Undergraduate Texts in Mathematics

Serge Lang

A First Course in Calculus

Fifth Edition

Springer

Undergraduate Texts in Mathematics

Springer
New York
Berlin
Heidelberg
Hong Kong
London
Milan
Paris
Tokyo

Books of Related Interest by Serge Lang

Short Calculus
2002, ISBN 0-387-95327-2

Calculus of Several Variables, Third Edition
1987, ISBN 0-387-96405-3

Undergraduate Analysis, Second Edition
1996, ISBN 0-387-94841-4

Introduction to Linear Algebra
1997, ISBN 0-387-96205-0

Math Talks for Undergraduates
1999, ISBN 0-387-98749-5

Other Books by Lang Published by Springer-Verlag

Math! Encounters with High School Students • The Beauty of Doing Mathematics • Geometry: A High School Course • Basic Mathematics • Short Calculus • A First Course in Calculus • Introduction to Linear Algebra • Calculus of Several Variables • Linear Algebra • Undergraduate Analysis • Undergraduate Algebra • Complex Analysis • Math Talks for Undergraduates • Algebra • Real and Functional Analysis • Introduction to Differentiable Manifolds • Fundamentals of Differential Geometry • Algebraic Number Theory • Cyclotomic Fields I and II • Introduction to Diophantine Approximations • $SL_2(\mathbf{R})$ • Spherical Inversion on $SL_n(\mathbf{R})$ (with Jay Jorgenson) • Elliptic Functions • Elliptic Curves: Diophantine Analysis • Introduction to Arakelov Theory • Riemann-Roch Algebra (with William Fulton) • Abelian Varieties • Introduction to Algebraic and Abelian Functions • Complex Multiplication • Introduction to Modular Forms • Modular Units (with Daniel Kubert) • Introduction to Complex Hyperbolic Spaces • Number Theory III • Survey on Diophantine Geometry

Collected Papers I-V, including the following: Introduction to Transcendental Numbers in volume I, Frobenius Distributions in GL2-Extensions (with Hale Trotter in volume II, Topics in Cohomology of Groups in volume IV, Basic Analysis of Regularized Series and Products (with Jay Jorgenson) in volume V and Explicit Formulas for Regularized Products and Series (with Jay Jorgenson) in volume V

THE FILE • CHALLENGES

Serge Lang

A First Course in Calculus

Fifth Edition

With 367 Illustrations

Springer

Serge Lang
Department of Mathematics
Yale University
New Haven, CT 06520
USA

Mathematics Subjects Classifications (2000): 26-01, 26A06

Library of Congress Cataloging-in-Publication Data
Lang, Serge, 1927–
A first course in calculus.
(Undergraduate texts in mathematics)
Includes index.
1. Calculus. I. Title. II. Series.
QA303.L26 1986 515 85-17181

Printed on acid-free paper.

Previous editions of this book were published in 1978, 1973, 1968, 1964 by Addison-Wesley, Publishing Company, Inc.

Typeset by Composition House Ltd., Salisbury, England.
Printed and bound by Sheridan Books, Ann Arbor, MI.
Printed in the United States of America.

9 8

ISBN 0-387-96201-8
ISBN 3-540-96201-8 SPIN 10965351

Springer-Verlag New York Berlin Heidelberg
A member of BertelsmannSpringer Science+Business Media GmbH

Foreword

The purpose of a first course in calculus is to teach the student the basic notions of derivative and integral, and the basic techniques and applications which accompany them. The very talented students, with an obvious aptitude for mathematics, will rapidly require a course in functions of one real variable, more or less as it is understood by professional mathematicians. This book is not primarily addressed to them (although I hope they will be able to acquire from it a good introduction at an early age).

I have not written this course in the style I would use for an advanced monograph, on sophisticated topics. One writes an advanced monograph for oneself, because one wants to give permanent form to one's vision of some beautiful part of mathematics, not otherwise accessible, somewhat in the manner of a composer setting down his symphony in musical notation.

This book is written for the students to give them an immediate, and pleasant, access to the subject. I hope that I have struck a proper compromise, between dwelling too much on special details and not giving enough technical exercises, necessary to acquire the desired familiarity with the subject. In any case, certain routine habits of sophisticated mathematicians are unsuitable for a first course.

Rigor. This does not mean that so-called rigor has to be abandoned. The logical development of the mathematics of this course from the most basic axioms proceeds through the following stages:

Set theory	Numbers (i.e. real numbers)
Integers (whole numbers)	Limits
Rational numbers (fractions)	Derivatives and forward.

No one in his right mind suggests that one should begin a course with set theory. It happens that the most satisfactory place to jump into the subject is between limits and derivatives. In other words, any student is ready to accept as intuitively obvious the notions of numbers and limits and their basic properties. Experience shows that the students do *not* have the proper psychological background to accept a theoretical study of limits, and resist it tremendously.

In fact, it turns out that one can have the best of both ideas. The arguments which show how the properties of limits can be reduced to those of numbers form a self-contained whole. Logically, it belongs *before* the subject matter of our course. Nevertheless, we have inserted it as an appendix. If some students feel the need for it, they need but read it and visualize it as Chapter 0. In that case, everything that follows is as rigorous as any mathematician would wish it (so far as objects which receive an analytic definition are concerned). Not one word need be changed in any proof. I hope this takes care once and for all of possible controversies concerning so-called rigor.

Most students will not feel any need for it. My opinion is that epsilon-delta should be entirely left out of ordinary calculus classes.

Language and logic. It is not generally recognized that some of the major difficulties in teaching mathematics are analogous to those in teaching a foreign language. (The secondary schools are responsible for this. Proper training in the secondary schools could entirely eliminate this difficulty.) Consequently, I have made great efforts to carry the student verbally, so to say, in using proper mathematical language. It seems to me essential that students be required to write their mathematics papers in full and coherent sentences. A large portion of their difficulties with mathematics stems from their slapping down mathematical symbols and formulas isolated from a meaningful sentence and appropriate quantifiers. Papers should also be required to be neat and legible. They should not look as if a stoned fly had just crawled out of an inkwell. Insisting on reasonable standards of expression will result in drastic improvements of mathematical performance. The systematic use of words like "let," "there exists," "for all," "if ... then," "therefore" should be taught, as in sentences like:

Let $f(x)$ be the function such that ....
There exists a number such that ....
For all numbers x with $0 < x < 1$, we have ....
If f is a differentiable function and K a constant such that $f'(x) = Kf(x)$, then $f(x) = Ce^{Kx}$ for some constant C.

Plugging in. I believe that it is unsound to view "theory" as adversary to applications or "computations." The present book treats both as complementary to each other. Almost always a theorem gives a tool for more efficient computations (e.g. Taylor's formula, for computing values of functions). Different classes will of course put different emphasis on them, omitting some proofs, but I have found that if no excessive pedantry is introduced, students are willing, and even eager, to understand the reasons for the truth of a result, i.e. its proof.

It is a disservice to students to teach calculus (or other mathematics, for that matter) in an exclusive framework of "plugging in" ready-made formulas. Proper teaching consists in making the student adept at handling a large number of techniques in a routine manner (in particular, knowing how to plug in), but it also consists in training students in knowing some general principles which will allow them to deal with new situations for which there are no known formulas to plug in.

It is impossible in one semester, or one year, to find the time to deal with all desirable applications (economics, statistics, biology, chemistry, physics, etc.). On the other hand, covering the proper balance between selected applications and selected general principles will equip students to deal with other applications or situations by themselves.

Worked-out problems and exercises. For the convenience of both students and instructors, a large number of worked-out problems has been added in the present edition. Many of these have been put in the answer section, to be referred to as needed. I did this for at least two reasons. First, in the text, they might obscure the main ideas of the course. Second, it is a good idea to make students think about a problem before they see it worked out. They are then much more receptive, and will retain the methods better for having encountered the difficulties (whatever they are, depending on individual students) by themselves. Both the inclusion of worked-out examples and their placement in the answer section was requested by students. Unfortunately, the requirements for good teaching, testing, and academic pressures are in conflict here. The *de facto* tendency is for students to object to being asked to think (even if they fail), because they are afraid of being penalized with bad grades for homework. Instructors may either make too strong requirements on students, or may take the path of least resistance and never require anything beyond plugging in new numbers in a type of exercise which has already been worked out (in class or in the book). I believe that testing conditions (limited time, pressures of other courses and examinations) make it difficult (if not unreasonable) to *test* students other than with basic, routine problems. I do not conclude that the course should consist only of this type of material. Some students often take the attitude that if something is not on tests, then why should it be covered in the course? I object very much to this attitude. I have no global solution to these conflicting pressures.

General organization. I have made no great innovations in the exposition of calculus. Since the subject was discovered some 300 years ago, such innovations were out of the question.

I have cut down the amount of analytic geometry to what is both necessary and sufficient for a general first course in this type of mathematics. For some applications, more is required, but these applications are fairly specialized. For instance, if one needs the special properties concerning the focus of a parabola in a course on optics, then that is the place to present them, not in a general course which is to serve mathematicians, physicists, chemists, biologists, and engineers, to mention but a few. I regard the tremendous emphasis on the analytic geometry of conics which has been the fashion for many years as an unfortunate historical accident. What is important is that the basic idea of representing a graph by a figure in the plane should be thoroughly understood, together with basic examples. The more abstruse properties of ellipses, parabolas, and hyperbolas should be skipped.

Differentiation and the elementary functions are covered first. Integration is covered second. Each makes up a coherent whole. For instance, in the part on differentiation, rate problems occur three times, illustrating the same general principle but in the contexts of several elementary functions (polynomials at first, then trigonometric functions, then inverse functions). This repetition at brief intervals is pedagogically sound, and contributes to the coherence of the subject. It is also natural to slide from integration into Taylor's formula, proved with remainder term by integrating by parts. It would be slightly disagreeable to break this sequence.

Experience has shown that Chapters III through VIII make up an appropriate curriculum for **one term (differentiation and elementary functions)** while Chapters IX through XIII make up an appropriate curriculum for a **second term (integration and Taylor's formula)**. The first two chapters may be used for a quick review by classes which are not especially well prepared.

I find that all these factors more than offset the possible disadvantage that for other courses (physics, chemistry perhaps) integration is needed early. This may be true, but so are the other topics, and unfortunately the course has to be projected in a totally ordered way on the time axis.

In addition to this, studying the log and exponential before integration has the advantage that we meet in a special concrete case the situation where we find an antiderivative by means of area: $\log x$ is the area under $1/x$ between 1 and x. We also see in this concrete case how $dA(x)/dx = f(x)$, where $A(x)$ is the area. This is then done again in full generality when studying the integral. Furthermore, inequalities involving lower sums and upper sums, having already been used in this concrete case, become more easily understandable in the general case. Classes which start the term on integration without having gone through the part on differentiation might well start with the last section of the chapter on logarithms, i.e. the last section of Chapter VIII.

Taylor's formula is proved with the integral form of the remainder, which is then properly estimated. The proof with integration by parts is more natural than the other (differentiating some complicated expression pulled out of nowhere), and is the one which generalizes to the higher dimensional case. I have placed integration after differentiation, because otherwise one has no technique available to evaluate integrals.

I personally think that the computations which arise naturally from Taylor's formula (computations of values of elementary functions, computation of e, π, $\log 2$, computations of definite integrals to a few decimals, traditionally slighted in calculus courses) are important. This was clear already many years ago, and is even clearer today in the light of the pocket computer proliferation. Designs of such computers rely precisely on effective means of computation by means of the Taylor polynomials. Learning how to estimate effectively the remainder term in Taylor's formula gives a very good feeling for the elementary functions, not obtainable otherwise.

The computation of integrals like

$$\int_0^1 e^{-x^2}\,dx \qquad \text{or} \qquad \int_0^{0.1} e^{-x^2}\,dx$$

which can easily be carried out numerically, without the use of a simple form for the indefinite integral, should also be emphasized. Again it gives a good feeling for an aspect of the integral not obtainable otherwise. Many texts slight these applications in favor of expanded treatment of applications of integration to various engineering situations, like fluid pressure on a dam, mainly by historical accident. I have nothing against fluid pressure, but one should keep in mind that too much time spent on some topics prevents adequate time being spent on others. For instance, Ron Infante tells me that numerical computations of integrals like

$$\int_0^1 \frac{\sin x}{x}\,dx,$$

which we carry out in Chapter XIII, occur frequently in the study of communication networks, in connection with square waves. Each instructor has to exercise some judgment as to what should be emphasized at the expense of something else.

The chapters on functions of several variables are included for classes which can proceed at a faster rate, and therefore have time for additional material during the first year. **Under ordinary circumstances, these chapters will not be covered during a first-year course. For instance, they are not covered during the first-year course at Yale.**

Induction. I think the first course in calculus is a good time to learn induction. However, an attempt to teach induction without having met natural examples first meets with very great psychological difficulties. Hence throughout the part on differentiation, I have not mentioned induction formally. Whenever a situation arises where induction may be used, I carry out stepwise procedures illustrating the inductive procedure. After enough repetitions of these, the student is then ready to see a pattern which can be summarized by the formal "induction," which just becomes a name given to a notion which has already been understood.

Review material. The present edition also emphasizes more review material. Deficient high school training is responsible for many of the difficulties experienced at the college level. These difficulties are not so much due to the problem of understanding calculus as to the inability to handle elementary algebra. A large group of students cannot automatically give the expansion for expressions like

$$(a+b)^2, \qquad (a-b)^2, \qquad \text{or} \qquad (a+b)(a-b).$$

The answers should be memorized like the multiplication table. To memorize by rote such basic formulas is not incompatible with learning general principles. It is complementary.

To avoid any misunderstandings, I wish to state explicitly that the poor preparation of so many high school students cannot be attributed to the "new math" versus the "old math." When I started teaching calculus as a graduate student in 1950, I found the quasi-totality of college freshmen badly prepared. Today, I find only a substantial number of them (it is hard to measure how many). On the other hand, a sizable group at the top has had the opportunity to learn some calculus, even as much as one year, which would have been inconceivable in former times. As bad as the situation is, it is nevertheless an improvement.

I wish to thank my colleagues at Yale and others in the past who have suggested improvements in the book: Edward Bierstone (University of Toronto), Folke Eriksson (University of Gothenburg), R. W. Gatterdam (University of Wisconsin, Parkside), and George Metakides (University of Rochester). I thank Ron Infante for assisting with the proofreading.

I am much indebted to Anthony Petrello for checking worked-out examples and answers in past editions. I am also much indebted to Allan Altman for a long list of corrections.

S. Lang

Contents

Part One

Review of Basic Material

If you are already at ease with the elementary properties of numbers and if you know about coordinates and the graphs of the standard equations (linear equations, parabolas, ellipses), then you should start immediately with Chapter III on derivatives.

CHAPTER I

Numbers and Functions

In starting the study of any sort of mathematics, we cannot prove everything. Every time that we introduce a new concept, we must define it in terms of a concept whose meaning is already known to us, and it is impossible to keep going backwards defining forever. Thus we must choose our starting place, what we assume to be known, and what we are willing to explain and prove in terms of these assumptions.

At the beginning of this chapter, we shall describe most of the things which we assume known for this course. Actually, this involves very little. Roughly speaking, we assume that you know about numbers, addition, subtraction, multiplication, and division (by numbers other than 0). We shall recall the properties of inequalities (when a number is greater than another). On a few occasions we shall take for granted certain properties of numbers which might not have occurred to you before and which will always be made precise. Proofs of these properties will be supplied in the Appendix for those of you who are interested.

I, §1. INTEGERS, RATIONAL NUMBERS, AND REAL NUMBERS

The most common numbers are the numbers $1, 2, 3, \ldots$ which are called **positive integers**.

The numbers $-1, -2, -3, \ldots$ are called **negative integers**. When we want to speak of the positive integers together with the negative integers and 0, we call them simply **integers**. Thus the integers are $0, 1, -1, 2, -2, 3, -3, \ldots$.

The sum and product of two integers are again integers.

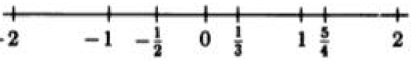
In addition to the integers we have **fractions**, like $\frac{3}{4}$, $\frac{5}{7}$, $-\frac{1}{8}$, $-\frac{101}{27}$, $\frac{8}{16}, \ldots$, which may be positive or negative, and which can be written as quotients m/n, where m, n are integers and n is not equal to 0. Such fractions are called **rational numbers**. Every integer m is a rational number, because it can be written as $m/1$, but of course it is not true that every rational number is an integer. We observe that the sum and product of two rational numbers are again rational numbers. If a/b and m/n are two rational numbers (a, b, m, n being integers and b, n unequal to 0), then their sum and product are given by the following formulas, which you know from elementary school:

$$\frac{a}{b}\frac{m}{n} = \frac{am}{bn},$$

$$\frac{a}{b} + \frac{m}{n} = \frac{an + bm}{bn}.$$

In this second formula, we have simply put the two fractions over the common denominator bn.

We can represent the integers and rational numbers geometrically on a straight line. We first select a unit length. The integers are multiples of this unit, and the rational numbers are fractional parts of this unit. We have drawn a few rational numbers on the line below.



Observe that the negative integers and rational numbers occur to the left of zero.

Finally, we have the numbers which can be represented by infinite decimals, like $\sqrt{2} = 1.414\ldots$ or $\pi = 3.14159\ldots$, and which will be called **real numbers** or simply **numbers**.

The integers and rational numbers are special cases of these infinite decimals. For instance,

$$3 = 3.000000\ldots,$$

and

$$\tfrac{3}{4} = 0.7500000\ldots,$$

$$\tfrac{1}{3} = 0.3333333\ldots.$$

We see that there may be several ways of denoting the same number, for instance as the fraction $\frac{1}{3}$ or as the infinite decimal $0.33333\ldots$. We have written the decimals with dots at the end. If we stop the decimal expansion at any given place, we obtain an approximation to the number. The further off we stop the decimal, the better approximation we obtain.

Finding the decimal expansion for a fraction is easy by the process of long division which you should know from high school.

Later in the course we shall learn how to find decimal expansions for other numbers which you may have heard about, like π. You were probably told that $\pi = 3.14 \ldots$ but were not told why. You will learn how to compute arbitrarily many decimals for π in Chapter XIII.

Geometrically, the numbers are represented as the collection of all points on the above straight line, not only those which are a rational part of the unit length or a multiple of it.

We note that the sum and product of two numbers are numbers. If a is a number unequal to zero, then there is a unique number b such that $ab = ba = 1$, and we write

$$b = \frac{1}{a} \qquad \text{or} \qquad b = a^{-1}.$$

We say that b is the **inverse** of a, or "a inverse." We emphasize that the expression

$$\mathbf{1/0} \qquad \textbf{or} \qquad \mathbf{0^{-1}} \quad \textbf{is not defined.}$$

In other words, we cannot divide by zero, and we do not attribute any meaning to the symbols $1/0$ or 0^{-1}.

However, if a is a number, then the prodct $0 \cdot a$ is defined and is equal to 0. The product of any number and 0 is 0. Furthermore, if b is any number unequal to 0, then $0/b$ is defined and equal to 0. It can also be written $0 \cdot (1/b)$.

If a is a rational number $\neq 0$, then $1/a$ is also a rational number. Indeed, if we can write $a = m/n$, with integers m, n both different from 0, then

$$\frac{1}{a} = \frac{n}{m}$$

is also a rational number.

I, §2. INEQUALITIES

Aside from addition, multiplication, subtraction, and division (by numbers other than 0), we shall now discuss another important feature of the real numbers.

We have the **positive numbers**, represented geometrically on the straight line by those numbers unequal to 0 and lying to the right of 0. If a is a positive number, we write $a > 0$. You have no doubt already worked with positive numbers, and with inequalities. The next two properties are the most basic ones, concerning positivity.

POS 1. *If a, b are positive, so is the product ab and the sum $a + b$.*

POS 2. *If a is a number, then either a is positive, or $a = 0$, or $-a$ is positive, and these possibilities are mutually exclusive.*

If a number is not positive and not 0, then we say that this number is **negative**. By **POS 2**, if a is negative, then $-a$ is positive.

Although you know already that the number 1 is positive, it can in fact be **proved** from our two properties. It may interest you to see the proof, which runs as follows and is very simple. By **POS 2**, we know that either 1 or -1 is positive. If 1 is not positive, then -1 is positive. By **POS 1**, it must then follow that $(-1)(-1)$ is positive. But this product is equal to 1. Consequently, it must be 1 which is positive, and not -1. Using property **POS 1**, we could now conclude that $1 + 1 = 2$ is positive, that $2 + 1 = 3$ is positive, and so forth.

If $a > 0$, we shall say that a is **greater than** 0. If we wish to say that a is positive or equal to 0, we write

$$a \geqq 0$$

and read this "a greater than or equal to zero."

Given two numbers a, b we shall say that a is **greater than** b and write $a > b$ if $a - b > 0$. We write $a < 0$ (a is **less than** 0) if $-a > 0$ and $a < b$ if $b > a$. Thus $3 > 2$ because $3 - 2 > 0$.

We shall write $a \geqq b$ when we want to say that a is **greater than or equal to** b. Thus $3 \geqq 2$ and $3 \geqq 3$ are both true inequalities.

Other rules concerning inequalities are valid.

In what follows, let a, b, c be numbers.

Rule 1. *If $a > b$ and $b > c$, then $a > c$.*

Rule 2. *If $a > b$ and $c > 0$, then $ac > bc$.*

Rule 3. *If $a > b$ and $c < 0$, then $ac < bc$.*

Rule 2 expresses the fact that an inequality which is multiplied by a positive number is **preserved**. Rule 3 tells us that if we multiply both sides of an inequality by a negative number, then the inequality gets **reversed**. For instance, we have the inequality

$$1 < 3$$

Since $2 > 0$ we also have $2 \cdot 1 < 2 \cdot 3$. But -2 is negative, and if we multiply both sides by -2 we get

$$-2 > -6.$$

In the geometric representation of the real numbers on the line, -2 lies to the right of -6. This gives us the geometric representation of the fact that -2 is greater than -6.

If you wish, you may assume these three rules just as you assume **POS 1** and **POS 2**. All of these are used in practice. It turns out that the three rules can be proved in terms of **POS 1** and **POS 2**. We cannot assume all the inequalities which you will ever meet in practice. Hence just to show you some techniques which might recur for other applications, we show how we can deduce the three rules from **POS 1** and **POS 2**. You may omit these (short) proofs if you wish.

To prove Rule 1, suppose that $a > b$ and $b > c$. By definition, this means that $(a - b) > 0$ and $(b - c) > 0$. Using property **POS 1**, we conclude that

$$a - b + b - c > 0,$$

and canceling b gives us $(a - c) > 0$. By definition, this means $a > c$, as was to be shown.

To prove Rule 2, suppose that $a > b$ and $c > 0$. By definition,

$$a - b > 0.$$

Hence using the property of **POS 1** concerning the product of positive numbers, we conclude that

$$(a - b)c > 0.$$

The left-hand side of this inequality is none other than $ac - bc$, which is therefore > 0. Again by definition, this gives us

$$ac > bc.$$

We leave the proof of Rule 3 as an exercise.

We give an example showing how to use the three rules.

Example. Let a, b, c, d be numbers with $c, d > 0$. Suppose that

$$\frac{a}{c} < \frac{b}{d}.$$

We wish to prove the "cross-multiplication" rule that

$$ad < bc.$$

Using Rule 2, multiplying each side of the original inequality by c, we obtain

$$a < bc/d.$$

Using Rule 2 again, and multiplying each side by d, we obtain

$$ad < bc,$$

as desired.

Let a be a number > 0. Then there exists a number whose square is a. If $b^2 = a$ then we observe that

$$(-b)^2 = b^2$$

is also to a. Thus either b or $-b$ is positive. We agree to denote by $\sqrt{a}$ the **positive** square root and call it simply **the square root of** a. Thus $\sqrt{4}$ is equal to 2 and not -2, even though $(-2)^2 = 4$. This is the most practical convention about the use of the $\sqrt{\ }$ sign that we can make. Of course, the square root of 0 is 0 itself. A negative number does *not* have a square root in the real numbers.

There are thus two solutions to an equation

$$x^2 = a$$

with $a > 0$. These two solutions are $x = \sqrt{a}$ and $x = -\sqrt{a}$. For instance, the equation $x^2 = 3$ has the two solutions

$$x = \sqrt{3} = 1.732\ldots \qquad \text{and} \qquad x = -\sqrt{3} = -1.732\ldots.$$

The equation $x^2 = 0$ has exactly one solution, namely $x = 0$. The equation $x^2 = a$ with $a < 0$ has no solution in the real numbers.

Definition. Let a be a number. We define the **absolute value** of a to be

$$\boxed{|a| = \sqrt{a^2}.}$$

In particular,

$$\boxed{|a|^2 = a^2.}$$

Thus the absolute value of a number is always $\geqq 0$. The absolute value of a positive number is always positive.

Example. We have

$$|3| = \sqrt{3^2} = \sqrt{9} = 3,$$

but

$$|-3| = \sqrt{(-3)^2} = \sqrt{9} = 3.$$

Also for any number a we get

$$|-a| = \sqrt{(-a)^2} = \sqrt{a^2} = |a|.$$

Theorem 2.1. *If a is any number, then*

$$|a| = \begin{cases} a & \text{if } a \geqq 0, \\ -a & \text{if } a < 0. \end{cases}$$

Proof. If $a \geqq 0$ then a is the unique number $\geqq 0$ whose square is a^2, so $|a| = \sqrt{a^2} = a$. If $a < 0$ then $-a > 0$ and

$$(-a)^2 = a^2,$$

so this time $-a$ is the unique number > 0 whose square is a^2, whence $|a| = -a$. This proves the theorem.

Theorem 2.2. *If a, b are numbers, then*

$$|ab| = |a|\,|b|.$$

Proof. We have:

$$|ab| = \sqrt{(ab)^2} = \sqrt{a^2b^2} = \sqrt{a^2}\sqrt{b^2} = |a|\ |b|.$$

As an example, we see that

$$|-6| = |(-3)\cdot 2| = |-3|\ |2| = 3\cdot 2 = 6.$$

There is one final inequality which is extremely important.

Theorem 2.3. *If a, b are two numbers, then*

$$|a + b| \leqq |a| + |b|.$$

Proof. We first observe that either ab is positive, or it is negative, or it is 0. In any case, we have

$$ab \leqq |ab| = |a|\ |b|.$$

Hence, multiplying both sides by 2, we obtain the inequality

$$2ab \leqq 2|a|\ |b|.$$

Using this inequality we find:

$$\begin{aligned}(a + b)^2 &= a^2 + 2ab + b^2 \\ &\leqq a^2 + 2|a|\,|b| + b^2 \\ &= (|a| + |b|)^2.\end{aligned}$$

We can take the square root of both sides and use Theorem 2.1 to conclude that

$$|a + b| \leqq |a| + |b|,$$

thereby proving our theorem.

You will find plenty of exercises below to give you practice with inequalities. We shall work out some numerical examples to show you the way.

Example 1. Determine the numbers satisfying the equality

$$|x + 1| = 2.$$

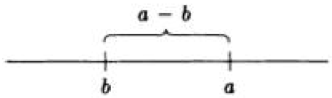
This equality means that either $x + 1 = 2$ or $-(x + 1) = 2$, because the absolute value of $x + 1$ is either $(x + 1)$ itself or $-(x + 1)$. In the first case, solving for x gives us $x = 1$, and in the second case, we get $-x - 1 = 2$ or $x = -3$. Thus the answer is $x = 1$ or $x = -3$.

Let a, b be numbers. We may interpret

$$|a - b| = \sqrt{(a - b)^2}$$

as the distance between a and b.

For instance, if $a > b$ then this is geometrically clear from the figure.



On the other hand, if $a < b$ we have

$$|a - b| = |-(b - a)| = |b - a|,$$

and $b > a$, so again we see that $|a - b| = |b - a|$ is the distance between a and b.

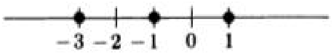
In the preceding example, the set of numbers x such that

$$|x + 1| = 2$$

is the set of numbers whose distance from -1 is 2, because we can write

$$x + 1 = x - (-1).$$

Hence we see again geometrically that this set of numbers consists of 1 and -3, as shown on the figure.



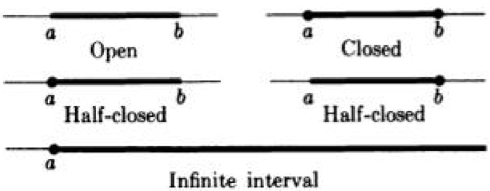
We shall also give an example showing how to determine numbers satisfying certain inequalities. For this we need some terminology. Let a, b be numbers, and assume $a < b$.

The collection of numbers x such that $a < x < b$ is called the **open interval** between a and b, and is sometimes denoted by (a, b).

The collection of numbers x such that $a \leqq x \leqq b$ is called the **closed interval** between a and b, and is sometimes denoted by $[a, b]$. A single point will also be called a closed interval.

In both above cases, the numbers a, b are called the **end points** of the intervals. Sometimes we wish to include only one of them in an interval, and so we define the collection of numbers x such that $a \leqq x < b$ to be **half-closed interval**, and similarly for those numbers x such that $a < x \leqq b$.

Finally, if a is a number, we call the collection of numbers $x > a$, or $x \geqq a$, or $x < a$, or $x \leqq a$ an **infinite interval**. Pictures of intervals are shown below.



Example 2. Determine all intervals of numbers satisfying

$$|x| \leqq 4.$$

We distinguish two cases. The first case is $x \geqq 0$. Then $|x| = x$, and in this case, our inequality amounts to

$$0 \leqq x \leqq 4.$$

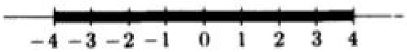
The second case is $x < 0$. In this case, $|x| = -x$, and our inequality amounts to $-x \leqq 4$, or in other words, $-4 \leqq x$. Thus in the second case, the numbers satisfying our inequality are precisely those in the interval.

$$-4 \leqq x < 0.$$

Considering now both cases together, we see that the interval of numbers satisfying our inequality $|x| \leqq 4$ is the interval

$$-4 \leqq x \leqq 4.$$

We can also phrase the answer in terms of distance. The numbers x such that $|x| \leqq 4$ are precisely those numbers whose distance from the origin is $\leqq 4$. Thus they constitute the closed interval between -4 and 4 as shown on the figure.



More generally, let a be a positive number. A number x satisfies the inequality $|x| < a$ if and only if

$$-a < x < a.$$

The argument to prove this is the same as in the special case $a = 4$ worked out above.

Example 3. Determine all intervals of numbers satisfying the inequality

$$|x + 1| > 2.$$

This inequality is equivalent with the two inequalities

$$x + 1 > 2 \quad \text{or} \quad -(x + 1) > 2.$$

From the first we get the condition $x > 1$, and from the second, we get the condition $-x - 1 > 2$, or in other words, $x < -3$. Thus there are two (infinite) intervals, namely

$$x > 1 \qquad \text{and} \qquad x < -3.$$

Example 4. On the other hand, we wish to determine the interval of numbers x such that

$$|x + 1| < 2.$$

These are the numbers x whose distance from -1 is < 2, because we can write

$$x + 1 = x - (-1).$$

Hence it is the interval of numbers satisfying

$$-3 < x < 1$$

as shown on the figure.

I, §2. EXERCISES

Determine all intervals of numbers x satisfying the following inequalities.

1. $|x| < 3$
2. $|2x + 1| \leqq 1$
3. $|x^2 - 2| \leqq 1$
4. $|x - 5| > 2$
5. $(x + 1)(x - 2) < 0$
6. $(x - 1)(x + 1) > 0$
7. $(x - 5)(x + 5) < 0$
8. $x(x + 1) \leqq 0$
9. $x^2(x - 1) \geqq 0$
10. $(x - 5)^2(x + 10) \leqq 0$
11. $(x - 5)^4(x + 10) \leqq 0$
12. $(2x + 1)^6(x - 1) \geqq 0$
13. $(4x + 7)^{20}(2x + 8) < 0$
14. $|x + 4| < 1$
15. $0 < |x + 2| < 1$
16. $|x| < 2$
17. $|x - 3| < 5$
18. $|x - 3| < 1$
19. $|x - 3| < 7$
20. $|x - 3| > 7$
21. $|x + 3| > 7$

Prove the following inequalities for all numbers x, y.

22. $|x + y| \geqq |x| - |y|$ [*Hint*: Write $x = x + y - y$, and apply Theorem 2.3, together with the fact that $|-y| = |y|$.]

23. $|x - y| \geqq |x| - |y|$

24. $|x - y| \leqq |x| + |y|$

25. Let a, b be positive numbers such that $a < b$. Show that $a^2 < b^2$.

26. Let a, b, c, d be numbers > 0, such that $a/b < c/d$. Show that

$$\frac{a}{b} < \frac{a + c}{b + d} \quad \text{and} \quad \frac{a + c}{b + d} < \frac{c}{d}.$$

27. Let a, b be numbers > 0. Show that

$$\sqrt{ab} \leqq \frac{a + b}{2}.$$

28. Let $0 < a < b$ and $0 < c < d$. Prove that

$$ac < bd.$$

I, §3. FUNCTIONS

A function, defined for all numbers, is an association which to any given number associates another number.

It is customary to denote a function by some letter, just as a letter "x" denotes a number. Thus if we denote a given function by f, and x is a number, then we denote by $f(x)$ the number associated with x by the function. This of course does not mean "f times x." There is no multiplication involved here. The symbols $f(x)$ are read "f of x." The association of the number $f(x)$ to the number x is sometimes denoted by a special arrow, namely

$$x \mapsto f(x).$$

For example, consider the function which associates to each number x the number x^2. If f denotes this function, then we have $f(x) = x^2$. In particular, the square of 2 is 4 and hence $f(2) = 4$. The square of 7 is 49 and thus $f(7) = 49$. The square of $\sqrt{2}$ is 2, and hence $f(\sqrt{2}) = 2$. The square of $(x + 1)$ is

$$x^2 + 2x + 1$$

and thus $f(x+1) = x^2 + 2x + 1$. If h is any number,

$$f(x+h) = x^2 + 2xh + h^2.$$

To take another example, let g be the function which to each number x associates the number $x + 1$. Then we may describe g by the symbols

$$x \mapsto x + 1$$

and write $g(x) = x + 1$. Therefore, $g(1) = 2$. Also $g(2) = 3$, $g(3) = 4$, $g(\sqrt{2}) = \sqrt{2} + 1$, and $g(x+1) = x + 2$ for any number x.

We can view the **absolute value** as a function,

$$x \mapsto |x|$$

defined by the rule: Given any number a, we associate the number a itself if $a \geqq 0$, and we associate the number $-a$ if $a < 0$. Let F denote the absolute value function. Then $F(x) = |x|$ for any number x. We have in particular $F(2) = 2$, and $F(-2) = 2$ also. The absolute value is not defined by means of a formula like x^2 or $x + 1$. We give you another example of such a function which is not defined by a formula.

We consider the function G described by the following rule:

$$G(x) = 0 \quad \text{if } x \text{ is a rational number.}$$

$$G(x) = 1 \quad \text{if } x \text{ is not a rational number.}$$

Then in particular, $G(2) = G(\frac{2}{3}) = G(-\frac{3}{4}) = 0$ but

$$G(\sqrt{2}) = 1.$$

You must be aware that you can construct a function just by prescribing arbitrarily the rule associating a number to a given one.

If f is a function and x a number, then $f(x)$ is called the **value** of the function at x. Thus if f is the function

$$x \mapsto x^2,$$

the value of f at 2 is 4 and the value of f at $\frac{1}{2}$ is $\frac{1}{4}$.

In order to describe a function, we need simply to give its value at any number x. That is the reason why we use the notation $x \mapsto f(x)$. Sometimes, for brevity, we speak of a function $f(x)$, meaning by that the function f whose value at x is $f(x)$. For instance, we would say "Let $f(x)$ be the function $x^3 + 5$" instead of saying "Let f be the function which to each number x associates $x^3 + 5$." Using the special arrow $\mapsto$, we could say also "Let f be the function $x \mapsto x^3 + 5$."

We would also like to be able to define a function for some numbers and leave it undefined for others. For instance we would like to say that $\sqrt{x}$ is a function (the square root function, whose value at a number x is the square root of that number), but we observe that a negative number does not have a square root. Hence it is desirable to make the notion of function somewhat more general by stating explicitly for what numbers it is defined. For instance, the square root function is defined only for numbers $\geqq 0$. This function is denoted by $\sqrt{x}$. The value $\sqrt{x}$ is the unique number $\geqq 0$ whose square is x.

Thus in general, let S be a collection of numbers. By a **function, defined on** S, we mean an association, which to each number x in S associates a number. We call S the **domain of definition** of the function. For example, the domain of definition of the square root function is the collection of all numbers $\geqq 0$.

Let us give another example of a function which is not defined for all numbers. Let S be the collection of all numbers $\neq 0$. The function

$$f(x) = \frac{1}{x}$$

is defined for numbers $x \neq 0$, and is thus defined on the domain S. For this particular function, we have $f(1) = 1$, $f(2) = \frac{1}{2}$, $f(\frac{1}{2}) = 2$, and

$$f(\sqrt{2}) = \frac{1}{\sqrt{2}}.$$

In practice, functions are used to denote the dependence of one quantity with respect to another.

Example. The area inside a circle of radius r is given by the formula

$$A = \pi r^2.$$

Thus the area is a function of the radius r, and we can also write

$$A(r) = \pi r^2.$$

If the radius is 2, then the area inside a circle of radius 2 is given by

$$A(2) = \pi 2^2 = 4\pi.$$

Example. A car moves at a constant speed of 50 km/hr. If time is measured in hours, the distance traveled is a function of time, namely if we denote distance by s, then

$$s(t) = 50t.$$

The distance is the product of the speed by the time traveled. Thus after two hours, the distance is

$$s(2) = 50 \cdot 2 = 100 \text{ km}.$$

One final word before we pass to the exercises: There is no magic reason why we should always use the letter x to describe a function $f(x)$. Thus instead of speaking of the function $f(x) = 1/x$ we could just as well say $f(y) = 1/y$ or $f(q) = 1/q$. Unfortunately, the most neutral way of writing would be $f(\text{blank}) = 1/\text{blank}$, and this is really not convenient.

I, §3. EXERCISES

1. Let $f(x) = 1/x$. What is $f(-\frac{2}{3})$?

2. Let $f(x) = 1/x$ again. What is $f(2x + 1)$ (for any number x such that $x \neq -\frac{1}{2}$)?

3. Let $g(x) = |x| - x$. What is $g(1)$, $g(-1)$, $g(-54)$?

4. Let $f(y) = 2y - y^2$. What is $f(z)$, $f(w)$?

5. For what numbers could you define a function $f(x)$ by the formula

$$f(x) = \frac{1}{x^2 - 2}?$$

   What is the value of this function for $x = 5$?

6. For what numbers could you define a function $f(x)$ by the formula $f(x) = \sqrt[3]{x}$ (cube root of x)? What is $f(27)$?

7. Let $f(x) = x/|x|$, defined for $x \neq 0$. What is:
   (a) $f(1)$ (b) $f(2)$ (c) $f(-3)$ (d) $f(-\frac{4}{3})$

8. Let $f(x) = x + |x|$. What is:
   (a) $f(\frac{1}{2})$ (b) $f(2)$ (c) $f(-4)$ (d) $f(-5)$

9. Let $f(x) = 2x + x^2 - 5$. What is:
   (a) $f(1)$ (b) $f(-1)$ (c) $f(x + 1)$

10. For what numbers could you define a function $f(x)$ by the formula $f(x) = \sqrt[4]{x}$ (fourth root x)? What is $f(16)$?

11. A function (defined for all numbers) is said to be an **even** function if $f(x) = f(-x)$ for all x. It is said to be an **odd** function if $f(x) = -f(-x)$ for all x. Determine which of the following functions are odd or even.
(a) $f(x) = x$ (b) $f(x) = x^2$ (c) $f(x) = x^3$
(d) $f(x) = 1/x$ if $x \neq 0$, and $f(0) = 0$.

12. Let f be any function defined for all numbers. Show that the function $g(x) = f(x) + f(-x)$ is even. What about the function

$$h(x) = f(x) - f(-x),$$

is it even, odd, or neither?

I, §4. POWERS

In this section we just summarize some elementary arithmetic.

Let n be an integer $\geqq 1$ and let a be any number. Then a^n is the product of a with itself n times. For example, let $a = 3$. If $n = 2$, then $a^2 = 9$. If $n = 3$, then $a^3 = 27$. Thus we obtain a function which is called the n-th **power**. If f denotes this function, then $f(x) = x^n$.

We recall the rule

$$x^{m+n} = x^m x^n$$

for any number x and integers $m, n \geqq 1$.

Again, let n be an integer $\geqq 1$, and let a be a positive number. We define $a^{1/n}$ to be the unique positive number b such that $b^n = a$. (That there exists such a unique number b is taken for granted as part of the properties of numbers.) We get a function called the n-th **root**. Thus if f is the 4th root, then $f(16) = 2$ and $f(81) = 3$.

The n-th root function can also be defined at 0, the n-th root of 0 being 0 itself.

If a, b are two numbers $\geqq 0$ and n is an integer $\geqq 1$, then

$$(ab)^{1/n} = a^{1/n}b^{1/n}.$$

There is another useful and elementary rule. Let m, n be integers $\geqq 1$ and let a be a number $\geqq 0$. We define $a^{m/n}$ to be $(a^{1/n})^m$ which is also equal to $(a^m)^{1/n}$. This allows us to define fractional powers, and gives us a function

$$f(x) = x^{m/n}$$

defined for $x \geqq 0$.

We now come to powers with negative numbers or 0. We want to define x^a when a is a negative rational number or 0 and $x > 0$. We want the fundamental rule

$$x^{a+b} = x^a x^b$$

to be true. This means that we must define x^0 to be 1. For instance, since

$$2^3 = 2^{3+0} = 2^3 2^0,$$

we see from this example that the only way in which this equation holds is if $2^0 = 1$. Similarly, in general, if the relation

$$x^a = x^{a+0} = x^a x^0$$

is true, then x^0 must be equal to 1.

Suppose finally that a is a positive rational number, and let x be a number > 0. We **define**

$$\boxed{x^{-a} = \frac{1}{x^a}.}$$

Thus

$$2^{-3} = \frac{1}{2^3} = \frac{1}{8}, \qquad \text{and} \qquad 4^{-2/3} = \frac{1}{4^{2/3}}.$$

We observe that in this special case,

$$(4^{-2/3})(4^{2/3}) = 4^0 = 1.$$

In general,

$$x^a x^{-a} = x^0 = 1.$$

We are tempted to define x^a even when a is not a rational number. This is more subtle. For instance, it is absolutely meaningless to say that $2^{\sqrt{2}}$ is the product of 2 square root of 2 times itself. The problem of defining 2^a (or x^a) when a is not rational will be postponed to a later chapter. Until that chapter, when we deal with such a power, we shall assume that there is a function, written x^a, described as we have done above for rational numbers, and satisfying the fundamental relation

$$x^{a+b} = x^a x^b, \qquad x^0 = 1.$$

Example. We have a function $f(x) = x^{\sqrt{2}}$ defined for all $x > 0$. It is actually hard to describe its values for special numbers, like $2^{\sqrt{2}}$. It was unknown for a very long time whether $2^{\sqrt{2}}$ is a rational number or not. The solution (*it is not*) was found only in 1927 by the mathematician Gelfond, who became famous for solving a problem that was known to be very hard.

Warning. Do not confuse a function like x^2 and a function like 2^x. Given a number $c > 0$, we can view c^x as a function defined for all x. (It will be discussed in detail in Chapter VIII.) This function is called an **exponential function**. Thus 2^x and 10^x are exponential functions. We shall select a number

$$e = 2.718\ldots$$

and the exponential function e^x as having special properties which make it better than any other exponential function. The meaning of our use of the word "better" will be explained in Chapter VIII.

I, §4. EXERCISES

Find a^x and x^a for the following values of x and a.

1. $a = 2$ and $x = 3$
2. $a = 5$ and $x = -1$
3. $a = \frac{1}{2}$ and $x = 4$
4. $a = \frac{1}{3}$ and $x = 2$
5. $a = -\frac{1}{2}$ and $x = 4$
6. $a = 3$ and $x = 2$
7. $a = -3$ and $x = -1$
8. $a = -2$ and $x = -2$
9. $a = -1$ and $x = -4$
10. $a = -\frac{1}{2}$ and $x = 9$
11. If n is an odd integer like $1, 3, 5, 7, \ldots$, can you define an n-th root function for all numbers?

CHAPTER II

Graphs and Curves

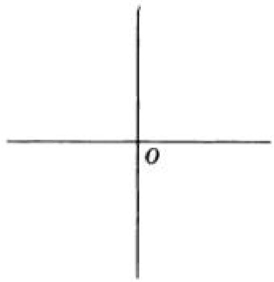
The ideas contained in this chapter allow us to translate certain statements backwards and forwards between the language of numbers and the language of geometry.

It is extremely basic for what follows, because we can use our geometric intuition to help us solve problems concerning numbers and functions, and conversely, we can use theorems concerning numbers and functions to yield results about geometry.

II, §1. COORDINATES

Once a unit length is selected, we can represent numbers as points on a line. We shall now extend this procedure to the plane, and to pairs of numbers.

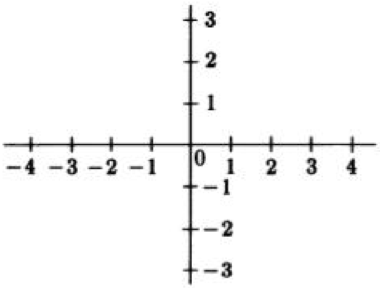
We visualize a horizontal line and a vertical line intersecting at an origin O.



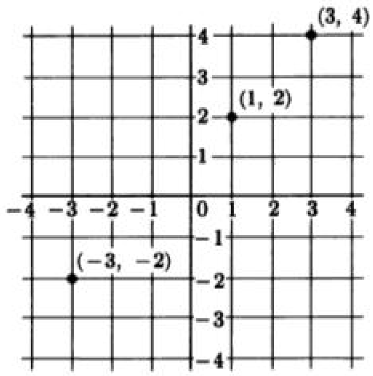
These lines will be called **coordinate axes** or simply **axes**.

We select a unit length and cut the horizontal line into segments of lengths 1, 2, 3,... to the left and to the right, and do the same to the vertical line, but up and down, as indicated in the next figure.

On the vertical line we visualize the points going below 0 as corresponding to the negative integers, just as we visualized points on the left of the horizontal line as corresponding to negative integers. We follow the same idea as that used in grading a thermometer, where the numbers below zero are regarded as negative. See figure.



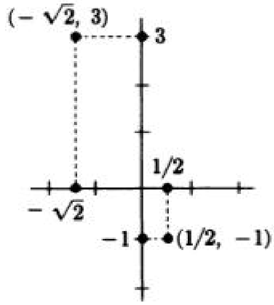
We can now cut the plane into squares whose sides have length 1.



We can describe each point where two lines intersect by a pair of integers. Suppose that we are given a pair of integers like (1, 2). We go to the right of the origin 1 unit and vertically up 2 units to get the point (1, 2) which has been indicated above. We have also indicated the point (3, 4). The diagram is just like a map.

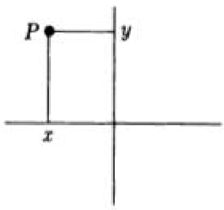
Furthermore, we could also use negative numbers. For instance, to describe the point $(-3, -2)$ we go to the left of the origin 3 units and vertically downwards 2 units.

There is actually no reason why we should limit ourselves to points which are described by integers. For instance we can also have the point $(\frac{1}{2}, -1)$ and the point $(-\sqrt{2}, 3)$ as on the figure below.



We have not drawn all the squares on the plane. We have drawn only the relevant lines to find our two points.

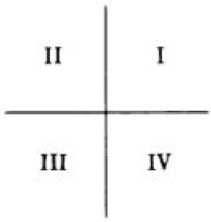
In general, if we take any point P in the plane and draw the perpendicular lines to the horizontal axis and to the vertical axis, we obtain two numbers x, y as in the figure below.



The perpendicular line from P to the horizontal axis determines a number x which is negative in the figure because it lies to the left of the origin. The number y determined by the perpendicular from P to the vertical axis is positive because it lies above the origin. The two numbers x, y are called the **coordinates** of the point P, and we can write $P = (x, y)$.

Every pair of numbers (x, y) determines a point of the plane. We find the point by going a distance x from the origin O in the horizontal direction and then a distance y in the vertical direction. If x is positive we go to the right of O. If x is negative, we go to the left of O. If y is positive we go vertically upwards, and if y is negative we go vertically downwards. The coordinates of the origin are $(0, 0)$. We usually call the horizontal axis the **x-axis** and the vertical axis the **y-axis**. If a point P is described by two numbers, say $(5, -10)$, it is customary to call the first number its x-coordinate and the second number its y-coordinate. Thus 5 is the x-coordinate, and -10 the y-coordinate of our point. Of course, we could use other letters besides x and y, for instance t and s, or u and v.

Our two axes separate the plane into four **quadrants** which are numbered as indicated in the figure:



If (x, y) is a point in the first quadrant, then both x and y are > 0. If (x, y) is a point in the fourth quadrant, then $x > 0$ but $y < 0$.

II, §1. EXERCISES

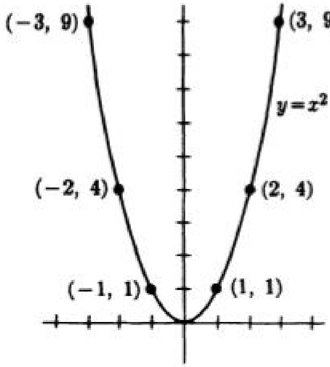
1. Plot the following points: $(-1, 1)$, $(0, 5)$, $(-5, -2)$, $(1, 0)$.
2. Plot the following points: $(\frac{1}{2}, 3)$, $(-\frac{1}{3}, -\frac{1}{2})$, $(\frac{4}{3}, -2)$, $(-\frac{1}{4}, \frac{1}{2})$.
3. Let (x, y) be the coordinates of a point in the second quadrant. Is x positive or negative? Is y positive or negative?
4. Let (x, y) be the coordinates of a point in the third quadrant. Is x positive or negative? Is y positive or negative?
5. Plot the following points: $(1.2, -2.3)$, $(1.7, 3)$.
6. Plot the following points: $(-2.5, \frac{1}{3})$, $(-3.5, \frac{5}{4})$.
7. Plot the following points: $(1.5, -1)$, $(-1.5, -1)$.

II, §2. GRAPHS

Let f be a function. We define the **graph** of f to be the collection of all pairs of numbers $(x, f(x))$ whose first coordinate is any number for which f is defined and whose second coordinate is the value of the function at the first coordinate.

For example, the graph of the function $f(x) = x^2$ consists of all pairs (x, y) such that $y = x^2$. In other words, it is the collection of all pairs (x, x^2), like $(1, 1)$, $(2, 4)$, $(-1, 1)$, $(-3, 9)$, etc.

Since each pair of numbers corresponds to a point on the plane (once a system of axes and a unit length have been selected), we can view the graph of f as a collection of points in the plane. The graph of the function $f(x) = x^2$ has been drawn in the following figure, together with the points which we gave above as examples.



To determine the graph, we plot a lot of points making a table giving the x- and y-coordinates.

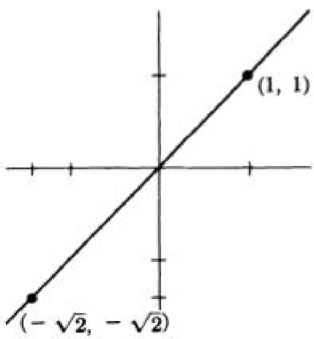
x	$f(x)$
1	1
2	4
3	9
$\frac{1}{2}$	$\frac{1}{4}$

x	$f(x)$
-1	1
-2	4
-3	9
$-\frac{1}{2}$	$\frac{1}{4}$

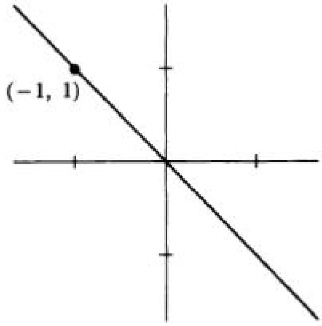
At this stage of the game there is no other way for you to determine the graph of a function other than this trial and error method. Later, we shall develop techniques which give you greater efficiency in doing it.

We shall now give several examples of graphs of functions which occur very frequently in the sequel.

Example 1. Consider the function $f(x) = x$. The points on its graph are of type (x, x). The first coordinate must be equal to the second. Thus $f(1) = 1$, $f(-\sqrt{2}) = -\sqrt{2}$, etc. The graph looks like this:

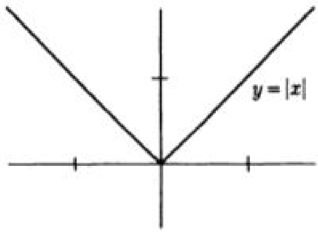


Example 2. Let $f(x) = -x$. Its graph looks this this:



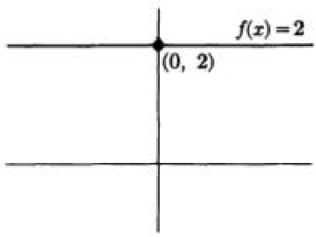
Observe that the graphs of the preceding two functions are straight lines. We shall study the general case of a straight line later.

Example 3. Let $f(x) = |x|$. When $x \geqq 0$, we know that $f(x) = x$. When $x \leqq 0$, we know that $f(x) = -x$. Hence the graph of $|x|$ is obtained by combining the preceding two, and looks like this:



All values of $f(x)$ are $\geqq 0$, whether x is positive or negative.

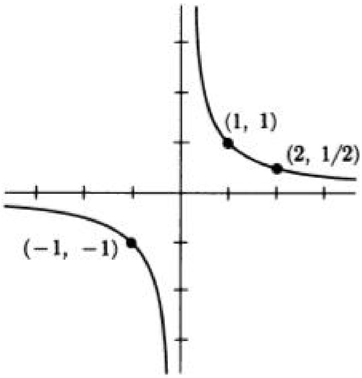
Example 4. There is an even simpler type of function than the ones we have just looked at, namely the constant functions. For instance, we can define a function f such that $f(x) = 2$ for all numbers x. In other words, we associate the number 2 to any number x. It is a very simple association, and the graph of this function is a horizontal line, intersecting the vertical axis at the point $(0, 2)$.



If we took the function $f(x) = -1$, then the graph would be a horizontal line intersecting the vertical axis at the point $(0, -1)$.

In general, let c be a fixed number. The graph of any function $f(x) = c$ is the horizontal line intersecting the vertical axis at the point $(0, c)$. The function $f(x) = c$ is called a **constant** function.

Example 5. The last of our examples is the function $f(x) = 1/x$ (defined for $x \neq 0$). By plotting a few points of the graph, you will see that it looks like this:



For instance, you can plot the following points:

x	$1/x$
1	1
2	$\frac{1}{2}$
3	$\frac{1}{3}$
$\frac{1}{2}$	2
$\frac{1}{3}$	3

x	$1/x$
-1	-1
-2	$-\frac{1}{2}$
-3	$-\frac{1}{3}$
$-\frac{1}{2}$	-2
$-\frac{1}{3}$	-3

As x becomes very large positive, $1/x$ becomes very small. As x approaches 0 from the right, $1/x$ becomes very large. A similar phenomenon occurs when x approaches 0 from the left; then x is negative and $1/x$ is negative. Hence in that case, $1/x$ is very large negative.

In trying to determine how the graph of a function looks, you can already watch for the following:

The points at which the graph intersects the two coordinate axes.

What happens when x becomes very large positive and very large negative.

On the whole, however, in working out the exercises, your main technique is just to plot a lot of points until it becomes clear to you what the graph looks like.

II, §2. EXERCISES

Sketch the graphs of the following functions and plot at least three points on each graph. In each case we give the value of the function at x.

1. $x + 1$
2. $2x$
3. $3x$
4. $4x$
5. $2x + 1$
6. $5x + \frac{1}{2}$
7. $\frac{x}{2} + 3$
8. $-3x + 2$
9. $2x^2 - 1$
10. $-3x^2 + 1$
11. x^3
12. x^4
13. $\sqrt{x}$
14. $x^{-1/2}$
15. $2x + 1$
16. $x + 3$
17. $|x| + x$
18. $|x| + 2x$
19. $-|x|$
20. $-|x| + x$
21. $\frac{1}{x+2}$
22. $\frac{1}{x-2}$
23. $\frac{1}{x+3}$
24. $\frac{1}{x-3}$
25. $\frac{2}{x-2}$
26. $\frac{2}{x+2}$
27. $\frac{2}{x}$
28. $\frac{-2}{x+5}$
29. $\frac{3}{x+1}$
30. $\frac{x}{|x|}$

(In Exercises 13, 14, and 21 through 30, the functions are not defined for all values of x.)

31. Sketch the graph of the function $f(x)$ such that:
$f(x) = 0$ if $x \leqq 0$. $f(x) = 1$ if $x > 0$.

32. Sketch the graph of the function $f(x)$ such that:
$f(x) = x$ if $x < 0$. $f(0) = 2$. $f(x) = x$ if $x > 0$.

33. Sketch the graph of the function $f(x)$ such that:
$f(x) = x^2$ if $x < 0$. $f(x) = x$ if $x \geqq 0$.

34. Sketch the graph of the function $f(x)$ such that:
$f(x) = |x| + x$ if $-1 \leqq x \leqq 1$.
$f(x) = 3$ if $x > 1$. [$f(x)$ is not defined for other values of x.]

35. Sketch the graph of the function $f(x)$ such that:
$f(x) = x^3$ if $x \leqq 0$. $f(x) = 1$ if $0 < x < 2$.
$f(x) = x^2$ if $x \geqq 2$.

36. Sketch the graph of the function $f(x)$ such that:
$f(x) = x$ if $0 < x \leqq 1$. $f(x) = x - 1$ if $1 < x \leqq 2$.
$f(x) = x - 2$ if $2 < x \leqq 3$. $f(x) = x - 3$ if $3 < x \leqq 4$.

[We leave $f(x)$ undefined for other values of x, but try to define it yourself in such a way as to preserve the symmetry of the graph.]

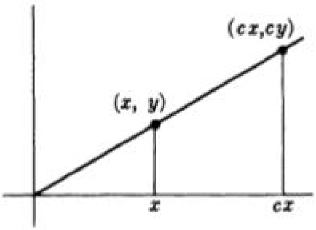
II, §3. THE STRAIGHT LINE

One of the most basic types of functions is the type whose graph represents a straight line. We have already seen that the graph of the function $f(x) = x$ is a straight line. If we take $f(x) = 2x$, then the line slants up much more steeply, and even more so for $f(x) = 3x$. The graph of the function $f(x) = 10{,}000x$ would look almost vertical. In general, let a be a positive number $\neq 0$. Then the graph of the function

$$f(x) = ax$$

represents a straight line. The point $(2, 2a)$ lies on the line because $f(2) = 2a$. The point $(\sqrt{2}, \sqrt{2}a)$ also lies on the line, and if c is any number, the point (c, ca) lies on the line. The (x, y) coordinates of these points are obtained by making a similarity transformation, starting with the coordinates $(1, a)$ and multiplying them by some number c.

We can visualize this procedure by means of similar triangles. In the figure below, we have a straight line. If we select a point (x, y) on the line and drop the perpendicular from this point to the x-axis, we obtain a right triangle.



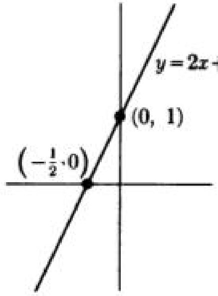
If x is the length of the base of the smaller triangle in the figure, and y its height, and if cx is the length of the base of the bigger triangle, then cy is the height of the bigger triangle: The smaller triangle is similar to the bigger one.

If a is a number < 0, then the graph of the function $f(x) = ax$ is also a straight line, which slants up to the left. For instance, the graphs of

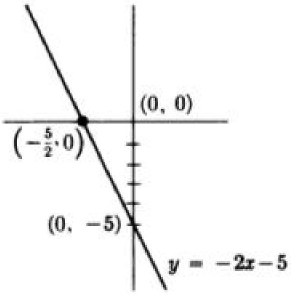
$$f(x) = -x \qquad \text{or} \qquad f(x) = -2x.$$

We now give examples of more general lines, not passing through the origin.

Example 1. Let $g(x) = 2x + 1$. When $x = 0$, then $g(x) = 1$. When $g(x) = 0$, then $x = -\frac{1}{2}$. The graph looks as on the following figure.



Example 2. Let $g(x) = -2x - 5$. When $x = 0$, then $g(x) = -5$. When $g(x) = 0$, then $x = -\frac{5}{2}$. The graph looks like this:



We shall frequently speak of a function $f(x) = ax + b$ as a straight line (although of course, it is its graph which is a straight line).

The number a which is the coefficient of x is called the **slope** of the line. It determines how much the line is slanted. As we have already seen in examples, when the slope is positive, the line is slanted to the right, and when the slope is negative, the line is slanted to the left. The relationship $y = ax + b$ is also called the **equation** of the line. It gives us the relation between the x- and y-coordinates of a point on the line.

Let $f(x) = ax + b$ be a straight line, and let (x_1, y_1) and (x_2, y_2) be two points of the line. It is easy to find the slope of the line in terms of the coordinates of these two points. By definition, we know that

$$y_1 = ax_1 + b$$

and

$$y_2 = ax_2 + b.$$

Subtracting, we get

$$y_2 - y_1 = ax_2 - ax_1 = a(x_2 - x_1).$$

Consequently, if the two points are distinct, $x_2 \neq x_1$, then we can divide by $x_2 - x_1$ and obtain

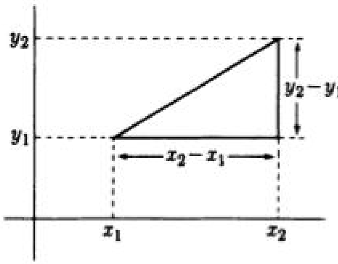
$$\boxed{\text{slope of line} = a = \frac{y_2 - y_1}{x_2 - x_1}.}$$

This formula gives us the slope in terms of the coordinates of two distinct points on the line.

