inverse hyperbolic functions are algebraic functions, we have available two more tools for anti-differentiation. For example, if the derivative $f'()$ of the function $f()$ is given by

$$f'(x) = \frac{1}{4x - x^2},$$

then we have

$$\begin{aligned} f'(x) &= \frac{1}{4 - 4 + 4x - x^2} \\ &= \frac{1}{4 - (x^2 - 4x + 4)} \\ &= \frac{1}{4 - (x - 2)^2} \\ &= \frac{1}{4} \left(\frac{1}{1 - \frac{(x - 2)^2}{4}} \right) \\ &= \frac{1}{4} \left(\frac{1}{1 - \left(\frac{x - 2}{2}\right)^2} \right) \end{aligned}$$

If $g()$ is any function whatsoever, then

$$\{f(g)\}'(x) = \frac{1}{4} \left[\frac{1}{1 - \left(\frac{g(x) - 2}{2}\right)^2} \right] g'(x).$$

A most reasonable choice for $g()$ would then be the function given by

$$\frac{g(x) - 2}{2} = x, \quad -1 < x < 1$$

or, in other words, by

$$g(x) = 2x + 2, \quad -1 < x < 1.$$

For this function, we have

$$g'(x) = 2, \quad -1 < x < 1,$$

and then

$$\{f(g)\}'(x) = \frac{1}{4} \left(\frac{1}{1-x^2} \right) 2, \quad -1 < x < 1.$$

$$\{f(g)\}'(x) = \frac{1}{2} \left(\frac{1}{1-x^2} \right), \quad -1 < x < 1.$$

Then an anti-derivative of the composite, $\{f(g)\}()$ will be given by

$$\{f(g)\}(x) = \frac{1}{2} \tanh^{-1} x, \quad -1 < x < 1.$$

If $g()$ is the chosen function, we also have

$$f(g(x)) = \frac{1}{2} \tanh^{-1} \left(\frac{g(x) - 2}{2} \right), \quad -1 < x < 1,$$

which, upon the substitution of z for $g(x)$, gives

$$f(z) = \frac{1}{2} \tanh^{-1} \left(\frac{z - 2}{2} \right), \quad 0 < z < 4,$$

so that the function $f()$ is specified by

$$f(x) = \frac{1}{2} \tanh^{-1} \left(\frac{x - 2}{2} \right) + C, \quad 0 < x < 4.$$

where C is some constant.

Exercises

1. Prove Theorem 11.15.
2. Prove Theorem 11.16.

Find the derivatives of the function specified by:

- | | |
|----------------------------|--------------------------------------|
| 3. $\sinh^{-1}(r^3)$. | 8. $\text{Log}(\sinh^{-1} w)$. |
| 4. $\tanh^{-1}(\sinh s)$. | 9. $\exp(\tanh^{-1} x)$. |
| 5. $\sinh^{-1}(\cosh t)$. | 10. $\tan(\tanh^{-1} y)$. |
| 6. $\coth(\tanh^{-1} u)$. | 11. $\text{arc Tan}(\tanh^{-1} z)$. |
| 7. $\cosh(\sinh^{-1} v)$. | |

Using the facts,

$$1 + \sinh^2 r = \cosh^2 r,$$

$$\cosh^2 s - 1 = \sinh^2 s,$$

$$\sinh^2 t = \frac{1}{2} \cosh 2t - \frac{1}{2},$$

$$\cosh^2 u = \frac{1}{2} \cosh 2u + \frac{1}{2};$$

find anti-derivatives of the functions specified by:

12. $\sqrt{1 + r^2}.$

13. $\sqrt{s^2 - 1}.$

14. $(4t^2 + 4t + 4)^{1/2}.$

15. $(9u^2 + 6u)^{-1/2}.$

16. $(v^4 + 1)^{3/2} v.$

17. $(2 - \cos^2 w) \cos w.$

18. $x \sqrt{1 + x^2}.$

19. $(y^2 + 2y)^{-5/2}.$

20. $(z^2 + 3z + 3)^{3/2}.$



APPLICATIONS OF INTEGRATION

12.1. VOLUME

In the beginning of Chapter 7, we examined the idea of the area of a portion of the plane. One is taught (and, hopefully, learns) at a relatively early age that triangles, rectangles, and circles have area, and one is taught how to find that area. It was seen that for parts of the plane which are not simple it is necessary to make a formal definition of the word “area” itself, and that this definition uses the concept of the integral of a function, over an interval in the domain of that function. This new way of looking at area has two important properties: first, we can now find the area of parts of the plane which area could not be found before; second, this new way of finding area yields, for the simple parts of the plane already found, a result no different from that originally learned.

When it comes to the volume of three-dimensional objects, one is taught a number of “formulas” which give the volume of various kinds of such objects. For instance, the volume of a sphere is given by the equation

$$V = \frac{4}{3} \pi r^3,$$

in which r is the number of (linear) units which number gives the length of the radius of the sphere and V is the number of (cubic) units which gives the volume of the sphere.

One also learns that the volume of a rectangular parallelepiped is given by the equation

$$V = bwh,$$

where b is the number of (linear) units which gives the length of the base, w is the number of linear units which gives the width of the base, and h is the number of linear units which gives the height. V is then the number of cubic units which gives the volume. (A rectangular parallelepiped is more familiarly known as a box.) A further formula,

$$V = \pi r^2 h$$

gives the volume of a right circular cylinder, with r giving the radius of the cylinder, h the height, and V the volume.

These, and other formulas are learned as elementary facts of life about the "real world," since we think that we live in a world containing three-dimensional solid objects. Physicists, however, inform us that a solid three-dimensional object such as a right circular cylinder exists only in our imagination, saying that matter is actually composed of things called molecules, which in turn are composed of things called atoms, and so on. It would appear then that such things as spheres, cylinders, and rectangular parallelepipeds are merely figments of our imagination.

But every mathematical entity is just such an imagined thing, so that to talk about volumes of three-dimensional objects, it is not necessary that we concern ourselves with whether or not such things actually exist in any "real" sense, but merely to make sure that our discussion does not become unreasonable and self-contradictory.

To give a general definition of the volume of a three-dimensional object requires the use of the idea of a "function of two variables," which idea would take us beyond the scope of this book. We will therefore restrict ourselves to but one "instinctive" kind of volume—that of a rectangular box—and try to make the numbers we call volumes be not unreasonable ones.

We take, then, as known, that the volume of a rectangular box is the product of the length, width, and height.

12.2. RIGHT CIRCULAR CYLINDERS

In this section, we will derive the formula

$$V = r^2h$$

for the volume of a right circular cylinder. In this formula, if r is the number which gives the length of the radius of the cylinder, and h is the number which gives the height of the cylinder, then V will be the number which gives the volume of the cylinder.

To be exact, a right circular cylinder is a three-dimensional object obtained as follows: We start with a line segment of a certain length—say h units of length. This line segment is called the *axis* of the cylinder and the length of this axis is called the *height* of the cylinder. We start also with a fixed positive number—say r —which is to be called the *radius* of the cylinder. At each point of the axis of the cylinder we pass a plane perpendicular to this axis. In each of these planes we construct a circle of radius r . The right circular cylinder will then consist of all points which are inside or on all of the circles thus constructed. The circles whose planes pass through the end points of the axis are then called the top and bottom of the cylinder.

To find the volume of such a right circular cylinder, we will consider the quarter-cylinder formed by slicing the full cylinder with two mutually perpendicular planes passing through the axis (Fig. 12.1).

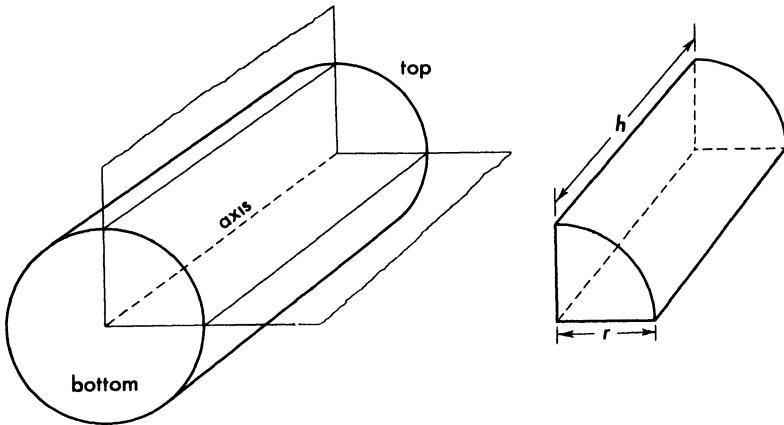


Fig. 12.1

We will attempt to find reasonable approximations to the volume of this quarter-cylinder in the same way in which we found approximations to the almost-triangle encountered in the beginning of Chapter 7. To this end, we will fix our attention on the straight line which appears as the horizontal radius of the quarter-circle which is the end of the quarter-cylinder. This line segment has length r , and, for definiteness, we will call this line segment our "base line."

If we have a new line segment, lying in the plane of this quarter-circle, which line segment is inside the quarter-circle, perpendicular to our base line, with end points on the base line and on the circumference of the quarter-circle, and which new line is located t units away from the center of the quarter-circle (Fig. 12.2), then the Pythagorean Theorem of elementary geometry tells us that the length of this new line segment must be

$$\sqrt{r^2 - t^2}.$$

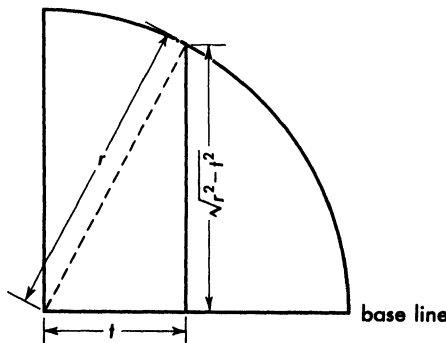


Fig. 12.2

If n is a natural number, and we divide our base line into n segments of equal length, then each segment will have length r/n . We call the two ends of our base line P_0 (for the center of the quarter-circle) and P_n , and call the $(n - 1)$ points of division

$$P_1, P_2, P_3, \dots, P_{n-2}, P_{n-1}.$$

Each of these $(n + 1)$ points on the base line is located at a certain distance from the center of the quarter-circle.

P_0 is 0 units distant from the center.

P_1 is $\frac{r}{n}$ units distant from the center.

P_2 is $2\frac{r}{n}$ units distant from the center.

P_3 is $3\frac{r}{n}$ units distant from the center.

$\vdots$

P_{n-2} is $(n - 2)\frac{r}{n}$ units distant from the center.

P_{n-1} is $(n - 1)\frac{r}{n}$ units distant from the center.

P_n is $n\frac{r}{n} (= r)$ units distant from the center.

At each of the $(n + 1)$ points we have on the base line, we now construct a line segment, perpendicular to the base line, with end points on the base line and on the circumference of the quarter-circle (Fig. 12.3).

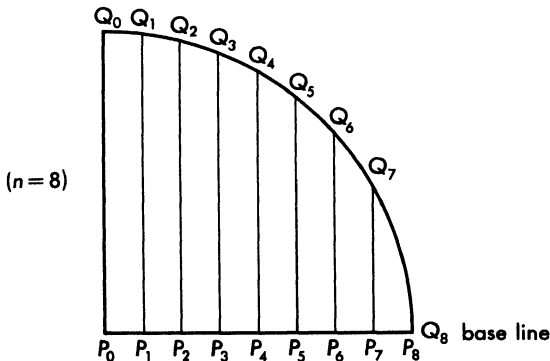


Fig. 12.3

We call the end points of these new line segments

$$Q_0, Q_1, Q_2, Q_3, \dots, Q_{n-2}, Q_{n-1}, \text{ and } Q_n,$$

respectively, noting that the line segment P_0Q_0 is a radius of the quarter-circle (and thus is r units long) and that the line segment P_nQ_n is 0 units long.

The lengths of these line segments can now be found:

$$P_0Q_0 \quad \text{is } \sqrt{r^2 - 0^2} \text{ units long.}$$

$$P_1Q_1 \quad \text{is } \sqrt{r^2 - \left(\frac{r}{n}\right)^2} \text{ units long.}$$

$$P_2Q_2 \quad \text{is } \sqrt{r^2 - \left(2\frac{r}{n}\right)^2} \text{ units long.}$$

$$P_3Q_3 \quad \text{is } \sqrt{r^2 - \left(3\frac{r}{n}\right)^2} \text{ units long.}$$

⋮

$$P_{n-2}Q_{n-2} \text{ is } \sqrt{r^2 - \left([n-2]\frac{r}{n}\right)^2} \text{ units long.}$$

$$P_{n-1}Q_{n-1} \text{ is } \sqrt{r^2 - \left([n-1]\frac{r}{n}\right)^2} \text{ units long.}$$

$$P_nQ_n \quad \text{is } \sqrt{r^2 - \left(n\frac{r}{n}\right)^2} (= 0) \text{ units long.}$$

We now extend the line segments $P_1Q_1, P_2Q_2, P_3Q_3, \dots, P_{n-2}Q_{n-2}, P_{n-1}Q_{n-1}$ slightly beyond the circumference of our quarter-circle, and draw a line segment perpendicular to the base line through P_n . Horizontal line segments of length r/n drawn from the points $Q_0, Q_1, Q_2, Q_3, \dots, Q_{n-2}$, and Q_{n-1} will end at these newly drawn vertical lines (Fig. 12.4).

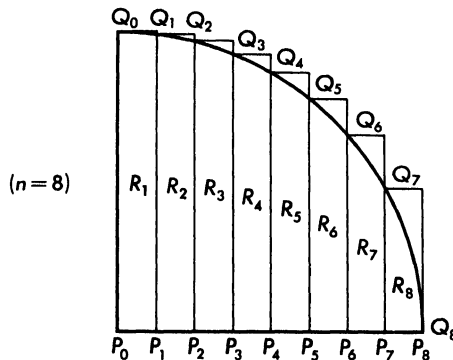


Fig. 12.4

In this manner we form n (thin) rectangles, each one of which has its base of length r/n on the base line, and each one of which has an area greater than the area of that piece of the quarter-circle this rectangle covers.

In fact, the heights of these rectangles are precisely the lengths of the line segments $P_0Q_0, P_1Q_1, P_2Q_2, P_3Q_3, \dots, P_{n-2}Q_{n-2}$, and $P_{n-1}Q_{n-1}$. If we call these rectangles $\overline{R}_1, \overline{R}_2, \overline{R}_3, \dots, \overline{R}_{n-1}$, and $\overline{R}_n$, respectively, then:

$\overline{R}_1$ has base $\frac{r}{n}$ and height $r\sqrt{1-0^2}$.

$\overline{R}_2$ has base $\frac{r}{n}$ and height $r\sqrt{1-\left(\frac{1}{n}\right)^2}$.

$\overline{R}_3$ has base $\frac{r}{n}$ and height $r\sqrt{1-\left(\frac{2}{n}\right)^2}$.

$\vdots$

$\overline{R}_{n-1}$ has base $\frac{r}{n}$ and height $r\sqrt{1-\left(\frac{n-2}{n}\right)^2}$.

$\overline{R}_n$ has base $\frac{r}{n}$ and height $r\sqrt{1-\left(\frac{n-1}{n}\right)^2}$.

We now recall that this quarter-circle is one end of our quarter-cylinder. If we construct (thin) boxes using these n rectangles as ends, each of which boxes has depth h , then each box will contain within itself a small slab of the quarter-cylinder, and the sum of the volumes of the boxes will be a number (slightly) larger than the volume the quarter-cylinder (Fig. 12.5).

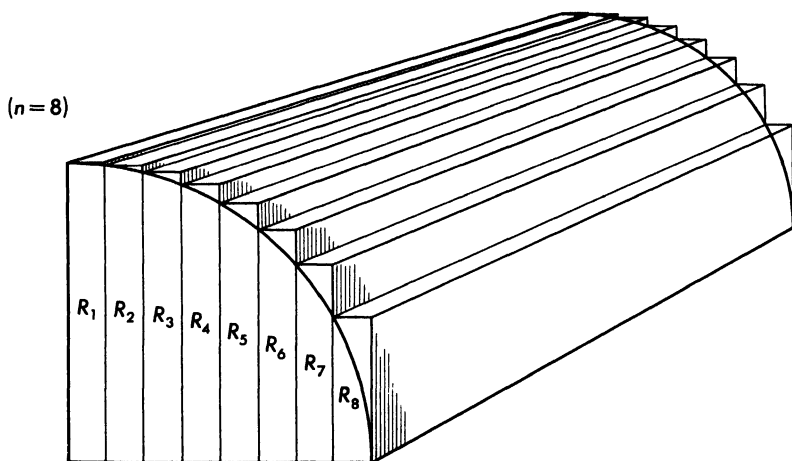


Fig. 12.5

We then have:

The box with base $\overline{R}_1$ has volume $r^2h\sqrt{1-(0)^2}\frac{1}{n}$.

The box with base $\overline{R_2}$ has volume $r^2h \sqrt{1 - \left(\frac{1}{n}\right)^2} \frac{1}{n}$.

The box with base $\overline{R_3}$ has volume $r^2h \sqrt{1 - \left(\frac{2}{n}\right)^2} \frac{1}{n}$.

⋮

The box with base $\overline{R_{n-1}}$ has volume $r^2h \sqrt{1 - \left(\frac{n-2}{n}\right)^2} \frac{1}{n}$.

The box with base $\overline{R_n}$ has volume $r^2h \sqrt{1 - \left(\frac{n-1}{n}\right)^2} \frac{1}{n}$.

If we call the sum of the volumes of the n boxes $\overline{V_n}$, we have

$$\begin{aligned} \overline{V_n} = r^2h \left[\sqrt{1 - (0)^2} + \sqrt{1 - \left(\frac{1}{n}\right)^2} + \sqrt{1 - \left(\frac{2}{n}\right)^2} + \dots \right. \\ \left. + \sqrt{1 - \left(\frac{n-2}{n}\right)^2} + \sqrt{1 - \left(\frac{n-1}{n}\right)^2} \right] \frac{1}{n}. \end{aligned}$$

This sum is exactly of the form used in the definition of the upper integral

$$\int_0^1 f() d()$$

for the function specified by

$$f(t) = r^2h \sqrt{1 - t^2}$$

over the interval:

the set of all t , $0 \leq t \leq 1$.

Since this function is continuous at every number in this interval, the integral exists, so that the upper and lower integrals also exist, and are the same number. If we were to form the sum appearing in the definition of the lower integral for this function, over this interval, we would find that we would have the sum of volumes of boxes which were obtained by drawing line segments left-ward from the points $Q_1, Q_2, Q_3, \dots, Q_{n-1}, Q_n$, and use the resulting thin rectangles as ends of the (thin) boxes. In such a case, the resulting (thin) boxes would all lie inside of our quarter-cylinder, and the sum of their volumes would be a number (slightly) less than the volume of the quarter-cylinder.

Consequently, since the numbers which form the sequence whose limit is the upper integral are all (slightly) larger than the volume of the quarter-cylinder, since the numbers which form the sequence whose limit is the lower integral are all (slightly) smaller than the volume of the quarter-

cylinder, and since the upper and lower integrals are the same number, regardless of the formal definition of volume for three-dimensional objects, the volume of this quarter-cylinder must be given by the number

$$\int_0^1 [r^2 h \sqrt{1-t^2}] dt.$$

To evaluate this integral, we have:

$$\int_0^1 [r^2 h \sqrt{1-t^2}] dt = r^2 h \int_0^1 \sqrt{1-t^2} dt.$$

If $f'(t) = \sqrt{1-t^2}$, then for any function $g()$,

$$\{f(g)\}'(t) = \sqrt{1-[g(t)]^2} g'(t).$$

Letting $g()$ be specified by

$$g(t) = \sin t,$$

we have

$$\begin{aligned}\{f(g)\}'(t) &= \sqrt{1-[\sin t]^2} \cos t \\ &= \cos^2 t \\ &= \frac{1}{2} + \frac{1}{2} \cos 2t.\end{aligned}$$

Thus, ignoring the possible addition of a constant,

$$\begin{aligned}\{f(g)\}(t) &= \frac{1}{2} t + \frac{1}{4} \sin 2t \\ &= \frac{1}{2} t + \frac{1}{2} \sin t \cos t \\ f(\sin t) &= \frac{1}{2} t + \frac{1}{2} \sin t \cos t \\ &= \frac{1}{2} \arcsin(\sin t) + \frac{1}{2} \sin t \sqrt{1-[\sin t]^2} \\ f(z) &= \frac{1}{2} \arcsin z + \frac{1}{2} z \sqrt{1-z^2}, \quad 0 \leq z \leq 1.\end{aligned}$$

We then have, by the Fundamental Theorem,

$$\begin{aligned}r^2 h \int_0^1 \sqrt{1-t^2} dt &= r^2 h \left[\frac{1}{2} \arcsin z + \frac{1}{2} z \sqrt{1-z^2} \right]_{z=0}^{z=1} \\ &= r^2 h \left(\frac{\pi}{4} \right).\end{aligned}$$

Consequently, the volume of the entire cylinder, V , must be given by

$$V = \pi r^2 h.$$

Exercises

1. Show how underestimates of the volume of the quarter-cylinder obtained by drawing line segments of length r/n to the left of the points $Q_1, Q_2, \dots, Q_n$ yield numbers in the sequence whose limit is

$$r^2 h \int_0^1 \sqrt{1-t^2} dt.$$

2. Show that if $f()$ is specified by

$$f(x) = \frac{1}{2} \arcsin x + \frac{1}{2} x \sqrt{1-x^2}, \quad 0 \leq x \leq 1,$$

then $f'(x) = \sqrt{1-x^2}$, $0 \leq x \leq 1$.

3. Given a part of the plane for which the area is defined, show that the volume of the cylinder constructed by raising this part of the plane, perpendicular to the plane, a distance of h units, can be reasonably defined to be the product

$$h (\text{area of base}).$$

12.3. SPHERES

In this section, we derive the formula

$$V = \frac{4}{3} \pi r^3,$$

which gives the volume V of a sphere of radius r . It will be assumed that the reader knows what a sphere is.

We will content ourselves with finding the volume of a hemisphere. If we start with a quarter-circle of radius r , consider the straight line, a segment of which forms one side of this quarter-circle, and revolve the quarter-circle (out of its own plane) around this straight line, then the quarter-circle will sweep out a piece of our three-dimensional space which piece is a hemisphere (Fig. 12.6).

We concentrate our attention, as before, on this quarter-circle. We call the vertical edge of the quarter-circle our "base line," divide this base line into n segments of equal length (where n is a natural number), draw line segments perpendicular to the base line from each point of division, and use these lines to form (thin) horizontal rectangles, each one of which extends just outside of the quarter-circle (Fig. 12.7).

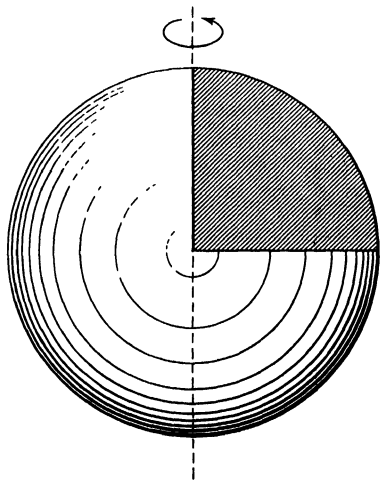


Fig. 12.6

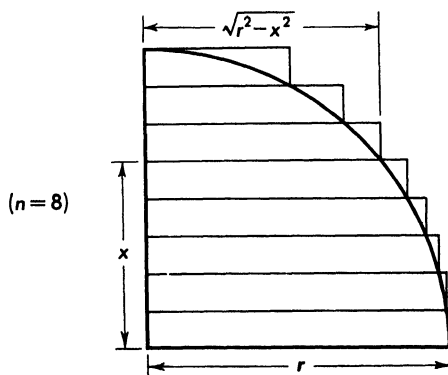


Fig. 12.7

That rectangle whose top is x units from the center of the quarter-circle will then have length

$$\sqrt{r^2 - x^2}.$$

Here, we have chosen x as a variable whose range is the interval:

$$\text{the set of all } x, \quad 0 \leq x \leq r.$$

If we revolve all these rectangles about the base line, then we will form n right circular cylinders, the volume of each of which is (slightly) more than the volume of the part of the hemisphere covered by that (thin) cylinder. To examine a typical one of these (thin) cylinders, we take k to be a natural number between 1 and n , so that the k -th cylinder away from the center will have radius

$$\sqrt{r^2 - \left(\frac{k}{n}r\right)^2}$$

and height $\frac{r}{n}$. The volume of this k -th cylinder then must be

$$\pi \left[\sqrt{r^2 - \left(\frac{k}{n}r\right)^2} \right]^2 \frac{r}{n},$$

which is to say

$$\pi \left[r^2 - \left(k \frac{r}{n}\right)^2 \right] \frac{r}{n}.$$

This expression is precisely the k -th term in the sum which forms one of the numbers in the sequence whose limit is the upper integral of the function specified by

$$f(x) = \pi[r^2 - x^2]$$

over the interval:

$$\text{the set of all } x, \quad 0 \leq x \leq r.$$

Since this (polynomial) function is continuous everywhere, the upper and lower integrals are the same, and the volume of our hemisphere must be the number

$$\begin{aligned} & \int_0^r \pi[r^2 - x^2] dx \\ &= \pi \left[r^2 x - \frac{x^3}{3} \right]_{x=0}^{x=r} \\ &= \frac{2}{3} \pi r^3. \end{aligned}$$

The volume of a sphere of radius r must then be

$$\frac{4}{3} \pi r^3.$$

Exercises

1. Find the volume of a right circular cone whose base has radius r and whose height is h .
2. The almost-triangle bounded by the graphs of $y = x^2$, $y = 0$ and $x = 1$ is revolved about the x -axis. Find the volume of the solid thus formed.
3. The part of the plane bounded by the graphs of $y = 8 - x^2$ and $y = 4$ is revolved about the line with equation $y = 4$. Find the volume of the solid thus formed.

Find the volume of the solid formed by revolving the portion of the plane bounded by the graphs of the given equation about the indicated straight line:

4. $y = x^2$, $y = 0$, $x = 1$; about the x -axis.
5. $y = x^2$, $x = 0$, $y = 1$; about the y -axis.
6. $y = \sqrt{4 - x}$, $x = 0$, $y = 0$; about the x -axis.
7. $y = \sqrt{4 - x}$, $x = 0$, $y = 0$; about the y -axis.
8. $y = 1/x$, $y = 0$, $x = 1$, $x = 2$; about the x -axis.
9. $y = \sin x$, $y = 0$, $x = \pi/2$; about the x -axis.
10. $y = e^x$, $x = 0$, $y = e$; about $y = e$.
11. $y = e^x$, $x = 0$, $y = e$; about the y -axis.

12.4. VOLUME OF A SOLID OF REVOLUTION

Spheres, right circular cylinders, and right circular cones are examples of three-dimensional objects which result from revolving a particular kind of portion of a plane (semicircles, rectangles, triangles) about a straight line which forms a side of that portion of the plane. We have noted that slicing the generating portion of the plane into thin almost-rectangles is equivalent to slicing the three-dimensional object into thin almost-cylinders, provided that, in the plane, the almost-rectangles are formed by partitioning a certain "base line" into segments, and then constructing lines through the points of division, perpendicular to this base line. A reasonable name for this base line is "the axis of revolution" of the three-dimensional object.

In this section, we will examine the general case, in which the axis of revolution of a three-dimensional solid of revolution need not form a side of the generating portion of the plane. In order to aid us in establishing general methods of procedure, we will re-examine the method of finding volume in the simple case that the axis of revolution actually does form a side of the generating plane object.

We consider a function $f()$, and an interval of numbers:

$$\text{the set of all } t, \quad a \leq t \leq b,$$

lying entirely within the domain of $f()$. We assume that the part of the range of values of $f()$ corresponding to this interval contains no negative numbers. This assumption means that the part of the graph of $f()$ between the points with coordinates $(a, f(a))$ and $(b, f(b))$ is never below the horizontal axis (Fig. 12.8).

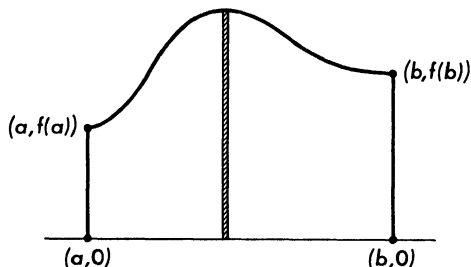


Fig. 12.8

We now proceed to find the volume of the solid of revolution obtained by revolving that portion of the plane bounded by the part of the graph of $f()$ between $(a, f(a))$ and $(b, f(b))$, the vertical line segments joining $(a, 0)$ to $(a, f(a))$, and $(b, 0)$ to $(b, f(b))$, and the horizontal line segment joining $(a, 0)$ to $(b, 0)$, about the x -axis. The x -axis is then the axis of revolution of this solid, and the line segment joining $(a, 0)$ to $(b, 0)$ is the base line for our analysis.