Geometrically, our quotient

$$\frac{y_2 - y_1}{x_2 - x_1}$$

is simply the ratio of the vertical side and horizontal side of the triangle in the next diagram:



In general, let a be a number, and (x_1, y_1) some point.

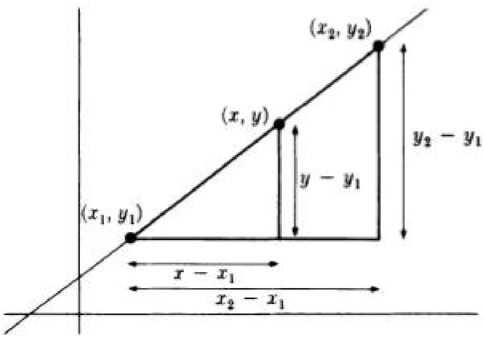
We wish to find the equation of the line having slope equal to a, and passing through the point (x_1, y_1).

The condition that a point (x, y) with $x \neq x_1$ be on this line is equivalent with the condition that

$$\frac{y - y_1}{x - x_1} = a.$$

Thus the equation of the desired line is

$$\boxed{y - y_1 = a(x - x_1).}$$



Example 3. Let $(1, 2)$ and $(2, -1)$ be the two points. What is the slope of the line between them? What is the equation of the line?

We first find the slope. We have:

$$\text{slope} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{-1 - 2}{2 - 1} = -3.$$

The line must pass through the given point $(1, 2)$. Hence its equation is

$$y - 2 = -3(x - 1).$$

This is a correct answer. Sometimes it may be useful to put the equation in the form

$$y = -3x + 5,$$

but it is equally valid to leave it in the first form.

Observe that it does not matter which point we call (x_1, y_1) and which we call (x_2, y_2). We would get the same answer for the slope.

We can also determine the equation of a line provided we know the slope and one point.

Example 4. Find the equation of the line having slope -7 and passing through the point $(-1, 2)$.

The equation is

$$y - 2 = -7(x + 1).$$

Example 5. In general, let (x_1, y_1) and (x_2, y_2) be two distinct points with $x_1 \neq x_2$. We wish to find the equation of the line passing through these two points. Its slope must then be equal to

$$\frac{y_2 - y_1}{x_2 - x_1}.$$

Therefore the equation of the line can be expressed by the formula

$$\boxed{\frac{y - y_1}{x - x_1} = \frac{y_2 - y_1}{x_2 - x_1}}$$

for all points (x, y) such that $x \neq x_1$, or for all points by

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

Finally, we should mention vertical lines. These cannot be represented by equations of type $y = ax + b$. Suppose that we have a vertical line intersecting the x-axis at the point $(2, 0)$. The y-coordinate of any point on the line can be arbitrary. Thus the equation of the line is simply $x = 2$. In general, the equation of the vertical line intersecting the x-axis at the point $(c, 0)$ is $x = c$.

We can find the point of intersection of two lines by solving simultaneously two linear equations.

Example 6. Find the point of intersection of the two lines

$$y = 3x - 5 \quad \text{and} \quad y = -4x + 1.$$

We solve

$$3x - 5 = -4x + 1$$

or equivalently, $7x = 6$. This gives $x = \frac{6}{7}$, whence

$$y = 3 \cdot \frac{6}{7} - 5 = \frac{18}{7} - 5.$$

Hence the common point is

$$\left(\frac{6}{7}, \frac{18}{7} - 5\right).$$

II, §3. EXERCISES

Sketch the graphs of the following lines:

1. $y = -2x + 5$
2. $y = 5x - 3$
3. $y = \frac{x}{2} + 7$
4. $y = -\frac{x}{3} + 1$

What is the equation of the line passing through the following points?

5. $(-1, 1)$ and $(2, -7)$
6. $(3, \frac{1}{2})$ and $(4, -1)$
7. $(\sqrt{2}, -1)$ and $(\sqrt{2}, 1)$
8. $(-3, -5)$ and $(\sqrt{3}, 4)$

What is the equation of the line having the given slope and passing through the given point?

9. slope 4 and point $(1, 1)$
10. slope -2 and point $(\frac{1}{2}, 1)$
11. slope $-\frac{1}{2}$ and point $(\sqrt{2}, 3)$
12. slope $\sqrt{3}$ and point $(-1, 5)$

Sketch the graphs of the following lines:

13. $x = 5$
14. $x = -1$
15. $x = -3$
16. $y = -4$
17. $y = 2$
18. $y = 0$

What is the slope of the line passing through the following points?

19. $(1, \frac{1}{2})$ and $(-1, 1)$
20. $(\frac{1}{4}, 1)$ and $(\frac{1}{2}, -1)$

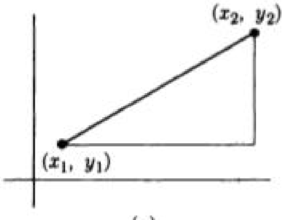
21. $(2, 3)$ and $(\sqrt{2}, 1)$
22. $(\sqrt{3}, 1)$ and $(3, 2)$

What is the equation of the line passing through the following points?

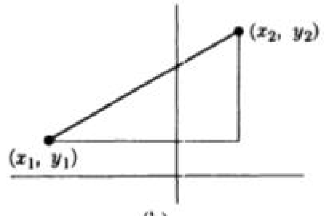
23. $(\pi, 1)$ and $(\sqrt{2}, 3)$
24. $(\sqrt{2}, 2)$ and $(1, \pi)$
25. $(-1, 2)$ and $(\sqrt{2}, -1)$
26. $(-1, \sqrt{2})$ and $(-2, -3)$
27. Sketch the graphs of the following lines:
    (a) $y = 2x$ (b) $y = 2x + 1$ (c) $y = 2x + 5$
    (d) $y = 2x - 1$ (e) $y = 2x - 5$
28. Two straight lines are said to be **parallel** if they have the same slope. Let $y = ax + b$ and $y = cx + d$ be the equations of two straight lines with $b \neq d$.
    (a) If they are parallel, show that they have no point in common.
    (b) If they are not parallel, show that they have exactly one point in common.
29. Find the common point of the following pairs of lines:
    (a) $y = 3x + 5$ and $y = 2x + 1$ (b) $y = 3x - 2$ and $y = -x + 4$
    (c) $y = 2x$ and $y = -x + 2$ (d) $y = x + 1$ and $y = 2x + 7$

II, §4. DISTANCE BETWEEN TWO POINTS

Let (x_1, y_1) and (x_2, y_2) be two points in the plane, for instance as in the following diagrams.



(a)



(b)

We can then make up a right triangle. By the Pythagoras theorem, the length of the line segment joining our two points can be determined from the lengths of the two sides. The square of the bottom side is $(x_2 - x_1)^2$, which is equal to $(x_1 - x_2)^2$.

The square of the length of the vertical side is $(y_2 - y_1)^2$, which is equal to $(y_1 - y_2)^2$. If L denotes the length of the line segment, then by Pythagoras,

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

and consequently,

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

Example 1. Let the two points be $(1, 2)$ and $(1, 3)$. Then the length of the line segment between them is

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

The length L is also called the **distance** between the two points.

Example 2. Find the distance between the points $(-1, 5)$ and $(4, -3)$.

The distance is

$$\sqrt{(4-(-1))^2 + (-3-5)^2} = \sqrt{89}.$$

II, §4. EXERCISES

Find the distance between the following points:

1. The points $(-3, -5)$ and $(1, 4)$
2. The points $(1, 1)$ and $(0, 2)$
3. The points $(-1, 4)$ and $(3, -2)$
4. The points $(1, -1)$ and $(-1, 2)$
5. The points $(\frac{1}{2}, 2)$ and $(1, 1)$
6. Find the coordinates of the fourth corner of a rectangle, three of whose corners are $(-1, 2)$, $(4, 2)$, $(-1, -3)$.
7. What are the lengths of the sides of the rectangle in Exercise 6?
8. Find the coordinates of the fourth corner of a rectangle, three of whose corners are $(-2, -2)$, $(3, -2)$, $(3, 5)$.
9. What are the lengths of the sides of the rectangle in Exercise 8?
10. If x, y are numbers, define the distance between these two numbers to be $|x - y|$. Show that this is the same as the distance between the points $(x, 0)$ and $(y, 0)$ in the plane.

II, §5. CURVES AND EQUATIONS

Let $F(x, y)$ be an expression involving a pair of numbers (x, y). Let c be a number. We consider the equation

$$F(x, y) = c.$$

Definition. The **graph** of the equation is the collection of points (a, b) in the plane satisfying the equation, that is such that

$$F(a, b) = c$$

This graph is also known as a **curve**, and we will usually not make a distinction between the equation

$$F(x, y) = c$$

and the curve which represents the equation.

For example,

$$x + y = 2$$

is the equation of a straight line, and its graph is the straight line. We shall study below important examples of equations which arise frequently.

If f is a function, then we can form the expression $y - f(x)$, and the graph of the **equation**

$$y - f(x) = 0$$

is none other than the graph of the **function** f as we discussed it in §2.

You should observe that there are equations of type

$$F(x, y) = c$$

which are not obtained from a function $y = f(x)$, i.e. from an equation

$$y - f(x) = 0.$$

For instance, the equation $x^2 + y^2 = 1$ is such an equation.

We shall now study important examples of graphs of equations

$$F(x, y) = 0 \qquad \text{or} \qquad F(x, y) = c.$$

II, §6. THE CIRCLE

The expression $F(x, y) = x^2 + y^2$ has a simple geometric interpretation. By Pythagoras' theorem, it is the square of the distance of the point (x, y) from the origin $(0, 0)$. Thus the points (x, y) satisfying the equation

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

are simply those points whose distance from the origin is 1. They form the circle of radius 1, with center at the origin.

Similarly, the points (x, y) satisfying the equation

$$x^2 + y^2 = 4$$

are those points whose distance from the origin is 2. They constitute the circle of radius 2. In general, if c is any number > 0, then the graph of the equation

$$x^2 + y^2 = c^2$$

is the circle of radius c, with center at the origin.

We have already remarked that the **equation**

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

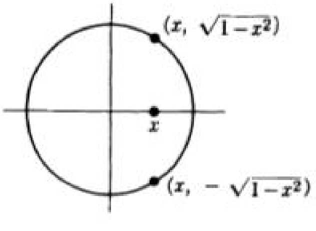
or $x^2 + y^2 - 1 = 0$ is not of the type $y - f(x) = 0$. However, we can write our equation in the form

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

For any value of x between -1 and $+1$, we can solve for y and get

$$y = \sqrt{1 - x^2} \qquad \text{or} \qquad y = -\sqrt{1 - x^2}.$$

If $x \neq 1$ and $x \neq -1$, then we get two values of y for each value of x. Geometrically, these two values correspond to the points indicated on the following diagram.



There is a function, defined for $-1 \leqq x \leqq 1$, such that

$$f(x) = \sqrt{1 - x^2},$$

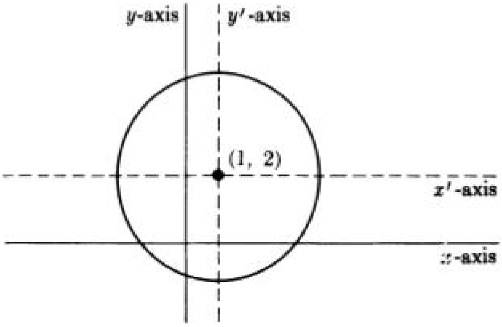
and the graph of this function is the upper half of our circle. Similarly, there is another function

$$g(x) = -\sqrt{1 - x^2},$$

also defined for $-1 \leqq x \leqq 1$, whose graph is the lower half of the circle. Neither of these functions is defined for other values of x.

We now ask for the equation of the circle whose center is $(1, 2)$ and whose radius has length 3. It consists of the points (x, y) whose distance from $(1, 2)$ is 3. These are the points satisfying the equation

$$(x - 1)^2 + (y - 2)^2 = 9.$$



The graph of this equation has been drawn above. We may also put

$$x' = x - 1 \qquad \text{and} \qquad y' = y - 2$$

In the new coordinate system (x', y') the equation of the circle is then

$$x'^2 + y'^2 = 9.$$

We have drawn the (x', y')-axes as dotted lines in the figure.

To pick another example, we wish to determine those points at a distance 2 from the point $(-1, -3)$. They are the points (x, y) satisfying the equation

$$(x - (-1))^2 + (y - (-3))^2 = 4$$

or, in other words,

$$(x + 1)^2 + (y + 3)^2 = 4.$$

(Observe carefully the cancellation of minus signs!) Thus the graph of this equation is the circle of radius 2 and center $(-1, -3)$.

In general, let a, b be two numbers and r a number > 0. Then the circle of radius r and center (a, b) is the graph of the equation

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

We may put

$$x' = x - a \qquad \text{and} \qquad y' = y - b.$$

Then in the new coordinates x', y' the equation of the circle is

$$x'^2 + y'^2 = r^2.$$

Completing the square

Example. Suppose we are given an equation

$$x^2 + y^2 + 2x - 3y - 5 = 0,$$

where x^2 and y^2 occur with the same coefficient 1. We wish to see that this is the equation of a circle, and we use the method of **completing the square**, which we now review.

We want the equation to be of the form

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

because then we know immediately that it represents a circle centered at (a, b) and of radius r. Thus we need $x^2 + 2x$ to be the first two terms of the expansion

$$(x - a)^2 = x^2 - 2ax + a^2.$$

Similarly, we need $y^2 - 3y$ to be the first two terms of the expansion

$$(y - b)^2 = y^2 - 2by + b^2.$$

This means that $a = -1$ and $b = 3/2$. Then

$$x^2 + 2x + y^2 - 3y - 5 = (x + 1)^2 - 1 + \left(y - \frac{3}{2}\right)^2 - \frac{9}{4} - 5.$$

Thus $x^2 + y^2 + 2x - 3y - 5 = 0$ is equivalent with

$$(x + 1)^2 + \left(y - \frac{3}{2}\right)^2 = 5 + 1 + \frac{9}{4} = \frac{33}{4}.$$

Consequently our given equation is the equation of a circle of radius $\sqrt{33/4}$, with center at $(-1, 3/2)$.

II, §6. EXERCISES

Sketch the graph of the following equations:

1. (a) $(x - 2)^2 + (y + 1)^2 = 25$ (b) $(x - 2)^2 + (y + 1)^2 = 4$
   (c) $(x - 2)^2 + (y + 1)^2 = 1$ (d) $(x - 2)^2 + (y + 1)^2 = 9$

2. (a) $x^2 + (y - 1)^2 = 9$ (b) $x^2 + (y - 1)^2 = 4$
   (c) $x^2 + (y - 1)^2 = 25$ (d) $x^2 + (y - 1)^2 = 1$

3. (a) $(x+1)^2 + y^2 = 1$ (b) $(x+1)^2 + y^2 = 4$
(c) $(x+1)^2 + y^2 = 9$ (d) $(x+1)^2 + y^2 = 25$

4. $x^2 + y^2 - 2x + 3y - 10 = 0$

5. $x^2 + y^2 + 2x - 3y - 15 = 0$

6. $x^2 + y^2 + x - 2y = 16$

7. $x^2 + y^2 - x + 2y = 25$

II, §7. DILATIONS AND THE ELLIPSE

Dilations

Before studying the ellipse, we have to make some remarks on "stretching," or to use a more standard word, dilations.

Let (x, y) be a point in the plane. Then $(2x, 2y)$ is the point obtained by stretching both its coordinates by a factor of 2, as illustrated on Fig. 1, where we have also drawn $(3x, 3y)$ and $(\frac{1}{2}, \frac{1}{2}y)$.

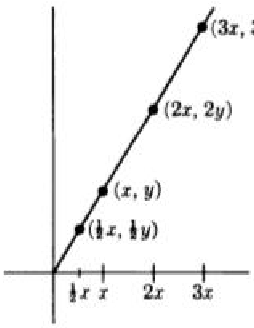


Figure 1

Definition. In general, if $c > 0$ is a positive number, we call (cx, cy) the **dilation** of (x, y) by a factor c.

Example. Let

$$u^2 + v^2 = 1$$

be the equation of the circle of radius 1. Put

$$x = cu \quad \text{and} \quad y = cv.$$

Then

$$u = x/c \quad \text{and} \quad v = y/c.$$

Hence x and y satisfy the equation

$$\frac{x^2}{c^2} + \frac{y^2}{c^2} = 1,$$

or equivalently,

$$x^2 + y^2 = c^2.$$

The set of points (x, y) satisfying this equation is the circle of radius c. Thus we may say:

The dilation of the circle of radius 1 by a factor of $c > 0$ is the circle of radius c.

This is illustrated on Fig. 2, with $c = 3$.

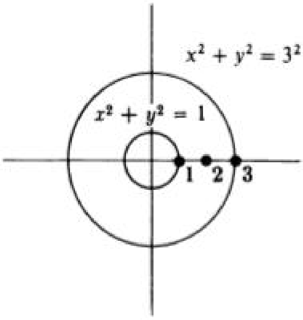


Figure 2

The Ellipse

There is no reason why we should dilate the first and second coordinates by the same factor. We may use different factors. For instance, if we put

$$x = 2u \qquad \text{and} \qquad y = 3v$$

we are dilating the first coordinate by a factor of 2, and we are dilating the second coordinate by a factor of 3. In that case, suppose that (u, v) is a point on the circle of radius 1, in other words suppose we have

$$u^2 + v^2 = 1.$$

Then (x, y) satisfies the equation

$$\frac{x^2}{4} + \frac{y^2}{9} = 1.$$

We interpret this as the equation of a "stretched out circle," as shown on Fig. 3.

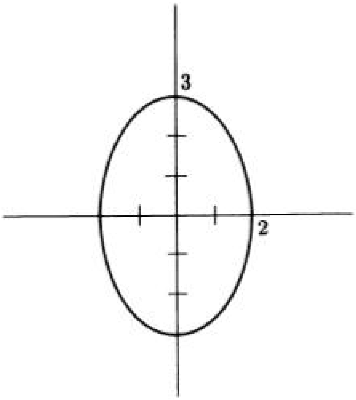


Figure 3

More generally, let a, b be numbers > 0. Let us put

$$x = au \qquad \text{and} \qquad y = bv.$$

If (u, v) satisfies

$$(*) \qquad u^2 + v^2 = 1,$$

then (x, y) satisfies

$$(**) \qquad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

Conversely, we may put $u = x/a$ and $v = y/b$ to see that the points satisfying equation $(*)$ correspond to the points of $(**)$ under this transformation, and vice versa.

Definition. An **ellipse** is the set of points satisfying an equation $(**)$ in some coordinate system of the plane. We have just seen that an ellipse is a dilated circle, by means of a dilation by factors $a, b > 0$ in the first and second coordinates respectively.

Example. Sketch the graph of the ellipse

$$\frac{x^2}{4} + \frac{y^2}{25} = 1.$$

This ellipse is a dilated circle by the factors 2 and 5, respectively. Note that

when $x = 0$ we have $\dfrac{y^2}{25} = 1$, so $y^2 = 25$ and $y = \pm 5$.

Also

$$\text{when} \quad y = 0, \quad \text{we have} \quad \frac{x^2}{4} = 1, \quad \text{so} \quad x^2 = 4 \quad \text{and} \quad x = \pm 2.$$

Hence the graph of the ellipse looks like Fig. 4.

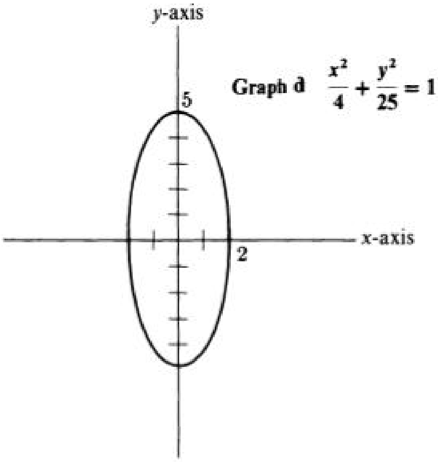


Figure 4

Example. Sketch the graph of the ellipse

$$\frac{(x-1)^2}{25} + \frac{(y+2)^2}{4} = 1.$$

In this case, let us put

$$x' = x - 1 \quad \text{and} \quad y' = y + 2.$$

We know that in (u, v) coordinates

$$u^2 + v^2 = 1$$

is the equation of a circle with center $(1, -2)$ and radius 1. Next we put

$$u = \frac{x'}{5} \quad \text{and} \quad v = \frac{y'}{2}.$$

The original equation is of the form

$$\frac{x'^2}{5^2} + \frac{y'^2}{2^2} = 1,$$

which in terms of u and v can be written

$$u^2 + v^2 = 1.$$

Thus our ellipse is obtained from the circle $u^2 + v^2 = 1$ by the dilation

$$u = x'/5 \qquad \text{and} \qquad v = y'/2,$$

or equivalently,

$$x' = 5u \qquad \text{and} \qquad y' = 2v.$$

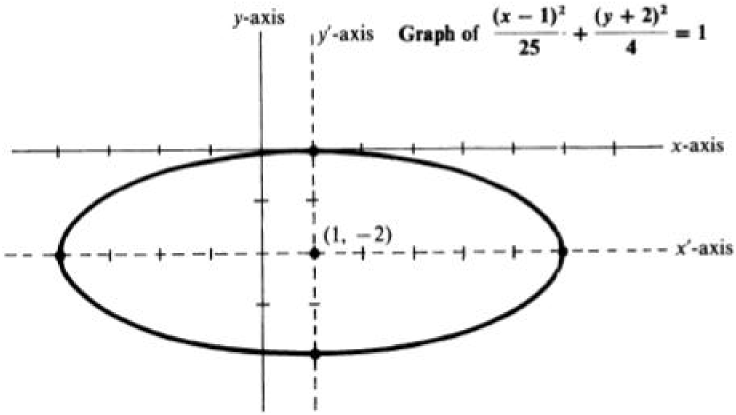
The easiest way to sketch its graph is to draw the new coordinate system with coordinates x', y'. To find the intercepts of the ellipse with these new axes, we see that when $y' = 0$, then

$$\frac{x'^2}{5^2} = 1, \qquad \text{so that} \qquad x' = \pm 5.$$

Similarly, when $x' = 0$, then

$$\frac{y'^2}{2^2} = 1, \qquad \text{so that} \qquad y' = \pm 2.$$

Thus the graph looks like:



II, §7. EXERCISES

Sketch the graphs of the following curves.

1. $\frac{x^2}{9} + \frac{y^2}{16} = 1$

2. $\frac{x^2}{4} + \frac{y^2}{9} = 1$

3. $\frac{x^2}{5} + \frac{y^2}{16} = 1$

4. $\frac{x^2}{4} + \frac{y^2}{25} = 1$

5. $\frac{(x-1)^2}{9} + \frac{(y+2)^2}{16} = 1$

6. $4x^2 + 25y^2 = 100$

7. $\frac{(x+1)^2}{3} + \frac{(y+2)^2}{4} = 1$

8. $25x^2 + 16y^2 = 400$

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

II, §8. THE PARABOLA

A **parabola** is a curve which is the graph of a function

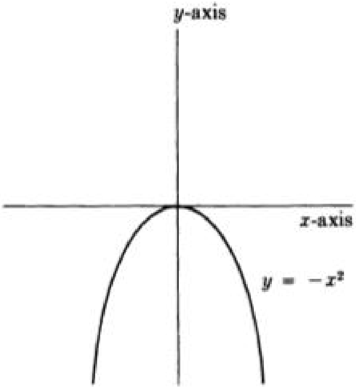
$$y = ax^2$$

in some coordinate system, with $a \neq 0$.

Example. We have already seen what the graph of the function $y = x^2$ looks like. Consider now

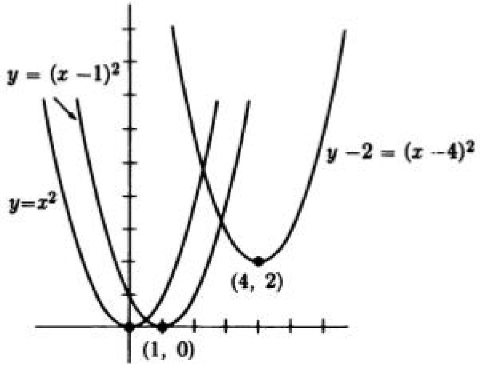
$$y = -x^2.$$

Then by symmetry, you can easily see that the graph looks as on the figure.



Suppose that we graph the equation $y = (x-1)^2$. We shall find that it looks exactly the same, but as if the origin were placed at the point $(1, 0)$.

Similarly, the curve $y - 2 = (x-4)^2$ looks again like $y = x^2$ except that the whole curve has been moved as if the origin were the point $(4, 2)$. The graphs of these equations have been drawn on the next diagram.



We can formalize these remarks as follows. Suppose that in our given coordinate system we pick a point (a, b) as a new origin. We let new coordinates be $x' = x - a$ and $y' = y - b$. Thus when $x = a$ we have $x' = 0$ and when $y = b$ we have $y' = 0$. If we have a curve

$$y' = x'^2$$

in the new coordinate system whose origin is at the point (a, b), then it gives rise to the equation

$$(y - b) = (x - a)^2$$

in terms of the old coordinate system. This type of curve is known as a **parabola**.

We can apply the same technique of completing the square that we did for the circle.

Example. What is the graph of the equation

$$2y - x^2 - 4x + 6 = 0?$$

Completing the square, we can write

$$x^2 + 4x = (x + 2)^2 - 4.$$

Thus our equation can be rewritten

$$2y = (x + 2)^2 - 10$$

or

$$2(y + 5) = (x + 2)^2.$$

We choose a new coordinate system

$$x' = x + 2 \qquad \text{and} \qquad y' = y + 5$$

so that our equation becomes

$$2y' = x'^2 \qquad \text{or} \qquad y' = \tfrac{1}{2}x'^2.$$

This is a function whose graph you already know, and whose sketch we leave to you.

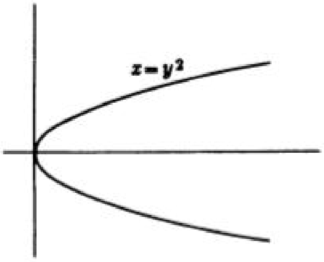
We remark that if we have an equation

$$x - y^2 = 0$$

or

$$x = y^2,$$

then we get a parabola which is tilted horizontally.



We can then apply the technique of changing the coordinate system to see what the graph of a more general equation is like.

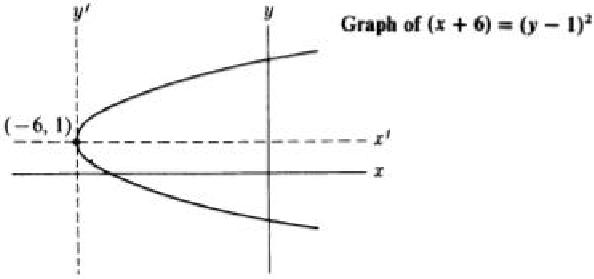
Example. Sketch the graph of

$$x - y^2 + 2y + 5 = 0.$$

We can write this equation in the form

$$(x + 6) = (y - 1)^2$$

and hence its graph looks like this:



Suppose we are given the equation of a parabola

$$y = f(x) = ax^2 + bx + c,$$

with $a \neq 0$. We wish to determine where this parabola intersects the x-axis. These are the values for which $f(x) = 0$ and are called the **roots** of f. It is shown in high school that the roots of f are given by the **quadratic formula**:

$$\boxed{x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.}$$

You should read this formula out loud enough times so that it is memorized, just like the multiplication table. *It should be used automatically*, without further thinking, to find the roots of a quadratic equation.

Example. We want to find the roots of the equation

$$-2x^2 + 5x - 1 = 0.$$

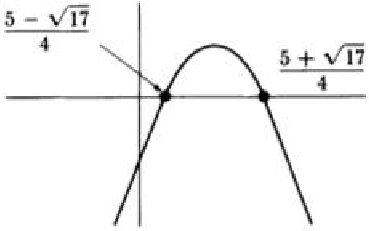
The roots are

$$x = \frac{-5 \pm \sqrt{25 - 8}}{2(-2)} = \frac{-5 \pm \sqrt{17}}{-4} = \frac{5 \pm \sqrt{17}}{4}.$$

Thus the two roots are

$$\frac{5 + \sqrt{17}}{4} \quad \text{and} \quad \frac{5 - \sqrt{17}}{4}.$$

These are the two points where the parabola $y = -2x^2 + 5x - 1$ crosses the x-axis, and its graph is shown on the figure.



Proof of the quadratic formula. We shall now give the proof of the quadratic formula, to convince you that it is true. So we want to solve

$$(*) \qquad \boxed{ax^2 + bx + c = 0.}$$

Since we assumed $a \neq 0$ this amounts to solving the equation

$$(**) \qquad x^2 + \frac{b}{a}x + \frac{c}{a} = 0$$

obtained by dividing by a. Recall the formula

$$(x + t)^2 = x^2 + 2tx + t^2.$$

We want to find t such that $x^2 + \left(\frac{b}{a}\right)x$ has the form $x^2 + 2tx$. This means that we let

$$\frac{b}{a} = 2t, \qquad \text{that is} \quad t = \frac{b}{2a}.$$

We now add $\left(\frac{b}{2a}\right)^2$ to both sides of equation (**), and obtain

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

This can be rewritten in the form

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

or equivalently,

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

Taking square roots yields

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

whence

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a},$$

thus proving the quadratic formula.

Remark. It may happen that $b^2 - 4ac < 0$, in which case the quadratic equation does not have a solution in the real numbers.

Example. Find the roots of the equation

$$3x^2 - 2x + 1 = 0.$$

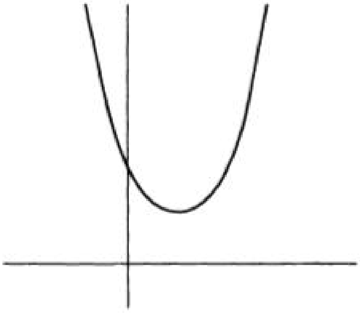
The roots are

$$x = \frac{-(-2) \pm \sqrt{4 - 12}}{6}.$$

Since $4 - 12 = -8 < 0$, the equation does not have roots in the real numbers. The equation

$$y = 3x^2 - 2x + 1$$

is the equation of a parabola, whose graph looks like that on the figure. The graph does not cross the x-axis.



The discussion of the quadratic formula in this section illustrates some general pedagogical principles concerning the relationship of rote memorization and the role of proofs in learning mathematics.

1. You should memorize the quadratic formula aurally:

x equals minus b plus or minus square root of b square minus $4ac$ over $2a$

as you would memorize a poem, by repeating it out loud. Such memorizing is necessary for a few important items in basic math, to induce getting the right answer as a conditioned reflex, without wasting any time.

2. Independently of using the formula as a conditioned reflex, you should see the proof by completing the square. Learning to handle logic and the English language in establishing theorems is also part of mathematics. In addition, the technique of completing the square arises often in the context of graphing circles, ellipses, parabolas, in addition to the quadratic formula. Knowing the formula as a conditioned reflex and knowing the proof serve two different complementary functions in mathematical training, neither of which eliminates the other.

II, §8. EXERCISES

Sketch the graph of the following equations:

1. $y = -x + 2$
2. $y = 2x^2 + x - 3$
3. $x - 4y^2 = 0$
4. $x - y^2 + y + 1 = 0$

Complete the square in the following equations and change the coordinate system to put them into the form

$$x'^2 + y'^2 = r^2 \qquad \text{or} \qquad y' = cx'^2 \qquad \text{or} \qquad x' = cy'^2$$

with a suitable constant c.

5. $x^2 + y^2 - 4x + 2y - 20 = 0$
6. $x^2 + y^2 - 2y - 8 = 0$
7. $x^2 + y^2 + 2x - 2 = 0$
8. $y - 2x^2 - x + 3 = 0$
9. $y - x^2 - 4x - 5 = 0$
10. $y - x^2 + 2x + 3 = 0$
11. $x^2 + y^2 + 2x - 4y = -3$
12. $x^2 + y^2 - 4x - 2y = -3$
13. $x - 2y^2 - y + 3 = 0$
14. $x - y^2 - 4y = 5$

II, §9. THE HYPERBOLA

We have already seen what the graph of the equation

$$xy = 1 \qquad \text{or} \qquad y = 1/x$$

looks like. It is of course the same as the graph of the function

$$f(x) = 1/x$$

(defined for $x \neq 0$). If we pick a coordinate system whose origin is at the point (a, b), the equation

$$y - b = \frac{1}{x - a}$$

is known as a **hyperbola**. In terms of the new coordinate system

$$x' = x - a$$

and $y' = y - b$, our hyperbola has the old type of equation

$$x'y' = 1.$$

If we are given an equation like

$$xy - 2x + 3y = 1$$

we want to put this equation in the form

$$(x - a)(y - b) = c,$$

or expanding out,

$$xy - ay - bx + ab = c.$$

This tells us what a and b must be. Thus

$$xy - 2x + 3y = (x + 3)(y - 2) + 6.$$

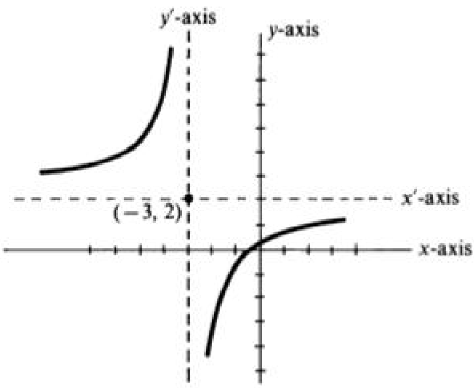
Hence $xy - 2x + 3y = 1$ is equivalent with

$$(x + 3)(y - 2) + 6 = 1$$

or in other words

$$(x + 3)(y - 2) = -5.$$

The graph of this equation has been drawn on the following diagram.



There is another form for the hyperbola. Let us try to graph the equation

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

If we solve for y we obtain

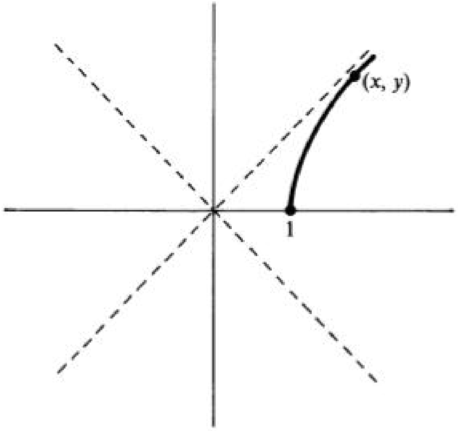
$$y^2 = x^2 - 1$$

so

$$y = \pm\sqrt{x^2 - 1}.$$

The graph has symmetry, because if (x, y) is a point on the graph then $(-x, y)$, $(x, -y)$ and $(-x, -y)$ is also a point on the graph. So let us look at the graph in the first quadrant when $x \geqq 0$ and $y \geqq 0$.

Since $x^2 - 1 = y^2$, it follows that $x^2 - 1 \geqq 0$ so $x^2 \geqq 1$. Hence the graph exists only for $x \geqq 1$. We claim that it looks like this in the first quadrant.



To see this, we could of course make a table of a few values first, to see experimentally what the graph is like. Do this yourself. Here we describe it theoretically.

As x increases, the expression $x^2 - 1$ increases, so $\sqrt{x^2 - 1}$ increases. Thus y also increases.

Also, since $y^2 = x^2 - 1$ it follows that $y^2 < x^2$ so $y < x$ for x, y in the first quadrant. We have drawn the line $y = x$. The set of points with $y < x$ lies below this line in the first quadrant.

Let us divide the equation $y^2 = x^2 - 1$ by x^2. We obtain

$$\frac{y^2}{x^2} = 1 - \frac{1}{x^2}.$$

When x becomes large, then

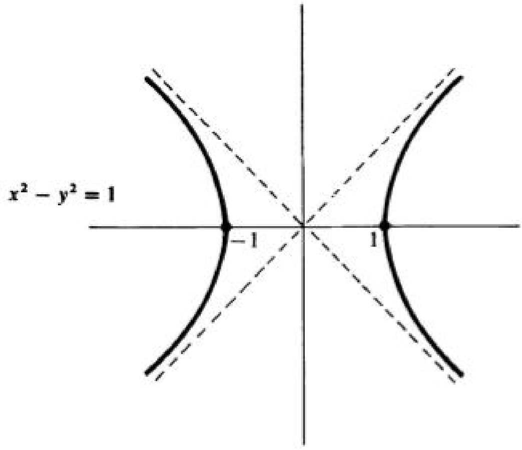
$$1 - \frac{1}{x^2} \quad \text{approaches } 1.$$

The ratio y/x is the slope of the line from the origin to the point (x, y). Hence this slope approaches 1 when x becomes large. Also from the expression

$$y = \sqrt{x^2 - 1},$$

we see that when x is large, $x^2 - 1$ is nearly equal to x^2, and so its square root is nearly equal to x. Thus the graph of the hyperbola comes closer and closer to the graph of the line $y = x$. This justifies drawing this graph as we have done.

Finally, by symmetry, the full graph of the hyperbola looks like this.



It is obtained by reflecting the graph in the first quadrant over the x-axis and over the y-axis, and also reflecting the graph through the origin.

II, §9. EXERCISES

Sketch the graphs of the following curves:

1. $(x - 1)(y - 2) = 2$
2. $x(y + 1) = 3$
3. $xy - 4 = 0$
4. $y = \dfrac{2}{1 - x}$
5. $y = \dfrac{1}{x + 1}$
6. $(x + 2)(y - 1) = 1$
7. $(x - 1)(y - 1) = 1$
8. $(x - 1)(y - 1) = 1$
9. $y = \dfrac{1}{x - 2} + 4$
10. $y = \dfrac{1}{x + 1} - 2$
11. $y = \dfrac{4x - 7}{x - 2}$
12. $y = \dfrac{-2x - 1}{x + 1}$
13. $y = \dfrac{x + 1}{x - 1}$
14. $y = \dfrac{x - 1}{x + 1}$
15. Graph the equation $y^2 - x^2 = 1$.
16. Graph the equation $(y - 1)^2 - (x - 2)^2 = 1$.
17. Graph the equation $(y + 1)^2 - (x - 2)^2 = 1$.

Part Two

Differentiation and Elementary Functions

In this part, we learn how to differentiate. Geometrically speaking, this amounts to finding the slope of a curve, or its rate of change. We analyze systematically the techniques for doing this, and how they apply to the elementary functions: polynomials, trigonometric functions, exponential and logarithmic functions, and inverse functions.

One of the reasons why we postpone integration till after this section is that the techniques of integration, up to a point, depend on our knowing the derivatives of certain functions, because one of the properties of integration is that it is the inverse operation to differentiation.

You will find applied rate problems similar to each other, but with different types of functions in Chapter III, §9, Chapter IV, §4, and Chapter VII, §4. This is an example of seeing the same idea threaded in a coherent manner through the part on differentiation.

CHAPTER III

The Derivative

The two fundamental notions of this course are those of the derivative and the integral. We take up the first one in this chapter.

The derivative will give us the slope of a curve at a point. It has also applications to physics, where it can be interpreted as the rate of change.

We shall develop some basic techniques which will allow you to compute the derivative in all the standard situations which you are likely to encounter in practice.

III, §1. THE SLOPE OF A CURVE

Consider a curve, and take a point P on the curve. We wish to define the notions of slope of the curve at that point, and tangent line to the curve at that point. Sometimes the statement is made that the tangent to the curve at the point is the line which touches the curve only at that point. This is pure nonsense, as the subsequent pictures will convince you.

Consider a straight line:

Don't you want to say that the line is tangent to itself? If yes, then this certainly contradicts that the tangent is the line which touches the curve at only one point, since the line touches itself at all of its points.

In Figs, 1, 2, and 3, we look at the tangent line to the curve at the point P. In Fig. 1 the line cuts the curve at the other point Q. In Fig. 2 the line is also tangent to the curve at the point Q. In Fig. 3, the

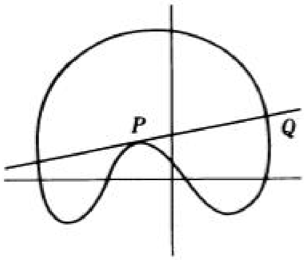


Figure 1

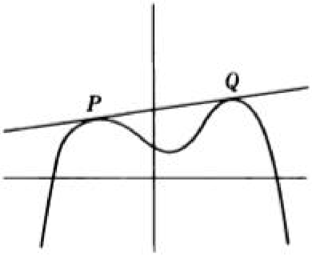


Figure 2

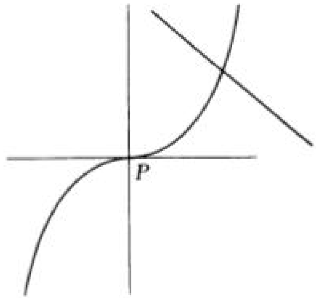


Figure 3

horizontal line is tangent to the curve and "cuts" the curve. The vertical line and the slanted line intersect the curve in only one point, but are not tangent.

Observe also that you cannot get out of the difficulties by trying to distinguish a line "cutting", the curve, or "touching the curve," or by saying that the line should lie on one side of the curve (cf. Fig. 1).

We therefore have to give up the idea of touching the curve only at one point, and look for another idea.

We have to face two problems. One of them is to give the correct geometric idea which allows us to define the tangent to the curve, and the other is to test whether this idea allows us to compute effectively this tangent line when the curve is given by a simple equation with numerical coefficients. It is a remarkable thing that our solution of the first problem will in fact give us a solution to the second.

In Chapter II, we have seen that knowing the slope of a straight line and one point on the straight line allows us to determine the equation of the line. We shall therefore define the slope of a curve at a point and then get its tangent afterward by using the method of Chapter II.

Our examples show us that to define the slope of the curve at P, we should not consider what happens at a point Q which is far removed from P. Rather, it is what happens near P which is important.

Let us therefore take any point Q on the given curve $y = f(x)$, and assume that $Q \neq P$. Then the two points P, Q determine a straight line with a certain slope which depends on P, Q and which we shall write as $S(P, Q)$. Suppose that the point Q approaches the point P on the curve (but stays distinct from P). Then, as Q comes near P, the slope $S(P, Q)$ of the line passing through P and Q should approach the (unkown) slope of the (unknown) tangent line to the curve at P. In the following diagram, we have drawn the tangent line to the curve at P and two lines between P and another point on the curve close to P (Fig. 4). The point

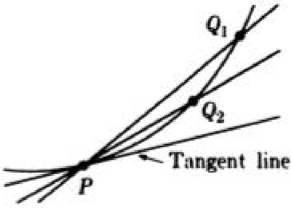


Figure 4

Q_2 is closer to P on the curve and so the slope of the line between P and Q_2 is closer to the slope of the tangent line than is the slope of the line between P and Q_1.

If the limit of the slope $S(P, Q)$ exists as Q approaches P, then it should be regarded as the slope of the curve itself at P. This is the basic idea behind our definition of the slope of the curve at P. We take it as a definition, perhaps the most important definition in this book. To repeat:

Definition. Given a curve $y = f(x)$, let P be a point on the curve. The **slope** of the curve at P is the limit of the slopes of lines through P and another point Q on the curve, as Q approaches P.

The idea of defining the slope in this manner was discovered in the seventeenth century by Newton and Leibnitz. We shall see that this definition allows us to determine the slope effectively in practice.

First we observe that when $y = ax + b$ is a straight line, then the slope of the line between any two distinct points on the curve is always the same, and is the slope of the line as we defined it in the preceding chapter.

Example. Let us now look at the next simplest example,

$$y = f(x) = x^2.$$

We wish to determine the slope of this curve at the point $(1, 1)$.

We look at a point near (1, 1), for instance a point whose x-coordinate is 1.1. Then $f(1.1) = (1.1)^2 = 1.21$. Thus the point (1.1, 1.21) lies on the curve. The slope of the line between two points (x_1, y_1) and (x_2, y_2) is

$$\frac{y_2 - y_1}{x_2 - x_1}.$$

Therefore the slope of the line between (1, 1) and (1.1, 1.21) is

$$\frac{1.21 - 1}{1.1 - 1} = \frac{0.21}{0.1} = 2.1.$$

In general, the x-coordinate of a point near (1, 1) can be written $1 + h$, where h is some small number, positive or negative, but $h \neq 0$. We have

$$f(1 + h) = (1 + h)^2 = 1 + 2h + h^2.$$

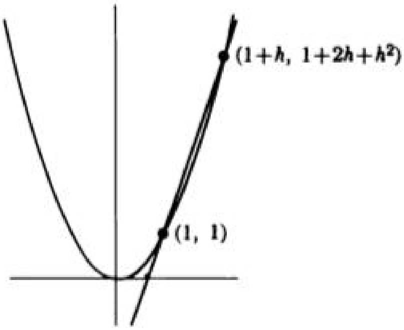


Figure 5

Thus the point $(1 + h, 1 + 2h + h^2)$ lies on the curve. When h is positive, the line between our two points would look like that in Fig. 5. When h is negative, then $1 + h$ is smaller than 1 and the line would look like this:

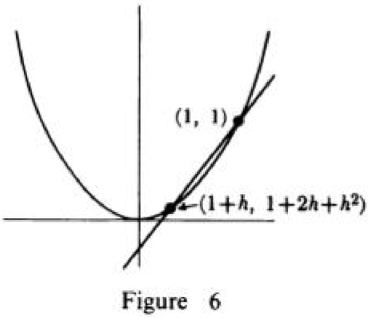


Figure 6

For instance, h could be -0.1 and $1 + h = 0.9$.

The slope of the line between our two points is therefore the quotient

$$\frac{(1 + 2h + h^2) - 1}{(1 + h) - 1},$$

which is equal to

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

As the point whose x-coordinate is $1 + h$ approaches our point $(1, 1)$, the number h approaches 0. As h approaches 0, the slope of the line between our two points approaches 2, which is therefore the slope of the curve at the point $(1, 1)$ by definition.

You will appreciate how simple the computation turns out to be, and how easy it was to get this slope!

Let us take another example. We wish to find the slope of the same curve $f(x) = x^2$ at the point $(-2, 4)$. Again we take a nearby point whose x-coordinate is $-2 + h$ for small $h \neq 0$. The y-coordinate of this nearby point is

$$f(-2 + h) = (-2 + h)^2 = 4 - 4h + h^2.$$

The slope of the line between the two points is therefore

$$\frac{4 - 4h + h^2 - 4}{-2 + h - (-2)} = \frac{-4h + h^2}{h} = -4 + h.$$

As h approaches 0, the nearby point approaches the point $(-2, 4)$ and we see that the slope approaches -4.

III, §1. EXERCISES

Find the slopes of the following curves at the indicated points:

1. $y = 2x^2$ at the point $(1, 2)$
2. $y = x^2 + 1$ at the point $(-1, 2)$
3. $y = 2x - 7$ at the point $(2, -3)$
4. $y = x^3$ at the point $(\frac{1}{2}, \frac{1}{8})$
5. $y = 1/x$ at the point $(2, \frac{1}{2})$
6. $y = x^2 + 2x$ at the point $(-1, -1)$
7. $y = x^2$ at the point $(2, 4)$
8. $y = x^2$ at the point $(3, 9)$
9. $y = x^3$ at the point $(1, 1)$
10. $y = x^3$ at the point $(2, 8)$
11. $y = 2x + 3$ at the point whose x-coordinate is 2.

12. $y = 3x - 5$ at the point whose x-coordinate is 1.

13. $y = ax + b$ at an arbitrary point.

(In Exercises 11, 12, 13 use the h-method, and verify that this method gives the same answer for the slope as that stated in Chapter II, §3.)

III, §2. THE DERIVATIVE

We continue to consider the function $y = x^2$. Instead of picking a definite numerical value for the x-coordinate of a point, we could work at an arbitrary point on the curve. Its coordinates are then (x, x^2). We write the x-coordinate of a point nearby as $x + h$ for some small h, positive or negative, but $h \neq 0$. The y-coordinate of this nearby point is

$$(x + h)^2 = x^2 + 2xh + h^2.$$

Hence the slope of the line between them is

$$\begin{aligned}\frac{(x + h)^2 - x^2}{(x + h) - x} &= \frac{x^2 + 2xh + h^2 - x^2}{x + h - x} \\ &= \frac{2xh + h^2}{h} \\ &= 2x + h.\end{aligned}$$

As h approaches 0, $2x + h$ approaches $2x$. Consequently, the slope of the curve $y = x^2$ at an arbitrary point (x, y) is $2x$. In particular, when $x = 1$ the slope is 2 and when $x = -2$ the slope is -4, as we found out before by the explicit computation using the special x-coordinates 1 and -2.

This time, however, we have found out a general formula giving us the slope for any point on the curve. Thus when $x = 3$ the slope is 6 and when $x = -10$ the slope is -20.

The example we have just worked out gives us the procedure for treating more general functions.

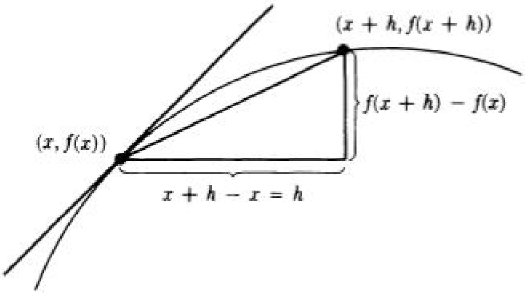
Given a function $f(x)$, we let the **Newton quotient** be

$$\boxed{\frac{f(x + h) - f(x)}{x + h - x} = \frac{f(x + h) - f(x)}{h}.}$$

This quotient is the slope of the line between the points

$$(x, f(x)) \quad \text{and} \quad (x + h, f(x + h)).$$

It is illustrated on the figure.



Definition. If the Newton quotient approaches a limit as h approaches 0, then we define the **derivative of** f **at** x to be this limit, that is

$$\boxed{\textbf{derivative of } f \textbf{ at } x = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}.}$$

The derivative of f at x will be denoted briefly by any one of the notations $f'(x)$, or df/dx, or $df(x)/dx$. Thus by definition

$$\boxed{f'(x) = \frac{df}{dx} = \frac{df(x)}{dx} = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}.}$$

Thus the two expressions $f'(x)$ and df/dx mean the same thing. We emphasize however that in the expression df/dx we do not multiply f or x by d or divide df by dx. The expression is to be read *as a whole*. We shall find out later that the expression, under certain circumstances, behaves *as if* we were dividing, and it is for this reason that we adopt this classical way of writing the derivative.

The derivative may thus be viewed as a function f', which is defined at all numbers x such that the Newton quotient approaches a limit as h tends to 0. Observe that when taking the limit, both the numerator $f(x + h) - f(x)$ and the denominator h approaches 0. However, their quotient

$$\frac{f(x + h) - f(x)}{h}$$

approaches the slope of the curve at the point $(x, f(x))$.

Definition. The function f is **differentiable** if it has a derivative at all the points for which it is defined.

Example 1. The function $f(x) = x^2$ is differentiable and its derivative is $2x$. Thus we have in this case

$$f'(x) = \frac{df}{dx} = 2x.$$

We have been using systematically the letter x. One can use any other letter: **The truth of mathematical statements is invariant under permutations of the alphabet**. Thus, for instance, if $f(u) = u^2$ then

$$f'(u) = \frac{df}{du} = 2u.$$

What is important here is that the *same letter* u should occur in each one of these places, and that the letter u should be different from the letter f.

We work out some examples before giving you exercises on this section.

Example 2. Let $f(x) = 2x + 1$. Find the derivative $f'(x)$.

We form the Newton quotient. We have $f(x + h) = 2(x + h) + 1$. Thus

$$\frac{f(x+h) - f(x)}{h} = \frac{2x + 2h + 1 - (2x + 1)}{h} = \frac{2h}{h} = 2.$$

As h approaches 0 (which we write also $h \to 0$), this Newton quotient is equal to 2 and hence the limit is 2. Thus

$$\frac{df}{dx} = f'(x) = 2$$

for all values of x. The derivative is constant.

Example 3. Find the slope of the graph of the function $f(x) = 2x^2$ at the point whose x-coordinate is 3, and find the equation of the tangent line at that point.

We may just as well find the slope at an arbitrary point on the graph. It is the derivative $f'(x)$. We have

$$f(x + h) = 2(x + h)^2 = 2(x^2 + 2xh + h^2).$$

The Newton quotient is

$$\frac{f(x+h)-f(x)}{h} = \frac{2(x^2+2xh+h^2)-2x^2}{h}$$
$$= \frac{4xh+2h^2}{h}$$
$$= 4x+2h.$$

Hence by definition

$$f'(x) = \lim_{h\to 0}(4x+2h) = 4x.$$

Thus $f'(x) = 4x$. At the point $x = 3$ we get

$$f'(3) = 12,$$

which is the desired slope.

As for the equation of the tangent line, when $x = 3$ we have $f(3) = 18$. Hence we must find the equation of the line passing through the point $(3, 18)$, having slope 12. This is easy, namely the equation is

$$y - 18 = 12(x - 3).$$

Remark on notation. In the preceding example, we have

$$\frac{df}{dx} = 4x = f'(x).$$

We wanted the derivative when $x = 3$. Here we see the advantage of the $f'(x)$ notation instead of the df/dx notation. We can substitute 3 for x in $f'(x)$ to write

$$f'(3) = 12.$$

We cannot substitute 3 for x in the df/dx notation, because writing

$$\frac{df}{d3}$$

would be very confusing. If we want to use the df/dx notation in such a context, we would have to use some device like writing

$$\frac{df}{dx} \quad \text{at} \quad x = 3 \quad \text{is equal to} \quad 12$$

or sometimes

$$\left.\frac{df}{dx}\right|_{x=3} = 12.$$

Still, it is clearly better to use the $f'(x)$ notation in such a context.

Example 4. Find the equation of the tangent line to the curve $y = 2x^2$ at the point whose x-coordinate is -2.

In the preceding example we have computed the general formula for the slope of the tangent line. It is

$$f'(x) = 4x.$$

At the point $x = -2$ the slope is therefore

$$f'(-2) = -8.$$

On the other hand, $f(-2) = 8$. Hence the equation of the tangent line is

$$y - 8 = -8(x + 2).$$

Example 5. Find the derivative of $f(x) = x^3$. We use the Newton quotient, and first write down its numerator:

$$\begin{aligned} f(x+h) - f(x) &= (x+h)^3 - x^3 \\ &= x^3 + 3x^2h + 3xh^2 + h^3 - x^3 \\ &= 3x^2h + 3xh^2 + h^3. \end{aligned}$$

Then

$$\frac{f(x+h) - f(x)}{h} = 3x^2 + 3xh + h^2.$$

As h approaches 0 the right-hand side approaches $3x^2$, so

$$\frac{d(x^3)}{dx} = 3x^2.$$

In defining the Newton quotient, we can take h positive or negative. It is sometimes convenient when taking the limit to look only at values of h which are positive. We are then looking only at points on the curve which approach the given point from the right. In this manner we get what is called the **right derivative**. If in taking the limit of the Newton quotient we took only negative values for h, we would get the **left derivative**.

Example 6. Let $f(x) = |x|$. Find its right derivative and its left derivative when $x = 0$.

The right derivative is the limit

$$\lim_{\substack{h \to 0 \\ h > 0}} \frac{f(0 + h) - f(0)}{h}.$$

When $h > 0$, we have

$$f(0 + h) = f(h) = h,$$

and $f(0) = 0$. Thus

$$\frac{f(0 + h) - f(0)}{h} = \frac{h}{h} = 1.$$

The limit as $h \to 0$ and $h > 0$ is therefore 1.

The left derivative is the limit

$$\lim_{\substack{h \to 0 \\ h < 0}} \frac{f(0 + h) - f(0)}{h}.$$

When $h < 0$ we have

$$f(0 + h) = f(h) = -h.$$

Hence

$$\frac{f(0 + h) - f(0)}{h} = \frac{-h}{h} = -1.$$

The limit as $h \to 0$ and $h < 0$ is therefore -1.

We see that the right derivative at 0 is 1 and the left derivative is -1. They are not equal. This is illustrated by the graph of our function

$$f(x) = |x|,$$

which looks like that in Fig. 7.

Both the right derivative of f and the left derivative of f exist but they are not equal.

We can rephrase our definition of the derivative and say that the derivative of a function $f(x)$ is defined when the right derivative and the left derivative exist and they are equal, in which case this common value is simply called the **derivative**.

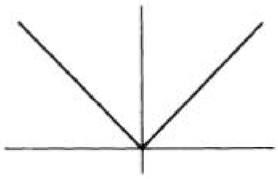


Figure 7

Thus the derivative of $f(x) = |x|$ is not defined at $x = 0$.

Example 7. Let $f(x)$ be equal to x if $0 < x \leqq 1$ and $x - 1$ if $1 < x \leqq 2$. We do not define f for other values of x. Then the graph of f looks like this:

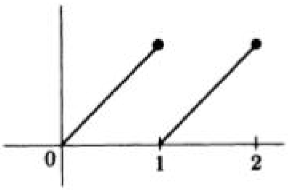


Figure 8

The left derivative of f at 1 exists and is equal to 1, but the right derivative of f at 1 does not exist. We leave the verification of the first assertion to you. To verify the second assertion, we must see whether the limit

$$\lim_{\substack{h \to 0 \\ h > 0}} \frac{f(1 + h) - f(1)}{h}$$

exists. Since $1 + h > 1$ we have

$$f(1 + h) = 1 + h - 1 = h.$$

Also $f(1) = 1$. Thus the Newton quotient is

$$\frac{f(1 + h) - f(1)}{h} = \frac{h - 1}{h} = 1 - \frac{1}{h}.$$

As h approaches 0 the quotient $1/h$ has no limit since it becomes arbitrarily large. Thus the Newton quotient has no limit for $h > 0$ and the function does not have a right derivative when $x = 1$.

III, §2. EXERCISES

Find (a) the derivatives of the following functions, (b) the slope of the graph at the point whose x-coordinate is 2, and find the equation of the tangent line at that point.

1. $x^2 + 1$
2. x^3
3. $2x^3$
4. $3x^2$
5. $x^2 - 5$
6. $2x^2 + x$
7. $2x^2 - 3x$
8. $\frac{1}{2}x^3 + 2x$
9. $\dfrac{1}{x+1}$
10. $\dfrac{2}{x+1}$

III, §3. LIMITS

In defining the slope of a curve at a point, or the derivative, we used the notion of limit, which we regarded as intuitively clear. It is indeed. You can see in the Appendix at the end of Part Four how one may define limits using only properties of numbers, but we do not worry about this here. However, we shall make a list of the properties of limits which will be used in the sequel, just to be sure of what we assume about them, and also to give you a technique for computing limits.

We consider functions $F(h)$ defined for all sufficiently small values of h, except that $h \neq 0$. We write

$$\boxed{\lim_{h \to 0} F(h) = L}$$

to mean that $F(h)$ **approaches** L **as** h **approaches** 0.

First, we note that if F is a constant function, $F(x) = c$ for all x, then

$$\lim_{h \to 0} F(h) = c$$

is the constant itself.

If $F(h) = h$, then

$$\lim_{h \to 0} F(h) = 0.$$

The next properties relate limits with addition, subtraction, multiplication, division, and inequalities.

Suppose that we have two functions $F(x)$ and $G(x)$ which are defined for the same numbers. Then we can form the sum of the two functions

$F + G$, whose value at a point x is $F(x) + G(x)$. Thus when $F(x) = x^4$ and $G(x) = 5x^{3/2}$ we have

$$F(x) + G(x) = x^4 + 5x^{3/2}.$$

The value $F(x) + G(x)$ is also written $(F + G)(x)$. The first property of limits concerns the sum of two functions.

Property 1. *Suppose that we have two functions F and G defined for small values of h, and assume that the limits*

$$\lim_{h\to 0} F(h) \quad \text{and} \quad \lim_{h\to 0} G(h)$$

exist. Then

$$\lim_{h\to 0}[F(h) + G(h)]$$

exists and

$$\lim_{h\to 0}(F + G)(h) = \lim_{h\to 0} F(h) + \lim_{h\to 0} G(h).$$

In other words the limit of a sum is equal to the sum of the limits.

A similar statement holds for the difference $F - G$, namely

$$\lim_{h\to 0} (F(h) - G(h)) = \lim_{h\to 0} F(h) - \lim_{h\to 0} G(h).$$

After the sum we discuss the product. Suppose we have two functions F and G defined for the same numbers. Then we can form their product FG whose value at a number x is

$$(FG)(x) = F(x)G(x).$$

For instance if $F(x) = 2x^2 - 2^x$ and $G(x) = x^2 + 5x$, then the product is

$$(FG)(x) = (2x^2 - 2^x)(x^2 + 5x).$$

Property 2. *Let F, G be two functions for small values of h, and assume that*

$$\lim_{h\to 0} F(h) \quad \text{and} \quad \lim_{h\to 0} G(h)$$

exist. Then the limit of the product exists and we have

$$\begin{aligned}\lim_{h\to 0}(FG)(h) &= \lim_{h\to 0}[F(h)G(h)] \\ &= \lim_{h\to 0} F(h)\cdot \lim_{h\to 0} G(h).\end{aligned}$$

In words, we can say that the product of the limits is equal to the limit of the product.

As a special case, suppose that $F(x)$ is the constant function $F(x) = c$. Then we can form the function cG, product of the constant by G, and we have

$$\lim_{h \to 0} cG(h) = c \cdot \lim_{h \to 0} G(h).$$

Example. Let $F(h) = 3h + 5$. Then $\lim_{h \to 0} F(h) = 5$.