A vertical line passing through the given portion of the plane will be cut at the top by the graph of $f()$ and at the bottom by the x -axis. If this vertical line has equation $x = t$, then these points of intersection will have coordinates $(t, f(t))$ and $(t, 0)$ and the length of the line segment thus cut off will be $f(t)$ units of length.

We now divide the base line into n pieces of equal length (where n is a natural number). The length of each piece is then $\frac{b-a}{n}$.

If k is a natural number between 1 and n , then the k -th point of division of the base line will have coordinates

$$\left(a + k \left(\frac{b-a}{n}\right), 0\right)$$

and the $(k+1)$ -th point of division will have coordinates

$$\left(a + (k+1) \left(\frac{b-a}{n}\right), 0\right).$$

Vertical line segments through these two points of division will then form the sides of a typical almost-rectangle into which the generating portion of the plane is sliced. The left-hand edge of this almost-rectangle ends at the points with coordinates

$$\left(a + k \left(\frac{b-a}{n}\right), 0\right)$$

and

$$\left(a + k \left(\frac{b-a}{n}\right), f\left(a + k \left(\frac{b-a}{n}\right)\right)\right),$$

and the right-hand edge of this almost-rectangle ends at the points with coordinates

$$\left(a + (k+1) \left(\frac{b-a}{n}\right), 0\right)$$

and

$$\left(a + (k+1) \left(\frac{b-a}{n}\right), f\left(a + (k+1) \left(\frac{b-a}{n}\right)\right)\right).$$

(See Fig. 12.9.)

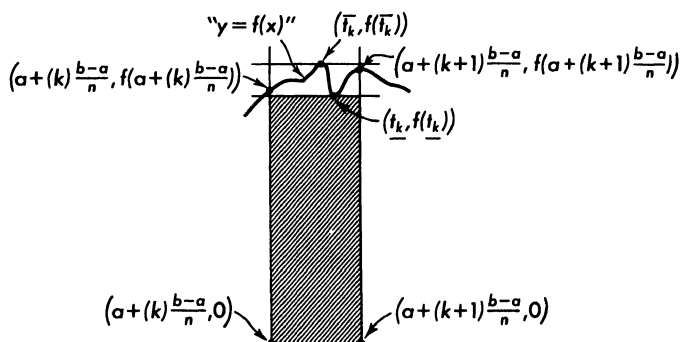


Fig. 12.9

If the function $f()$ is continuous, then there will be a number, $\bar{t}_k$, in the subinterval:

$$\text{the set of all } t, \quad a + k \left(\frac{b-a}{n}\right) \leq t \leq a + (k+1) \left(\frac{b-a}{n}\right)$$

for which $f()$ has an absolute maximum value for this subinterval, and another number, t_k , in this subinterval, for which $f()$ has an absolute minimum value for this subinterval.

The rectangle formed by the lines with equations $y = 0$,

$$x = a + k \left(\frac{b-a}{n}\right), \quad x = a + (k+1) \left(\frac{b-a}{n}\right), \quad \text{and} \quad y = f(\bar{t}_k)$$

is then larger than this k -th almost-rectangle, and the rectangle formed by the lines with equations $y = 0$,

$$x = a + k \left(\frac{b-a}{n}\right), \quad x = a + (k+1) \left(\frac{b-a}{n}\right), \quad \text{and} \quad y = f(t_k)$$

is then smaller than this k -th almost rectangle.

Revolving the two rectangles and the one almost-rectangle about the x -axis yields two right circular cylinders and one almost-cylinder, with the volume of the almost-cylinder a number between the volumes of the actual cylinders. Thus, we have

$$\pi [f(\underline{t}_k)]^2 \left(\frac{b-a}{n} \right) \leq \text{Volume of the } k\text{-th almost-cylinder} \leq \pi [f(\bar{t}_k)]^2 \left(\frac{b-a}{n} \right).$$

This set of inequalities is true for $k = 1, 2, \dots, n$, so that if we add the numbers in the respective parts, we have

$$\begin{aligned} & \pi ([f(\underline{t}_1)]^2 + [f(\underline{t}_2)]^2 + \dots + [f(\underline{t}_n)]^2) \left(\frac{b-a}{n} \right) \\ & \leq \text{Volume of the solid of revolution} \\ & \leq \pi ([f(\bar{t}_1)]^2 + [f(\bar{t}_2)]^2 + \dots + [f(\bar{t}_n)]^2) \left(\frac{b-a}{n} \right). \end{aligned}$$

The number on the left-hand side of this last inequality is one of the numbers in the sequence appearing in the definition of the lower integral of the function $\{ \pi f^2 \}()$ over the interval:

$$\text{the set of all } t, \quad a \leq t \leq b,$$

and the number on the right-hand side of this last inequality is one of the numbers in the sequence appearing in the definition of the upper integral of the same function over the same interval. But, if $f()$ is a continuous function, then so is $\{ \pi f^2 \}()$, so that

$$\overline{\int_a^b \{ \pi f^2 \}() d()} = \underline{\int_a^b \{ \pi f^2 \}() d()} = \int_a^b \{ \pi f^2 \}() d(),$$

and we conclude that the volume of the solid of revolution generated by revolving about the x -axis that portion of the plane bounded by the part of the graph of $f()$ between $(a, f(a))$ and $(b, f(b))$, the vertical lines with equations $x = a$ and $x = b$, and the x -axis is the number

$$\pi \int_a^b [f(t)]^2 dt.$$

We now consider the solid of revolution obtained by revolving about the x -axis a portion of the plane which does not have a piece of the axis as a side. Let $f()$ and $g()$ be two functions, both continuous, and let the interval:

$$\text{the set of all } t, \quad a \leq t \leq b$$

be in the domain of both functions. Further, we assume that, if $a \leq t \leq b$, then $f(t) \leq g(t)$. Thus, the part of the graph of $f()$ between $(a, f(a))$ and $(b, f(b))$ is not above the part of the graph of $g()$ between $(a, f(a))$ and $(b, f(b))$. Again, taking the part of the x -axis between $(a, 0)$ and $(b, 0)$ as the base line, we take two successive points of division, draw lines perpendi-

cular to the base line through these points of division, and examine the almost-rectangles thus formed in this slicing of the portion of the plane to be revolved: the portion of the plane bounded by the part of the graph of $f()$ between $(a, f(a))$ and $(b, f(b))$, the part of the graph of $g()$ between $(a, g(a))$ and $(b, g(b))$, and the vertical lines with equations $x = a$ and $x = b$ (Fig. 12.10).

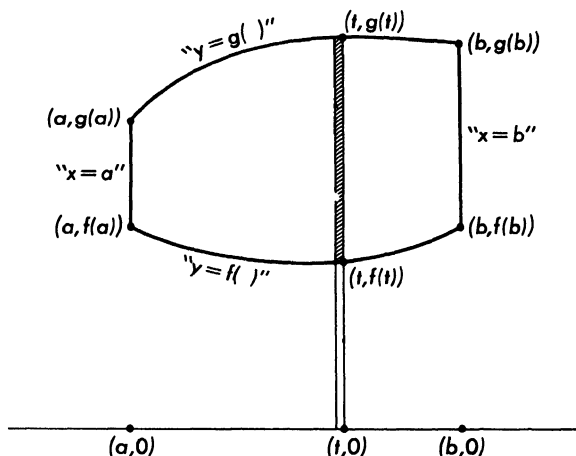


Fig. 12.10

If we revolve a typical almost-rectangle about the x -axis, we will have a washer-shaped disk, whose inside radius is some number in the range of values of $f()$ and whose outside radius is some number in the range of values of $g()$. But the volume of a washer-shaped disk is simply the difference between the volume of the cylinder this disk would be if it had no hole and the volume of the cylinder of empty space which constitutes the hole. Thus, the volume of a typical slice of the solid of revolution will be given by

$$\pi [g(t)]^2 \left(\frac{b-a}{n} \right) - \pi [f(t)]^2 \left(\frac{b-a}{n} \right) = \pi [(g(t))^2 - (f(t))^2] \left(\frac{b-a}{n} \right),$$

where n is the number of subdivisions of the base line and t has values giving maximum and/or minimum values to the expression $\{g(t) - f(t)\}$ in this typical slice. Whether we take overestimates or underestimates of the washer-shaped slices of the solid of revolution, increasing the number of subdivisions gives numbers closer to the upper and lower integrals of the function

$$\pi \{g^2 - f^2\}()$$

over the interval:

$$\text{the set of all } t, \quad a \leq t \leq b.$$

If $f()$ and $g()$ are both continuous, then so is $\{g^2 - f^2\}()$, and the upper and lower integrals are the same number. We then conclude that

the volume of the solid of revolution obtained by revolving about the x -axis that portion of the plane bounded by the graphs of the equations $y = f(x)$, $y = g(x)$, $x = a$, and $x = b$ is the number

$$\pi \int_a^b ([g(t)]^2 - [f(t)]^2) dt.$$

Finally, if we were to revolve this same portion of the plane, not about the x -axis, but about a line with equation $y = c$, which line is below the generating portion of the plane, then the inside and outside radii of the washer-shaped typical slice would be

$$f(t) - c \quad \text{and} \quad g(t) - c,$$

respectively, and the volume of the solid of revolution so generated would be the number

$$\pi \int_a^b ([g(t) - c]^2 - [f(t) - c]^2) dt.$$

Exercises

Find the volume of the solid generated by revolving that part of the plane bounded by the graphs of the given equations about the given line:

1. $y = x^2$, $y = 0$, $x = 1$; about $x = 2$.
2. $y = x^2$, $y = 0$, $x = 1$; about $x = -1$.
3. $y = x^2$, $y = 0$, $x = 1$; about $y = 1$.
4. $y = x^2$, $y = 0$, $x = 1$; about $y = 2$.
5. $y^2 + x = 4$; $x = 0$; about $x = -1$.
6. $y^2 + x = 4$; $x = 0$; about $y = 2$.
7. $y = \sin x$, $y = 0$, $x = \pi/2$; about $y = 1$.
8. $y = e^x$, $x = 0$, $y = e$; about the x -axis.
9. $y = e^x$, $x = 0$, $y = e$; about $y = 1$.
10. $y = 25 - x^2$; $2x - y + 10 = 0$; about the x -axis.

12.5. SOLIDS WITH REGULAR CROSS-SECTIONS

In a solid which is the result of revolving a given portion of the plane about a line which is either a side of that portion of the plane or lies entirely outside that portion of the plane, we found volume to be the number which was the value of a certain integral of a certain function over a certain interval of numbers.

The interval of numbers, over which the integration was performed, was determined by examining the length of that part of the axis of revolution which we called the base line. The function to be integrated was found by examining the sums which were approximations to the volume, which

sums were the sums of volumes of cylindrical or washer-shaped slabs. The volumes of such slices of the solid were found by using either of the two formulas,

$$\pi r^2 h \quad \text{or} \quad \pi(r_o^2 - r_i^2) h,$$

for the volume of a cylinder, or of a cylinder with a hole in it. In these formulas the variable h gave the thickness of the slice of our solid. This thickness turned out to be the length of the part of the base line we were examining, which in turn was the length of one of the sub-intervals of the interval of integration of the integral we found. Then the factors, πr^2 or $\pi(r_o^2 - r_i^2)$, in these formulas must also have some geometric significance, and, in fact, they are expressions which give us the area of a cross-section of our solid of revolution. But these factors which multiply the thickness of the slabs also give us the function which is to be integrated.

For example, the sphere which results in revolving about the x -axis the portion of the plane bounded by the x -axis and the graph of

$$y = \sqrt{9 - x^2},$$

has volume

$$\pi \int_{-3}^3 [\sqrt{9 - t^2}]^2 dt = 36\pi.$$

The function being integrated is that given by the equation

$$A(t) = 9 - t^2, \quad -3 \leq t \leq 3.$$

If we have this sphere, and do slice through the sphere by a plane perpendicular to the axis of revolution (the x -axis) through the point with coordinates $(t, 0)$, then the cross-section of this sphere will be a circle whose radius is

$$\sqrt{9 - t^2}.$$

The area of the circular cross-section thus found will be

$$\pi(9 - t^2).$$

Further, in order actually to cut the sphere (whose radius is 3 units), the variable t must have, as its range, the interval:

$$\text{the set of all } t, \quad -3 \leq t \leq 3.$$

But then the function $A()$ is precisely that function which has the property that, for any given value of t , the cross-section of the solid of revolution found by slicing this solid by a plane perpendicular to the axis of revolution, through the point on that axis of revolution given by that value of t , will have area $A(t)$.

In order, then, to find the volume of any three-dimensional object which has the property that the area of any cross-section taken perpendicular to some base line can be found by evaluating some function, we need merely integrate that function over an interval which interval is determined by the base line we use.

For example, suppose we consider the following three-dimensional object: The portion of the plane bounded by the graphs of $y = x^2$ and $y = x^{1/2}$ is the base of this solid. Every cross-section taken perpendicular to the x -axis is a square (Fig. 12.11).

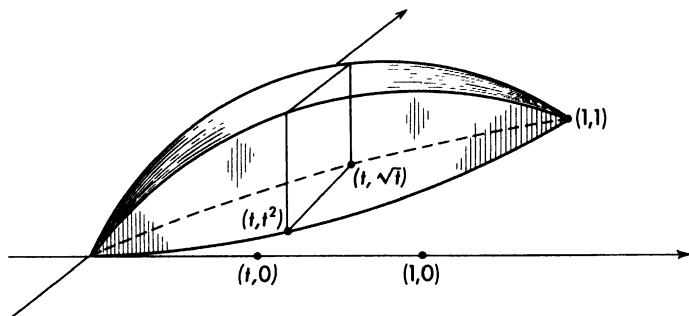


Fig. 12.11

The cross-section formed by taking a plane through the point $(t, 0)$ (where $0 \leq t \leq 1$), perpendicular to the x -axis will then be a square whose base is the line segment joining (t, t^2) to $(t, t^{1/2})$. This line segment has length $(t^{1/2} - t^2)$. The square cross-section then has area $(t^{1/2} - t^2)^2$. Thus, the area of a cross-section of this wedge-shaped object with a flat bottom and curved top and sides will be given by the function specified by

$$A(t) = t - 2t^{5/2} + t^4, \quad 0 \leq t \leq 1,$$

where the interval:

$$\text{the set of all } t, \quad 0 \leq t \leq 1$$

determines the x -coordinates of points on the base line which is the line segment joining the origin to the point $(1, 0)$. The volume of this wedge-shaped object then must be

$$\begin{aligned} & \int_0^1 A(t) \, dt \\ &= \int_0^1 [t - 2t^{5/2} + t^4] \, dt \\ &= \left[\frac{t^2}{2} - \frac{4}{7} t^{7/2} + \frac{t^5}{5} \right]_{t=0}^{t=1} \\ &= \frac{1}{2} - \frac{4}{7} + \frac{1}{5} \\ &= \frac{9}{70}. \end{aligned}$$

Exercises

1. The part of the plane that is inside the graph of $9x^2 + 16y^2 = 144$ is the base of a solid. Every cross-section of this solid taken perpendicular to the x -axis is a square. Find the volume of this solid.

2. The base of a solid is that part of the plane inside the triangle formed by the points $(1, 1)$, $(4, 5)$, and $(7, 1)$. Cross-sections perpendicular to the y -axis are semicircles. Find the volume of this solid.
3. Find the volume of a pyramid with square base whose height is h feet and whose base measures b feet on each side.
4. The base of a solid is a circle of radius r . If one of the diameters of the circle is fixed, all cross-sections perpendicular to this diameter will be isosceles right triangles with their hypotenuses in the plane of the given circle. Find the volume of this solid.
5. The solid in exercise 1 has cross-sections which are equilateral triangles rather than squares. Find the volume.
6. A cylinder has height h and radius of the base r . One diameter is drawn on the base, with ends labeled P and Q . A radius of the base, perpendicular to the line PQ , is drawn, and the end of this radius is called R . The point on the top of the cylinder directly above R is called S . The plane PQS cuts the cylinder into two pieces. Find the volume of the smaller piece.
7. Inside the cylinder of exercise 6, a hemisphere is constructed whose base is the base of the cylinder. Find the volume of the wedge bounded by the plane PQS , the hemisphere, and the cylinder.

12.6. SHELL METHOD FOR VOLUMES OF REVOLUTION

As we have seen, the volume of a solid of revolution, which results from revolving about some straight line any portion of the plane whose area can be found, is found by evaluating the integral of a particular function over a particular interval. It may well happen that the function to be integrated is one for which an anti-derivative cannot be found readily, or cannot be found at all. If this occurs, then we have two choices. We can use the definition of integral in order to find an approximation to the value of the integral or we can look for a completely different way of finding an integral which gives us the volume.

If we are content with an approximation to the volume of a solid of revolution—to within some desired degree of accuracy—and have available rapid computing facilities, then the first choice can be made. If we are not content with an approximation, or if the computations involved in getting one are too onerous, then we must make the second choice and seek some new way of approaching the volume of a solid of revolution.

The method of using cylindrical or washer-shaped disks to find an integral whose value is the desired volume considers the given solid of revolution to be approximately constructed from many of these thin disks. The examination of the form of the sum of the expressions giving the volumes of these thin disks then suggested a function whose integral, over an appropriate interval, was reasonably called the volume of the solid.

In this section, we will see what happens if we imagine the solid to be approximately constructed from concentric thin cylindrical shells. For the

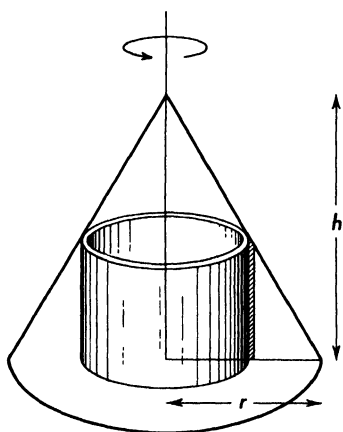


Fig. 12.12

purposes of illustration, we consider a right circular cone, of height h , whose base has radius r . Such a cone will result from revolving a right triangle whose sides have lengths r and h about the line forming the side of length h (Fig. 12.12). The result we arrive at is correct, but we are not actually giving a proof, merely a justification.

If we use the side of this right triangle whose length is r as our base line, then thin rectangles formed by drawing lines perpendicular to the points of division of the base line will form thin cylindrical shells when revolved about the axis of revolution.

Since the base line has length r , we let the variable t have as range the interval:

$$\text{the set of all } t, \quad 0 \leq t \leq r,$$

and we interpret a value of t geometrically as fixing the point on the base line which is t units from the right angle.

A line in the plane of the triangle, perpendicular to the base line, at the point on the base line t units from the right angle, will form a smaller right triangle which is similar to the original triangle. If the length of this line segment is x units, then the similarity of the triangles tells us that

$$\frac{x}{r-t} = \frac{h}{b},$$

or, in other words that

$$x = \frac{h}{b}(r-t).$$

A typical one of these cylindrical shells will have volume approximately that of the portion of the cone which it covers, with the approximation improving as we increase the number of subdivisions of the base line. If we take two successive points of division—say the k -th and $(k+1)$ -th points—the cylindrical shell they produce will have inside radius of

$$k \frac{r}{n},$$

outside radius of

$$(k+1) \frac{r}{n}$$

(where n is the number of subdivisions),

and height of either

$$\frac{h}{b} \left(r - k \frac{r}{n} \right) \quad \text{or} \quad \frac{h}{b} \left(r - [k + 1] \frac{r}{n} \right),$$

depending on whether we are taking an overestimate or an underestimate.

Trusting that the function we finally find will be continuous, so that upper and lower integrals will be the same, we will choose an underestimate in order to simplify the algebraic manipulations. The volume of this underestimating cylindrical shell must then be

$$\begin{aligned} & \pi \left(\left[(k+1) \frac{r}{n} \right]^2 - \left[k \frac{r}{n} \right]^2 \right) \left(\frac{h}{r} \left[r - k \frac{r}{n} \right] \right) \\ &= \pi r^2 h \left([k+1]^2 - k^2 \right) \left(1 - \frac{k}{n} \right) \frac{1}{n^2} \\ &= \pi r^2 h (2k+1) \left(1 - \frac{k}{n} \right) \frac{1}{n^2} \\ &= 2\pi r^2 h \frac{k}{n} \left(1 - \frac{k}{n} \right) \frac{1}{n} + \pi r^2 h \left(1 - \frac{k}{n} \right) \frac{1}{n} \frac{1}{n}. \end{aligned}$$

Let k have the successive values $1, 2, \dots, n$, and adding the resulting volumes, we have

$$\begin{aligned} & 2\pi r^2 h \left[\left(\frac{1}{n} \right) \left(1 - \frac{1}{n} \right) + \left(\frac{2}{n} \right) \left(1 - \frac{2}{n} \right) + \dots + \left(\frac{n-1}{n} \right) \left(1 - \frac{n-1}{n} \right) \right. \\ & \quad \left. + \left(\frac{n}{n} \right) \left(1 - \frac{n}{n} \right) \right] \left(\frac{1}{n} \right) \\ & + \left(\pi r^2 h \left[\left(1 - \frac{1}{n} \right) + \left(1 - \frac{2}{n} \right) + \dots + \left(1 - \frac{n-1}{n} \right) + \left(1 - \frac{n}{n} \right) \right] \frac{1}{n} \right) \left(\frac{1}{n} \right). \end{aligned}$$

The first term above is a number in the sequence whose limit is

$$\begin{aligned} & 2\pi r^2 h \int_0^1 t(1-t) dt \\ &= 2\pi r^2 h \int_0^1 (t - t^2) dt \\ &= 2\pi r^2 h \left[\frac{t^2}{2} - \frac{t^3}{3} \right]_{t=0}^{t=1} \\ &= 2\pi r^2 h \left(\frac{1}{2} - \frac{1}{3} \right) \\ &= \frac{1}{3} \pi r^2 h. \end{aligned}$$

The first factor in the second term is a number in the sequence whose limit is

$$\pi r^2 h \int_0^1 (1-t) dt$$

and the second factor in the second term is a number in the sequence

$$\left\{ \frac{1}{n}, n = 1, 2, \dots \right\},$$

whose limit is 0.

We conclude, of course, that the volume of a right circular cone of height h , whose base has radius r , is

$$\frac{1}{3} \pi r^2 h,$$

but, of more importance, we conclude that we did not need to find the actual volume of the small cylindrical shells, since one of the terms in the expression giving this actual volume had limit 0.

Looking again at the expression for the volume of a typical cylindrical shell, we have

$$\begin{aligned} & 2\pi r^2 h \frac{k}{n} \left(1 - \frac{k}{n}\right) \frac{1}{n} \\ &= [2\pi] \left[\left(k \frac{r}{n}\right) \right] \left[\frac{h}{r} \left(r - k \frac{r}{n}\right) \right] \left(\frac{r}{n}\right). \end{aligned}$$

The factor $\frac{r}{n}$ is the thickness of this shell. The factor $\frac{h}{r} \left(r - k \frac{r}{n}\right)$ is the height of the shell. The remaining factor $2\pi k \frac{r}{n}$ is the circumference of the circle which is the inner circle of the base of the shell.

The only term of the expression giving the volume of a typical shell, which term yields a non-zero limit for the volume of the cone, is then the product of three numbers. The volume of a rectangular box is also the product of three numbers. We can conclude that it is sufficient for us to estimate the volume of the cylindrical shell by considering this shell as having the same volume as a rectangular box whose height is the height of the shell, whose thickness is the thickness of the shell, and whose length is the length of the circumference of the inside of the shell. That is, we now imagine our solid of revolution as being made up of cylindrical shells, and we will find the volumes of these shells by slicing each one down its side, flattening them into thin rectangular boxes, and finding the volumes of these boxes.

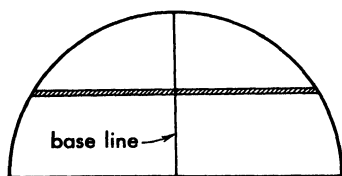


Fig. 12.13

As an illustration, we will recalculate the volume of a sphere. We start with a semicircle, taking the radius perpendicular to the base of the semicircle as our base line (Fig. 12.13).

If the radius of the sphere is r , then the length of the base line is r . If the variable t has, as its range, the interval:

$$\text{the set of all } t, \quad 0 \leq t \leq r,$$

then the thin rectangle t units from the center, along the base line will have length

$$2 \sqrt{r^2 - t^2}.$$

When revolved about the axis of revolution which is the base of the semi-circle, this thin rectangle forms a cylindrical shell whose height is

$$2 \sqrt{r^2 - t^2},$$

whose (inside) radius is t , and whose thickness is the length of the sub-interval into which the base line has been divided. Flattening this typical shell gives a thin rectangular box whose volume is

$$4\pi t \sqrt{r^2 - t^2} \frac{r}{n},$$

where n is the number of subdivisions. If this typical shell is the k -th one out from the axis of revolution, then the variable t has the value of $k \frac{r}{n}$, so that a typical term in the sum which approximates the volume of the sphere is

$$\begin{aligned} & 4\pi \left(k \frac{r}{n}\right) \sqrt{r^2 - \left(k \frac{r}{n}\right)^2} \frac{r}{n} \\ &= 4\pi r^3 \left(\frac{k}{n}\right) \sqrt{1 - \left(\frac{k}{n}\right)^2} \frac{1}{n}. \end{aligned}$$

When added, these typical terms give a sum whose limit is

$$4\pi r^3 \int_0^1 t \sqrt{1 - t^2} dt,$$

which integral is then the volume of our sphere.

In order to evaluate this integral, we have

$$\begin{aligned} f'(t) &= t \sqrt{1 - t^2}. \\ G_1(t) &= (1 - t^2)^{3/2}. \\ G_1'(t) &= \frac{3}{2} (1 - t^2)^{1/2} (-2t) \\ &= -3t \sqrt{1 - t^2}. \\ f(t) &= G_2(t) = -\frac{1}{3} (1 - t^2)^{3/2}. \\ \int_0^1 t \sqrt{1 - t^2} dt &= \left[-\frac{1}{3} (1 - t^2)^{3/2} \right]_{t=0}^{t=1} \\ &= \frac{1}{3}. \end{aligned}$$

Thus, the volume of a sphere of radius r is

$$\frac{4}{3} \pi r^3.$$

Exercises

Using the method of cylindrical shells, find the volumes of the solids obtained by revolving the portion of the plane bounded by the given equations, about the given line:

1. $y = x^2$, $y = 0$, $x = 1$, $x = 2$; about the x -axis.
2. $x^2 - y^2 = 1$, $x = 2$; about the y -axis.
3. $y = \sin x$, $y = 0$, $x = \pi/2$; about the y -axis.
4. $x^2 + y^2 = a^2$; about $x = b$ (where $b > a > 0$).
5. $y = x^{3/2}$, $y = 0$, $x = 4$; about y -axis.
6. $xy = 1$, $y = 1$, $y = 2$, $x = 0$; about x -axis.
7. $b^2x^2 + a^2y^2 = a^2b^2$; about $y = k$ (where $k > b > 0$).
8. A hole of radius R is drilled through the center of a sphere of radius r . Find the volume remaining.

12.7. ALGEBRAIC SIMPLICITY AND GEOMETRIC COMPLEXITY

In finding the area of a portion of the plane, or the volume of a solid of revolution, we used a particular line segment as a base line, divided this base line into (small) pieces, and used these small pieces to construct (small) rectangles, disks, washers, or cylindrical shells. The sum of the areas or volumes of these slices of the portion of the plane or solid of revolution was then approximated by considering these small slices as geometric objects whose area or volume we could find readily. The algebraic form of this resulting approximation to area or volume was then used to determine a function and an interval, so constructed that the integral of that function over that interval could reasonably and properly be called the area or volume of the geometric object in question. Finally, the integral obtained was evaluated.

It is possible to give formulas which treat areas and volumes derived from portions of the plane bounded by graphs of various functions. These formulas—which, of course, are theorems—normally contain many restrictions on the behavior of the functions involved, which restrictions tend to be overlooked in the applications. For example, a formula for the area bounded by the graph of a function, $f()$, the vertical lines with equations $x = a$ and $x = b$, and the x -axis is

$$A = \int_a^b f() d().$$

This formula looks good, and seems to give just what we have been doing. However, closer examination of all of the words shows that this formula, by itself, is nonsense. If this formula is to be valid for *all* portions of the plane bounded by the x -axis, two vertical lines, and the graph of a function,

then, among other things, this formula tells us that the area of that portion of the plane bounded by the graphs of $x = -1$, $x = 1$, the x -axis, and the graph of the equation $y = x^3$ will be the number

$$\begin{aligned} & \int_{-1}^1 x^3 dx \\ &= \frac{x^4}{4} \Big|_{x=-1}^{x=1} \\ &= 0, \end{aligned}$$

and 0 is a most unreasonable number for such an area. The error occurs in forgetting that area is defined in terms of the integral of a function over an interval of numbers, only when that function has non-negative values for that interval.