Example. Let $F(h) = 4h^3 - 5h + 1$. Then $\lim_{h \to 0} F(h) = 1$. We can see this by considering the limits

$$\lim_{h \to 0} 4h^3 = 0, \qquad \lim_{h \to 0} 5h = 0, \qquad \lim_{h \to 0} 1 = 1,$$

and taking the appropriate sum.

Example. We have $\lim_{h \to 0} 3xh = 0$, and

$$\lim_{h \to 0}(3xh - 7y) = -7y.$$

Thirdly, we come to quotients. Let F, G be as before, but assume that $G(x) \neq 0$ for any x. Then we can form the quotient function F/G whose value at x is

$$\frac{F}{G}(x) = \frac{F(x)}{G(x)}.$$

Example. Let $F(x) = 2x^3 - 4x$ and $G(x) = x^4 + x^{1/3}$. Then

$$\frac{F}{G}(x) = \frac{F(x)}{G(x)} = \frac{2x^3 - 4x}{x^4 + x^{1/3}}.$$

Property 3. *Assume that the limits*

$$\lim_{h \to 0} F(h) \qquad \text{and} \qquad \lim_{h \to 0} G(h)$$

exist, and that

$$\lim_{h \to 0} G(h) \neq 0.$$

Then the limit of the quotient exists and we have

$$\lim_{h \to 0} \frac{F(h)}{G(h)} = \frac{\lim F(h)}{\lim G(h)}.$$

In words, the quotient of the limits is equal to the limit of the quotient.

As we have done above, we shall sometimes omit writing $h \to 0$ for the sake of simplicity.

The next property is stated here for completeness. It will not be used until we find the derivative of sine and cosine, and consequently should be skipped until then.

Property 4. *Let F, G be two functions defined for small values of h, and assume that $G(h) \leqq F(h)$. Assume also that*

$$\lim_{h \to 0} F(h) \quad \text{and} \quad \lim_{h \to 0} G(h)$$

exist. Then

$$\lim_{h \to 0} G(h) \leqq \lim_{h \to 0} F(h).$$

Property 5. *Let the assumptions be as in Property 4, and in addition, assume that*

$$\lim_{h \to 0} G(h) = \lim_{h \to 0} F(h).$$

Let E be another function defined for same numbers as F, G such that

$$G(h) \leqq E(h) \leqq F(h)$$

for all small values of h. Then

$$\lim_{h \to 0} E(h)$$

exists and is equal to the limits of F and G.

Property 5 is known as the **squeezing process**. You will find many applications of it in the sequel.

Example. Find the limit

$$\lim_{h\to 0} \frac{2xh + 3}{x^2 - 4h}$$

when $x \neq 0$.

The numerator of our quotient approaches 3 when $h \to 0$ and the denominator approaches x^2. Thus the quotient approaches $3/x^2$. We can justify these steps more formally by applying our three properties. For instance:

$$\begin{aligned}\lim_{h\to 0}(2xh + 3) &= \lim_{h\to 0}(2xh) + \lim_{h\to 0} 3 \\ &= \lim(2x)\lim(h) + \lim 3 \\ &= 2x\cdot 0 + 3 \\ &= 3.\end{aligned}$$

For the denominator, we have

$$\begin{aligned}\lim(x^2 - 4h) &= \lim x^2 + \lim(-4h) \\ &= x^2 + \lim(-4)\lim(h) \\ &= x^2 + (-4)\cdot 0 \\ &= x^2.\end{aligned}$$

Using the rule for the quotient, we see that the desired limit is equal to $3/x^2$.

Example. In the previous examples, it turns out that we could substitute the value 0 for h and find the appropriate limit. This cannot be done in general. For instance, suppose that we want to find the limit

$$\lim_{h\to 0} \frac{h^2 - h}{h^3 + 2h}.$$

If we substitute $h = 0$ we get a meaningless expression $0/0$, and hence we do not get information on the limit. However, for $h \neq 0$ we can cancel h from the quotient, and we see that

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

From the expression on the right we can determine the limit. Indeed, $h - 1$ approaches -1 as h approaches 0. Also, $h^2 + 2$ approaches 2 as h approaches 0. Hence by the rule for quotients of limits, we conclude that

$$\lim_{h \to 0} \frac{h^2 - h}{h^3 + 2h} = \lim_{h \to 0} \frac{h - 1}{h^2 + 2} = \frac{-1}{2}.$$

Observe that in this example, both the numerator and the denominator approach 0 as h approaches 0. However, the limit exists, and we see that it is $-\frac{1}{2}$.

Example. Find the limit

$$\lim_{h \to 0} \frac{x^2h^3 - h^2}{3xh - h}.$$

In this case, we can factor h from the numerator and denominator, so that the quotient is equal to

$$\frac{x^2h^2 - h}{3x - 1}.$$

The numerator is now seen to approach 0 and the denominator approaches $3x - 1$ as h approaches 0. Hence the quotient approaches 0. This is the desired limit.

The properties of limits which we have stated above will allow you to compute limits in determining derivatives. We illustrate this by an example.

Example. Let $f(x) = 1/x$ (defined for $x \neq 0$). Find the derivative df/dx.

The Newton quotient is

$$\begin{aligned} \frac{f(x + h) - f(x)}{h} &= \frac{\dfrac{1}{x + h} - \dfrac{1}{x}}{h} \\ &= \frac{x - (x + h)}{(x + h)xh} \\ &= \frac{-h}{(x + h)xh} = \frac{-1}{(x + h)x}. \end{aligned}$$

Then we take the limit:

$$\lim_{h\to 0}\frac{f(x+h)-f(x)}{h} = \lim_{h\to 0}\frac{-1}{(x+h)x}$$

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

(by the rule for the limit of a quotient)

$$= \frac{-1}{x^2}.$$

Thus we have proved:

$$\boxed{\frac{d\left(\frac{1}{x}\right)}{dx} = -\frac{1}{x^2}.}$$

III, §3. EXERCISES

Find the derivatives of the following functions, justifying the steps in taking limits by means of the first three properties:

1. $f(x) = 2x^2 + 3x$
2. $f(x) = \dfrac{1}{2x+1}$
3. $f(x) = \dfrac{x}{x+1}$
4. $f(x) = x(x+1)$
5. $f(x) = \dfrac{x}{2x-1}$
6. $f(x) = 3x^3$
7. $f(x) = x^4$
8. $f(x) = x^5$

(It is especially important that you should work out Exercises 7 and 8 to see a developing pattern, to be followed in the next section.)

9. $f(x) = 2x^3$
10. $f(x) = \frac{1}{2}x^3 + x$
11. $2/x$
12. $3/x$
13. $\dfrac{1}{2x-3}$
14. $\dfrac{1}{3x+1}$
15. $\dfrac{1}{x+5}$
16. $\dfrac{1}{x-2}$
17. $1/x^2$
18. $1/(x+1)^2$

III, §4. POWERS

We have seen that the derivative of the function x^2 is $2x$.

Let us consider the function $f(x) = x^3$ and find its derivative. We have

$$f(x + h) = (x + h)^3 = x^3 + 3x^2h + 3xh^2 + h^3.$$

Hence the Newton quotient is

$$\begin{aligned}\frac{f(x+h) - f(x)}{h} &= \frac{x^3 + 3x^2h + 3xh^2 + h^3 - x^3}{h} \\ &= \frac{3x^2h + h^2(3x + h)}{h} \qquad \text{(after cancellations)} \\ &= 3x^2 + 3xh + h^2 \qquad \text{(after cancelling } h\text{)}.\end{aligned}$$

Using the properties of limits of sums and products, we see that $3x^2$ remains equal to itself as h approaches 0, that $3xh$ and h^3 both approach 0. Hence

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = 3x^2.$$

This suggests that in general, whenever $f(x) = x^n$ for some positive integer n, the derivative $f'(x)$ should be nx^{n-1}. This is indeed the case, and will be proved in the next theorem. The proof will follow the same pattern as in the case $f(x) = x^3$ above. You should work out the case $f(x) = x^4$ and probably also the case $f(x) = x^5$ in detail (these were exercises for the preceding section) in order to confirm in these special cases that the pattern remains the same, before seeing the general case. Note that when we treated $f(x) = x^3$ we obtained an expression for the Newton quotient which had a numerator

$$3x^2h + h^2(3x + h),$$

containing the term $3x^2h$, and another term containing h^2 as a factor. When we divide by h, $3x^2$ yields the value of the derivative, and the remaining term $h(3x + h)$ still contains h as a factor, and consequently approaches 0 as h approaches 0. You will find explicitly a similar phenomenon for x^4 and x^5.

Theorem 4.1. *Let n be an integer $\geqq 1$ and let $f(x) = x^n$. Then*

$$\frac{df}{dx} = nx^{n-1}.$$

Proof. We have

$$f(x + h) = (x + h)^n = (x + h)(x + h)\cdots(x + h),$$

the product being taken n times. Selecting x from each factor gives us a term x^n. If we take x from all but one factor and h from the remaining factors, we get hx^{n-1} taken n times. This gives us a term $nx^{n-1}h$. All other terms will involve selecting h from at least two factors, and the corresponding terms will be divisible by h^2. Thus we get

$$f(x + h) = (x + h)^n = x^n + nx^{n-1}h + h^2g(x, h),$$

where $g(x, h)$ is simply some expression involving powers of x and h with numerical coefficients which, as we shall see later in the proof, it is unnecessary for us to determine. However, using the rules for limits of sums and products we can conclude that

$$\lim_{h \to 0} g(x, h)$$

will be some number which it is unnecessary for us to determine.

The Newton quotient is therefore

$$\begin{aligned}\frac{f(x + h) - f(x)}{h} &= \frac{x^n + nx^{n-1}h + h^2g(x, h) - x^n}{h} \\ &= \frac{nx^{n-1}h + h^2g(x, h)}{h} && \text{(because } x^n \text{ cancels)} \\ &= nx^{n-1} + hg(x, h) && \text{(divide numerator and denominator by } h\text{).}\end{aligned}$$

As h approaches 0, the term nx^{n-1} remains unchanged. The limit of h as h tends to 0 is 0, and hence by the product rule, the term $hg(x, h)$ approaches 0 when h approaches 0. Thus finally

$$\lim_{h \to 0} \frac{f(x + h) - f(x)}{h} = nx^{n-1},$$

which proves our theorem.

For another proof, see the end of the next section.

Theorem 4.2. *Let a be any number and let $f(x) = x^a$ (defined for $x > 0$). Then $f(x)$ has a derivative, which is*

$$f'(x) = ax^{a-1}.$$

It would not be difficult to prove Theorem 4.2 when a is a negative integer. It is best, however, to wait until we have a rule giving us the derivative of a quotient before doing it. We could also give a proof when a is a rational number. However, we shall prove the general result in a later chapter, and thus we prefer to wait until then, when we have more techniques available.

Examples. If $f(x) = x^{10}$, then $f'(x) = 10x^9$.
If $f(x) = x^{3/2}$ (for $x > 0$), then $f'(x) = \frac{3}{2}x^{1/2}$.
If $f(x) = x^{-5/4}$, then $f'(x) = -\frac{5}{4}x^{-9/4}$.
If $f(x) = x^{\sqrt{2}}$, then $f'(x) = \sqrt{2}x^{\sqrt{2}-1}$.

Note especially the special case when $f(x) = x$. Then $f'(x) = 1$.

Example. We can now find the equations of tangent lines to certain curves which we could not do before. Consider the curve

$$y = x^5$$

We wish to find the equation of its tangent line at the point (2, 32). By Theorem 4.1, if $f(x) = x^5$, then $f'(x) = 5x^4$. Hence the slope of the tangent line at $x = 2$ is

$$f'(2) = 5 \cdot 2^4 = 80.$$

On the other hand, $f(2) = 2^5 = 32$. Hence the equation of the tangent line is

$$y - 32 = 80(x - 2).$$

III, §4. EXERCISES

1. Write out the expression of $(x + h)^4$ in terms of powers of x and h.

2. Find the derivative of the function x^4 directly, using the Newton quotient.

3. What are the derivatives of the following functions?
   (a) $x^{2/3}$ (b) $x^{-3/2}$ (c) $x^{7/6}$

4. What is the equation of the tangent line to the curve $y = x^9$ at the point (1, 1)?

5. What is the slope of the curve $y = x^{2/3}$ at the point (8, 4)? What is the equation of the tangent line at that point?

6. Give the slope and equation of the tangent line to the curve $y = x^{-3/4}$ at the point whose x-coordinate is 16.

7. Give the slope and equation of the tangent line to the curve $y = \sqrt{x}$ at the point whose x-coordinate is 3.

8. Give the derivatives of the following functions at the indicated points:
   (a) $f(x) = x^{1/4}$ at $x = 5$
   (b) $f(x) = x^{-1/4}$ at $x = 7$
   (c) $f(x) = x^{\sqrt{2}}$ at $x = 10$
   (d) $f(x) = x^{\pi}$ at $x = 7$

III, §5. SUMS, PRODUCTS, AND QUOTIENTS

In this section we shall derive several rules which allow you to find the derivatives for sums, products, and quotients of functions when you know the derivative of each factor.

We begin with a definition of continuous functions and the reason why a differentiable function is continuous.

Definition. A function is said to be **continuous at a point** x if and only if

$$\lim_{h \to 0} f(x + h) = f(x).$$

A function is said to be **continuous** if it is continuous at every point of its domain of definition.

Let f be a function having a derivative $f'(x)$ at x. Then f is continuous at x.

Proof. The quotient

$$\frac{f(x + h) - f(x)}{h}$$

approaches the limit $f'(x)$ as h approaches 0. We have

$$h\,\frac{f(x + h) - f(x)}{h} = f(x + h) - f(x).$$

Therefore using the rule for the limit of a product, and noting that h approaches 0, we find

$$\lim_{h \to 0} f(x + h) - f(x) = 0f'(x) = 0.$$

This is another way of stating that

$$\lim_{h \to 0} f(x + h) = f(x).$$

In other words, f is continuous.

Of course, we can never substitute $h = 0$ in our quotient, because then it becomes 0/0, which is meaningless. Geometrically, letting $h = 0$ amounts to taking the two points on the curve equal to each other. It is then impossible to have a unique straight line through one point. Our procedure of taking the limit of the Newton quotient is meaningful only if $h \neq 0$.

Observe that in the Newton quotient, both the numerator and the denominator approach 0. The quotient itself, however, need not approach 0.

Example. Let $f(x) = |x|$. Then the absolute value function f is continuous at 0, even though it is not differentiable at 0. It is still true that

$$f(0 + h) = f(h) = |h|$$

approaches 0 as h approaches 0, even though the function is not differentiable at 0. As we saw in §2, the function $f(x) = |x|$ is right differentiable at 0 and left differentiable at 0, but not differentiable at 0.

Example. Let $f(x) = 0$ if $x < 0$, and $f(x) = 1$ if $x \geqq 0$. The graph of f is shown on Fig. 9.

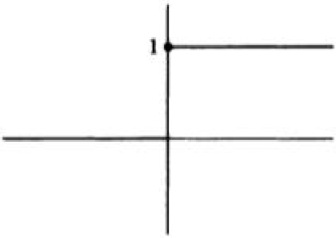


Figure 9

The function f is not continuous at 0. Roughly speaking, the function is not continuous at 0 because its graph has a "break" at 0. In Fig. 10, we show other examples of graphs with discontinuities.

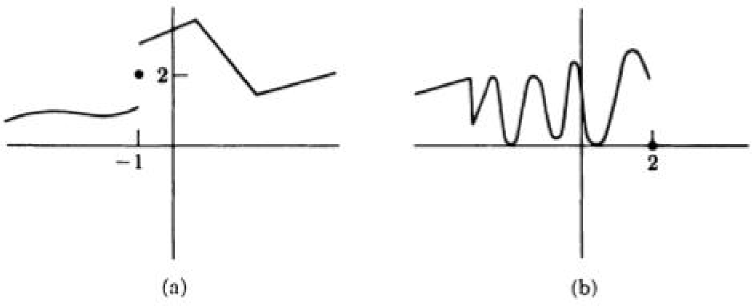


Figure 10

In Fig. 10(a), the function is continuous except at $x = -1$. In Fig. 10(b), the function is continuous except at $x = 2$.

The remark at the beginning of this section shows that if a function is differentiable, then it is continuous. Since we are concerned principally about differentiable functions at present, we do not go any deeper into continuous functions, and wait till later, until the notion becomes more relevant to us.

Let c be a number and $f(x)$ a function which has a derivative $f'(x)$ for all values of x for which it is defined. We can multiply f by the constant c to get another function cf whose value at x is $cf(x)$.

Constant times a function. *The derivative of cf is then given by the formula*

$$(cf)'(x) = c \cdot f'(x).$$

In other words, the derivative of a constant times a function is the constant times the derivative of the function.

In the other notation, this reads

$$\boxed{\frac{d(cf)}{dx} = c\,\frac{df}{dx}.}$$

To prove this rule, we use the definition of derivative. The Newton quotient for the function cf is

$$\frac{(cf)(x+h) - (cf)(x)}{h} = \frac{cf(x+h) - cf(x)}{h} = c\,\frac{f(x+h) - f(x)}{h}.$$

Let us take the limit as h approaches 0. Then c remains fixed, and

$$\frac{f(x+h) - f(x)}{h}$$

approaches $f'(x)$. According to the rule for the product of limits, we see that our Newton quotient approaches $cf'(x)$, as was to be proved.

Example. Let $f(x) = 3x^2$. Then $f'(x) = 6x$. If $f(x) = 17x^{1/2}$, then $f'(x) = \frac{17}{2}x^{-1/2}$. If $f(x) = 10x^a$, then $f'(x) = 10ax^{a-1}$.

Next we look at the sum of two functions.

Sum. *Let $f(x)$ and $g(x)$ be two functions which have derivatives $f'(x)$ and $g'(x)$, respectively. Then the sum $f(x) + g(x)$ has a derivative, and*

$$(f + g)'(x) = f'(x) + g'(x).$$

The derivative of a sum is equal to the sum of the derivatives.

In the other notation, this reads:

$$\boxed{\frac{d(f+g)}{dx} = \frac{df}{dx} + \frac{dg}{dx}.}$$

To prove this, we have by definition

$$(f+g)(x+h) = f(x+h) + g(x+h)$$

and

$$(f+g)(x) = f(x) + g(x).$$

Therefore the Newton quotient for $f+g$ is

$$\frac{(f+g)(x+h)-(f+g)(x)}{h} = \frac{f(x+h)+g(x+h)-f(x)-g(x)}{h}.$$

Collecting terms and separating the fraction, we see that this expression is equal to

$$\frac{f(x+h)-f(x)+g(x+h)-g(x)}{h}$$

$$= \frac{f(x+h)-f(x)}{h} + \frac{g(x+h)-g(x)}{h}.$$

Taking the limit as h approaches 0 and using the rule for the limit of a sum, we see that this last sum approaches $f'(x)+g'(x)$ as h approaches 0. This proves what we wanted.

Example.

$$\frac{d}{dx}(x^3 + x^2) = 3x^2 + 2x,$$

$$\frac{d}{dx}(4x^{1/2} + 5x^{-10}) = 2x^{-1/2} - 50x^{-11}.$$

Carried away by our enthusiasm at determining so easily the derivative of functions built up from others by means of constants and sums, we might now be tempted to state the rule that the derivative of a product is the product of the derivatives. Unfortunately, this is false. To see that the rule is false, we look at an example.

Let $f(x) = x$ and $g(x) = x^2$. Then $f'(x) = 1$ and $g'(x) = 2x$. Therefore $f'(x)g'(x) = 2x$. However, the derivative of the product $(fg)(x) = x^3$ is

$3x^2$, which is certainly not equal to $2x$. Thus the product of the derivatives is not equal to the derivative of the product.

Through trial and error the correct rule was discovered. It can be stated as follows:

Product. *Let $f(x)$ and $g(x)$ be two functions having derivatives $f'(x)$ and $g'(x)$. Then the product function $f(x)g(x)$ has a derivative, which is given by the formula*

$$(fg)'(x) = f(x)g'(x) + f'(x)g(x).$$

In words, the derivative of the product is equal to the first times the derivative of the second, plus the derivative of the first times the second.

In the other notation, this reads

$$\boxed{\frac{d(fg)}{dx} = f(x)\frac{dg}{dx} + \frac{df}{dx}g(x).}$$

The proof is not very much more difficult than the proofs we have already encountered. By definition, we have

$$(fg)(x + h) = f(x + h)g(x + h)$$

and

$$(fg)(x) = f(x)g(x).$$

Consequently the Newton quotient for the product function fg is

$$\frac{(fg)(x + h) - (fg)(x)}{h} = \frac{f(x + h)g(x + h) - f(x)g(x)}{h}.$$

At this point, it looks a little hopeless to transform this quotient in such a way that we see easily what limit it approaches as h approaches 0. But we use a trick, and rewrite our quotient by inserting

$$-f(x + h)g(x) + f(x + h)g(x)$$

in the numerator. This certainly does not change the value of our quotient, which now looks like

$$\frac{f(x + h)g(x + h) - f(x + h)g(x) + f(x + h)g(x) - f(x)g(x)}{h}.$$

We can split this fraction into a sum of two fractions:

$$\frac{f(x+h)g(x+h)-f(x+h)g(x)}{h}+\frac{f(x+h)g(x)-f(x)g(x)}{h}.$$

We can factor $f(x+h)$ in the first term, and $g(x)$ in the second term, to obtain

$$f(x+h)\frac{g(x+h)-g(x)}{h}+\frac{f(x+h)-f(x)}{h}g(x).$$

The situation is now well under control. As h approaches 0, $f(x+h)$ approaches $f(x)$, and the two quotients in the expression we have just written approach $g'(x)$ and $f'(x)$ respectively. Thus the Newton quotient for fg approaches

$$f(x)g'(x)+f'(x)g(x),$$

thereby proving our assertion.

Example. Applying the product rule, we find:

$$\frac{d}{dx}(x+1)(3x^2)=(x+1)6x+1\cdot 3x^2.$$

Similarly,

$$\frac{d}{dx}[(2x^5+5x^4)(2x^{1/2}+x^{-1})]$$
$$=(2x^5+5x^4)\left(x^{-1/2}-\frac{1}{x^2}\right)+(10x^4+20x^3)(2x^{1/2}+x^{-1})$$

which you may and should leave just like that without attempting to simplify the expression.

A special case of the product rule is used all the time, that of the square of a function.

Example.

$$\boxed{\frac{d(f(x)^2)}{dx}=2f(x)f'(x).}$$

Indeed, differentiating the product $y=f(x)f(x)$, we obtain

$$\frac{dy}{dx}=f(x)f'(x)+f'(x)f(x)=2f(x)f'(x).$$

The last rule of this section concerns the derivative of a quotient. We begin with a special case.

Let $g(x)$ be a function having a derivative $g'(x)$, and such that $g(x) \neq 0$. Then the derivative of the quotient $1/g(x)$ exists, and is equal to

$$\boxed{\frac{d}{dx}\frac{1}{g(x)} = \frac{-1}{g(x)^2}\, g'(x).}$$

To prove this, we look at the Newton quotient

$$\frac{\dfrac{1}{g(x+h)} - \dfrac{1}{g(x)}}{h}$$

which is equal to

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

Letting h approach 0 we see immediately that our expression approaches

$$\frac{-1}{g(x)^2}\, g'(x)$$

as desired.

Example.

$$\frac{d}{dx}\frac{1}{(x^5 - 3x)} = \frac{-1}{(x^5 - 3x)^2}(5x^4 - 3).$$

The general case of the rule for quotients can now be easily stated and proved.

Quotient. *Let $f(x)$ and $g(x)$ be two functions having derivatives $f'(x)$ and $g'(x)$ respectively, and such that $g(x) \neq 0$. Then the derivative of the quotient $f(x)/g(x)$ exists, and is equal to*

$$\frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2}.$$

In words this yields:

The bottom times the derivative of the top, minus the top times the derivative of the bottom, over the bottom squared

(*which you should memorize like a poem*).

In the other notation, this reads:

$$\boxed{\frac{d(f/g)}{dx} = \frac{g(x)\,df/dx - f(x)\,dg/dx}{g(x)^2}.}$$

To prove this rule, we write our quotient in the form

$$\frac{f(x)}{g(x)} = f(x)\,\frac{1}{g(x)}$$

and use the rule for the derivative of a product, together with the special case we have just proved. We obtain its derivative:

$$\begin{aligned}\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) &= f'(x)\,\frac{1}{g(x)} + f(x)\,\frac{-1}{g(x)^2}\,g'(x)\\ &= \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2}.\end{aligned}$$

by putting this expression over the common denominator $g(x)^2$. This is the desired derivative.

We work out some examples.

Example.

$$\frac{d}{dx}\left(\frac{x^2+1}{3x^4-2x}\right) = \frac{(3x^4-2x)2x - (x^2+1)(12x^3-2)}{(3x^4-2x)^2}.$$

Example.

$$\frac{d}{dx}\left(\frac{2x}{x+4}\right) = \frac{(x+4)\cdot 2 - 2x\cdot 1}{(x+4)^2}.$$

Example. Find the equation of the tangent line to the curve

$$y = x/(x^2+4)$$

at the point $x = -3$.

We let $f(x) = x/(x^2+4)$. Then

$$f'(x) = \frac{(x^2+4)\cdot 1 - x\cdot 2x}{(x^2+4)^2}.$$

Hence

$$f'(-3) = \frac{13 - 18}{13^2} = \frac{-5}{169}.$$

This is the slope at the given point. Furthermore,

$$f(-3) = \frac{-3}{13}.$$

Hence the coordinates of the given point are $(-3, -3/13)$. The equation of the tangent line is therefore

$$y + \frac{3}{13} = \frac{-5}{169}(x + 3).$$

Example. Show that the two curves

$$y = x^2 + 1 \quad \text{and} \quad y = \tfrac{2}{3}x^3 + \tfrac{4}{3}$$

have a common tangent line at the point $(1, 2)$. We let

$$f(x) = x^2 + 1 \quad \text{and} \quad g(x) = \tfrac{2}{3}x^3 + \tfrac{4}{3}.$$

Then

$$f'(x) = 2x \quad \text{and} \quad g'(x) = 2x^2.$$

Since $f'(1) = g'(1) = 2$, the slope of the tangent lines to the graph of f and the graph of g is the same at the point $(1, 2)$. Hence the tangent lines are the same since they pass through the same point, namely

$$y - 2 = 2(x - 1).$$

Appendix. Another proof that $dx^n/dx = nx^{n-1}$

There is another proof that

$$\frac{dx^n}{dx} = nx^{n-1}$$

for any positive integer n, as follows. First we note that when $n = 1$ we have proved directly

$$\frac{dx}{dx} = 1.$$

by using the h-method.

Now we use the rule for the derivative of a product. To get the derivative of x^2, we have:

$$\frac{d(x^2)}{dx} = \frac{d(x \cdot x)}{dx} = x\frac{dx}{dx} + \frac{dx}{dx}x = 2x.$$

Next, we get the derivative of x^3:

$$\begin{aligned}\frac{d(x^3)}{dx} = \frac{d(x^2 \cdot x)}{dx} &= x^2\frac{dx}{dx} + \frac{dx^2}{dx}x \\ &= x^2 + 2x^2 \qquad \text{(by the preceding step)} \\ &= 3x^2.\end{aligned}$$

Next we get the derivative of x^4:

$$\begin{aligned}\frac{d(x^4)}{dx} = \frac{d(x^3 \cdot x)}{dx} &= x^3\frac{dx}{dx} + \frac{d(x^3)}{dx}x \\ &= x^3 + 3x^3 \qquad \text{(by the preceding step)} \\ &= 4x^3.\end{aligned}$$

We can proceed like this for any integer n. Suppose we have proved the formula up to some integer n. Then

$$\begin{aligned}\frac{d(x^{n+1})}{dx} = \frac{d(x^n \cdot x)}{dx} &= x^n\frac{dx}{dx} + \frac{d(x^n)}{dx}x \\ &= x^n + nx^n \qquad \text{(by the preceding step)} \\ &= (n+1)x^n.\end{aligned}$$

This shows how to proceed from one step to the next.

Such a procedure is called **induction**. You will find several instances of such a procedure throughout the course. In each case, work out the steps for $n = 1$, $n = 2$, $n = 3$, $n = 4$, and so forth, in succession. Go as far as you need to recognize the pattern and to be comfortable with it. Then try to formulate and carry out the final step with n instead of a specific number. After you have met enough examples of this type, you will have understood what induction means, and in particular, you will understand the following formal definition.

Suppose we want to prove an assertion $A(n)$ for every positive integer n. Then a **proof by induction** consists in proving:

(1) The assertion is true when $n = 1$.
(2) If the assertion is true for a given integer n, then it is true for $n + 1$.

The first step allows us to start the procedure; and the second step allows us to go from one integer to the next, just as we did in the above example.

III, §5. EXERCISES

Find the derivatives of the following functions:

1. (a) $2x^{1/3}$ (b) $3x^{3/4}$ (c) $\frac{1}{2}x^2$ (d) $\frac{3}{4}x^2$
2. (a) $5x^{11}$ (b) $4x^{-2}$ (c) $\frac{1}{3}x^4 - 5x^3 + x^2 - 2$
3. (a) $\frac{1}{2}x^{-3/4}$ (b) $3x - 2x^3$ (c) $4x^5 - 7x^3 + 2x - 1$
4. (a) $7x^3 + 4x^2$ (b) $4x^{2/3} + 5x^4 - x^3 + 3x$
5. (a) $25x^{-1} + 12x^{1/2}$ (b) $2x^3 + 5x^7$ (c) $4x^4 - 7x^3 + x - 12$
6. (a) $\frac{3}{5}x^2 - 2x^8$ (b) $3x^4 - 2x^2 + x - 10$ (c) $\pi x^7 - 8x^5 + x + 1$
7. $(x^3 + x)(x - 1)$
8. $(2x^2 - 1)(x^4 + 1)$
9. $(x + 1)(x^2 + 5x^{3/2})$
10. $(2x - 5)(3x^4 + 5x + 2)$
11. $(x^{-2/3} + x^2)\left(x^3 + \dfrac{1}{x}\right)$
12. $(2x + 3)\left(\dfrac{1}{x^2} + \dfrac{1}{x}\right)$
13. $\dfrac{2x + 1}{x + 5}$
14. $\dfrac{2x}{x^2 + 3x + 1}$

To break the monotony of the letter x, let us use another.

15. $f(t) = \dfrac{t^2 + 2t - 1}{(t + 1)(t - 1)}$
16. $\dfrac{t^{-5/4}}{t^2 + t - 1}$
17. What is the slope of the curve
$$y = \frac{t}{t + 5}$$
at the point $t = 2$? What is the equation of the tangent line at this point?
18. What is the slope of the curve?
$$y = \frac{t^2}{t^2 + 1}$$
at $t = 1$? What is the equation of the tangent line?

III, §5. SUPPLEMENTARY EXERCISES

Find the derivatives of the following functions. Do not simplify your answers!

1. $3x^3 - 4x + 5$
2. $x^2 + 2x + 27$
3. $x^2 + x - 1$
4. $x^{1/2} - 8x^4 + x^{-1}$

5. $x^{5/2} + x^{-5/2}$

6. $x^7 + 15x^{-1/5}$

7. $(x^2 - 1)(x + 5)$

8. $\left(x^5 + \dfrac{1}{x}\right)(x^5 + 1)$

9. $(x^{3/2} + x^2)(x^4 - 99)$

10. $(x^2 + x + 1)(x^5 - x - 25)$

11. $(2x^2 + 1)\left(\dfrac{1}{x^2} + 4x + 8\right)$

12. $(x^4 - x^2)(x^2 - 1)$

13. $(x + 1)(x + 2)(x + 3)$

14. $5(x - 1)(x - 2)(x^2 + 1)$

15. $x^3(x^2 + 1)(x + 1)$

16. $(x^4 + 1)(x + 5)(2x + 7)$

17. $\dfrac{1}{2x + 3}$

18. $\dfrac{1}{7x + 27}$

19. $\dfrac{-5}{x^3 + 2x^2}$

20. $\dfrac{3}{2x^4 + x^{3/2}}$

21. $\dfrac{-2x}{x + 1}$

22. $\dfrac{x + 1}{x - 5}$

23. $\dfrac{3x^{1/2}}{(x + 1)(x - 1)}$

24. $\dfrac{2x^{1/2} + x^{3/4}}{(x + 1)x^3}$

25. $\dfrac{x^5 + 1}{(x^2 + 1)(x + 7)}$

26. $\dfrac{(x + 1)(x + 5)}{x - 4}$

27. $\dfrac{x^3}{1 - x^2}$

28. $\dfrac{x^5}{x^{3/2} + x}$

29. $\dfrac{x^2 - x}{x^2 + 1}$

30. $\dfrac{x^2 + 2x + 7}{8x}$

31. $\dfrac{2x + 1}{x^2 + x - 4}$

32. $\dfrac{x^5}{x^2 + 3}$

33. $\dfrac{4x - x^3}{x^2 + 2}$

34. $\dfrac{x^3}{x^2 - 5x + 7}$

35. $\dfrac{1 - 5x}{x}$

36. $\dfrac{1 + 6x + x^{3/4}}{7x - 2}$

37. $\dfrac{x^2}{(x + 1)(x - 2)}$

38. $\dfrac{x^{1/2} - x^{-1/2}}{x^{3/4}}$

39. $\dfrac{3x^4 + x^{5/4}}{4x^3 - x^5 + 1}$

40. $\dfrac{x - 1}{(x - 2)(x - 3)}$

Find the equations of the tangent lines to the following curves at the given point.

41. $y = x^{1/4} + 2x^{3/4}$ at $x = 16$

42. $y = 2x^3 + 3$ at $x = \frac{1}{2}$

43. $y = (x - 1)(x - 3)(x - 4)$ at $x = 0$

44. $y = 2x^2 + 5x - 1$ at $x = 2$

45. $y = (x^2 + 1)(2x + 3)$ at $x = 1$

46. $y = \dfrac{x - 1}{x + 5}$ at $x = 2$

47. $y = \dfrac{x^2}{x^3 + 1}$ at $x = 2$

48. $y = \dfrac{x^2 + 1}{x^3 + 1}$ at $x = 2$

49. $y = \dfrac{x^2}{x^2 - 1}$ at $x = 2$

50. $y = \dfrac{x-1}{x^2+1}$ at $x = 1$

51. Show that the line $y = -x$ is tangent to the curve given by the equation

$$y = x^3 - 6x^2 + 8x.$$

Find the point of tangency.

52. Show that the line $y = 9x - 15$ is tangent to the curve

$$y = x^3 - 3x + 1.$$

Find the point of tangency.

53. Show that the graphs of the equations

$$y = 3x^2 \qquad \text{and} \qquad y = 2x^3 + 1$$

have the common tangent line at the point $(1, 3)$. Sketch the graphs.

54. Show that there are exactly two tangent lines to the graph of $y = (x + 1)^2$ which pass through the origin, and find their equations.

55. Find all the points (x_0, y_0) on the curve

$$y = 4x^4 - 8x^2 + 16x + 7$$

such that the tangent line to the curve at (x_0, y_0) is parallel to the line

$$16x - y + 5 = 0.$$

Find the tangent line to the curve at each of these points.

III, §6. THE CHAIN RULE

We know how to build up new functions from old ones by means of sums, products, and quotients. There is one other important way of building up new functions. We shall first give examples of this new way.

Consider the function $(x + 2)^{10}$. We can say that this function is made up of the 10-th power function, and the function $x + 2$. Namely, given a number x, we first add 2 to it, and then take the 10-th power. Let

$$g(x) = x + 2$$

and let f be the 10-th power function. Then we can take the value of f at $x + 2$, namely

$$f(x + 2) = (x + 2)^{10}$$

and we can also write it as

$$f(x+2) = f(g(x)).$$

Another example: Consider the function $(3x^4 - 1)^{1/2}$. If we let

$$g(x) = 3x^4 - 1$$

and we let f be the square root function, then

$$f(g(x)) = \sqrt{3x^4 - 1} = (3x^4 - 1)^{1/2}.$$

In order not to get confused by the letter x, which cannot serve us any more in all contexts, we use another letter to denote the values of g. Thus we may write $f(u) = u^{1/2}$.

Similarly, let $f(u)$ be the function $u + 5$ and $g(x) = 2x$. Then

$$f(g(x)) = f(2x) = 2x + 5.$$

One more example of the same type: Let

$$f(u) = \frac{1}{u+2}$$

and

$$g(x) = x^{10}.$$

Then

$$f(g(x)) = \frac{1}{x^{10} + 2}.$$

In order to give you sufficient practice with many types of functions, we now mention several of them whose definitions will be given later. These will be sin and cos (which we read sine and cosine), log (which we read logarithm or simply log), and the exponential function exp. We shall select a special number e (whose value is approximately $2.718\ldots$), such that the function exp is given by

$$\exp(x) = e^x.$$

We now see how we make new functions out of these:

Let $f(u) = \sin u$ and $g(x) = x^2$. Then

$$f(g(x)) = \sin(x^2).$$

Let $f(u) = e^u$ and $g(x) = \cos x$. Then

$$f(g(x)) = e^{\cos x}.$$

Let $f(v) = \log v$ and $g(t) = t^3 - 1$. Then

$$f(g(t)) = \log(t^3 - 1).$$

Let $f(w) = w^{10}$ and $g(z) = \log z + \sin z$. Then

$$f(g(z)) = (\log z + \sin z)^{10}.$$

Whenever we have two functions f and g such that f is defined for all numbers which are values of g, then we can build a new function denoted by $f \circ g$ whose value at a number x is

$$(f \circ g)(x) = f(g(x)).$$

The rule defining this new function is: Take the number x, find the number $g(x)$, and then take the value of f at $g(x)$. This is the value of $f \circ g$ at x. The function $f \circ g$ is called the **composite function** of f and g. We say that g is the **inner** function and that f is the **outer** function. For example, in the function $\log \sin x$, we have

$$f \circ g = \log \circ \sin.$$

The outer function is the log, and the inner function is the sine.

It is important to keep in mind that we can compose two functions only when the outer function is defined at all values of the inner function. For instance, let $f(u) = u^{1/2}$ and $g(x) = -x^2$. Then we cannot form the composite function $f \circ g$ because f is defined only for positive numbers (or 0) and the values of g are all negative, or 0. Thus $(-x^2)^{1/2}$ does not make sense.

However, for the moment you are asked to learn the mechanism of composite functions just the way you learned the multiplication table, in order to acquire efficient conditioned reflexes when you meet composite functions. Hence for the drills given by the exercises at the end of the section, you should forget for a while the meaning of the symbols and operate with them formally, just to learn the formal rules properly.

Even though we have not defined the exponential function e^x nor have we dealt formally with $\sin x$ or other of the functions just mentioned, nevertheless, we don't need to know their definitions in order to manipulate them. If we limited ourselves only to those functions which we have explicitly dealt with so far, we would be restricted in seeing how composite functions work, and how the chain rule works below.

We come to the problem of taking the derivative of a composite function.

We start with an example. Suppose we want to find the derivative of the function $(x + 1)^{10}$. The Newton quotient would be a very long expression, which it would be essentially hopeless to disentangle by brute force, the way we have up to now. It is therefore a pleasant surprise that there will be an easy way of finding the derivative. We tell you the answer right away: The derivative of this function is $10(x + 1)^9$. This looks very much related to the derivative of powers.

Before stating and proving the general theorem, we give you other examples.

Example.

$$\frac{d}{dx}(x^2 + 2x)^{3/2} = \frac{3}{2}(x^2 + 2x)^{1/2}(2x + 2).$$

Observe carefully the extra term $2x + 2$, which is the derivative of the expression $x^2 + 2x$. We may describe the answer in the following terms. We let $u = x^2 + x$ so that $du/dx = 2x + 1$. Let $f(u) = u^{3/2}$. Then we have

$$\frac{d(f(u(x)))}{dx} = \frac{df}{du}\frac{du}{dx}.$$

Example.

$$\frac{d}{dx}(x^2 + x)^{10} = 10(x^2 + x)^9(2x + 1).$$

Observe again the presence of the term $2x + 1$, which is the derivative of $x^2 + x$. Again, if we let $u = x^2 + x$ and $f(u) = u^{10}$, then

$$\frac{df(u(x))}{dx} = \frac{df}{du}\frac{du}{dx}, \qquad \text{where} \quad \frac{df}{du} = 10u^9 \qquad \text{and} \qquad \frac{du}{dx} = 2x + 1.$$

Can you guess the general rule from the preceding assertions? The general rule was also discovered by trial and error, but we profit from three centuries of experience, and thus we are able to state it and prove it very simply, as follows.

Chain rule. *Let f and g be two functions having derivatives, and such that f is defined at all numbers which are values of g. Then the composite function $f \circ g$ has a derivative, given by the formula*

$$(f \circ g)'(x) = f'(g(x))g'(x).$$

This can be expressed in words by saying that we take the *derivative of the outer function times the derivative of the inner function* (*or the derivative of what's inside*).

The preceding assertion is known as the **chain rule**.

If we put $u = g(x)$, then we may express the chain rule in the form

$$\boxed{\frac{df(u(x))}{dx} = \frac{df}{du}\frac{du}{dx},}$$

or also

$$\frac{d(f \circ g)}{dx} = \frac{df}{du}\frac{du}{dx}.$$

Thus the derivative behaves *as if* we could cancel the du. As long as we have proved this result, there is nothing wrong with working like a machine in computing derivatives of composite functions, and we shall give you several examples before the exercises.

Example. Let $F(x) = (x^2 + 1)^{10}$. Then

$$F(x) = f(g(x)),$$

where

$$u = g(x) = x^2 + 1 \qquad \text{and} \qquad f(u) = u^{10}.$$

Then

$$f'(u) = 10u^9 = \frac{df}{du} \qquad \text{and} \qquad g'(x) = 2x = \frac{dg}{dx}.$$

Thus

$$\frac{d(f \circ g)}{dx} = \frac{df}{du}\frac{du}{dx} = 10u^9 \cdot 2x = 10(x^2 + 1)^9 2x.$$

Example. *Let* $f(u) = 2u^{1/2}$ and $g(x) = 5x + 1$. Then

$$f'(u) = u^{-1/2} = \frac{df}{du} \qquad \text{and} \qquad g'(x) = 5 = \frac{dg}{dx}.$$

Thus

$$\frac{d}{dx}\, 2(5x + 1)^{1/2} = 2 \cdot \tfrac{1}{2}(5x + 1)^{-1/2} \cdot 5 = (5x + 1)^{-1/2} \cdot 5.$$

(Pay attention to the constant 5, which is the derivative of $5x + 1$. You are very likely to forget it.)

In order to give you more extensive drilling than would be afforded by the functions we have considered, like powers, we summarize the derivatives of the elementary functions which are to be considered later.

$$\frac{d(\sin x)}{dx} = \cos x, \qquad \frac{d(\cos x)}{dx} = -\sin x.$$

$$\frac{d(e^x)}{dx} = e^x \qquad \text{(yes, } e^x\text{, the same as the function!).}$$

$$\frac{d(\log x)}{dx} = \frac{1}{x}.$$

Example.

$$\frac{d}{dx}(\sin x)^7 = 7(\sin x)^6 \cos x.$$

In this example, $f(u) = u^7$ and $df/du = 7u^6$. Also

$$u = \sin x, \qquad \text{and} \qquad \frac{du}{dx} = \cos x.$$

Example.

$$\frac{d}{dx}(\sin 3x)^7 = 7(\sin 3x)^6 \cos 3x \cdot 3.$$

The last factor of 3 occurring on the right-hand side is the derivative $d(3x)/dx$.

Example. Let n be any integer. For any differentiable function $f(x)$,

$$\boxed{\frac{d}{dx} f(x)^n = nf(x)^{n-1} \frac{df}{dx}.}$$

Example.

$$\frac{de^{\sin x}}{dx} = \frac{df}{du}\frac{du}{dx} = e^{\sin x}(\cos x).$$

In this example, $f(u) = e^u$, $df/du = e^u$, and $u = \sin x$.

Example.

$$\frac{d \cos(2x^2)}{dx} = \frac{df}{du}\frac{du}{dx} = -\sin(2x^2) \cdot 4x.$$

In this example, $f(u) = \cos u$ and $df/du = -\sin u$. Also $u = 2x^2$ so that $du/dx = 4x$.

Example.

$$\frac{d\cos 4x}{dx} = -\sin(4x)\cdot 4.$$

In this example, $f(u) = \cos u$ and $u = 4x$ so $du/dx = 4$.

We emphasize what we have already stated. *If we limited ourselves just to polynomials or quotients of polynomials, we would not have enough examples to drill the mechanism of the chain rule. There is nothing wrong in **using** the properties of functions which have not yet been formally defined in the course.* We could in fact create totally imaginary functions to achieve the same end.

Example. Suppose there is a function called schmoo(x), whose derivative is given by

$$\frac{d\,\text{schmoo}(x)}{dx} = \frac{1}{x + \sin x}.$$

Then

$$\begin{aligned}\frac{d}{dx}\,\text{schmoo}(x^3 + 4x) &= \frac{d\,\text{schmoo}(u)}{du}\frac{du}{dx}\\ &= \frac{1}{(x^3 + 4x) + \sin(x^3 + 4x)}(3x^2 + 4).\end{aligned}$$

Example. Suppose there is a function cow(x) such that $\text{cow}'(x) = \text{schmoo}(x)$. Then

$$\frac{d\,\text{cow}(x^2)}{dx} = \text{schmoo}(x^2)\cdot 2x.$$

Proof of the Chain Rule. We must consider the Newton quotient of the composite function $f\circ g$. By definition, it is

$$\frac{f[g(x+h)] - f[g(x)]}{h}.$$

Put $u = g(x)$, and let

$$k = g(x+h) - g(x).$$

Then k depends on h, and tends to 0 as h approaches 0. Our Newton quotient is equal to

$$\frac{f(u+k)-f(u)}{h}.$$

For the present argument *suppose that k is unequal to* 0 for all small values of h. Then we can multiply and divide this quotient by k, and obtain

$$\frac{f(u+k)-f(u)}{k}\frac{k}{h}=\frac{f(u+k)-f(u)}{k}\frac{g(x+h)-g(x)}{h}.$$

If we let h approach 0 and use the rule for the limit of a product, we see that our Newton quotient approaches

$$f'(u)g'(x),$$

and this would prove our chain rule, under the assumption that k is not 0.

It does not happen very often that $k=0$ for arbitrarily small values of h, but when it does happen, the preceding argument must be refined. For those of you who are interested, we shall show you how the argument can be slightly modified so as to be valid in all cases. **The uninterested reader can just skip it.**

We distinguish two kinds of numbers h. The first kind, those for which $g(x+h)-g(x)\neq 0$, and the second kind, those for which

$$g(x+h)-g(x)=0.$$

Let H_1 be the set of h of the first kind, and H_2 the set of h of the second kind. We assume that we have

$$g(x+h)-g(x)=0$$

for arbitrarily small values of h. Then the Newton quotient

$$\frac{g(x+h)-g(x)}{h}$$

is 0 for such values, that is for h in H_2, and consequently

$$g'(x)=\lim_{h\to 0}\frac{g(x+h)-g(x)}{h}=0.$$

Furthermore,

$$\lim_{\substack{h\to 0 \\ h \text{ in } H_2}} \frac{f(g(x+h)) - f(g(x))}{h} = 0,$$

because h is of second kind, so $g(x+h) - g(x) = 0$, $g(x+h) = g(x)$ and therefore $f(g(x+h)) - f(g(x)) = 0$. The limit here is taken with h approaching 0, but h of the second kind.

On the other hand, if we take the limit with h of the first kind, then the original argument applies, i.e. we can divide and multiply by

$$k = g(x+h) - g(x),$$

and we find

$$\lim_{\substack{h\to 0 \\ h \text{ in } H_1}} \frac{f(g(x+h)) - f(g(x))}{h} = f'(g(x))g'(x)$$

as before. But $g'(x) = 0$. Hence the limit is $0 = f'(g(x))g'(x)$ when h approaches 0, whether h is of first kind or second kind. This concludes the proof.

III, §6. EXERCISES

Find the derivatives of the following functions.

1. $(x+1)^8$
2. $(2x-5)^{1/2}$
3. $(\sin x)^3$
4. $(\log x)^5$
5. $\sin 2x$
6. $\log(x^2+1)$
7. $e^{\cos x}$
8. $\log(e^x + \sin x)$
9. $\sin\left(\log x + \dfrac{1}{x}\right)$
10. $\dfrac{x+1}{\sin 2x}$
11. $(2x^2+3)^3$
12. $\cos(\sin 5x)$
13. $\log(\cos 2x)$
14. $\sin[(2x+5)^2]$
15. $\sin[\cos(x+1)]$
16. $\sin(e^x)$
17. $\dfrac{1}{(3x-1)^4}$
18. $\dfrac{1}{(4x)^3}$
19. $\dfrac{1}{(\sin 2x)^2}$
20. $\dfrac{1}{(\cos 2x)^2}$
21. $\dfrac{1}{\sin 3x}$
22. $(\sin x)(\cos x)$

23. $(x^2 + 1)e^x$

24. $(x^3 + 2x)(\sin 3x)$

25. $\dfrac{1}{\sin x + \cos x}$

26. $\dfrac{\sin 2x}{e^x}$

27. $\dfrac{\log x}{x^2 + 3}$

28. $\dfrac{x + 1}{\cos 2x}$

29. $(2x - 3)(e^x + x)$

30. $(x^3 - 1)(e^{3x} + 5x)$

31. $\dfrac{x^3 + 1}{x - 1}$

32. $\dfrac{x^2 - 1}{2x + 3}$

33. $(x^{4/3} - e^x)(2x + 1)$

34. $(\sin 3x)(x^{1/4} - 1)$

35. $\sin(x^2 + 5x)$

36. e^{3x^2+8}

37. $\dfrac{1}{\log(x^4 + 1)}$

38. $\dfrac{1}{\log(x^{1/2} + 2x)}$

39. $\dfrac{2x}{e^x}$

40. Let f be a function such that $f'(u) = \dfrac{1}{1 + u^3}$. Let $g(x) = f(x^2)$. Find $g'(x)$ and $g'(2)$. Do not attempt to evaluate $f(u)$.

41. Relax.

III, §6. SUPPLEMENTARY EXERCISES

Find the derivatives of the following functions.

1. $(2x + 1)^2$

2. $(2x + 5)^3$

3. $(5x + 3)^7$

4. $(7x - 2)^{81}$

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

6. $(2x^3 - 3x)^4$

7. $(3x + 1)^{1/2}$

8. $(2x - 5)^{5/4}$

9. $(x^2 + x - 1)^{-2}$

10. $(x^4 + 5x + 6)^{-1}$

11. $(x + 5)^{-5/3}$

12. $(x^3 + 2x + 1)^3$

13. $(x - 1)(x - 5)^3$

14. $(2x^2 + 1)^2(x^2 + 3x)$

15. $(x^3 + x^2 - 2x - 1)^4$

16. $(x^2 + 1)^3(2x + 5)^2$

17. $\dfrac{(x + 1)^{3/4}}{(x - 1)^{1/2}}$

18. $\dfrac{(2x + 1)^{1/2}}{(x + 5)^5}$

19. $\dfrac{(2x^2 + x - 1)^{5/2}}{(3x + 2)^9}$

20. $\dfrac{(x^2 + 1)(3x - 7)^8}{(x^2 + 5x - 4)^3}$

21. $\sqrt{2x+1}$

22. $\sqrt{x+3}$

23. $\sqrt{x^2+x+5}$

24. $\sqrt{2x^3-x+1}$

In the following exercises, we may assume that there are functions $\sin u$, $\cos u$, $\log u$, and e^u whose derivatives are given by the following formulas:

$$\frac{d \sin u}{du} = \cos u, \qquad \frac{d \cos u}{du} = -\sin u,$$

$$\frac{d(e^u)}{du} = e^u, \qquad \frac{d \log u}{du} = \frac{1}{u}.$$

Find the derivative of each function (with respect to x):

25. $\sin(x^3+1)$

26. $\cos(x^3+1)$

27. e^{x^3+1}

28. $\log(x^3+1)$

29. $\sin(\cos x)$

30. $\cos(\sin x)$

31. $e^{\sin(x^3+1)}$

32. $\log[\sin(x^3+1)]$

33. $\sin[(x+1)(x^2+2)]$

34. $\log(2x^2+3x+5)$

35. $e^{(x+1)(x-3)}$

36. e^{2x+1}

37. $\sin(2x+5)$

38. $\cos(7x+1)$

39. $\log(2x+1)$

40. $\log\dfrac{2x+1}{x+3}$

41. $\sin\dfrac{x-5}{2x+4}$

42. $\cos\dfrac{2x-1}{x+3}$

43. e^{2x^2+3x+1}

44. $\log(4x^3-2x)$

45. $\sin[\log(2x+1)]$

46. $\cos(e^{2x})$

47. $\cos(3x^2-2x+1)$

48. $\sin\left(\dfrac{x^2-1}{2x^3+1}\right)$

49. $(2x+1)^{80}$

50. $(\sin x)^{50}$

51. $(\log x)^{49}$

52. $(\sin 2x)^4$

53. $(e^{2x+1}-x)^5$

54. $(\log x)^{20}$

55. $(3\log(x^2+1)-x^3)^{1/2}$

56. $(\log(2x+3))^{4/3}$

57. $\dfrac{\sin 2x}{\cos 3x}$

58. $\dfrac{\sin(2x+5)}{\cos(x^2-1)}$

59. $\dfrac{\log 2x^2}{\sin x^3}$

60. $\dfrac{e^{x^3}}{x^2-1}$

61. $\dfrac{x^4+4}{\cos 2x}$

62. $\dfrac{\sin(x^3-2)}{\sin 2x}$

63. $\dfrac{(2x^2+1)^4}{(\cos x^3)}$

64. $\dfrac{e^{-x}}{\cos 2x}$

65. e^{-3x}

66. e^{-x^2}

67. e^{-4x^2+x}

68. $\sqrt{e^x+1}$

69. $\dfrac{\log(x^2+2)}{e^{-x}}$

70. $\dfrac{\log(2x+1)}{\sin(4x+5)}$

III, §7. HIGHER DERIVATIVES

Given a differentiable function f defined on an interval, its derivative f' is also a function on this interval. If it turns out to be also differentiable (this being usually the case), then its derivative is called the **second derivative** of f and is denoted by $f''(x)$.

Example. Let $f(x) = (x^3 + 1)^2$. Then

$$f'(x) = 2(x^3 + 1)3x^2 = 6x^5 + 6x^2 \qquad \text{and} \qquad f''(x) = 30x^4 + 12x.$$

There is no reason to stop at the second derivative, and one can of course continue with the third, fourth, etc. provided they exist. Since it is notationally inconvenient to pile up primes after f to denote successive derivatives, one writes

$$f^{(n)}$$

for the n-th derivative of f. Thus f'' is also written $f^{(2)}$. If we wish to refer to the variable x, we also write

$$\boxed{f^{(n)}(x) = \frac{d^n f}{dx^n}.}$$

Example. Let $f(x) = x^3$. Then

$$\frac{df}{dx} = 3x^2, \qquad \frac{d^2 f}{dx^2} = f''(x) = f^{(2)}(x) = 6x,$$

$$\frac{d^3 f}{dx^3} = f^{(3)}(x) = 6, \qquad \frac{d^4 f}{dx^4} = f^{(4)}(x) = 0.$$

Example. Let $f(x) = 5x^3$. Then

$$f'(x) = 15x^2 = \frac{df}{dx},$$

$$f^{(2)}(x) = 30x = \frac{d^2 f}{dx^2},$$

$$f^{(3)}(x) = 30 = \frac{d^3 f}{dx^3},$$

$$f^{(4)}(x) = 0 = \frac{d^4 f}{dx^4}.$$

Example. Let $f(x) = \sin x$. Then:

$$f'(x) = \cos x,$$

$$f^{(2)}(x) = -\sin x.$$

III, §7. EXERCISES

Find the second derivatives of the following functions:

1. $3x^3 + 5x + 1$

2. $(x^2 + 1)^5$

3. Find the 80-th derivative of $x^7 + 5x - 1$.

4. Find the 7-th derivative of $x^7 + 5x - 1$.

5. Find the third derivative of $x^2 + 1$.

6. Find the third derivative of $x^3 + 2x - 5$.

7. Find the third derivative of the function $g(x) = \sin x$.

8. Find the fourth derivative of the function $g(x) = \cos x$.

9. Find the 10-th derivative of $\sin x$.

10. Find the 10-th derivative of $\cos x$.

11. Find the 100-th derivative of $\sin x$.

12. Find the 100-th derivative of $\cos x$.

13. (a) Find the 5-th derivative of x^5.
(b) Find the 7-th derivative of x^7.
(c) Find the 13-th derivative of x^{13}.

In the process of finding these derivatives, you should observe a pattern. Let n be a positive integer. Define $n!$ to be the product of the first n integers. Thus

$$n! = n(n-1)(n-2)\cdots 2\cdot 1.$$

This is called n **factorial**. For example:

$$2! = 2, \qquad 3! = 3\cdot 2 = 6, \qquad 4! = 4\cdot 3\cdot 2 = 24.$$

Compute 5!, 6!, 7!. You will find $n!$ used especially in the chapter on Taylor's formula, much later in the course.

14. In general, let k be a positive integer. Let

$$f(x) = x^k.$$

(a) What is $f^{(k)}(x)$?
(b) What is $f^{(k)}(0)$?
(c) Let n be a positive integer $> k$. What is $f^{(n)}(0)$?
(d) Let n be a positive integer $< k$. What is $f^{(n)}(0)$?

III, §8. IMPLICIT DIFFERENTIATION

Suppose that a curve is defined by an equation

$$F(x, y) = 0$$

like a circle, $x^2 + y^2 = 7$, or an ellipse, or more generally an equation like

$$3x^3y - y^4 + 5x^2 + 5 = 0.$$

It is usually the case that for most values of x, one can solve back for y as a function of x, that is find a differentiable function

$$y = f(x)$$

satisfying the equation. For example, in the case of the circle,

$$x^2 + y^2 = 7,$$

we have

$$y^2 = 7 - x^2$$

and hence we get two possibilities for y,

$$y = \sqrt{7 - x^2} \qquad \text{or} \qquad y = -\sqrt{7 - x^2}.$$

The graph of the first function is the upper semicircle, and the graph of the second function is the lower semicircle.

In the example $3x^3y - y^4 + 5x^2 + 5 = 0$, it is a mess to solve for y, and we don't do it. However, assuming that $y = f(x)$ is a differentiable function satisfying this equation, we can find an expression for the derivative more easily, and we shall do so in an example below.

Example. Find the derivative dy/dx if $x^2 + y^2 = 7$, in terms of x and y.

We differentiate both sides of the equation using the chain rule, and the fact that $dx/dx = 1$. We then obtain:

$$2x\frac{dx}{dx} + 2y\frac{dy}{dx} = 0, \qquad \text{that is} \qquad 2x + 2y\frac{dy}{dx} = 0.$$

Hence,

$$\frac{dy}{dx} = \frac{-2x}{2y} = \frac{-x}{y}.$$

Example. Find the tangent line to the circle $x^2 + y^2 = 7$ at the point $x = 2$ and $y = \sqrt{3}$.

The slope of the line at this point is given by

$$\left.\frac{dy}{dx}\right|_{(2,\sqrt{3})} = \frac{-2\cdot 2}{2\sqrt{3}} = \frac{-2}{\sqrt{3}}.$$

Hence the equation of the tangent line is

$$y - \sqrt{3} = -\frac{2}{\sqrt{3}}(x - 2).$$

Example. Find the derivative dy/dx in terms of x and y if

$$3x^3y - y^4 + 5x^2 = -5.$$

Again we assume that y is a function of x. We differentiate both sides using the rule for derivative of a product, and the chain rule. We then obtain:

$$3x^3\frac{dy}{dx} + 9x^2y - 4y^3\frac{dy}{dx} + 10x = 0,$$

or factoring out,

$$\frac{dy}{dx}(3x^3 - 4y^3) = -10x - 9x^2y.$$

This yields

$$\frac{dy}{dx} = -\frac{10x + 9x^2y}{3x^3 - 4y^3}.$$

Example. Find the equation of the tangent line in the preceding example at the point $x = 1$, $y = 2$.

We first find the slope at the given point. This is obtained by substituting $x = 1$ and $y = 2$ in the expression for dy/dx obtained in the preceding example. Thus:

$$\left.\frac{dy}{dx}\right|_{(1,2)} = -\frac{10 + 18}{3 - 32} = \frac{28}{29}.$$

Then the equation of the tangent line at $(1, 2)$ is

$$y - 2 = \tfrac{28}{29}(x - 1).$$

III, §8. EXERCISES

Find dy/dx in terms of x and y in the following problems.

1. $x^2 + xy = 2$
2. $(x-3)^2 + (y+1)^2 = 37$
3. $x^3 - xy + y^3 = 1$
4. $y^3 - 2x^3 + y = 1$
5. $2xy + y^2 = x + y$
6. $\frac{1}{x} + \frac{1}{y} = 1$
7. $y^2 + 2x^2y + x = 0$
8. $x^2y^2 = x^2 + y^2$

Find the tangent line of the following curves at the indicated point.

9. $x^2y^2 = 9$ at $(-1, 3)$
10. $x^2 + y^3 + 2x - 5y - 19 = 0$ at $(3, -1)$
11. $(y - x)^2 = 2x + 4$ at $(6, 2)$
12. $2x^2 - y^3 + 4xy - 2x = 0$ at $(1, -2)$
13. $x^2 + y^2 = 25$ at $(3, -4)$
14. $x^2 - y^2 + 3xy + 12 = 0$ at $(-4, 2)$
15. $x^2 + xy - y^2 = 1$ at $(2, 3)$

III, §9. RATE OF CHANGE

The derivative has an interesting physical interpretation, which was very closely connected with it in its historical development, and is worth mentioning.