Similarly, even though a particular portion of the plane may be easily described by a few simple equations, this does not mean that one can find area or volume by blindly using formulas. For example, the portion of the plane lying above the x -axis, bounded by the graphs of $y = 2 - x^2$, $y = x^3$, and $y = -x$ appears to be bounded by the graphs of two functions (Fig. 12.14).

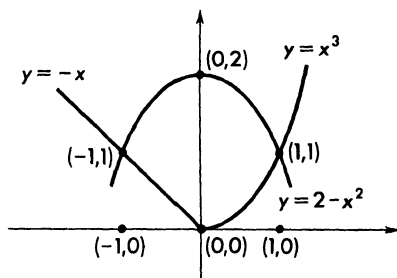


Fig. 12.14

If we use the vertical line segment joining $(0, 0)$ to $(0, 2)$ as our base line, and form thin horizontal rectangles perpendicular to this base line, then some of these rectangles will extend from the left side of the graph of $y = 2 - x^2$ to the right side of this graph, and others of these rectangles will extend from the graph of $y = -x$ to the graph of $y = x^3$.

If we use the horizontal line segment joining $(-1, 0)$ to $(1, 0)$ as our base line, and form vertical rectangles perpendicular to this base line, then some of these rectangles will extend from the graph of $y = -x$ to the graph of $y = 2 - x^2$ and others of these rectangle will extend from the graph of $y = x^3$ to the graph of $y = 2 - x^2$.

Then whichever base line we use (and either one will give us the area of this portion of the plane), we must be careful in examining the areas of typical rectangles to take into account the fact that the typical character of the rectangles differs in different parts of this portion of the plane.

By far the most convenient method of handling this geometric complexity is to consider the given portion of the plane to be made up of two or more (in this example, two) parts, each one of which parts does not have any such complex situation. Exactly, how we do this dividing depends on whether we want to use a horizontal or vertical base line. Whether we want to use a horizontal or vertical base line depends on how pleasant or unplea-

sant the anti-differentiation involved will be. Since there is no immediate way of determining which kind of base line gives a function easiest to integrate, a usual method of procedure is to set up the integral which will give area for every base line we can think of, examine the integrals found to see which will be easiest to evaluate, and then evaluate that one. If the problem is really complicated, then evaluation of the other integral or integrals found will provide a check on the accuracy of our first evaluation.

For the given example, we first split this portion of the plane into two subportions by the horizontal line segment joining $(-1, 1)$ to $(1, 1)$ (Fig. 12.15).

For the upper subportion, we let the vertical line segment joining $(0, 1)$ to $(0, 2)$ be the base line. If we let the variable t denote the y -coordinate of a point on the base line, then a typical rectangle through the point $(0, t)$ will have left-hand end at the point $(-\sqrt{2-t}, t)$ and right-hand end at the point $(\sqrt{2-t}, t)$. The length of this rectangle is then $2\sqrt{2-t}$ units. Since the thickness of the rectangle is the length of the subinterval into which the base line has been divided, we have the area of the upper subportion given by

$$2 \int_1^2 \sqrt{2-t} \, dt.$$

For the lower subportion, we use the vertical line segment joining $(0, 0)$ to $(0, 1)$. The variable s will denote the y -coordinate of a point on this base line. A typical rectangle through the point $(0, s)$ will have left-hand end at the point $(-s^{1/3}, s)$ and right-hand end at the point $(s^{1/3}, s)$. The thickness of the rectangle is the length of the subinterval into which the base line has been divided. Thus, the area of the lower subportion is given by

$$\int_0^1 [s^{1/3} + s] \, ds.$$

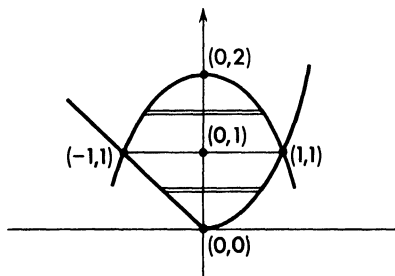


Fig. 12.15

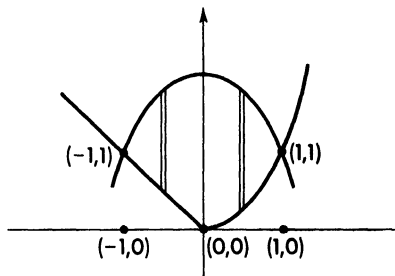


Fig. 12.16

The area of this portion of the plane is then

$$\int_0^1 [s^{1/3} + s] \, ds + 2 \int_1^2 \sqrt{2-t} \, dt.$$

Alternately, we split the given portion of the plane into two subportions by the vertical line segment joining $(0, 0)$ to $(0, 2)$ (Fig. 12.16).

For the left subportion, we use the horizontal line segment joining $(-1, 0)$ to $(0, 0)$ as the base line. If v denotes the x -coordinate of a point on this base line, then a typical vertical rectangle, perpendicular to the base line at the point $(v, 0)$, will have top end $(v, 2 - v^2)$ and bottom end $(v, -v)$. The length of this typical rectangle is then $(2 - v^2 + v)$ units. The thickness of the rectangle is the length of the subinterval into which the base line has been divided. Thus, the area of the left subportion is given by

$$\int_{-1}^0 (2 - v^2 + v) dv.$$

For the right subportion, we use the horizontal line segment joining $(0, 0)$ to $(1, 0)$ as the base line. If w denotes the x -coordinate of a point on this base line, then a typical vertical rectangle, perpendicular to the base line at the point $(w, 0)$, will have top end $(w, 2 - w^2)$ and bottom end (w, w^3) . The length of this typical rectangle is then $(2 - w^2 - w^3)$ units. The thickness of the rectangle is the length of the subinterval into which the base line has been divided. Thus, the area of the right subportion is given by

$$\int_0^1 (2 - w^2 - w^3) dw.$$

The area of this portion of the plane is then

$$\int_{-1}^0 (2 - v^2 + v) dv + \int_0^1 (2 - w^2 - w^3) dw.$$

Of the two expressions found for the area of the given portion of the plane—the first by using a vertical base line and the second by using a horizontal base line—the second is obviously easier to evaluate. (Both expressions, of course, give the same number.)

Again looking at this same portion of the plane, let us find the volume of the solid of revolution generated by revolving this portion of the plane about the x -axis.

If we take a vertical base line, the horizontal rectangular strips form cylindrical shells. The volume of this solid is then

$$\begin{aligned} & \int_1^2 2\pi [2\sqrt{2-t^2}] t dt + \int_0^1 2\pi [s^{1/3} + s] ds \\ &= 4\pi \int_1^2 \sqrt{2-t^2} + dt + 2\pi \int_0^1 [s^{4/3} + s^2] ds. \end{aligned}$$

If we take a horizontal base line, the vertical rectangular strips form washer-shaped disks. The volume of this solid is then

$$\begin{aligned} & \int_{-1}^0 \pi [(2 - v^2)^2 - (-v)^2] dv + \int_0^1 \pi [2 - w^2)^2 - (w^3)^2] dw \\ &= \pi \int_{-1}^0 [4 - 3v^2 + v^4] dv + \pi \int_0^1 [4 - 2w^2 + w^4 - w^6] dw. \end{aligned}$$

Both of these expressions for volume are complicated, but the evaluation of the first is somewhat easier than that of the second, since less arithmetic is involved. Thus to find the volume of the solid of revolution generated by the given portion of the plane, a vertical base line would be used.

Exercises

In each of the following, find an integral of the sum of integrals which allows the simplest determination of the area or volume mentioned.

1. The area of that part of the plane to the right of the y -axis bounded by the graphs of $y = x^2$, $y = 2 - x^2$, and $x + y = 0$.
2. The area of that part of the plane to the left of the y -axis bounded by the graphs in exercise 1.
3. The volume of the solid generated by revolving the area in exercise 1 about the y -axis.
4. The volume of the solid generated by revolving the area in exercise 1 about the line with equation $y = 1$.
5. The volume of the solid generated by revolving the area in exercise 2 about the x -axis.
6. The area of that part of the plane bounded by the graphs of $y = e^{-x}$, $y = \text{Log}(x + 1)$, and the line segment joining $(2, \text{Log } 3)$ to $(5, e^{-5})$.
7. The volume of the solid generated by revolving the area in exercise 6 about
 - a) the y axis.
 - b) the x -axis.
 - c) the line with equation $y = \text{Log } 3$.
 - d) the line with equation $x = 5$.
 - e) the line with equation $y = 4$.
 - f) the line with equation $x = -1$.

12.8. ARC LENGTH

The definitions of the trigonometric functions, $\sin()$ and $\cos()$, involved the length of portions of the circumference of a circle. Before this, all lengths had been measured along straight lines, and had been determined by the theorem (3.3) giving the distance between two points in a plane with a coordinate system. These two ways of measuring length—that of a straight line segment and that of an arc of the circumference of a circle—give the only kind of lengths which can be measured until we give ourselves a definition for length of a curve which is not a straight line segment nor an arc of the circumference of a circle.

This inability to find length is exactly the same as our previous inability to find area, and we will resolve the difficulty in the same manner: by giving a definition for the length of certain kinds of curves, which definition is a reasonable one.

To begin with, the only kind of "curve" which we will consider as having a "length" will be a segment of the graph of a function, which function has analytic properties sufficient to make the graph "nice." In this case, as usual, "nice" is defined to mean "having the properties we want when we want them."

Let us examine a function $f()$, whose domain includes the interval:

$$\text{the set of all } t, \quad a \leq t \leq b,$$

and look closely at the part of the graph of $f()$ which lies between the points $(a, f(a))$ and $(b, f(b))$.

For a given natural number n , we subdivide the interval:

$$\text{the set of all } t, \quad a \leq t \leq b,$$

into n parts of equal length. The points of division of this interval are then the numbers

$$a + \frac{b-a}{n}, \quad a + 2\frac{b-a}{n}, \quad a + 3\frac{b-a}{n}, \quad \dots, \quad a + (n-1)\frac{b-a}{n}.$$

We now examine two successive points of division:

$$\left(a + k\frac{b-a}{n}\right) \quad \text{and} \quad \left(a + (k+1)\frac{b-a}{n}\right),$$

where k is some natural number between 1 and n .

On the graph of $f()$, the points with these two numbers as x -coordinates will be

$$\left(a + k\frac{b-a}{n}, \quad f\left(a + k\frac{b-a}{n}\right)\right)$$

and

$$\left(a + (k+1)\frac{b-a}{n}, \quad f\left(a + (k+1)\frac{b-a}{n}\right)\right).$$

The (straight line) distance between these two points is

$$\begin{aligned} & \sqrt{\left(\left[a + (k+1)\frac{b-a}{n}\right] - \left[a + k\frac{b-a}{n}\right]\right)^2 + \left(f\left(a + (k+1)\frac{b-a}{n}\right) - f\left(a + k\frac{b-a}{n}\right)\right)^2} \\ &= \sqrt{\left(\frac{b-a}{n}\right)^2 + \left(f\left(a + (k+1)\frac{b-a}{n}\right) - f\left(a + k\frac{b-a}{n}\right)\right)^2} \\ &= \sqrt{1 + \left[\frac{f\left(a + (k+1)\frac{b-a}{n}\right) - f\left(a + k\frac{b-a}{n}\right)}{\frac{b-a}{n}}\right]^2} \left(\frac{b-a}{n}\right). \end{aligned}$$

The expression

$$\frac{f\left(a + (k+1)\frac{b-a}{n}\right) - f\left(a + k\frac{b-a}{n}\right)}{\frac{b-a}{n}}$$

is precisely the slope of the (short) line segment joining these two points. If the natural number n is large, then the second point will be close to this first, and this quotient will be a close approximate to the slope of the tangent line to the graph of $f()$, with the first point of the pair being the point of tangency. Replacing this expression for the slope of the (small) line segment by an expression for the slope of the tangent line at this point, we have

$$\sqrt{1 + \left[f'\left(a + k\frac{b-a}{n}\right)\right]^2} \left(\frac{b-a}{n}\right).$$

Letting k have values $1, 2, 3, \dots, n$ and adding, we have terms in a sequence whose limit will be

$$\int_a^b \{\sqrt{1 + \{f'\}^2}\} () d(),$$

provided that this integral exists. Since the original small line segments joining successive points on the graph are reasonable approximations to whatever we might define for the length of these small pieces of the graph, we have

Definition 12.1: If $f()$ is a function, and the interval:

the set of all $t, \quad a \leq t \leq b$

is in the domain of $f()$, and

$$\int_a^b \sqrt{1 + [f'(t)]^2} dt$$

exists, then this number is the *length* of that part of the graph of $f()$ between the points $(a, f(a))$ and $(b, f(b))$.

This definition will give lengths the same as those previously determined for straight lines and circles. For example, the length of that part of the graph of $y = x$ between $(0, 0)$ and $(1, 1)$ is

$$\int_0^1 \sqrt{1 + (1)^2} dt = \sqrt{2}$$

which is the distance between $(0, 0)$ and $(1, 1)$. Also, the length of that part of the graph of $y = \sqrt{1 - x^2}$ between $(0, 1)$ and $(1, 0)$ is

$$\begin{aligned} & \int_0^1 \sqrt{1 + \left[\frac{-t}{\sqrt{1-t^2}}\right]^2} dt \\ &= \int_0^1 \sqrt{1 + \frac{t^2}{1-t^2}} dt \end{aligned}$$

$$\begin{aligned}
 &= \int_0^1 \frac{1}{\sqrt{1-t^2}} dt \\
 &= \text{arc Sin } t \Big|_{t=0}^{t=1} \\
 &= \pi/2.
 \end{aligned}$$

As an example of a "curve" whose length can be found only by using this definition, the length of that part of the graph of $y = x^2$ between $(0, 0)$ and $(1, 1)$ is

$$\begin{aligned}
 &\int_0^1 \sqrt{1 + (2t)^2} dt \\
 &= \int_0^1 \sqrt{1 + 4t^2} dt.
 \end{aligned}$$

To evaluate this integral, we have

$$\begin{aligned}
 f'(z) &= \sqrt{1 + 4z^2} \\
 \{f(g)\}'(z) &= \sqrt{1 + 4[g(z)]^2} g'(z).
 \end{aligned}$$

If we let $g()$ be specified by

$$g(z) = \frac{1}{2} \sinh z,$$

then

$$\begin{aligned}
 \{f(g)\}'(z) &= \sqrt{1 + \sinh^2 z} \cosh z \\
 &= |\cosh z| \cosh z \\
 &= \cosh^2 z, \text{ since } \cosh z \geq 0 \\
 &= \frac{1}{2} + \frac{1}{2} \cosh 2z
 \end{aligned}$$

Thus,

$$\begin{aligned}
 \{f(g)\}(z) &= \frac{1}{2} z + \frac{1}{4} \sinh 2z \\
 f\left(\frac{\sinh z}{2}\right) &= \frac{1}{2} \sinh^{-1}(\sinh z) + \frac{1}{2} \sinh z \sqrt{1 + \sinh^2 z} \\
 &= \frac{1}{2} \sinh^{-1}\left(2 \frac{\sinh z}{2}\right) + \frac{\sinh z}{2} \sqrt{1 + 4 \left[\frac{\sinh z}{2}\right]^2} \\
 f(t) &= \frac{1}{2} \sinh^{-1} 2t + t \sqrt{1 + 4t^2}.
 \end{aligned}$$

We then have

$$\begin{aligned}
 \int_0^1 \sqrt{1 + 4t^2} dt &= \left[\frac{1}{2} \sinh^{-1} 2t + t \sqrt{1 + 4t^2} \right]_{t=0}^{t=1} \\
 &= \frac{1}{2} \sinh^{-1} 2 + \sqrt{5} \\
 &= \frac{1}{2} \text{Log}(2 + \sqrt{5}) + \sqrt{5}.
 \end{aligned}$$

Exercises

Find the length of that part of the indicated graph.

1. $x^{1/2} = y^{1/3}$ from $(0, 0)$ to $(1, 1)$.
2. The circumference of a circle centered at the origin, of radius a ($a > 0$).
3. $y = p \cosh (x/p)$ from $(-p, ?)$ to $(p, ?)$.
4. All of $x^{2/3} + y^{2/3} = a^{2/3}$.
5. $8y^2 = x(x - 4)^2$ from $(0, 0)$ to $(4, 0)$.
6. $y = \sin x$ from $(0, 0)$ to $(2\pi, 0)$.
7. $y = \text{Log } x$ from $(1, 0)$ to $(p, \text{Log } p)$ (where $p > 1$).
8. $y = x^3$ from $(0, 0)$ to $(1, 1)$.
9. $y = x$ from $(0, 0)$ to $(1, 1)$.
10. Find the most general function for whose graph the integral yielding arc length is the integral of a constant function.



INFINITE SERIES

13.1. SEQUENCES OF NUMBERS

In Chapter 7, we defined the integral of a function, $f()$, over an interval of numbers in the domain of that function, as the common value of the upper and lower integrals of that function. Each of these was defined to be the limit of a certain sequence of numbers. In this chapter, we will examine sequences of numbers, which sequences are formed in various ways.

It will be recalled that a sequence is defined to be a function whose domain is the set of natural numbers (1, 2, 3, 4, and so on). Further, if $\{a_n \mid n = 1, 2, 3, \dots\}$ is the range of values of a given sequence, and there is one fixed number A , to which the numbers a_n “get close” as we take “larger and larger” values of n , then we called this number, A the limit of the given sequence. To be precise, we defined “ $\lim_n a_n$ is the number A ” to mean: given any $\epsilon > 0$, there will exist an integer N , such that, if $n \geq N$, then $|a_n - A| < \epsilon$.

As an introductory example, we examine a sequence created in a not unusual, though perhaps not everyday, manner. We give a rule for this sequence as follows:

$$\begin{aligned}b_1 &= 1, \\b_2 &= 1 + 1/2, \\b_3 &= 1 + 1/2 + 1/4, \\b_4 &= 1 + 1/2 + 1/4 + 1/8,\end{aligned}$$

and, for any natural number n ,

$$b_n = 1 + 1/2 + 1/4 + 1/8 + \dots + 1/2^{n-1}.$$

We can give this rule in words: “Given any natural number n , find the sum of 1 and the reciprocals of all of the natural number powers of 2 up to and including the power $n - 1$.”

Perhaps a simpler formulation of this rule is that which refers to a linear coordinate system. We can consider a point which moves from the origin

to the right with the motion consisting of successive jumps from one point to another. The first jump lands us on the point with coordinate "1." The second jump covers half of the distance which the first jump covered and lands us on the point with coordinate " $1 + 1/2$." The third jump covers half of the distance which the second jump covered and lands us on the point with coordinate " $1 + 1/2 + 1/4$." The fourth jump covers half of the distance which the preceding jump covered and lands us on the point with coordinate " $1 + 1/2 + 1/4 + 1/8$." This process can be continued for n jumps.

With this geometric picture, we can note that each jump covers exactly half of the distance between the point we start from and the point with coordinate "2." There is then an extremely heavy hint of the fact that this sequence has the number 2 as its limit.

Before proceeding with an analysis of this sequence, it will be well to straighten out a matter of notation. Each of the numbers in the range of values of this sequence is the sum of certain other numbers. With the seeming exception of 1, each of these other numbers is the reciprocal of a natural number power of 2. Things will look neater if we can express all of the terms in these sums as the reciprocals of powers of 2. Accordingly, we choose to write a typical number in the range of values of our original sequence as

$$b_n = \frac{1}{2^0} + \frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^3} + \dots + \frac{1}{2^{n-1}}.$$

We can now consider each number of the form b_n as the sum of n numbers, *each* of which is the reciprocal of an integer power of 2.

There remains in our notation the non-neatness which results from using the subscript n on the left-hand side of the above equation, while the exponents on the right-hand side stop at $n - 1$. To avoid this awkwardness, we make use of the following device. We will take

$$s_0 = b_1,$$

$$s_1 = b_2,$$

$$s_2 = b_3,$$

and, in general,

$$s_n = b_{n+1}.$$

We then have

$$s_n = \frac{1}{2^0} + \frac{1}{2^1} + \frac{1}{2^2} + \dots + \frac{1}{2^n}.$$

The numbers in the set $\{s_n \mid n = 0, 1, 2, \dots\}$ do not seem to form the range of values of a sequence since they do not appear to be the range of values of a function whose domain is the set of all natural numbers. But only the presence of the subscript 0 prevents this set of numbers from being

the range of values of a sequence, so that all will be well if we merely amend the definition (7.1) of the word sequence to read:

“A sequence is a function whose domain is the set of natural numbers or the set of non-negative integers.”*

The individual terms in each of the sums, s_n , are themselves numbers from the range of values of the sequence

$$\{(n, 1/2^n) \mid n = 0, 1, 2, \dots\}.$$

In fact, in order to form the numbers in the sequence $\{(n, s_n) \mid n = 0, 1, 2, \dots\}$ we had to have available this other sequence.

As a further notational device, rather than use the space required to indicate our sequences by expressions of the form

$$\{(n, s_n) \mid n = 0, 1, 2, 3, \dots\},$$

we will use

$$\{s_n\}_{n=0}^{\infty}$$

To recapitulate in our new notation, we have begun with the sequence

$$\{1/2^n\}_{n=0}^{\infty},$$

and then formed the new sequence

$$\{s_n\}_{n=0}^{\infty} = \{1 + 1/2 + 1/2^2 + \dots + 1/2^n\}_{n=0}^{\infty},$$

and discovered strong evidence that

$$\lim_n s_n = 2.$$

We note that the term with index n in the sequence $\{s_n\}_{n=0}^{\infty}$ is the result of adding together all those terms in the sequence $\{1/2^n\}_{n=0}^{\infty}$ whose index does not exceed n . It sounds extremely plausible to say that the number 2, which is the limit of the sequence $\{s_n\}_{n=0}^{\infty}$ is the result of adding together *all* of the numbers in the sequence $\{1/2^n\}_{n=0}^{\infty}$, but we must tread very warily when we tamper with the basic laws governing the algebraic properties of numbers.

The first addition law, A-1, tells us that we can always find the sum of two numbers. Repeated application of this law and the law A-3 tells us that we can always add together the numbers in any *finite* set. It is not reasonable, however to expect that we can always add together *all* of the numbers in *any* infinite set. Further, even in those cases (such as the present example) in which we might expect to be able to add together all of the numbers in a given infinite set of numbers, we must be very careful to define exactly what we mean by this new kind of addition. We base this chapter on the following definitions:

* This is not really necessary, since we can always consider the set of ordered pairs of the form (n, s_{n-1}) .

Definition 13.1: An *infinite series* is the indicated sum of the numbers in a sequence. If the sequence is written $\{a_n\}_{n=0}^{\infty}$, then the infinite series arising from this sequence is written $\sum_{n=0}^{\infty} a_n$.

(The most significant word in this definition is "indicated.")

Definition 13.2: If the infinite series $\sum_{n=0}^{\infty} a_n$ is formed from the sequence $\{a_n\}_{n=0}^{\infty}$, and the new sequence $\{s_n\}_{n=0}^{\infty}$ is formed by

$$s_n = a_0 + a_1 + a_2 + \dots + a_n,$$

and has, as its limit, the number S , then this number S is called the *sum of the infinite series* $\sum_{n=0}^{\infty} a_n$, and this infinite series is said to be *convergent* to S . If the sequence $\{s_n\}_{n=0}^{\infty}$ formed above does not have a limit, then the infinite series $\sum_{n=0}^{\infty} a_n$ is said to be *divergent*.

(The notation $\sum_{n=0}^{\infty} a_n$ is read "the summation, from $n = 0$, of a_n .")

The example under consideration gives us the statement that the infinite series $\sum_{n=0}^{\infty} \frac{1}{2^n}$ is convergent to 2. This infinite series is but one example of the type known as a "geometric series." The following theorem applies to all such series:

Theorem 13.1:

If $|r| < 1$, then the infinite series

$$\sum_{n=0}^{\infty} r^n$$

is convergent to $\frac{1}{1-r}$.

If $|r| \geq 1$, then the infinite series

$$\sum_{n=0}^{\infty} r^n$$

is divergent.

Proof:

Part I, $|r| \neq 1$.

1. Let $s_n = r^0 + r^1 + r^2 + \dots + r^n$, $n = 0, 1, 2, \dots$

2. $s_n = \frac{1 - r^{n+1}}{1 - r}$, $n = 0, 1, 2, \dots$ by factoring $1 - r^{n+1}$.

3. $s_n = \frac{1}{1-r} + \frac{1}{1-r} r^{n+1}$, $n = 0, 1, 2, \dots$
by step 2.

4. If $|r| < 1$, then $\lim_n r^{n+1} = 0$ by an elementary theorem on limits of sequences.

5. If $|r| > 1$, then $\lim_n r^{n+1}$
does not exist since then $|r|^0 < |r|^1 < |r|^2 < \dots$

6. If $|r| < 1$, then $\lim_n s_n = \frac{1}{1-r}$ by steps 3 and 4.

7. If $|r| > 1$, then $\lim_n s_n$ does not exist by steps 3 and 5.

Part II, $r = -1$.

8. $s_n = (-1)^0 + (-1)^1 + (-1)^2 + \dots + (-1)^n$ by step 1.

9. If n is even, $s_n = 1$ by step 8.

10. If n is odd, $s_n = 0$ by step 9.

11. If $r = -1$, $\lim_n s_n$ does not exist by steps 8-10.

Part III, $r = 1$.

12. $s_n = (1)^0 + (1)^1 + (1)^2 + \dots + (1)^n$ by step 1.

13. $s_n = n + 1$ by step 12.

14. If $r = 1$, then $\lim_n s_n$ does not exist by step 13.

15. If $|r| < 1$, then $\sum_{n=0}^{\infty} r^n$ is convergent to $\frac{1}{1-r}$ by step 5 and the definition (13.2).

16. If $|r| \geq 1$, then $\sum_{n=0}^{\infty} r^n$ is divergent by steps 7, 11, and 14, and the definition (13.2).

Exercises

1. Perform the division $\frac{a^p - b^p}{a - b}$, where p is a natural number. Then substitute $a = 1$, $b = n + 1$, $b = r$.

2. Show that the set of numbers $\{c_n \mid n = 0, 1, 2, \dots\}$ is the range of values of the composite function $\{f(g)\}()$, where $g()$ is the sequence $\{(k, k - 1) \mid k = 1, 2, 3, \dots\}$, and $f()$ is the function given by $f(n) = c_n$, $n = 0, 1, 2, \dots$

3. Show how the set of numbers $\{d_p \mid p = 3, 4, 5, \dots\}$ may be considered as the range of values of a sequence which is the composite of two functions $\{f(g)\}()$, where the domain of $g()$ is the set of natural numbers.

4. Find the solution of the inequality $|r|^{x+1} < e$, $x \geq 0$, where e is a fixed number, $0 < e < 1$, and r is a fixed number, $|r| < 1$. (Hint: if $0 < e < 1$, then $\log(1/e) > 0$.)

5. How can the result of exercise 4 be used to show that, if

$$|r| < 1, \text{ then } \lim_n |r|^{n+1} = 0?$$

6. Show that

$$(-1)^0 + (-1)^1 + (-1)^2 + \dots + (-1)^n = \frac{1}{2}[1 + (-1)^n].$$

13.2. TESTS FOR CONVERGENCE AND DIVERGENCE

If we are presented with an infinite series of the form $\sum_{n=0}^{\infty} a_n$, then there are several theorems which enable us to determine whether the given series is convergent or divergent. Most of the tests do not tell us, in the case of convergence, what the sum is. In this section, we prove some of these theorems.

Theorem 13.2:

If $\sum_{n=0}^{\infty} a_n$ is convergent, then $\lim_n a_n = 0$.

Proof:

1. There is some number S , such that $\sum_{n=0}^{\infty} a_n$ is convergent to S by hypothesis.
2. Let $\{s_n\}_{n=0}^{\infty}$ be the sequence given by

$$s_n = a_0 + a_1 + \dots + a_n.$$
3. $\lim_n s_n = S$ by steps 2 and 3 and the definition (13.2).
4. Let $\{t_n\}_{n=1}^{\infty}$ be the sequence given by

$$t_n = a_0 + a_1 + \dots + a_{n-1}.$$
5. $a_n = s_n - t_n$, for $n = 1, 2, 3, \dots$ by steps 2 and 4.
6. $\lim_n a_n = \lim_n s_n - \lim_n t_n$ by step 5 and the theorem on the limit of the difference of two sequences.
7. $\lim_n t_n = S$ by step 1 and the definition (13.2).
8. $\lim_n a_n = 0$ by steps 3, 6, and 7.

This theorem is most often used to determine that a given infinite series is divergent. Since we cannot have $\sum_{n=0}^{\infty} a_n$ convergent unless we also have $\lim_n a_n = 0$, if $\lim_n a_n \neq 0$, then $\sum_{n=0}^{\infty} a_n$ must be divergent. However, this necessary condition for convergence is not sufficient. As will be seen, there are sequences with limit zero whose indicated sums are not convergent series. For example, the infinite series

$$\sum_{n=0}^{\infty} n, \quad \sum_{n=0}^{\infty} n^2, \quad \sum_{n=0}^{\infty} n^0, \quad \sum_{n=0}^{\infty} (-1)^n, \quad \text{and} \quad \sum_{n=0}^{\infty} \frac{n+1}{2n+3}$$

are all divergent.

Theorem 13.3:

If $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ are sequences of positive numbers; if $a_n \leq b_n$ for $n = 0, 1, 2, \dots$; and if $\sum_{n=0}^{\infty} b_n$ is convergent; then $\sum_{n=0}^{\infty} a_n$ is convergent.

Proof:

1. Let $\{s_n\}$ be the sequence given by

$$s_n = a_0 + a_1 + \dots + a_n.$$
2. Let $\{t_n\}$ be the sequence given by

$$t_n = b_0 + b_1 + \dots + b_n.$$
3. $\lim_n t_n$ exists by hypothesis.
4. $s_n \leq t_n, n = 0, 1, 2, \dots$ by hypothesis.
5. $s_{n+1} = s_n + a_{n+1}, n = 0, 1, 2, \dots$ by step 1.
6. $s_{n+1} > s_n, n = 0, 1, 2, \dots$ by step 5 and the hypothesis that $\{a_n\}$ is a sequence of positive numbers.
7. $\lim_n s_n$ exists by steps 4 and 6 and the fact that an increasing sequence of numbers which is bounded must have a limit.
8. $\sum_{n=0}^{\infty} a_n$ is convergent by step 7.

The usual name for the result of this theorem is "the comparison test for convergence." In order to use such a test to determine the convergence of a given series, it is necessary that the given series contain only positive terms, and it is necessary to know of some convergent series. Thus far, we know that certain geometric series are convergent.

Theorem 13.4:

If $\{a_n\}_{n=0}^{\infty}$ is a sequence of positive numbers;

if $\lim_n \frac{a_{n+1}}{a_n}$ exists and is the number L ;

and if $L < 1$;

then $\sum_{n=0}^{\infty} a_n$ is convergent.

Proof:

1. Let $\{s_n\}_{n=0}^{\infty}$ be the sequence given by

$$s_n = a_0 + a_1 + \dots + a_n.$$
2. $L < 1$ by hypothesis.
3. $\frac{1-L}{2} > 0$ by step 1.

4. There is a natural number N ,
such that, if $n \geq N$, then

$$\left| \frac{a_{n+1}}{a_n} - L \right| < \frac{1-L}{2}$$

by step 3, the hypothesis, and the definition of limit of a sequence.

5. If $n \geq N$, then

$$L - \frac{1-L}{2} < \frac{a_{n+1}}{a_n} < L + \frac{1-L}{2},$$

$$\text{so that } \frac{a_{n+1}}{a_n} < \frac{1+L}{2}$$

by step 4.

6. Let $t = \frac{1+L}{2}$; $t < 1$

by step 2.

7. If k is any natural number,

$$\frac{a_{N+k}}{a_N} = \frac{a_{N+1}}{a_N} \frac{a_{N+2}}{a_{N+1}} \frac{a_{N+3}}{a_{N+2}} \cdots \frac{a_{N+k-1}}{a_{N+k-2}} \frac{a_{N+k}}{a_{N+k-1}}$$

by the hypothesis that a_n is a sequence of positive numbers.

8. If k is any natural number,

$$\frac{a_{N+k}}{a_N} < t^k$$

by steps 5, 6, and 7.

9. If $n \geq N$, then there is a natural
number (or 0) p , such that

$$n = N + p$$

by O-2.

10. If $n \geq N$, then

$$s_n = a_0 + a_1 + \cdots + a_{N-1} + a_N + a_{N+1} + \cdots + a_{N+p}$$

by steps 1 and 9.

11. If $n \geq N$, then

$$s_n = s_{N-1} + a_N + a_{N+1} + a_{N+2} + \cdots + a_{N+p}$$

by steps 1 and 10.

12. If $n \geq N$, then

$$s_n = s_{N-1} + a_N \left[1 + \frac{a_{N+1}}{a_N} + \frac{a_{N+2}}{a_N} + \cdots + \frac{a_{N+p}}{a_N} \right]$$

by step 11 and the fact that $a_N \neq 0$

13. If $n \geq N$, then

$$s_n < s_{N-1} + a_N [1 + t + t^2 + \cdots + t^p]$$

by steps 8 and 12.

14. Let b_n be the sequence given by

$$b_n = s_{N-1} + a_N [1 + t + t^2 + \cdots + t^p], \text{ where } n = N + p.$$

15. $\{b_n\}_{n=N}^{\infty}$ is convergent

by step 6 and Theorem 13.1.

16. $\sum_{n=0}^{\infty} a_n$ is convergent

by steps 13 and 15 and Theorem 13.3

The result of this theorem is called the ratio test for determining convergence of an infinite series. (It is also true that, if $L > 1$, then the given series diverges.) As an example of the use of this test, the series

$\sum_{n=0}^{\infty} \frac{n}{3^n}$ is convergent since

$$\begin{aligned} \lim_n \frac{\frac{n+1}{3^{n+1}}}{\frac{n}{3^n}} &= \lim_u \frac{n+1}{3^{n+1}} \frac{3^n}{n} \\ &= \lim_n \frac{1}{3} \frac{n+1}{n} \\ &= 1/3 < 1. \end{aligned}$$

This ratio test for convergence is usable only for series made up of positive terms. To deal with series which may have an infinite number of negative terms, we have

Theorem 13.5:

If $\sum_{n=0}^{\infty} |a_n|$ is convergent, then $\sum_{n=0}^{\infty} a_n$ is convergent.

The proof of this theorem will be omitted.

Our purpose in this chapter is not that of giving an exhaustive treatment of the subject of infinite series, but merely to attain one major theorem which has to do with the calculus. For this purpose, we will need to use only the convergence tests given here.

To determine that a given series is divergent, we will use either of the following theorems.

Theorem 13.6:

If $\{c_n\}_{n=0}^{\infty}$ and $\{d_n\}_{n=0}^{\infty}$ are sequences of positive numbers; if $c_n \leq d_n$, for $n = 0, 1, 2, \dots$; and if $\sum_{n=0}^{\infty} c_n$ is divergent; then $\sum_{n=0}^{\infty} d_n$ is divergent

The proof of this theorem is similar to that for the comparison tests for convergence and is left as an exercise.

Theorem 13.7:

$\sum_{n=1}^{\infty} \frac{1}{n}$ is divergent.

(Note: It is necessary to begin the summation at $n = 1$, since we must never divide by 0.)

Proof of 13.7:

1. Let $c_n = \int_n^{n+1} \frac{1}{t} dt$, $n = 1, 2, 3, \dots$
2. Let $s_n = c_1 + c_2 + \dots + c_n$, $n = 1, 2, 3, \dots$

3. $s_n = \int_1^{n+1} \frac{1}{t} dt$ by steps 1 and 2 and the theorem that $\int_a^b f = \int_a^c f + \int_c^b f$.
4. $s_n = L(n+1)$ by theorems in Chapter 12.
5. $\lim_n s_n$ does not exist by properties of the function $L()$.
6. $\sum_{n=1}^{\infty} c_n$ is divergent by step 5 and definition.
7. If $n \leq t \leq n+1$, then $\frac{1}{t} \leq \frac{1}{n}$ by the ordering laws.
8. $\int_n^{n+1} \frac{1}{t} dt < \int_n^{n+1} \frac{1}{n} dt$ by steps 1 and 7.
9. $c_n < \frac{1}{n}$ by steps 1 and 8 and the evaluation of the integral of a constant function.
10. $\sum_{n=1}^{\infty} \frac{1}{n}$ is divergent by steps 6 and 9 and Theorem 13.6.

The series in this theorem is known as the harmonic series. The divergence of the harmonic series raises the question of the convergence or divergence of series of the form $\sum_{n=1}^{\infty} \frac{1}{n^p}$, where $p > 1$. (Comparison with the harmonic series shows that, if $p < 1$, then $\sum_{n=1}^{\infty} \frac{1}{n^p}$ is divergent.) We have

Theorem 13.8:

If $p > 1$, then $\sum_{n=1}^{\infty} \frac{1}{n^p}$ is convergent.

The proof of this theorem is similar to that of the theorem on the divergence of the harmonic series and is left as an exercise.

Exercises

Use the comparison test to determine the convergence or divergence of the following series:

1. $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$.
2. $\sum_{n=3}^{\infty} \frac{\sqrt{n}}{2n+1}$.
3. $\sum_{n=7}^{\infty} \frac{1}{(2n+1)3^n}$.
4. $\sum_{n=2}^{\infty} \frac{1}{\sqrt{n^2+n+1}}$.
5. $\sum_{n=1}^{\infty} \frac{1}{n(n+1)(n+2)}$.
6. $\sum_{n=0}^{\infty} \frac{1}{n^2+3}$.
7. $\sum_{n=6}^{\infty} \frac{1}{4^n+5}$.
8. $\sum_{n=7}^{\infty} \frac{1}{n^n}$.
9. $\sum_{n=31}^{\infty} \frac{2}{3+\sqrt{n}}$.
10. $\sum_{n=510}^{\infty} \frac{n}{n+1}$.

Use the ratio test to determine the convergence of the following series: (Remark: $n!$ is read “ n factorial” and $n! = (1)(2)(3)(4) \dots (n-1)(n)$, is its definition, so that $3! = 6$, $4! = 24$, and so on.)

$$11. \sum_{n=1}^{\infty} \frac{4n^2 - 1}{2^n}.$$

$$13. \sum_{n=0}^{\infty} \frac{1}{n!}.$$

$$12. \sum_{n=0}^{\infty} n \left(\frac{3}{4}\right)^n.$$

$$14. \sum_{n=0}^{\infty} \frac{2^n}{n!}.$$

$$15. \sum_{n=1}^{\infty} \frac{1(3)(5)(7) \dots (2n-1)}{3(6)(9)(12) \dots (3n)}$$

16. Prove Theorem 13.6.

17. Prove Theorem 13.8.

13.3. SERIES OF FUNCTIONS

As should be well-known by now, a function is a set of ordered pairs of numbers with no two ordered pairs having the same first element. A series of functions is simply the indicated sum of the functions in a sequence of functions. A sequence of functions is a collection of functions which is put into one-to-one correspondence with the set of natural numbers, or with the set of all integers not less than some fixed integer.

Let us consider the function $f()$, which will be given by the rule:

$$f(t) = \sum_{n=0}^{\infty} t^n, \quad 0 \leq t \leq 1/2.$$

We can consider this function as the indicated sum of the sequence of functions given by the expressions

$$\{t^n, 0 \leq t \leq 1/2\}_{n=0}^{\infty}.$$

We can also consider this function as having a rule which includes the instruction: find the sum of an infinite series.

If a is any given number in the domain of $f()$, we already have determined that

$$f(a) = \sum_{n=0}^{\infty} a^n = \frac{1}{1-a}.$$

Then the function $f()$ is a restriction of the function prescribed by the algebraic expression

$$\frac{1}{1-t}.$$

If we concentrate on the series $\sum_{n=0}^{\infty} a^n$, we will be able to obtain approximations to the number $f(a)$ by adding together some of the numbers of the sequence $\{a^n\}_{n=0}^{\infty}$.

Suppose we want these approximations to $f(a)$ to be numbers which are within 0.005 units of the number $f(a)$ itself. We have

$$\begin{aligned}
 f(0.1) &= 1.111\dots; & (0.1)^0 + (0.1)^1 + (0.1)^2 &= 1.11. \\
 f(0.2) &= 1.25; & (0.2)^0 + (0.2)^1 + (0.2)^2 + (0.2)^3 &= 1.248. \\
 f(0.3) &= 1.428\dots; & (0.3)^0 + (0.3)^1 + (0.3)^2 + (0.3)^3 + (0.3)^4 &= 1.4251. \\
 f(0.4) &= 1.6666\dots; & (0.4)^0 + (0.4)^1 + (0.4)^2 + (0.4)^3 + (0.4)^4 \\
 & & + (0.4)^5 + (0.4)^6 &= 1.663936. \\
 f(0.5) &= 2; & (0.5)^0 + (0.5)^1 + (0.5)^2 + (0.5)^3 + (0.5)^4 \\
 & & + (0.5)^5 + (0.5)^6 + (0.5)^7 + (0.5)^8 \\
 & & &= 1.9921875.
 \end{aligned}$$

In each of the sums on the right above, if we had taken fewer terms, our approximation to $f(a)$ would not be within 0.005 units of the actual value of $f(a)$. In general, we have

$$f(a) = \frac{1}{1-a}.$$

$$a^0 + a^1 + a^2 + \dots + a^n = \frac{1 - a^{n+1}}{1 - a}.$$

We want

$$\left| \left(\frac{1}{1-a} \right) - \left(\frac{1 - a^{n+1}}{1-a} \right) \right| < 0.005,$$

which is equivalent to

$$\frac{a^{n+1}}{1-a} < 0.005.$$

Since $a > 0$, $1 - a < 1$ and the truth of this last inequality will be implied by

$$a^{n+1} < 0.005.$$

But this is equivalent to saying

$$(n+1) L(a) < L(0.005),$$

and for $a < 1$, $L(a)$ is negative, so that we have

$$n+1 > \frac{L(0.005)}{L(a)}.$$

If we examine the function given by

$$n(t) = \frac{L(0.005)}{L(t)} - 1, \quad 0 \leq t \leq 1/2,$$

we find that

$$n'(t) = - \frac{L(0.005)}{(L(t))^2} \frac{1}{t} \quad 0 < t < 1/2,$$

or, since $L(0.005)$ is negative, $n'(t) > 0$ and $n(t)$ is an increasing function.

As a result, taking a to be a number near $1/2$, we need to take more terms to maintain the same degree of accuracy of our approximation to $f(a)$. We can say that of any two series $\sum a^n$ and $\sum b^n$, where $0 \leq a \leq 1/2$ and $0 \leq b \leq 1/2$, if $a < b$, then $\sum a_n$ converges more rapidly than $\sum b_n$, or conversely, that $\sum b_n$ converges more slowly than $\sum a_n$. Of the collection of series

$$\left\{ \sum_{n=0}^{\infty} t^n \mid 0 \leq t \leq 1/2 \right\},$$

the particular series $\sum_{n=0}^{\infty} (1/2)^n$ converges more slowly than any of the others. Since we have determined that, for $t = 1/2$, $t^0 + t^1 + t^2 + \dots + t^8$ will be a number within 0.005 units of $f(t)$, the fact that $\sum_{n=0}^{\infty} (1/2)^n$ is the most slowly convergent of all the series in this collection will imply that taking sums of this form (that is, up to and including t^8) will give us an approximation within 0.005 units of $f(t)$ for all values of t in the domain of this function.

If we now create a new function by extending the domain, by taking

$$g(t) = \sum_{n=0}^{\infty} t^n, \quad 0 \leq t < 1,$$

an examination of the function given by $n(t)$ shows that the collection of series of numbers

$$\sum_{n=0}^{\infty} t^n, \quad 0 \leq t < 1,$$

has no one most slowly convergent series, and that, given a desired degree of accuracy (say 0.005), we cannot say that taking finite sums of a certain length will give us the desired degree of accuracy for all allowable values of t .

These two different situations lead to:

Definition 13.3: The series of functions given by $\sum_{n=0}^{\infty} g_n(t)$ is *convergent pointwise* to the function $g()$ on an interval means that for each specific number a of this interval, the series of numbers $\sum_{n=0}^{\infty} g_n(a)$ is convergent to the number $g(a)$.

Definition 13.4: The series of functions given by $\sum_{n=0}^{\infty} f_n(t)$ is *convergent uniformly* to the function $f()$ on an interval means that

- 1) $\sum_{n=0}^{\infty} f_n(t)$ is convergent pointwise to $f()$ on that interval.
- 2) Given $\epsilon > 0$, there is a natural number N , such that, if $n > N$, then for every t in that interval,

$$|f(t) - [f_0(t) + f_1(t) + \dots + f_n(t)]| < \epsilon.$$

In the example under consideration, $\sum_{n=0}^{\infty} t^n$ converges pointwise to the function given by $\frac{1}{1-t}$ on the interval:

$$\text{the set of all } t, \quad 0 \leq t < 1,$$

and converges uniformly to the function given by $\frac{1}{1-t}$ on the interval:

$$\text{the set of all } t, \quad 0 \leq t \leq 1/2.$$

While the determination of intervals of pointwise or uniform convergence for a given series of functions is sometimes extremely difficult, we are going to consider only series of the form

$$\sum a_n t^n,$$

where $\{a_n\}$ is some sequence of numbers. Such a series is called a power series in t . A series of the form

$$\sum a_n (t - b)^n$$

is called a power series, centered at b . The following most powerful theorem is used for power series:

Theorem 13.9:

Let $\{a_n\}_{n=0}^{\infty}$ be a sequence of numbers:

- 1) If $\lim_n \frac{|a_{n+1}|}{|a_n|}$ exists and is the number $L > 0$,

then $\sum a_n t^n$ is convergent pointwise on the interval:

$$\text{the set of all } t, \quad |t| < 1/L.$$

- 2) Further, if $r = 1/L$, and d is any number, such that $0 < d < r$, then $\sum a_n t^n$ is uniformly convergent on the interval:

$$\text{the set of all } t, \quad |t| \leq r - d.$$

- 3) If $\lim_n \frac{|a_{n+1}|}{|a_n|} = 0$, then $\sum a_n t^n$ converges pointwise on the set of all (real) numbers and converges uniformly on any bounded set of numbers.

- 4) If $\lim_n \frac{|a_{n+1}|}{|a_n|} = +\infty$, then $\sum a_n t^n$ diverges for any $t \neq 0$.

- 5) If r is as in part 1, and c is any number such that $|c| > r$, then the series of numbers $\sum a_n c^n$ is divergent.

The proof of this theorem results immediately from the ratio test. The number r is called the "radius of convergence of the power series $\sum a_n t^n$." The interval mentioned is called the "interval of convergence." Shifting to $\sum a_n (t - b)^n$, we have, similarly, a radius of convergence, and the intervals involved center around the number b .

As a result of this theorem, any power series in the variable t , centered around 0 or around any other number, is an expression (but not an algebraic one necessarily) which can be used to prescribe a function whose domain is the interval of convergence of that series. If we discover that the function

is actually the same as a known function (or a restriction of that known function to some interval), then we have found an alternate way of finding numbers in the range of values of that known function or, more importantly, of finding approximations to numbers in the range of values of that known function.

The example with which we started showed how to get approximations to the value of $\frac{1}{1-t}$ for values of t between 0 and 1, without doing any division. Thus, the power series provided us with an alternate to the rule given by this expression.

Exercises

- What values of n will make $(1 + a + a^2 + \dots + a^n)$ be within 0.05 units of $\frac{1}{1-a}$ for $a = 0.6, 0.7, 0.8, 0.9, 0.99, 0.999$?
- If $g(t) = \sum_{n=0}^{\infty} t^n$, $0 \leq t < 1$, find $g'(t)$.

Determine the intervals of convergence for the functions given by:

- $\sum_{n=2}^{\infty} ns^n.$
- $\sum_{n=1}^{\infty} \frac{1(3)(5)(7)\dots(2n-1)}{2(4)(6)(8)\dots(2n)} t^n.$
- $\sum_{n=3}^{\infty} \frac{u^n}{n}.$
- $\sum_{n=7}^{\infty} \frac{(-1)^n}{(2n)!} v^{2n}.$
- $\sum_{n=5}^{\infty} \frac{(n+2)!}{2n!} w^n.$
- $\sum_{n=4}^{\infty} \frac{x^n}{\sqrt{n}}.$
- $\sum_{n=6}^{\infty} \frac{(-1)^n}{(2n+1)!} y^{2n+1}.$
- $\sum_{n=8}^{\infty} \frac{z^n}{n^2}.$
- Find the new series which results from term-by-term differentiation of the series in exercises 3-7. Find the radius of convergence of this new series in each case, and compare this interval of convergence with the original interval of convergence.

13.4. DIFFERENTIATION, ANTI-DIFFERENTIATION, AND INTEGRATION

We state without proof a theorem which makes calculus with functions given by power series very simple.

Theorem 13.10:

If $\sum_{n=0}^{\infty} a_n (t - b)^n$ has radius of convergence r and gives the function $f()$, then

- 1) $\sum_{n=1}^{\infty} n a_n (t - b)^{n-1}$ has radius of convergence r and gives the function $f'()$.
- 2) If $g()$ is any anti-derivative of $f()$, then there is a constant c , such that

$$\sum_{n=0}^{\infty} \frac{a_n}{n+1} (t - b)^{n+1}$$

has radius of convergence r and gives the function $\{g + c\}()$.

- 3) If p and q are any two numbers in the interval of convergence of $\sum_{n=0}^{\infty} a_n (t - b)^n$, then

$$\int_p^q f(t) dt = \sum_{n=0}^{\infty} \left\{ a_n \int_p^q (t - b)^n dt \right\}.$$

Thus, a power series may be differentiated term-by-term, integrated term-by-term, and anti-derivatives may be found term-by-term, provided only that all procedures do not lead us out of the interval of convergence.

We noted in the preceding section that a power series might give a new expression for a function already known. Theorem 13.10 provides us with the tools to give an example of such a situation: We consider the function $f()$, given by

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

To find the radius of convergence of this power series, we have:

$$|a_n| = a_n = \frac{1}{n!}.$$

$$\frac{|a_{n+1}|}{|a_n|} = \frac{n!}{(n+1)!} = \frac{1}{n+1}.$$

$$\lim_n \frac{|a_{n+1}|}{|a_n|} = \lim_n \frac{1}{n+1} = 0.$$

Thus, this power series converges pointwise for all numbers, and converges uniformly on every bounded set of numbers. For the derivative, we have

$$\begin{aligned} f'(t) &= \sum_{n=0}^{\infty} n \frac{1}{n!} t^{n-1} \\ &= \sum_{n=1}^{\infty} \frac{t^{n-1}}{(n-1)!} \\ &= \sum_{n=0}^{\infty} \frac{t^n}{n!} \\ &= f(t). \end{aligned}$$

The function $f()$ is then its own derivative. $f()$ is not a constant function, so that we must have $f() = \exp()$, since the function $\exp()$ is the only non-constant function which is its own derivative. We then have the power series representation

$$e^t = \sum_{n=0}^{\infty} \frac{t^n}{n!}.$$

With this manner of expressing e^t , we can easily compute approximations of e^t for values of t near 0. For example,

$$e^{0.1} = \sum_{n=0}^{\infty} \frac{(0.1)^n}{n!} = 1 + 0.1 + \frac{(0.1)^2}{2} + \frac{(0.1)^3}{6} + \dots,$$

which is approximately the number 1.1053. That is, the number 1.1053 is within 0.00005 units of the number $e^{0.1}$.

As an example of integration term-by-term:

$$\begin{aligned} \int_0^{0.1} \frac{dt}{1+t^4} &= \int_0^{0.1} \frac{1}{1-(-t^4)} dt \\ &= \int_0^{0.1} \sum_{n=0}^{\infty} (-t^4)^n dt \\ &= \sum_{n=0}^{\infty} (-1)^n \int_0^{0.1} t^{4n} dt \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{4n+1} t^{4n+1} \Big|_0^{0.1} \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{4n+1} (0.1)^{4n+1} \\ &= 0.1 - \frac{1}{5} (0.1)^5 + \frac{1}{9} (0.1)^9 - \dots, \end{aligned}$$

which is approximately 0.999998. That is, the number 0.999998 is within 0.0000005 units of the integral given.

We can also use Theorem 13.10 to evaluate familiar functions. By definition of the function $\text{Log } ()$,

$$\text{Log } t = \int_1^t \frac{dv}{v}.$$

$$\frac{1}{v} = \frac{1}{1 - (1 - v)} = \sum_{n=0}^{\infty} (1 - v)^n, \quad |1 - v| < 1.$$

Thus,

$$\begin{aligned} \text{Log } t &= \int_1^t \sum_{n=0}^{\infty} (1 - v)^n dv \\ &= \sum_{n=0}^{\infty} \int_1^t (1 - v)^n dv \\ &= \sum_{n=0}^{\infty} \left. \frac{-1}{n+1} (1 - v)^{n+1} \right|_1^t \\ &= - \sum_{n=0}^{\infty} \frac{(1 - t)^{n+1}}{n+1}, \quad |1 - t| < 1. \end{aligned}$$

Then

$$\begin{aligned} \text{Log } (1/2) &= - \sum_{n=0}^{\infty} \frac{1}{2^n(n+1)} \\ &= - \left[1 + \frac{1}{2(2)} + \frac{1}{3(4)} + \frac{1}{4(8)} + \frac{1}{5(16)} \right. \\ &\quad \left. + \frac{1}{6(32)} + \dots \right], \end{aligned}$$

which is approximately -0.693 . Thus, $\text{Log } 2$ is approximately 0.693 .

A further result which is preliminary to the main theorem of the chapter is:

Theorem 13.11:

If $\sum_{n=0}^{\infty} a_n(t - b)^n$ has radius of convergence r and gives the function $f()$, then $f()$ has derivatives of all order at b and

$$f^{(k)}(b) = a_k k!, \quad k = 1, 2, \dots$$

Proof:

1. $f(t) = \sum_{n=0}^{\infty} a_n(t - b)^n, \quad |t - b| < r$
by hypothesis.
2. $f'(t) = \sum_{n=1}^{\infty} n a_n(t - b)^{n-1}, \quad |t - b| < r$
by Theorem 13.10.

3. $f''(t) = \sum_{n=2}^{\infty} n(n-1) a_n(t-b)^{n-2}, \quad |t-b| < r$
by Theorem 13.10.
4. $f^{(k)}(t) = \sum_{n=k}^{\infty} n(n-1)(n-2)\dots(n-k+1) a_n(t-b)^{n-k}, \quad |t-b| < r$
by k applications of Theorem 13.10.
5. $f^{(k)}(b) = k(k-1)(k-2)\dots(3)(2)(1) a_k$
by step 4 and the fact that $0^0 = 1$.
6. $f^{(k)}(b) = k! a_k$
by step 5 and the definition of factorial.

This theorem has the curious consequence that we can find the values of the derivative of a function given by a power series without doing any differentiating. For example, if $f()$ is given by

$$f(t) = \frac{1}{1-t^2} = \sum_{n=0}^{\infty} t^{2n}, \quad |t| < 1,$$

For any natural number k ,

$$f^{(2k)}(0) = (2k)! \quad \text{and} \quad f^{(2k-1)}(0) = 0.$$

Exercises

1. Use term-by-term anti-differentiation to find a power series expansion, centered at 0, of the functions given by
 - a) $(1-t)^{-2}$.
 - b) $(1-t)^{-3}$.
 - c) $(1-t)^{-k}$, where k is some natural number.
2. Show that the functions, given by

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} u^{2n+1} \quad \text{and} \quad \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} v^{2n},$$

have the property $\{f'' + f\}() = \theta()$. (Recall that $\theta()$ is the constant function whose domain is the set of all numbers and whose range of values consists of the single number 0.)

3. a) Find a power series expansion, centered at 0, of the function given by $\frac{1}{a+bw}$, where a and b are constants. For what values of w will this expansion be valid?
- b) For $g()$ the function given by $g(x) = (a+bx)^{-1}$, what is $g^{(k)}(0)$?

Find approximations to the following numbers by using three terms of the appropriate power series expansions:

4. $\int_0^{0.2} \frac{dy}{1+y}$.
5. $\int_0^{0.2} \frac{dz}{1-z^2}$.
6. $\int_0^{0.2} \frac{dr}{1+r^3}$.

7. $e^{0.2}$.
8. $e^{-0.2}$.
9. $\cosh(0.2)$.
10. $\text{Log}(\cosh(0.2))$.