Suppose that a particle moves along some straight line a certain distance depending on time t. Then the distance s is a function of t, which we write $s = f(t)$.

For two values of the time, t_1 and t_2, the quotient

$$\frac{f(t_2) - f(t_1)}{t_2 - t_1}$$

can be regarded as a sort of average speed of the particle since it gives the total distance covered divided by the total time elapsed. At a given time t_0, it is therefore reasonable to regard the limit

$$\lim_{t \to t_0} \frac{f(t) - f(t_0)}{t - t_0}$$

as the rate of change of s with respect to t. This is none other than the derivative $f'(t)$, which is called the **speed**.

Let us denote the **speed** by $v(t)$. Then

$$v(t) = \frac{ds}{dt}.$$

The rate of change of the speed is called the **acceleration**. Thus

$$\frac{dv}{dt} = a(t) = \text{acceleration} = \frac{d^2s}{dt^2}.$$

Example. If the particle is an object dropping under the influence of gravity, then experimental data show that

$$s = \tfrac{1}{2}Gt^2,$$

where G is the gravitational constant. In that case,

$$\frac{ds}{dt} = Gt$$

is its speed. The acceleration is then

$$\frac{d^2s}{dt^2} = G = \text{gravitational constant.}$$

Example. A particle is moving so that at time t the distance traveled is given by the function

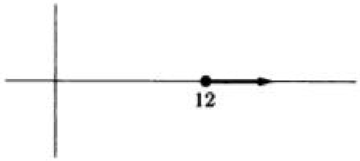
$$s(t) = t^2 + 1.$$

The derivative $s'(t)$ is equal to $2t$. Thus the speed of the particle is equal to 0 at time $t = 0$. Its speed is equal to 4 at time $t = 2$.

In general, given a function $y = f(x)$, the derivative $f'(x)$ is interpreted as the **rate of change of y with respect to x**. Thus f' is also a function. If y is increasing as x is increasing, this means that the derivative is positive, in other words $f'(x) > 0$. If y is decreasing, this means that the rate of change of y with respect to x is negative, that is $f'(x) < 0$.

Example. Suppose that a particle moves with uniform speed along a straight line, say along the x-axis toward the right, away from the origin. Suppose the speed is 5 cm/sec. We may then write

$$\frac{dx}{dt} = 5.$$



Suppose that at some time the particle is 12 cm to the right of the origin. After each further second of motion, the distance increases by another 5 cm, so after 3 seconds, the distance of the particle from the origin will be

$$12 + 3 \cdot 5 = 12 + 15 = 27 \text{ cm}.$$

One can find the x-coordinate as a function of time. Under uniform speed, distance traveled is equal to the product of speed with time. Thus if the particle starts from the origin at time $t = 0$ then

$$x(t) = 5t.$$

If, on the other hand, the particle starts from another point x_0, then

$$x(t) = 5t + x_0.$$

Indeed, if we substitute 0 for t in this equation, we find

$$x(0) = 5 \cdot 0 + x_0 = x_0.$$

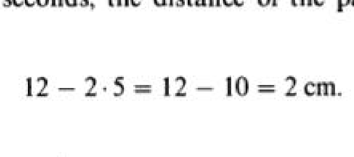
Hence x_0 is the value of x when $t = 0$.

Example. Suppose that a particle is moving to the left at a rate of 5 cm/sec. Then we write

$$\frac{dx}{dt} = -5.$$

Again suppose that at some time the particle is 12 cm to the right of the origin. Then after 2 seconds, the distance of the particle from the origin will be

$$12 - 2 \cdot 5 = 12 - 10 = 2 \text{ cm}.$$



Finally, suppose the particle does not start when $t = 0$ but starts later, say after 25 seconds, but still moves with the same constant speed. We could measure time in terms of a new coordinate t'. In terms of t', the x-coordinate is given by

$$x = -5t'.$$

We can give t' as a function of t, by

$$t' = t - 25.$$

Then

$$x = -5(t - 25)$$

gives x as a function of t.

In many applications, we have to consider related rates of change, which involve the chain rule. Suppose, that y is a function of x, and also that x is given as some function of time, say, $x = g(t)$, then we can determine both the rate of change of y with respect to x, namely dy/dx, but also the rate of change of y with respect to t, namely

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt}$$

by the chain rule.

Example. A square is expanding in such a way that its edge is changing at a rate of 2 cm/sec. When its edge is 6 cm long, find the rate of change of its area.

The area of a square as a function of its side is given by the function

$$A(x) = x^2.$$

If the side x is given as a function of time t, say $x = x(t)$, then the rate of change of the area with respect to time is by definition

$$\frac{d(A(x(t)))}{dt}.$$

Thus we use the chain rule, and if we denote the area by A, we find

$$\frac{dA}{dt} = \frac{dA}{dx}\frac{dx}{dt} = 2x\,\frac{dx}{dt}.$$

We are told that x increases at a rate of 2 cm/sec. This means that

$$\frac{dx}{dt} = 2.$$

Thus when $x(t) = 6$, we find that

$$\frac{dA}{dt} = 2 \cdot 6 \cdot 2 = 24 \text{ cm}^2/\text{sec}.$$

Example. A point moves along the graph of $y = x^3$ so that its x-coordinate is decreasing at the rate of 2 units per second. What is the rate of change of its y-coordinate when $x = 3$?

We have by the chain rule,

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt} = 3x^2 \frac{dx}{dt}.$$

We are told that x is *decreasing*. This means that

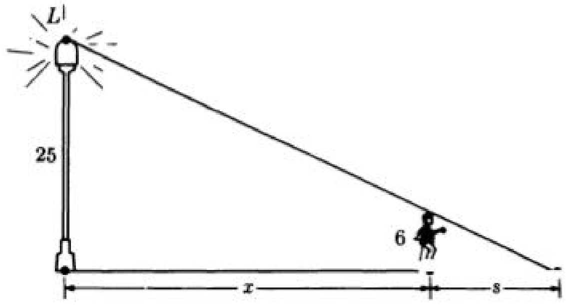
$$\frac{dx}{dt} = -2.$$

Hence the rate of change of y when $x = 3$ is equal to

$$\left.\frac{dy}{dt}\right|_{x=3} = 3(3^2)(-2) = -54 \text{ units per second.}$$

That dy/dt comes out negative means that y is decreasing when $x = 3$.

Example. A light on top of a lamppost shines 25 ft above the ground. A man 6 ft tall is walking away from the light. What is the length of his shadow when he is 40 ft away from the base of the lamppost? If he is walking away at the rate of 5 ft/sec, how fast is his shadow increasing at this point?



We need to establish a relationship between the length of the shadow and the distance of the man from the post. Let s be the length of the shadow, and let x be the distance between the man and the base of the post. Then using similar triangles, we see that

$$\frac{25}{x+s} = \frac{6}{s}.$$

After cross-multiplying we see that $25s = 6x + 6s$, whence

$$s = \tfrac{6}{19}x.$$

Therefore, $ds/dx = 6/19$, and

$$\frac{ds}{dt} = \frac{6}{19}\frac{dx}{dt}.$$

Since $dx/dt = 5$, we get what we want:

$$\frac{ds}{dt} = \frac{6}{19}\cdot 5 = \frac{30}{19} \text{ ft/sec.}$$

Also, when $x = 40$, we find that the length of his shadow is given by

$$s = \frac{6\cdot 40}{19} = \frac{240}{19} \text{ ft.}$$

Remark. If the man is walking toward the post, then the distance x is decreasing, and hence in this case,

$$\frac{dx}{dt} = -5.$$

Consequently, a similar argument shows that

$$\frac{ds}{dt} = -\frac{30}{19} \text{ ft/sec.}$$

Example. The area of a disc of radius r is given by the formula

$$A = \pi r^2,$$

where r is the radius. Let s be the diameter. Then $s = 2r$ so $r = s/2$ and we can give A as a function of s by

$$A = \pi\left(\frac{s}{2}\right)^2 = \frac{\pi s^2}{4}.$$

Hence the rate of change of A with respect to s is

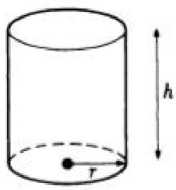
$$\frac{dA}{ds} = \frac{2\pi s}{4} = \frac{\pi s}{2}.$$

Example. A cylinder is being compressed from the side and stretched, so that the radius of the base is decreasing at a rate of 2 cm/sec and the height is increasing at a rate of 5 cm/sec. Find the rate at which the volume is changing when the radius is 6 cm and the height is 8 cm.

The volume is given by the formula

$$V = \pi r^2 h,$$

where r is the radius of the base and h is the height.



We are given $dr/dt = -2$ (note the negative sign because the radius is decreasing), and $dh/dt = 5$. Hence using the formula for the derivative of a product, we find

$$\frac{dV}{dt} = \pi\left[r^2 \frac{dh}{dt} + h2r \frac{dr}{dt}\right] = \pi[5r^2 - 4hr].$$

When $r = 6$ and $h = 8$ we get

$$\left.\frac{dV}{dt}\right|_{\substack{r=6\\h=8}} = \pi(5 \cdot 6^2 - 4 \cdot 8 \cdot 6) = -12\pi.$$

The negative sign in the answer means that the volume is decreasing when $r = 6$ and $h = 8$.

Example. Two trains leave a station 3 hours apart. The first one moves north at a speed of 100 km/hr. The second moves east at a speed of 60 km/hr. The second leaves 3 hours after the first. At what rate is the distance between the trains changing 2 hours after the second train has left?

Let y be the distance of the first train and x the distance of the second train from the station. Then

$$\frac{dy}{dt} = 100 \qquad \text{and} \qquad \frac{dx}{dt} = 60.$$

We have $y = 100t$, and since the second train leaves three hours later, we have

$$x = 60(t - 3).$$

Let $f(t)$ be the distance between them. Then

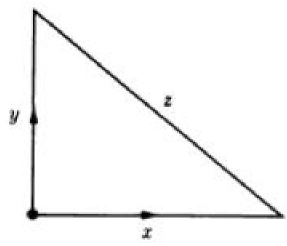
$$f(t) = \sqrt{60^2(t-3)^2 + 100^2t^2}.$$

Hence

$$f'(t) = \tfrac{1}{2}[3600(t-3)^2 + 10000t^2]^{-1/2}[2\cdot 60^2(t-3) + 2\cdot 100^2 t].$$

The time 2 hours after the second train leaves is $t = 2 + 3 = 5$. The desired rate is therefore $f'(5)$, so

$$f'(5) = \tfrac{1}{2}[14400 + 250000]^{-1/2}[14400 + 100000].$$



Example. We shall work the preceding example by using another method. Let z be the distance between the trains. We have

$$z^2 = x^2 + y^2.$$

Furthermore, x, y, z are functions of t. Differentiating with respect to t yields:

$$2z\frac{dz}{dt} = 2x\frac{dx}{dt} + 2y\frac{dy}{dt}.$$

We cancel 2 to get

$$z\frac{dz}{dt} = x\frac{dx}{dt} + y\frac{dy}{dt}.$$

Write $x(t)$, $y(t)$, $z(t)$ for x, y, z as functions of t. Then

$x(5) = 120$, $\quad y(5) = 500$, $\quad$ and by Pythagoras, $\quad z(5) = \sqrt{120^2 + 500^2}$.

Substituting $t = 5$, we find

$$z(5)\left.\frac{dz}{dt}\right|_{t=5} = x(5)\left.\frac{dx}{dt}\right|_{t=5} + y(5)\left.\frac{dy}{dt}\right|_{t=5}.$$

Dividing by $z(5)$ yields:

$$\left.\frac{dz}{dt}\right|_{t=5} = \frac{120\cdot 60 + 500\cdot 100}{\sqrt{120^2 + 500^2}}.$$

III, §9. EXERCISES

For some of these exercises, the following formulas can be used.

> Volume of a sphere of radius r is $\frac{4}{3}\pi r^3$.
> Area of a sphere of radius r is $4\pi r^2$.
> Volume of a cone of height h and radius of base r is $\frac{1}{3}\pi r^2 h$.
> Area of circle of radius r is πr^2.
> Circumference of circle of radius r is $2\pi r$.

1. A particle is moving so that at time t, the distance is given by $s(t) = t^3 - 2t$. At what time is the acceleration equal to
   (a) 1 (b) 0 (c) -5?

2. A particle is moving so that at time t, the distance is given by the function $s(t) = 2t^4 + t^2$. At what time is the speed equal to 0?

3. An object travels on a straight line with speed given by the function $v(t) = 4t^5$. Find the acceleration at time $t = 2$.

4. A particle is moving so that at time t, the distance traveled is given by $s(t) = t^3 - 2t + 1$. At what time is the acceleration equal to 0?

5. A cube is expanding in such a way that its edge is changing at a rate of 5 m/sec. When its edge is 4 m long, find the rate of change of its volume.

6. A sphere is increasing so that its radius increases at the rate of 1 cm/sec. How fast is its volume changing when its radius is 3 cm?

7. What is the rate of change of the area of a circle with respect to its radius, diameter, circumference?

8. A point moves along the graph of $y = 1/(x^2 + 4)$ in a way such that $dx/dt = -3$ units per second. What is the rate of change of its y-coordinate when $x = 2$?

9. A light shines on top of a lamppost 20 ft above the ground. A woman 5 ft tall walks away from the light. Find the rate at which her shadow is increasing if she walks at the rate of
   (a) 4 ft/sec (b) 3 ft/sec.

10. If in Exercise 9 the woman walks toward the light, find the rate at which her shadow is decreasing if she walks at the rate of
    (a) 5 ft/sec (b) 6 ft/sec.

11. A particle moves differentiably on the parabola $y = x^2$. At what point on the curve are its x- and y-coordinates moving at the same rate? (You may assume dx/dt and $dy/dt \neq 0$ for all t.)

12. One side of a right triangle decreases 1 cm/min and the other side increases 2 cm/min. At some time the first side is 8 cm and the second side is 6 cm long. At what rate is the area increasing 2 min after this time?

13. The length of the side of a square is increasing at the rate of 3 cm per second. Find the rate at which the area is increasing when the side is 15 cm long.

14. A ladder 17 ft long leans against a vertical wall. If the lower end of the ladder is being moved away from the foot of the wall at the rate of 3 ft/sec, how fast is the top descending when the lower end is 8 ft from the wall?

15. A swimming pool is 25 ft wide, 40 ft long, 3 ft deep at one end and 9 ft deep at the other, the bottom being an inclined plane. If water is pumped into the pool at the rate of 10 ft^3/min, how fast is the water level rising when it is 4 ft deep at the deep end?

16. A reservoir has the shape of a cone, vertex down, 10 ft high. The radius of the top is 4 ft. Water is poured into the reservoir at the rate of 5 ft^3/min. How fast is the water rising when the depth of the water is 5 ft?

17. A particle is moving so that at time t the distance traveled is given by $s(t) = 2t^2 - t$. At what time is the speed equal to 0? What is the acceleration of the particle?

18. A point moves on the parabola with equation $y = x^2 - 6x$. Find the point on the curve at which the rate of change of the y-coordinate is four times the rate of change of the x-coordinate. (Assume $dx/dt \neq 0$ for all x.)

19. Water is flowing into a tank in the form of a hemisphere of radius 10 ft with flat side up at the rate of 4 ft^3/min. Let h be the depth of the water, r the radius of the surface of the water, and V the volume of the water in the tank. Assume that $dV/dt = \pi r^2\, dh/dt$. Find how fast the water level is rising when $h = 5$ ft.

20. A train leaves a station at a certain time and travels north at the rate of 50 mi/hr. A second train leaves the same station 2 hr after the first train leaves, and goes east at the rate of 60 mi/hr. Find the rate at which the two trains are separating 1.5 hr after the second train leaves the station.

21. Sand is falling on a pile, always having the shape of a cone, at the rate of 3 ft^3/min. Assume that the diameter at the base of the pile is always three times the altitude. At what rate is the altitude increasing when the altitude is 4 ft?

22. The volume of a sphere is decreasing at the rate of 12π cm^3/min. Find the rate at which the radius and the surface area of the sphere are changing when the radius is 20 cm.

23. Water runs into a conical reservoir at the constant rate of 2 m^3/min. The vertex is 18 m down and the radius of the top is 24 m. How fast is the water level rising when it is 6 m deep?

24. Sand is falling on a pile, forming a cone. Let V be the volume of sand, so $V = \frac{1}{3}\pi r^2 h$. What is the rate of change of the volume of sand when $r = 10$ ft, if the radius of the base is expanding at a rate of 2 ft/sec, and the height is increasing at a rate of 1 ft/sec? You may assume $r = h = 0$ when $t = 0$.

CHAPTER IV

Sine and Cosine

From the sine of an angle and the cosine of an angle, we shall define functions of numbers, and determine their derivatives.

It is convenient to recall all the facts about trigonometry which we need in the sequel, especially the formula giving us the sine and cosine of the sum of two angles. Thus our treatment of the trigonometric functions is self-contained—you do not need to know anything about sine and cosine before starting to read this chapter. However, most of the proofs of statements in §1 come from plane geometry and will be left to you.

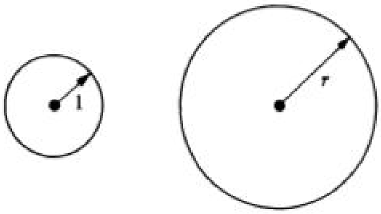
IV, §0. REVIEW OF RADIAN MEASURE

In order to eliminate some confusion of terminology, it is often convenient to use two different words for a circle, and a circle together with the region lying inside. Thus we reserve the word **circle** for the former, and call the circle together with its inside a **disc**. Thus we speak of the length of a circle, but the area of a disc.

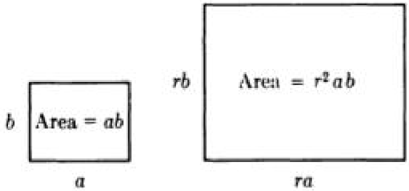
We suppose fixed a measure of length. This determines a measure of area. For instance, if length is measured in meters, then area is measured in square meters.

For our immediate purposes, we define π to be the **area of the disc of radius** 1. It is, of course, a problem to find a decimal expansion for π, which you probably have been told is approximately equal to 3.14159.... Later in the course, you will learn to compute π to any degree of accuracy.

The disc of radius r is obtained by a dilation (blow up) of the disc of radius 1, as shown on the figure at the top of the next page.



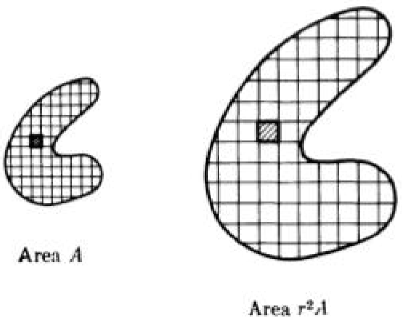
How does area change under dilation? Let us first look at rectangles. Let R be a rectangle whose sides have length a, b. Let r be a positive number. Let rR be the rectangle whose sides have length ra, rb, as shown on the figure. Then the area of rR is $rarb = r^2ab$. If A is the area of R, then the area of rR is r^2A.



Under dilation by a factor r the area of a rectangle changes by a factor r^2. This applies as a general principle to any region:

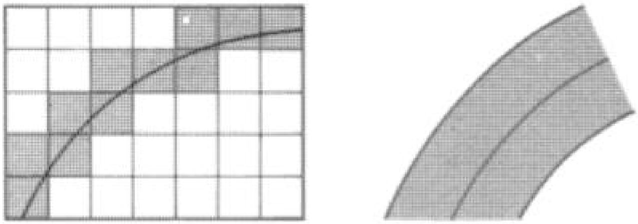
> **Let S be any region in the plane. Let A be the area of S. Under dilation by a factor r, the dilated region has area r^2A.**

We shall prove this by approximating S by the squares of a grid, as shown on the figure. If we blow up S by a factor of r, we obtain a region rS. Let A be the area of S. Then the area of rS will again be r^2A, because each small square is blown up by a factor of r, so the area of a small square changes by a factor of r^2. The sum of the areas of the squares gives an approximation to the area of the figure. We want to



estimate how good the approximation is. The difference between the sum of the areas of all the little squares contained in the figure and the area of the figure itself is at most the area of all the small squares which touch the boundary of the figure. We can give an estimate for this as follows.

Suppose we make a grid so that the squares have sides of length c. Then the diagonal of such a square has length $c\sqrt{2}$. If a square intersects the boundary, then any point on the square is at distance at most $c\sqrt{2}$ from the boundary. Look at the figure. This is because the distance between any two points of the square is at most $c\sqrt{2}$. Let us



(a) (b)

draw a band of width $c\sqrt{2}$ on each side of the boundary, as shown in (b) of the above figure. Then all the squares which intersect the boundary must lie within that band. It is very plausible that the area of the band is at most equal to

$$2c\sqrt{2} \quad \text{times the length of the boundary.}$$

Thus if we take c to be very small, i.e. if we take the grid to be a very fine grid, then the area of the band is small, and the area of the figure is approximated by the area covered by the squares lying entirely inside the figure. Under dilation, a similar argument applies to the dilated band for the dilated figure, so that the area of the dilated band is at most

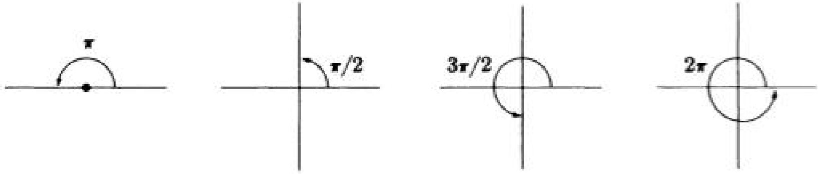
$$2r^2 \cdot c\sqrt{2} \quad \text{times the length of the boundary.}$$

As c approaches 0, the areas of these bands approach 0. This justifies our assertion that area changes by a factor of r^2 under dilation by a factor of r.

Since we defined π to be the area of the disc of radius 1, we now obtain:

The area of a disc of radius r is πr^2.

We select a unit of measurement for angles such that the flat angle is equal to π times the unit angle. (See the following figure.) The right angle has measure $\pi/2$. The full angle going once around has then measure 2π. This unit of measurement for which the flat angle has measure π is called the **radian**. Thus the right angle has $\pi/2$ radians.



There is another current unit of measurement for which the flat angle is 180. This unit is called the **degree**. Thus the flat angle has 180 degrees, and the right angle has 90 degrees. We also have

$$360 \text{ degrees} = 2\pi \text{ radians},$$

$$60 \text{ degrees} = \pi/3 \text{ radians},$$

$$45 \text{ degrees} = \pi/4 \text{ radians},$$

$$30 \text{ degrees} = \pi/6 \text{ radians}.$$

We shall deal mostly with radian measure, which makes some formulas come out more easily later on. It is easy to convert from one measure to the other.

Example. A wheel is turning at the rate of 50° per minute. Find its rate in rad/min and rpm (revolutions per minute).

We have

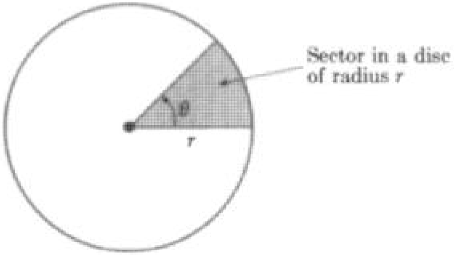
$$1^\circ = \frac{2\pi}{360} \text{ radians} = \frac{\pi}{180} \text{ radians}.$$

Hence 50° per minute is equal to $50\pi/180 = 5\pi/18$ radians per minute.

On the other hand, a full revolution is 2π radians, so 1 radian is $1/2\pi$ revolutions. Hence the wheel is turning at the rate of

$$\frac{5\pi}{18} \cdot \frac{1}{2\pi} = \frac{5}{36} \text{ rpm}.$$

A **sector** is the region of the plane lying inside an angle. Often we also speak of a **sector in a disc**, meaning the portion of the sector lying inside the disc, as illustrated on the figure.



A sector is measured by its angle. In the figure, we have labeled this angle θ (theta), with $0 \leqq \theta \leqq 2\pi$. It is measured in radians. The area of the sector is a certain fraction of the total area of the disc. Namely, we know that the total area is πr^2. The fraction is $\theta/2\pi$. Hence if we let S be the sector having angle θ in a disc of radius r, then the area of S is

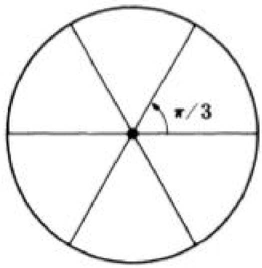
$$\frac{\theta}{2\pi} \cdot \pi r^2 = \frac{\theta r^2}{2}.$$

We box this for reference:

$$\boxed{\textbf{Area of sector of angle } \theta \textbf{ radians in disc of radius } r = \frac{\theta r^2}{2}.}$$

If the radius is 1, then the area of the sector is $\theta/2$. We shall use this in §4.

Example. The area of a sector of angle $\pi/3$ in a disc of radius 1 is $\pi/6$, because the total area of the disc is π, and the sector represents one-sixth of the total area. This is illustrated on the figure.



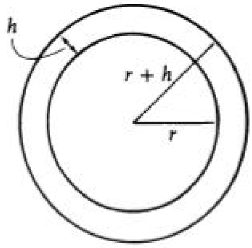
The disc in the figure is cut up into six sectors, each angle measuring $\pi/3$ radians, for a total of 2π radians.

Next we come to the length of arc of the circle.

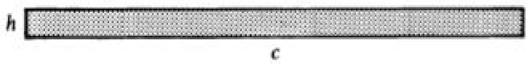
Let c be the circumference of a circle of radius r. Then

$$\boxed{c = 2\pi r.}$$

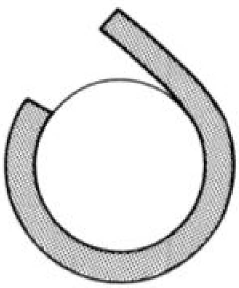
Proof. The proof will be a beautiful example of the idea of differentiation. We consider a circle of radius r, and a circle of slightly bigger radius, which we write $r + h$. We suppose that these circles have the same center, and so obtain a circular band between them.



Let c be the length of the small circle. If we had a rectangular band of length c and height h then its area would be ch.



Suppose we wrap this band around the small circle.



Since the circle curves outward, the rectangular band has to be stretched if it is to cover the band between the circle of radius r and the circle of radius $r + h$. Thus we have an inequality for areas:

$$ch < \text{area of circular band}.$$

Similarly, if C is the circumference of the bigger circle, then

$$\text{area of circular band} < Ch.$$

But the area of the circular band is the difference between the areas of the discs, which is

$$\begin{aligned}\text{area of circular band} &= \pi(r+h)^2 - \pi r^2 \\ &= \pi(2rh + h^2).\end{aligned}$$

Therefore we obtain the inequalities

$$ch < \pi(2rh + h^2) < Ch.$$

We divide these inequalities by the positive number h to obtain

$$c < \pi(2r + h) < C.$$

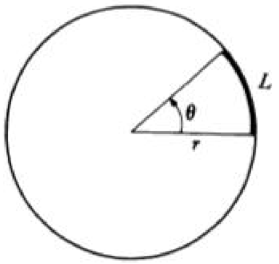
Now let h approach 0. Then the circumference of the big circle C approaches the circumference of the small circle c, and $\pi(2r + h)$ approaches $2\pi r$. It follows that

$$c = 2\pi r,$$

thus proving our formula.

Observe that the length of the circumference is just the derivative of the area.

An arc on a circle can be measured by its angle in radians. What is the length L of this arc, as on the following figure?



The answer is as follows.

Let L be the length of an arc of θ radians on a circle of radius r. Then

$$\boxed{L = r\theta.}$$

Proof. The total length of the circle is $2\pi r$, and L is the fraction $\theta/2\pi$ of $2\pi r$, which gives precisely $r\theta$.

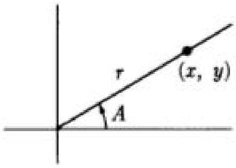
Example. The length of arc of $\pi/3$ radians in a circle of radius r is $r\pi/3$.

IV, §0. EXERCISE

Suppose you are given that the volume of a sphere of radius r is $\frac{4}{3}\pi r^3$. Can you figure out an argument to obtain the area of the sphere?

IV, §1. THE SINE AND COSINE FUNCTIONS

Suppose that we have given coordinate axes, and a certain angle A, as shown on the figure.



We select a point (x, y) (not the origin) on the ray determining our angle A. We let $r = \sqrt{x^2 + y^2}$. Then r is the distance from $(0, 0)$ to the point (x, y). We define

$$\boxed{\begin{aligned} \text{sine } A &= \frac{y}{r} = \frac{y}{\sqrt{x^2 + y^2}}, \\ \text{cosine } A &= \frac{x}{r} = \frac{x}{\sqrt{x^2 + y^2}}. \end{aligned}}$$

If we select another point (x_1, y_1) on the ray determining our angle A and use its coordinates to get the sine and cosine, then we shall obtain the same values as with (x, y). Indeed, there is a positive number c such that

$$x_1 = cx \quad \text{and} \quad y_1 = cy.$$

Thus we can substitute these values in

$$\frac{y_1}{\sqrt{x_1^2 + y_1^2}}$$

to obtain

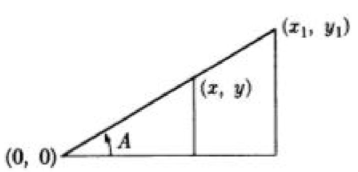
$$\frac{y_1}{\sqrt{x_1^2 + y_1^2}} = \frac{cy}{\sqrt{c^2x^2 + c^2y^2}}.$$

We can factor c from the denominator, and then cancel c in both the numerator and denominator to get

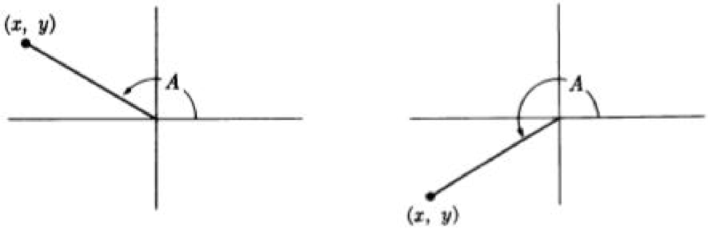
$$\frac{y}{\sqrt{x^2 + y^2}}.$$

In this way we see that sine A does not depend on the choice of the point (x, y).

The geometric interpretation of the above argument simply states that the triangles in the following diagram are similar.



The angle A can go all the way around. For instance, we could have an angle determined by a point (x, y) in the second or third quadrant.



When the angle A is in the first quadrant, then its sine and cosine are positive because both coordinates x, y are positive. When the angle A is in the second quadrant, its sine is positive because y is positive, but its cosine is negative because x is negative.

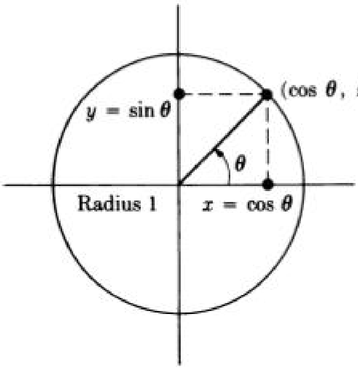
When A is in the third quadrant, sine A is negative and cosine A is negative also.

As mentioned in the introductory section, we use radian measure for angles. Suppose A is an angle of θ radians, and let (x, y) be a point on the circle of radius 1, also lying on the line determining the angle A as shown on the figure. Then in this case,

$$r = \sqrt{x^2 + y^2} = 1$$

and therefore

$$\boxed{(x, y) = (\cos\theta,\ \sin\theta).}$$



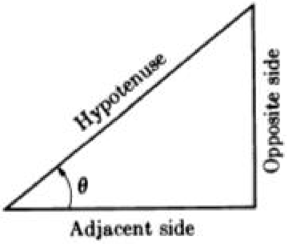
In general for arbitrary radius r, we have the relations:

$$\boxed{\begin{aligned} x &= r\cos\theta, \\ y &= r\sin\theta. \end{aligned}}$$

We can also define the sine and cosine of angles in a right triangle as shown on the figure, by the formulas:

$$\sin\theta = \frac{\text{Opposite side}}{\text{Hypotenuse}}$$

$$\cos\theta = \frac{\text{Adjacent side}}{\text{Hypotenuse}}$$

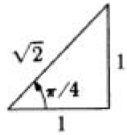


We make a table of the sines and cosines of some angles.

Angle	Sine	Cosine
$\pi/6$	$1/2$	$\sqrt{3}/2$
$\pi/4$	$1/\sqrt{2}$	$1/\sqrt{2}$
$\pi/3$	$\sqrt{3}/2$	$1/2$
$\pi/2$	1	0
π	0	-1
2π	0	1

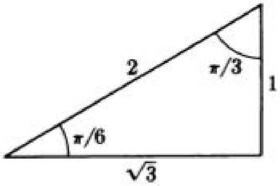
Unless otherwise specified, we *always use the radian measure*, and our table is given for this measure.

The values of this table are easily determined, using properties of similar triangles and plane geometry. For instance, we get the sine of the angle $\pi/4$ radians $= 45°$ from a right triangle with two equal sides:



We can determine the sine of $\pi/4$ by means of the point $(1, 1)$. Then $r = \sqrt{2}$ and sine $\pi/4$ radians is $1/\sqrt{2}$. Similarly for the cosine.

We get the sine of an angle of $\pi/6$ radians by considering a triangle such that two angles have $\pi/6$ and $\pi/3$ radians (that is, 30° and 60° respectively).



If we let the side opposite the angle of 30° have length 1, then the hypotenuse has length 2, and the side adjacent to the angle of 30° has length $\sqrt{3}$.

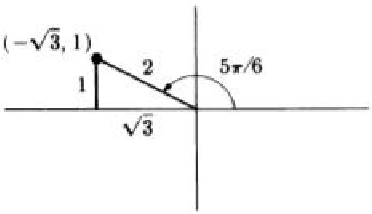
Hence we find that

$$\boxed{\sin \pi/6 = \frac{1}{2} \quad \text{and} \quad \cos \pi/6 = \frac{\sqrt{3}}{2}.}$$

On the other hand, we have

$$\boxed{\sin 5\pi/6 = \frac{1}{2}, \qquad \cos 5\pi/6 = \frac{-\sqrt{3}}{2}}$$

as is clear from the figure.



The choice of radian measure allows us to define the **sine of a number** rather than the sine of an angle as follows.

Let x be a number with $0 \leqq x < 2\pi$. We define **sin** $\boldsymbol{x}$ to be the sine of x radians.

For an arbitrary number x, we write

$$x = x_0 + 2\pi n$$

where n is an integer, $0 \leqq x_0 < 2\pi$, and define

$$\sin x = \sin x_0.$$

From this definition, we see that

$$\boxed{\sin x = \sin(x + 2\pi) = \sin(x + 2\pi n)}$$

for any integer n.

Of course, we call this function the **sine function**. Thus $\sin \pi = 0$, $\sin \pi/2 = 1$, $\sin 2\pi = 0$, $\sin 0 = 0$.

Similarly, we have the **cosine function**, which is defined for all numbers x by the rule:

> $\cos x$ is the number which is the cosine of the angle x radians.

Thus $\cos 0 = 1$ and $\cos \pi = -1$.

If we had used the measure of angles in degrees we would obtain *another* sine function which is not equal to the sine function which we defined in terms of radians. Suppose we call this other sine function $\sin^*$. Then

$$\sin^*(180) = \sin \pi,$$

and in general

$$\sin^*(180x) = \sin \pi x$$

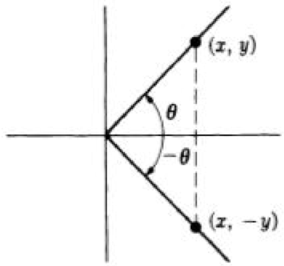
for any number x. Thus

$$\sin^* x = \sin\left(\frac{\pi}{180} x\right)$$

is the formula relating our two sine functions. It will become clear later why we always pick the radian measure instead of any other.

At present we have no means of computing values for the sine and cosine other than the very special cases listed above (and similar ones, based on simple symmetries of right triangles). In Chapter XIII we shall develop a method which will allow us to find $\sin x$ and $\cos x$ for any value of x, up to any degree of accuracy that you wish.

On the next figure we illustrate an angle of θ radians, and by convention the angle of $-\theta$ radians is the reflection around the x-axis.



If (x, y) are the coordinates of a point on the ray defining the angle of θ radians, then $(x, -y)$ are the coordinates of a point on the reflected ray. Thus

$$\boxed{\begin{aligned} \sin(-\theta) &= -\sin\theta, \\ \cos(-\theta) &= \cos\theta. \end{aligned}}$$

Finally we recall the definitions of the other trigonometric functions:

$$\tan\theta = \frac{\sin\theta}{\cos\theta}, \qquad \cot\theta = \frac{1}{\tan\theta} = \frac{\cos\theta}{\sin\theta},$$

$$\sec\theta = \frac{1}{\cos\theta}, \qquad \csc\theta = \frac{1}{\sin\theta}.$$

The most important among these are the sine, cosine, and tangent. We make a few additional remarks on the tangent.

The tangent is of course defined for all numbers θ such that

$$\cos\theta \neq 0.$$

These are the numbers unequal to

$$\pm\frac{\pi}{2}, \quad \pm\frac{3\pi}{2}, \quad \pm\frac{5\pi}{2}, \ldots$$

in general, $\theta \neq (2n + 1)/2$ for some integer n. We make a table of some values of the tangent.

θ	$\tan\theta$
0	0
$\pi/6$	$1/\sqrt{3}$
$\pi/4$	1
$\pi/3$	$\sqrt{3}$

You should complete this table for all similar values of θ in all four quadrants.

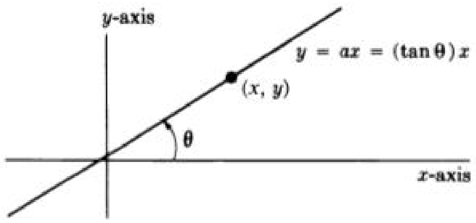
Consider an angle of θ radians, and let (x, y) be a point on the ray determining this angle, with $x \neq 0$. Then

$$\tan\theta = y/x$$

so that

$$y = (\tan\theta)x.$$

Conversely, any point (x, y) satisfying this equation is a point on the line making an angle θ with the x-axis, as shown on the next figure.



In an equation $y = ax$ where a is the slope, we can say that

$$a = \tan\theta,$$

where θ is the angle which the line makes with the x-axis.

Example. Take $\theta = \pi/6$. Then $\tan\theta = 1/\sqrt{3}$. Hence the line making an angle of θ with the x-axis has the equation

$$y = \left(\tan\frac{\pi}{6}\right)x, \qquad \text{or also} \qquad y = \frac{1}{\sqrt{3}}x.$$

Example. Take $\theta = 1$. There is no easier way to express $\tan 1$ other than writing $\tan 1$. In Chapter XIII you will learn how to compute arbitrarily close decimal approximations. Here we don't care. We simply point out that the equation of the line making an angle of 1 radian with the x-axis is given by

$$y = (\tan 1)x.$$

Similarly, the equation of the line making an angle of 2 radians with the x-axis is given by

$$y = (\tan 2)x.$$

Suppose given that π is approximately equal to 3.14. Then 1 is approximately equal to $\pi/3$. Thus the line making an angle of 1 radian with the x-axis, as shown on Fig. 1(a) is close to the line making an angle of $\pi/3$ radians. Similarly, the line making an angle of 2 radians with the x-axis, as shown on Fig. 1(b), is close to the line making an angle of $2\pi/3$ radians.

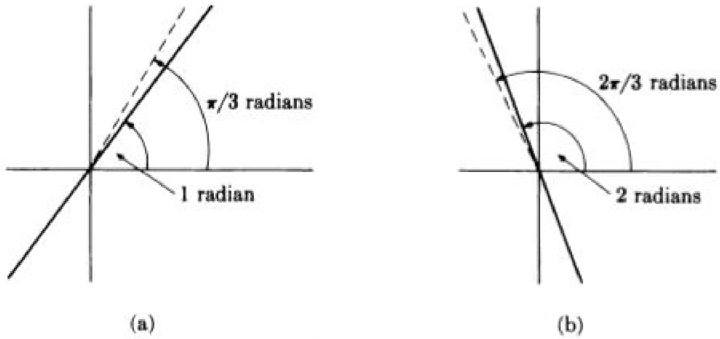


Figure 1

IV, §1. EXERCISES

Find the following values of the sin function and cos function:

1. $\sin 3\pi/4$
2. $\sin 2\pi/6$
3. $\sin \dfrac{2\pi}{3}$
4. $\sin\left(\pi - \dfrac{\pi}{6}\right)$
5. $\cos\left(\pi + \dfrac{\pi}{6}\right)$
6. $\cos\left(\pi + \dfrac{2\pi}{6}\right)$
7. $\cos\left(2\pi - \dfrac{\pi}{6}\right)$
8. $\cos \dfrac{5\pi}{4}$

Find the following values:

9. $\tan \dfrac{\pi}{4}$
10. $\tan \dfrac{2\pi}{6}$
11. $\tan \dfrac{5\pi}{4}$
12. $\tan\left(2\pi - \dfrac{\pi}{4}\right)$
13. $\sin \dfrac{7\pi}{6}$
14. $\cos \dfrac{7\pi}{6}$
15. $\cos 2\pi/3$
16. $\cos(-\pi/6)$
17. $\cos(-5\pi/6)$
18. $\cos(-\pi/3)$

19. Complete the following table.

θ	$\sin\theta$	$\cos\theta$	$\tan\theta$
$2\pi/3$			
$3\pi/4$			
$5\pi/6$			
π			
$7\pi/6$			
$5\pi/4$			
$7\pi/4$			

IV, §2. THE GRAPHS

sin *x*

We wish to sketch the graph of the sine function.

We know that $\sin 0 = 0$. As x goes from 0 to $\pi/2$, the sine of x increases until x reaches $\pi/2$, at which point the sine is equal to 1.

As x ranges from $\pi/2$ to π, the sine decreases until it becomes

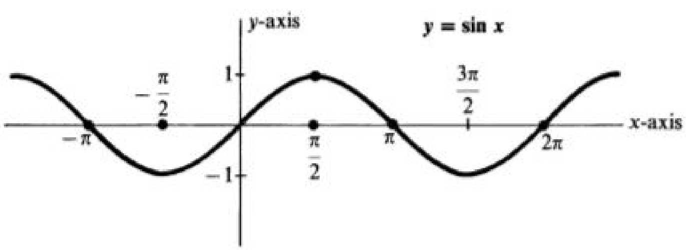
$$\sin \pi = 0.$$

As x ranges from π to $3\pi/2$ the sine becomes negative, but otherwise behaves in a similar way to the first quadrant, until it reaches

$$\sin \frac{3\pi}{2} = -1.$$

Finally, as x goes from $3\pi/2$ to 2π, the sine of x goes from -1 to 0, and we are ready to start all over again.

The graph looks like this:



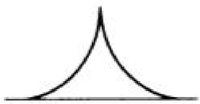
If we go once around by 2π, the sine and cosine each take on the same values they had originally, in other words

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

$$\cos(x + 2\pi) = \cos x$$

for all x. This holds whether x is positive or negative, and the same would be true if we took $x - 2\pi$ instead of $x + 2\pi$.

You might legitimately ask why one arch of the sine (or cosine) curve looks the way we have drawn it, and not the following way:

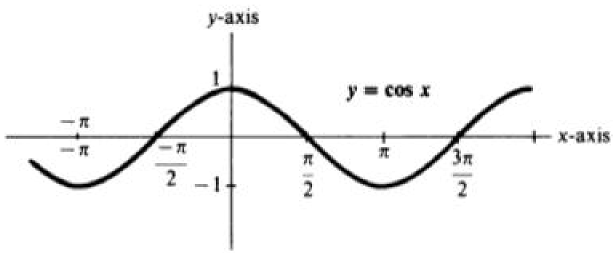


In the next section, we shall find the slope of the curve $y = \sin x$. It is equal to $\cos x$. Thus when $x = 0$, the slope is $\cos 0 = 1$. Furthermore, when $x = \pi/2$, we have $\cos \pi/2 = 0$ and hence the slope is 0. This means that the curve becomes horizontal, and cannot have a peak the way we have drawn it above.

At present we have no means for computing more values of $\sin x$ and $\cos x$. However, using the few that we know and the derivative, we can convince ourselves that the graph looks as we have drawn it.

cos *x*

The graph of the cosine will look like that of the sine, but it starts with $\cos 0 = 1$.

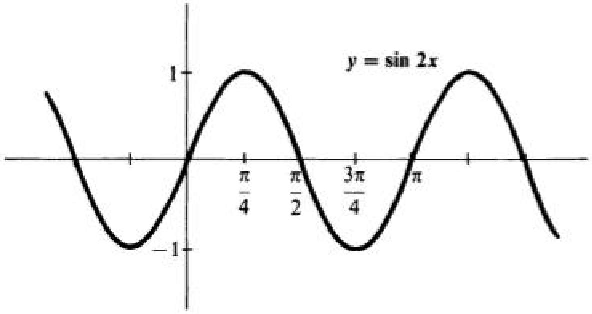


sin 2*x*

Next let us graph $y = f(x) = \sin 2x$. Since the sine changes its behavior in intervals of length $\pi/2$, $\sin 2x$ will change behavior in intervals of length $\pi/4$. Thus we make a table with x ranging over intervals of length $\pi/4$.

x	$2x$	$\sin 2x$
inc. 0 to $\pi/4$	inc. 0 to $\pi/2$	inc. 0 to 1
inc. $\pi/4$ to $\pi/2$	inc. $\pi/2$ to π	dec. 1 to 0
inc. $\pi/2$ to $3\pi/4$	inc. π to $3\pi/2$	dec. 0 to -1
inc. $3\pi/4$ to π	inc. $3\pi/2$ to 2π	inc. -1 to 0

Then the graph repeats. Hence the graph of $\sin 2x$ looks like this.



We see that the graph of $y = \sin 2x$ has twice as many wiggles over an interval as the graph of $y = \sin x$.

Similarly, you would see that $y = \sin \frac{1}{2}x$ has half as many wiggles as the graph of $y = \sin x$.

tan x

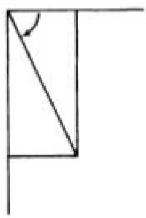
Finally, let us graph $y = \tan x$. Note that $\tan 0 = 0$. We take the interval

$$-\frac{\pi}{2} < x < \frac{\pi}{2}.$$

If x is near $-\pi/2$ then the tangent is very large negative. Namely

$$\tan x = \frac{\sin x}{\cos x}.$$

When x is near $-\pi/2$ then $\cos x$ is near 0 and $\sin x$ is near -1. You can also see this from a right triangle.



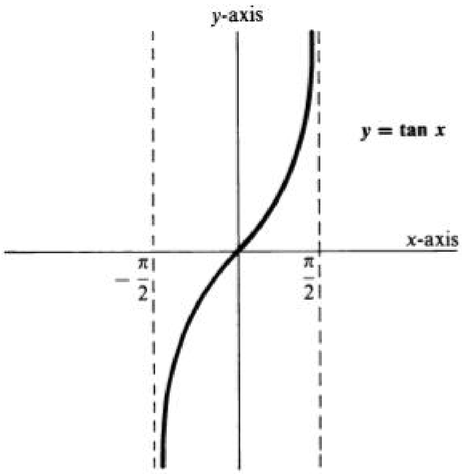
As x increases from $-\pi/2$ to 0, $\sin x$ increases from -1 to 0. On the other hand, as x increases from $-\pi/2$ to 0, $\cos x$ increases from 0 to 1. Hence $1/\cos x$ decreases from very large negative to 1. Hence as x increases from $-\pi/2$ to 0,

$$\tan x = \frac{\sin x}{\cos x} \quad \text{increases from large negative to 0.}$$

Similarly, as x increases from 0 to $\pi/2$, $\sin x$ increases from 0 to 1 and $\cos x$ decreases from 1 to 0. Hence $1/\cos x$ increases from 1 to very large positive, and so

$$\tan x = \frac{\sin x}{\cos x} \quad \text{increases from 0 to large positive.}$$

Thus the graph of $y = \tan x$ looks like this.



Remark on notation. Do not confuse $\sin 2x$ and $(\sin 2)x$. We usually write

$$\sin(2x) = \sin 2x$$

without parentheses to mean the sine of $2x$. On the other hand, $(\sin 2)x$ is the number $\sin 2$ times x. The graph of

$$y = (\sin 2)x$$

is a straight line, just like $y = cx$ for some fixed number c.

IV, §2. EXERCISES

1. Draw the graph of $\tan x$ for all values of x.

2. Let $\sec x = 1/\cos x$ be defined when $\cos x \neq 0$. Draw the graph of $\sec x$.

3. Let $\cot x = 1/\tan x$. Draw the graph of $\cot x$.
(Sec and cot are abbreviations for the secant and cotangent.)

4. Sketch the graphs of the following functions:
   (a) $y = \sin 2x$ (b) $y = \sin 3x$
   (c) $y = \cos 2x$ (d) $y = \cos 3x$

5. Sketch the graphs of the following functions:
   (a) $y = \sin \frac{1}{2}x$ (b) $y = \sin \frac{1}{3}x$ (c) $y = \sin \frac{1}{4}x$
   (d) $y = \cos \frac{1}{2}x$ (e) $y = \cos \frac{1}{3}x$ (f) $y = \cos \frac{1}{4}x$

6. Sketch the graphs of:
   (a) $y = \sin \pi x$ (b) $y = \cos \pi x$
   (c) $y = \sin 2\pi x$ (d) $y = \cos 2\pi x$

7. Sketch the graph of the following functions:
   (a) $y = |\sin x|$ (b) $y = |\cos x|$.

8. Let $f(x) = \sin x + \cos x$. Plot approximate values of $f(n\pi/4)$, for

$$n = 0, 1, 2, 3, 4, 5, 6, 7, 8.$$

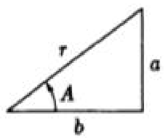
IV, §3. ADDITION FORMULA

In this section we shall state and prove the most important formulas about sine and cosine.

To begin with, using the Pythagoras theorem, we shall see that

$$\boxed{(\sin \theta)^2 + (\cos \theta)^2 = 1}$$

for all numbers θ. To show this, we take an angle A and we determine its sine and cosine from the right triangle, as in the following figure.



Then by Pythagoras we have

$$a^2 + b^2 = r^2.$$

Dividing by r^2 yields

$$\left(\frac{a}{r}\right)^2 + \left(\frac{b}{r}\right)^2 = 1.$$

We drew the angle A to be between 0 and $\pi/2$. But in general, referring back to the definitions, we have

$$\sin^2\theta + \cos^2\theta = \frac{y^2}{r^2} + \frac{x^2}{r^2}$$
$$= \frac{x^2 + y^2}{r^2} = \frac{r^2}{r^2} = 1.$$
$$= \frac{r^2}{r^2} = 1.$$

It is customary to write the square of the sine and cosine as $\sin^2 A$ and $\cos^2 A$. In the second case note that b is negative.

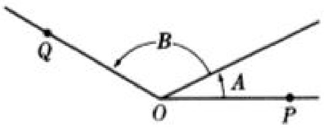
Our main result is the **addition formula**, which should be memorized.

Theorem 3.1. *For any angles A and B, we have*

$$\boxed{\begin{aligned}\sin(A+B) &= \sin A\cos B + \cos A\sin B,\\ \cos(A+B) &= \cos A\cos B - \sin A\sin B.\end{aligned}}$$

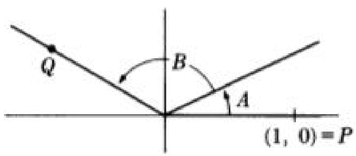
Proof. We shall prove the second formula first.

We consider two angles A, B and their sum:



We take two points P, Q as indicated, at a distance 1 from the origin O. We shall now compute the distance from P to Q, using two different coordinate systems.

First, we take a coordinate system as usual:



Then $P = (1, 0)$ and

$$Q = (\cos(A+B), \sin(A+B)).$$

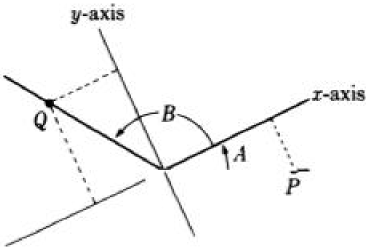
The square of the distance between $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ is

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

Hence

$$\begin{aligned}\operatorname{dist}(P, Q)^2 &= \sin^2(A + B) + (\cos(A + B) - 1)^2,\\ &= -2\cos(A + B) + 2.\end{aligned}$$

Next we place the coordinate system as shown in the figure below.



Then the coordinates of P become

$$(\cos(-A), \sin(-A)) = (\cos A, -\sin A).$$

Those of Q are simply $(\cos B, \sin B)$. Hence

$$\begin{aligned}\operatorname{dist}(P, Q)^2 &= (\sin B + \sin A)^2 + (\cos B - \cos A)^2,\\ &= \sin^2 B + 2\sin B \sin A + \sin^2 A\\ &\quad + \cos^2 B - 2\cos B\cos A + \cos^2 A\\ &= 2 + 2\sin A \sin B - 2\cos A\cos B.\end{aligned}$$

If we set the squares of the two distances equal to each other, we get our formula.

From the addition formula for the cosine, we get some formulas relating sine and cosine.

$$\boxed{\begin{aligned}\sin x &= \cos\left(x - \frac{\pi}{2}\right),\\ \cos x &= \sin\left(x + \frac{\pi}{2}\right).\end{aligned}}$$

To prove the first one, we start with the right-hand side:

$$\begin{aligned}\cos\left(x - \frac{\pi}{2}\right) &= \cos x \cos\left(-\frac{\pi}{2}\right) - \sin x \sin\left(-\frac{\pi}{2}\right)\\ &= 0 + \sin x = \sin x\end{aligned}$$

because $\cos(-\pi/2) = 0$ and $-\sin(-\theta) = \sin\theta$ (with $\theta = \pi/2$).

The second relation follows from the first, namely in the first relation put $x = z + \pi/2$. Then

$$\sin\left(z + \frac{\pi}{2}\right) = \cos\left(z + \frac{\pi}{2} - \frac{\pi}{2}\right) = \cos z.$$

This proves the second relation.

Similarly, you can prove

$$\sin\left(\frac{\pi}{2} - x\right) = \cos x.$$

This will be used in the next argument.

The addition formula for the sine can be obtained from the addition formula for the cosine by the following device:

$$\begin{aligned}\sin(A + B) &= \cos\left(A + B - \frac{\pi}{2}\right) \\ &= \cos A \cos\left(B - \frac{\pi}{2}\right) - \sin A \sin\left(B - \frac{\pi}{2}\right) \\ &= \cos A \sin B + \sin A \sin\left(\frac{\pi}{2} - B\right) \\ &= \cos A \sin B + \sin A \cos B,\end{aligned}$$

thereby proving the addition formula for the sine.

Example. Find $\sin(\pi/12)$.

We write

$$\frac{\pi}{12} = \frac{\pi}{3} - \frac{\pi}{4}.$$

Then $\sin(\pi/12) = \sin(\pi/3)\cos(\pi/4) - \cos(\pi/3)\sin(\pi/4)$, and substituting the known values we find

$$\sin(\pi/12) = \frac{\sqrt{3} - 1}{2\sqrt{2}}.$$

Example. We have

$$\cos\left(x + \frac{\pi}{2}\right) = -\sin x,$$

because

$$\cos\left(x + \frac{\pi}{2}\right) = \cos x \cos \frac{\pi}{2} - \sin x \sin \frac{\pi}{2} = -\sin x,$$

since $\cos \pi/2 = 0$ and $\sin \pi/2 = 1$.

In the exercises, you will derive a few more useful formulas for the sine and cosine, notably:

$$\boxed{\begin{gathered} \sin 2x = 2 \sin x \cos x, \\ \cos 2x = \cos^2 x - \sin^2 x, \\ \cos^2 x = \frac{1 + \cos 2x}{2} \quad \text{and} \quad \sin^2 x = \frac{1 - \cos 2x}{2}. \end{gathered}}$$

You will remember them better for having worked them out, so we don't spoil this in the text.

IV, §3. EXERCISES

1. Find $\sin 7\pi/12$. [*Hint*: Write $7\pi/12 = 4\pi/12 + 3\pi/12$.]

2. Find $\cos 7\pi/12$.

3. Find the following values:
   (a) $\sin \pi/12$ (b) $\cos \pi/12$
   (c) $\sin 5\pi/12$ (d) $\cos 5\pi/12$
   (e) $\sin 11\pi/12$ (f) $\cos 11\pi/12$
   (g) $\sin 10\pi/12$ (h) $\cos 10\pi/12$

4. Prove the following formulas.
   (a) $\sin 2x = 2 \sin x \cos x$ (b) $\cos 2x = \cos^2 x - \sin^2 x$
   (c) $\cos^2 x = \dfrac{1 + \cos 2x}{2}$ (d) $\sin^2 x = \dfrac{1 - \cos 2x}{2}$

   [*Hint*: For (c) and (d), start with the special case of the addition formula for $\cos 2x$. Then use the identity

   $$\sin^2 x + \cos^2 x = 1.]$$

5. Find a formula for $\sin 3x$ in terms of $\sin x$ and $\cos x$. Similarly for $\cos 3x$.

6. Prove that $\sin(\pi/2 - x) = \cos x$, using only the addition formula for the cosine.

7. Prove the formulas

$$\sin mx \sin nx = \tfrac{1}{2}[\cos(m - n)x - \cos(m + n)x],$$

$$\sin mx \cos nx = \tfrac{1}{2}[\sin(m + n)x + \sin(m - n)x],$$

$$\cos mx \cos nx = \tfrac{1}{2}[\cos(m + n)x + \cos(m - n)x].$$

[*Hint*: Expand the right-hand side by the addition formula, and then cancel as much as you can. The left-hand side should be forthcoming. Note that, for instance,

$$\begin{aligned}\cos(m-n)x &= \cos(mx - nx)\\ &= \cos mx \cos nx + \sin mx \sin nx.]\end{aligned}$$

IV, §4. THE DERIVATIVES

We shall prove:

Theorem 4.1. *The functions* $\sin x$ *and* $\cos x$ *have derivatives and*

$$\boxed{\begin{aligned}\frac{d(\sin x)}{dx} &= \cos x,\\ \frac{d(\cos x)}{dx} &= -\sin x.\end{aligned}}$$

Proof. We shall first determine the derivative of $\sin x$. We have to look at the Newton quotient of $\sin x$. It is

$$\frac{\sin(x+h) - \sin x}{h}.$$

Using the addition formula to expand $\sin(x+h)$, we see that the Newton quotient is equal to

$$\frac{\sin x \cos h + \cos x \sin h - \sin x}{h}.$$

We put together the two terms involving $\sin x$:

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

and separate our quotient into a sum of two terms:

$$\cos x\,\frac{\sin h}{h} + \sin x\,\frac{\cos h - 1}{h}.$$

We now face the problem of finding the limit of

$$\frac{\sin h}{h} \quad \text{and} \quad \frac{\cos h - 1}{h} \quad \text{as } h \text{ approaches } 0.$$

This is a somewhat more difficult problem than those we encountered previously. We cannot tell right away what these limits will be. In the next section, we shall prove that

$$\lim_{h \to 0} \frac{\sin h}{h} = 1 \quad \text{and} \quad \lim_{h \to 0} \frac{\cos h - 1}{h} = 0.$$

Once we know these limits, then we see immediately that the first term approaches $\cos x$ and the second term approaches

$$(\sin x) \cdot 0 = 0.$$

Hence,

$$\lim_{h \to 0} \frac{\sin(x + h) - \sin x}{h} = \cos x.$$

This proves that

$$\frac{d(\sin x)}{dx} = \cos x.$$

To find the derivative of $\cos x$, we could proceed in the same way, and we would encounter the same limits. However, there is a trick which avoids this.

We know that $\cos x = \sin\left(x + \frac{\pi}{2}\right)$. Let $u = x + \frac{\pi}{2}$ and use the chain rule. We get

$$\frac{d(\cos x)}{dx} = \frac{d(\sin u)}{du} \frac{du}{dx}.$$

However, $du/dx = 1$. Hence

$$\frac{d(\cos x)}{dx} = \cos u = \cos\left(x + \frac{\pi}{2}\right) = -\sin x,$$

thereby proving our theorem.

Remark. It is not true that the derivative of the function $\sin^* x$ is $\cos^* x$. Using the chain rule, find out what its derivative is. The reason for using the radian measure of angles is to get a function $\sin x$ whose derivative is $\cos x$.

Example. Find the tangent line of the curve $y = \sin 4x$ at the point $x = \pi/16$. This is easily done. Let $f(x) = \sin 4x$. Then

$$f'(x) = 4 \cos 4x.$$

Hence the slope of the tangent line at $x = \pi/16$ is equal to

$$f'\left(\frac{\pi}{16}\right) = 4\cos\left(\frac{4\pi}{16}\right) = 4\cos\left(\frac{\pi}{4}\right) = \frac{4}{\sqrt{2}}.$$

On the other hand, we have

$$f\left(\frac{\pi}{16}\right) = \sin\left(\frac{4\pi}{16}\right) = \sin\frac{\pi}{4} = \frac{1}{\sqrt{2}}.$$

Hence the equation of the tangent line is

$$y - \frac{1}{\sqrt{2}} = \frac{4}{\sqrt{2}}\left(x - \frac{\pi}{16}\right).$$

Theorem 4.2. *We have*

$$\boxed{\frac{d(\tan x)}{dx} = 1 + \tan^2 x = \sec^2 x.}$$

Proof. We use the rule for the derivative of a quotient. So:

$$\begin{aligned}\frac{d}{dx}\left(\frac{\sin x}{\cos x}\right) &= \frac{(\cos x)(\cos x) - (\sin x)(-\sin x)}{(\cos x)^2}\\ &= \frac{\cos^2 x + \sin^2 x}{\cos^2 x}\\ &= \frac{1}{\cos^2 x} = \sec^2 x.\end{aligned}$$

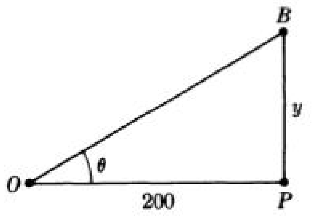
But also

$$\frac{\cos^2 x + \sin^2 x}{\cos^2 x} = 1 + \frac{\sin^2 x}{\cos^2 x} = 1 + \tan^2 x.$$

This proves the theorem.