13.5. TAYLOR'S THEOREM AND TAYLOR'S SERIES

Among the theorems of elementary calculus is

Theorem 13.12:

(The Mean Value Theorem)

If $f()$ is a function and the interval:

the set of all t , $a \leq t \leq b$,

is in the domain of $f'()$, then there exists a number c between a and b , for which

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Geometrically, this theorem says that there is at least one point $(c, f(c))$ on the graph of $f()$ such that the tangent line to the graph at this point is parallel to the line joining $(a, f(a))$ and $(b, f(b))$. We use the mean value theorem to prove:

Theorem 13.13:

(Taylor's Theorem)

If $f()$ is a function; if the interval:

the set of all t , $|t - b| < r$,

is in the domains of $f()$, $f'()$, $f''()$, ..., $f^{(n+1)}()$, and x is a number such that $0 < |x - b| < r$, then there is a number c between b and x , such that

$$\begin{aligned} f(x) = & f(b) + (x - b)f'(b) + \frac{(x - b)^2}{2!}f''(b) + \dots \\ & \dots + \frac{(x - b)^n}{n!}f^{(n)}(b) + \frac{(x - b)^{n+1}}{(n + 1)!}f^{(n+1)}(c). \end{aligned}$$

Proof:

1. Let $F()$ be the function given by

$$\begin{aligned} F(t) = & f(x) - \left\{ f(t) + (x - t)f'(t) + \frac{(x - t)^2}{2!}f''(t) + \dots \right. \\ & \left. \dots + \frac{(x - t)^{n-1}}{(n - 1)!}f^{(n-1)}(t) + \frac{(x - t)^n}{n!}f^{(n)}(t) \right\} \\ & - \left[\frac{(x - t)^{n+1}}{(x - b)^{n+1}} \right] \left[f(x) - \left\{ f(b) + (x - b)f'(b) \right. \right. \\ & \left. \left. + \frac{(x - b)^2}{2!}f''(b) + \dots + \frac{(x - b)^n}{n!}f^{(n)}(b) \right\} \right], \\ & |t - b| < r. \end{aligned}$$

2. $F'(t)$ is given by

$$\begin{aligned} F'(t) = & - \left\{ f'(t) + [-f'(t) + (x-t)f''(t)] \right. \\ & + \left[-(x-t)f''(t) + \frac{(x-t)^2}{2!}f'''(t) \right] + \dots \\ & + \dots + \left[-\frac{(x-t)^{n-2}}{(n-2)!}f^{(n-1)}(t) + \frac{(x-t)^{n-1}}{(n-1)!}f^{(n)}(t) \right] \\ & \left. + \left[-\frac{(x-t)^{n-1}}{(n-1)!}f^{(n)}(t) + \frac{(x-t)^n}{n!}f^{(n+1)}(t) \right] \right\} \\ & + \left[(n+1) \frac{(x-t)^n}{(x-b)^{n+1}} \right] \left[f(x) - \left\{ f(b) + (x-b)f'(b) + \dots \right. \right. \\ & \left. \left. \dots + \frac{(x-b)^n}{n!}f^{(n)}(b) \right\} \right] \end{aligned}$$

by step 1.

$$\begin{aligned} 3. F'(t) = & - \left[\frac{(x-t)^n}{n!}f^{(n+1)}(t) \right] \\ & + \left[(n+1) \frac{(x-t)^n}{(x-b)^{n+1}} \right] \left[f(x) - \left\{ f(b) + \dots + \frac{(x-b)^n}{n!}f^{(n)}(b) \right\} \right] \end{aligned}$$

by step 2 and cancellation.

$$4. F(x) = 0 \quad \text{by step 1.}$$

$$5. F(b) = 0 \quad \text{by step 1.}$$

6. There is a number c between b and x , such that $F'(c) = 0$ by steps 4 and 5 and the mean value theorem.

7. There is a number c between b and x , such that

$$\begin{aligned} 0 = & - \left[\frac{(x-c)^n}{n!}f^{(n+1)}(c) \right] \\ & + \left[(n+1) \frac{(x-c)^n}{(x-b)^{n+1}} \right] \left[f(x) - \left\{ f(b) + (x-b)f'(b) + \dots \right. \right. \\ & \left. \left. + \frac{(x-b)^n}{n!}f^{(n)}(b) \right\} \right] \end{aligned}$$

by steps 3 and 6.

8. There is a number c between b and x , such that

$$\begin{aligned} 0 = & - \frac{(x-b)^{n+1}}{(n+1)!}f^{(n+1)}(c) + f(x) - f(b) - (x-b)f'(b) \\ & - \frac{(x-b)^2}{2!}f''(b) - \dots - \frac{(x-b)^n}{n!}f^{(n)}(b) \end{aligned}$$

by step 7 and algebraic manipulation

9. There is a number c between b and x , such that

$$f(x) = f(b) + (x-b)f'(b) + \frac{(x-b)^2}{2!}f''(b) + \dots \\ \dots + \frac{(x-b)^n}{n!}f^{(n)}(b) + \frac{(x-b)^{n+1}}{(n+1)!}f^{(n+1)}(c).$$

As a result of this theorem, we have a way of estimating the error involved in replacing a series of the form

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

with the finite sum of the terms whose index does not exceed some fixed natural number. For example, if $f(\cdot)$ is the function $\cos(\cdot)$, we have

$$\begin{array}{ll} f(t) = \cos t & f(0) = 1 \\ f'(t) = -\sin t & f'(0) = 0 \\ f''(t) = -\cos t & f''(0) = -1 \\ f'''(t) = \sin t & f'''(0) = 0 \\ f^{(4)}(t) = \cos t & \end{array}$$

If x is a number near 0, we have, from Taylor's theorem

$$\cos x = \cos 0 + x \sin 0 - \frac{x^2}{2!} \cos 0 + \frac{x^3}{3!} \sin 0 + \frac{x^4}{4!} \cos c,$$

where c is a number between 0 and x . That is,

$$\cos x = 1 - \frac{x^2}{2} + \frac{x^4}{24} \cos c.$$

For $x = 0.1$,

$$\cos 0.1 = 1 - \frac{0.01}{2} + \frac{0.0001}{24} \cos c, \quad \text{where } 0 < c < 0.1$$

Since $|\cos c| < 1$, the third term above is less than 0.00005 and we have that 0.995 (the sum of the first two terms) is a number within 0.00005 units of $\cos 0.1$.

If, in Taylor's theorem, we are dealing with a function and an interval in the domain of the derivatives of all orders, then we have

Theorem 13.14:

If $f(\cdot)$ is a function and b is a number in the domain of the derivatives of all orders of $f(\cdot)$, then there is a positive number r , such that

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(b)}{n!} (x-b)^n, \quad \text{for } |x-b| < r.$$

The series in this theorem is called the Taylor's series of $f()$, centered at the b . Theorem 13.14 says that any function which possesses derivatives of all orders can be represented—at least in some interval—by a power series. Theorem 13.11 tells us that this power series expansion is unique—that is it will agree with any power series centered about the same number, which other power series is found in any other way.

Exercises

Find the Taylor's series, centered at 0, of the functions given by the following expressions. Then find the interval of convergence for each series found.

1. $1 + r$.
2. $(1 - s)^{-1}$.
3. $\text{Log } (1 + t)$.
4. e^u .
5. $\sin v$.
6. $\cos w$.
7. $\text{arc Tan } x$.
8. $\sinh y$.
9. $(1 + z)^{-2/3}$.
10. Find an estimate of the error introduced when only the first four terms of the Taylor's series found above in exercises 1-9 are used to evaluate these functions at 0.1.

13.6. RATIONAL FUNCTIONS

The possibility of expanding a differentiable function as a Taylor's series allows us to re-examine all functions more complicated than polynomial functions by considering these functions as "polynomials of infinite degree." Of course, we must be careful never to emerge from the intervals of convergence which the series have, but—at least piecewise—we should be able to treat by means of series any function which has derivatives of all orders.

These last few sections of this book will examine a method for treating all of the functions previously considered either as polynomials or as power series. We will start with the following differentiation theorems:

$$\begin{aligned}
 k'() &= \theta(); \\
 \{I^n\}'() &= \{nI^{n-1}\}(), \text{ where } n \text{ is a natural number;} \\
 \{f + g\}'() &= \{f' + g'\}(); \\
 \{fg\}'() &= \{fg' + f'g\}(); \\
 \{f(g)\}'() &= \{f'(g)g'\}();
 \end{aligned}$$

and derive all others from these and from the results established for power series. We consider only the differentiation theorems for the new functions we are going to create, since the Fundamental Theorem of Calculus allows us to apply these theorems to integrals.

The differentiation theorems above allow us to find the derivative of any polynomial function. What we intend to do in this section is to give a definition for rational functions and then to show how the derivatives of such functions can be found.

From algebra, we know that any expression of the form $P(x)/Q(x)$, where $P()$ and $Q()$ are polynomial functions, will be a symbol for some number whenever x is any number other than a zero of the function $Q()$. Excluding such non-allowed values of x , we have

$$\frac{P(x)}{Q(x)} = P(x)f(Q(x)),$$

where $f(t) = 1/t$. For the functions involved, we would be able to use the differentiation theorems above to get to

$$\{P/Q\}'(x) = P'(x)f(Q(x)) + P(x)f'(Q(x))Q'(x).$$

The theorems above tell us what $P'(x)$ and $Q'(x)$ are, but do not give us a rule for $f'()$. In fact, if we are allowing only polynomials and series, we must first give a definition for this function $f()$. Our point is that we are going to find a way to take reciprocals which never require us to do arithmetic with unpleasant and complicated numbers. We will consider integers as pleasant and uncomplicated numbers. In order to accomplish this, we create a new function, which we will call $\text{Rec}()$, and show that this function is, in fact, the function which we previously called $\{\frac{1}{t}\}()$.

If we were allowed to use $f(t) = 1/t$, we would have

$$f'(t) = -\frac{1}{t^2};$$

$$f''(t) = \frac{2}{t^3};$$

$$f'''(t) = -\frac{3!}{t^4};$$

$$\vdots$$

$$f^{(n)}(t) = (-1)^n \frac{n!}{t^{n+1}}.$$

Naturally, we cannot allow $t = 0$. If $\text{Rec}()$ is to be the same function as $\{\frac{1}{t}\}()$, then the Taylor's series must be the same for these two functions. Accordingly, we define $\text{Rec}()$ by

Definition 13.5: The function $\text{Rec}()$ is given by:

- 1) The domain of $\text{Rec}()$ is the set of all numbers other than 0.
- 2) If k is any integer, $\text{Rec}(k) = 1/k$.
- 3) If $t > 1/2$ or $t \leq -1/2$, and t is not an integer, then we find the integer, k_t , nearest to t (if t is midway between two integers, we take the smaller of the two), then

$$\text{Rec}(t) = \frac{1}{k_t} \sum_{n=0}^{\infty} \frac{(k_t - t)^n}{k_t^n}.$$

It should be noted that the interval of convergence in the series given in part 3 of this definition is the set of all t , $|t - k_t| < |k_t|$. According to this rule, we find, for example, $\text{Rec}(7/4)$ as follows:

2 is the integer nearest to $7/4$,

$$\begin{aligned} \text{Rec}(7/4) &= \frac{1}{2} \sum_{n=0}^{\infty} \frac{1}{2^n} \left(\frac{1}{4}\right)^n \\ &= \frac{1}{2} \sum_{n=0}^{\infty} \frac{1}{8^n}. \end{aligned}$$

If we are content with an approximation to $\text{Rec}(7/4)$, we need merely take a sufficient number of terms in this series to give us our desired degree of accuracy. (We note that

$$\frac{1}{2} \sum_{n=0}^{\infty} \frac{1}{8^n} = \frac{1}{2} \left(\frac{1}{1 - 1/8} \right) = \frac{1}{2} \cdot \frac{8}{7} = \frac{4}{7},$$

which is indeed the reciprocal of $7/4$.)

For the derivative, we have

Theorem 13.15:

$$\{\text{Rec}\}'() = \{- (\text{Rec})^2\}().$$

Proof:

(In these last few sections, we will dispense with numbering the steps of our proofs, and will give only a few significant reasons.)

$$[\text{Rec}(t)]^2 = \left[\frac{1}{k} \sum_{n=0}^{\infty} \frac{(k - t)^n}{k^n} \right]^2 = \frac{1}{k^2} \left[\sum_{n=0}^{\infty} \left(\frac{k - t}{k} \right)^n \right]^2$$

For any number s , with $|s| < 1$,

$$\begin{aligned}
 \left[\sum_{n=0}^{\infty} s^n \right]^2 &= [1 + s + s^2 + s^3 + \dots + s^n + \dots] \\
 &\quad [1 + s + s^2 + s^3 + \dots + s^n + \dots] \\
 &= 1 + s + s^2 + s^3 + \dots + s^n + \dots \\
 &\quad + s + s^2 + s^3 + \dots + s^n + \dots \\
 &\quad + s^2 + s^3 + \dots + s^n + \dots \\
 &\quad + s^3 + \dots + s^n + \dots \\
 &\quad \cdot \\
 &\quad \cdot \\
 &\quad + s^n + \dots \\
 &\quad + \dots \\
 &= 1 + 2s + 3s^2 + 4s^3 + \dots + (n+1)s^n + \dots \\
 &= \sum_{n=1}^{\infty} ns^{n-1}.
 \end{aligned}$$

Thus,

$$\frac{1}{k^2} \left[\sum_{n=0}^{\infty} \left(\frac{k-t}{k} \right)^n \right]^2 = \frac{1}{k^2} \sum_{n=1}^{\infty} n \left(\frac{k-t}{k} \right)^{n-1}.$$

But

$$\begin{aligned}
 \text{Rec}'(t) &= \frac{1}{k} \sum_{n=0}^{\infty} \frac{n(k-t)^{n-1}}{k^n} (-1) \text{ by Theorem 13.9,} \\
 &= -\frac{1}{k^2} \sum_{n=1}^{\infty} n \left(\frac{k-t}{k} \right)^{n-1}
 \end{aligned}$$

Theorem 13.16:

$R()$ is a rational function if and only if there are polynomial functions, $P()$ and $Q()$, such that $R() = \{P \cdot \text{Rec}(Q)\}()$.

Proof:

$\text{Rec}()$ is identical with $\left\{ \frac{1}{x} \right\}()$.

We have thus succeeded in creating rational functions by using only polynomials and power series.

Exercises

1. Find approximations, accurate to within 0.00005 units, of $\text{Rec}(t)$ for $t = 5/2, -3/4, 1/5, 20.1/49.9, 500.3$.
2. Use the method of proof of Theorem 13.15 to show that

$$\{\text{Rec}''\}() = \{2 \text{Rec}^3\}().$$

(Hint: $x^3 = x \cdot x^2$.)

13.7. ALGEBRAIC FUNCTIONS

Any function whose rule involves only the algebraic operations of addition, subtraction, multiplication, division, and raising to rational number powers is an algebraic function. In order to construct such a function by using only polynomials and power series, we need to find power series' expressions for functions of the form $\{t^{p/q}\}()$, where p and q are natural numbers with no common divisor. If we do succeed in constructing such a function, calling it $f()$, the successive derivatives will be given by

$$\begin{aligned} f(t) &= t^{p/q}; \\ f'(t) &= \frac{p}{q} t^{p/q-1}; \\ f''(t) &= \frac{p}{q} \left(\frac{p}{q} - 1 \right) t^{p/q-2}; \\ f'''(t) &= \frac{p}{q} \left(\frac{p}{q} - 1 \right) \left(\frac{p}{q} - 2 \right) t^{p/q-3}; \\ &\vdots \\ f^{(n)}(t) &= \frac{p}{q} \left(\frac{p}{q} - 1 \right) \left(\frac{p}{q} - 2 \right) \dots \left(\frac{p}{q} - n + 1 \right) t^{p/q-n}. \end{aligned}$$

Clearly, with p/q a fixed positive rational number, there will be a natural number n sufficiently large so that all derivatives of order n or greater will not have 0 in their domains. If $a \neq 0$, the Taylor's series, centered at a , for $f()$ will be

$$\begin{aligned} &\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (t-a)^n \\ &= a^{p/q} + \sum_{n=1}^{\infty} \frac{\frac{p}{q} \left(\frac{p}{q} - 1 \right) \left(\frac{p}{q} - 2 \right) \dots \left(\frac{p}{q} - n + 1 \right)}{n! a^n} a^{p/q} (t-a)^n. \end{aligned}$$

The radius of convergence of this series is $|a|$. To get the coefficients of even the first few terms in this series we need to evaluate $a^{p/q}$. It is then most reasonable to center our power series at a number for which we can find the p/q -power. Among the most convenient possible centers will be the numbers of the form $k^{q/p}$, where k is an integer. Thus, we have:

Definition 13.6: Given the natural numbers, p and q , with no common divisor, the function

$$\left\{ \frac{p}{q} \text{ Power} \right\} () \text{ is defined by}$$

- 1) If q is odd, the domain is the set of all numbers.
- 2) If q is even, the domain is the set of all non-negative numbers.

$$3) \left\{ \frac{p}{q} \text{ Power} \right\} (0) = 0.$$

$$4) \text{ If } k \text{ is an integer, and } k^{q/p} \text{ is a number in the domain of } \left\{ \frac{p}{q} \text{ Power} \right\} (), \text{ then}$$

$$\left\{ \frac{p}{q} \text{ Power} \right\} (k^{q/p}) = k.$$

$$5) \text{ If } t \text{ is not a number of the form } k^{q/p}, \text{ where } k \text{ is an integer, let } k_t \text{ be any integer, such that}$$

$$|(k_t)^{q/p} - t| < |k_t|^{q/p},$$

then

$$\left\{ \frac{p}{q} \text{ Power} \right\} (t) = k_t \left[1 + \sum_{n=1}^{\infty} \frac{\frac{p}{q} \left(\frac{p}{q} - 1 \right) \left(\frac{p}{q} - 2 \right) \dots \left(\frac{p}{q} - n + 1 \right)}{n! (k_t^{q/p})^n} (t - k_t^{q/p})^n \right].$$

With this definition, then $\left\{ \frac{p}{q} \text{ Power} \right\} ()$ is the same function as $I^{p/q}()$.

As an example, we find $5^{2/3}$.

5 is not of the form $k^{3/2}$, where k is an integer.

$4^{3/2} = 8$, so we take $k_5 = 4$ (with $k_5^{3/2} = 8$), and our power series is centered at 8.

$$\left\{ \frac{2}{3} \text{ Power} \right\} (t) = 4 \left[1 + \sum_{n=1}^{\infty} \frac{1}{n! 8^n} \left(\frac{2}{3} \right) \left(-\frac{1}{3} \right) \left(-\frac{4}{3} \right) \dots \left(\frac{2}{3} - n + 1 \right) (t - 8)^n \right],$$

$$|t - 8| < 8.$$

$$\left\{ \frac{2}{3} \text{ Power} \right\} (5) = 4 \left[1 + \sum_{n=1}^{\infty} \frac{1}{n! 8^n} (-1)^{n-1} \frac{2(1)(4)(7) \dots (3n-5)}{3^n} (-3)^n \right]$$

$$= 4 \left[1 - \sum_{n=1}^{\infty} \frac{2(1)(4)(7) \dots (3n-5)}{n! 8^n} \right]$$

$$= 4 \left[1 - \frac{2}{1! 8} - \frac{2(1)}{2! 8^2} - \frac{2(1)(4)}{3! 8^3} - \dots \right]$$

$$\approx 4(0.7319) = 2.928.$$

For the derivative, we have

Theorem 13.17:

$$\left\{ \frac{p}{q} \text{ Power} \right\}'() = \left\{ \frac{p}{q} \left\{ \left(\frac{p}{q} - 1 \right) \text{ Power} \right\} \right\} ().$$

Proof:

$$\left\{ \frac{p}{q} \text{Power} \right\} (t) = a^{p/q} \left[1 + \sum_{n=1}^{\infty} \frac{r(r-1) \dots (r-n+1)}{n! a^n} (t-a)^n \right],$$

$$\text{where } a = (k_t)^{q/p}; \quad r = \frac{p}{q}.$$

$$\begin{aligned} \left\{ \frac{p}{q} \text{Power} \right\}'(t) &= a^r \left[\sum_{n=1}^{\infty} \frac{r(r-1) \dots (r-n+1)}{n! a^n} n(t-a)^{n-1} \right] \\ &= a^r \sum_{n=1}^{\infty} r \frac{(r-1)(r-2) \dots (r-1-[n-1]+1)}{(n-1)! a^{n-1} a} (t-a)^{n-1} \\ &= r a^{r-1} \sum_{n=0}^{\infty} \frac{(r-1)(r-2) \dots (r-1-n+1)}{n! a^n} (t-a)^n \\ &= \frac{p}{q} \left\{ \left(\frac{p}{q} - 1 \right) \text{Power} \right\} (t). \end{aligned}$$

Exercises

Use the power series expansion to find approximations to:

1. $7^{2/3}$.
2. $9^{3/4}$.
3. $(125)^{3/2}$.
4. $5^{-2/3}$.
5. $\sqrt{4.1}$.
6. $\sqrt[3]{7.99}$.

13.8. TRIGONOMETRIC FUNCTIONS

When it comes to finding cosines and sines of numbers, those numbers which are integer multiples of 0, $\pi/6$, $\pi/4$, $\pi/3$, and $\pi/2$ are particularly pleasant. For other numbers, some labor is involved. In order to create a power series which will replace the cosine function, we will need,

$$\begin{aligned} f(t) &= \cos t; \\ f'(t) &= -\sin t; \\ f''(t) &= -\cos t; \\ f'''(t) &= \sin t; \\ f^{(4)}(t) &= \cos t; \end{aligned}$$

.

.

.

$$f^{(n)}(t) = f^{(n-4)}(t).$$

For $t = 0$, we get

$$\begin{aligned}f(0) &= 1; \\f'(0) &= 0; \\f''(0) &= -1; \\f'''(0) &= 0; \\f^{(4)}(0) &= 1;\end{aligned}$$

$$f^{(n)}(0) = f^{(n-4)}(0).$$

This means that the Taylor's series, centered at 0, for the cosine function contains only even powers, and we have

Definition 13.7: COSINE () is given by

$$\text{COSINE } (t) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} t^{2n}.$$

This power series converges for all values of t .

In like manner, we get

Definition 13.8: SINE () is defined by

$$\text{SINE } (t) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} t^{2n+1}.$$

This power series converges for all values of t .

We can use these power series to evaluate these functions for any number. For numbers near 0, the series converge moderately rapidly. Each of the remaining trigonometric functions may now be defined in terms of SINE () and COSINE (), exactly as was done in Chapter 10. These functions are in fact indentially the functions $\sin ()$ and $\cos ()$, which were defined in Chapter 10, but the new definitions have the advantage of requiring no geometry, and no concept of arc length.

For the derivative, we have

Theorem 13.18:

$$\text{COSINE}'() = \{-\text{SINE}\}().$$

Proof:

$$\text{COSINE } (t) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} t^{2n}.$$

$$\begin{aligned}
 \{\text{COSINE}\}'(t) &= \sum_{n=1}^{\infty} \frac{(-1)^n}{(2n)!} 2nt^{2n-1} \\
 &= - \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)!} t^{2n-1} \\
 &= - \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} t^{2n+1} \\
 &= \{-\text{SINE}\}(t).
 \end{aligned}$$

Theorem 13.19:

$$\text{SINE}'(t) = -\text{COSINE}(t).$$

Proof:

$$\begin{aligned}
 \text{SINE}(t) &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} t^{2n+1} \\
 \text{SINE}'(t) &= \sum_{n=0}^{\infty} \frac{(-1)^n (2n+1)}{(2n+1)!} t^{2n} \\
 &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} t^{2n} \\
 &= \text{COSINE}(t).
 \end{aligned}$$

Theorem 13.20:

$$\text{COSINE}^2(t) + \text{SINE}^2(t) = 1.$$

Proof:

Let the function $F(t)$ be given by

$$F(t) = \text{COSINE}^2(t) + \text{SINE}^2(t).$$

Then $F'(t)$ is given by

$$\begin{aligned}
 F'(t) &= 2 \text{COSINE}(t) (-\text{SINE}(t)) + 2 \text{SINE}(t) \text{COSINE}(t) \\
 &= 0.
 \end{aligned}$$

Thus, $F(t)$ is a constant function. Further, $F(0) = 1$. Thus, $F(t)$ is the function given by

$$F(t) = 1.$$

In a similar manner, we can find the Taylor's series for the various inverse trigonometric functions. We will not do all of the inverse trigonometric functions here, but we will present one particularly neat series expansion.

Definition 13.9: $\text{ARCTAN } (x)$ is defined by

$$\text{ARCTAN } (x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} x^{2n+1}, \quad |x| < 1.$$

Theorem 13.21:

$$\{\text{ARCTAN}\}'(x) = \frac{1}{1+x^2}, \quad |x| < 1.$$

Proof:

$$\begin{aligned} \{\text{ARCTAN}\}'(x) &= \sum_{n=0}^{\infty} \frac{(-1)^n (2n+1)}{2n+1} x^{2n}, \quad |x| < 1, \\ &= \sum_{n=0}^{\infty} (-x^2)^n, \quad |x| < 1. \\ &= \frac{1}{1+x^2}, \quad |x| < 1. \end{aligned}$$

Exercises

Use the power series expansions given to find:

1. $\cos 0.1$.
2. $\cos 0.2$.
3. $\cos 0.3$.
4. $\sin 0.1$.
5. $\sin 0.2$.
6. $\sin 0.3$.
7. Use the power series expansion for $\frac{1}{\sqrt{1-x^2}}$ to find a power series expansion for the function $\text{arc Sin } (x)$.
8. $\tan \pi/6 = 1/\sqrt{3}$. Use the expansion for $\text{arc Tan } (x)$ given in the definition (13.9) to find an approximation to this value.
9. Find expansions for the sine and cosine functions which are power series centered at $\pi/4$.

13.9. LOGARITHMIC AND EXPONENTIAL FUNCTIONS

As with the trigonometric functions, if we construct functions by means of power series, and those functions and their derivatives have the same properties as the functions $\text{Log } (x)$ and $\exp (x)$, then we have arrived at alternate definitions of the functions $\text{Log } (x)$ and $\exp (x)$, which new definitions involve no operations other than finding the sum of power series.

For the function $\text{Log } ()$, we proceed as follows:

$$\frac{1}{1+x} = \sum_{n=0}^{\infty} (-1)^n x^n, \quad |x| < 1.$$

$$\log(1+x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n+1} x^{n+1}, \quad |x| < 1,$$

$$= \sum_{n=1}^{\infty} \frac{(-1)^{n+1} x^n}{n}, \quad |x| < 1.$$

$$\log t = \log(a + [t-a])$$

$$= \log a + \log \left[1 + \frac{t-a}{a} \right]$$

$$= \log a + \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{na^n} (t-a)^n, \quad |t-a| < |a|.$$

If $a = e^k$, where k is an integer, then $\text{Log } a = k$.

Definition 13.10: The function $\log ()$ is defined by

- 1) The domain of $\log ()$ is the set of all positive numbers.
- 2) If $t = e^k$, where k is an integer, then $\log t = k$.
- 3) If t is not of the form e^k , where k is an integer, let k_t be an integer, such that $|t - e^{k_t}| < e^{k_t}$.

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{ne^{k_t n}} (t - e^{k_t})^n$$

Then

$$\log t = k_t + \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{ne^{k_t n}} (t - e^{k_t})^n$$

From this definition, we find $\log 2$:

$|e - 2| < e$, so that we may take $k_2 = 1$.

$$\begin{aligned} \log 2 &= 1 + \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{ne^n} (2 - e)^n \\ &= 1 - \sum_{n=1}^{\infty} \frac{(e - 2)^n}{ne^n} \\ &= 1 - \sum_{n=1}^{\infty} \frac{1}{n} \left(1 - \frac{2}{e}\right)^n \\ &= 1 - \left(1 - \frac{2}{e}\right) - \frac{1}{2} \left(1 - \frac{2}{e}\right)^2 - \frac{1}{3} \left(1 - \frac{2}{e}\right)^3 - \dots \\ &\approx 1 - (0.26424) - \frac{1}{2} (0.26424)^2 - \frac{1}{3} (0.26424)^3 - \dots \\ &\approx 0.693. \end{aligned}$$

For the derivative, we have

Theorem 13.22:

$$\{\log\}'(t) = \frac{1}{t}, \quad t > 0.$$

Proof:

$$\begin{aligned} \log t &= k_t + \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{ne^{k_t n}} (t - e^{k_t})^n. \\ \log'(t) &= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{e^{k_t n}} (t - e^{k_t})^{n-1} \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{e^{k_t(n+1)}} (t - e^{k_t})^n \\ &= \frac{1}{e^{k_t}} \sum_{n=0}^{\infty} \left(\frac{e^{k_t} - t}{e^{k_t}} \right)^n \\ &= \left(\frac{1}{e^{k_t}} \right) \left(\frac{1}{1 - \frac{e^{k_t} - t}{e^{k_t}}} \right) \\ &= \left(\frac{1}{e^{k_t}} \right) \left(\frac{e^{k_t}}{e^{k_t} - e^{k_t} + t} \right) \\ &= \frac{1}{t}. \end{aligned}$$

With this theorem on the derivative of $\log()$, we can see that $\log()$ and $\text{Log}()$ are precisely the same functions. The logarithmic properties of $\log()$ then follow in the manner given in Chapter 11.

The exponential function, $\exp()$, is given by:

Definition 13.11: $\exp()$ is defined by

$$\exp t = \sum_{n=0}^{\infty} \frac{t^n}{n!}.$$

(This power series converges for all values of t .)