Example. A balloon is going up, starting at a point P. An observer standing 200 ft away looks at the balloon, and the angle θ which the balloon makes increases at the rate of $\frac{1}{20}$ rad/sec. Find the rate at which the distance of the balloon from the ground is increasing when $\theta = \pi/4$.

The picture looks as follows, where y is the distance from the balloon to the ground.



We have $\tan \theta = y/200$, whence

$$y = 200 \cdot \tan \theta.$$

We want to find the rate at which y is increasing, i.e. we want to find dy/dt. Taking the derivative with respect to time t yields

$$\begin{aligned}\frac{dy}{dt} &= 200 \frac{d \tan \theta}{dt}.\\ &= 200(1 + \tan^2 \theta) \frac{d\theta}{dt}\end{aligned}$$

by Theorem 4.2 and the chain rule. Hence

$$\begin{aligned}\left.\frac{dy}{dt}\right|_{\theta = \pi/4} &= 200\left(1 + \tan^2 \frac{\pi}{4}\right)\frac{1}{20}\\ &= 10(1 + 1)\\ &= 20 \text{ ft/sec.}\end{aligned}$$

This is our answer.

Example. In the preceding example, find the rate at which the distance of the balloon from the ground is increasing when $\sin \theta = 0.2$.

We have

$$\frac{dy}{dt} = 200 \frac{d \tan \theta}{d\theta} \frac{d\theta}{dt} = 200(1 + \tan^2 \theta) \frac{d\theta}{dt}.$$

When $\sin \theta = 0.2$, we have $\sin^2 \theta = 0.04$, and

$$\cos^2 \theta = 1 - \sin^2 \theta = 0.96.$$

Hence

$$\tan^2\theta = \frac{\sin^2\theta}{\cos^2\theta} = \frac{4}{96} = \frac{1}{24}.$$

Since we are given $d\theta/dt = 1/20$, we find

$$\left.\frac{dy}{dt}\right|_{\sin\theta=0.2} = 200\left(1+\frac{1}{24}\right)\frac{1}{20}$$

$$= \frac{10\cdot 25}{24}$$

$$= \frac{125}{12}\text{ ft/sec.}$$

IV, §4. EXERCISES

1. What is the derivative of $\cot x$?

Find the derivative of the following functions:

2. $\sin(3x)$
3. $\cos(5x)$
4. $\sin(4x^2 + x)$
5. $\tan(x^3 - 5)$
6. $\tan(x^4 - x^3)$
7. $\tan(\sin x)$
8. $\sin(\tan x)$
9. $\cos(\tan x)$
10. What is the slope of the curve $y = \sin x$ at the point whose x-coordinate is π?

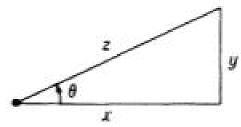
Find the slope of the following curves at the indicated point (we just give the x-coordinate of the point):

11. $y = \cos(3x)$ at $x = \pi/3$
12. $y = \sin x$ at $x = \pi/6$
13. $y = \sin x + \cos x$ at $x = 3\pi/4$
14. $y = \tan x$ at $x = -\pi/4$
15. $y = \dfrac{1}{\sin x}$ at $x = -\pi/6$
16. Give the equation of the tangent line to the following curves at the indicated point.
    (a) $y = \sin x$ at $x = \pi/2$
    (b) $y = \cos x$ at $x = \pi/6$
    (c) $y = \sin 2x$ at $x = \pi/4$
    (d) $y = \tan 3x$ at $x = \pi/4$
    (e) $y = 1/\sin x$ at $x = \pi/2$
    (f) $y = 1/\cos x$ at $x = \pi/4$
    (g) $y = 1/\tan x$ at $x = \pi/4$
    (h) $y = \tan\dfrac{x}{2}$ at $x = \pi/2$

(i) $y = \sin \dfrac{x}{2}$ at $x = \pi/3$ (j) $y = \cos \dfrac{\pi x}{3}$ at $x = 1$

(k) $y = \sin \pi x$ at $x = \frac{1}{2}$ (l) $y = \tan \pi x$ at $x = \frac{1}{6}$

17. In the following right triangle, suppose that θ is decreasing at the rate of $\frac{1}{30}$ rad/sec. Find each one of the indicated derivatives:
 (a) dy/dt, when $\theta = \pi/3$ and x is constant, $x = 12$.
 (b) dz/dt, when $\theta = \pi/4$ if y is constant, $y = 10\sqrt{2}$.
 (c) dx/dt, when $x = 1$ if x and y are both changing, but z is constant, $z = 2$.
 Remember that θ is decreasing, so $d\theta/dt = -\frac{1}{30}$.



18. A Ferris wheel 50 ft in diameter makes 1 revolution every 2 min. If the center of the wheel is 30 ft above the ground, how fast is a passenger in the wheel moving vertically when he is 42.5 ft above the ground?

19. A balloon is going up, starting at a point P. An observer O, standing 300 ft away, looks at the balloon, and the angle θ which the balloon makes increases at the rate of 0.3 rad/sec. Find the rate at which the distance of the balloon from the ground is increasing, when
 (a) $\theta = \pi/4$, (b) $\theta = \pi/3$, (c) $\cos \theta = 0.2$,
 (d) $\sin \theta = 0.3$, (e) $\tan \theta = 4$.

20. An airplane is flying horizontally on a straight line at a speed of 1,000 km/hr, at an elevation of 10 km. An automatic camera is photographing a point directly ahead on the ground. How fast must the camera be turning when the angle between the path of the plane and the line of sight to the point is 30°?

21. A Beacon light is located 1000 ft from a sea wall, and rotates at the constant rate of 2 revolutions per minute.
 (a) How fast is the lighted spot on the wall moving along the wall at the nearest point to the beacon?
 (b) How fast is the spot of light moving at a point 500 ft from this nearest point?

22. An airplane flying at an altitude of 20,000 ft and on a horizontal course passes directly over an observer on the ground below. The observer notes that when the angle between the ground and his line of sight is 60°, the angle is decreasing at a rate of 2° per second. What is the speed of the airplane?

23. A vertical pole is 30 ft high and is located 30 ft east of a tall building. If the sun is rising at the rate of 18° per hour, how fast is the shadow of the pole on the building shortening when the elevation of the sun is 30°? [*Hint*: The rate of rise is the rate of change of the angle of elevation θ of the sun. First convert the degrees to radians per hour, namely

$$18 \text{ deg/hr} = 18\,\pi/180 = \pi/10 \text{ rad/hr}.$$

If s is the length of the shadow, you then have $\tan \theta = (30 - s)/30$.]

24. A weather balloon is released on the ground 1500 ft from an observer and rises vertically at the constant rate 250 ft/min. How fast is the angle between

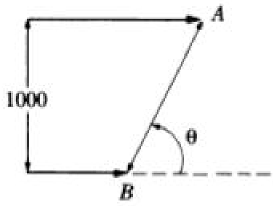
the observer's line of sight and the ground increasing when the balloon is at an altitude of 2000 ft? Give the answer in degrees per minute.

25. A ladder 30 ft long leans against a wall. Suppose the bottom of the ladder slides away from the wall at the rate of 3 ft/sec. How fast is the angle between the ladder and the ground changing when the bottom of the ladder is 15 ft from the wall?

26. A rocket leaves the ground 2000 ft from an observer and rises vertically at the constant rate of 100 ft/sec. How fast is the angle between the observer's line of sight and the ground increasing after 20 sec? Give the answer in degrees per second.

27. A kite at an altitude of 200 ft moves horizontally at the rate of 20 ft/sec. At what rate is the angle between the line and the ground changing when 400 ft of line is out?

28. Two airplanes are flying in the same direction, and at constant altitude. At $t = 0$ airplane A is 1000 ft vertically above airplane B. Airplane A travels at a constant speed of 600 ft/sec, and B at a constant speed of 400 ft/sec. Find the rate of change of the angle of elevation θ of A relative to B at time $t = 10$ sec.



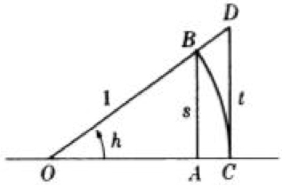
IV, §5. TWO BASIC LIMITS

We shall first prove that

$$\boxed{\lim_{h \to 0} \frac{\sin h}{h} = 1.}$$

Both the numerator and the denominator approach 0 as h approaches 0, and we get no information by trying some cancellation procedure, the way we did it for powers.

Let us assume first that h is positive, and look at the following diagram.



We take a circle of radius 1 and an angle of h radians. Let s be the altitude of the small triangle OAB, and t that of the big triangle OCD. Then, using the small triangle OAB,

$$\sin h = \frac{s}{1} = s$$

and using the large triangle, OCD,

$$\tan h = \frac{t}{1} = t = \frac{\sin h}{\cos h}.$$

We see that:

area of triangle $OAB <$ area of sector $OCB <$ area of triangle OCD.

The base OA of the small triangle is equal to $\cos h$ and its altitude AB is $\sin h$.

The base OC of the big triangle is equal to 1. Its altitude CD is

$$t = \frac{\sin h}{\cos h}.$$

The area of each triangle is $\frac{1}{2}$ the base times the altitude.

The area of the sector is the fraction $h/2\pi$ of the area of the circle, which is π. Hence the area of the sector is $h/2$. Thus we obtain:

$$\frac{1}{2} \cos h \sin h < \frac{1}{2}\, h < \frac{1}{2} \frac{\sin h}{\cos h}.$$

We multiply everywhere by 2 and get

$$\cos h \sin h < h < \frac{\sin h}{\cos h}.$$

Since we assumed that $h > 0$ it follows that $\sin h > 0$ and we divide both the inequalities by $\sin h$ to obtain

$$\cos h < \frac{h}{\sin h} < \frac{1}{\cos h}.$$

As h approaches 0, both $\cos h$ and $1/\cos h$ approach 1. Thus $h/\sin h$ is squeezed between two quantities which approach 1, and therefore $h/\sin h$ must approach 1 also. Thus we may write

$$\lim_{h \to 0} \frac{h}{\sin h} = 1.$$

Since

$$\frac{\sin h}{h} = \frac{1}{h/\sin h}$$

and the limit of a quotient is the quotient of the limits, it follows that

$$\lim_{h \to 0} \frac{\sin h}{h} = 1,$$

as was to be shown.

We computed our limit when $h > 0$. Suppose that $h < 0$. We can write

$$h = -k$$

with $k > 0$. Then

$$\frac{\sin(-k)}{-k} = \frac{-\sin k}{-k} = \frac{\sin k}{k}.$$

As h tends to 0, so does k. Hence we are reduced to our previous limit because $k > 0$.

We still have to prove the limit

$$\boxed{\lim_{h \to 0} \frac{\cos h - 1}{h} = 0.}$$

We have

$$\begin{aligned}
\frac{\cos h - 1}{h} &= \frac{(\cos h - 1)(\cos h + 1)}{h(\cos h + 1)} \\
&= \frac{\cos^2 h - 1}{h(\cos h + 1)} \\
&= \frac{-\sin^2 h}{h(\cos h + 1)} \\
&= -\frac{\sin h}{h}(\sin h)\frac{1}{\cos h + 1}.
\end{aligned}$$

We shall use the property concerning the product of limits. We have a product of three factors. The first is

$$-\frac{\sin h}{h}$$

and approaches -1 as h approaches 0.

The second is $\sin h$ and approaches 0 as h approaches 0.

The third is

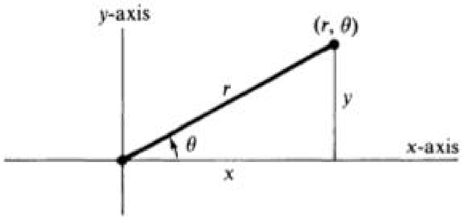
$$\frac{1}{\cos h + 1}$$

and its limit is $\frac{1}{2}$ as h approaches 0.

Therefore the limit of the product is 0, and everything is proved!

IV, §6. POLAR COORDINATES

Instead of describing a point in the plane by its coordinates with respect to two perpendicular axes, we can also describe it as follows. We draw a ray between the point and a given origin. The **angle** θ which this ray makes with the horizontal axis and the **distance** r between the point and the origin determine our point. Thus the point is described by a pair of numbers (r, θ), which are called its **polar coordinates**.



If we have our usual axes and x, y are the ordinary coordinates of our point, then we see that

$$\boxed{\frac{x}{r} = \cos\theta \qquad \text{and} \qquad \frac{y}{r} = \sin\theta,}$$

whence

$$\boxed{x = r\cos\theta \qquad \text{and} \qquad y = r\sin\theta.}$$

This allows us to change from polar coordinates to ordinary coordinates.

It is to be understood that r is always supposed to be $\geqq 0$. In terms of the ordinary coordinates, we have

$$\boxed{r = \sqrt{x^2 + y^2}.}$$

By Pythagoras, r is the distance of the point (x, y) from the origin $(0, 0)$. Note that distance is always $\geqq 0$.

Example 1. Find polar coordinates of the point whose ordinary coordinates are $(1, \sqrt{3})$.

We have $x = 1$ and $y = \sqrt{3}$, so that $r = \sqrt{1 + 3} = 2$. Also

$$\cos \theta = \frac{x}{r} = \frac{1}{2}, \qquad \sin \theta = \frac{y}{r} = \frac{\sqrt{3}}{2}.$$

Hence $\theta = \pi/3$, and the polar coordinates are $(2, \pi/3)$.

We observe that we may have several polar coordinates corresponding to the same point. The point whose polar coordinates are $(r, \theta + 2\pi)$ is the same as the point (r, θ). Thus in our example above, $(2, \pi/3 + 2\pi)$ would also be polar coordinates for our point. In practice, we usually use the value for the angle which lies between 0 and 2π.

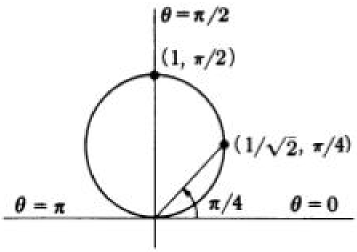
Suppose a bug is traveling in the plane. Its position is completely determined if we know the angle θ and the distance of the bug from the origin, that is if we know the polar coordinates. If the distance r from the origin is given as a function of θ, then the bug is traveling along a curve and we can sketch this curve.

Example 2. Sketch the graph of the function $r = \sin \theta$ for $0 \leqq \theta \leqq \pi$.

If $\pi < \theta < 2\pi$, then $\sin \theta < 0$ and hence for such θ we don't get a point on the curve. Next, we make a table of values. We consider intervals of θ such that $\sin \theta$ is always increasing or always decreasing over these intervals. This tells us whether the point is moving further away from the origin, or coming closer to the origin, since r is the distance of the point from the origin. Intervals of increase and decrease for $\sin \theta$ can be taken to be of length $\pi/2$. Thus we find the following table:

θ	$\sin \theta = r$
inc. 0 to $\pi/2$	inc. 0 to 1
inc. $\pi/2$ to π	dec. 1 to 0
$\pi/6$	$1/2$
$\pi/4$	$1\sqrt{2}$
$\pi/3$	$\sqrt{3}/2$

Put in words: as θ increases from 0 to $\pi/2$, then $\sin \theta$ and therefore r increases until r reaches 1. As θ increases from $\pi/2$ to π then $\sin \theta$ and thus r decreases from 1 to 0. Hence the graph looks like this.



We have drawn the graph like a circle. Actually, we don't know whether it is a circle or not. The graph could be flatter in one direction than in another. In the next example, we shall see that it actually must be a circle.

Example 3. Change the equation

$$\boxed{r = \sin\theta}$$

to rectangular coordinates.

We substitute the expressions

$$r = \sqrt{x^2 + y^2}$$

and

$$\sin\theta = y/r = y/\sqrt{x^2 + y^2}$$

in the polar equation, to obtain

$$\sqrt{x^2 + y^2} = \frac{y}{\sqrt{x^2 + y^2}}.$$

Of course, this substitution is valid only when $r \neq 0$, i.e. $r > 0$. We can then simplify the equation we have just obtained, multiplying both sides by $\sqrt{x^2 + y^2}$. We then obtain

$$x^2 + y^2 = y.$$

You should know from Chapter II that this is the equation of a circle, by completing the square. We recall here how this is done. We write the equation in the form

$$x^2 + y^2 - y = 0.$$

We would like this equation to be of the form

$$x^2 + (y - b)^2 = c^2,$$

because then we know immediately that this is a circle of center $(0, b)$ and radius c. We know that

$$(y - b)^2 = y^2 - 2by + b^2.$$

Therefore we let $2b = 1$ and $b = \frac{1}{2}$. Then

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

because the $\frac{1}{4}$ cancels. Thus the equation

$$\boxed{x^2 + y^2 - y = 0}$$

is equivalent with

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

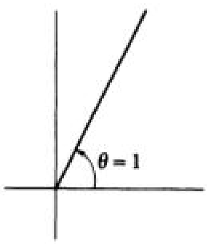
This is the equation of a circle of center $(0, \frac{1}{2})$ and radius $\frac{1}{2}$. The point corresponding to the polar coordinate $r = 0$ is the point with rectangular coordinates $x = 0$ and $y = 0$.

Example 4. The equation of the circle of radius 3 and center at the origin in polar coordinates is simply

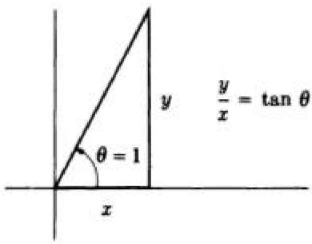
$$r = 3 \qquad \text{or} \qquad \sqrt{x^2 + y^2} = 3 \qquad \text{or} \qquad x^2 + y^2 = 9.$$

This expresses the condition that that distance of the point (x, y) from the origin is the constant 3. The angle θ can be arbitrary.

Example 5. Consider the equation $\theta = 1$ in polar coordinates. A point with polar coordinates (r, θ) satisfies this equation if and only if its angle θ is 1 and there is no restriction on its r-coordinate, i.e. $r \geqq 0$. Thus geometrically, this set of points can be described as a half line, or a ray, as on the figure (a).



(a)



(b)

By the definition of the tangent, if (x, y) are the ordinary coordinates of a point on this ray, and $y \neq 0$, then

$$y/x = \tan 1 \qquad \text{and} \qquad x > 0,$$

whence

$$y = (\tan 1)x \qquad \text{and} \qquad x > 0.$$

Of course, the point with $x = y = 0$ also lies on the ray. Conversely, any point whose ordinary coordinates (x, y) satisfy

$$y = (\tan 1)x \qquad \text{and} \qquad x \geqq 0$$

lies on the ray. Hence the ray defined in polar coordinates by the equation $\theta = 1$ is defined in ordinary coordinates by the pair of conditions

$$y = (\tan 1)x \qquad \text{and} \qquad x \geqq 0.$$

Instead of 1 we could take any number. For instance, the ray defined by the equation $\theta = \pi/6$ in polar coordinates is also defined by the pair of conditions

$$y = (\tan \pi/6)x \qquad \text{and} \qquad x \geqq 0.$$

Since $\tan \pi/6 = 1/\sqrt{3}$, we may write the equivalent pair of conditions

$$y = \frac{1}{\sqrt{3}}\,x \qquad \text{and} \qquad x \geqq 0.$$

Note that there is no simpler way of expressing $\tan 1$ than just writing $\tan 1$. Only when dealing with fractional multiples of π do we have a way of writing the trigonometric functions in terms of roots, like $\tan \pi/6 = 1/\sqrt{3}$.

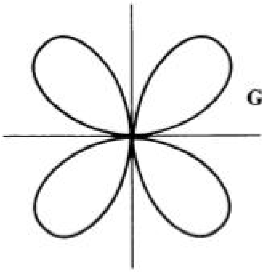
Example 6. Let us sketch the curve given in polar coordinates by the equation

$$r = |\sin 2\theta|.$$

The absolute value sign makes the right-hand side always $\geqq 0$, and so there is a value of r for every value of θ. Regions of increase and decrease for $\sin 2\theta$ will occur when 2θ ranges over intervals of length $\pi/2$. Hence it is natural to look at intervals for θ of length $\pi/4$. We now make a table of the increasing and decreasing behavior of $|\sin 2\theta|$ and r over such intervals.

θ	$r = \lvert\sin 2\theta\rvert$
inc. 0 to $\pi/4$	inc. 0 to 1
inc. $\pi/4$ to $\pi/2$	dec. 1 to 0
inc. $\pi/2$ to $3\pi/4$	inc. 0 to 1
inc. $3\pi/4$ to π	dec. 1 to 0
and so forth	

The graph therefore looks like this:



Graph of $r = |\sin 2\theta|$

Because of the absolute value sign, for any value of θ we obtain a value for r which is $\geqq 0$. According to our convention, if we wanted to graph

$$r = \sin 2\theta$$

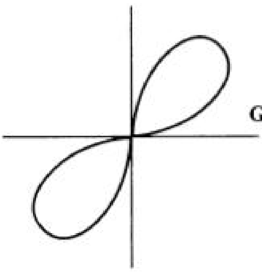
without the absolute value sign, then we would have to omit those portions of the above graph for which $\sin 2\theta$ is negative, i.e. those portions of the graph for which

$$\frac{\pi}{2} < \theta < \pi$$

and

$$\frac{3\pi}{2} < \theta < 2\pi.$$

Thus the graph of $r = \sin 2\theta$ would look like that in the next figure.



Graph of $r = \sin 2\theta$

Example 7. We want to sketch the curve given in polar coordinates by the equation

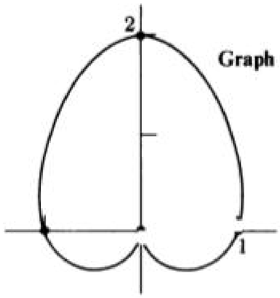
$$r = 1 + \sin\theta$$

We look at the behavior of r when θ ranges over the intervals.

$$[0, \pi/2], \qquad [\pi/2, \pi], \qquad [\pi, 3\pi/2], \qquad [3\pi/2, 2\pi].$$

θ	$\sin\ \theta$	r
inc. from 0 to $\pi/2$	inc. 0 to 1	inc. 1 to 2
inc. from $\pi/2$ to π	dec. 1 to 0	dec. 2 to 1
inc. from π to $3\pi/2$	dec. 0 to -1	dec. 1 to 0.
inc. from $3\pi/2$ to 2π	inc. -1 to 0	inc. 0 to 1

Thus the graph looks roughly like this:



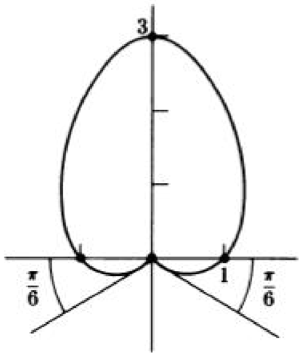
Example 8. Let us look at the slightly different equation

$$r = 1 + 2\sin\theta.$$

A similar analysis will work, but we must be careful of the possibility that the expression on the right-hand side is negative. According to our convention, this will not yield a point since we assume that $r \geqq 0$ for polar coordinates. Thus when $2\sin\theta < -1$, we do not get any point. This occurs precisely in the interval

$$\frac{7\pi}{6} < \theta < \frac{11\pi}{6}.$$

The graph will therefore look like the next figure.



We have also drawn the rays determining angles of $\frac{7\pi}{6}$ and $\frac{11\pi}{6}$.

IV, §6. EXERCISES

1. Plot the following points in polar coordinates:
   (a) $(2, \pi/4)$ (b) $(3, \pi/6)$ (c) $(1, -\pi/4)$ (d) $(2, -3\pi/6)$

2. Same directions as in Exercise 1.
   (a) $(1, 1)$ (b) $(4, -3)$
   (These are polar coordinates. Just show approximately the angle represented by the given coordinates.)

3. Find polar coordinates for the following points given in the usual x- and y-coordinates:
   (a) $(1, 1)$ (b) $(-1, -1)$ (c) $(3, 3\sqrt{3})$ (d) $(-1, 0)$

4. Sketch the following curves and put the equation in rectangular coordinates.
   (a) $r = 2 \sin \theta$ (b) $r = 3 \cos \theta$

5. Change the following to rectangular coordinates and sketch the curve. We assume $a > 0$.
   (a) $r = a \sin \theta$ (b) $r = a \cos \theta$
   (c) $r = 2a \sin \theta$ (d) $r = 2a \cos \theta$

Sketch the graphs of the following curves given in polar coordinates.

6. $r^2 = \cos \theta$
7. $r^2 = \sin \theta$
8. (a) $r = \sin^2 \theta$ (b) $r = \cos^2 \theta$
9. $r = 4 \sin^2 \theta$
10. $r = 5$
11. $r = 4$
12. (a) $r = \dfrac{1}{\cos \theta}$ (b) $r = \dfrac{1}{\sin \theta}$
13. $r = 3/\cos \theta$
14. $r = 1 + \cos \theta$
15. $r = 1 - \sin \theta$
16. $r = 1 - \cos \theta$
17. $r = 1 - 2 \sin \theta$
18. $r = \sin 3\theta$
19. $r = \sin 4\theta$

20. $r = \cos 2\theta$ 21. $r = \cos 3\theta$ 22. $r = |\cos 2\theta|$

23. $r = |\sin 3\theta|$ 24. $r = |\cos 3\theta|$ 25. $r = \theta$

26. $r = 1/\theta$

In the next three problems put the equation in rectangular coordinates and sketch the curve.

27. $r = \dfrac{1}{1 - \cos \theta}$ 28. $r = \dfrac{2}{2 - \cos \theta}$ 29. $r = \dfrac{4}{1 + 2\cos \theta}$

Sketch the following curves given in polar coordinates.

30. $r = \tan \theta$

31. $r = 5 + 2\sin \theta$

32. $r = |1 + 2\cos \theta|$

33. (a) $r = 2 + \sin 2\theta$
(b) $r = 2 - \sin 2\theta$

34. $\theta = \pi$

35. $\theta = \pi/2$

36. $\theta = -\pi/2$

37. $\theta = 5\pi/4$

38. $\theta = 3\pi/2$

39. $\theta = 3\pi/4$

CHAPTER V

The Mean Value Theorem

Given a curve, $y = f(x)$, we shall use the derivative to give us information about the curve. For instance, we shall find the maximum and minimum of the graph, and regions where the curve is increasing or decreasing. We shall use the mean value theorem, which is basic in the theory of derivatives.

V, §1. THE MAXIMUM AND MINIMUM THEOREM

Definition. Let f be a differentiable function. **A critical point** of f is a number c such that

$$f'(c) = 0.$$

The derivative being zero means that the slope of the tangent line is 0 and thus that the tangent line itself is horizontal. We have drawn three examples of this phenomenon.

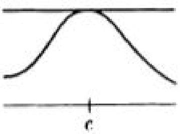


Figure 1

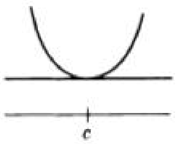


Figure 2

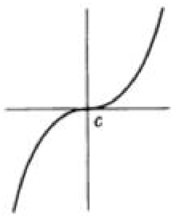


Figure 3

The third example is that of a function like $f(x) = x^3$. We have $f'(x) = 3x^2$ and hence when $x = 0$, $f'(0) = 0$.

The other two examples are those of a maximum and a minimum, respectively, if we look at the graph of the function only near our point c. We shall now formalize these notions.

Let a, b be two numbers with $a < b$. We shall repeatedly deal with the interval of numbers between a and b. Sometimes we want to include the end points a and b, and sometimes we do not. We recall the standard terminology.

The collection of numbers x such that $a < x < b$ is called the **open interval** between a and b.

The collection of numbers x such that $a \leqq x \leqq b$ is called the **closed interval** between a and b. We denote this closed interval by the symbols $[a, b]$. (A single point will also be called a closed interval.)

If we wish to include only one end point, we shall say that the interval is **half-closed**. We have of course two half-closed intervals, namely the one consisting of the numbers x with $a \leqq x < b$, and the other one consisting of the numbers x with $a < x \leqq b$.

Sometimes, if a is a number, we call the collection of numbers $x > a$ (or $x < a$) an open interval. The context will always make this clear.

Let f be a function, and c a number at which f is defined.

Definition. We shall say that c is a **maximum point** of the function f if and only if

$$f(c) \geqq f(x)$$

for all numbers x at which f is defined. If the condition $f(c) \geqq f(x)$ holds for all numbers x in some interval, then we say that **the function has a maximum at c in that interval**. We call $f(c)$ a **maximum value**.

Example 1. Let $f(x) = \sin x$. Then f has a maximum at $\pi/2$ because $f(\pi/2) = 1$ and $\sin x \leqq 1$ for all values of x. This is illustrated in Fig. 4. Note that $-3\pi/2$ is also a maximum for $\sin x$.

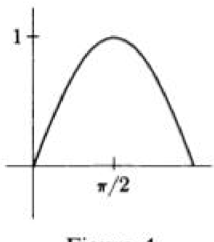


Figure 4

Example 2. Let $f(x) = 2x$, and view f as a function defined only on the interval

$$0 \leqq x \leqq 2.$$

Then the function has a maximum at 2 in this interval because $f(2) = 4$ and $f(x) \leqq 4$ for all x in the interval. This is illustrated in Fig. 5.

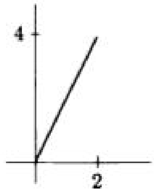


Figure 5

Example 3. Let $f(x) = 1/x$. We know that f is not defined for $x = 0$. This function has no maximum. It becomes arbitrarily large when x comes close to 0 and $x > 0$. This is illustrated in Fig. 6.

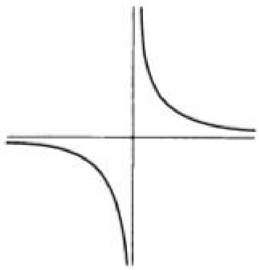


Figure 6

Definition. A **minimum point** for f is a number c such that

$$f(x) \geqq f(c) \qquad \text{for all } x \text{ where } f \text{ is defined.}$$

A **minimum value** for the function is the value $f(c)$, taken at a minimum point.

We illustrate various minima with the graphs of certain functions.

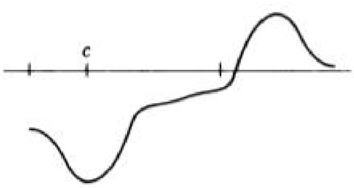


Figure 7

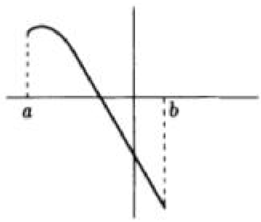


Figure 8

In Fig. 7 the function has a minimum. In Fig. 8 the minimum is at the end point of the interval. In Figs. 3 and 6 the function has no minimum.

In the following picture, the point c_1 looks like a maximum and the point c_2 looks like a minimum, provided we stay close to these points, and don't look at what happens to the curve farther away.

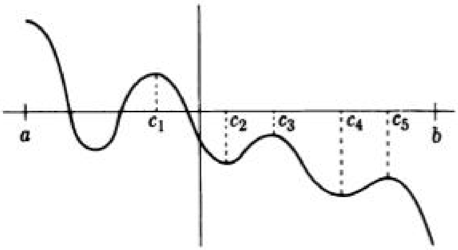


Figure 9

There is a name for such points. We shall say that a point c is a **local minimum** or **relative minimum** of the function f if there exists an interval

$$a_1 < c < b_1$$

such that $f(c) \leqq f(x)$ for all numbers x with $a_1 \leqq x \leqq b_1$.

Similarly, we define the notion of **local maximum** or **relative maximum**. (Do it yourself.) In Fig. 9, the point c_3 is a local maximum, c_4 is a local minimum, and c_5 is a local maximum.

The actual maximum and minimum occur at the end points.

Using basic properties of numbers, one can prove the next theorem which is, however, rather obvious, and so we omit the proof.

Theorem 1.1. *Let f be a continuous function over a closed interval $[a, b]$. Then there exists a point in the interval where f has a maximum, and there exists a point where f has a minimum.*

We wish to have some ideas of the range of values of the given function. The next theorem tells us this information.

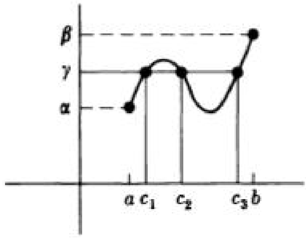
Theorem 1.2. Intermediate value theorem. *Let f be a continuous function on the interval $[a, b]$. Let $\alpha = f(a)$ and $\beta = f(b)$. Let γ be a number between α and β. For instance, if $\alpha < \beta$, let $\alpha < \gamma < \beta$, and if $\alpha > \beta$ then let*

$$\alpha > \gamma > \beta.$$

Then there exists a number c such that $a < c < b$ and such that

$$f(c) = \gamma.$$

The theorem is intuitively obvious since a continuous function has no breaks. It is illustrated on the figure.



The proof belongs to the range of ideas in the appendix on epsilon-delta and can safely be omitted. Observe that there may be several points c in the interval $[a, b]$ such that $f(c) = \gamma$. In the figure, there are three such points, labeled c_1, c_2, c_3.

As we have mentioned, the point where f has a maximum may occur at the end points of the interval. However, when such a point is not an end point, and the function is differentiable, then we are in a situation similar to that of Fig. 4 or 9, where we see that the tangent to the curve at that point is a horizontal line; in other words the derivative of the function is 0. We can prove this as a theorem.

Theorem 1.3. *Let f be a function which is defined and differentiable in the open interval $a < x < b$. Let c be a number in the interval at which the function has a local maximum or a local minimum. Then*

$$f'(c) = 0.$$

Proof. We give the proof in the case of a local maximum. If we take small values of h (positive or negative), the number $c + h$ will lie in the interval. We are going to find the limit of the Newton quotient as we approach c from the right and from the left, and in that way, determine the value $f'(c)$.

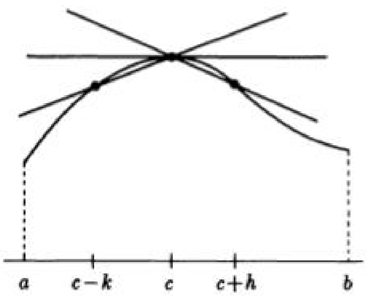


Figure 10

Let us first take h positive (see Fig. 10). We must have

$$f(c) \geqq f(c + h)$$

no matter what h is (provided h is small) since $f(c)$ is the local maximum. Therefore $f(c+h) - f(c) \leqq 0$. Since $h > 0$, the Newton quotient satisfies

$$\frac{f(c+h) - f(c)}{h} \leqq 0.$$

Hence the limit is $\leqq 0$, or in symbols:

$$\lim_{\substack{h \to 0 \\ h > 0}} \frac{f(c+h) - f(c)}{h} \leqq 0.$$

Now take h negative, say $h = -k$ with $k > 0$. Then

$$f(c-k) - f(c) \leqq 0, \qquad f(c) - f(c-k) \geqq 0$$

and the quotient is

$$\frac{f(c-k) - f(c)}{-k} = \frac{f(c) - f(c-k)}{k}.$$

Thus the Newton quotient is $\geqq 0$. Taking the limit as h (or k) approaches 0, we see that

$$\lim_{\substack{h \to 0 \\ h < 0}} \frac{f(c+h) - f(c)}{h} \geqq 0.$$

The only way in which our two limits can be equal is that they should both be 0. Therefore $f'(c) = 0$. This concludes the proof.

We can interpret our arguments geometrically by saying that the line between our two points slants up to the left when we take $h > 0$ and slants up to the right when we take $h < 0$. As h approaches 0, both lines must approach the tangent line to the curve. The only way this is possible is for the tangent line at the point whose x-coordinate is c to be horizontal. This means that its slope is 0, i.e. $f'(c) = 0$.

In practice, a function usually has only a finite number of critical points, and it is easy to find all points c such that $f'(c) = 0$. One can then determine by inspection which of these are maxima, which are minima, and which are neither.

Example 4. Find the critical points of the function $f(x) = x^3 - 1$.

We have $f'(x) = 3x^2$. Hence there is only one critical point, namely $x = 0$, since $3x^2 = 0$ only when $x = 0$.

Example 5. Find the critical points of the function

$$y = x^3 - 2x + 1.$$

The derivative is $3x^2 - 2$. It is equal to 0 precisely when

$$x^2 = \tfrac{2}{3},$$

which means $x = \sqrt{2/3}$ or $-\sqrt{2/3}$. These are the critical points.

We shall find various ways in the next sections to see if the critical point is a local maximum or minimum. In simple cases, sketching the curve, you can often see it by inspection.

V, §1. EXERCISES

Find the critical points of the following functions.

1. $x^2 - 2x + 5$
2. $2x^2 - 3x - 1$
3. $3x^2 - x + 1$
4. $-x^2 + 2x + 2$
5. $-2x^2 + 3x - 1$
6. $x^3 + 2$
7. $x^3 - 3x$
8. $\sin x + \cos x$
9. $\cos x$
10. $\sin x$

V, §2. INCREASING AND DECREASING FUNCTIONS

Let f be a function defined on some interval (which may be open or closed).

Definition. We shall say that f is **increasing** over this interval if

$$f(x_1) \leqq f(x_2)$$

whenever x_1 and x_2 are two points of the interval such that

$$x_1 \leqq x_2.$$

Thus, if a number lies to the right of another, the value of the function at the larger number must be greater than or equal to the value of the function at the smaller number.

In the next figure, we have drawn the graph of an increasing function.

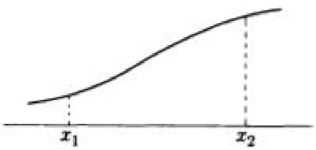


Figure 11

We say that a function defined on some interval is **decreasing** over this interval if

$$f(x_1) \geqq f(x_2)$$

whenever x_1 and x_2 are two points of the interval such that $x_1 \leqq x_2$.

Observe that a constant function (whose graph is horizontal) is both increasing and decreasing.

If we want to omit the equality sign in our definitions, we shall use the word **strictly** to qualify decreasing or increasing. Thus a function f is **strictly increasing** if

$$x_1 < x_2 \quad \text{implies} \quad f(x_1) < f(x_2)$$

and f is **strictly decreasing** if

$$x_1 < x_2 \quad \text{implies} \quad f(x_1) > f(x_2)$$

Suppose that a function has a positive derivative throughout an interval, as shown for instance on Fig. 11. Then we can interpret this as meaning that the rate of change of the function is always positive, and therefore that the function is increasing. We state this as a theorem.

Theorem 2.1. *Let f be a function which is continuous in some interval, and differentiable in the interval (excluding the end points).*

If $f'(x) > 0$ in the interval (excluding the end points), then f is strictly increasing.

If $f'(x) < 0$ in the interval (excluding the end points), then f is strictly decreasing.

If $f'(x) = 0$ in the interval (excluding the end points), then f is constant.

In this last statement, the hypothesis that $f'(x) = 0$ in the interval means that the rate of change is 0, and so it is quite plausible that the function is constant. To see how these statements fit into a more formal context, see §3.

Application. Graphs of parabolas

Example. Let us graph the curve

$$y = f(x) = x^2 - 3x + 5,$$

which you should know is a parabola as in Chapter II. We treat the graph here by the method which works in more general cases. First, we have

$$f'(x) = 2x - 3,$$

and

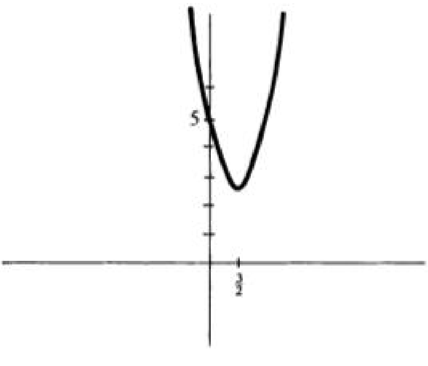
$f'(x) = 0$	if and only if	$x = 3/2,$	so $x = 3/2$ is the only critical point.
$f'(x) > 0$	if and only if	$2x - 3 > 0$	
	if and only if	$x > 3/2.$	
$f'(x) < 0$	if and only if	$x < 3/2.$	

Thus the function is strictly increasing for $x > 3/2$ and strictly decreasing for $x < 3/2$. Thus by using the derivative, we are able to find the peak of the parabola.

The points where $f(x) = 0$, that is where the graph crosses the x-axis, are given by the quadratic formula:

$$x = \frac{3 \pm \sqrt{9 - 20}}{2} = \frac{3 \pm \sqrt{-11}}{2}.$$

There are no such points. The graph therefore looks like this.



Observe that even if we did not know before the general shape of a parabola, we could deduce it now, and we would know that $x = 3/2$ is a minimum point of the graph. This is because $f(x)$ is strictly decreasing for $x < 3/2$ and strictly increasing for $x > 3/2$. Thus $x = 3/2$ must be a minimum.

Example. Sketch the graph of

$$y = f(x) = x^2 - 5x + 9/4.$$

This time, we have

$$f'(x) = 2x - 5.$$

Hence there is exactly one critical point, namely:

$$\begin{aligned} f'(x) = 0 \quad & \text{if and only if} \quad x = 5/2. \\ f'(x) > 0 \quad & \text{if and only if} \quad 2x - 5 > 0 \\ & \text{if and only if} \quad x > 5/2. \\ f'(x) < 0 \quad & \text{if and only if} \quad 2x - 5 < 0 \\ & \text{if and only if} \quad x < 5/2. \end{aligned}$$

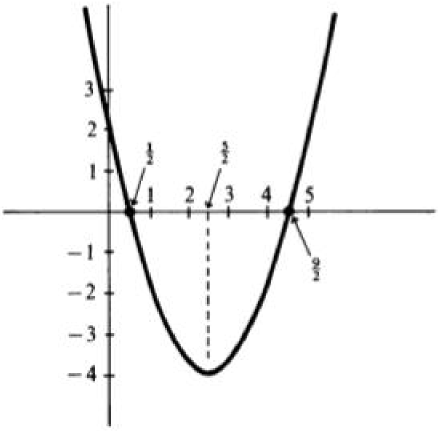
So f is strictly increasing for $x > 5/2$ and strictly decreasing for $x < 5/2$. Hence f has a minimum at $x = 5/2$.

The x-intercepts of the graph of f are

$$x = \frac{5 \pm \sqrt{25 - 9}}{2} = \frac{9}{2} \quad \text{and} \quad \frac{1}{2}.$$

Definition. The x-intercepts of the graph of f are also called the **roots** of f. In the case of a quadratic polynomial the roots are computed by the quadratic formula.

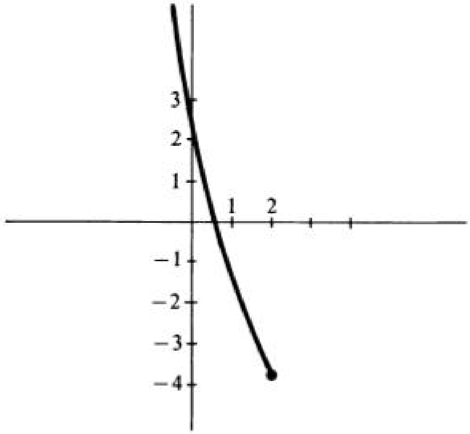
Therefore the graph of f looks like this.



In the above examples, the function was defined for all numbers. In the next example, we look at a function defined only over an interval.

Example. Let $f(x) = x^2 - 5x + 9/4$. Find the minimum and maximum of f for $x \leqq 2$.

By the preceding example, we know that f is strictly decreasing for $x \leqq 2$. Therefore the minimum is at the end point of the interval, that is $x = 2$, as shown on the figure. Note that at this end point, $f'(2) \neq 0$. Thus the test with critical points is valid only on open intervals.



The expression "if and only if" will recur quite frequently. We shall therefore use an abbreviating symbol for it, and we write

$\Leftrightarrow$ to mean, "if and only if."

Thus we could write the assertion:

$$x^2 = 3 \quad \Leftrightarrow \quad x = \sqrt{3} \quad \text{or} \quad x = -\sqrt{3}.$$

Similarly,

$$x^2 - 3x + 1 = 0 \quad \Leftrightarrow \quad x = \frac{3 \pm \sqrt{5}}{2}.$$

Example. Show that among all rectangles of given area, the one with least perimeter is a square.

Let a be the given area, and let x be the length of one side of the possible rectangle with area a. We shall express the perimeter as a function $f(x)$. We then differentiate with respect to x, keeping in mind that a is constant, and this will give us a value for x which will show that the rectangle is a square. To carry this out, we take $0 < x$ because the side of an actual rectangle cannot be 0 or have negative length. If y is the length of the other side, then $xy = a$, so that $y = a/x$ is the length of the other side. Hence the perimeter is

$$f(x) = 2\left(x + \frac{a}{x}\right)$$

We have

$$f'(x) = 2\left(1 - \frac{a}{x^2}\right),$$

and:

$$\begin{aligned} f'(x) = 0 \quad \Leftrightarrow \quad & \frac{x^2 - a}{x^2} = 0 \\ & \Leftrightarrow x^2 - a = 0 \\ & \Leftrightarrow x^2 = a \\ & \Leftrightarrow x = \sqrt{a}, \end{aligned}$$

because we consider only $x > 0$. Thus the only critical point of f for $x > 0$ is when $x = \sqrt{a}$. Furthermore $x^2 > 0$ for all $x \neq 0$. Hence the fraction $(x^2 - a)/x^2$ is positive if and only if its numerator $x^2 - a$ is positive. Thus:

$$f'(x) > 0 \quad \Leftrightarrow \quad x^2 - a > 0 \quad \Leftrightarrow \quad x^2 > a \quad \Leftrightarrow \quad x > \sqrt{a},$$

$$f'(x) < 0 \quad \Leftrightarrow \quad x^2 - a < 0 \quad \Leftrightarrow \quad x^2 < a \quad \Leftrightarrow \quad x < \sqrt{a},$$

Hence:

$$f \text{ is strictly increasing} \quad \Leftrightarrow \quad x > \sqrt{a},$$

$$f \text{ is strictly decreasing} \quad \Leftrightarrow \quad x < \sqrt{a}.$$

Hence finally $x = \sqrt{a}$ is a minimum for f. When $x = \sqrt{a}$ we have $y = \sqrt{a}$ also, because

$$y = a/x = a/\sqrt{a}.$$

This proves that the rectangle is a square.

Example. Show that among all rectangular fences with given length, the one encompassing the largest area must be a square.

To do this, let c be the fixed length, and let x be one of the sides. If y is the other side, then

$$2x + 2y = c,$$

so that $y = (c - 2x)/2$. Therefore the area encompassed by the fence is equal to

$$xy = \frac{x(c - 2x)}{2} = \frac{xc - 2x^2}{2} = A(x).$$

This area $A(x)$ is a function of x, which has a critical point when $A'(x) = 0$. But

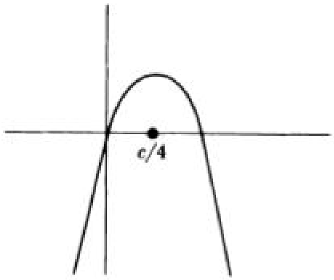
$$A'(x) = \tfrac{1}{2}(c - 4x).$$

Thus $A'(x) = 0$ if and only if $c = 4x$, that is, $x = c/4$ is the only critical point.

We must now see that this is a maximum. The function

$$A(x) = \frac{xc - 2x^2}{2}$$

has a graph which is a parabola. When x becomes large positive or negative, then $A(x)$ becomes large negative. Hence the parabola is shaped as in the following figure.



It has only one maximum, and that maximum must therefore be at $x = c/4$. We then find that $y = c/4$ also. In other words, the fence must be a square.

Inequalities

The derivative test for increasing and decreasing functions can also be used to prove inequalities.

Example. Prove that $\sin x < x$ for all $x > 0$.
Let $f(x) = x - \sin x$. Then

$$f'(x) = 1 - \cos x.$$

First take $0 < x < \pi/2$. Then $f'(x) > 0$ because $\cos x < 1$ in this interval. Therefore $f(x)$ is strictly increasing for $0 \leqq x \leqq \pi/2$. But

$$f(0) = 0 - \sin 0 = 0.$$

Hence we must have $f(x) > 0$ for $0 < x \leqq \pi/2$.

If $x \geqq \pi/2$, then $x > 1$ (because π is approximately 3.14), and so $\sin x < x$ whenever $x > \pi/2$. Thus the desired inequality holds for simpler reasons when $x > \pi/2$.

The preceding example illustrates a technique which is used for proving certain inequalities between functions. In general:

Suppose we have two functions f and g over a certain interval $[a, b]$ and we assume that f, g are differentiable. Suppose that

$$f(a) \leqq g(a),$$

and that $f'(x) \leqq g'(x)$ throughout the interval. Then $f(x) \leqq g(x)$ in the interval.

Proof. We let

$$h(x) = g(x) - f(x).$$

Then by assumption,

$$h'(x) = g'(x) - f'(x) \geqq 0,$$

so h is increasing throughout the interval. Since

$$h(a) = g(a) - f(a) \geqq 0,$$

it follows that $h(x) \geqq 0$ throughout the interval, whence

$$g(x) \geqq f(x).$$

The principle just stated can be visualized in the following picture, drawn for the case when $f(a) = g(a)$.

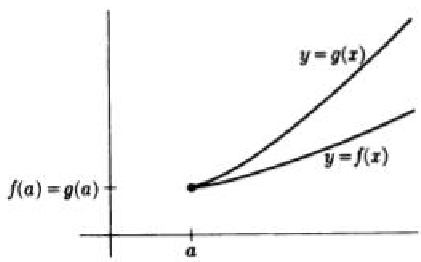


Figure 12

In other words, if g is bigger than or equal to f at $x = a$, and if g grows faster than f, then $g(x)$ is bigger than $f(x)$ for all $x > a$.

Example. Show that for any integer $n \geqq 1$ and any number $x \geqq 1$ one has the inequality

$$x^n - 1 \geqq n(x - 1).$$

Let $f(x) = x^n - 1 - n(x - 1)$. Then

$$f'(x) = nx^{n-1} - n.$$

Since $x \geqq 1$ it follows that $x^{n-1} \geqq 1$ and so $f'(x) \geqq 0$. Hence f is increasing for $x \geqq 1$. But $f(1) = 0$. Hence $f(x) \geqq 0$ for $x \geqq 1$. This is equivalent to the desired inequality.

On the other hand, the next theorem tells us what happens if two functions have the same derivative throughout an interval.

Constants

Theorem 2.2. *Let $f(x)$ and $g(x)$ be two functions which are differentiable in some interval and assume that*

$$f'(x) = g'(x)$$

for all x in the interval. Then there is a constant C such that

$$f(x) = g(x) + C$$

for all x in the interval.

Proof. Let $h(x) = f(x) - g(x)$ be the difference of our two functions. Then

$$h'(x) = f'(x) - g'(x) = 0.$$

Hence $h(x)$ is constant by Theorem 2.1, that is $h(x) = C$ for some number C and all x. This proves the theorem.

Remark. The theorem is the *converse* of the statement:

If a function is constant, then its derivative is equal to 0.

We shall use Theorem 2.2 in a fundamental way when we come to the chapter on integration.

For the applications of the theorem, see the beginning of the chapter on logarithms, and also the beginning of Chapter X, §1. We give simpler applications here.

Example. Let f be a function of x such that $f'(x) = 5$. Suppose that $f(0) = 2$. Determine $f(x)$ completely.

We know from past experience that the function

$$g(x) = 5x$$

has the derivative

$$g'(x) = 5.$$

Hence there is constant C such that

$$f(x) = 5x + C.$$

We are also given $f(0) = 2$. Hence

$$2 = f(0) = 0 + C.$$

Therefore $C = 2$. Thus finally

$$f(x) = 5x + 2.$$

Example. A particle moves on the x-axis toward the left at a rate of 5 cm/sec. At time $t = 5$ the particle is at the point 8 cm to the right of the origin. Determine the x-coordinate $x = f(t)$ completely as a function of time.

We are given

$$\frac{dx}{dt} = f'(t) = -5.$$

Let $g(t) = -5t$. Then $g'(t) = -5$ also. Hence there is a constant C such that

$$f(t) = -5t + C.$$

But we are also given $f(5) = 8$. Hence

$$8 = -5 \cdot 5 + C = -25 + C.$$

Therefore $C = 8 + 25 = 33$, so finally

$$f(t) = -5t + 33.$$

V, §2. EXERCISES

Determine the intervals on which the following functions are increasing and decreasing.

1. $f(x) = x^3 + 1$
2. $f(x) = x^2 - x + 5$
3. $f(x) = x^3 + x - 2$
4. $f(x) = -x^3 + 2x + 1$
5. $f(x) = 2x^3 + 5$
6. $f(x) = 5x^2 + 1$
7. $f(x) = -4x^3 - 2x$
8. $f(x) = 5x^3 + 6x$

Sketch the graphs of the following parabolas. Determine the critical point in each case.

9. $f(x) = x^2 - x - 1$
10. $f(x) = x^2 + x + 1$
11. $f(x) = -x^2 + x - 1$
12. $f(x) = -x^2 - x - 1$
13. $f(x) = x^2 + 3x + 1$
14. $f(x) = x^2 - 5x + 1$
15. $f(x) = -2x^2 + 4x - 1$
16. $f(x) = 2x^2 - 4x - 3$

For each of the following functions, find the maximum, minimum for all x in the given interval.

17. $x^2 - 2x - 8$, $[0, 4]$
18. $x^2 - 2x + 1$, $[-1, 4]$
19. $4 - 4x - x^2$, $[-1, 4]$
20. $x - x^2$, $[-1, 2]$
21. $3x - x^3$, $[-2, \sqrt{3}]$
22. $(x - 4)^5$, $[3, 6]$
23. The following steps show how to prove inequalities for the sine and cosine. We start with the inequality proved as an example in the text, namely

(a) $\sin x \leqq x$ for all $x \geqq 0$.

Let $f_1(x) = x - \sin x$. Then this inequality is equivalent with the inequality

$$f_1(x) \geqq 0 \quad \text{for all } x \geqq 0. \tag{1}$$

Now prove:

(b) $1 - \dfrac{x^2}{2} \leqq \cos x$ for $x \geqq 0$.

[*Hint*: Let $f_2(x) = \cos x - \left(1 - \dfrac{x^2}{2}\right)$ and use (1), to prove

$$f_2(x) \geqq 0 \quad \text{for all } x \geqq 0.] \tag{2}$$

(c) $x - \dfrac{x^3}{3 \cdot 2} \leqq \sin x$. $\left[\text{\textit{Hint}: Let } f_3(x) = \sin x - \left(x - \dfrac{x^3}{3 \cdot 2}\right).\right]$

(d) $\cos x \leqq 1 - \dfrac{x^2}{2} + \dfrac{x^4}{4 \cdot 3 \cdot 2}$

(e) $\sin x \leqq x - \dfrac{x^3}{3 \cdot 2} + \dfrac{x^5}{5 \cdot 4 \cdot 3 \cdot 2}$

24. Prove that $\tan x > x$ if $0 < x < \pi/2$.

25. (a) Prove that

$$t + \frac{1}{t} \geqq 2 \qquad \text{for} \quad t > 0.$$

[*Hint*: Let $f(t) = t + 1/t$. Show that f is strictly decreasing for $0 < t \leqq 1$ and f is strictly increasing for $1 \leqq t$. What is $f(1)$?]

(b) Let a, b be two positive numbers. Let

$$f(x) = ax + \frac{b}{x} \qquad \text{for} \quad x > 0.$$

Show that the minimum value of f is $2\sqrt{ab}$.

26. A box with open top is to be made with a square base and a constant surface C. Determine the sides of the box if the volume is to be a maximum.

27. A container in the shape of a cylinder with open top is to have a fixed surface area C. Find the radius of its base and its height if it is to have maximum volume.

28. Do the above two problems when the box and the container are closed at the top. (The area of a circle of radius x is πx^2 and its length is $2\pi x$. The volume of a cylinder whose base has radius x and of height y is $\pi x^2 y$.)

29. Assume that there is a function $f(x)$ such that $f(x) \neq 0$ for all x, and $f'(x) = f(x)$. Let $g(x)$ be any function such that $g'(x) = g(x)$. Show that there is a constant C such that $g(x) = Cf(x)$. [*Hint*: Differentiate the quotient g/f.]

30. Suppose that f is a differentiable function of t such that (a) $f'(t) = -3$, (b) $f'(t) = 2$. What can you say about $f(t)$?

31. Suppose that $f'(t) = -3$ and $f(0) = 1$. Determine $f(t)$ completely.

32. Suppose that $f'(t) = 2$ and $f(0) = -5$. Determine $f(t)$ completely.

33. A particle is moving on the x-axis toward the right at a constant speed of 7 ft/sec. If at time $t = 9$ the particle is at a distance 2 ft to the right of the origin, find its x-coordinate as a function of t.

34. Water is dripping out of a vertical tank so that the height of the water is falling at a rate of 2 ft/day. When the tank is full, the height of the water is 30 ft. Find explicitly the height of the water as a function of time.

V, §3. THE MEAN VALUE THEOREM

The theorems in this section are fairly obvious intuitively, and therefore you might omit the proofs of Rolle's theorem and Theorem 3.2 if you wish, **after understanding their statement**.

First suppose we have a function over a closed interval $[a, b]$, whose graph looks like this.

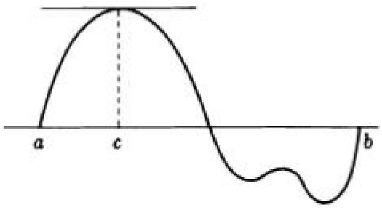


Figure 13

Then we have the following theorem about this function.

Theorem 3.1. Rolle's theorem. *Let a, b be two numbers, $a < b$. Let f be a function which is continuous over the closed interval*

$$a \leqq x \leqq b$$

and differentiable on the open interval $a < x < b$. Assume that

$$f(a) = f(b) = 0.$$

Then there exists a point c such that

$$a < c < b$$

and such that $f'(c) = 0$.

Proof. If the function is constant in the interval, then its derivative is 0 and any point in the open interval $a < x < b$ will do.

If the function is not constant, then there exists some point in the interval where the function is not 0, and this point cannot be one of the end points a or b. Suppose that some value of our function is positive. By Theorem 1.1, the function has a maximum at a point c. Then $f(c)$ must be greater than 0, and c cannot be either one of the end points because $f(a) = f(b) = 0$. Consequently

$$a < c < b.$$

By Theorem 1.3, we must have $f'(c) = 0$. This proves our theorem in case the function is positive somewhere in the interval.

If the function is negative for some number in the interval, then we use Theorem 1.1 to get a minimum, and we argue in a similar way, using Theorem 1.3 (applied to a minimum). (Write out the argument in full as an exercise.)

Let $f(x)$ be a function which is differentiable for $a < x < b$, and continuous in the closed interval

$$a \leqq x \leqq b.$$

We continue to assume throughout that $a < b$. This time we do not assume, as in Theorem 3.1, that $f(a) = f(b) = 0$. We shall prove that there exists a point c between a and b such that the slope of the tangent line at $(c, f(c))$ is the same as the slope of the line between the end points of our graph. In other words, the tangent line is parallel to the line passing through the end points of our graph.

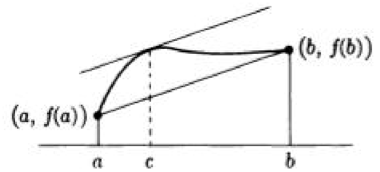


Figure 14

The slope of the line between the two end points is

$$\frac{f(b) - f(a)}{b - a}$$

because the coordinates of the end points are $(a, f(a))$ and $(b, f(b))$ respectively. Thus we have to find a point c such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Theorem 3.2. Mean value theorem. *Let $a < b$ as before. Let f be a function which is continuous in the closed interval $a \leqq x \leqq b$, and differentiable in the interval $a < x < b$. Then there exists a point c such that $a < c < b$ and*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Proof. The equation of the line between the two end points is

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

Indeed, the slope

$$\frac{f(b) - f(a)}{b - a}$$

is the coefficient of x. When $x = a$, $y = f(a)$. Hence we have written down the equation of the line having the given slope and passing through a given point. When $x = b$, we note that $y = f(b)$.

We now consider geometrically the difference between $f(x)$ and the straight line. This difference becomes 0 at the end points. This geometric idea will allow us to apply Rolle's theorem. In other words, we consider the function

$$g(x) = f(x) - \frac{f(b) - f(a)}{b - a}(x - a) - f(a).$$

Then

$$g(a) = f(a) - f(a) = 0$$

and

$$g(b) = f(b) - f(b) = 0$$

also.

We can therefore apply Theorem 3.1 to the function $g(x)$. We know that there is a point c between a and b, and not equal to a or b, such that

$$g'(c) = 0.$$

But

$$g'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}.$$

Consequently

$$0 = g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a}.$$

This gives us the desired value for $f'(c)$, and concludes the proof.

The point of the mean value theorem is not so much to find explicitly a value c such that

$$f'(c) = \frac{f(b) - f(a)}{b - a},$$

as to use it for theoretical considerations.