The derivative is given by

Theorem 13.23:

$$\exp'() = \exp().$$

Proof:

$$\exp t = \sum_{n=0}^{\infty} \frac{t^n}{n!}.$$

$$\begin{aligned}
 \exp'(t) &= \sum_{n=1}^{\infty} \frac{n t^{n-1}}{n!} \\
 &= \sum_{n=1}^{\infty} \frac{t^{n-1}}{(n-1)!} \\
 &= \sum_{n=0}^{\infty} \frac{t^n}{n!} \\
 &= \exp t.
 \end{aligned}$$

Since the hyperbolic functions, $\cosh()$ and $\sinh()$, are defined in terms of $\exp()$, we have the following power series expansions for them:

$$\begin{aligned}
 \cosh x &= \frac{1}{2} (e^x + e^{-x}) \\
 &= \frac{1}{2} \left\{ \sum_{n=0}^{\infty} \frac{x^n}{n!} + \sum_{n=0}^{\infty} (-1)^n \frac{x^n}{n!} \right\} \\
 &= \frac{1}{2} \sum_{n=0}^{\infty} \frac{1 + (-1)^n}{n!} x^n \\
 &= \frac{1}{2} \sum_{k=0}^{\infty} \frac{2}{(2k)!} x^{2k} \\
 &= \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!} . \\
 \sinh x &= \frac{1}{2} (e^x - e^{-x}) \\
 &= \frac{1}{2} \left\{ \sum_{n=0}^{\infty} \frac{x^n}{n!} - \sum_{n=0}^{\infty} (-1)^n \frac{x^n}{n!} \right\} \\
 &= \frac{1}{2} \sum_{n=0}^{\infty} \frac{1 - (-1)^n}{n!} x^n \\
 &= \sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!} .
 \end{aligned}$$

Exercises

Find approximations to:

1. $\log 3$.
2. $\log 5$.
3. $\log 9$.
4. $\log (1/10)$.
5. Find a power series expansion for e^t which is centered at 2.

Find approximations to:

6. $e^{0.1}$.
7. $e^{0.2}$.
8. $e^{0.3}$.
9. Find the Taylor's series, centered at 0, for the functions $\cosh ()$ and $\sinh ()$.
10. Find a power series expansion for $\cosh^{-1} (x)$ by finding an anti-derivative of the powers series for the derivative of $\cosh^{-1} ()$.

13.10. CONCLUSION

In Chapters 1 through 7, we developed the calculus of polynomial functions. Chapter 8 introduced the concept of the quotient of two functions and allowed us to extend the operations of calculus to rational functions. Chapter 9 introduced the composite of two functions, and the chain rule. With this machinery, we could perform all the operations of calculus on all algebraic functions.

Trigonometric and inverse trigonometric functions were introduced in Chapter 10, and the basic new concept was that of arc length measured along the circumference of a circle. Chapter 12 used the Fundamental Theorem of Calculus to define $\text{Log} ()$ by means of an integral. The exponential function was then defined in terms of $\text{Log} ()$. Each of the new kinds of functions introduced in Chapters 9 through 11 required a return to the definition of derivative, or—in the case of $\text{Log} ()$ —a use of the Fundamental Theorem of Calculus, in order to differentiate. In this chapter we have given a method for finding all of these non-polynomial functions by means of power series. Since any power series may be differentiated term-by-term within its interval of convergence, we have shown that it was not really necessary to go through a new differentiation theorem each time we came across a new kind of function. Moreover, this tool of power series expansions is so powerful that we can handle every new function which we create by any method whatsoever, provided only that all successive derivatives of this new function exist, and can be evaluated for at least some numbers.



CLASSIFICATION OF THE GRAPHS OF SECOND DEGREE EQUATIONS; CONIC SECTIONS

A.1. DEFINITIONS OF CIRCLE, PARABOLA, ELLIPSE, AND HYPERBOLA

In Chapter 3, we examined the graphs of equations of the form

$$Ax + By + C = 0,$$

and discovered that, no matter what numbers A , B , and C were, the graph of such an equation was always a straight line. In this appendix, we will examine the graphs of equations of the form

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$

and find exactly what kind of geometric entities the graph of such equations could be.

In order that we have some geometric entities other than straight lines, it will be necessary to consider pure Euclidean plane geometry. We have the following definitions:

Definition A.1: A *circle* is the collection of all points in a plane such that the distances of all the points from a given, fixed point are the same. This given, fixed point is called the *center* of the circle.

Definition A.2: A *parabola* is the collection of all points in a plane such that each one of these points is exactly as far from a given fixed point, as it is from a given, fixed line. The given, fixed point is called the *focus* of the parabola and the given, fixed line is called the *directrix* of the parabola.

Definition A.3: An *ellipse* is the collection of all points in a plane such that the sum of the distances from two given, fixed points is always the same. The given, fixed points are called the *foci* of the ellipse.

(Remark: The word “foci” is the plural of the word “focus.”)

Definition A.4: A *hyperbola* is the collection of all points in a plane such that the absolute value of the difference of the distances from two given, fixed points is always the same. The given, fixed points are called the *foci* of the hyperbola.

In order to construct each of these “curves,” we can apply the definitions directly. Thus, to construct a circle, we need only have a means of measuring length. It will be assumed that the reader knows how to construct a circle.

To construct a parabola, we start with a fixed point and a fixed line. If P in Fig. A.1 is the fixed point, and l is the fixed line, then the line through P , perpendicular to l , will contain one point of the parabola, namely the point labeled Q , which is the mid-point of the line segment PR .

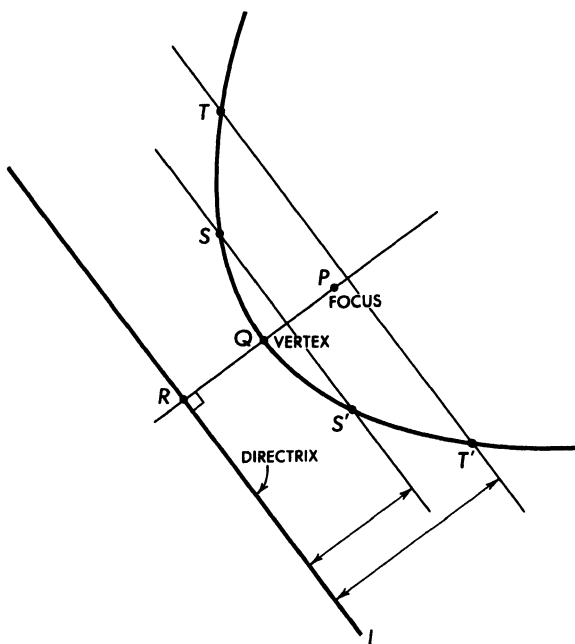


Fig. A.1

If s is a number greater than the distance between P and Q , then there will be two points of the parabola which are s units away from both P and the line l . We can find these two points, S and S' , by drawing a line parallel to l , s units from l in the direction of P , and drawing arcs of a circle centered at P , of radius s . Similarly, if t is a number greater than the distance between P and Q , then drawing a line parallel to l , t units away, and drawing arcs of a circle centered at P , of radius t , will give two more points, T and T' , of the parabola. We can continue to construct more and more pairs of points, and a curve which we draw through these points will give us an approximation to an actual parabola. (This curve will be only an approximation since we can construct only a finite number of points of the parabola

in a finite period of time.) Once we have constructed a parabola, that point which is on the parabola, and on the line through the focus, perpendicular to the directrix, is called the *vertex* of the parabola.

To construct an ellipse directly from the definition, we let the constant mentioned in the definition be the number d . If d is less than the distance between the two fixed points, the ellipse will consist of no points at all. If d is precisely the distance between two fixed points, the ellipse consists of just one point, namely that point midway between the two fixed points. If d is greater than the distance between the two fixed points, called F and F' on Fig. A.2, then for the number p , which is less than d , the intersection of an arc of a circle centered at F' , of radius p , with an arc of a circle centered at F , of radius $d - p$, will be a point on the ellipse, labeled P in Fig. A.2. Similarly, the point R is a distance r units from F' and $(d - r)$ units from F : the point S is at a distance s units from F and $(d - s)$ units from F' ; the point Q is at a distance q units from F' and $(d - q)$ units from F . The mid-point of the line segment joining the two foci is called the *center* of the ellipse (labeled C on Fig. A.2). The intersections of the line through the foci with the ellipse itself are called the *vertices* (plural of vertex) of the ellipse (V and V' on Fig. A.2). The line segment joining the two vertices (V and V') is called the *major axis* of the ellipse. If we draw a line through the center of an ellipse, perpendicular to the major axis, the segment of this line inside the ellipse (the line segment joining U and U' on Fig. A.2) is called the *minor axis* of the ellipse. It should be noted that the major axis always passes through the foci, and is longer than the minor axis. Finally, since the point V is on the ellipse, the sum of the distances between V and F and V and F' is the constant mentioned in

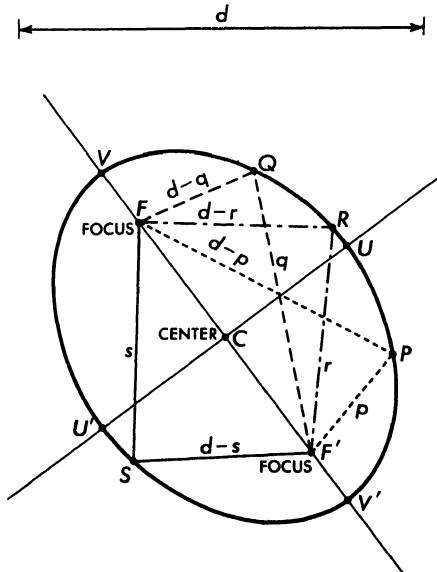


Fig. A.2

the definition, and since the distance between V and F is precisely the same as the distance between V' and F' , the constant mentioned in the definition must be the length of the major axis of the ellipse.

In the construction of a hyperbola (Fig. A.3), we start with the two fixed points, F and F' . The line passing through these points is called the *transverse axis* of the hyperbola. The point C , midway between the two foci, F and F' , is called the *center* of the hyperbola. If d is the constant

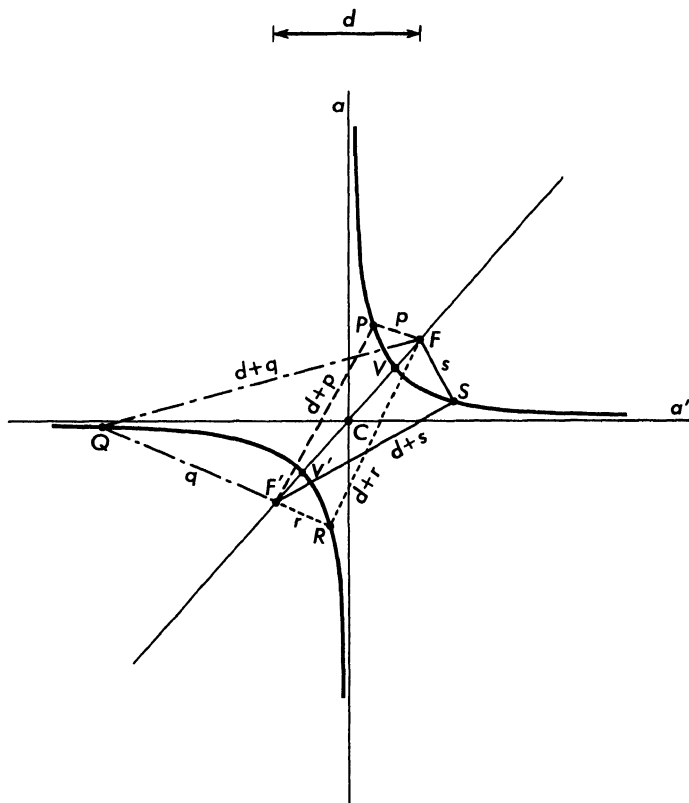


Fig. A.3

mentioned in the definition of a hyperbola, then every point on this hyperbola must have the property that its distance from one focus is d units more than its distance from the other focus. On Fig. A.3, the point P is p units from F and $(d + p)$ units from F' ; the point Q is q units from F' and $(q + d)$ units from F ; the point R is r units from F' and $(d + r)$ units from F ; the point S is s units from F and $(d + s)$ units from F' . The hyperbola itself will then have two branches. The points, V and V' , at which the hyperbola intersects the transverse axis, are called the *vertices* of the hyperbola. Points on the hyperbola which are far from the center actually are all close to two fixed lines, a and a' , which have their intersection at the center C . These lines are called the *asymptotes* of the hyperbola.

Exercises

1. Construct an ellipse whose foci are 4 units apart, and in which the sum of the distances from any point to the two foci is 10 units.
2. Construct a parabola whose focus is 6 units from its directrix.
3. Construct a hyperbola whose foci are 10 units apart, in which the absolute value of the difference of the distances from any point to the two foci is always 6 units.

A.2. ROTATION PRINCIPLE

In this section, we give a basic principle of analytic geometry which will allow us to investigate the graph of

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$

merely by investigating the graph of the simpler equation

$$ax^2 + cy^2 + dx + ey + f = 0.$$

Rotation principle:

If $f(x, y)$ is an algebraic expression involving the two variables, x and y ; if $g(r, s)$ is an algebraic expression involving the two variables, r and s ; if the variables r and s are given by

$$\begin{aligned} r &= \frac{mx - y}{1 + m^2}, \\ s &= \frac{x + my}{1 + m^2}, \end{aligned}$$

where m is a constant, not zero; if the algebraic expressions $f(x, y)$ and $g(r, s)$ are identical; then the graph of the equation $f(x, y) = 0$ is precisely that set of points which results from rotating the graph of $g(x, y) = 0$ counterclockwise about the origin until the x -axis coincides with the line whose equation is $mx - y = 0$, and the y -axis coincides with the line whose equation is $x + my = 0$.

As an example of the application of this principle, the equation

$$(3x + 4y) = 2(4x - 3y)^2$$

is equivalent to

$$(3x + 4y) - 2(4x - 3y)^2 = 0,$$

which, in turn, is equivalent to

$$\frac{3x + 4y}{5} - 10 \left[\frac{4x - 3y}{5} \right]^2 = 0$$

Then, if we let

$$f(x, y) = \left(\frac{3x + 4y}{5} \right)^2 - 10 \left(\frac{4x - 3y}{5} \right)^2$$

$$r = \frac{4x - 3y}{5}, \quad s = \frac{3x + 4y}{5}$$

and

$$g(x, y) = y - 10x^2,$$

then we have

$$g(r, s) \text{ is identical with } f(x, y).$$

The rotation principle then says that the graph of

$$(3x + 4y) = 2(4x - 3y)^2$$

will have the same relation to the lines $4x - 3y = 0$ and $3x + 4y = 0$ as the graph of

$$y = 10x^2$$

has to the coordinate axes.

The graph of $y = 10x^2$ appears in Fig. A.4. The other part of Fig. A.4 shows what the graph of $(3x + 4y) = 2(4x - 3y)^2$ must be. It should be noted that the rotation principle allows us to recognize that the graph of the complicated equation is merely the result of rotating about the origin the graph of the simple one.

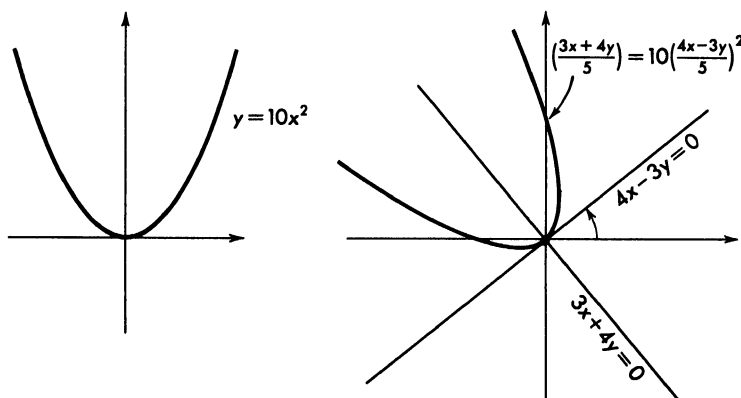


Fig. A.4

We will use this rotation principle to prove the following theorem:

Theorem A.1:

The graph of the equation

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$

is the result of rotating the graph of a new equation

$$ax^2 + cy^2 + dx + ey + f = 0$$

about the origin.

Proof:

We need only find the non-zero number m , such that $Ax^2 + Bxy + Cy^2 + Dx + Ey + F$ is the expression $f(x, y)$ and $ax^2 + cy^2 + dx + ey + f$ is the expression $g(r, s)$ of the rotation principle.

Thus, we need only find the non-zero number m , such that the expressions $f(x, y)$ and

$$a \left(\frac{x + my}{\sqrt{m^2 + 1}} \right)^2 + c \left(\frac{mx - y}{\sqrt{m^2 + 1}} \right)^2 + d \left(\frac{x + my}{\sqrt{m^2 + 1}} \right) + e \left(\frac{mx - y}{\sqrt{m^2 + 1}} \right) + f$$

are identically the same. If we call this last expression $g(r, s)$, then we have:

$$\begin{aligned} g(r, s) &= \frac{a}{m^2 + 1} (x^2 + 2mxy + m^2y^2) + \frac{c}{m^2 + 1} (m^2x^2 - 2mxy + y^2) \\ &\quad + \frac{d}{\sqrt{m^2 + 1}} (x + my) + \frac{e}{\sqrt{m^2 + 1}} (mx - y) + f \\ &= \frac{a + m^2c}{m^2 + 1} x^2 + \frac{2m(a - c)}{m^2 + 1} xy + \frac{am^2 + c}{m^2 + 1} y^2 \\ &\quad + \frac{d + me}{\sqrt{m^2 + 1}} x + \frac{md - e}{\sqrt{m^2 + 1}} y + f \end{aligned}$$

We then need only choose the number m to be such that the coefficients of the various expressions involving x^2 , xy , y^2 , x , and y in $g(r, s)$ are exactly the same as the coefficients of the same terms in $f(x, y)$. Since this can always be done, then the rotation principle will always apply.

Exercises

1. Solve the equations:

$$\begin{aligned} \frac{a + m^2c}{m^2 + 1} &= A; & \frac{2m(a - c)}{m^2 + 1} &= B; \\ \frac{am^2 + c}{m^2 + 1} &= C; & \frac{d + me}{\sqrt{m^2 + 1}} &= D; \\ \frac{md - e}{\sqrt{m^2 + 1}} &= E; \end{aligned}$$

for m , a , c , d , and e in terms of A , B , C , D , and E .

2. Apply the rotation principle to the equation

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

A.3. JUSTIFICATION OF ROTATION PRINCIPLE

To justify the rotation principle, suppose we have a point (p, q) on the graph of

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0.$$

If we draw the lines through the origin making an angle of θ° with the x and y -axes, these lines will have equations

$$-x \sin \theta + y \cos \theta = 0$$

and

$$x \cos \theta + y \sin \theta = 0.$$

By Section 6 of Chapter 3, the distance between (p, q) and the first line is

$$-p \sin \theta + q \cos \theta$$

and the distance between (p, q) and the second line is

$$p \cos \theta + q \sin \theta.$$

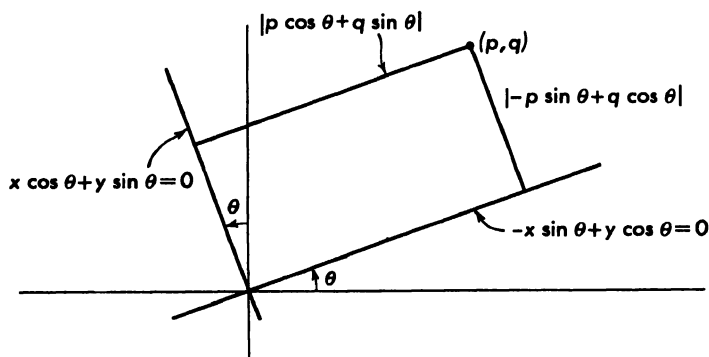


Fig. A.5

Let

$$\begin{aligned} s &= p \cos \theta + q \sin \theta, \\ t &= -p \sin \theta + q \cos \theta. \end{aligned}$$

Solving these equations for p and q in terms of s and t , we have

$$\begin{aligned} p &= s \cos \theta - t \sin \theta, \\ q &= s \sin \theta + t \cos \theta. \end{aligned}$$

(p, q) is on the graph of $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$, so that

$$Ap^2 + Bpq + Cq^2 + Dp + Eq + F = 0;$$

$$A(s \cos \theta - t \sin \theta)^2 + B(s \cos \theta - t \sin \theta)(s \sin \theta + t \cos \theta)$$

$$+ C(s \sin \theta + t \cos \theta)^2 + D(s \cos \theta - t \sin \theta)$$

$$+ E(s \sin \theta + t \cos \theta) + F = 0.$$

$$(A \cos^2 \theta + B \sin \theta \cos \theta + C \sin^2 \theta) s^2$$

$$+ ((-A + C) 2 \sin \theta \cos \theta + B(\cos^2 \theta - \sin^2 \theta)) st$$

$$+ (A \sin^2 \theta - B \sin \theta \cos \theta + C \cos^2 \theta) t^2$$

$$+ (D \cos \theta + E \sin \theta) s$$

$$+ (-D \sin \theta + E \cos \theta) t + F = 0.$$

If θ is chosen so that

$$(-A + C) 2 \sin \theta \cos \theta + B(\cos^2 \theta - \sin^2 \theta) = 0,$$

i.e.

$$-(A - C) \sin 2\theta + B \cos 2\theta = 0,$$

then the term in st becomes 0. For $A \neq C$, we need

$$\tan 2\theta = \frac{B}{A - C},$$

and for $A = C$, we need $\theta = \pi/4$.

Now, for this choice of θ , we let a be the coefficient of s^2 , c be the coefficient of t^2 , d the coefficient of s , e the coefficient of t , and $f = F$. We then have

$$as^2 + ct^2 + ds + et + f = 0,$$

or, the point with coordinates (s, t) is on the graph of

$$ax^2 + cy^2 + dx + ey + f = 0.$$

If we do rotate the graph of this equation counterclockwise about the origin, through the chosen angle, we will get precisely the graph of

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0.$$

A.4. TRANSLATION PRINCIPLE

As a consequence of the rotation principle, we need only consider, for our classification purpose, equations of the form

$$Ax^2 + Cy^2 + Dx + Ey + F = 0.$$

In this section, we will enunciate a second principle which will simplify even further our task;

Translation principle:

If $f(x, y)$ is an algebraic expression involving the two variables x and y ; if $g(u, v)$ is an algebraic expression involving the two variables u and v ; if the variables u and v are defined by

$$\begin{aligned} u &= x - a \\ v &= y - b, \end{aligned}$$

where a and b are constants; if the algebraic expression $f(x, y)$ and $g(u, v)$ are identically the same; then the graph of the equation $f(x, y) = 0$ bears the same relation to the lines with equations $y - b = 0$ and $x - a = 0$ as the graph of $g(x, y) = 0$ does to the coordinate axes.

As an example of the application of this principle, the equation:

$$y + 1 = (x - 3)^2$$

bears the same relation to the lines with equations $x = 3$ and $y = -1$ as the graph of $y = x^2$ bears to the axes. Fig. A.6 shows then how we get the graph $y + 1 = (x - 3)^2$.

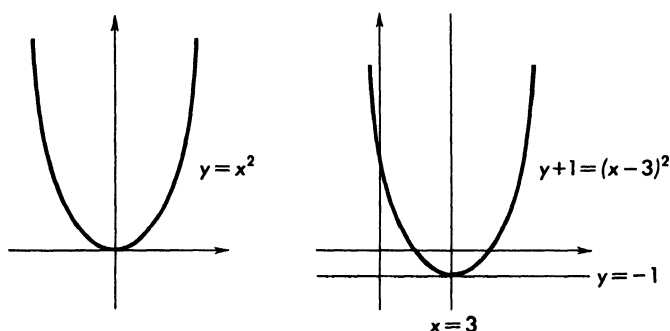


Fig. A.6

We will use the translation principle to prove the following theorems:

Theorem A.2:

The graph of the equation

$$Ax^2 + Cy^2 + Dx + Ey + F = 0,$$

in which A and C are both not zero, bears the same relation to some pair of vertical and horizontal lines as the graph of a new equation

$$ax^2 + cy^2 + f = 0$$

does to the coordinate axes.

Proof:

The equation

$$Ax^2 + Cy^2 + Dx + Ey + F = 0$$

is equivalent to the equation

$$A \left(x^2 + \frac{D}{A}x \right) + c \left(y^2 + \frac{E}{c}y \right) + F = 0.$$

Using the technique known as “completing the square,” we find that this last equation is equivalent to

$$A \left(x^2 + \frac{D}{A}x + \frac{D^2}{4A^2} \right) + C \left(y^2 + \frac{E}{C}y + \frac{E^2}{4C^2} \right) + F - \frac{D^2}{4A^2} - \frac{E^2}{4C^2} = 0,$$

which is

$$A \left(x + \frac{D}{2A} \right)^2 + C \left(y + \frac{E}{2C} \right)^2 + \left(F - \frac{D^2}{4A^2} - \frac{E^2}{4C^2} \right) = 0,$$

an equation precisely in the form required by the translation principle. Consequently, the graph of this equation bears the same relation to the lines $y + E/2C = 0$ and $x + D/2A = 0$ as the graph of

$$Ax^2 + Cy^2 + f = 0$$

(where f is the constant term above) does to the coordinate axes.

Theorem A.2 then tells us that, whenever both an x^2 and y^2 are actually present in an equation, the graph of this equation can be described merely by considering an appropriate equation of the form

$$Ax^2 + Cy^2 + F = 0,$$

where the constants A and C are both not zero.

For the case in which one of the squared terms is missing, we have

Theorem A.3:

The graph of the equation

$$Ax^2 + Dx + Ey + F = 0,$$

in which A is not zero, bears the same relation to some pair of vertical and horizontal lines as the graph of a new equation

$$x^2 = ky$$

does to the coordinate axes.

Proof:

The equation

$$Ax^2 + Dx + Ey + F = 0$$

is equivalent to the equation

$$A \left(x^2 + \frac{D}{A} x \right) = -Ey - F.$$

Using the technique known as completing the square again, we find this last equation equivalent to

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

which is an equation precisely in the form required by the translation principle. Consequently, the graph of this equation bears the same relation to the lines $y + F/E - D^2/4AE = 0$ and $x + D/2A = 0$ as the graph of $x^2 = ky$ (where $k = -E/A$) does to the coordinate axes.

In precisely the same manner, we can prove

Theorem A.4:

The graph of the equation

$$Cy^2 + Dx + Ey + F = 0$$

in which C is not zero, bears the same relation to some pair of vertical and horizontal lines as the graph of a new equation

$$y^2 = Kx$$

does to the coordinate axes.

The proof is left as an exercise for the reader.

Exercise

1. Prove Theorem A.4.

A.5. REDUCTION OF THE CLASSIFICATION PROBLEM

The successive use of the rotation and translation principles has shown us that every equation of the form

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0,$$

in which A , B , and C are not all 0, has a graph which is precisely the same kind of geometric entity as the graph of one of the equations:

$$\begin{aligned} ax^2 + cy^2 + f &= 0 & (a \neq 0, \ c \neq 0) \\ x^2 &= ky & (k \neq 0) \\ y^2 &= Kx & (K \neq 0). \end{aligned}$$

Our problem of classification is then reduced to the relatively simple task of classifying the graphs of these three equations.

Let us first dispose of the equation $x^2 = ky$.

Theorem A.5:

The graph of $x^2 = ky$ is a parabola, if $k \neq 0$.

Proof:

The following equations are all equivalent:

$$\begin{aligned} x^2 &= ky, \\ x^2 &= \frac{k}{2}y + \frac{k}{2}y, \\ x^2 - \frac{k}{2}y &= \frac{k}{2}y, \\ x^2 + y^2 - \frac{k}{2}y + \frac{k^2}{16} &= y^2 + \frac{k}{2}y + \frac{k^2}{16}, \\ x^2 + \left(y - \frac{k}{4}\right)^2 &= \left(y + \frac{k}{4}\right)^2, \\ \sqrt{(x-0)^2 + \left(y - \frac{k}{4}\right)^2} &= \left|y - \left(-\frac{k}{4}\right)\right|. \end{aligned}$$

The last of these equations will be satisfied by those points, and only those points whose distance from the point $(0, k/4)$ is exactly the same as the distance from the line $y = -k/4$. By the definition (A.2), this collection of points is a parabola.

Next we have

Theorem A.6:

The graph of $y^2 = Kx$ is a parabola, if $K \neq 0$.

The proof is analogous to the proof of the preceding theorem and is left as an exercise for the reader.

For the remaining equation, $ax^2 + cy^2 + f = 0$, in which neither a nor c is 0, we must examine the various possibilities of the signs and relative sizes of the constants a , c , and f . In order to do this, we will use a notation which allows non-negative numbers to be immediately identified. All possible circumstances are then given by the equations:

$$A^2x^2 + C^2y^2 = F,$$

where F can be positive, negative, or zero,

$$\begin{aligned} A^2x^2 - C^2y^2 &= F^2, \\ -A^2x^2 + C^2y^2 &= F^2. \end{aligned}$$

We have the elementary cases:

Theorem A.7:

The graph of $A^2x^2 + C^2y^2 = F$, where F is negative, consists of no points whatsoever.