Corollary 3.3. *Let f be a function which is differentiable in some interval, and such that $f'(x) = 0$ for all x in the interval. Then f is constant.*

Proof. Let a, b be distinct numbers in the interval. By the mean value theorem, there exists c between a and b such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

But $f'(c) = 0$ by assumption. Hence $f(b) - f(a) = 0$ and therefore $f(b) = f(a)$. Hence f has the same value at all points of the interval, so f is constant, as was to be shown.

Similarly, we may now give the rest of the proof of Theorem 2.1 from the preceding section:

Corollary 3.4. *If $f'(x) > 0$ for x in an interval, excluding the end points, and f is continuous in the interval, then f is strictly increasing.*

Proof. Let x_1 and x_2 be two points of the interval, and suppose $x_1 < x_2$. By the mean value theorem, there exists a point c such that $x_1 < c < x_2$ and

$$f'(c) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

The difference $x_2 - x_1$ is positive, and we have

$$f(x_2) - f(x_1) = (x_2 - x_1)f'(c).$$

If the derivative $f'(x)$ is > 0 for all x in the interval, excluding the end points, then $f'(c) > 0$ (because c is in the interval). Hence the product $(x_2 - x_1)f'(c)$ is positive, and $f(x_2) - f(x_1) > 0$, so that

$$f(x_1) < f(x_2).$$

This proves that the function is increasing.

We leave the proof of the assertion concerning decreasing functions as an exercise.

When we study Taylor's formula, we shall use mean value theorems to estimate various functions. Even though we don't know the exact value $f'(c)$, we still may have an estimate for the derivative. For instance the functions $\sin x$ and $\cos x$ can both be estimated in absolute value by 1. In many applications, this is all that matters.

CHAPTER VI

Sketching Curves

We have developed enough techniques to be able to sketch curves and graphs of functions much more efficiently than before. We shall investigate systematically the behavior of a curve, and the mean value theorem will play a fundamental role.

We shall especially look for the following aspects of the curve:

1. Intersections with the coordinate axes.
2. Critical points.
3. Regions of increase.
4. Regions of decrease.
5. Maxima and minima (including the local ones).
6. Behavior as x becomes large positive and large negative.
7. Values of x near which y becomes large positive or large negative.

These seven pieces of information will be quite sufficient to give us a fairly accurate idea of what the graph looks like. We shall devote a section to considering one other aspect, namely:

8. Regions where the curve is bending up or down.

VI, §1. BEHAVIOR AS x BECOMES VERY LARGE

Suppose we have a function f defined for all sufficiently large numbers. Then we get substantial information concerning our function by investigating how it behaves as x becomes large.

For instance, $\sin x$ oscillates between -1 and $+1$ no matter how large x is.

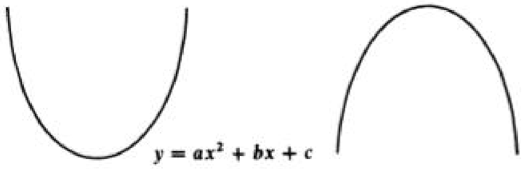
However, polynomials don't oscillate. When $f(x) = x^2$, as x becomes large positive, so does x^2. Similarly with the function x^3, or x^4 (etc.). We consider this systematically.

Parabolas

Example 1. Consider a parabola,

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

with $a \neq 0$. There are two essential cases, when $a > 0$ and $a < 0$. We shall see that the parabola looks like those drawn in the figure.



Case $a > 0$ **Case $a < 0$**

We look at numerical examples.

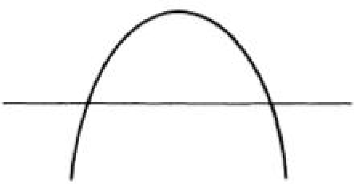
Example 2. Sketch the graph of the curve

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

We recognize this as a parabola. Factoring out x^2 shows that

$$f(x) = x^2\left(-3 + \frac{5}{x} - \frac{1}{x^2}\right).$$

When x is large positive or negative, then x^2 is large positive and the factor on the right is close to -3. Hence $f(x)$ is large negative. This means that the parabola has the shape as shown on the figure.



We have $f'(x) = -6x + 5$. Thus $f'(x) = 0$ if and only if $-6x + 5 = 0$, or in other words,

$$x = \tfrac{5}{6}.$$

There is exactly one critical point. We have

$$f(\tfrac{5}{6}) = -3(\tfrac{5}{6})^2 + \tfrac{25}{6} - 1 > 0.$$

The critical point is a maximum, because we have already seen that the parabola bends down.

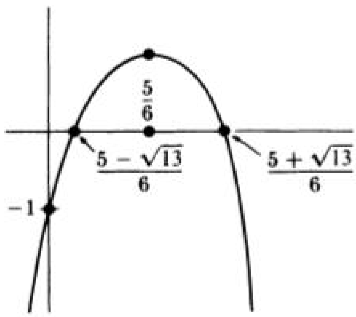
The curve crosses the x-axis exactly when

$$-3x^2 + 5x - 1 = 0.$$

By the quadratic formula (see Chapter II, §8), this is the case when

$$x = \frac{-5 \pm \sqrt{25 - 12}}{-6} = \frac{5 \pm \sqrt{13}}{6}.$$

Hence the graph of the parabola looks as on the figure.



The same principle applies to sketching any parabola.

(i) Looking at what happens when x becomes large positive or negative tells us whether the parabola bends up or down.

(ii) A quadratic function

$$f(x) = ax^2 + bx + c, \qquad \text{with} \quad a \neq 0$$

has only one critical point, when

$$f'(x) = 2ax + b = 0$$

so when

$$x = -b/2a.$$

Knowing whether the parabola bends up or down tells us whether the critical point is a maximum or minimum, and the value $x = -b/2a$ tells us exactly where this critical point lies.

(iii) The points where the parabola crosses the x-axis are determined by the quadratic formula.

Example 3. Cubics. Consider a polynomial

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

We can write it in the form

$$x^3\left(1 + \frac{2}{x^2} - \frac{1}{x^3}\right).$$

When x becomes very large, the expression

$$1 + \frac{2}{x^2} - \frac{1}{x^3}$$

approaches 1. In particular, given a small number $\delta > 0$, we have, for all x sufficiently large, the inequality

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

Therefore $f(x)$ satisfies the inequality

$$x^3(1 - \delta) < f(x) < x^3(1 + \delta).$$

This tells us that $f(x)$ behaves very much like x^3 when x is very large. In particular:

If x becomes large positive, then $f(x)$ becomes large positive.

If x becomes large negative, then $f(x)$ becomes large negative.

A similar argument can be applied to any polynomial.

It is convenient to use an abbreviation for the expression "become large positive." Instead of saying x becomes large positive, we write

$$x \to \infty$$

and also say that x **approaches,** or **goes to infinity**. **Warning: there is no number called infinity**. The above symbols merely abbreviate the notion of becoming large positive. We have a similar notation for x becoming large negative, when we write

$$x \to -\infty$$

and say that x **approaches minus infinity**. Thus in the case when

$$f(x) = x^3 + 2x - 1,$$

we can assert:

$$\text{If } x \to \infty \quad \text{then } f(x) \to \infty.$$
$$\text{If } x \to -\infty \text{ then } f(x) \to -\infty.$$

Example 4. Consider a quotient of polynomials like

$$Q(x) = \frac{x^3 + 2x - 1}{2x^3 - x + 1}.$$

We factor out the highest power of x from the numerator and denominator, and therefore write $Q(x)$ in the form

$$Q(x) = \frac{x^3(1 + 2/x^2 - 1/x^3)}{x^3(2 - 1/x^2 + 1/x^3)} = \frac{1 + 2/x^2 - 1/x^3}{2 - 1/x^2 + 1/x^3}.$$

As x becomes very large, the numerator approaches 1 and the denominator approaches 2. Thus our fraction approaches $\frac{1}{2}$. We may express this in the form

$$\lim_{x \to \infty} Q(x) = \frac{1}{2}.$$

Or we may write:

$$\text{If } x \to \pm\infty \text{ then } Q(x) \to \tfrac{1}{2}.$$

Example 5. Consider the quotient

$$Q(x) = \frac{x^2 - 1}{x^3 - 2x + 1}.$$

Does it approach a limit as x becomes very large?

We write

$$Q(x) = \frac{x^2(1 - 1/x^2)}{x^3(1 - 2/x^2 + 1/x^3)}$$
$$= \frac{1}{x}\frac{1 - 1/x^2}{1 - 2/x^2 + 1/x^3}.$$

As x becomes large, the term $1/x$ approaches 0, and the other factor approaches 1. Hence $Q(x)$ approaches 0 as x becomes large negative or positive.

We may also write

$$\text{If } x \to \pm\infty \text{ then } Q(x) \to 0,$$

or

$$\lim_{x \to \pm\infty} Q(x) = 0.$$

Example 6. Consider the quotient

$$Q(x) = \frac{x^3 - 1}{x^2 + 5}$$

and determine what happens when x becomes large.

We write

$$\begin{aligned} Q(x) &= \frac{x^3(1 - 1/x^3)}{x^2(1 + 5/x^2)} \\ &= x\,\frac{1 - 1/x^3}{1 + 5/x^2}. \end{aligned}$$

As x becomes large, positive or negative, the quotient

$$\frac{1 - 1/x^3}{1 + 5/x^2}$$

approaches 1. Hence $Q(x)$ differs from x by a factor near 1. Hence $Q(x)$ becomes large positive when x is large positive, and becomes large negative when x is large negative. We may express this by saying:

$$\text{If } x \to \infty \quad \text{then } Q(x) \to \infty.$$
$$\text{If } x \to -\infty \text{ then } Q(x) \to -\infty.$$

We may also write these assertions in the form of a limit:

$$\lim_{x \to \infty} Q(x) = \infty \qquad \text{and} \qquad \lim_{x \to -\infty} Q(x) = -\infty.$$

However, even though we use this notation, and may say that the limit of $Q(x)$ is $-\infty$ when x becomes large negative, we emphasize that $-\infty$ is not a number, and so this limit is not quite the same as when the limit is a number. It is correct to say that there is no number which is the limit of $Q(x)$ as $x \to \infty$ or as $x \to -\infty$.

These four examples are typical of what happens when we deal with quotients of polynomials.

Later when we deal with exponents and logarithms, we shall again meet the problem of comparing the quotient of two expressions which become large. There will be a common ground for some of the arguments, summarized by the following table:

> Large positive times large positive is large positive.
> Large positive times large negative is large negative.
> Large negative times large negative is large positive.
> Small positive times large positive: you can't tell without knowing more information.

VI, §1. EXERCISES

Find the limits of the following quotients $Q(x)$ as x becomes large positive or negative. In other words, find

$$\lim_{x\to\infty} Q(x) \quad \text{and} \quad \lim_{x\to-\infty} Q(x).$$

1. $\dfrac{2x^3 - x}{x^4 - 1}$
2. $\dfrac{\sin x}{x}$
3. $\dfrac{\cos x}{x}$
4. $\dfrac{x^2 + 1}{\pi x^2 - 1}$
5. $\dfrac{\sin 4x}{x^3}$
6. $\dfrac{5x^4 - x^3 + 3x + 2}{x^3 - 1}$
7. $\dfrac{-x^2 + 1}{x + 5}$
8. $\dfrac{2x^4 - 1}{-4x^4 + x^2}$
9. $\dfrac{2x^4 - 1}{-4x^3 + x^2}$
10. $\dfrac{2x^4 - 1}{-4x^5 + x^2}$

Describe the behavior of the following polynomials as x becomes large positive and large negative.

11. $x^3 - x + 1$
12. $-x^3 - x + 1$
13. $x^4 + 3x^3 + 2$
14. $-x^4 + 3x^3 + 2$
15. $2x^5 + x^2 - 100$
16. $-3x^5 + x + 1000$
17. $10x^6 - x^4$
18. $-3x^6 + x^3 + 1$

19. A function $f(x)$ which can be expressed as follows:

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0,$$

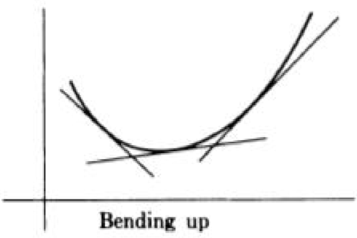
where n is a positive integer and the $a_n, a_{n-1}, \ldots, a_0$ are numbers, is called a polynomial. If $a_n \neq 0$, then n is called the **degree** of the polynomial. Describe the behavior of $f(x)$ as x becomes large positive or negative, n is odd or even, and $a_n > 0$ or $a_n < 0$. You will have eight cases to consider. Fill out the following table.

n	a_n	$x \to \infty$	$x \to -\infty$
Odd	> 0	$f(x) \to ?$	$f(x) \to ?$
Odd	< 0	$f(x) \to ?$	$f(x) \to ?$
Even	> 0	$f(x) \to ?$	$f(x) \to ?$
Even	< 0	$f(x) \to ?$	$f(x) \to ?$

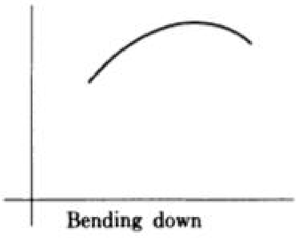
20. Using the intermediate value theorem, show that any polynomial of odd degree has a root.

VI, §2. BENDING UP AND DOWN

Let a, b be numbers, $a < b$. Let f be a continuous function defined on the interval $[a, b]$. Assume that f' and f'' exist on the interval $a < x < b$. We view the second derivative f'' as the rate of change of the slope of the curve $y = f(x)$ over the interval. If the second derivative is positive in the interval $a < x < b$, then the slope of the curve is increasing, and we interpret this as meaning that the curve is **bending up**. If the second derivative is negative, we interpret this as meaning that the curve is **bending down**. The following two figures illustrate this.



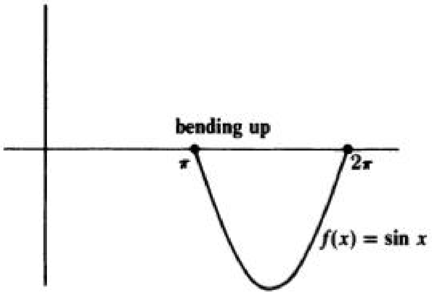
Bending up

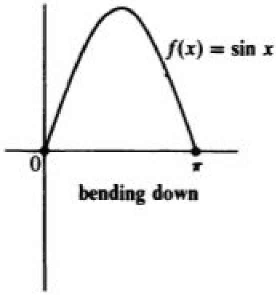


Bending down

Example 1. The curve $y = x^2$ is bending up. We can see this using the second derivative. Let $f(x) = x^2$. Then $f''(x) = 2$, and the second derivative is always positive. The present considerations justify drawing the curve as we have always done, i.e. bending up.

Example 2. Let $f(x) = \sin x$. We have $f''(x) = -\sin x$, and thus $f''(x) > 0$ on the interval $\pi < x < 2\pi$. Hence the curve is bending up on this interval. Similarly, $f''(x) < 0$ on the interval $0 < x < \pi$. Hence the curve is bending down on this interval, as shown on the next figures. Of course, this merely justifies the drawings which we have always made for the graph of the sine function.





Example 3. Determine the intervals where the curve

$$y = -x^3 + 3x - 5$$

is bending up and bending down.

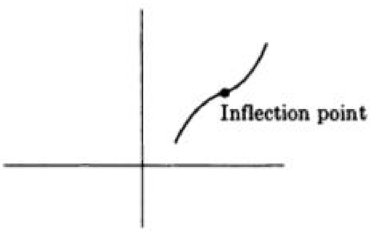
Let $f(x) = -x^3 + 3x - 5$. Then $f''(x) = -6x$. Thus:

$$f''(x) > 0 \iff x < 0,$$

$$f''(x) < 0 \iff x > 0.$$

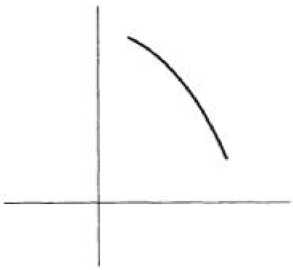
Hence f is bending up if and only if $x < 0$; and f is bending down if and only if $x > 0$. The graph of this curve will be discussed fully in the next section when we graph cubics systematically.

A point where a curve changes its behavior from bending up to down (or vice versa) is called an **inflection point**. If the curve is the graph of a function f whose second derivative exists and is continuous, then we must have $f''(x) = 0$ at that point. The following picture illustrates this:



In Example 3 above, the point $(0, -5)$ is an inflection point.

The determination of regions of bending up or down and inflection points gives us worthwhile pieces of information concerning curves. For instance, knowing that a curve in a region of decrease is actually bending down tells us that the decrease occurs essentially as in this example:



and not as in these examples:

The second derivative can also be used as a test whether a critical point is a **local** maximum or minimum.

Second derivative test. *Let f be twice continuously differentiable on an open interval, and suppose that c is a point where*

$$f'(c) = 0 \qquad \text{and} \qquad f''(c) > 0.$$

Then c is a local minimum point of f. On the other hand, if

$$f''(c) < 0$$

then c is a local maximum point of f.

To see this, suppose that $f''(c) > 0$. Then $f''(x) > 0$ for all x close to c because we assumed that the second derivative is continuous. Thus the curve is bending up. Consequently the picture of the graph of f is as on Fig. 1(a) and c is a local minimum.

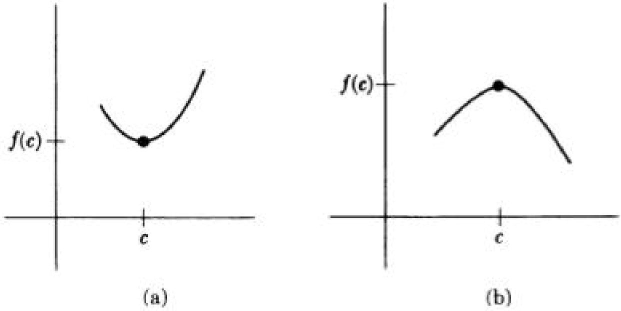


Figure 1

A similar argument shows that if $f''(c) < 0$ then c is a local maximum as on Fig. 1(b).

VI, §2. EXERCISES

1. Determine all inflection points of $\sin x$.
2. Determine all inflection points of $\cos x$.
3. Determine the inflection points of $f(x) = \tan x$ for $-\pi/2 < x < \pi/2$.
4. Sketch the curve $y = \sin^2 x$. Determine the critical points and the inflection points. Compare with the graph of $|\sin x|$.
5. Sketch the curve $y = \cos^2 x$. Determine the critical points and the inflection points. Compare with the graph of $|\cos x|$.

Determine the inflection points and the intervals of bending up and bending down for the following curves.

6. $y = x + \dfrac{1}{x}$
7. $y = \dfrac{x}{x^2 + 1}$
8. $y = \dfrac{x}{x^2 - 1}$
9. Sketch the curve $y = f(x) = \sin x + \cos x$ for $0 \leqq x \leqq 2\pi$. First plot all values $f(n\pi/4)$ with $n = 0, 1, 2, 3, 4, 5, 6, 7, 8$. Then determine all the critical points. Then determine the regions of increase and decrease. Then determine the inflection points, and the regions where the curve bends up or down.

VI, §3. CUBIC POLYNOMIALS

We can now sketch the graphs of cubic polynomials systematically.

Example 1. Sketch the graph of $f(x) = x^3 - 2x + 1$.

1.

$$\text{If } x \to \infty \quad \text{then } f(x) \to \infty \quad \text{by §1.}$$

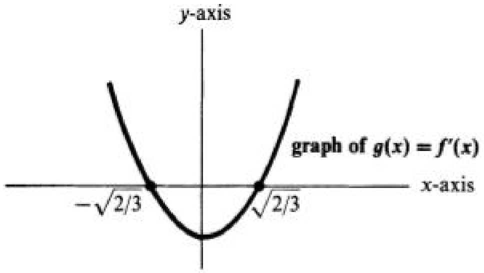
$$\text{If } x \to -\infty \text{ then } f(x) \to -\infty \text{ by §1.}$$

2. We have $f'(x) = 3x^2 - 2$. Thus

$$f'(x) = 0 \quad \Leftrightarrow \quad x = \pm\sqrt{2/3}.$$

The critical points of f are $x = \sqrt{2/3}$ and $x = -\sqrt{2/3}$.

3. Let $g(x) = f'(x) = 3x^2 - 2$. Then the graph of g is a parabola, and the x-intercepts of the graph of g are precisely the critical points of f. (Do not confuse the functions f and $f' = g$.) The graph of g is a parabola bending up, as follows.



Therefore:

$$f'(x) > 0 \quad \Leftrightarrow \quad x > \sqrt{2/3} \quad \text{and} \quad x < -\sqrt{2/3}, \quad \text{where} \quad g(x) > 0$$

and f is strictly increasing on the intervals

$$x \geqq \sqrt{2/3} \qquad \text{and} \qquad x \leqq -\sqrt{2/3}.$$

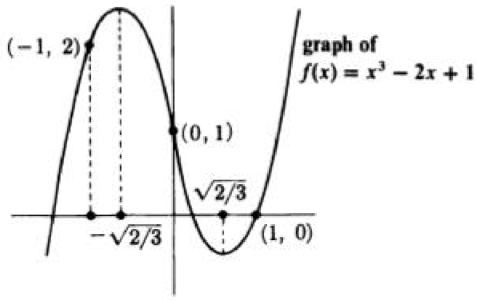
Similarly:

$$f'(x) < 0 \quad \Leftrightarrow \quad -\sqrt{2/3} < x < \sqrt{2/3}, \quad \text{where} \quad g(x) < 0,$$

and f is strictly decreasing on this interval. Therefore $-\sqrt{2/3}$ is a local maximum for f, and $\sqrt{2/3}$ is a local minimum.

4. $f''(x) = 6x$, and $f''(x) > 0$ if and only if $x > 0$. Also $f''(x) < 0$ if and only if $x < 0$. Therefore f is bending up for $x > 0$ and bending down for $x < 0$. There is an inflection point at $x = 0$.

Putting all this together, we find that the graph of f looks like this.



Observe how we used a quadratic polynomial, namely $f'(x) = 3x^2 - 2$, as an intermediate step in the arguments.

Remark 1. Instead of using the quadratic polynomial, we can also argue as follows, after we know that the only critical points of f are $x = \sqrt{2/3}$ and $x = -\sqrt{2/3}$. Consider the interval $x < -\sqrt{2/3}$. Then $f'(x) \neq 0$ for all $x < \sqrt{-2/3}$. Hence $f'(x)$ is either > 0 for all $x < -\sqrt{2/3}$, or $f'(x) < 0$ for all $x < -\sqrt{2/3}$ by the intermediate value theorem. Which is it? We just try one value, say with $x = -10$, to see that $f'(x) > 0$ for $x < -\sqrt{2/3}$, because $f'(-10) = 3 \cdot 10^2 - 2 = 298$. Hence we must have $f'(x) > 0$ for $x < -\sqrt{2/3}$.

Remark 2. For a cubic polynomial it is much more difficult to determine the roots, that is the x-intercepts, and we usually do not do so, unless there is a simple way of doing it, by accident. In the above case when

$$f(x) = x^3 - 2x + 1,$$

there is such an accident, since $f(1) = 0$. Therefore 1 is a root of f. Hence $f(x)$ factors

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

The other roots of f are the roots of $x^2 + x - 1$, which can be found by the quadratic formula:

$$x = \frac{-1 \pm \sqrt{5}}{2}.$$

In the next example, however, there is no such simple way of finding the roots, and we do not find them.

Example 2. Sketch the graph of the curve

$$y = -x^3 + 3x - 5.$$

1. When $x = 0$, we have $y = -5$. With polynomials of degree $\geqq 3$ there is in general no simple formula for those x such that $f(x) = 0$, so we do not give explicitly the intersection of the graph with the x-axis.
2. The derivative is

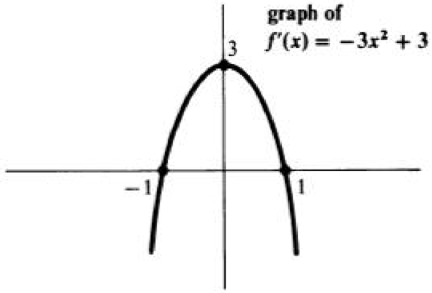
$$f'(x) = -3x^2 + 3.$$

The graph of $f'(x)$ is a parabola bending down, as you should know from previous experience with parabolas. We have

$$f'(x) = 0 \quad \Leftrightarrow \quad x^2 = 1 \quad \Leftrightarrow \quad x = 1 \text{ and } x = -1.$$

Thus there are two critical points of f, namely $x = 1$ and $x = -1$.

3. The graph of $f'(x)$ looks like a parabola bending down, as follows.



Then:

$$\begin{aligned} f \text{ is strictly decreasing} \quad &\Leftrightarrow \quad f'(x) < 0 \\ &\Leftrightarrow \quad x < -1 \quad \text{and} \quad x > 1. \\ f \text{ is strictly increasing} \quad &\Leftrightarrow \quad f'(x) > 0 \\ &\Leftrightarrow \quad -1 < x < 1. \end{aligned}$$

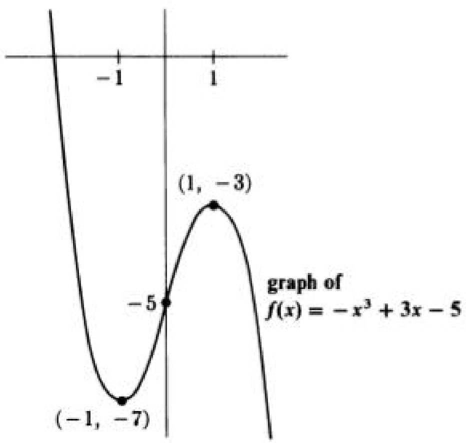
Therefore f has a local minimum at $x = -1$, and has a local maximum at $x = 1$.

4.

$$\begin{aligned} &\text{If } x \to \infty \quad \text{then } f(x) \to -\infty \text{ by §1.} \\ &\text{If } x \to -\infty \text{ then } f(x) \to +\infty \text{ by §1.} \end{aligned}$$

5. We have $f''(x) = -6x$. Hence $f''(x) > 0$ if and only if $x < 0$ and $f''(x) < 0$ if and only if $x > 0$. There is an inflection point at $x = 0$.

Putting all this information together, we see that the graph of f looks like this:



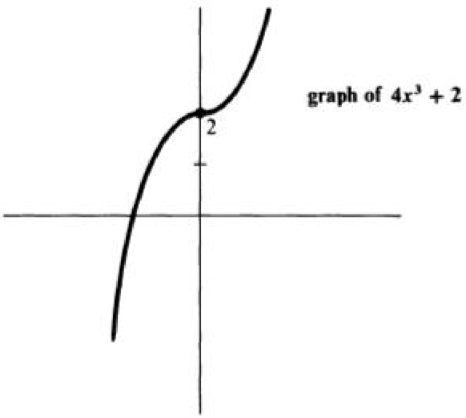
Remark. When f is a polynomial of degree 3, its derivative $f'(x)$ is a polynomial of degree 2, and in general this polynomial has two roots, giving the two critical points of the curve $y = f(x)$. In the preceding example, these critical points are at $(-1, -7)$ and $(1, -3)$.

Again note how we used the graph of a parabola, namely the graph of $f'(x)$, in the process of determining the graph of f itself.

In the last two examples, the cubic polynomial had two bumps, at the two critical points. This is the most general form of cubic polynomials. However, there may be special cases, when there is no critical point, or only one critical point.

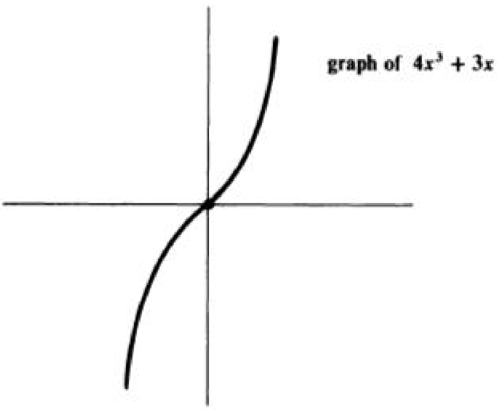
Example 3(a). Let $f(x) = 4x^3 + 2$. Sketch the graph of f.

Here we have $f'(x) = 12x^2 > 0$ for all $x \neq 0$. There is only one critical point, when $x = 0$. Hence the function is strictly increasing for all x, and its graph looks like this.



Example 3(b). Let $f(x) = 4x^3 + 3x$. Sketch the graph of f.

Here we have $f'(x) = 12x^2 + 3 > 0$ for *all* x. Therefore the graph of f looks like this. There is no critical point.



In both examples, we have

$$f''(x) = 24x.$$

Thus in both examples, there is an inflection point at $x = 0$. The graph of f bends down for $x < 0$ and bends up for $x > 0$. The difference between case (a) and case (b) is that in case (a) the inflection point is a critical point, where the derivative of f is equal to 0, so that the curve is flat at the critical point. In case (b), the derivative at the inflection point is

$$f'(0) = 3,$$

so in case (b) the derivative at the inflection point is positive.

VI, §3. EXERCISES

1. Show that a curve

$$y = ax^3 + bx^2 + cx + d$$

with $a \neq 0$ has exactly one inflection point.

Sketch the graphs of the following curves.

2. $x^3 - 2x^2 + 3x$
3. $x^3 + x^2 - 3x$
4. $2x^3 - x^2 - 3x$
5. $\frac{1}{3}x^3 + x^2 - 2x$
6. $x^3 - 3x^2 + 6x - 3$
7. $x^3 + x - 1$
8. $x^3 - x - 1$
9. $-x^3 + 2x + 5$
10. $-2x^3 + x + 2$
11. $x^3 - x^2 + 1$
12. $y = x^4 + 4x$
13. $y = x^5 + x$
14. $y = x^6 + 6x$
15. $y = x^7 + x$
16. $y = x^8 + x$
17. Which of the following polynomials have a minimum (for all x)?
    (a) $x^6 - x + 2$
    (b) $x^5 - x + 2$
    (c) $-x^6 - x + 2$
    (d) $-x^5 - x + 2$
    (e) $x^6 + x + 2$
    (f) $x^5 + x + 2$

    Sketch the graphs of these polynomials.
18. Which of the polynomials in Exercise 17 have a maximum (for all x)?

In the following two problems:
(a) Show that f has exactly two inflection points.
(b) Sketch the graph of f. Determine the critical points explictly. Determine the regions of bending up or down.

19. $f(x) = x^4 + 3x^3 - x^2 + 5$

20. $f(x) = x^4 - 2x^3 + x^2 + 3$

21. Sketch the graph of the function

$$f(x) = x^6 - \tfrac{3}{2}x^4 + \tfrac{9}{16}x^2 - \tfrac{1}{32}.$$

Find the critical points. Find the values of f at these critical points. Sketch the graph of f. It will come out much neater than may be apparent at first.

VI, §4. RATIONAL FUNCTIONS

We shall now consider quotients of polynomials.

Example. Sketch the graph of the curve

$$y = f(x) = \frac{x-1}{x+1}$$

and determine the eight properties stated in the introduction.

1. When $x = 0$, we have $f(x) = -1$. When $x = 1$, $f(x) = 0$.
2. The derivative is

$$f'(x) = \frac{2}{(x+1)^2}.$$

(You can compute it using the quotient rule.) It is never 0, and therefore the function f has no critical points.

3. The denominator is a square and hence is always positive, wherever it is defined, that is for $x \neq -1$. Thus $f'(x) > 0$ for all $x \neq -1$. The function is increasing for all x, Of course, the function is not defined for $x = -1$ and neither is the derivative. Thus it would be more accurate to say that the function is increasing in the region

$$x < -1$$

and is increasing in the region $x > -1$.

4. There is no region of decrease.
5. Since the derivative is never 0, there is no relative maximum or minimum.

6. The second derivative is

$$f''(x) = \frac{-4}{(x+1)^3}.$$

There is no inflection point since $f''(x) \neq 0$ for all x where the function is defined. If $x < -1$, then the denominator $(x+1)^3$ is negative, and $f''(x) > 0$, so the graph is bending upward. If $x > -1$, then the denominator is positive, and $f''(x) < 0$ so the graph is bending downward.

7. As x becomes large positive, our function approaches 1 (using the method of §1). As x becomes large negative, our function also approaches 1.

There is one more useful piece of information which we can look into, when $f(x)$ itself becomes large positive or negative. This occurs near points where the denominator of $f(x)$ is 0. On the present instance, $x = -1$.

8. As x approaches -1, the denominator approaches 0 and the numerator approaches -2. If x approaches -1 from the right so $x > -1$, then the denominator is positive, and the numerator is negative. Hence the fraction

$$\frac{x-1}{x+1}$$

is negative, and is large negative.

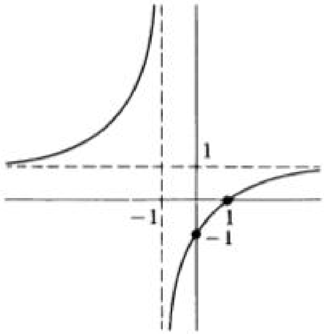


Figure 2

If x approaches -1 from the left so $x < -1$, then $x - 1$ is negative, but $x + 1$ is negative also. Hence $f(x)$ is positive and large, since the denominator is small when x is close to -1.

Putting all this information together, we see that the graph looks like that in the preceding figure.

We have drawn the two lines $x = -1$ and $y = 1$, as these play an important role when x approaches -1 and when x becomes large, positive or negative.

Remark. Again let

$$y = f(x) = \frac{x-1}{x+1}.$$

Then we can rewrite this relation to see directly that the graph of f is a hyperbola, as follows. We write the relation in the form

$$y = \frac{x+1-2}{x+1} = 1 - \frac{2}{x+1},$$

that is, $y - 1 = -2/(x+1)$. Clearing denominators, this gives

$$(y-1)(x+1) = -2.$$

By Chapter II, you should know that this is a hyperbola. We worked out a sketch by a more general method, because it also works in cases when you cannot reduce the equation to one of the standard curves, like circles, parabolas, or hyperbolas.

Example. Sketch the graph of $f(x) = \dfrac{x^2+x}{x-1}$.

Note that f is not defined at $x = 1$. We can rewrite

$$f(x) = \frac{x(x+1)}{x-1}.$$

We have $f(x) = 0$ if and only if the numerator $x(x+1) = 0$. Thus:

$$f(x) = 0 \quad \text{if and only if} \quad x = 0 \text{ or } x = -1.$$

Next we look at the derivative, which is

$$f'(x) = \frac{x^2-2x-1}{(x-1)^2}.$$

(Compute it using the quotient rule.) Then

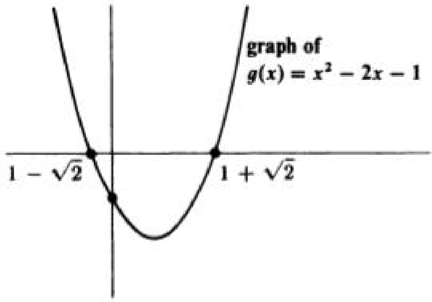
$$f'(x) = 0 \iff x^2 - 2x - 1 = 0,$$
$$\iff x = 1 \pm \sqrt{2} \qquad \text{(by the quadratic formula)}$$

These are the critical points of f.

The denominator $(x-1)^2$ in $f'(x)$ is a square and hence is always positive, wherever it is defined, that is for $x \neq 1$. Therefore the sign of $f'(x)$ is the same as the sign of its numerator $x^2 - 2x - 1$. Let

$$g(x) = x^2 - 2x - 1.$$

The graph of g is a parabola, and since the coefficient of x^2 is $1 > 0$, this parabola is bending up as shown on the figure.



The two roots of $g(x) = 0$ are $x = 1 - \sqrt{2}$ and $1 + \sqrt{2}$. From the graph of $g(x)$ we see that

$$g(x) < 0 \quad \text{when} \quad 1 - \sqrt{2} < x < 1 + \sqrt{2},$$
$$g(x) > 0 \quad \text{when} \quad x < 1 - \sqrt{2} \text{ or } x > 1 + \sqrt{2}.$$

This gives us the regions of increase and decrease for $f(x)$.

For $x \leqq 1 - \sqrt{2}$, $f(x)$ is strictly increasing.

For $1 - \sqrt{2} \leqq x < 1$, $f(x)$ is strictly decreasing.

For $1 < x \leqq 1 + \sqrt{2}$, $f(x)$ is strictly decreasing.

For $1 + \sqrt{2} \leqq x$, $f(x)$ is strictly increasing.

It follows that f has a local maximum at $x = 1 - \sqrt{2}$ and f has a local minimum at $x = 1 + \sqrt{2}$.

As x becomes large positive, $f(x)$ becomes large positive as is seen from the expression

$$f(x) = \frac{x^2 + x}{x - 1} = \frac{x^2(1 + 1/x)}{x(1 - 1/x)} = x\,\frac{1 + 1/x}{1 - 1/x}.$$

As x becomes large negative, $f(x)$ becomes large negative.

As x approaches 1 and $x < 1$, the function $f(x)$ becomes large negative because the denominator $x - 1$ approaches 0, and is negative, while the numerator $x^2 + x$ approaches 2.

As x approaches 1 and $x > 1$, the function $f(x)$ becomes large positive because the denominator $x - 1$ approaches 0 and both numerator and denominator are positive, while the numerator approaches 2.

Hence the graph looks as drawn on Fig. 3.

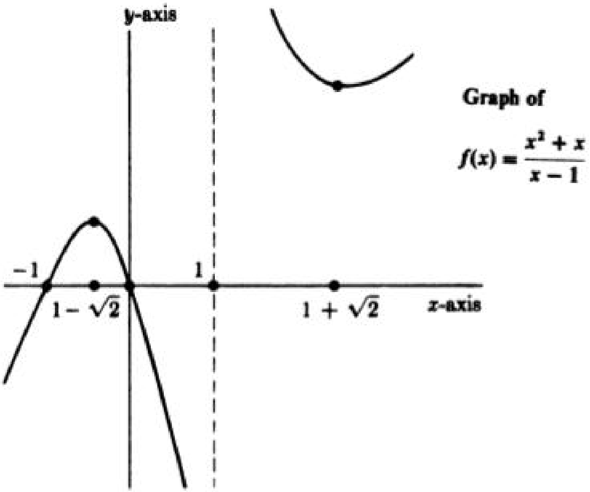


Figure 3

VI, §4. EXERCISES

Sketch the following curves, indicating all the information stated in the introduction. Regard convexity possibly as optional.

1. $y = \dfrac{x^2 + 2}{x - 3}$
2. $y = \dfrac{x - 3}{x^2 + 1}$
3. $y = \dfrac{x + 1}{x^2 + 1}$
4. $y = \dfrac{x^2 - 1}{x} = x - \dfrac{1}{x}$
5. $\dfrac{x}{x^3 - 1}$
6. $y = \dfrac{2x^2 - 1}{x^2 - 2}$
7. $y = \dfrac{2x - 3}{3x + 1}$
8. $\dfrac{4x}{x^2 - 9}$
9. $x + \dfrac{3}{x}$
10. $\dfrac{x^2 - 4}{x^3}$
11. $\dfrac{3x - 2}{2x + 3}$
12. $\dfrac{x}{3x - 5}$

13. $\dfrac{2x}{x+4}$

14. $\dfrac{x^2}{\sqrt{x+1}}$

15. $\dfrac{x+1}{x^2+5}$

16. $\dfrac{x+1}{x^2-5}$

17. $\dfrac{x^2+1}{x^2-1}$

18. $\dfrac{x^2-1}{x^2-4}$

19. Sketch the graph of $f(x) = x + 1/x$.

20. Let a, b be two positive numbers. Let

$$f(x) = ax + \frac{b}{x}$$

Show that the minimum value of $f(x)$ for $x > 0$ is $2\sqrt{ab}$. Give reasons for your assertions. Deduce that $\sqrt{ab} \leqq (a+b)/2$. Sketch the graph of f for $x > 0$.

VI, §5. APPLIED MAXIMA AND MINIMA

This section deals with word problems concerning maxima and minima, and applies the techniques discussed previously. In each case, we want to maximize or minimize a function, which is at first given in terms of perhaps two variables. We proceed as follows.

1. Enough data are given so that one of these variables can be expressed in terms of the other, by some relation. We end up dealing with a function of only one variable.
2. We then find its critical points, setting the derivative equal to 0, and then determine whether the critical points are local maxima or minima.
3. We verify if these local maxima or minima are also maxima or minima for the whole interval of definition of the function. If the function is given only on some finite interval, it may happen that the maximum, say, occurs at an end point, where the derivative test does not apply.

Example 1. Find the point on the graph of the equation $y^2 = 4x$ which is nearest to the point $(2, 3)$.

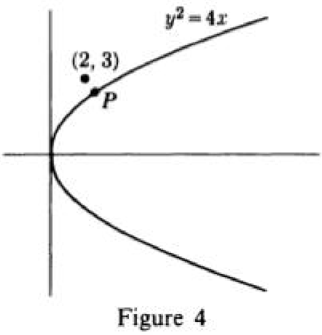


Figure 4

To minimize the distance between a point (x, y) and $(2, 3)$, it suffices to minimize the square of the distance, which has the advantage that no square root occurs in its formula. Indeed, suppose z_0^2 is a minimum value for the square of the distance, with z_0 positive. Then z_0 itself is a minimum value for the distance, because a positive number has a unique positive square root. The square of the distance is equal to

$$z^2 = (2 - x)^2 + (3 - y)^2.$$

Thus z^2 is expressed in terms of the two variables x, y. But we know that the point (x, y) lies on the curve whose equation is $y^2 = 4x$. Hence we can solve for one variable in terms of the other, namely $y = 2\sqrt{x}$. Substituting $y = 2\sqrt{x}$, we find an expression for the square of the distance only in terms of x, namely

$$\begin{aligned} f(x) &= (2 - x)^2 + (3 - 2\sqrt{x})^2 \\ &= 4 - 4x + x^2 + 9 - 12\sqrt{x} + 4x \\ &= 13 + x^2 - 12\sqrt{x}. \end{aligned}$$

We now determine the critical points of f. We have

$$f'(x) = 2x - \frac{6}{\sqrt{x}},$$

so

$$\begin{aligned} f'(x) = 0 \quad &\Leftrightarrow \quad 2x\sqrt{x} = 6 \\ &\Leftrightarrow \quad x = \sqrt[3]{9}. \end{aligned}$$

Furthermore, we have

$$\begin{aligned} f'(x) > 0 \quad &\Leftrightarrow \quad 2x\sqrt{x} > 6 \\ &\Leftrightarrow \quad x^3 > 9. \\ f'(x) < 0 \quad &\Leftrightarrow \quad x^3 < 9. \end{aligned}$$

Hence $f(x)$ is strictly increasing when $x > \sqrt[3]{9}$ and strictly decreasing when $x < \sqrt[3]{9}$. Hence $\sqrt[3]{9}$ is a minimum. When $x = \sqrt[3]{9}$ the corresponding value for y is

$$y = 2\sqrt{x} = 2\sqrt[3]{3}.$$

Hence the point on the graph of $y^2 = 4x$ closest to $(2, 3)$ is the point

$$P = (\sqrt[3]{9}, 2\sqrt[3]{3}).$$

Example 2. An oil can is to be made in the form of a cylinder to contain one quart of oil. What dimensions should it have so that the surface area is minimal (in other words, minimize the cost of material to make the can)?

Let r be the radius of the base of the cylinder and let h be its height. Then the volume is

$$V = \pi r^2 h.$$

The total surface area is the sum of the top, bottom, and circular sides, namely

$$A = 2\pi r^2 + 2\pi r h.$$

Thus the area is given in terms of the two variables r and h. However, we are also given that the volume V is constant, $V = 1$. Thus we get a relation between r and h,

$$\pi r^2 h = 1,$$

and we can solve for h in terms of r, namely

$$h = 1/\pi r^2.$$

Hence the area can be expressed entirely in terms of r, that is

$$A = 2\pi r^2 + 2\pi r/\pi r^2 = 2\pi r^2 + 2/r.$$

We want the area to be minimum. We first find the critical points of A. We have:

$$\begin{aligned} A'(r) = 4\pi r - 2/r^2 = 0 \quad &\Leftrightarrow \quad 4\pi r = 2/r^2 \\ &\Leftrightarrow \quad \pi r^3 = \tfrac{1}{2}. \end{aligned}$$

Thus we find exactly one critical point

$$r = \left(\frac{1}{2\pi}\right)^{1/3}$$

By physical considerations, we could see that this corresponds to a minimum, but we can also argue as follows. When r becomes large positive, or when r approaches 0, the function $A(r)$ becomes large, and so there has to be a minimum of the function for some value $r > 0$. This minimum is a critical point, and we have found that there is only one critical point. Hence we have found that the minimum occurs when r is the critical point. In this case we can solve back for h, namely

$$h = \frac{1}{\pi r^2} = \frac{(2\pi)^{2/3}}{\pi} = \frac{2^{2/3}}{\pi^{1/3}}.$$

This gives us the required dimensions.

Example 3. A truck is to be driven 200 mi at constant speed x mph. Speed laws require $30 \leqq x \leqq 60$. Assume that gasoline costs 50 cents/gallon and is consumed at the rate of

$$3 + \frac{x^2}{500} \text{ gal/hr.}$$

If the driver's wages are \$8 per hour, find the most economical speed.

We express the total cost as a sum of the cost of gasoline and the wages. The total time taken for the trip will be

$$\frac{200}{x}$$

because (time)(speed) = (distance) if the speed is constant. The cost of gas is then equal to the product of

(price per gallon)(number of gallons used per hr)(total time)

so that the cost of gasoline is

$$G(x) = \frac{1}{2}\left(3 + \frac{x^2}{500}\right)\frac{200}{x}.$$

(We write 1/2 because 50 cents $= \frac{1}{2}$ dollar.) On the other hand, the wages are given by the product

(wage per hour)(total time),

so that the cost of wages is

$$W(x) = 8 \cdot \frac{200}{x}.$$

Hence the total cost of the trip is

$$\begin{aligned} f(x) &= G(x) + W(x) \\ &= \frac{1}{2}\left(3 + \frac{x^2}{500}\right)\frac{200}{x} + \frac{8 \cdot 200}{x} \\ &= 100\left(\frac{3}{x} + \frac{x}{500}\right) + \frac{1600}{x}. \end{aligned}$$

We have

$$\begin{aligned} f'(x) &= -\frac{300}{x^2} + \frac{1}{5} - \frac{1600}{x^2} \\ &= -\frac{1900}{x^2} + \frac{1}{5}. \end{aligned}$$

Therefore $f'(x) = 0$ if and only if

$$\frac{1900}{x^2} = \frac{1}{5}$$

or in other words,

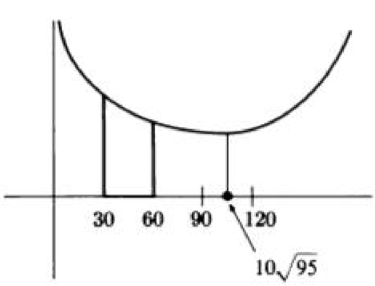
$$x^2 = 9500.$$

Thus $x = 10\sqrt{95}$. [We take x positive since this is the solution which has physical significance.]

Now we observe that $10\sqrt{95}$ is approximately equal to 10×10, and in any case is >60, so is beyond the speed limit of 60 which was assigned to begin with. Furthermore, if $0 < x < 10\sqrt{95}$ then

$$f'(x) < 0.$$

Hence the function $f(x)$ is decreasing for $0 \leqq x \leqq 10\sqrt{95}$. Its graph may be sketched as on the figure.



Since to begin with we restricted the possible speed to the interval $30 \leqq x \leqq 60$, it follows that the minimum of f over this interval must occur when $x = 60$. This is therefore the speed which minimizes the total cost.

Example 4. In the preceding example, suppose there is no speed limit. Then we see that if $x > 10\sqrt{95}$ then $f'(x) > 0$ so that $f(x)$ is increasing for

$$x > 10\sqrt{95}.$$

Therefore $10\sqrt{95}$ is a minimum point for f when no restriction is placed on x. Consequently, in this case, the speed which minimizes the cost is

$$x = 10\sqrt{95}.$$

Example 5. When light from a point source strikes a plane surface, the intensity of illumination is proportional to the cosine of the angle of incidence, and inversely proportional to the square of the distance from the source. How high should a light be located above the center of a circle of radius 12 ft to give the best illumination along the circumference?

The **angle of incidence** is measured from the perpendicular to the plane. The picture is as follows.

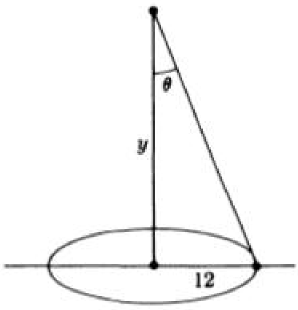


Figure 5

We denote by θ the angle of incidence, and by y the height of the light. Let I be the intensity of illumination. Two quantities are proportional means that there is a constant such that one is equal to the constant times the other. Thus there is a constant c such that

$$\begin{aligned} I(y) &= c\cos\theta\,\frac{1}{12^2 + y^2} \\ &= c\,\frac{y}{\sqrt{12^2 + y^2}}\,\frac{1}{12^2 + y^2} \\ &= \frac{cy}{(12^2 + y^2)^{3/2}}. \end{aligned}$$

The critical points of $I(y)$ are those points where $I'(y) = 0$. We have:

$$I'(y) = c\left[\frac{(12^2 + y^2)^{3/2} - y \cdot \frac{3}{2}(12^2 + y^2)^{1/2}(2y)}{(12^2 + y^2)^3}\right],$$

and this expression is equal to 0 precisely when the numerator is equal to 0, that is,

$$(12^2 + y^2)^{3/2} = 3y^2(12^2 + y^2)^{1/2}.$$

Canceling $(12^2 + y^2)^{1/2}$, we see that this is equivalent to

$$12^2 + y^2 = 3y^2,$$

or in other words,

$$12^2 = 2y^2.$$

Solving for y yields

$$y = \pm\frac{12}{\sqrt{2}}.$$

Only the positive value of y has physical significance, and thus the height giving maximum intensity is $12/\sqrt{2}$ ft, provided that we know that this critical point is a maximum for the function $I(y)$, for $y > 0$. This can be seen as follows.

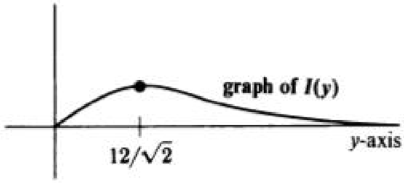
If y is very close to 0, then the numerator cy of $I(y)$ is close to 0, and the denominator $(12^2 + y^2)^{3/2}$ is close to $(12^2)^{3/2}$ so $I(y)$ tends to 0 when y is near 0. For large y, the denominator of $I(y)$ is analyzed by factoring out y^2, namely

$$(12^2 + y^2)^{3/2} = \left(\frac{12^2}{y^2} + 1\right)^{3/2} y^3.$$

Therefore $I(y)$ tends to 0 as y becomes large, because

$$I(y) = \frac{cy}{(\text{term near } 1)y^3} = \frac{c}{(\text{term near } 1)} \frac{1}{y^2}$$

if y is large positive. Hence we have shown that $I(y)$ tends to 0 when y approaches 0 or y becomes large. It follows that $I(y)$ reaches a maximum for some value of $y > 0$, and this maximum must be a critical point. On the other hand, we have also proved that there is only one critical point. Hence this critical point is the maximum, as desired. Thus $y = 12/\sqrt{2}$ is a maximum for the function. In view of the preceding discussion, the graph can be sketched as on the figure.



Example 6. A business makes automobile transmissions selling for \$400. The total cost of marketing x units is

$$f(x) = 0.02x^2 + 160x + 400{,}000.$$

How many transmissions should be sold for maximum profit?

Let $P(x)$ be the profit coming from selling x units. Then $P(x)$ is the difference between the total receipts and the cost of marketing. Hence

$$\begin{aligned} P(x) &= 400x - (0.02x^2 + 160x + 400{,}000) \\ &= -0.02x^2 + 240x - 400{,}000. \end{aligned}$$

We want to know when $P(x)$ is maximum. We have:

$$P'(x) = -0.04x + 240$$

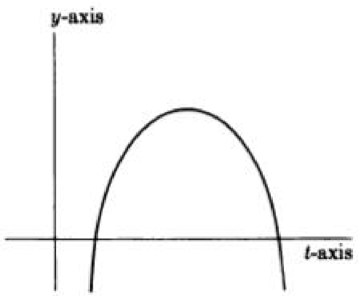
so the derivative is 0 when

$$0.04x = 240,$$

or in other words,

$$x = \frac{240}{0.04} = 6{,}000.$$

The equation $y = P(x)$ is a parabola, which bends down because the leading coefficient is -0.02 (negative). Hence the critical point is a maximum, and the answer is therefore 6,000 units.



Parabola $y = at^2 + bt + c$ with $a < 0$.

Example 7. A farmer buys a bull weighing 600 lbs, at a cost of \$180. It costs 15 cents per day to feed the animal, which gains 1 lb per day. Every day that the bull is kept, the sale price per pound declines according to the formula

$$B(t) = 0.45 - 0.00025t$$

where t is the number of days. How long should the farmer wait to maximize profits?

To do this, note that the total cost after time t is given by

$$f(t) = 180 + 0.15t.$$

The total sales amount to the product of the price $B(t)$ per pound times the weight of the animal, in other words,

$$\begin{aligned} S(t) &= (0.45 - 0.00025t)(600 + t) \\ &= -0.00025t^2 + 0.30t + 270. \end{aligned}$$

Hence the profit is

$$\begin{aligned} P(t) &= S(t) - f(t) \\ &= -0.00025t^2 + 0.15t + 90. \end{aligned}$$

Therefore

$$P'(t) = -0.0005t + 0.15$$

and $P'(t) = 0$ exactly when $0.0005t = 0.15$, or in other words,

$$t = \frac{0.15}{0.0005} = 300.$$

Hence the answer is 300 days for the farmer to wait before selling the bull, provided we can show that this value of t gives a maximum. But the formula for the profit is a quadratic expression in t, of the form

$$P(t) = at^2 + bt + c,$$

and $a < 0$. Hence $P(t)$ is a parabola, and since $a < 0$ this parabola opens downward as on the figure. Hence the critical point must be a maximum, as desired.

VI, §5. EXERCISES

1. Find the length of the sides of a rectangle of largest area which can be inscribed in a semicircle, the lower base being on the diameter.

2. A rectangular box has a square base and no top. The combined area of the sides and bottom is 48 ft^2. Find the dimensions of the box of maximum volume meeting these requirements.

3. Prove that, among all rectangles of given area, the square has the least perimeter.

4. A truck is to be driven 300 km at a constant speed of x km/hr. Speed laws require $30 \leqq x \leqq 60$. Assume that gasoline costs 30 cents/gallon and is consumed at the rate of $2 + x^2/600$ gal/hr. If the driver's wages are D dollars per hour, find the most economical speed and the cost of the trip if (a) $D = 0$, (b) $D = 1$, (c) $D = 2$, (d) $D = 3$, (e) $D = 4$.

5. A rectangle is to have an area of 64 m^2. Find its dimensions so that the distance from one corner to the mid-point of a non-adjacent side shall be a minimum.

6. Express the number 4 as the sum of two positive numbers in such a way that the sum of the square of the first and the cube of the second is as small as possible.

7. A wire 24 cm long is cut in two, and one part is bent into the shape of a circle, and the other into the shape of a square. How should it be cut if the sum of the areas of the circle and the square is to be (a) minimum, (b) maximum?

8. Find the point on the graph of the equation $y^2 = 4x$ which is nearest to the point $(2, 1)$.

9. Find the points on the hyperbola $x^2 - y^2 = 1$ nearest to the point $(0, 1)$.

10. Show that $(2, 2)$ is the point on the graph of the equation $y = x^3 - 3x$ that is nearest the point $(11, 1)$.

11. Find the coordinates of the points on the curve $x^2 - y^2 = 16$ which are nearest to the point $(0, 6)$.

12. Find the coordinates of the points on the curve $y^2 = x + 1$ which are nearest to the origin.

13. Find the coordinates of the point on the curve $y^2 = \frac{5}{2}(x + 1)$ which is nearest to the origin.

14. Find the coordinates of the points on the curve $y = 2x^2$ which are closest to the point $(9, 0)$.

15. A circular ring of radius b is uniformly charged with electricity, the total charge being Q. The force exerted by this charge on a particle at a distance x from the center of the ring, in a direction perpendicular to the plane of the ring, is given by $F(x) = Qx(x^2 + b^2)^{-3/2}$. Find the maximum of F for all $x \geqq 0$.

16. Let F be the rate of flow of water over a certain spillway. Assume that F is proportional to $y(h-y)^{1/2}$, where y is the depth of the flow, and h is the height, and is constant. What value of y makes F a maximum?

17. Find the point on the x-axis the sum of whose distances from (2, 0) and (0, 3) is a minimum.

18. A piece of wire of length L is cut into two parts, one of which is bent into the shape of an equilateral triangle and the other into the shape of a circle. How should the wire be cut so that the sum of the enclosed areas is:
(a) a minimum, (b) a maximum?

19. A fence $13\frac{1}{2}$ ft high is 4 ft from the side wall of a house. What is the length of the shortest ladder, one end of which will rest on the level ground outside the fence and the other on the side wall of the house?

20. A tank is to have a given volume V and is to be made in the form of a right circular cylinder with hemispheres attached to each end. The material for the ends costs twice as much per square meter as that for the sides. Find the most economical proportions. [You may assume that the area of a sphere is $4\pi r^2$.]

21. Find the length of the longest rod which can be carried horizontally around a corner from a corridor 8 ft wide into one 4 ft wide.

22. Let P, Q be two points in the plane on the same side of the x-axis. Let R be a point on the x-axis (Fig. 6). Show that the sum of the distances PR and QR is smallest when the angles θ_1 and θ_2 are equal.

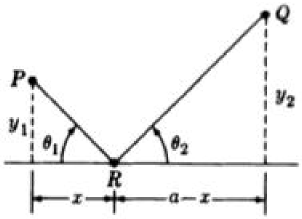


Figure 6

[*Hint*: First use the Pythagoras theorem to give an expression for the distances PR and RQ in terms of x and the fixed quantities y_1, y_2. Let $f(x)$ be the sum of the distances. Show that the condition $f'(x) = 0$ means that $\cos\theta_1 = \cos\theta_2$. Using values of x near 0 and a, show that $f(x)$ is decreasing near $x = 0$ and increasing near $x = a$. Hence the minimum must be in the open interval $0 < x < a$, and is therefore the critical point.]

23. Suppose the velocity of light is v_1 in air and v_2 in water. A ray of light traveling from a point P_1 above the surface of water to a point P_2 below the surface will travel by the path which requires the least time. Show that the ray will cross the surface at the point Q in the vertical plane through P_1 and P_2 so placed that

$$\frac{\sin\theta_1}{v_1} = \frac{\sin\theta_2}{v_2},$$

where θ_1 and θ_2 are the angles shown in the following figure:

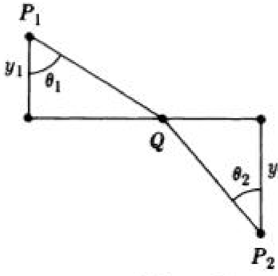


Figure 7

(You may assume that the light will travel in the vertical plane through P_1 and P_2. You may also assume that when the velocity is constant, equal to v, throughout a region, and s is the distance traveled, then the time t is equal to $t = s/v$.)

24. Let p be the probability that a certain event will occur, at any trial. In n trials, suppose that s successes have been observed. The likelihood function L is defined as $L(p) = p^s(1-p)^{n-s}$. Find the value of p which maximizes the likelihood function. (Take $0 \leqq p \leqq 1$.) View n, s as constants.

25. Find an equation for the line through the following points making with the coordinate axes a triangle of minimum area in the first quadrant:
    (a) through the point (3, 1).
    (b) through the point (3, 2).

26. Let $a_1, \ldots, a_n$ be numbers. Show that there is a single number x such that

$$(x-a_1)^2 + \cdots + (x-a_n)^2$$

is a minimum, and find this number.

27. When light from a point source strikes a plane surface, the intensity of illumination is proportional to the cosine of the angle of incidence and inversely proportional to the square of the distance from the source. How high should a light be located above the center of a circle of radius 25 cm to give the best illumination along the circumference? (The angle of incidence is measured from the perpendicular to the plane.)

28. A horizontal reservoir has a cross section which is an inverted isosceles triangle, where the length of a leg is 60 ft. Find the angle between the equal legs to give maximum capacity.

29. A reservoir has a horizontal plane bottom and a cross section as shown on the figure. Find the angle of inclination of the sides from the horizontal to give maximum capacity.

Figure 8

30. Determine the constant a such that the function

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

has (a) a local minimum at $x = 2$, (b) a local minimum at $x = -3$. (c) Show that the function cannot have a local maximum for any value of a.

31. The intensity of illumination at any point is proportional to the strength of the light source and varies inversely as the square of the distance from the source. If two sources of strengths a and b respectively are a distance c apart, at what point on the line joining them will the intensity be a minimum?

32. A window is in the shape of a rectangle surmounted by a semicircle. Find the dimensions when the perimeter is 12 ft and the area is as large as possible.

33. Find the radius and angle of the circular sector of maximum area if the perimeter is
(a) 20 cm (b) 16 cm.

34. You are watering the lawn and aiming the hose upward at an angle of inclination θ. Let r be the range of the hose, that is, the distance from the hose to the point of impact of the water. Then r is given by

$$r = \frac{2v^2}{g} \sin\theta \cos\theta,$$

where v, g are constants. For what angle is the range a maximum?

35. A ladder is to reach over a fence 12 ft high to a wall 2 ft behind the fence. What is the length of the shortest ladder that can be used?

36. A cylindrical tank is supposed to have a given volume V. Find the dimension of the radius of the base and the height in terms of V so that the surface area is minimal. The tank should be open on top, but closed at the bottom.

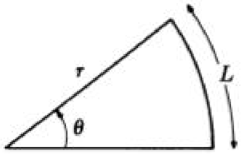
37. A flower bed is to have the shape of a circular sector of radius r and central angle θ. Find r and θ if the area is fixed and the perimeter is a minimum in case:
(a) $0 < \theta \leq \pi$ and (b) $0 < \theta \leq \pi/2$.

Recall that the area of a sector is

$$A = \pi r^2 \cdot \frac{\theta}{2\pi} = \theta r^2/2.$$

The length of an arc of a circle of radius r is

$$L = 2\pi r \cdot \frac{\theta}{2\pi} = r\theta.$$



38. A firm sells a product at $50 per unit. The total cost of marketing x units is given by the function

$$f(x) = 5000 + 650x - 45x^2 + x^3.$$

How many units should be produced per day to maximize profits? What is the daily profit for this number of units?

39. The daily cost of producing x units of a product is given by the formula

$$f(x) = 2002 + 120x - 5x^2 + \tfrac{1}{3}x^3.$$

Each unit sells for $264. How many units should be produced per day to maximize profits? What is the daily profit for this number of units?

40. A product is marketed at 50 dollars per unit. The total cost of marketing x units of the product is

$$f(x) = 1000 + 150x - 100x^2 + 2x^3.$$

How many units should be produced to maximize the profit? What is the daily profit for this x?

41. A company is the sole producer of a product, whose cost function is

$$f(x) = 100 + 20x + 2x^2.$$

If the company increases the price, then fewer units are sold, and in fact if we express the price $p(x)$ as a function of the number of units x, then

$$p(x) = 620 - 8x.$$

How many units should be produced to maximize profits? What is this maximum profit? [*Hint*: The total revenue is equal to the product $xp(x)$, number of units times the price.]

42. The cost of producing x units of a product is given by the function

$$f(x) = 10x^2 + 200x + 6{,}000.$$

If p is the price per unit, then the number of units sold at that price is given by

$$x = \frac{1000 - p}{10}.$$

For what value of x will the profit be positive? How many units should be produced to give maximum profits?

CHAPTER VII

Inverse Functions

Suppose that we have a function, for instance

$$y = 3x - 5.$$

Then we can solve for x in terms of y, namely

$$x = \tfrac{1}{3}(y + 5).$$

Thus x can be expressed as a function of y.

Although we are able to solve by means of an explicit formula, there are interesting cases where x can be expressed as a function of y, but without such an explicit formula. In this chapter, we shall investigate such cases.

VII, §1. DEFINITION OF INVERSE FUNCTIONS

Let $y = f(x)$ be a function, defined for all x in some interval. If, for each value y_1 of y, there exists exactly *one* value x_1 of x in the interval such that $f(x_1) = y_1$, then we can define the **inverse function**

$$x = g(y) = \textit{the unique number } x \textit{ such that } y = f(x).$$

Our inverse function is defined only at those numbers which are values of f. We have the fundamental relation

$$f(g(y)) = y \qquad \text{and} \qquad g(f(x)) = x.$$

Example 1. Consider the function $y = x^2$, which we view as being defined only for $x \geqq 0$. Every positive number (or 0) can be written uniquely as the square of a positive number (or 0). Hence we can define the inverse function, which will also be defined for $y \geqq 0$, but not for $y < 0$. It is the square root function, $x = \sqrt{y}$.

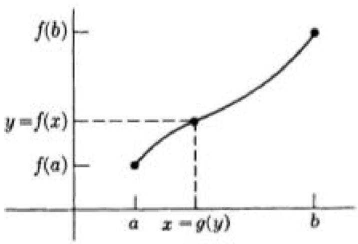
Example 2. Suppose that $y = 5x - 7$. Then we can solve for x in terms of y, namely

$$x = \tfrac{1}{5}(y + 7).$$

If $f(x) = 5x - 7$, then its inverse function is the function $g(y)$ such that

$$g(y) = \tfrac{1}{5}(y + 7).$$

In these examples, we could write down the inverse function by explicit formulas. In general, this is not possible, but there are criteria which tell us when the inverse function exists, for instance when the graph of the function f looks like this.



In this case, f is strictly increasing, and is defined on the interval $[a, b]$. To each point x in this interval, there is a value $f(x) = y$, and to each y between $f(a)$ and $f(b)$, there is a unique x between a and b such that $f(x) = y$. We formalize this in the next theorem.

Theorem 1.1. *Let $f(x)$ be a function which is strictly increasing. Then the inverse function exists, and is defined on the set of values of f.*

Proof. This is practically obvious: Given a number y_1 and a number x_1 such that $f(x_1) = y_1$, there cannot be another number x_2 such that $f(x_2) = y_1$ unless $x_2 = x_1$, because if $x_2 \neq x_1$, then

$$\begin{aligned} &\text{either} \quad x_2 > x_1, \text{ in which case } f(x_2) > f(x_1), \\ &\text{or} \quad x_2 < x_1, \text{ in which case } f(x_2) < f(x_1). \end{aligned}$$

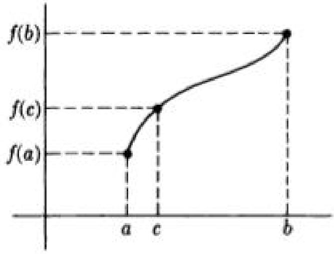
Since the positivity of the derivative gives us a good test when a function is strictly increasing, we are able to define inverse functions whenever the function is differentiable and its derivative is positive.

As usual, what we have said above applies as well to functions which are strictly decreasing, and whose derivatives are negative.

The following theorem is intuitively clear and is proved in an appendix. We already recalled it as Theorem 1.2 of Chapter V.

Intermediate value theorem. *Let f be a continuous function on a closed interval $[a, b]$. Let v be a number between $f(a)$ and $f(b)$. Then there exists a point c between a and b such that $f(c) = v$.*

This theorem says that the function f takes on every intermediate value between the values at the end points of the interval, and is illustrated on the next figure.



Using the intermediate value theorem, we now conclude:

Theorem 1.2. *Let f be a continuous function on the closed interval*

$$a \leqq x \leqq b$$

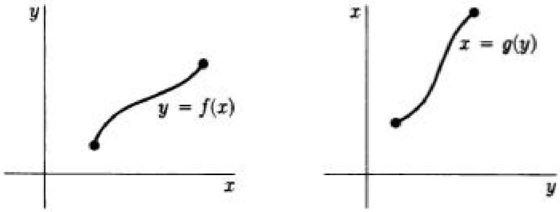
and assume that f is strictly increasing. Let $f(a) = \alpha$ and $f(b) = \beta$. Then the inverse function is defined on the closed interval $[\alpha, \beta]$.

Proof. Given any number γ between α and β, there exists a number c between a and b such that $f(c) = \gamma$, by the intermediate value theorem. Our assertion now follows from Theorem 1.1.

If we let g be this inverse function, then $g(\alpha) = a$ and $g(\beta) = b$. Furthermore, the inverse function is characterized by the relation

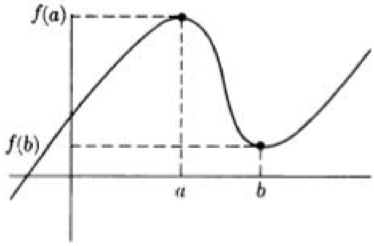
$$f(x) = y \qquad \text{if and only if} \qquad x = g(y).$$

Note that we can easily visualize the graph of an inverse function. If we want x in terms of y, we just flip the page over a 45° angle, reversing the roles of the x- and y-axes. Thus the graph of $y = f(x)$ gets reflected across the slanted line at 45° to give us the graph of $x = g(y)$.

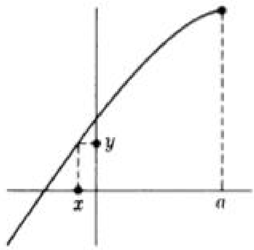


We shall now give some examples of how to define inverse functions over certain intervals.

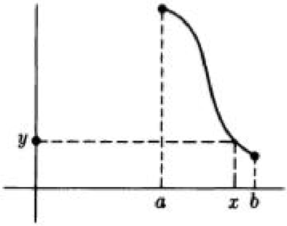
Example 3. Take for $f(x)$ a polynomial of degree 3. When the coefficient of x^3 is positive, and when f has local maxima and minima, its graph looks like this:

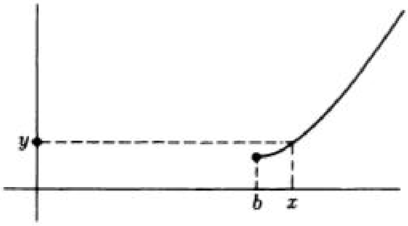


To any given value of y between $f(a)$ and $f(b)$ there correspond three possible values for x, and hence the inverse function cannot be defined unless we make other specifications. To do this, suppose first that we view f as defined only for those numbers $\leqq a$. Then the graph of f looks like this:



The inverse function is defined in this case. Similarly, we could view f as defined on the interval $[a, b]$, or on the interval of all $x \geqq b$. In each one of these cases, illustrated on the next figures, the inverse function would be defined.





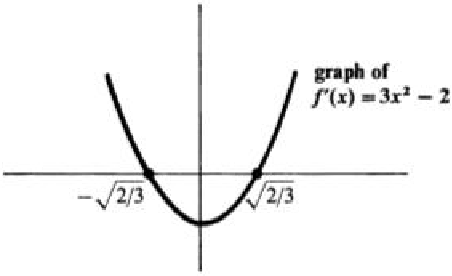
In each case, we have drawn a point y and the corresponding value x of the inverse function. They are different in the three cases.

Example 4. Let us consider a numerical example. Let

$$f(x) = x^3 - 2x + 1,$$

viewed as a function on the interval $x > \sqrt{2/3}$. Can we define the inverse function? For what numbers? If g is the inverse function, what is $g(0)$? What is $g(5)$?

We have $f'(x) = 3x^2 - 2$. The graph of $f'(x)$ is a parabola bending up, which crosses the x-axis at $x = \pm\sqrt{2/3}$.



Hence

$$f'(x) > 0 \quad \Leftrightarrow \quad x > \sqrt{2/3} \text{ and } x < -\sqrt{2/3}.$$

Consider the interval $x > \sqrt{2/3}$. Then f is strictly increasing on this interval, and so the inverse function g is defined. Since $f(x) \to \infty$ when $x \to \infty$, it follows that the inverse function $g(y)$ is defined for all $y > f(\sqrt{2/3})$, that is

$$y > f(\sqrt{2/3}) = (2/3)^{3/2} - 2(2/3)^{1/2} + 1.$$

Now $1 > \sqrt{2/3}$, so 1 lies in the interval $x > \sqrt{2/3}$, and $f(1) = 0$. Therefore $g(0) = 1$.

Similarly $f(2) = 5$ and 2 lies in the interval $x > \sqrt{2/3}$, so $g(5) = 2$.

Note that we do not give an explicit formula for our inverse function. When dealing with polynomials of degree $\geqq 3$, no single formula can be given.

Example 5. On the other hand, take f defined by the same formula,

$$f(x) = x^3 - 2x + 1,$$

but viewed as a function on the interval

$$-\sqrt{\tfrac{2}{3}} \leqq x \leqq \sqrt{\tfrac{2}{3}}.$$

The derivative of f is given by $f'(x) = 3x^2 - 2$. We have

$$f'(x) < 0 \quad \Leftrightarrow \quad -\sqrt{\tfrac{2}{3}} < x < \sqrt{\tfrac{2}{3}}.$$

Hence f is strictly decreasing on this interval, and the inverse function is defined, but is quite different from that of Example 4. For instance, 0 is in the interval, and $f(0) = 1$, so if h denotes the inverse function, we have $h(1) = 0$.

Example 6. Let $f(x) = x^n$ (n being a positive integer). We view f as defined only for numbers $x > 0$. Since $f'(x)$ is nx^{n-1}, the function is strictly increasing. Hence the inverse function exists. This inverse function g is in fact what we mean by the n-th root.

In all the exercises of the previous chapter you determined intervals over which certain functions increase and decrease. You can now define inverse functions for such intervals. In most cases, you cannot write down a simple explicit formula for such inverse functions.

VII, §1. EXERCISES

For each of the following functions, determine whether there is an inverse function g, and determine those numbers at which g is defined.

1. $f(x) = 3x + 2$, all x
2. $f(x) = x^2 + 2x - 3$, $0 \leqq x$
3. $f(x) = x^3 + 4x - 5$, all x
4. $f(x) = \dfrac{x}{x+1}$, $-1 < x$
5. $f(x) = \dfrac{x}{x+2}$, $-2 < x$
6. $f(x) = \dfrac{x+1}{x-1}$, $1 < x$
7. $f(x) = \dfrac{1}{x^2}$, $0 < x \leqq 1$
8. $f(x) = \dfrac{x^2}{x^2+1}$, $0 \leqq x \leqq 5$

9. $f(x) = \dfrac{x+2}{x-2}$, $0 \leqq x < 2$

10. $f(x) = x + \dfrac{1}{x}$. $1 \leqq x \leqq 10$

11. $f(x) = x + \dfrac{1}{x}$, $0 < x \leqq 1$

12. $f(x) = x - \dfrac{1}{x}$, $0 < x \leqq 1$

13. $f(x) = \dfrac{2x}{1+x^2}$, $-1 \leqq x \leqq 1$

14. $f(x) = \dfrac{2x}{1+x^2}$, $1 \leqq x$

VII, §2. DERIVATIVE OF INVERSE FUNCTIONS

We shall state a theorem which allows us to determine the derivative of an inverse function when we know the derivative of the given function.

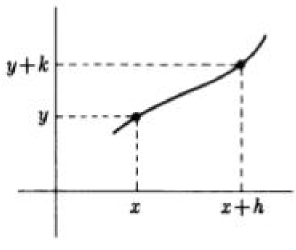
Theorem 2.1. *Let a, b be two numbers, $a < b$. Let f be a function which is differentiable on the interval $a < x < b$ and such that its derivative $f'(x)$ is > 0 for all x in this open interval. Then the inverse function $x = g(y)$ exists, and we have*

$$\boxed{g'(y) = \frac{1}{f'(x)} = \frac{1}{f'(g(y))}.}$$

Proof. We are supposed to investigate the Newton quotient

$$\frac{g(y+k) - g(y)}{k}.$$

The following picture illustrates the situation:



By the intermediate value theorem, every number of the form $y + k$ with small values of k can be written as a value of f. We let $x = g(y)$ and we let $h = g(y + k) - g(y)$. Then

$$x = g(y) \qquad \text{and} \qquad g(y+k) = x + h.$$

Furthermore, $y + k = f(x + h)$ and hence

$$k = f(x + h) - f(x).$$

The Newton quotient for g can therefore be written

$$\frac{g(y + k) - g(y)}{k} = \frac{x + h - x}{f(x + h) - f(x)} = \frac{h}{f(x + h) - f(x)},$$

and we see that it is the reciprocal of the Newton quotient for f, namely

$$\frac{1}{\dfrac{f(x + h) - f(x)}{h}}.$$

As h approaches 0, we know that k approaches 0 since

$$k = f(x + h) - f(x).$$

Conversely, as k approaches 0, we know that there exists exactly one value of h such that $f(x + h) = y + k$, because the inverse function is defined. Consequently, the corresponding value of h must also approach 0.

If we now take the limit of the reciprocal of the Newton quotient of f, as h (or k) approaches 0, we get

$$\frac{1}{f'(x)}.$$

By definition, this is the derivative $g'(y)$ and our theorem is proved.

Example 1. Let $f(x) = x^3 - 2x + 1$. Find an interval such that the inverse function g of f is defined, and find $g'(0)$, $g'(5)$.

By inspection, we see that

$$f(1) = 0 \quad \text{and} \quad f(2) = 5.$$

We must therefore find an interval containing 1 and 2 such that the inverse function of f is defined for that interval. But

$$f'(x) = 3x^2 - 2$$

and $f'(x) > 0$ if and only if $x > \sqrt{2/3}$ or $x < -\sqrt{2/3}$. (See Example 4 of the preceding section.) We select the interval $x > \sqrt{2/3}$, which contains both 1 and 2. Then we can apply the general theorem on the derivative of the inverse function, which states that if $y = f(x)$ then

$$g'(y) = \frac{1}{f'(x)}.$$

This gives:

$$g'(0) = \frac{1}{f'(1)} = 1 \qquad \text{and} \qquad g'(5) = \frac{1}{f'(2)} = \frac{1}{10}.$$

Please note that the derivative $g'(y)$ is given in terms of $f'(x)$. We don't have a formula in terms of y.

The theorem giving us the derivative of the inverse function could also be expressed by saying that

$$\boxed{\frac{dx}{dy} = \frac{1}{dy/dx}.}$$

Here also, the derivative behaves *as if* we were taking a quotient. Thus the notation is very suggestive and we can use it from now on without thinking, because we proved a theorem justifying it.

Remark. In Theorem 2.1, we have proved that in fact, the derivative of the inverse function g exists, and is given by $g'(y) = 1/f'(x)$. If one *assumes* that this derivative exists, then one can give a much shorter argument to find its value, using the chain rule. Indeed, we have

$$f(g(y)) = y, \qquad \text{since } g(y) = x.$$

Differentiating with respect to y, we find by the chain rule,

$$f'(g(y))g'(y) = 1, \qquad \text{because } \frac{dy}{dy} = 1.$$

Hence

$$g'(y) = \frac{1}{f'(g(y))},$$

as was to be shown.

VII, §2. EXERCISES

In each one of the exercises from 1 through 10, restrict f to an interval so that the inverse function g is defined in an interval containing the indicated point, and find the derivative of the inverse function at the indicated point.

0. $f(x) = -x^3 + 2x + 1$. Find $g'(2)$.

1. $f(x) = x^3 + 1$. Find $g'(2)$.
2. $f(x) = (x-1)(x-2)(x-3)$. Find $g'(6)$.
3. $f(x) = x^2 - x + 5$. Find $g'(7)$.
4. $f(x) = \sin x + \cos x$. Find $g'(-1)$.
5. $f(x) = \sin 2x$ $(0 \leqq x \leqq 2\pi)$. Find $g'(\sqrt{3}/2)$.
6. $f(x) = x^4 - 3x^2 + 1$. Find $g'(-1)$.
7. $f(x) = x^3 + x - 2$. Find $g'(0)$.
8. $f(x) = -x^3 + 2x + 1$. Find $g'(2)$.
9. $f(x) = 2x^3 + 5$. Find $g'(21)$.
10. $f(x) = 5x^2 + 1$. Find $g'(11)$.
11. Let f be a continuous function on the interval $[a, b]$. Assume that f is twice differentiable on the open interval $a < x < b$, and that $f'(x) > 0$ and $f''(x) > 0$ on this interval. Let g be the inverse function of f.
    (a) Find an expression for the second derivative of g.
    (b) Show that $g''(y) < 0$ on its interval of definition. Thus g bends in the opposite direction to f.

VII, §3. THE ARCSINE

It is impossible to define an inverse function for the function $y = \sin x$ because to each value of y there correspond infinitely many values of x because $\sin(x + 2\pi) = \sin x = \sin(\pi - x)$. However, if we restrict our attention to special intervals, we can define the inverse function.

We restrict the sine function to the interval

$$-\frac{\pi}{2} \leqq x \leqq \frac{\pi}{2}.$$

The derivative of $\sin x$ is $\cos x$ and in that interval, we have

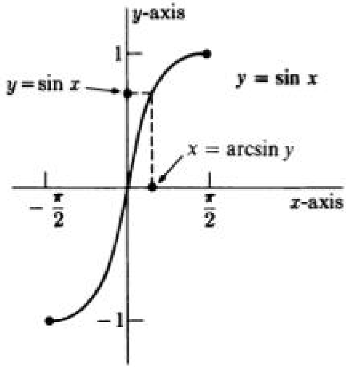
$$0 < \cos x, \qquad \text{so the derivative is positive,}$$

except when $x = \pi/2$ or $x = -\pi/2$ in which case the cosine is 0.

Therefore, in the interval

$$-\frac{\pi}{2} \leqq x \leqq \frac{\pi}{2}$$

the function is strictly increasing. The inverse function exists, and is called the **arcsine**.



Let $f(x) = \sin x$, and $x = \arcsin y$, the inverse function. Since $f(0) = 0$ we have $\arcsin 0 = 0$. Furthermore, since

$$\sin(-\pi/2) = -1 \qquad \text{and} \qquad \sin(\pi/2) = 1,$$

we know that the inverse function is defined over the interval going from -1 to $+1$, that is for

$$-1 \leqq y \leqq 1.$$

In words, we can say loosely that $\arcsin x$ *is the angle whose sine is* x. (We throw in the word *loosely* because, strictly speaking, $\arcsin x$ is a number, and not an angle, and also because we mean the angle between $-\pi/2$ and $\pi/2$.)

Example 1. Let $f(x) = \sin x$ and let $g(y) = \arcsin y$. Then

$$\arcsin(1/\sqrt{2}) = \pi/4$$

because

$$\sin \pi/4 = 1/\sqrt{2}.$$

Similarly,

$$\arcsin 1/2 = \pi/6$$

because

$$\sin \pi/6 = 1/2.$$

For any value of x *in the interval* $-\pi/2 \leqq x \leqq \pi/2$ *we have*

$$\boxed{\arcsin \sin x = x,}$$

by definition of the inverse function. However, if x is not in this interval, then we do **not** have

$$\arcsin \sin x = x.$$

Example 2. Let $x = -\pi$. Then

$$\sin(-\pi) = 0,$$

and

$$\arcsin(\sin(-\pi)) = \arcsin 0 = 0 \neq -\pi.$$

Example 3. We have

$$\arcsin \sin(3\pi/4) = \pi/4,$$

because $\sin 3\pi/4 = 1/\sqrt{2}$, and $\arcsin 1/\sqrt{2} = \pi/4$.

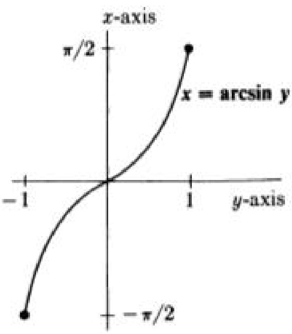
We now consider the derivative and the graph of the inverse function. The derivative of $\sin x$ is positive in the interval

$$-\pi/2 < x < \pi/2.$$

Since the derivative of the inverse function $x = g(y)$ is $1/f'(x)$, the derivative of $\arcsin y$ is also positive, in the interval

$$-1 < y < 1.$$

Therefore the inverse function is strictly increasing in that interval. Its graph looks like the figure shown below.



According to the general rule for the derivative of inverse functions, we know that when $y = \sin x$ and $x = \arcsin y$ the derivative is

$$\frac{dx}{dy} = \frac{1}{dy/dx} = \frac{1}{\cos x}.$$

When x is very close to $\pi/2$, we know that $\cos x$ is close to 0. Therefore the derivative is very large. Hence the curve is almost vertical. Similarly, when x is close to $-\pi/2$ and y is close to -1, the curve is almost vertical, as drawn.

Finally, it turns out that we can express our derivative explicitly as a function of y. Indeed, we have the relation

$$\sin^2 x + \cos^2 x = 1,$$

whence

$$\cos^2 x = 1 - \sin^2 x.$$

In the interval between $-\pi/2$ and $\pi/2$, the cosine is $\geqq 0$. Hence we can take the square root, and we get

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

in that interval. Since $y = \sin x$, we can write our derivative in the form

$$\boxed{\frac{dx}{dy} = \frac{1}{\sqrt{1 - y^2}}}$$

which is expressed entirely in terms of y.

Example 4. Let $x = \arcsin y$. Find dx/dy when $y = 1/\sqrt{2}$. This is easily done, namely if $g(y) = \arcsin y$, then

$$g'(1/\sqrt{2}) = \frac{1}{\sqrt{1 - (1/\sqrt{2})^2}} = \sqrt{2}.$$

Having now obtained all the information we want concerning the arcsine, we shift back our letters to the usual ones. We state the main properties as a theorem.

Theorem 3.1. *View the sine function as defined on the interval*

$$[-\pi/2, \pi/2].$$

Then the inverse function is defined on the interval $[-1, 1]$. *Call it*

$$g(x) = \arcsin x.$$

That is, $\arcsin x$ *is the unique number* y *such that* $\sin(y) = x$ *and* $-\pi/2 \leqq y \leqq \pi/2$. *Then* g *is differentiable in the open interval* $-1 < x < 1$, *and*

$$g'(x) = \frac{1}{\sqrt{1 - x^2}}.$$

VII, §3. EXERCISES

1. Viewing the cosine as defined only on the interval $[0, \pi]$, prove that the inverse function arccos exists. On what interval is it defined? Sketch the graph.
2. What is the derivative of arccosine?
3. Let $g(x) = \arcsin x$. Find the following values:

   (a) $g'(1/2)$ (b) $g'(1/\sqrt{2})$ (c) $g(1/2)$
   (d) $g(1/\sqrt{2})$ (e) $g'(\sqrt{3}/2)$ (f) $g(\sqrt{3}/2)$

4. Let $g(x) = \arccos x$. What is $g'(\frac{1}{2})$? What is $g'(1/\sqrt{2})$? What is $g(\frac{1}{2})$? What is $g(1/\sqrt{2})$?
5. Let $\sec x = 1/\cos x$. Define the inverse function of the secant over a suitable interval and obtain a formula for the derivative of this inverse function.

Find the following numbers.

6. $\arcsin(\sin 3\pi/2)$
7. $\arcsin(\sin 2\pi)$
8. $\arccos(\cos 3\pi/2)$
9. $\arccos(\cos -\pi/2)$
10. $\arcsin(\sin -3\pi/4)$

Find the derivatives of the following functions.

11. $\arcsin(x^2 - 1)$
12. $\arccos(2x + 5)$
13. $\dfrac{1}{\arcsin x}$
14. $\dfrac{2}{\arccos 2x}$
15. Determine the intervals over which the function arcsin is bending upward, and bending downward.

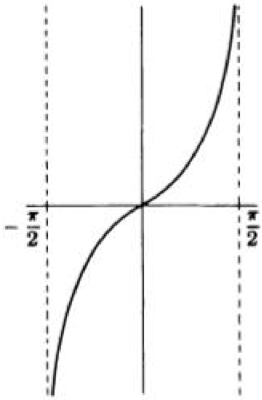
VII, §4. THE ARCTANGENT

Let $f(x) = \tan x$ and view this function as defined over the interval

$$-\frac{\pi}{2} < x < \frac{\pi}{2}.$$

As x goes from $-\pi/2$ to $\pi/2$, the tangent goes from very large negative values to very large positive values. As x approaches $\pi/2$, the tangent has in fact arbitrarily large positive values, and similarly when x approaches $-\pi/2$, the tangent has arbitrarily large negative values.

We recall that the **graph of the tangent** looks like that in the following figure.



The derivative of $\tan x$ is

$$\frac{d(\tan x)}{dx} = 1 + \tan^2 x.$$

Hence the derivative is always positive, and the tan function is strictly increasing. Furthermore when x ranges over the interval

$$-\pi/2 < x < \pi/2,$$

$\tan x$ ranges from large negative to large positive values. Hence $\tan x$ ranges over all numbers. Therefore the inverse function is defined for all numbers. We call it the **arctangent**.

As with the arcsin and arccos, we may say roughly that $\arctan y$ is the angle whose tangent is y. We put "roughly" in our statement, because as pointed out before, we really mean the number of radians of the angle whose tangent is y, such that this number lies between $-\pi/2$ and $\pi/2$.

Example 1. We have $\arctan(-1/\sqrt{3}) = -\pi/6$, but

$$\arctan(-1/\sqrt{3}) \neq 5\pi/6,$$

even though $\tan 5\pi/6 = -1/\sqrt{3}$.

Example 2. In the same vein, we find that

$$\arctan \tan 5\pi/6 = \arctan(-1/\sqrt{3}) = -\pi/6.$$

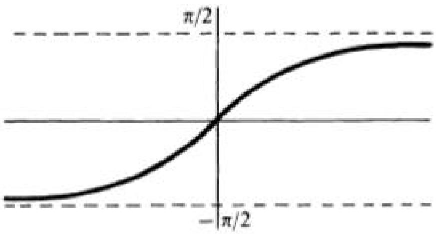
The reason why $\arctan \tan x \neq x$ in this case is due to our choice of interval of definition for the tangent, when we wish to have an inverse function, and the fact that $x = 5\pi/6$ does not lie in this interval. Of course, if x lies between $-\pi/2$ and $\pi/2$, then we must have

$$\arctan \tan x = x$$

Thus

$$\arctan \tan(-\pi/6) = -\pi/6.$$

The **graph of the arctan** looks like this:



Now for the derivative. Let $x = g(y) = \arctan y$. Then

$$g'(y) = \frac{1}{f'(x)} = \frac{1}{1 + \tan^2 x}$$

so that

$$g'(y) = \frac{1}{1 + y^2}.$$

Here again we are able to get an explicit formula for the derivative of the inverse function.

As with the arcsine, when dealing simultaneously with the function and its inverse function, we have to keep our letters x, y separate. However, we now summarize the properties of the arctan in terms of our usual notation.

Theorem 4.1. *The inverse function of the tangent is defined for all numbers. Call it the arctangent. Thus* $\arctan x$ *is the unique number* y *such that* $\tan y = x$ *and* $-\pi/2 < y < \pi/2$. *Then arctan has a derivative, and that derivative is given by the relation*

$$\frac{d(\arctan x)}{dx} = \frac{1}{1+x^2}.$$

As x *becomes very large positive,* $\arctan x$ *approaches* $\pi/2$.
As x *becomes very large negative,* $\arctan x$ *approaches* $-\pi/2$.
The arctangent is strictly increasing for all x.

Example 3. Let $h(x) = \arctan 2x$. To find the derivative, we use the chain rule, letting $u = 2x$. Then

$$h'(x) = \frac{1}{1+(2x)^2} \cdot 2 = \frac{2}{1+4x^2}.$$

Example 4. Let g be the arctan function. Then

$$g'(5) = \frac{1}{1+5^2} = \frac{1}{26}.$$

Example 5. Find the equation of the tangent line to the curve

$$y = \arctan 2x$$

at the point $x = 1/(2\sqrt{3})$.

Let $h(x) = \arctan 2x$. Then

$$h'\left(\frac{1}{2\sqrt{3}}\right) = \frac{2}{1+4\left(\dfrac{1}{2\sqrt{3}}\right)^2} = \frac{3}{2}.$$

When $x = 1/2\sqrt{3}$ we have

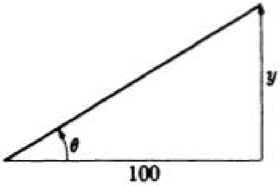
$$y = \arctan\left(2 \cdot \frac{1}{2\sqrt{3}}\right) = \arctan \frac{1}{\sqrt{3}} = \pi/6.$$

Hence we must find the equation of the line with slope $3/2$, passing through the point $(1/2\sqrt{3}, \pi/6)$. We know how to do this; the equation is

$$y - \frac{\pi}{6} = \frac{3}{2}\left(x - \frac{1}{2\sqrt{3}}\right).$$

Example 6. A balloon leaves the ground 100 m from an observer at the rate of 50 m/min. How fast is the angle of elevation of the observer's line of sight increasing when the balloon is at an altitude of 100 m?

The figure is as follows:



We have to determine $d\theta/dt$. We know that $dy/dt = 50$. We have

$$\frac{y}{100} = \tan\theta, \qquad \text{whence} \qquad \theta = \arctan\left(\frac{y}{100}\right).$$

Since

$$\frac{d\theta}{dt} = \frac{d\theta}{dy}\frac{dy}{dt}$$

we get:

$$\frac{d\theta}{dt} = \frac{1}{1 + \left(\frac{y}{100}\right)^2}\frac{1}{100}50.$$

Hence

$$\left.\frac{d\theta}{dt}\right|_{y=100} = \frac{1}{1+1^2}\frac{1}{100}50 = \frac{1}{4}\ \text{rad/min}.$$

VII, §4. EXERCISES

1. Let g be the arctan function. What is $g(1)$? What is $g(1/\sqrt{3})$? What is $g(-1)$? What is $g(\sqrt{3})$?

2. Let g be the arctan function. What is $g'(1)$? What is $g'(1/\sqrt{3})$? What is $g'(-1)$? What is $g'(\sqrt{3})$?

3. Suppose you were to define an inverse function for the tangent in the interval $\pi/2 < x < 3\pi/2$. What would be the derivative of this inverse function?

4. What is
   (a) $\arctan(\tan 3\pi/4)$? (b) $\arctan(\tan 2\pi)$?
   (c) $\arctan(\tan 5\pi/6)$? (d) $\arctan(\tan(-5\pi/6))$?

Find the derivatives of the following functions.

5. $\arctan 3x$ 6. $\arctan \sqrt{x}$ 7. $\arcsin x + \arccos x$

8. $x \arcsin x$
9. $\arctan(\sin 2x)$
10. $x^2 \arctan 2x$
11. $\dfrac{\sin x}{\arcsin x}$
12. $\arcsin(\cos x - x^2)$
13. $\arctan \dfrac{1}{x}$
14. $\arctan \dfrac{1}{2x}$
15. $(1 + \arcsin 3x)^3$
16. $(\arcsin 2x + \arctan x^2)^{3/2}$

Find the equation of the tangent line at the indicated point for the following curves.

17. $y = \arcsin x, \quad x = 1/\sqrt{2}$
18. $y = \arccos x, \quad x = 1/\sqrt{2}$
19. $y = \arctan 2x, \quad x = \sqrt{3}/2$
20. $y = \arctan x, \ x = -1$
21. $y = \arcsin x, \quad x = -\frac{1}{2}$
22. A balloon leaves the ground 500 ft from an observer at the rate of 200 ft/min. How fast is the angle of elevation of the observer's line of sight increasing when the balloon is at an altitude of 1000 ft?
23. An airplane at an altitude of 4400 ft is flying horizontally directly away from an observer. When the angle of elevation is $\pi/4$, the angle is decreasing at the rate of 0.05 rad/sec. How fast is the airplane flying at that instant?
24. A man is walking along a sidewalk at the rate of 5 ft/sec. A searchlight on the ground 30 ft from the walk is kept trained on him. At what rate is the searchlight revolving when the man is 20 ft from the point on the sidewalk nearest the light?
25. A tower stands at the end of a street. A man drives toward the tower at the rate of 50 ft/sec. The tower is 500 ft tall. How fast is the angle subtended by the tower at the man's eye increasing when the man is 1000 ft from the tower?
26. A police car approaches an intersection at 80 ft/sec. When it is 200 ft from the intersection, a car crosses the intersection traveling at a right angle from the police car at the rate of 60 ft/sec. If the policeman directs his beam of light on this second car, how fast is the light beam turning 2 sec later, assuming that both cars continue at their original rate?
27. A weight is drawn along a level floor by means of a rope which passes over a hook 6 ft above the floor. If the rope is pulled over the hook at the rate of 1 ft/sec find an expression for the rate of change of the angle θ between the rope and the floor as a function of the angle θ.
28. A man standing at a fixed point on a wharf pulls in a small boat. The wharf is 20 ft above the level of the water. If he is pulling the rope at 2 ft/sec, how fast is the angle that the rope makes with the water increasing when the distance from the man to the boat is 50 ft?
29. A helicopter leaves the ground 1000 ft from an observer and rises vertically at 20 ft/sec. At what rate is the observer's angle of elevation of the helicopter changing when the helicopter is 800 ft above the ground?

30. Determine those intervals where arctan is bending upward and bending downward.

31. A helicopter leaves a base, rising straight up at a speed of 15 ft/sec. At the same time that the helicopter leaves, an observer starts from a point 100 ft away from the base, and moves on a straight line away from the base at a speed of 80 ft/sec. How fast is the angle of elevation from the observer to the helicopter increasing when the observer is (a) 400 ft from the base? (b) 600 ft from the base?

32. A train is moving on a straight line away from the station at a speed of 20 ft/sec. A cameraman starts from a point 50 ft away from the station at the same time that the train leaves, and, directing the camera toward the train, moves away from the station perpendicularly to the line made by the tracks, at a speed of 10 ft/sec. At what rate is the angle of the camera turning after the train has moved (a) 80 ft? (b) 100 ft?

33. A car is moving on a straight line toward the point where a rocket is being launched. The car is traveling at 50 ft/sec. When the car is 300 ft from the launching site, the rocket starts going up, and its height is given as function of time by $y = t^3$ ft. A person in the car is photographing the rocket. How fast is the angle of elevation of the camera turning 5 seconds after the rocket has started?

CHAPTER VIII

Exponents and Logarithms

We remember that we had trouble at the very beginning with the function 2^x (or 3^x, or 10^x). It was intuitively very plausible that there should be such functions, satisfying the fundamental equation

$$2^{x+y} = 2^x 2^y$$

for all numbers x, y, and $2^0 = 1$, but we had difficulties in saying what we meant by $2^{\sqrt{2}}$ (or 2^π).

It is the purpose of this chapter to study this function, and others like it.

VIII, §1. THE EXPONENTIAL FUNCTION

If n is a positive integer, we know what 2^n means: It is the product of 2 with itself n times. For instance, 2^8 is the product of 2 with itself eight times.

Furthermore, we also know that $2^{1/n}$ is the n-th root of 2; it is that number whose n-th power is 2. Thus $2^{1/8}$ is that number whose 8-th power is 2.

If $x = m/n$ is a quotient of two positive integers, then

$$2^{m/n} = (2^{1/n})^m = (2^m)^{1/n}$$

can be expressed in terms of roots and powers, so fractional powers of 2 are also easily understood. The problem arises in understanding 2^x when

x is not a quotient of two positive integers. We leave this problem aside for the moment, and assume that there is a function defined for all x, denoted by 2^x, which is differentiable. We shall now see how to find its derivative.

We form the Newton quotient. It is

$$\frac{2^{x+h} - 2^x}{h}.$$

Using the fundamental equation we see that this quotient is equal to

$$\frac{2^x 2^h - 2^x}{h} = 2^x \frac{2^h - 1}{h}.$$

As h approaches 0, 2^x remains fixed, but it is very difficult to see what happens to

$$\frac{2^h - 1}{h}.$$

It is not at all clear that this quotient approaches a limit. Roughly speaking, we meet a difficulty which is analogous to the one we met when we tried to find the derivative of $\sin x$. However, in the present situation, a direct approach would lead to much greater difficulties than those which we met when we discussed

$$\lim_{h \to 0} \frac{\sin h}{h}.$$

It is, in fact, true that

$$\lim_{h \to 0} \frac{2^h - 1}{h}$$

exists. We see that it does not depend on x. It depends only on 2.

If we tried to take the derivative of 10^x, we would end up with the problem of determining the limit

$$\lim_{h \to 0} \frac{10^h - 1}{h},$$

which is also independent of x.

In general, we shall assume the following.

Let a be a number >1. *There exists a function a^x, defined for all numbers x, satisfying the following properties:*

Property 1. The fundamental equation

$$a^{x+y} = a^x a^y$$

holds for all numbers x, y.

Property 2. If x is a rational number, $x = m/n$ with m, n positive integers, then $a^{m/n}$ has the usual meaning:

$$a^{m/n} = (a^{1/n})^m = (a^m)^{1/n}.$$

Property 3. The function a^x is differentiable.

The function a^x is called an **exponential function**.

We can then apply the same procedure to a^x that we applied to 2^x. We form the Newton quotient

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

Since we assumed that a^x is differentiable, it follows that

$$\boxed{\frac{da^x}{dx} = \lim_{h\to 0} \frac{a^{x+h} - a^x}{h} = a^x \lim_{h\to 0} \frac{a^h - 1}{h}.}$$

Thus we meet the mysterious limit

$$\lim_{h\to 0} \frac{a^h - 1}{h}.$$

This is a similar situation to the differentiation of $\sin x$, but previously we were able to find the limit of $(\sin h)/h$ as h approaches 0. Here we cannot take a direct approach. The limit will be clarified later when we study the log.

However, we can analyze this limit a little more. Let

$$f(x) = a^x.$$

We claim that

$$\boxed{a^0 = 1.}$$

This is because

$$a = a^{1+0} = a \cdot a^0.$$

If we multiply by a^{-1} on both sides, we get $1 = a^0$.

Similarly, we find

$$\boxed{a^{-x} = \frac{1}{a^x},}$$

because

$$1 = a^0 = a^{x-x} = a^x a^{-x}.$$

Now if we put $x = 0$ in the formula

$$f'(x) = a^x \lim_{h \to 0} \frac{a^h - 1}{h},$$

then we find that

$$\boxed{f'(0) = \lim_{h \to 0} \frac{a^h - 1}{h}}$$

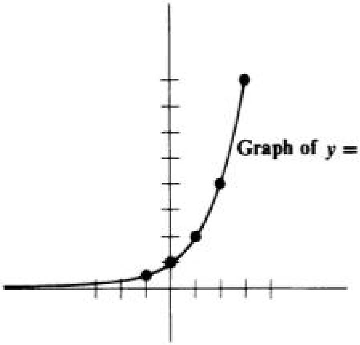
because $a^0 = 1$. *Consequently the mysterious limit on the right-hand side is the slope of the curve $y = a^x$ at $x = 0$.*

Let us try to get a feeling for curves like 2^x, or 3^x, or 10^x by plotting points. We give a table of values for 2^x.

x	2^x
1	2
2	4
3	8
4	16
5	32
10	1024
20	1048576

x	2^x
-1	1/2
-2	1/4
-3	1/8
-4	1/16
-5	1/32
-10	1/1024
-20	1/1048576

We see that the value $y = 2^x$ increases rapidly when x becomes large, and approaches 0 rapidly when x becomes large negative, as illustrated on the figure.



The behavior when x becomes large negative is due to the relation $2^{-x} = 1/2^x$. For instance, $2^{-10} = 1/2^{10}$ which is small. We can write

$$\lim_{x \to -\infty} 2^x = 0.$$

Observe that $2^x > 0$ for all numbers x. Similarly, if $a > 0$, then

$$\boxed{a^x > 0 \qquad \text{for all } x.}$$

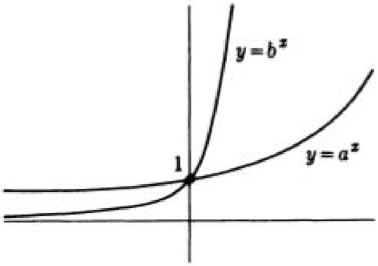
We can even prove this from what we have assumed explicitly. For suppose $a^c = 0$ with some number c. Then for all x we get

$$a^x = a^{x-c+c} = a^{x-c}a^c = 0,$$

which is not true since $a^1 = a \neq 0$.

Exercise. Make a similar table for 3^x, 10^x, and $(3/2)^x$.

Next suppose that $1 < a < b$. It is plausible that the curve b^x has a bigger slope than the curve a^x. We are especially interested in the slope when $x = 0$. If b is very large, then the curve $y = b^x$ will have a very steep slope at $x = 0$. If a is > 1 but close to 1, then the curve $y = a^x$ will have a small slope at $x = 0$. We have drawn these curves on the next figure.



Try for yourself plotting some points on the curves 2^x, 3^x, 10^x to see what happens in these concrete cases. It is plausible that as a increases fron numbers close to 1 (and > 1) to very large numbers, the slope of a^x at $x = 0$ increases continuously from values close to 0 to large values, and therefore for some value of a, which we call e, this slope is precisely equal to 1. Thus in this naive approach, e is the number such that the slope of e^x at $x = 0$ is equal to 1, that is for $f(x) = e^x$, we have

$$\boxed{\lim_{h\to 0} \frac{f(0 + h) - f(0)}{h} = \lim_{h\to 0} \frac{e^h - e^0}{h} = 1.}$$

So in addition to the three properties stated previously, we assume:

Property 4. There is a number $e > 1$ such that

$$\frac{de^x}{dx} = e^x,$$

or equivalently,

$$\lim_{h\to 0} \frac{e^h - 1}{h} = 1.$$

This number e is called the **natural base for exponential functions**.

Warning. Do not confuse the functions 2^x and x^2. The derivative of x^2 is $2x$. The derivative of 2^x is

$$\frac{d(2^x)}{dx} = 2^x \lim_{h\to 0} \frac{2^h - 1}{h}.$$

Similarly, when a is a fixed number, do not confuse the function a^x and x^a, where x is the variable.

At first we have no idea how big or small the number e may be. In Exercises 16 through 20, you will learn a very efficient way of finding a decimal expansion, or approximations for e by rational numbers. It turns out that e lies between 2 and 3, and in particular is approximately equal to $2.7183\ldots$.

Assuming the basic properties of e^x as we have done, we can apply some of our previous techniques in the context of this exponential function. First we show that e^x is the only function equal to its own derivative, up to a constant factor.

Theorem 1.1. *Let $g(x)$ be a function defined for all numbers and such that $g'(x) = g(x)$. Then there is a constant C such that $g(x) = Ce^x$.*

Proof. We have to prove that $g(x)/e^x$ is constant. We know how to do this. It suffices to prove that the derivative is 0. But we find:

$$\begin{aligned}\frac{d}{dx}\left(\frac{g(x)}{e^x}\right) &= \frac{e^x g'(x) - g(x)e^x}{e^{2x}}\\ &= \frac{e^x g(x) - g(x)e^x}{e^{2x}}\\ &= 0.\end{aligned}$$

Hence there is a constant C such that $g(x)/e^x = C$. Multiplying both sides by e^x we get

$$g(x) = Ce^x,$$

thus proving the theorem.

As a special case of the theorem we have:

Let g be a differentiable function such that $g'(x) = g(x)$ and $g(0) = 1$. Then $g(x) = e^x$.

Proof. Since $g(x) = Ce^x$ we get $g(0) = Ce^0 = C$. Hence $C = 1$ and $g(x) = e^x$.

Thus there is one and only one function g which is equal to its own derivative and such that $g(0) = 1$. This function is called **the exponential function**, and is sometimes denoted by **exp**. We may write

$$\exp'(x) = \exp(x) \qquad \text{and} \qquad \exp(0) = 1.$$

But usually we use the notation e^x as before, instead of $\exp(x)$.

There are several ways of *proving* the existence of a function $g(x)$ such that $g'(x) = g(x)$ and $g(0) = 1$, rather than giving the plausibility arguments as above.

In Chapter XIV we shall give a proof by infinite series. On the other hand, when we study the logarithm in §6, we shall first show that there exists a function $L(x)$ such that $L'(x) = 1/x$ and $L(1) = 0$. Then we can define the inverse function, and it is easy to see that this inverse function g satisfies $g'(y) = g(y)$ and $g(0) = 1$. Anyone interested in such theory can suit their tastes and look up these later sections as they see fit.

We now give examples and applications involving the function e^x.

Example. Find the derivative of e^{3x^2}.

We use the chain rule, with $u = 3x^2$. Then

$$\frac{d(e^u)}{dx} = \frac{de^u}{du}\frac{du}{dx} = e^{3x^2} \cdot 6x.$$

Example. Let $f(x) = e^{\cos 2x}$. We find the derivative of f by the chain rule, namely

$$f'(x) = e^{\cos 2x}(-\sin 2x)2.$$

There is no point simplifying this expression.

Example. Find the equation of the tangent line to the curve $y = e^x$ at $x = 2$.

Let $f(x) = e^x$. Then $f'(x) = e^x$ and $f'(2) = e^2$. When $x = 2$, $y = e^2$. Hence we must find the equation of the line with slope e^2, passing through the point $(2, e^2)$. This equation is

$$y - e^2 = e^2(x - 2).$$

Graph of e^x

Let us sketch the graph of e^x. We justify our statements by using only the four properties listed above. Since

$$\frac{de^x}{dx} = e^x > 0 \qquad \text{for all } x,$$

we conclude that the function $f(x) = e^x$ is strictly increasing. Since

$$f''(x) = f'(x) = f(x) > 0 \qquad \text{for all } x,$$

we conclude that the function is bending up.

Since $f(0) = 1$ and the function is strictly increasing, we conclude that

$$f(1) = e > 1.$$

Hence when n is a positive integer, $n = 1, 2, 3, \ldots$, the powers e^n become large as n becomes large. Since e^x is strictly increasing for all x, this also shows that e^x becomes large when x is a large real number.

We had also seen that

$$e^{-x} = (e^x)^{-1}.$$

Hence when x is large, the inverse

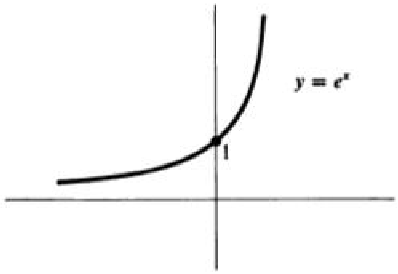
$$(e^x)^{-1} = 1/e^x$$

is small (positive).

Thus we may write:

$$\boxed{\begin{array}{llll} \text{If} & x \to \infty & \text{then} & e^x \to \infty. \\ \text{If} & x \to -\infty & \text{then} & e^x \to 0. \end{array}}$$

We are now in a position to see that the graph of e^x looks like this:



VIII, §1. EXERCISES

1. What is the equation of the tangent line to the curve $y = e^{2x}$ at the point whose x-coordinate is (a) 1, (b) -2, (c) 0?
2. What is the equation of the tangent line to the curve $y = e^{x/2}$ at the point whose x-coordinate is (a) -4, (b) 1, (c) 0?
3. What is the equation of the tangent line to the curve $y = xe^x$ at the point whose x-coordinate is 2?

4. Find the derivatives of the following functions:
   (a) $e^{\sin 3x}$ (b) $\sin(e^x + \sin x)$
   (c) $\sin(e^{x+2})$ (d) $\sin(e^{4x-5})$

5. Find the derivatives of the following functions:
   (a) $\arctan e^x$ (b) $e^x \cos(3x + 5)$
   (c) $e^{\sin 2x}$ (d) $e^{\arccos x}$
   (e) $1/e^x$ (f) x/e^x
   (g) e^{e^x} (h) $e^{-\arcsin x}$
   (i) $\tan(e^x)$ (j) $\arctan e^{2x}$
   (k) $1/(\sin e^x)$ (l) $\arcsin(e^x + x)$
   (m) $e^{\tan x}$ (n) $\tan e^x$

6. (a) Show that the n-th derivative of xe^x is $(x + n)e^x$ for $n = 1, 2, 3, 4, 5$.
   (b) Show that the n-th derivative of xe^{-x} is $(-1)^n(x - n)e^{-x}$ for $n = 1, 2, 3, 4, 5$.
   (c) Suppose you have already proved the above formulas for the n-th derivative of xe^x and xe^{-x}. How would you proceed to prove these formulas for the $(n + 1)$-th derivative?

7. Let $f(x)$ be a function such that $f'(x) = f(x)$ and $f(0) = 2$. Determine f completely in terms of e^x.

8. (a) Let $f(x)$ be a differentiable function over some interval satisfying the relation $f'(x) = Kf(x)$ for some constant K. Show that there is a constant C such that $f(x) = Ce^{Kx}$. [*Hint*: Show that the function $f(x)/e^{Kx}$ is constant.]
   (b) Let f be a differentiable function such that $f'(x) = -2xf(x)$. Show that there is a constant C such that $f(x) = Ce^{-x^2}$.
   (c) In general suppose there is a function h such that $f'(x) = h'(x)f(x)$. Show that $f(x) = Ce^{h(x)}$. [*Hint*: Show that the function $f(x)/e^{h(x)}$ is constant.] The technique of this exercise will be used in applications in the last section.

Find the tangent line to the curve at the indicated point.

9. $y = e^{2x}$, $x = 1$
10. $y = xe^x$, $x = 2$
11. $y = xe^x$, $x = 5$
12. $y = xe^{-x}$, $x = 0$
13. $y = e^{-x}$, $x = 0$
14. $y = x^2e^{-x}$, $x = 1$

15. Prove that there is a unique number x such that $e^x + x = 0$. [*Hint*: Show that the function is strictly increasing, and has positive and negative values.]

16. Prove the inequalities for $x > 0$:

   (a) $1 < e^x$ (b) $1 + x < e^x$ (c) $1 + x + \dfrac{x^2}{2} < e^x$

   [*Hint*: Using the method of Chapter V, §2, first prove (a). Then prove (b) using (a). Then prove (c) using (b).]

17. Let $x = 1$ in Exercise 16. Show that $2 < e$. Also show that $2.5 < e$.

18. Prove for $n = 3, 4, 5, 6$ that for $x > 0$ we have

$$1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} < e^x.$$

By $n!$ (read n **factorial**) we mean the product of the first n integers. For instance:

$$\begin{array}{ll} 1! = 1 & 4! = 24 \\ 2! = 2 & 5! = 120 \\ 3! = 6 & 6! = 720 \end{array}$$

19. For $x > 0$ prove:

(a) $1 - x < e^{-x}$ (b) $e^{-x} < 1 - x + \frac{x^2}{2}$

(c) $1 - x + \frac{x^2}{2} - \frac{x^3}{3 \cdot 2} < e^{-x}$

(d) $e^{-x} < 1 - x + \frac{x^2}{2} - \frac{x^3}{3 \cdot 2} + \frac{x^4}{4 \cdot 3 \cdot 2}$

20. (a) Let $x = 1/2$ in Exercise 19(a). Show that $e < 4$.
(b) Let $x = 1$ in Exercise 19(c). Show that $e < 3$.

Hyperbolic functions

21. (a) Define the functions **hyperbolic cosine** and **hyperbolic sine** by the formulas

$$\cosh t = \frac{e^t + e^{-t}}{2} \quad \text{and} \quad \sinh t = \frac{e^t - e^{-t}}{2}.$$

Show that their derivatives are given by

$$\cosh' = \sinh \quad \text{and} \quad \sinh' = \cosh.$$

(b) Show that for all t we have

$$\cosh^2 t - \sinh^2 t = 1.$$

Note: We see that the functions $\cosh t$ and $\sinh t$ satisfy the equation of a hyperbola, in a way similar to ordinary sine and cosine which satisfy the equation of a circle, namely

$$\sin^2 t + \cos^2 t = 1.$$

That is the reason $\cosh t$ and $\sinh t$ are called hyperbolic cosine and hyperbolic sine respectively.

22. Sketch the graph of the function

$$f(x) = \frac{e^x + e^{-x}}{2}.$$

Plot at least six points on this graph.

23. Sketch the graph of the function

$$f(x) = \frac{e^x - e^{-x}}{2}.$$

Plot at least six points on this graph.

24. Let $f(x) = \frac{1}{2}(e^x + e^{-x}) = \cosh x = y$.
    (a) Show that f is strictly increasing for $x \geqq 0$.
    Then the inverse function exists for this interval. Denote this inverse function by $x = \operatorname{arccosh} y = g(y)$.
    (b) For which numbers y is $\operatorname{arccosh} y$ defined?
    (c) Show that

$$g'(y) = \frac{1}{\sqrt{y^2 - 1}}.$$

25. Let $f(x) = \frac{1}{2}(e^x - e^{-x}) = \sinh x = y$.
    (a) Show that f is strictly increasing for all x. Let $x = \operatorname{arcsinh} y$ be the inverse function.
    (b) For which numbers y is $\operatorname{arcsinh} y$ defined?
    (c) Let $g(y) = \operatorname{arcsinh} y$. Show that

$$g'(y) = \frac{1}{\sqrt{1 + y^2}}.$$

VIII, §2. THE LOGARITHM

If $e^x = y$, then we define $x = \log y$. *Thus the* log, *here and thereafter, is what some call the natural* log. We don't deal with any other log. By definition, we therefore have:

$$\boxed{e^{\log y} = y \qquad \text{and} \qquad \log e^x = x.}$$

Thus log is the inverse function of the exponential function e^x. Since e^x is strictly increasing, the inverse function exists.

Examples. We have

$$\log e^2 = 2, \qquad \log e^{-\sqrt{2}} = -\sqrt{2},$$

$$\log e^{-3} = -3, \qquad \log e^{\pi} = \pi.$$

And the other way:

$$e^{\log 2} = 2, \qquad e^{\log \pi} = \pi.$$

Furthermore, the relation $e^0 = 1$ means that

$$\boxed{\log 1 = 0.}$$

Since all values e^x are positive for all numbers x, it follows that

$\log y$ *is defined only for positive numbers* y.

The rule $e^{a+b} = e^a e^b$ translates into a rule for the log, as follows.

Theorem 2.1. *If* u, v *are* > 0, *then*

$$\boxed{\log uv = \log u + \log v.}$$

Proof. Let $a = \log u$ and $b = \log v$. Then

$$e^{a+b} = e^a e^b = e^{\log u} e^{\log v} = uv.$$

By definition, the relation

$$e^{a+b} = uv$$

means that

$$\log uv = a + b = \log u + \log v$$

as was to be shown.

Theorem 2.2. *If* $u > 0$, *then*

$$\boxed{\log u^{-1} = -\log u.}$$

Proof. We have $1 = uu^{-1}$. Hence

$$0 = \log 1 = \log(uu^{-1}) = \log u + \log u^{-1}.$$

Adding $-\log u$ to both sides proves the theorem.

Examples. We have

$$\log(1/2) = -\log 2$$

$$\log(2/3) = \log 2 - \log 3.$$

Of course, we can take the log of a product with more than two terms, just as we can take the exponential of a sum of more than two terms. For instance

$$e^{a+b+c} = e^{a+b}e^c = e^a e^b e^c.$$

Similarly, if n is a positive integer, then

$$e^{na} = e^{a+a+\cdots+a} = e^a e^a \cdots e^a = (e^a)^n,$$

where the product on the right is taken n times.

We have the corresponding rule for the log, namely

$$\boxed{\log(u^n) = n \log u.}$$

For instance, by Theorem 2.1, we find:

$$\log(u^2) = \log(u \cdot u) = \log u + \log u = 2 \log u.$$

$$\begin{aligned}\log(u^3) = \log(u^2 u) &= \log u^2 + \log u \\ &= 2 \log u + \log u \\ &= 3 \log u.\end{aligned}$$

And so forth, to get $\log u^n = n \log u$.

It now follows that if n is a positive integer, then

$$\boxed{\log u^{1/n} = \frac{1}{n} \log u.}$$

Proof. Let $v = u^{1/n}$. Then $v^n = u$, and we have already seen that

$$\log v^n = n \log v.$$

Hence

$$\log v = \frac{1}{n} \log v^n,$$

which is precisely the relation $\log u^{1/n} = 1/n \log u$.

The same type of rule holds for fractional exponents, that is:

If m, n are positive integers, then

$$\log u^{m/n} = \frac{m}{n} \log u.$$

Proof. We write $u^{m/n} = (u^m)^{1/n}$. Then

$$\begin{aligned} \log u^{m/n} &= \log(u^m)^{1/n} \\ &= \frac{1}{n} \log u^m \\ &= \frac{m}{n} \log u \end{aligned}$$

by using the two cases separately.

Just to give you a feeling for the behavior of the log, we give a few approximate values:

$$\log 10 = 2.3\ldots, \qquad \log 10{,}000 = 9.2\ldots,$$

$$\log 100 = 4.6\ldots, \qquad \log 100{,}000 = 11.5\ldots,$$

$$\log 1000 = 6.9\ldots, \qquad \log 1{,}000{,}000 = 13.8\ldots.$$

You can see that if x grows like a geometric progression, then $\log x$ grows like an arithmetic progression. The above values illustrate the rule

$$\log 10^n = n \log 10,$$

where $\log 10$ is approximately 2.3.

In Exercises 17 and 19 of §1 you should have proved that $2.5 < e < 3$. Make up a table of the values e^n and $\log e^n = n$. You can then compare the growth of e^n with $\log e^n$ in a similar way. For positive integers you can then see that $\log e^n$ grows very slowly compared to e^n. For instance,

$$\log e^3 = 3,$$

$$\log e^4 = 4,$$

$$\log e^5 = 5,$$

$$\log e^{10} = 10.$$

Using the fact that e lies between 2 and 3, you can see that powers like e^5 or e^{10} are quite large compared to the values of the log, which are 5 and 10, respectively, in these cases. For instance, since $e > 2$ we have

$$e^{10} > 2^{10} > 1{,}000.$$

We have the same phenomenon in the opposite direction for negative powers of e. For instance:

$$\log \frac{1}{e} = -1,$$

$$\log \frac{1}{e^2} = -2,$$

$$\log \frac{1}{e^3} = -3,$$

$$\log \frac{1}{e^{10}} = -10.$$

Put $h = 1/e^y$. As h approaches 0, y becomes large positive, but rather slowly. Make a similar table for $\log(1/10^n)$ with $n = 1, 2, 3, 4, 5, 6$ to get a feeling for numerical examples.

Observe that if x is a small positive number and we write $x = e^y$ then $y = \log x$ is large negative. For instance

$$\text{if} \quad x = 1/e^{10^6} = e^{-10^6} \qquad \text{then} \qquad \log x = -10^6,$$

$$\text{if} \quad x = 1/e^{10^{100}} = e^{-10^{100}} \qquad \text{then} \qquad \log x = -10^{100}.$$

In short:

$$\boxed{\text{If} \quad x \to 0 \quad \text{then} \quad \log x \to -\infty.}$$

The reason comes from the behavior of e^y. If $y \to -\infty$ then $e^y \to 0$. Similarly, if $y \to \infty$ then $e^y \to \infty$. This translates into the corresponding property of the inverse function:

$$\boxed{\text{If} \quad x \to \infty \quad \text{then} \quad \log x \to \infty.}$$

The derivative of log

Next we consider the differentiation properties of the log function. Let

$$y = e^x \qquad \text{and} \qquad x = \log y.$$

By the rule for differentiating inverse functions, we find:

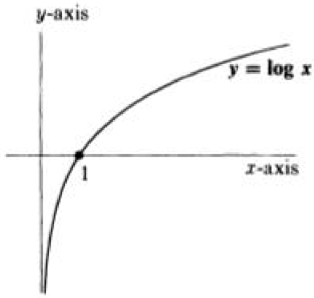
$$\frac{dx}{dy} = \frac{1}{dy/dx} = \frac{1}{e^x} = \frac{1}{y}.$$

Hence we have the formula:

Theorem 2.3.

$$\boxed{\frac{d \log y}{dy} = \frac{1}{y}.}$$

From the graph of e^x, we see that e^x takes on all values > 0. Hence the inverse function log is defined for all positive real numbers, and by the general way of finding the graph of an inverse function, we see that its graph looks like that in the figure.