Proof:

No numbers will make the left-hand side of this equation negative.

Theorem A.8:

The graph of $A^2x^2 + C^2y^2 = 0$ consists of the one point $(0, 0)$.

Proof:

Only these numbers can satisfy the equation.

Theorem A.9:

The graph of $A^2x^2 - C^2y^2 = 0$ or of $-A^2x^2 + C^2y^2 = 0$ consists of two intersecting straight lines.

Proof:

Both equations are equivalent to

$$(Ax + Cy)(Ax - Cy) = 0.$$

Now for the significant cases:

If, in the equation $A^2x^2 + C^2y^2 = F$, F is positive (for we have already considered the cases in which F is either negative or zero), then this equation is equivalent to

$$\frac{A^2}{F}x^2 + \frac{C^2}{F}y^2 = 1,$$

which, since $A \neq 0$ and $C \neq 0$, is in turn equivalent to

$$\frac{x^2}{F/A^2} + \frac{y^2}{F/C^2} = 1.$$

Since the two denominators are positive, we need to consider equations of the form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

Theorem A.10:

If $a = b > 0$, the graph of

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

is a circle.

Proof:

If $a = b$, let this positive number be r . Then the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

is equivalent to the equation

$$x^2 + y^2 = r^2.$$

This last equation is equivalent to

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

The graph of this last equation is the collection of all points in the plane which are r units from the point with coordinates $(0, 0)$. By the definition (A.1), this graph is a circle.

If we do not have the same denominators in the equation under investigation, then:

Theorem A.11:

If $a \neq b$, and both are positive, the graph of

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

is an ellipse.

Proof:

Case I: $a > b$.

The following equations are equivalent:

$$\begin{aligned}\frac{x^2}{a^2} + \frac{y^2}{b^2} &= 1, \\ b^2 x^2 + a^2 y^2 &= a^2 b^2, \\ (a^2 + b^2 - a^2) x^2 + a^2 y^2 &= a^2(b^2 - a^2 + a^2), \\ a^2 x^2 + (b^2 - a^2) x^2 + a^2 y^2 &= a^2(b^2 - a^2) + a^4, \\ a^2 x^2 + a^2(a^2 - b^2) + a^2 y^2 &= (a^2 - b^2) x^2 + a^4, \\ a^2[x^2 + (a^2 - b^2)] + a^2 y^2 &= (a^2 - b^2) x^2 + a^4, \\ a^2[x^2 - 2\sqrt{a^2 - b^2}x + (a^2 - b^2)] + a^2 y^2 &= (a^2 - b^2) x^2 - 2a^2\sqrt{a^2 - b^2}x + a^4, \\ a^2\{(x - \sqrt{a^2 - b^2})^2 + y^2\} &= (\sqrt{a^2 - b^2}x - a^2)^2.\end{aligned}$$

Since, in the original equation, the range of the variable x is the interval: all t , $|t| \leq a$, we must have

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

(i.e. $a^2 - b^2 < a^2$ implies $\sqrt{a^2 - b^2} < a$, so that $\sqrt{a^2 - b^2}x < a^2$ and the right-hand side must be negative.)

This last equation is equivalent to

$$\begin{aligned}4\sqrt{a^2 - b^2}x - 4a^2 &= -4a\sqrt{(x - \sqrt{a^2 - b^2})^2 + y^2}, \\ 2\sqrt{a^2 - b^2}x &= 4a^2 - 2\sqrt{a^2 - b^2}x - 4a\sqrt{(x - \sqrt{a^2 - b^2})^2 + y^2}, \\ x^2 + 2\sqrt{a^2 - b^2}x + a^2 - b^2 + y^2 &= 4a^2 + x^2 - 2\sqrt{a^2 - b^2}x + a^2 - b^2 \\ &\quad + y^2 - 4a\sqrt{(x - \sqrt{a^2 - b^2})^2 + y^2}, \\ (x + \sqrt{a^2 - b^2})^2 + y^2 &= 4a^2 + (x - \sqrt{a^2 - b^2})^2 + y^2 \\ &\quad - 4a\sqrt{(x - \sqrt{a^2 - b^2})^2 + y^2}, \\ (x + \sqrt{a^2 - b^2})^2 + y^2 &= \left\{2a - \sqrt{(x - \sqrt{a^2 - b^2})^2 + y^2}\right\}^2, \\ \sqrt{(x + \sqrt{a^2 - b^2})^2 + y^2} &= 2a - \sqrt{(x - \sqrt{a^2 - b^2})^2 + y^2}\end{aligned}$$

(the last occurring because of the restrictions on the ranges of x and y),

$$\sqrt{(x + \sqrt{a^2 - b^2})^2 + (y - 0)^2} + \sqrt{(x - \sqrt{a^2 - b^2})^2 + (y - 0)^2} = 2a.$$

The graph of the last equation is the collection of all the points in the plane which have the property that the sum of the distances to $(\sqrt{a^2 - b^2}, 0)$ and $(-\sqrt{a^2 - b^2}, 0)$ is always the constant $2a$. By the definition (A.3), this graph is an ellipse.

Case II: $b > a$.

If, in each and every step in case I, we interchange a and b and interchange x and y , the reasoning remains valid, and we find that the graph is the collection of all the points in the plane which have the property that the sum of the distances to $(0, \sqrt{b^2 - a^2})$ and $(0, -\sqrt{b^2 - a^2})$ is always the constant $2b$.

For the remaining two equations, we have

Theorem A.12:

The graph of the equation

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

is a hyperbola.

Proof:

Let c be the number $\sqrt{a^2 + b^2}$.

The hyperbola with foci $(c, 0)$ and $(-c, 0)$, for which the constant mentioned in the definition (A.4) is the number $2a$, is the graph of the equation

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

This equation is equivalent to the following equations:

$$\sqrt{(x+c)^2 + y^2} - \sqrt{(x-c)^2 + y^2} = \pm 2a,$$

$$\sqrt{(x+c)^2 + y^2} = \sqrt{(x-c)^2 + y^2} \pm 2a,$$

$$(x+c)^2 + y^2 = (x-c)^2 + y^2 + 4a^2 \pm 4a\sqrt{(x-c)^2 + y^2},$$

$$x^2 + 2cx + c^2 + y^2 = x^2 - 2cx + c^2 + y^2 + 4a^2 \pm 4a\sqrt{(x-c)^2 + y^2},$$

$$4cx - 4a^2 = \pm 4a\sqrt{(x-c)^2 + y^2},$$

$$cx - a^2 = \pm a\sqrt{(x-c)^2 + y^2},$$

$$c^2x^2 - 2a^2cx + a^4 = a^2(x-c)^2 + a^2y^2,$$

$$c^2x^2 - 2a^2cx + a^4 = a^2x^2 - 2a^2cx + a^2c^2 + a^2y^2,$$

$$c^2x^2 + a^4 = a^2x^2 + a^2c^2 + a^2y^2,$$

$$(c^2 - a^2)x^2 - a^2y^2 = a^2c^2 - a^4,$$

$$(c^2 - a^2)x^2 - a^2y^2 = a^2(c^2 - a^2)$$

$$\frac{x^2}{a^2} - \frac{y^2}{c^2 - a^2} = 1,$$

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \quad \text{since } a^2 + b^2 = c^2$$

In the proofs of Theorems A.5, A.11, and A.12, it is certainly most reasonable to ask how one should know which algebraic manipulations to perform in order to "come out with the right answer." Let us look at each.

The graph of the equations $x^2 = ky$ is certainly a collection of points none of whose y -coordinates can make the right-hand side a negative number. If $k > 0$, then the graph lies entirely above the x -axis. Further this graph is symmetric with respect to the y -axis. Thus, if the graph is actually to be a parabola, the origin must be the vertex of this parabola.

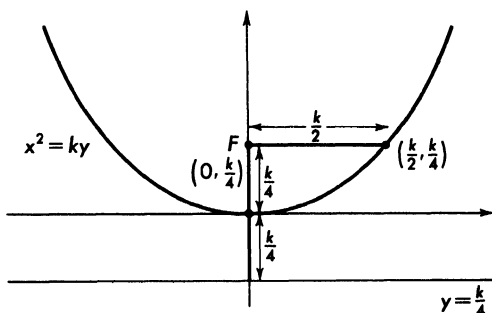


Fig. A.7

In any parabola, a line segment perpendicular to the axis, from the focus of the parabola, must be twice as long as the distance between the focus and vertex. If the graph of $x^2 = ky$ is to be a parabola with vertex at the origin, then there must be a point on this graph whose x -coordinate is twice its y -coordinate. This point can only be $(k/2, k/4)$. Then, if this graph is to be a parabola, the point $(0, k/4)$ must be the focus and the line $y = -k/4$ must be the directrix.

From Theorem A.11, if $a > b$, the graph of

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

will be a "closed curve" with intercepts at the points $(a, 0)$, $(-a, 0)$, $(0, b)$, and $(0, -b)$. If this graph is to be an ellipse, then the horizontal line segment between $(-a, 0)$ and $(a, 0)$ should be its major axis, so that

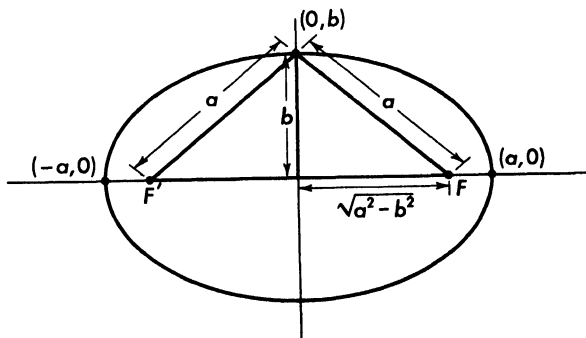


Fig. A.8

the constant mentioned in the definition of an ellipse must be the number $2a$. If $(0, b)$ is to be on this ellipse, then the foci must be located so that the distance from $(0, b)$ to each focus is precisely a units. This implies that the foci are at $(\sqrt{a^2 - b^2}, 0)$ and $(-\sqrt{a^2 - b^2}, 0)$.

A similar examination of the graph of the equation in Theorem A.12 shows that for this graph to be a hyperbola the distance between the center and focus must be $\sqrt{a^2 + b^2}$. If, in the statement and proof of this theorem, we interchange x and y , we can prove the final theorem:

Theorem A.13:

The graph of the equation

$$-\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

is a hyperbola.

Exercise

1. Prove Theorem A.13.

A.6. CONCLUSION AND SUMMARY

In this appendix, we have derived a complete classification of equations, involving the variables x and y , of the form

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0.$$

We noted first, by the rotation principle, that we may, without loss of generality, assume that $B = 0$. For equations of the form

$$Ax^2 + Cy^2 + Dx + Ey + F = 0,$$

we have:

$A = 0, C \neq 0$	parabola
$A \neq 0, C = 0$	parabola
$A \neq 0, C \neq 0, A = C$	circle, or one point, or no points
$A \neq C, AC > 0$	ellipse, or one point, or no points
$A \neq C, AC < 0$	hyperbola, or two intersecting lines

If we consider a generalized right circular cone, which is the locus of all points in a three-dimensional space swept out by a given line through a fixed point and a fixed circle, the point being not in the plane of the circle

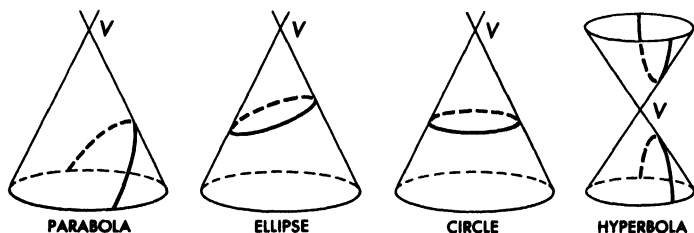


Fig. A.9

and the center of the circle being directly above or below the fixed point (Fig. A.9), then every one of these graphs can be found by intersecting this generalized cone by a plane. If the cutting plane is parallel to one position of the generating line, then the intersection will be a parabola. If the cutting plane cuts both parts of the generalized cone, the intersection will be a hyperbola. If the cutting plane cuts only one part of the generalized cone, the intersection will be either an ellipse or a circle. If the cutting plane passes through the vertex of the cone, the intersection will be either one point or a pair of intersecting lines.

ANSWERS TO ODD-NUMBERED EXERCISES

Section 1.1, PAGE 4

3. Look for empty seats, or standees, or both.
7. Count.
9. $2 \longleftrightarrow 7$; $12 \longleftrightarrow 17$;
 $22 \longleftrightarrow 27$; etc.
11. Two senators per state.

Section 1.3, PAGE 10

7. $[(3)(4)]6$; $[(3)(6)]4$; $[(4)(6)]3$;
 $[(4)(3)]6$; $[(6)(3)]4$; $[(6)(4)]3$.
13. $12 < 20$.
15. No.

Section 1.4, PAGE 16

1. $1/2$ is not included in the definition of integer.
3. $-q$.
5. 2.

Section 1.5, PAGE 18

3. $-3/2$.
5. $7/8 = 77/88$; $9/11 = 72/88$;
 $72 < 77$.

Section 1.6, PAGE 20

7. Yes. $3 = 3/1$; $5 = 5/1$.

Section 1.7, PAGE 22

5. False.
7. True.
9. True.
11. True.
13. $a - b = 0$.

Section 1.8, PAGE 24

1. a) $6/5$; b) $39/20$; c) $65/21$.
3. a) $3/4 < 9/2$.
b) $-9 < 3$.
c) $-2 < -1/2$.
7. Neither.
9. Smallest: 1; no largest.
11. Smallest: 0; no largest.
13. Largest: 3; no smallest.
15. Neither.
17. Neither.

Section 2.1, PAGE 34

1. All numbers.
3. All numbers ≥ 4 .
5. All numbers ≥ 5 .
7. All numbers ≥ -3 .
9. All numbers ≤ 1 .
11. All numbers other than 3.
13. All numbers; $r < 2$.
15. All numbers; $t < 3$.
17. All numbers; $-2 < v < 2$.
19. All numbers; $-2 < x < 2$.

Section 2.2, PAGE 40

1. $=$; yes; true.
3. $=$; yes; false.
5. Are; yes; false.
7. $t^2 = 3$.
9. $1/t = 1/t$.
11. $1/(t+1) + 1/(t-3) =$
 $1/(t+1) + 1/(t-3)$.
13. $t/t = 2t/t$.

15. a) $\{1/3\}$.
 b) $\{-3/2\}$.
 c) $\{3/40\}$.
 d) $\{(\sqrt{3} - 1)/\sqrt{5}\}$.

17. $\{1\}$.
 19. None.
 21. None.
 23. None.

Section 2.3, PAGE 44

1. They have the same solution.
 3. $x > 5/4$.
 5. $x > 3/2$.
 7. $x < 1/4$.
 9. $x < -9/2$.
 11. $x < 4/3$.
 13. $x < 6/5$.
 15. $x > 3/2$.
 17. $x > 6/5$.
 19. $x > 9/5$.
 21. $x < 1$ or $x > 2$.

Section 2.4, PAGE 51

1. $-1 < x < 3$.
 3. $-1 < x < 4$.
 5. $x < -1$ or $x > 5/2$.
 7. None.
 9. None.
 11. None.
 13. $-1 < x < 3$.
 15. $ax^2 + bx + c$ is the negative of the sum of two squares.
 17. $ax^2 + bx + c = a(x - b/2a)^2$.

Section 2.5, PAGE 55

1. $x < -1$ or $-1/2 < x < 3$.
 3. $-1/3 < x < 4/3$ or $x > 2$.
 5. $x \neq -3$.
 7. $-3 < x < -2$ or $2 < x < 3$.
 9. $x < 5$.
 11. $x < 0$ or $1/3 < x < 1$.

Section 2.6, PAGE 58

5. $0 \leq x^2 \leq 16$.
 7. $-4 < x < 0$.
 9. $19/5 < x < 21/5$.
 11. $3.998 < x < 4.002$.
 13. $4 - e/5 < x < 4 + e/5$.

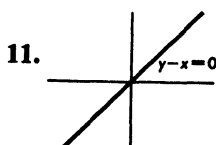
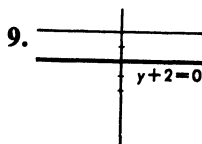
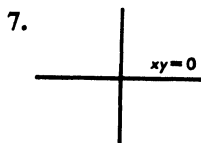
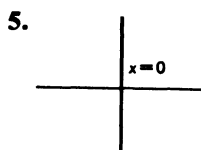
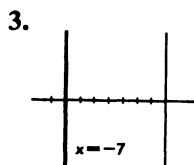
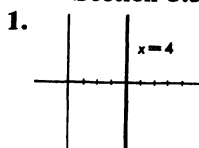
Section 3.1, PAGE 62

3. A vertical line 4 units to the right of the y -axis.
 5. The line through the origin and $(1, 1)$.
 7. (y -coordinate of top point)
 — (y -coordinate of bottom point).

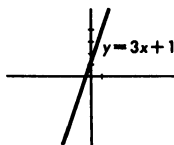
Section 3.2, PAGE 65

3. $\sqrt{17}$.
 5. $\sqrt{29}$.
 7. $\sqrt{10}$.
 9. $\sqrt{97}/2$.
 11. 3.

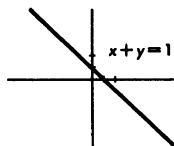
Section 3.3, PAGE 68



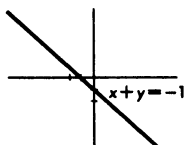
13.



15.



17.

19. $y = 2$.21. $y = -4$.23. $x = 0$.25. $y = x$.**Section 3.4, PAGE 77**

1. a) No points.
b) No points.
c) A vertical line, 1 unit to the right of the y -axis.
d) A horizontal line, 1 unit down from the x -axis.
e) A vertical line, 3 units to the left of the y -axis.
f) A horizontal line, 4 units up from the x -axis.
g) A vertical line, $5/2$ units to the left of the y -axis.
h) A horizontal line, $4/3$ units up from the x -axis.
3. $(-400, -64)$, $(-50, -92/3)$, $(-10, -4)$, $(-1, -2)$, $(0, 4/3)$, $(1, 10/3)$, $(2, 4)$, $(3, 14/3)$, $(10, 28/3)$, $(100, 208/3)$, $(1000, 2008/3)$.
7. a) $-2/3$.
b) $-1/2$.
c) 0.
d) None.
9. a) $4x + y = -1$.
b) $3x - 4y = -13$.

c) $y = 3$.d) $2x + 3y = 7$.e) $3x - 2y = -9$.f) $\sqrt{3}x - y = 3 - \sqrt{3}$.11. a) $1/3$. h) $3/4$.b) 4. i) $3/4$.c) -4 . j) $3/4$.d) -2 . k) $3/4$.e) 2. l) $\sqrt{2}$.f) $3/4$. m) $-1/\sqrt{3}$.g) $3/4$. n) $-\sqrt{5/7}$.**Section 3.5, PAGE 83**5. a) $4x + 3y = -9$.b) $x - 2y = -5$.c) $y = 1$.d) $x = -3$.e) $x - y = \frac{1}{4}$.7. $2/5 = 2/5$; $5/3 = 5/3$.**Section 3.6, PAGE 87**3. For such a point, $d = 0$.

5. a) 13.

b) 13.

c) $60/13$.

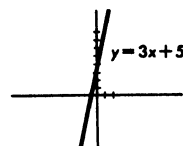
d) 0.

Section 3.7, PAGE 93

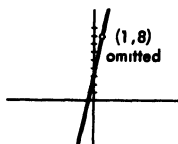
1. a) First and third.
b) First and fourth.
c) First and fourth.
d) All four.
e) First and third.
f) First.
3. a) Neither.
b) Symmetric to x -axis.
c) Symmetric to y -axis.
d) Symmetric to y -axis.
e) Symmetric to y -axis.
f) Symmetric to both axes.
g) Symmetric to x -axis.

Section 3.8, PAGE 98

1.



3.



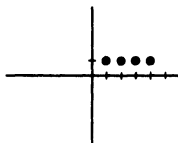
7. $4 - \sqrt{17} < h < \sqrt{19} - 4$, or any smaller interval.

Section 4.1, PAGE 104

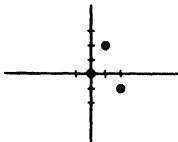
1. A function.
3. Not a function.
5. Not a function.
7. Not a function.
9. Not a function.
11. $1 < y < 4$.
13. Coordinates of those points on the graph of $y = x^2$ which are between $(-1, 1)$ and $(1, 1)$.
15. a) $(0, 0)$.
b) Coordinates of those points on the graph of $y = x^3$ which are between $(-1, 1)$ and $(0, 0)$.

Section 4.2, PAGE 110

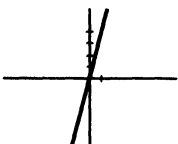
1. a)



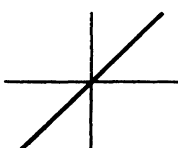
b)



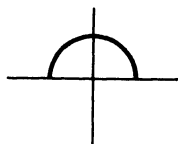
c)



d)



e)



3. a) 1.
b) 0, 2, -1.
c) All numbers.
d) All numbers.
e) $0 \leq y \leq 1$.
7. All numbers.
9. All numbers other than -1.
11. All numbers.
13. $-1 < w < 1$.
15. All numbers > 3 .

Section 4.4, PAGE 119

1. $x^2 + 3x$.
3. $7x$.
5. $5x^2 - 3x - 2$.
7. All numbers.
9. All numbers.
11. a) $(0, 3)$, $(1, 2)$, $(2, -2)$.
b) $(0, 1)$, $(1, 0)$, $(2, -6)$.
13. a) $(-2, 2)$.
b) $(-2, 8)$.

Section 4.5, PAGE 124

1. (1) $3x^3$.
(2) x^2 .
(3) $12x^2$.
(4) $5x^3 - 13x^2 - 6x$.
(5) $6x^5 - 8x^4 + 5x^3 - 13x^2 + 13x - 3$.
(6) $-3x^3 + 12x^2$.
(7) $x^2 - x$.
(8) $-x^2$.
(9) $-x^2 + 7x - 10$.
(10) $-x^2 + 11x - 28$.
(11) $(0, 2)$, $(1, 1)$, $(2, -8)$.
(12) $(0, -6)$, $(1, 15)$.
(13) $(-2, 15)$.
(14) None.
3. a) $(1, 1)$, $(2, 2)$, $(3, 3)$.
b) $g = f$.
c) $(1, 1)$, $(2, 2)$, $(4, 4)$.
d) $(1, 1)$, $(4, 4)$, $(5, 5)$.
e) $(4, 4)$, $(5, 5)$, $(6, 6)$.

5. $f(x) = 0$, x restricted in two ways.

7. a) $3 \leq x \leq 4$.

b) $3 \leq x < 4$.

c) $3 < x \leq 4$.

9. a) $x \leq -4$.

b) $x \leq -4$.

c) $x < -4$.

11. a) 0.

b) None.

c) None.

Section 4.6, PAGE 127

1. $x^3 - 4x^2 - 7x - 2$;

$t^3 - 4t^2 - 7t - 2$;
 -24 ; -12 ; -2 ; 0 .

3. $-x^3 - 2x^2 - 3x$;

$-t^3 - 2t^2 - 3t$;
 -22 ; -6 ; 0 ; 2 .

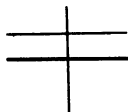
5. $\{I^3 - 3I^2 + 2I + 1\}()$.

7. $\{I^3 - 3I - 5\}()$.

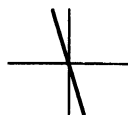
9. $\{7I^5 - 3I^2 + I\}()$.

11. $\left\{\frac{I+1}{I}\right\}()$.

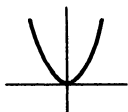
13. a)



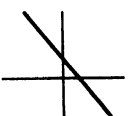
b)



c)



d)



e)



Section 5.1, PAGE 131

1. a) $9 - \sqrt{8}$, or any smaller number.

b) $9 - \sqrt{8.9}$, or any smaller number.

c) $9 - \sqrt{8.99}$, or any smaller number.

d) If $e < 9$, $9 - \sqrt{9 - e}$, or any smaller number.

3. $3 - \sqrt{9 - e} < s < 3 + \sqrt{9 - e}$.

5. If $m \neq 0$, $e/|m|$ or closer.
 If $m = 0$, any closeness.

7. a) $(\sqrt[3]{7.99}, 7.99)$ and $(\sqrt[3]{8.01}, 8.01)$.

b) $(\sqrt[3]{7.99}, 7.99)$.

c) $\sqrt[3]{7.99} < t < 4 - \sqrt[3]{7.99}$.

9. a) $(\sqrt[3]{8 - e}, 8 - e)$ and

$(\sqrt[3]{8 + e}, 8 + e)$.

b) $(\sqrt[3]{8 - e}, 8 - e)$.

c) $\sqrt[3]{8 - e} < t < 4 - \sqrt[3]{8 - e}$.

Section 5.2, PAGE 135

1. a) 9.

b) 9.

c) 9.

d) 9.

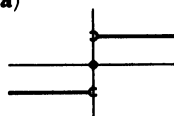
e) None.

f) None.

g) None.

3. $4 < x < 6$; $5 - e < x < 5 + e$.

5. a)



b) $h(-1) = -1$; $h(0) = 0$;

$h(1) = 1$.

c) -1 ; none; 1 .

7. Every i -interval containing a or b in its interior is not wholly in the domain of $f()$.

Section 5.3, PAGE 139

1. 6.

3. 14.

5. $11/6 < v < 13/6$.
 7. $15/8 < x < 17/8$.
 9. $d_1 = e/4$ or any smaller number.

Section 5.4, PAGE 142

1. - 55.
 3. 4.
 5. 1.
 7. 6.
 9. None.
 11. - 1.
 13. $1/2 \sqrt{a}$.
 15. $1/2$.

Section 5.5, PAGE 149

1. a) 9.
 b) $9 + 6h + h^2$.
 c) $6 + h$.
 d) 6.
 3. a) $ma + b$.
 b) $ma + mh + b$.
 c) m .
 d) m .
 5. $\{50I^9 - 84I^8 + 12I^7 - 6I^6 + 30I^5 + 6\}$ ().
 7. $\{20I^4 - 28I^3 - 6I^2\}$ ().
 9. $\{60I^{11} - 75I^4 + 8\}$ ().
 11. $\{I^3 - I + 1\}$ ().

Section 5.6, PAGE 155

1. $4\pi r^3$.
 3. $2s$.
 5. $3s^2$.
 7. a) $v = -32t + 3200$.
 b) $0 \leq s \leq 160,000$.
 c) $-3200 < v < 3200$.
 d) 160,000 feet.
 9. 25,314 miles.

Section 5.7, PAGE 158

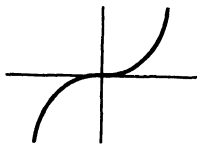
1. $4x - y = 4$.
 3. $y = 3x + r$.
 5. $2x - y = 10$; $3x + y = 4$.
 7. $18x - y = -81$ or $18x - y = 27$.
 9. $12x + y = 32$ or $12x + y = -32$.

Section 5.8, PAGE 165

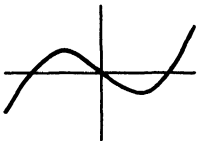
1. Yes, by definition of increasing.
 5. a) 4.
 b) 4.1, 3.9.

- c) $f(1.9) < f(2) < f(2.1)$.
 d) $3.61 < 4 < 4.41$.

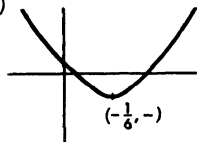
7. a)



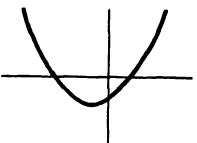
b)



c)



d)



Section 5.9, PAGE 169

1. $\{I^2 + C\}$ (); $f'(r) = 2$; $f''(r) = 0$;
 $f'''(r) = 0$.
 3. $\{I^3 - 2I^2 + C\}$ (); $f'(s) = 6s - 4$;
 $f''(s) = 6$; $f'''(s) = 0$.
 5. $\{(7/5)I^5 - (3/4)I^4 + (4/3)I^3 + (1/2)I^2 - I + C\}$ ();
 $f'(v) = 28v^3 - 9v^2 + 8v + 1$;
 $f''(v) = 84v^2 - 18v + 8$;
 $f'''(v) = 168v - 18$.
 7. $C()$; $f'(x) = f''(x) = f'''(x) = 0$.
 9. The function given by $(1/4)z^4$,
 $z \neq 1$; $f'(z) = 3z^3$, $z \neq 1$;
 $f''(z) = 6z$, $z \neq 1$; $f'''(z) = 6$,
 $z \neq 1$.
 11. The function given by $200 - 10t$
 $- 16t^2$, $0 \leq t \leq (5 + \sqrt{3225})/18$.

Section 6.1, PAGE 173

5. Yes. A polynomial function.
 7. Yes. Definition of continuous.
 9. Yes. Definition of continuous.
 11. No. 2 is not in the domain.
 13. Yes. A polynomial function.

Section 6.2, PAGE 177

5. By definition of a relative maximum value.
7. -2 and 4 are not in the range of values.
9. No maximum; minimum 0 .
11. Maximum 4 ; minimum 0 .
13. Maximum 2 ; minimum 1 .
15. Maximum 2 ; minimum 1 .
17. $f'()$ has value 0 throughout the domain of $f()$.

Section 6.3, PAGE 180

1. $(0, 0)$.
3. $(0, 4)$.
5. $(3/4, -23/8)$.
7. $(-3/4, -79/320)$.
9. $(1, 0)$ and $(2, -1)$.
11. $(-6, 125)$ and $(4, 13)$.
13. $(-2/3, 10177/27)$ and $(1/2, 1495/4)$.

Section 6.4, PAGE 184

1. $(0, 0)$ relative max.
3. $(0, -4)$ relative max.
5. $(-1/2, -1/4)$ relative min.
7. $(-2, 9)$ relative max; $(2, -13)$ relative min.
9. $(-1/3, 68/27)$ relative max; $(3, -16)$ relative min.
11. None.

Section 6.5, PAGE 187

1. 100 .
3. 6 .
5. $(15,000)^{3/64}$.
7. $25/8$.
9. $135/\sqrt{2}$.

Section 6.6, PAGE 191

3. $q \geq r$.
5. $y = G'(p)x + G(p) - pG'(p)$.
7. 0 .
9. Concave down everywhere.
11. Concave down everywhere.
13. Concave down to the left of $(3/4, -47/32)$.

Section 6.7, PAGE 192

1. $(0, 0)$ relative min.
3. $(-1, 2)$ relative max; $(0, 0)$ inflection; $(1, -2)$ relative min.
5. $(1, 0)$ inflection.
7. $(-2/3, 70/27)$ inflection.
9. $(0, 2)$ relative max; $(1/12, -)$ inflection; $(1/6, -)$ relative min.
11. $(0, 3)$ inflection; $(1, 9/4)$ relative min.

Section 6.8, PAGE 196

3. -2.2 ; 4.8 .

Section 7.1, PAGE 200

5. n^2 .
7. $n(2n+1)(3n+2)/3$.

Section 7.2, PAGE 207

3. $1/2 + 1/2n$.
5. $1/2$. A triangle.
7. $1/4$.

Section 7.3, PAGE 209

1. None.
3. 0 .
5. $1/2$.
7. 0 .
9. None.
11. $1/2$.

Section 7.4, PAGE 213

1. $1/2 \pm 1/2n$.
3. $-1 \pm 1/n$.
5. $8/3 \pm 1/4n + 4/3n^2$.
7. $2/3 \pm 1/n + 1/3n^2$.
9. $16/3 \pm 2/n + 8/3n^2$.
11. $-1/3 \pm 1/2n - 1/6n^2$.

Section 7.5, PAGE 215

1. $1/2$.
3. -1 .
5. $8/3$.
7. $2/3$.
9. $16/3$.
11. $-1/3$.

Section 7.6, PAGE 216

3. $7/4$; $5/4$.

Section 7.7, PAGE 221

3. 8.
5. $-9/2$.
7. $13/2$.
9. $85/2$.
11. 9.
13. $37/3$.
15. $8/3$.
17. $5/6$.
19. $55/3$.

Section 7.8, PAGE 223

1. $9/2$.
3. $49/2$.
5. $10/3$.
7. $1/4$.
9. 2.

Section 7.9, PAGE 229

5. Yes.
7. 0.35.
9. 0.45.
11. 0.35.
13. 0.45.
15. $\{I^3/3 + k\}()$; k constant.
17. 0.42.

Section 7.10, PAGE 232

1. $-81/4$.
3. 0.
5. $64/3$.
7. $14/3$.
9. $129/4$.
11. $27/2$.
13. 10.
15. 30.
17. -54 .
19. $151/64$.
21. $351/4$.

Section 8.1, PAGE 235

1. $1/9$; $1/8$; -8 ; $-1/9$.
3. 2; $1/27$; 2; 64.
5. 64; 512.
7. 3.
9. 0.

Section 8.2, PAGE 239

1. a) $t > -1$.
- b) $t < \sqrt{2}$.
5. None.
7. None.
9. $-11/4$.

Section 8.3, PAGE 241

1. $\{3I^3 - 4I - 8I^{-3} + 20I^{-5}\}()$.
3. $\{(3/2)I^3 - 2I - 1/3 - (1/4)I^{-3} + 4I^{-5}\}()$.
5. $\{-4\sqrt{2}I^{-5} - \sqrt{5}I^{-3} - 1\}()$.
7. $\{(1/3)I^3 + I^{-3} + (1/2)I^4 + k\}()$.
9. $\{(-1/2)I^3 - (\sqrt{3}/6)I^{-6} + (\sqrt{5}/3)I^3 + (1/18)I^{-9} + k\}()$.
11. $\{(1/5)I^5 + I^3 - I^{-1} + k\}()$.
13. $5/4$.
15. $(\sqrt[3]{2}, \sqrt[3]{2} + 1/\sqrt[3]{4})$, relative min.
17. $n(n-1)(n-2) \dots (4)(3)(2)(-1)^{n+1}t^{-n-1}$.

Section 8.4, PAGE 244

1. $\{(I+2)^{-2}\}()$.
3. $\{21/(4I+5)^2\}()$.
5. $\{[(2I^3 - 5I)(3I^3 - (I^3 + 1)(4I - 5))]/(3I + 5)^2\}()$.
7. $\{[(3I + 5)(3I^3 - 6I - (I^3 - 3I^2 + 1)(3))]/(3I + 5)^2\}()$.
9. $\left\{ \frac{\left[\begin{aligned} &(2I^8 - 2\sqrt{2}I^7)(4\sqrt{3}I) \\ &- (2\sqrt{3}I^3 + 1)(16I^7 - 14\sqrt{2}I^6) \end{aligned} \right]}{(2I^8 - 2\sqrt{2}I^7)^2} \right\}()$.

Section 8.5, PAGE 247

1. $\{(2I + 3)(5) + (5I - 1)(2)\}()$.
3. $\{(I^3 + 5I^2 - 7I + 1) + I(3I^3 + 10I - 7)\}()$.
5. $\left\{ \frac{\begin{aligned} &(6I - 7)[(2I - 3)(4) \\ &+ (4I + 5)(2)] \\ &- (2I - 3)(4I + 5)(6) \end{aligned}}{(6I - 7)^2} \right\}()$.
7. 0 relative max; $-238/27$ relative min.
9. -2 relative max; -25 relative min.

Section 8.6, PAGE 249

1. $2x^3 - 7x + 27$

$$+ \frac{-78x^2 + 113x - 109}{x^3 + 2x^2 - 3x + 4}.$$
3. $7x^5 + 12x^4 + 12x^3 - 11x^2 + 11x$

$$- 11 + \frac{8}{x + 1}.$$

Section 8.7, PAGE 253

1. a) $x > 1.1$ or $x < -1.1$.
 b) $x > 1.01$ or $x < -1.01$.
 c) $x > 1 + 1/\sqrt{B}$ or
 $x < -1 - 1/\sqrt{B}.$
3. a) $2.99 < x < 3.01$.
 b) $2.999999 < x < 3.000001$.
 c) $3 - 1/B < x < 3 + 1/B$.
9. None.
11. $+\infty$.

Section 8.8, PAGE 256

1. a) $x > 6$.
 b) $x > 501$.
 c) $x > 5/e + 1$.
3. a) 3.
 b) 3.
5. a) $y = 0$; b) $y = 2$; c) $y = -3$;
 d) $y = 2$; e) $y = 0$; f) $y = 0$;
 g) $y = 0$.
7. $1/2$.
9. None.
11. None.

Section 8.10, PAGE 261

1. a) 0.
 b) None.
 c) $+\infty$.
3. 2; 2; 3; 2.
5. $1/2$; $1/2$; $1/3$; $1/2$.
7. -2 ; 4; $-1/2$; $1/4$.
9. $\{(1/3)I^{-2/3} - (6/5)I^{-2/5} + 1\}()$.
11. $\{(3/8)I^{1/3} + (1/2)I^{-5/3}$
 $- (1/50)I^{-11/10}\}()$.
15. (1, 2) relative min.

Section 9.1, PAGE 267

1. $f(g) = g(f) = I^6$.
 3. $f(g) = 3I^{1/4} - 6$.
 5. $f(g) = g(f) = I^{1/6}$.

7. $f(g) = g(f) = I$.

9. $f(g) = g(f) = I$.

11. Add 5.

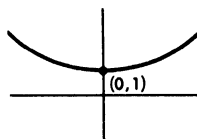
Section 9.2, PAGE 269

1. $9I^2 - 6I + 1$.
 9. a) $1, (1 + h)^6$.
 b) $6h^5 + 15h^4 + 20h^3 + 15h^2 + 6h$.
 c) 6.
 d) $2I$; $2I^3$.
 e) 2; 3.

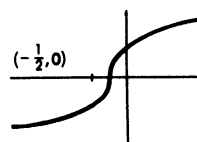
Section 9.3, PAGE 274

1. (1) $6I^5$.
 (2) $\{f(g)\}' = 6I^2$;
 $\{g(f)\}' = 24I^2 + 24I + 6$.
 (3) $(3/4)I^6$.
 (4) $(3/4)(3I - 6)$.
 (5) $(1/6)I^{-5/6}$.
 (6) $\{f(g)\}'(x) = x/x$;
 $\{g(f)\}'(x) = 1, x > 0$.
 (7) $U()$.
 (8) 12.
 (9) $U()$.
 (10) $\{f(g)\}' = 2I + 3$;
 $\{g(f)\}' = 2I + 1$.
 (11) $O()$.
3. $6(2I + 1)^2 - 4(2I + 1)$.
5. $(1/5)(I^{1/3} + I^{1/3})^{-4/5} [(1/3)I^{-2/3}$
 $+ (1/2)I^{-1/2}]$.
7. $-15(I^{-1} - I^{-2})^{-16} (-I^{-2} + 2I^{-3})$.
9. $942(2I + 3)^{470}$.

11.



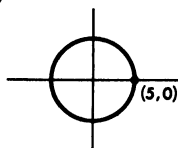
13.



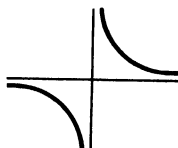
Section 9.4, PAGE 278

1. $f'(x) = -x/y$.
 3. $f'(x) = -(y/x)^{1/3}$.
 5. $f'(x) = (3x^2)/(2y)$.
 7. $x + 5y = 26$.

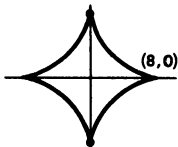
9. (1)



(2)



(3)



13. $bpx + aqy = bp^3 + aq^3.$

Section 9.5, PAGE 282

1. 50 miles per hour.
3. 54 units per hour.
5. 27.
7. $(10,368 - 120\pi)/4\pi.$
9. $(325/3)$ (height of man, in feet).

Section 9.6, PAGE 286

1. $I^{1/3}.$
3. $(I^3 + 2)^{1/3}.$
5. $-I^{-1/3}.$
7. $-I^{1/3}.$
9. None.

Section 9.7, PAGE 288

1. $2(2I + 10)^{1/3} + 2^2(2I + 10)^{-2/3} + 56.$
3. $4(I^3 + 2I - I)^3 (2I - 1)^5 \times (7I^3 + 13I - 6).$
5. $40(I^3 + 1)(I^3 + 1)^3 + 2^3[(I^3 + 1)^3 + 2^4 - 4].$
7. $3(I^9 + 3I^5 - 2I^2 + 3)^3 (4I^3 + I^3)^{-4} (24I^{11} + 7I^{10} + 36I^8 + 20I^4 - 36I^2 - 6I).$
9. $U () .$

Section 9.8, PAGE 292

1. $I^3 - I + C.$
3. $5I^4/4 - 4I^3/3 + 3I^2/2 - 2I + C.$

5. $7I^6/3 + 19I^5/5 - 31I^4/4 + I^3 - 11I^3/2 + 43I + C.$

7. $I^8/2 - 5I^6/6 + I^4/4 - 2I^3/3 + 7I^2/2 - 4I + C.$

9. $I^{16}/16 + I^4/2 + C.$

11. $4I^{4/3}/3 - 3I^2/2 - 4I^{-1} + C.$

13. $I^{-4} - 7I^{-3}/3 - 3I^{-2}/2 + C.$

15. The function given by

$$f(x) = \begin{cases} x^2, & x \geq 0 \\ -x^2, & x < 0 \end{cases}$$

Section 9.9, PAGE 296

1. $-(1/6)(3 - 4I)^{3/2} + C.$
3. $-(1/4)(I^3 + 4)^{-2} + C.$
5. $-(1/2)(I^3 + 2I + 3)^{-1} + C.$
7. $-(1/3)(I^{3/2} + 1)^{-2} + C.$
9. $-(1/192)(3I^{-4} + 1)^{16} + C.$
11. $-1.$
13. $77/1944.$

Section 9.10, PAGE 298

1. 0.
3. $-(51/56)(2)^{1/3} + 9/28.$
5. $\frac{(a+1)^{p+1}(p-a+1) + a^{p+2}}{(p+1)(p+2)}$
7. $n.$

Section 10.1, PAGE 303

1. t	$\cos t$	$\sin t$
2π	1	0
$-\pi$	-1	0
3π	-1	0
7π	-1	0
$n\pi$	$(-1)^n$	0
$2\pi/3$	$-1/2$	$\sqrt{3}/2$
$3\pi/4$	$-1/\sqrt{2}$	$1/\sqrt{2}$
$5\pi/6$	$-\sqrt{3}/2$	$1/2$
$7\pi/6$	$-\sqrt{3}/2$	$-1/2$
$5\pi/4$	$-1/\sqrt{2}$	$-1/\sqrt{2}$
$2\pi/3$	$-1/2$	$-\sqrt{3}/2$
$5\pi/3$	$1/2$	$-\sqrt{3}/2$
$7\pi/4$	$1/\sqrt{2}$	$-1/\sqrt{2}$
$11\pi/6$	$\sqrt{3}/2$	$-1/2$

Section 10.3, PAGE 308

1. $\sec t, t \neq (2k+1)\pi/2, k$ an integer.
 $\csc t, t \neq k\pi, k$ an integer.
 $\cot t, t \neq k\pi, k$ an integer.

Section 10.5, PAGE 316

3. 2.
5. 1.
7. 1.
9. 0.
11. 1.

Section 10.6, PAGE 318

5. $\sin 2s$.
7. $-3 \cos^2 u \sin u$.
9. $-(\csc^2 u)/2u$.
11. $-6w \csc 3w^3 \cot 3w^3$.
13. $-3 \cot^2(x-1) \csc^2(x-1)$.
15. Relative max at $x = 2k\pi$, k an integer;
relative min at $x = (2k+1)\pi$, k an integer.
17. Relative max at $x = (2k+1/4)\pi$, k an integer;
relative min at $x = (2k+5/4)\pi$, k an integer.

Section 10.7, PAGE 320

1. $(1/2) \sin^2 r + k$ or $(-1/2) \cos^2 r + C$.
3. $(1/2) \tan^2 t + k$.
5. $(1/3) \tan^3 t + k$.
7. $(-1/2) \cos x^2 + k$.
9. $\tan y + \sin y + k$.
11. 1.
13. $\pi/3$.
15. $(2-s^2) \cos s + 2s \sin s + k$.
17. $(u^3-2) \sin u + 2u \cos u + k$.

Section 10.8, PAGE 323

1. $\arcsin t$, $0 \leq t \leq 1$;
 $\arccos t$, $0 \leq t \leq 1$;
 $\arctan t$, $t > 0$;
 $\operatorname{arcsec} t$, $t > 1$.
3. $\pi/6$.
5. $\pi/4$.
7. $\arctan(1/\sqrt{2})$.
9. $\sqrt{3}$.
11. $\sqrt{1-s^2}$.
13. $\sqrt{u^2-1}$.
15. w , $0 \leq w \leq 1$.
17. $\pi/2 - y$.

Section 10.9, PAGE 326

3. $1/\sqrt{2x-x^2}$, $0 \leq x \leq 1$.
5. $\operatorname{Arc Sin} u$.
7. $2w/(1+w^4)$.
9. $4/(2-y)$.

Section 10.10, PAGE 330

1. $\operatorname{Arc Tan}(r/2) + k$.
3. $\operatorname{Arc Tan}(t+1) + k$.
5. $\operatorname{Arc Sec}(v+1) + k$.
7. $(1/a) \arcsin(ax/b) + k$.
9. $2 \arctan z < z + k$.
11. $\pi r^2 - r^2 \arcsin(a/r) - a \sqrt{r^2 - a^2}$.

Section 11.1, PAGE 334

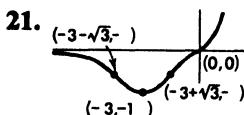
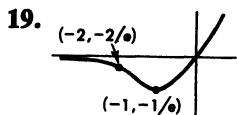
1. 0.7276.
5. $1/0729 < \log 3 < 1.1396$.
11. $I \log - I()$.
13. $\log(s+1) + k$.
15. $(1/2) \log(u^2+1) + k$.
17. $\log(-\cos w) + k$, $\pi/2 < w < \pi$.
19. $(3/2)y^2 + 17y + 81 \log(y-5) + k$.

Section 11.2, PAGE 337

1. 2.0793.
3. 2.3025.
5. 2.6390.
7. 2.9956.
9. 3.6887.
11. 4.0942.
13. -0.6931 .
15. -0.4055 .
17. -0.2876 .
19. -2.3025 .
21. -4.6050 .
23. -25.531 .

Section 11.3, PAGE 339

1. $3 \exp(3r)$.
3. $2(t+t^2) \exp(2t)$.
5. $(\cos v - \sin v) \exp v$.
7. $3 \exp(3x)$.
9. $-\exp(-z)$.
11. $(1/2) \exp(s^2) + k$.
13. $(x-1) \exp x + k$.
15. $< \exp(-w) + k$.
17. $y + k$, $y > 0$.



Section 11.4, PAGE 340

3. 7.389.
5. 54.60.
7. 0.1353.
9. 1.074 (10)³³.

Section 11.6, PAGE 345

1. $(\log 2) 2^r$.
3. $(\log 2) 2^{4t}$.
5. $\left[\log(v+2) + \frac{v}{v+2} \right] (v+2)^v$.
7. $\left[2 \log(x^3+1) + \frac{4x^3}{x^3+1} \right] (x^3+1)^{2x}$.
9. $4z^3 + 2z$.
11. $(\exp 2I)$.
13. $10 \exp(I \log \sqrt{10})$

Section 11.7, PAGE 349

5. $2r \sinh(s^3)$.
7. $[\cosh(\arcsin u)]/\sqrt{1-u^2}$.
9. $\tanh w$.
11. 0.
17. $\log(\cosh r) + k$.
19. $(1/2) \cosh t^2 + k$.

Section 11.8, PAGE 353

3. $2r/\sqrt{1+r^2}$.
5. $\cosh t \sqrt{1 + \cosh^2 t}$.
7. $v/\sqrt{v^2+1}$.
9. $(1/2)(1-x)^{-3/2}(1+x)^{-1/2}$.
11. $1/[1 + (\tanh^{-1} z)^2] (1-z^2)$.
13. $(1.2) s \sqrt{s^2-1} - \cosh^{-1} t + k$.
15. $(1/3) \cosh^{-1}(3u+1) + k$.
17. $\sin w + (1/3) \sin^3 w + k$.
19. $\int_k^t (t^3 + 2t)^{-5/2} dt$

Section 12.3, PAGE 365

1. $\pi r^2 h/3$.
3. $512\pi/15$.
5. $\pi/2$.
7. $256\pi/15$.
9. $\pi^2/4$.
11. $\pi(e+2)$.

Section 12.4, PAGE 370

1. $5\pi/6$.
3. $7\pi/15$.
5. $416\pi/15$.
7. $\pi(2-\pi)$.
9. $(e^2-3)/2$.

Section 12.5, PAGE 372

1. 254.
3. $b^2 h/3$.
5. $91\sqrt{3}$.

Section 12.6, PAGE 378

1. 3π .
3. 2π .
5. $512\pi/7$.
7. $\pi^2 abk$.

Section 12.7, PAGE 382

1. $\int_0^1 (2-2x)^2 dx$.
3. $2\pi \int_0^1 x(2-2x^2) dx$.
5. $\pi \int_{-1}^0 [(2-x^2)^2 - x^2] dx$.
7. a) $2\pi \int_0^2 [x \log(x+1) + 1 - e^{-x}] dx$
 $+ 2\pi \int_2^6 x[(\log(3e) - e^{-5})(x/3)$
 $- (2/3) \log(3e) + (5/3)e^{-5}$
 $- e^{-x}] dx$.

Section 12.8, PAGE 386

1. $(13\sqrt{13} - 8)/27$.
3. $(e^2 - 1)/e$.
5. $40/9 \sqrt{3}$.
7. $2(\sqrt{p+1} - 2)$
 $+ \log \frac{\sqrt{p+1}-1}{\sqrt{p+1}+1} + \log \frac{\sqrt{2}-1}{\sqrt{2}+1}$.
9. $\sqrt{2}$.

Section 13.1, PAGE 391

$$1. a^{p-1} + a^{p-2}b + a^{p-3}b^2 + \dots + a^2b^{p-3} + ab^{p-2} + b^{p-1}.$$

Section 13.2, PAGE 396

1. Convergent.
3. Convergent.
5. Convergent.
7. Convergent.
9. Divergent.
11. Convergent.
13. Convergent.
15. Convergent.

Section 13.3, PAGE 401

1. 12, 14, 24, 49, 476, 58888.
3. $|s| < 1$.
5. $-1 \leq u < 1$.
7. $|w| < 1$.
9. All y .
11. Same radius of convergence; 5 and 8 lose convergence at the end points.

Section 13.4, PAGE 405

$$1. a) (1-t)^{-2} = \sum_{n=0}^{\infty} (n+1) t^n.$$

$$b) (1-t)^{-3} = \frac{1}{2} \sum_{n=0}^{\infty} (n+2) \times (n+1) t^n.$$

$$c) (1-t)^{-k} = \frac{1}{k-1} \sum_{n=0}^{\infty} \times (n+k-1)(n+k-2) \dots (n+2)(n+1) t^n.$$

$$3. a) \frac{1}{a+bw} = \frac{1}{a} \sum_{n=0}^{\infty} (-1)^n \frac{b^n w^n}{a^n},$$

$$w < a/b.$$

$$b) g^{(k)}(0) = k! (-1)^k b^k / a^k.$$

5. 0.2027.
7. 1.22.
9. 1.10.

Section 13.5, PAGE 409

$$1. 1 + r.$$

$$3. \sum_{n=0}^{\infty} \frac{(-1)^n}{n+1} x^n, \quad |x| < 1.$$

$$5. \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} v^{2k+1}, \quad \text{all } v.$$

$$7. -1 + \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} x^{2n+1}, \quad |x| < 1.$$

$$9. 1 + \sum_{n=1}^{\infty} (-1)^n \frac{(2)(5)(8) \dots (3n-1)}{3^n n!} z^n,$$

$$|z| < 1.$$

Section 13.6, PAGE 412

$$1. 0.4, -1.333333, 5, 0.02004, 0.001998.$$

Section 13.7, PAGE 415

1. 3.66.
3. 1323.
5. 2.0205.

Section 13.8, PAGE 418

1. 0.995.
3. 0.296.
5. 0.199.

$$7. x + \sum_{n=1}^{\infty} \times \frac{(1)(3)(5)(7) \dots (2n-1)}{n! (2n+1)} x^{2n+1}.$$

$$9. \sin x = (1/\sqrt{2}) \left[\sum_{k=0}^{\infty} \frac{(x - \pi/4)^{4k}}{(4k)!} + \sum_{k=0}^{\infty} \frac{(x - \pi/4)^{4k+1}}{(4k+1)!} - \sum_{k=0}^{\infty} \frac{(x - \pi/4)^{4k+2}}{(4k+2)!} - \sum_{k=0}^{\infty} \frac{(x - \pi/4)^{4k+3}}{(4k+3)!} \right]$$

Section 13.9, PAGE 421

1. 1.0986.

3. 2.1972.

5. $e^3 \sum_{n=0}^{\infty} \frac{(t-2)^n}{n!}.$

7. 1.221.

9. $\cosh x = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!},$

$$\sinh x = \sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!}.$$

INDEX

- abscissa, 61
- absolute value, 56
- algebraic laws,
 - for rational numbers, 19
 - for real numbers, 27
- analytic geometry, 68
- anti-derivative, 167, 288
- anti-differentiation by parts, 297
- area, 201, 222
- associative laws,
 - for integers, 14
 - for natural numbers, 6, 7
 - for real numbers, 27
- asymptotes,
 - vertical, 251
 - horizontal, 255
- chain rule, 269
- circle, 423
- closure laws,
 - for integers, 12, 14
 - for natural numbers, 5, 7
 - for real numbers, 27
- commutative laws,
 - for integers, 13, 14
 - for natural numbers, 6, 7
 - for real numbers, 27
- complex numbers, 33
- composites of functions, 263
- concavity, 188
- conic sections, 440
- constant, 32
- continuous, 170
- convergence,
 - pointwise, 399
 - radius of, 400
 - uniform, 399
- coordinate,
 - of a point on the number line, 5
 - of a point in a plane, 61
- coordinate system,
 - linear, 30
 - plane, 61
- cosecant, 305
- cosh, 346
- cosine, 302
- cotangent, 305
- coth, 346
- counting, 1
- critical number, 52
- curve, 90
- cylinder, 356
- decreasing, 160, 165
- derivative, 143
 - of an inverse, 286
 - second, third, 166
 - of trigonometric functions, 316
- distance, 63
 - between a point and a line, 84
 - unit of, 3
- distributive law,
 - for integers, 14
 - for natural numbers, 8
 - for rational numbers, 22
 - for real numbers, 28

divisible by, 25

e , 341

ellipse, 423

end-point of an interval, 41

equality, properties of, 9, 34

equation, 35

of a collection of points, 68

equivalence of two, 39

solution of, 35

quadratic, 37

exponent, 234, 258

exponential function, 338

extent, 91

finite set, 3

first derivative test, 182

focus, foci, 423

function, 100

constant function, 111

defined by a rule, 108

domain of, 101

given by an equation, 274

graph of, 105

inverse of, 283

negative of, 118

range of values of, 101

rational function, 248

rule of, 102

sum of two, 113

function of, 103

function-of-a-function, 264

fundamental theorem of calculus, 223

horizontal axis, 61

horizontal coordinate, 61

hyperbola, 424, 426

hyperbolic functions, 345

inverse hyperbolic functions, 350

identity function, 125

identity laws,

for integers, 14, 15

for natural numbers, 8

for real numbers, 27, 28

imaginary numbers, 33

increasing, 159, 165

inequalities, 41

first degree, 43

quadratic, 45

infinite set, 3

inflection, point of, 192

integer, 12

integrable, 214

integral, 213

lower, 212

upper, 210

integration, limits of, 220

interval, 41

inverse laws,

for integers, 15

for rational numbers, 22

for real numbers, 27, 28

inverse trigonometric functions, 321

irrational numbers, 27

length of arc, 382

lines, 3, 69

equations for, 76

parallel, 78

perpendicular, 81

slope of, 73

limits, 132

infinite, 249

logarithm, 332

marginal cost, 154, 185

revenue, 185

maximum, 176

mean value theorem, 406

minimum, 176

natural numbers, 2

sums of, 197

Newton's method, 193

one-to-one correspondence, 1

ordered pair, 60

ordering laws,

for integers, 15

for natural numbers, 9

for rational numbers, 23

for real numbers, 28

ordinate, 61
origin,
 of the number line, 8
 of a plane coordinate system, 62

parabola, 423
pinching theorem, 142
polynomial function, 126
power series, 400
product,
 derivative of, 245
 of functions, 120
 of numbers, 7, 14
proof, 8

quotient,
 derivative of, 242
 of functions, 122, 235

rate,
 of change, 152
 related, 279
rational number, 16
reciprocal,
 of a function, 121
 of numbers, 20
relation, 101
rotation, 427

secant, 305
second derivative test, 183
sequences, 207, 387
 limits of, 208
series, 390
sine, 302
sinh, 346

sphere, 363
subtraction,
 of functions, 116
 of integers, 15
square roots, 33, 38
symmetry, 88, 91

tangent, 305
tangent line, 157
 vertical, 257
tanh, 346
Taylor's theorem, 406
theorem, 8
 θ ($\cdot$), 112
translation, 431
triangle inequality, 58

unit function, 112

variable, 31
 dependent and independent, 103
 range of, 31
velocity, 153
volume, 355

x -axis, 61
 x -intercept, 77

y -axis, 61
 y -intercept, 77

zero, 11
 function, 112
 of a polynomial, 127