In the figure, the graph crosses the horizontal axis at 1, because

$$e^0 = 1 \qquad \text{means} \qquad \log 1 = 0.$$

Note that the derivative satisfies

$$\frac{d \log x}{dx} = \frac{1}{x} > 0 \qquad \text{for all} \quad x > 0,$$

so the log function is strictly increasing.

Furthermore

$$\boxed{\frac{d^2 \log x}{dx^2} = -\frac{1}{x^2} < 0.}$$

We conclude that the log function is bending down as shown.

Remark. We shall sometimes consider composite functions of the type $\log(f(x))$. Since the log is not defined for numbers < 0, the expression $\log(f(x))$ is defined only for numbers x such that $f(x) > 0$. This is to be understood whenever we write such an expression.

Thus when we write $\log(x - 2)$, this is defined only when $x - 2 > 0$, in other words $x > 2$. When we write $\log(\sin x)$, this is meaningful only when $\sin x > 0$. It is not defined when $\sin x \leqq 0$.

Example. Find the tangent line to the curve $y = \log(x - 2)$ at the point $x = 5$.

Let $f(x) = \log(x - 2)$. Then $f'(x) = 1/(x - 2)$, and

$$f'(5) = 1/3.$$

When $x = 5$, $\log(x - 2) = \log 3$. We must find the equation of the line with slope $1/3$, passing through $(5, \log 3)$. This is easy, namely:

$$y - \log 3 = \tfrac{1}{3}(x - 5).$$

Example. Sketch the graph of the function $f(x) = x^2 + \log x$, for $x > 0$.

We begin by taking the derivative, namely

$$f'(x) = 2x + \frac{1}{x}.$$

The function f has a critical point precisely when $2x = -1/x$, that is $2x^2 = -1$. This can never be the case. Hence there is no critical point. When $x > 0$, the derivative is positive. Hence in this interval, the function is strictly increasing.

When x becomes large positive, both x^2 and $\log x$ become large positive. Hence

$$\text{if} \quad x \to \infty \quad \text{then} \quad f(x) \to \infty.$$

As x approaches 0 from the right, x^2 approaches 0, but $\log x$ becomes large negative. Hence

$$\text{if} \quad x \to 0 \quad \text{and} \quad x > 0 \quad \text{then} \quad f(x) \to -\infty.$$

Finally, to determine the regions where f is bending up or down, we take the second derivative, and find

$$f''(x) = 2 - \frac{1}{x^2} = \frac{2x^2 - 1}{x^2}.$$

Then:

$$f''(x) > 0 \quad \Leftrightarrow \quad 2x^2 - 1 > 0 \quad \Leftrightarrow \quad x > 1/\sqrt{2}$$

$$\Leftrightarrow \quad f \quad \text{is bending up.}$$

Similarly,

$$f''(x) < 0 \quad \Leftrightarrow \quad 2x^2 - 1 < 0 \quad \Leftrightarrow \quad x < 1/\sqrt{2}$$

$$\Leftrightarrow \quad f \quad \text{is bending down.}$$

Hence $1/\sqrt{2}$ is an inflection point. We claim that

$$f(1/\sqrt{2}) > 0.$$

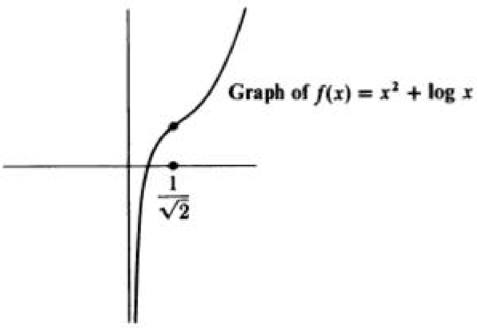
Indeed,

$$f(1/\sqrt{2}) = \frac{1}{2} - \log(\sqrt{2}) = \frac{1}{2} - \frac{1}{2}\log 2.$$

But the log is strictly increasing, and $2 < e$ so

$$\log 2 < \log e = 1.$$

Therefore $1 - \log 2 > 0$. This proves that $f(1/\sqrt{2}) > 0$. It follows that the graph of f looks like this.



VIII, §2. EXERCISES

1. What is the tangent line to the curve $y = \log x$ at the point whose x-coordinate is (a) 2, (b) 5, (c) $\frac{1}{2}$?

2. What is the equation of the tangent line of the curve $y = \log(x^2 + 1)$ at the point whose x-coordinate is (a) -1, (b) 2, (c) -3?

3. Find the derivatives of the following functions:

   (a) $\log(\sin x)$ (b) $\sin(\log(2x + 3))$ (c) $\log(x^2 + 5)$ (d) $\dfrac{\log 2x}{\sin x}$

4. What is the equation of the tangent line of the curve $y = \log(x + 1)$ at the point where x-coordinate is 3?

5. What is the equation of the tangent line of the curve $y = \log(2x - 5)$ at the point whose x-coordinate is 4?

6. (a) Prove that $\log(1 + x) < x$ for all $x > 0$. [*Hint*: Let $f(x) = x - \log(1 + x)$, find $f(0)$, and show that f is strictly increasing for $x \geqq 0$.]
   (b) For $x > 0$ show that
$$\frac{x}{1 + x} < \log(1 + x).$$

Find the tangent line to the curve at the indicated point.

7. $y = \log x$, at $x = e$
8. $y = x \log x$, at $x = e$
9. $y = x \log x$, at $x = 2$
10. $y = \log(x^3)$, at $x = e$
11. $y = \dfrac{1}{\log x}$, at $x = e$
12. $y = \dfrac{1}{\log x}$, at $x = 2$

Differentiate the following functions.

13. $\log(2x + 5)$
14. $\log(x^2 + 3)$
15. $\dfrac{1}{\log x}$
16. $\dfrac{x}{\log x}$
17. $x(\log x)^{1/3}$
18. $\log \sqrt{1 - x^2}$
19. Sketch the curve $y = x + \log x$, $x > 0$

20. Prove that there is a unique number $x > 0$ such that $\log x + x = 0$. [*Hint*: Show that the function is strictly increasing and takes on positive and negative values. Use the intermediate value theorem.]

VIII, §3. THE GENERAL EXPONENTIAL FUNCTION

Let a be a number > 0. In §1 we listed four properties of the function a^x, with x the variable. We shall now list one more:

Property 5. For all numbers x, y we have

$$(a^x)^y = a^{xy}.$$

For example, if $x = n$ and $y = m$ are positive integers, then $(a^m)^n$ is the product of a^m with itself n times, which is equal to a^{mn}. We shall now deduce some consequences from this property.

First, from the preceding section, we know that

$$a = e^{\log a}.$$

Therefore

$$a^x = (e^{\log a})^x.$$

By Property 5, this yields

$$\boxed{a^x = e^{x \log a}}$$

because $(\log a)x = x \log a$. Thus for instance,

$$2^x = e^{x \log 2}, \qquad \pi^x = e^{x \log \pi}, \qquad 10^x = e^{x \log 10}.$$

The above formula allows us to find the derivative of a^x. Remember that a is viewed as *constant*.

Theorem 3.1. *We have*

$$\boxed{\frac{d(a^x)}{dx} = a^x(\log a).}$$

Proof. We use the chain rule. Let $u = (\log a)x$. Then $du/dx = \log a$ and $a^x = e^u$. Hence

$$\frac{d(a^x)}{dx} = \frac{d(e^u)}{du}\frac{du}{dx} = e^u(\log a) = a^x \log a$$

as desired.

Example.

$$\frac{d(2^x)}{dx} = 2^x \log 2.$$

Warning. The derivative of x^x is **NOT** $x^x \log x$. Work it out in Exercise 7. The difference between x^x and a^x (like 2^x, or 10^x) is that a is constant, whereas in the expression x^x, the variable x appears twice.

The result of Theorem 3.1 clarifies the mysterious limit

$$\lim_{h \to 0} \frac{2^h - 1}{h}$$

which we encountered in §1. We shall now see that this limit is equal to $\log 2$. More generally, let $a > 0$ and let

$$f(x) = a^x.$$

In §1 we gave the argument that

$$f'(x) = a^x \lim_{h \to 0} \frac{a^h - 1}{h}.$$

By Theorem 3.1 we know also that

$$f'(x) = a^x \log a.$$

Therefore

$$\boxed{\lim_{h \to 0} \frac{a^h - 1}{h} = \log a.}$$

Since the log is strictly increasing, there is only one number a such that $\log a = 1$, and that number is $a = e$. Thus the exponential function e^x is the only one among all possible exponential functions a^x whose derivative is equal to itself. By Theorem 3.1, since

$$\frac{da^x}{dx} = a^x \log a,$$

if $a \neq e$ then we get the factor $\log a \neq 1$ coming into the formula for the derivative of a^x.

As an application of our theory of the exponential function, we also can take care of the general power function (which we had left dangling in Chapter III).

Theorem 3.2. *Let c be any number, and let*

$$f(x) = x^c$$

be defined for $x > 0$. Then $f'(x)$ exists and is equal to

$$f'(x) = cx^{c-1}.$$

Proof. Put $u = c \log x$. By definition,

$$f(x) = e^{c \log x} = e^u.$$

Then

$$\frac{du}{dx} = \frac{c}{x}.$$

Using the chain rule, we see that

$$f'(x) = e^u \cdot \frac{du}{dx} = e^{c \log x} \cdot \frac{c}{x} = x^c \cdot \frac{c}{x} = cx^c x^{-1} = cx^{c-1}.$$

This proves our theorem.

Warning. the number c in Theorem 3.2 is constant, and x is the variable. **Do not confuse**

$$\frac{dc^x}{dx} = c^x \log c \qquad \text{and} \qquad \frac{dx^c}{dx} = cx^{c-1}.$$

Example. Find the tangent line to the curve $y = x^x$ at $x = 2$.
We can write the function x^x in the form

$$y = f(x) = e^{x \log x}.$$

Then

$$\begin{aligned} f'(x) &= e^{x \log x}\left(x \cdot \frac{1}{x} + \log x\right) \\ &= x^x(1 + \log x). \end{aligned}$$

In particular, we get the derivative (slope) at $x = 2$,

$$f'(2) = 2^2(1 + \log 2) = 4(1 + \log 2).$$

We have $f(2) = 2^2 = 4$. Therefore the equation of the tangent line at $x = 2$ is

$$y - 4 = 4(1 + \log 2)(x - 2).$$

Definition. When x, y are two numbers such that $y = 2^x$, it is customary to say that x is the **log of y to the base 2**. Similarly, if a is a number > 0, and $y = a^x = e^{x \log a}$, we say that x is the **log of y to the base a**. When $y = e^x$, we simply say that $x = \log y$.

The log to the base a is sometimes written $\log_a$.

We conclude this section by discussing limits, which are now very easy to handle.

First, we have

Limit 1.

$$\boxed{\lim_{h \to 0} \frac{1}{h} \log(1 + h) = 1.}$$

Indeed, the limit on the left is nothing else but the limit of the Newton quotient

$$\lim_{h \to 0} \frac{\log(1 + h) - \log 1}{h} = \log'(1).$$

Since $\log'(x) = 1/x$, it follows that $\log'(1) = 1$, as desired.

Limit 2.

$$\boxed{\lim_{h \to 0} (1 + h)^{1/h} = e.}$$

Proof. We have

$$(1 + h)^{1/h} = e^{(1/h)\log(1 + h)}$$

because by definition $a^x = e^{x \log a}$. We have just seen that

$$\frac{1}{h} \log(1 + h) \to 1 \quad \text{as} \quad h \to 0.$$

Hence

$$e^{(1/h)\log(1 + h)} \to e^1 = e \quad \text{as} \quad h \to 0.$$

This proves the desired limit.

Remark. We are using here the *continuity* of the function e^x.

If x approaches a number x_0 then e^x approaches e^{x_0}.

That the function e^x is continuous follows from our assumption that e^x is differentiable.

For instance, if x approaches $\sqrt{2}$ then e^x approaches $e^{\sqrt{2}}$.
If x approaches 1 then e^x approaches $e^1 = e$.
If x approaches 0 then e^x approaches $e^0 = 1$.

Let us go back to Limit 2. We reformulate this limit. Write $h = 1/x$. When h approaches 0 then x becomes large, that is

$$h \to 0 \quad \text{if and only if} \quad x \to \infty.$$

Therefore we find the limit:

Limit 3.

$$\boxed{\lim_{x \to \infty} \left(1 + \frac{1}{x}\right)^x = e.}$$

In the exercises, you will deduce easily from this limit that for $r > 0$, we have

$$\boxed{\lim_{x \to \infty} \left(1 + \frac{r}{x}\right)^x = e^r.}$$

This has an interesting application.

Example. Compound interest. Let an amount of A dollars be invested at yearly compound interest of $100r$ per cent where $r > 0$. Thus r is the ratio of the rate of interest of 100 per cent. Then this original amount increases to the following after the indicated number of years:

After 1 year: $A + rA = (1 + r)A$.
After 2 years: $(1 + r)A + r(1 + r)A = (1 + r)^2 A$.
After 3 years: $(1 + r)^2 A + r(1 + r)^2 A = (1 + r)^3 A$.

Continuing in this way, we conclude that after n years the amount is

$$A_n = (1 + r)^n A.$$

Now suppose that this same interest rate $100r$ per cent is compounded every $1/m$ years, where m is a positive integer. This is equivalent to saying that the rate is $100r/m$ per cent per $1/m$ years, compounded every $1/m$ years. Let us apply the preceding formula to the case where the unit of time is $1/m$ year. Then q years are equal to $qm \cdot 1/m$ years. Therefore, if the interest is compounded every $1/m$ years, after q years the amount is

$$A_{q,m} = \left(1 + \frac{r}{m}\right)^{qm} A.$$

The amount one gets after q years if the interest is compounded continuously is the limit of $A_{q,m}$ as $m \to \infty$. In light of the limit which you should have determined, we see that after q years of continuous compounding, the amount is

$$\lim_{m \to \infty} \left(1 + \frac{r}{m}\right)^{qm} A = e^{rq}A.$$

To give a numerical case, suppose 1,000 dollars return 15 per cent compounded continuously. Then $r = 15/100$. After 10 years, the amount will be

$$e^{\frac{15}{100}10} \cdot 1{,}000 = e^{1.5}1{,}000.$$

Since e is approximately 2.7 you can get a definite numerical answer.

VIII, §3. EXERCISES

1. What is the derivative of 10^x? 7^x?

2. What is the derivative of 3^x? π^x?

3. Sketch the curves $y = 3^x$ and $y = 3^{-x}$. Plot at least five points.

4. Sketch the curves $y = 2^x$ and $y = 2^{-x}$. Plot at least five points.

5. Find the equation of the tangent line to the curve $y = 10^x$ at $x = 0$.

6. Find the equation of the tangent line to the curve of $y = \pi^x$ at $x = 2$.

7. (a) What is the derivative of the function x^x (defined for $x > 0$)? [*Hint*: $x^x = e^{x \log x}$]

   (b) What is the derivative of the function $x^{(x^x)}$?

8. Find the equation of the tangent line to the curve $y = x^x$
   (a) at the point $x = 1$ (b) at $x = 2$ (c) at $x = 3$.

Find the tangent lines of the following curves:

9. $y = x^{\sqrt{x}}$ (a) at $x = 2$ (b) at $x = 5$

10. $y = x^{\sqrt[3]{x}}$ (a) at $x = 2$ (b) at $x = 5$

11. If a is a number > 1 and $x > 0$, show that

$$x^a - 1 \geqq a(x - 1).$$

12. Let a be a number > 0. Find the critical points of the function $f(x) = x^2/a^x$.

13. Let $0 < r$. Using Limit 3, prove the limit

$$\lim_{x\to\infty}\left(1 + \frac{r}{x}\right)^x = e^r.$$

[*Hint*: Let $x = ry$ and let $y \to \infty$.]

14. Show that

$$\lim_{n\to\infty} n(\sqrt[n]{a} - 1) = \log a.$$

[*Hint*: Let $h = 1/n$.] This exercise shows how approximations of the log can be obtained just by taking ordinary n-th roots. In fact, if we take $n = 2^k$ and use large integers k, we obtain arbitrarily good approximations of the log by extracting a succession of square roots. Do it on a pocket calculator to check it out.

VIII, §4. SOME APPLICATIONS

It is known (from experimental data) that when a piece of radium is left to disintegrate, the rate of disintegration is proportional to the amount of radium left. Two quantities are proportional when one is a constant multiple of the other.

Suppose that at time $t = 0$ we have 10 grams of radium and let $f(t)$ be the amount of radium left at time t. Then

$$\boxed{\frac{df}{dt} = Kf(t)}$$

for some constant K. We take K negative since the physical interpretation is that the amount of substance decreases.

Let us show that there is a constant C such that

$$\boxed{f(t) = Ce^{Kt}.}$$

If we take the derivative of the quotient

$$\frac{f(t)}{e^{Kt}}$$

and use the rule for the derivative of a quotient, we find

$$\frac{d}{dt}\left(\frac{f(t)}{e^{Kt}}\right) = \frac{e^{Kt}f'(t) - Ke^{Kt}f(t)}{e^{2Kt}} = 0$$

because $f'(t) = Kf(t)$. Since the derivative is 0, the quotient $f(t)/e^{Kt}$ is constant, or equivalently, there is a constant C such that

$$f(t) = Ce^{Kt}.$$

Let $t = 0$. Then $f(0) = C$. Thus $C = 10$, if we assumed that we started with 10 grams.

In general, if $f(t) = Ce^{Kt}$ is the function giving the amount of substance as a function of time, then

$$\boxed{f(0) = C,}$$

and C is interpreted as the amount of substance when $t = 0$, that is the original amount.

Similarly, consider a chemical reaction. It is frequently the case that the rate of the reaction is proportional to the quantity of reacting substance present. If $f(t)$ denotes the amount of substance left after time t, then

$$\frac{df}{dt} = Kf(t)$$

for some constant K (determined experimentally in each case). We are therefore in a similar situation as before, and

$$f(t) = Ce^{Kt},$$

where C is the amount of substance at $t = 0$.

Example 1. Suppose $f(t) = 10e^{Kt}$ where K is constant. Assume that $f(3) = 5$. Find K.

We have

$$5 = 10e^{K3}$$

and therefore

$$e^{3K} = \tfrac{5}{10} = \tfrac{1}{2},$$

whence

$$3K = \log(1/2) \qquad \text{and} \qquad K = \frac{-\log 2}{3}.$$

Example 2. Sugar in water decomposes at a rate proportional to the amount still unchanged. If 50 lb of sugar reduce to 15 lb in 3 hr, when will 20 per cent of the sugar be decomposed?

Let $S(t)$ be the amount of sugar undecomposed, at time t. Then by hypothesis,

$$S(t) = Ce^{-kt},$$

for suitable constants C and k. Furthermore, since $S(0) = C$, we have $C = 50$. Thus

$$S(t) = 50e^{-kt}.$$

We also have

$$S(3) = 50e^{-3k} = 15$$

so

$$e^{-3k} = \tfrac{15}{50} = \tfrac{3}{10}.$$

Thus we can solve for k, namely we take the log and get

$$-3k = \log(3/10),$$

whence

$$-k = \tfrac{1}{3}\log(3/10).$$

When 20 per cent has decomposed then 80 per cent is left. Note that 80 per cent of 50 is 40. We want to find t such that

$$40 = 50e^{-kt},$$

or in other words,

$$e^{-kt} = \tfrac{40}{50} = \tfrac{4}{5}.$$

We obtain

$$-kt = \log(4/5),$$

whence

$$t = \frac{\log(4/5)}{-k} = 3\,\frac{\log(4/5)}{\log(3/10)}.$$

This is our answer.

Remark. It does not make any difference whether originally we let

$$S(t) = Ce^{-kt} \qquad \text{or} \qquad S(t) = Ce^{Kt}.$$

We could also have worked the problem the other way. For applications, when substances decrease, it is convenient to use a convention such that $k > 0$ so that the expression e^{-kt} decreases when t increases. But mathematically the procedures are equivalent, putting $K = -k$.

Example 3. A radioactive substance disintegrates proportionally to the amount of substance present at a given time, say

$$f(t) = Ce^{-kt}$$

for some positive constant k. At what time will there be exactly 1/4-th of the original amount left?

To do this, we want to know the value of t such that

$$f(t) = C/4.$$

Thus we want to solve

$$Ce^{-kt} = C/4.$$

Note that we can cancel C to get $e^{-kt} = 1/4$. Taking logs yields

$$-kt = -\log 4,$$

whence

$$t = \frac{\log 4}{k}.$$

Observe that the answer is independent of the original amount C. Experiments also allow us to determine the constant k. For instance, if we can analyze a sample, and determine that 1/4-th is left after 1000 years, then we find that

$$k = \frac{\log 4}{1000}.$$

Example 4. Exponential growth also reflects population explosion. If $P(t)$ is the population as a function of time t, then its rate of increase is proportional to the total population, in other words,

$$\frac{dP}{dt} = KP(t)$$

for some positive constant K. It then follows that

$$P(t) = Ce^{Kt}$$

for some constant C which is the population at time $t = 0$.

Suppose we ask at what time the population will double. We must then find t such that

$$Ce^{Kt} = 2C,$$

or equivalently

$$e^{Kt} = 2.$$

Taking the log yields

$$Kt = \log 2,$$

whence

$$t = \frac{\log 2}{K}.$$

Note that this time depends only on the rate of change of the population, not on the original value of C.

VIII, §4. EXERCISES

1. Let $f(t) = 10e^{Kt}$ for some constant K. Suppose you know that $f(1/2) = 2$. Find K.

2. Let $f(t) = Ce^{2t}$. Suppose that you know $f(2) = 5$. Determine the constant C.

3. One gram of radium is left to disintegrate. After one million years, there is 0.1 gram left. What is the formula giving the rate of disintegration?

4. A certain chemical substance reacts in such a way that the rate of reaction is equal to the quantity of substance present. After one hour, there are 20 grams of substance left. How much substance was there at the beginning?

5. A radioactive substance disintegrates proportionally to the amount of substance present at a given time, say

$$f(t) = Ce^{Kt}.$$

At what time will there be exactly half the original amount left?

6. Suppose $K = -4$ in the preceding exercise. At what time will there be one-third of the substance left?

7. If bacteria increase in number at a rate proportional to the number present, how long will it take before 1,000,000 bacteria increase to 10,000,000 if it takes 12 minutes to increase to 2,000,000?

8. A substance decomposes at a rate proportional to the amount present. At the end of 3 minutes, 10 per cent of the original substance has decomposed. When will half the original amount have decomposed?

9. Let f be a function of a variable t and increasing at the rate $df/dt = kf$ where k is a constant. Let $a_n = f(nt_1)$ where t_1 is a fixed value of t, $t_1 > 0$. Show that $a_0, a_1, a_2, \ldots$ is a geometric progression.

10. In 1900 the population of a city was 50,000. In 1950 it was 100,000. If the rate of increase of the population is proportional to the population, what is the population in 1984? In what year is it 200,000?

11. Assume that the rate of change with respect to height of atmospheric pressure at any height is proportional to the pressure there. If the barometer reads 30 at sea level and 24 at 6000 ft above sea level, find the barometric reading 10,000 ft above sea level.

12. Sugar in water decomposes at a rate proportional to the amount still unchanged. If 30 lb of sugar reduces to 10 lb in 4 hr, when will 95 per cent of the sugar be decomposed?

13. A particle moves with speed $s(t)$ satisfying $ds/dt = -ks$, where k is some constant. If the initial speed is 16 units/min and if the speed is halved in 2 min, find the value of t when the speed is 10 units/min.

14. Assume that the difference x between the temperature of a body and that of surrounding air decreases at a rate proportional to this difference. If $x = 100°$ when $t = 0$, and $x = 40°$ when $t = 40$ minutes, find t (a) when $x = 70°$, (b) when $x = 16°$, (c) the value of x when $t = 20$.

15. A moron loses money in gambling at a rate equal to the amount he owns at any given time. At what time t will he have lost half of his initial capital?

16. It is known that radioactive carbon has a half-life of 5568 years, meaning that it takes that long for one-half of the original amount to decompose. Also, the rate of decomposition is proportional to the amount present, so that by what we have seen in the text, we have the formula

$$f(t) = Ce^{Kt}$$

for this amount, where C and K are constants.
(a) Find the constant K explicitly.
(b) Some decomposed carbon is found in a cave, and an analysis shows that one-fifth of the original amount has decomposed. How long has the carbon been in the cave?

VIII, §5. ORDER OF MAGNITUDE

In this section we analyze more closely what we mean when we say that e^x grows much faster than x, and $\log x$ grows much slower than x, when x becomes large positive.

We consider the quotient

$$\frac{e^x}{x}$$

as x becomes large positive. Both the numerator and the denominator become large, and the question is, what is the behavior of the quotient?

First let us make a table for simple values $2^n/n$ when n is a *positive integer*, to see that $2^n/n$ becomes large as n becomes large, experimentally. We agree to the convention that n always denotes a positive integer, unless otherwise specified.

n	2^n	$2^n/n$
1	2	2
2	4	2
3	8	8/3
4	16	4
5	32	$32/5 > 6$
10	1,024	$102.4 > 100$
20	1,048,576	$52{,}428.8 > 5 \times 10^4$

Since $2 < e$, we have $2^n/n < e^n/n$, and we see *experimentally* that e^n/n becomes large. We now wish to *prove* this fact. We first prove some inequalities for e^x. We use techniques from the exercises of §1. We proceed stepwise. We consider $x \geqq 0$.

(a) We first show that

$$\boxed{1 + x < e^x \qquad \text{for} \quad x > 0.}$$

Let $f_1(x) = e^x - (1 + x)$. Then $f'_1(x) = e^x - 1$. Since $e^x > 1$ for $x > 0$, we conclude that

$$f'_1(x) > 0 \qquad \text{for} \quad x > 0.$$

Therefore $f_1(x)$ is strictly increasing for $x \geqq 0$. Since $f_1(0) = 0$, we conclude that $f_1(x) > 0$ for $x > 0$, which means

$$e^x - (1 + x) > 0, \qquad \text{or in other words,} \qquad e^x > 1 + x,$$

as was to be shown.

(b) Next we show that

$$\boxed{1 + x + \frac{x^2}{2} < e^x \qquad \text{for} \quad x > 0.}$$

Let $f_2(x) = e^x - (1 + x + x^2/2)$. Then $f_2(0) = 0$. Furthermore,

$$f'_2(x) = \mathrm{e}^x - (1 + x).$$

By part (a), we know that $f'_2(x) > 0$ for $x > 0$. Hence f_2 is strictly increasing, and it follows that $f_2(x) > 0$ for $x > 0$, or in other words,

$$e^x - \left(1 + x + \frac{x^2}{2}\right) > 0 \qquad \text{for} \quad x > 0.$$

This proves the desired inequality.

Theorem 5.1. *The function e^x/x becomes large as x becomes large.*

Proof. We divide both sides of inequality (b) by x we obtain

$$\frac{1}{x} + 1 + \frac{x}{2} < \frac{e^x}{x}.$$

As x becomes large, so does the left-hand side, and Theorem 5.1 is proved.

Theorem 5.2. *The function e^x/x^2 becomes large as x becomes large. More generally, let m be a positive integer. Then*

$$\frac{e^x}{x^m} \to \infty \qquad \text{as} \quad x \to \infty.$$

Proof. We use the same method. First we prove the inequality

$$\text{(c)} \qquad \boxed{1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} < e^x \qquad \text{for} \quad x > 0.}$$

Recall that by definition, $2! = 2$ and $3! = 3 \cdot 2 = 6$. This time we let

$$f_3(x) = e^x - \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!}\right).$$

Then $f_3(0) = 0$. Furthermore, using inequality (b) we find

$$f'_3(x) = e^x - \left(1 + x + \frac{x^2}{2!}\right) = f_2(x) > 0 \qquad \text{for} \quad x > 0.$$

Hence $f_3(x)$ is strictly increasing, and therefore $f_3(x) > 0$ for $x > 0$. This proves inequality (c).

If we divide both sides of inequality (c) by x^2, then we find

$$\frac{1}{x^2} + \frac{1}{x} + \frac{1}{2} + \frac{x}{6} < \frac{e^x}{x^2}.$$

As x becomes large, the left-hand side becomes large, so e^x/x^2 becomes large. This proves the first statement of Theorem 5.2.

We can continue the same method to prove the general statement about e^x/x^n. First you should prove that

$$\text{(d)} \qquad \boxed{1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} < e^x \qquad \text{for} \quad x > 0}$$

in order to get a good feeling for the stepwise procedure used. We shall now prove the general step using an arbitrary integer n. In general, let

$$P_n(x) = 1 + x + \frac{x^2}{2!} + \frac{x^2}{3!} + \cdots + \frac{x^n}{n!}.$$

Suppose we have already proved that

$$1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} < e^x \qquad \text{for} \quad x > 0,$$

or in other words, that

$$P_n(x) < e^x \qquad \text{for} \quad x > 0.$$

We shall then prove

$$P_{n+1}(x) < e^x \qquad \text{for} \quad x > 0.$$

To do this, we let

$$f_{n+1}(x) = e^x - P_{n+1}(x) \qquad \text{and} \qquad f_n(x) = e^x - P_n(x).$$

Then $f_{n+1}(0) = 0$ and $f'_{n+1}(x) = f_n(x) > 0$ for $x > 0$. Hence f_{n+1} is strictly increasing, and therefore $f_{n+1}(x) > 0$ for $x > 0$, as desired.

Therefore, given our integer m, we have an inequality

$$\boxed{1 + x + \frac{x^2}{2} + \cdots + \frac{x^{m+1}}{(m+1)!} < e^x \qquad \text{for} \quad x > 0.}$$

We divide both sides of this inequality by x^m. Then the left-hand side consists of a sum of positive terms, the last of which is

$$\frac{x}{(m+1)!}.$$

Hence we obtain the inequality

$$\frac{x}{(m+1)!} < \frac{e^x}{x^m} \qquad \text{for} \quad x > 0.$$

Since the left-hand side becomes large when x becomes large, so does the right-hand side, and Theorem 5.2 is proved.

Example. We sketch the graph of $f(x) = xe^x$. We have

$$f'(x) = xe^x + e^x = e^x(x + 1).$$

Since $e^x > 0$ for all x, we get:

$$f'(x) = 0 \quad \Leftrightarrow \quad x + 1 = 0 \quad \Leftrightarrow \quad x = -1,$$

$$f'(x) > 0 \quad \Leftrightarrow \quad x + 1 > 0 \quad \Leftrightarrow \quad x > -1,$$

$$f'(x) < 0 \quad \Leftrightarrow \quad x + 1 < 0 \quad \Leftrightarrow \quad x < -1.$$

There is only one critical point at $x = -1$, and the other inequalities give us the regions of increase and decrease for f.

As to the bending up or down:

$$f''(x) = e^x \cdot 1 + e^x(x + 1) = e^x(x + 2).$$

Therefore:

$$f''(x) = 0 \quad \Leftrightarrow \quad x = -2,$$

$$f''(x) > 0 \quad \Leftrightarrow \quad x > -2 \quad \Leftrightarrow \quad f \text{ is bending up},$$

$$f''(x) < 0 \quad \Leftrightarrow \quad x < -2 \quad \Leftrightarrow \quad f \text{ is bending down}.$$

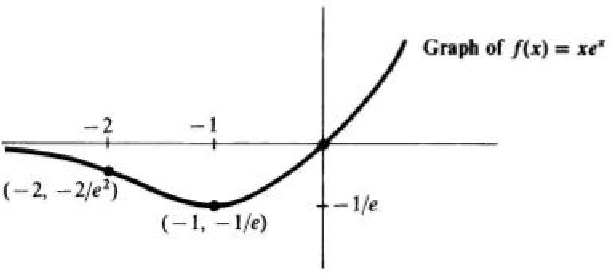
If $x \to \infty$ then $e^x \to \infty$ so $f(x) \to \infty$.

If $x \to -\infty$ then we put $x = -y$ with $y \to \infty$.

By Theorem 5.1,

$$xe^x = -ye^{-y} \to 0 \qquad \text{if} \quad y \to \infty.$$

Finally, $f(0) = 0$, $f(-1) = -1/e$, $f(-2) = -2/e^2$. Hence the graph looks like this.



Example. Let $0 < a < 1$ and a is a fixed number. Find the maximum of the function

$$f(x) = xa^x.$$

First, take the derivative:

$$\begin{aligned} f'(x) &= x \cdot a^x \log a + a^x \\ &= a^x(x \log a + 1). \end{aligned}$$

Since $a^x > 0$ for all x, we see that

$$f'(x) = 0 \quad \Leftrightarrow \quad x = -\frac{1}{\log a}.$$

Thus the function has exactly one critical point. Furthermore,

$$f'(x) > 0 \quad \Leftrightarrow \quad x \log a + 1 > 0 \quad \Leftrightarrow \quad x < -\frac{1}{\log a}$$

$$f'(x) < 0 \quad \Leftrightarrow \quad x \log a + 1 < 0 \quad \Leftrightarrow \quad x > -\frac{1}{\log a}.$$

(Remember that $0 < a < 1$, so that $\log a$ is negative.) Consequently the function is increasing on the interval to the left of the critical point, and decreasing on the interval to the right of the critical point. Therefore the critical point is the desired maximum. The value of f at this critical point is equal to

$$\begin{aligned} f(-1/\log a) &= -\frac{1}{\log a} a^{-1/\log a} = -\frac{1}{\log a} e^{-\log a/\log a} \\ &= -\frac{1}{e \log a}. \end{aligned}$$

Example. Show that the equation $3^x = 5x$ has at least one solution.

Let $f(x) = 3^x - 5x$. Then $f(0) = 1$, and by trial and error we find a value where f is negative, namely

$$f(2) = 9 - 10 < 0.$$

By the intermediate value theorem, there exists some number x between 2 and 0 such that $f(x) = 0$, and this number fulfills our requirement.

From Theorems 5.1 and 5.2, by means of a change of variable, we can analyze what happens when comparing $\log x$ with powers of x.

Theorem 5.3. *As x becomes large, the quotient $x/\log x$ also becomes large.*

Proof. Our strategy is to reduce this statement to Theorem 5.1. We make a change of variables. Let $y = \log x$. Then $x = e^y$ and our quotient has the form

$$\frac{x}{\log x} = \frac{e^y}{y}.$$

We know that $y = \log x$ becomes large when x becomes large. So does e^y/y by Theorem 5.1. This proves the theorem.

Corollary 5.4. *As x becomes large, the function $x - \log x$ also becomes large.*

Proof. We write

$$x - \log x = x\left(1 - \frac{\log x}{x}\right),$$

that is we factor x in the expression $x - \log x$. By Theorem 5.3, $(\log x)/x$ approaches 0 as x becomes large. Hence the factor

$$1 - \frac{\log x}{x}$$

approaches 1. The factor x becomes large. Hence the product becomes large. This proves the corollary.

Remark. We have just used the same factoring technique that was used in analyzing the behavior of polynomials, as when we wrote

$$x^3 - 2x^2 + 5 = x^3\left(1 - \frac{2}{x} + \frac{5}{x^3}\right)$$

to see that the x^3 term determines the behavior of the polynomial when x becomes large.

Corollary 5.5. *As x becomes large, $x^{1/x}$ approaches 1 as a limit.*

Proof. We write

$$x^{1/x} = e^{(\log x)/x}.$$

By Theorem 5.3 we know that $(\log x)/x$ approaches 0 when x becomes large. Hence

$$e^{(\log x)/x}$$

approaches 1, as desired.

Remark. In Corollary 5.5 we used the fact that the function e^u is continuous, because any differentiable function is continuous. If u approaches u_0 then e^u approaches e^{u_0}. Thus if $u = (\log x)/x$, then u approaches 0 as x becomes large, so e^u approaches $e^0 = 1$.

VIII, §5. EXERCISES

1. Sketch the graph of the curve $y = xe^{2x}$. In this and other exercises, you may treat the convexity properties as optional, but it usually comes out easily.

Sketch the graphs of the following functions. (In Exercises 6 through 8, $x \neq 0$.)

2. xe^{-x}
3. xe^{-x^2}
4. $x^2e^{-x^2}$
5. x^2e^{-x}
6. e^x/x
7. e^x/x^2
8. e^x/x^3
9. $e^x - x$
10. $e^x + x$
11. $e^{-x} + x$
12. Sketch the graph of $f(x) = x - \log x$.
13. Show that the equation $e^x = ax$ has at least one solution for any number a except when $0 \leqq a < e$.
14. (a) Give values of $x \log x$ when $x = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$, in general when $x = 1/2^n$ for some positive integer n.
    (b) Does $x \log x$ approach a limit as $x \to 0$? What about $x^2 \log x$? [*Hint*: Let $x = e^{-y}$ and let y become large.]
15. Let n be a positive integer. Prove that $x(\log x)^n \to 0$ as $x \to 0$.
16. Prove that $(\log x)^n/x \to 0$ as $x \to \infty$.
17. Sketch the following curves for $x > 0$.
    (a) $y = x \log x$
    (b) $y = x^2 \log x$
    (c) $y = x(\log x)^2$
    (d) $y = x/\log x$
18. Show that the function $f(x) = x^x$ is strictly increasing for $x > 1/e$.
19. Sketch the curve $f(x) = x^x$ for $x > 0$.
20. Sketch the curve $f(x) = x^{-x}$ for $x > 0$.

21. Let $f(x) = 2^x x^x$. Show that f is strictly increasing for $x > 1/2e$.

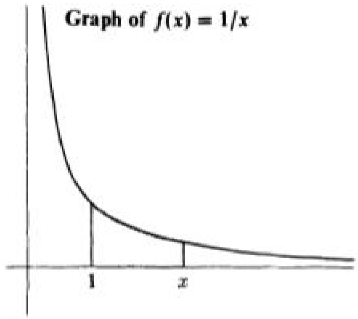
22. Find the following limits as $n \to \infty$
 (a) $(\log n)^{1/n}$ (b) $[(\log n)/n]^{1/n}$
 (c) $(n/e^n)^{1/n}$ (d) $(n \log n)^{1/n}$

VIII, §6. THE LOGARITHM AS THE AREA UNDER THE CURVE $1/x$

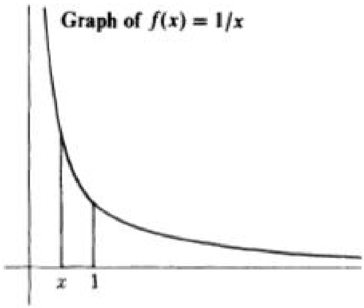
The present section is interesting for its own sake, because it gives us further insight into the logarithm. It also provides a very nice and concrete introduction to integration which is going to be covered in the next part. We shall give an interpretation of the logarithm as the area under a curve.

We define a function $L(x)$ to be the area under the curve $1/x$ between 1 and x if $x \geqq 1$, and the negative of the area under the curve $1/x$ between 1 and x if $0 < x < 1$. In particular, $L(1) = 0$.

The shaded portion of the picture that follows represents the area under the curve between 1 and x. On the left we have taken $x > 1$.



$L(x)$ for $x > 1$



$L(x)$ for $0 < x < 1$

If $0 < x < 1$, we would have the picture shown on the right. If $0 < x < 1$, we have said that $L(x)$ is equal to the negative of the area. Thus $L(x) < 0$ if $0 < x < 1$ and $L(x) > 0$ if $x > 1$.

We shall prove:

1. $L'(x) = 1/x$.
2. $L(x) = \log x$.

The first assertion that $L'(x) = 1/x$ is independent of everything else in this chapter, and we state it as a separate theorem.

Theorem 6.1. *The function $L(x)$ is differentiable, and*

$$\frac{dL(x)}{dx} = \frac{1}{x}.$$

Proof. We form the Newton quotient

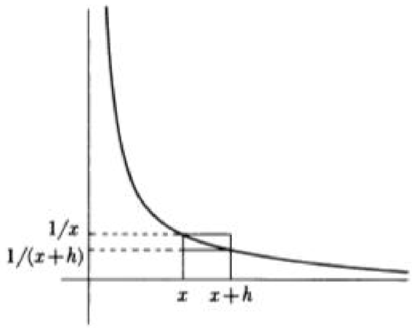
$$\frac{L(x+h) - L(x)}{h}$$

and have to prove that it approaches $1/x$ as a limit when h approaches 0.

Let us take $x \geqq 1$ and $h > 0$ for the moment. Then $L(x + h) - L(x)$ is the area under the curve between x and $x + h$. Since the curve $1/x$ is decreasing, this area satisfies the following inequalities:

$$h\frac{1}{x+h} < L(x+h) - L(x) < h\frac{1}{x}.$$

Indeed, $1/x$ is the height of the big rectangle as drawn on the next figure, and $1/(x + h)$ is the height of the small rectangle. Since h is the base of the rectangle, and since the area under the curve $1/x$ between x and



$x + h$ is in between the two rectangles, we see that it satisfies our inequalities. We divide both sides of our inequalities by the positive number h. Then the inequalities are preserved, and we get

$$\frac{1}{x+h} < \frac{L(x+h) - L(x)}{h} < \frac{1}{x}.$$

As h approaches 0, our Newton quotient is squeezed between $1/(x + h)$ and $1/x$ and consequently approaches $1/x$. This proves our theorem in case $h > 0$.

When $h < 0$ we use an entirely similar argument, which we leave as an exercise. (You have to pay attention to the sign of L. Also when you divide an inequality by h and $h < 0$, then the inequality gets reversed. However, you will again see that the Newton quotient is squeezed between $1/x$ and $1/(x + h)$.)

Theorem 6.2. *The function* $L(x)$ *is equal to* $\log x$.

Proof. Both functions $L(x)$ and $\log x$ have the same derivative, namely $1/x$ for $x > 0$. Hence there is a constant C such that

$$L(x) = \log x + C.$$

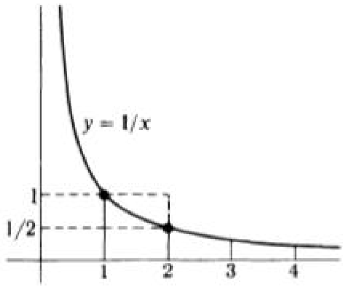
This is true for all $x > 0$. In particular, let $x = 1$. We obtain

$$0 = L(1) = \log 1 + C.$$

But $\log 1 = 0$. Hence $C = 0$, and the theorem is proved.

The identification of the log with the area under the curve $1/x$ can be used to give inequalities for the log. This is simple, and is given as an exercise. We can also obtain an estimate for e.

Example. The area under the curve $1/x$ between 1 and 2 is less than the area of a rectangle whose base is the interval [1, 2] and whose height is 1, as shown on the following figure.



Hence we get the inequality

$$\log 2 < 1.$$

Since $\log e = 1$, it follows that $2 < e$. This gives us a lower estimate for e.

Similarly we obtain an upper estimate as follows. The area under the curve $1/x$ between 1 and 2 is greater than the area of the rectangle whose base is the interval [1, 2] and whose height is 1/2, as shown on the above figure. Hence we get the inequality

$$\log 2 > \frac{1}{2}.$$

Then

$$\log 4 = \log(2^2) = 2 \log 2 > 2 \cdot \frac{1}{2} = 1.$$

Since $\log e = 1$ it follows that $e < 4$.

In Exercises 16 through 20 of §1 we had another method to obtain estimates for e. The method with the area under the curve can be used in other contexts and is independently useful.

VIII, §6. EXERCISES

1. Let h be a positive number. Compare the area under the curve $1/x$ between 1 and $1 + h$ with the area of suitable rectangles to show that

$$\frac{h}{1+h} < \log(1 + h) < h.$$

2. Prove by using Exercise 1 that

$$\lim_{h \to 0} \frac{1}{h} \log(1 + h) = 1.$$

3. Prove by comparing areas, that for every positive integer n, we have

$$\frac{1}{n+1} < \log\left(1 + \frac{1}{n}\right) < \frac{1}{n}.$$

4. Instead of using $\log 4 = \log(2^2)$ as in the text, use two rectangles under the graph of $1/x$, with bases [1, 2] and [2, 4] to show that $\log 4 > 1$.

VIII, APPENDIX. SYSTEMATIC PROOF OF THE THEORY OF EXPONENTIALS AND LOGARITHMS

Instead of assuming the five basic properties of the exponential function as in §1, and §3, we might have given a development of the log and exponential as follows. This is intended only for those interested in theory.

We *start* by defining $L(x)$ as we did in §6. The proof that $L'(x) = 1/x$ is self-contained, and yields a function L defined for all $x > 0$ and satisfying

$$L'(x) = 1/x \qquad \text{and} \qquad L(0) = 1.$$

Since $1/x > 0$ for all $x > 0$, it follows that the function L is strictly increasing, and so has an inverse function, which we denote by $x = E(y)$. Then using the rule for derivative of an inverse function, we find:

$$E'(y) = \frac{1}{L'(x)} = \frac{1}{1/x} = x = E(y).$$

Thus we have found a function E such that $E'(y) = E(y)$ for all y. In other words, we have found a function equal to its own derivative.

Since $L(0) = 1$ we find that $E(1) = 0$.

Next we prove:

For all numbers a, $b > 0$ we have

$$\boxed{L(ab) = L(a) + L(b).}$$

Proof. Fix the number a and let $f(x) = L(ax)$. By the chain rule, we obtain

$$f'(x) = \frac{1}{ax} \cdot a = \frac{1}{x}.$$

Since L and f have the same derivative, there is a constant C such that $f(x) = L(x) + C$ for all $x > 0$. In particular, for $x = 1$ we get

$$L(a) = f(1) = L(1) + C = 0 + C = C.$$

So $L(a) = C$ and $L(ax) = L(x) + L(a)$. This proves the first property.

Since $L'(x) = 1/x > 0$ for $x > 0$ it follows that L is strictly increasing. Since $L''(x) = -1/x^2 < 0$, it follows that L is bending down.

Let $a > 0$. We get:

$$L(a^2) = L(a) + L(a) = 2L(a),$$
$$L(a^3) = L(a^2 a) = L(a^2) + L(a) = 2L(a) + L(a) = 3L(a).$$

Continuing in this way, we get for all positive integers n:

$$L(a^n) = nL(a)$$

In particular, take $a > 1$. Since $L(1) = 0$ we conclude that $L(a) > 0$ because L is strictly increasing. Hence $L(a^n) \to \infty$ as $n \to \infty$. Again since L is strictly increasing, it follows that $L(x) \to \infty$ as $x \to \infty$.

From the formula

$$0 = L(1) = L(aa^{-1}) = L(a) + L(a^{-1})$$

we conclude that

$$\boxed{L(a^{-1}) = -L(a).}$$

Next, let x approach 0. Write $x = 1/y$ where $y \to \infty$. Then

$$L(x) = -L(y) \to -\infty \quad \text{as} \quad y \to \infty$$

so $L(x) \to -\infty$ as $x \to 0$. We may now write $\log x$ instead of $L(x)$.

Now let E be the inverse function of L. We already proved that $E' = E$. The inverse function of L is defined on the set of values of L, which is all numbers. The set of values of E is the domain of definition of L, which is the set of positive numbers. So $E(y) > 0$ for all y. Thus E is strictly increasing, and $E''(y) = E(y)$ for all y shows that the graph of E bends up.

We have $E(0) = 1$ because $L(1) = 0$.

Then we can prove

$$\boxed{E(u + v) = E(u)E(v).}$$

Namely, let $a = E(u)$ and $b = E(v)$. By the meaning of inverse function, $u = L(a)$ and $v = L(b)$. Then:

$$L(ab) = L(a) + L(b) = u + v.$$

Hence

$$E(u)E(v) = ab = E(u + v),$$

as was to be shown.

We now **define** $e = E(1)$. Since E is the inverse function of L we have $L(e) = 1$. From the rule

$$E(u + v) = E(u)E(v)$$

we now get for any positive integer n that

$$E(n) = E(1 + 1 + \cdots + 1) = E(1)^n = e^n.$$

Similarly,

$$E(nu) = E(u)^n.$$

Put $u = 1/n$. Then

$$e = E(1) = E\left(n \cdot \frac{1}{n}\right) = E\left(\frac{1}{n}\right)^n.$$

Hence $E(1/n)$ is the n-th root of e. From now on we write

$$e^u \quad \text{instead of} \quad E(u).$$

Next we deal with the general exponential function.

Let a be a positive number, and x any number. We **define**

$$\boxed{a^x = e^{x \log a}.}$$

Thus

$$a^{\sqrt{2}} = e^{\sqrt{2} \log a}.$$

If we put $u = x \log a$ and use $\log e^u = u$, we find the formula

$$\boxed{\log a^x = x \log a.}$$

For instance,

$$\log 3^{\sqrt{2}} = \sqrt{2} \log 3.$$

Having made the general definition of a^x, we must be sure that in those cases when we have a preconceived idea of what a^x should be, for instance when $x = n$ is a positive integer, then

$$e^{n \log a} \quad \text{is the product of } a \text{ with itself } n \text{ times.}$$

For instance, take $x = 2$. Then

$$e^{2 \log a} = e^{\log a + \log a} = e^{\log a} e^{\log a} = a \cdot a,$$

$$e^{3 \log a} = e^{\log a + \log a + \log a} = e^{\log a} e^{\log a} e^{\log a} = a \cdot a \cdot a$$

and so forth. For any positive integer n we have

$$\begin{aligned} e^{n \log a} &= e^{\log a + \log a + \cdots + \log a} \\ &= e^{\log a} e^{\log a} \ldots e^{\log a} \\ &= a \cdot a \cdots a \qquad \text{(product taken } n \text{ times).} \end{aligned}$$

Therefore, if n is a positive integer, $e^{n \log a}$ means the product of a with itself n times.

Similarly,

$$\begin{aligned} (e^{(1/n) \log a})^n &= e^{(1/n) \log a} e^{(1/n) \log a} \ldots e^{(1/n) \log a} \qquad \text{(product taken } n \text{ times)} \\ &= e^{(1/n) \log a + (1/n) \log a + \cdots + (1/n) \log a} \\ &= e^{\log a} \\ &= a. \end{aligned}$$

Hence the n-th power of $e^{(1/n)\log a}$ is equal to a, so

$$e^{(1/n)\log a} \quad \textit{is the } n\text{-th root of } a.$$

This shows that $e^{x\log a}$ is what we expect when x is a positive integer or a fraction.

Next we prove other properties of the function a^x. First:

$$\boxed{a^0 = 1.}$$

Proof. By definition, $a^0 = e^{0\log a} = e^0 = 1$.

For all numbers x, y we have

$$\boxed{a^{x+y} = a^x a^y.}$$

Proof. We start with the right-hand side to get:

$$\begin{aligned} a^x a^y = e^{x\log a}e^{y\log a} &= e^{x\log a + y\log a} \\ &= e^{(x+y)\log a} \\ &= a^{x+y}. \end{aligned}$$

This proves the formula.

For all numbers x, y,

$$\boxed{(a^x)^y = a^{xy}.}$$

Proof.

$$\begin{aligned} (a^x)^y &= e^{y\log a^x} && (\text{because } u^y = e^{y\log u} \quad \text{for } u > 0) \\ &= e^{yx\log a} && (\text{because } \log a^x = x\log a) \\ &= a^{xy} && (\text{because } a^t = e^{t\log a}, \text{ with the} \\ & && \ \text{special value } t = xy = yx) \end{aligned}$$

thus proving the desired property.

At this point we have recovered all five properties of the general exponential function which were used in §1, §2, and §3.

VIII, APPENDIX. EXERCISE

Suppose you did not know anything about the exponential and log functions. You are given a function E such that

$$E'(x) = E(x) \qquad \text{for all numbers } x, \qquad \text{and} \qquad E(0) = 1.$$

Prove:

(a) $E(x) \neq 0$ for all x. [*Hint*: Differentiate the product $E(x)E(-x)$ to show that this product is constant. Using $E(0) = 1$, what is this constant?]

(b) Let f be a function such that $f'(x) = f(x)$ for all x. Show that there exists a constant C such that $f(x) = CE(x)$.

(c) For all numbers u, v the function E satisfies

$$E(u + v) = E(u)E(v).$$

[*Hint*: Fix the number u and let $f(x) = E(u + x)$. Then apply (b).]

Part Three

Integration

CHAPTER IX

Integration

In this chapter, we solve, more or less simultaneously, the following problems:

(1) Given a function $f(x)$, find a function $F(x)$ such that

$$F'(x) = f(x).$$

This is the inverse of differentiation, and is called integration.

(2) Given a function $f(x)$ which is $\geqq 0$, give a definition of the area under the curve $y = f(x)$ which does not appeal to geometric intuition.

Actually, in this chapter, we give the ideas behind the solutions of our two problems. The techniques which allow us to compute effectively when specific data are given will be postponed to the next chapter.

In carrying out (2) we shall follow an idea of Archimedes. It is to approximate the function f by horizontal functions, and the area under f by the sum of little rectangles.

IX, §1. THE INDEFINITE INTEGRAL

Let $f(x)$ be a function defined over some interval.

Definition. An **indefinite integral** for f is a function F such that

$$F'(x) = f(x) \qquad \text{for all } x \text{ in the interval.}$$

If $G(x)$ is another indefinite integral of f, then $G'(x) = f(x)$ also. Hence the derivative of the difference $F - G$ is 0:

$$(F - G)'(x) = F'(x) - G'(x) = f(x) - f(x) = 0.$$

Consequently, by Corollary 3.3 of Chapter V, there is a constant C such that

$$F(x) = G(x) + C$$

for all x in the interval.

Example 1. An indefinite integral for $\cos x$ would be $\sin x$. But $\sin x + 5$ is also an indefinite integral for $\cos x$.

Example 2. $\log x$ is an indefinite integral for $1/x$. So is $\log x + 10$ or $\log x - \pi$.

In the next chapter, we shall develop techniques for finding indefinite integrals. Here, we merely observe that every time we prove a formula for a derivative, it has an analogue for the integral.

It is customary to denote an indefinite integral of a function f by

$$\int f \quad \text{or} \quad \int f(x)\, dx.$$

In this second notation, the dx is meaningless by itself. It is only the full expression $\int f(x)\, dx$ which is meaningful. When we study the method of substitution in the next chapter, we shall get further confirmation for the practicality of our notation.

We shall now make a table of some indefinite integrals, using the information which we have obtained about derivatives.

Let n be an integer, $n \neq -1$. Then we have

$$\boxed{\int x^n\, dx = \frac{x^{n+1}}{n+1}.}$$

If $n = -1$, then

$$\boxed{\int \frac{1}{x}\, dx = \log x.}$$

(This is true only in the interval $x > 0$.)

In the interval $x > 0$ we also have

$$\boxed{\int x^c \, dx = \frac{x^{c+1}}{c+1}}$$

for any number $c \neq -1$.

The following indefinite integrals are valid for all x.

$$\boxed{\begin{array}{ll} \displaystyle\int \cos x \, dx = \sin x, & \displaystyle\int \sin x \, dx = -\cos x, \\ \displaystyle\int e^x \, dx = e^x, & \displaystyle\int \frac{1}{1+x^2} \, dx = \arctan x. \end{array}}$$

Finally, for $-1 < x < 1$, we have

$$\int \frac{1}{\sqrt{1-x^2}} \, dx = \arcsin x.$$

In practice, one frequently omits mentioning over what interval the various functions we deal with are defined. However, in any specific problem, one has to keep it in mind. For instance, if we write

$$\int x^{-1/3} \, dx = \tfrac{3}{2} \cdot x^{2/3},$$

this is valid for $x > 0$ and is also valid for $x < 0$. But 0 cannot be in any interval of definition of our function. Thus we could have

$$\int x^{-1/3} \, dx = \tfrac{3}{2} \cdot x^{2/3} + 5$$

when $x < 0$ and

$$\int x^{-1/3} \, dx = \tfrac{3}{2} \cdot x^{2/3} - 2$$

when $x > 0$. Any other constants besides 5 and -2 could also be used.

We agree throughout that indefinite integrals are defined only over intervals. Thus in considering the function $1/x$, we have to consider **separately** the cases $x > 0$ and $x < 0$. For $x > 0$, we have already remarked that $\log x$ is an indefinite integral. It turns out that for the interval $x < 0$ we can also find an indefinite integral, and in fact we have

$$\boxed{\int \frac{1}{x}\, dx = \log(-x) \qquad \text{for} \quad x < 0.}$$

Observe that when $x < 0$, $-x$ is positive, and thus $\log(-x)$ is meaningful. The derivative of $\log(-x)$ is equal to $1/x$, by the chain rule, namely, let $u = -x$. Then $du/dx = -1$, and

$$\frac{d \log(-x)}{dx} = \frac{1}{-x}(-1) = \frac{1}{x}.$$

For $x < 0$, any other indefinite integral is given by

$$\log(-x) + C,$$

where C is a constant.

It is sometimes stated that in all cases,

$$\int \frac{1}{x}\, dx = \log|x| + C.$$

With our conventions, we do not attribute any meaning to this, because our functions are not defined over intervals (the missing point 0 prevents this). In any case, the formula would be **false**. Indeed, for $x < 0$ we have

$$\int \frac{1}{x}\, dx = \log|x| + C_1,$$

and for $x > 0$ we have

$$\int \frac{1}{x}\, dx = \log|x| + C_2.$$

However, the two constants need not be equal, and hence we cannot write

$$\int \frac{1}{x}\, dx = \log|x| + C$$

in all cases. This formula is true only over an interval not containing 0.

IX, §1. EXERCISES

Find indefinite integrals for the following functions:

1. $\sin 2x$
2. $\cos 3x$
3. $\dfrac{1}{x+1}$
4. $\dfrac{1}{x+2}$

(In these last two problems, specify the intervals over which you find an indefinite integral.)

IX, §2. CONTINUOUS FUNCTIONS

Definition. Let $f(x)$ be a function. We shall say that f is **continuous** if

$$\lim_{h \to 0} f(x+h) = f(x)$$

for all x at which the function is defined.

It is understood that in taking the limit, only values of h for which $f(x+h)$ is defined are considered. For instance, if f is defined on an interval

$$a \leqq x \leqq b$$

(assuming $a < b$), then we would say that f is continuous at a if

$$\lim_{\substack{h \to 0 \\ h > 0}} f(a+h) = f(a).$$

We cannot take $h < 0$, since the function would not be defined for $a + h$ if $h < 0$.

Geometrically speaking, a function is continuous if there is no break in its graph. All differentiable functions are continuous. We have already remarked this fact, because if a quotient

$$\frac{f(x+h) - f(x)}{h}$$

has a limit, then the numerator $f(x+h) - f(x)$ must approach 0, because

$$\begin{aligned}\lim_{h \to 0} f(x+h) - f(x) &= \lim_{h \to 0} h \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \to 0} h \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = 0.\end{aligned}$$

The following are graphs of functions which are not continuous.

In Fig. 1, we have the graph of a function like

$$f(x) = -1 \qquad \text{if} \quad x \leqq 0,$$
$$f(x) = 1 \qquad \text{if} \quad x > 0.$$

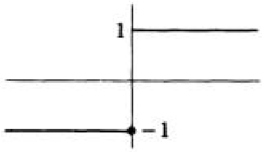


Figure 1

We see that

$$f(a + h) = f(h) = 1$$

for all $h > 0$. Hence

$$\lim_{\substack{h \to 0 \\ h > 0}} f(a + h) = 1,$$

which is unequal to $f(0)$.

A similar phenomenon occurs in Fig. 2 where there is a break. (Cf. Example 6 of Chapter III, §2.)

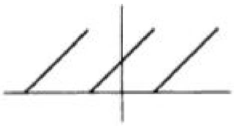


Figure 2

IX, §3. AREA

Let $a < b$ be two numbers, and let $f(x)$ be a continuous function defined on the interval $a \leqq x \leqq b$. This closed interval is denoted by $[a, b]$.

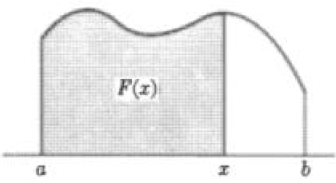
We wish to find a function $F(x)$ which is differentiable in this interval, and such that

$$F'(x) = f(x).$$

In this section, we appeal to our geometric intuition concerning area. *We assume that $f(x) \geqq 0$ for all x in the interval.* We let:

$F(x) =$ numerical measure of the area under the graph of f between a and x.

The following figure illustrates this.



We thus have $F(a) = 0$. The area between a and a is 0.

Theorem 3.1. *The function $F(x)$ is differentiable, and its derivative is $f(x)$.*

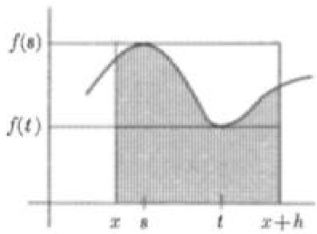
Proof. Since we defined F geometrically, we shall have to argue geometrically.

We have to consider the Newton quotient

$$\frac{F(x + h) - F(x)}{h}.$$

Suppose first that x is unequal to the end point b, and also suppose that we consider only values of $h > 0$.

Then $F(x + h) - F(x)$ is the area between x and $x + h$. A picture may look like this.



The shaded area represents $F(x + h) - F(x)$.

We let s be a point in the closed interval $[x, x + h]$ which is a maximum for our function f *in that small interval.* We let t be a point in the same closed interval which is a minimum for f in that interval. Thus

$$f(t) \leqq f(u) \leqq f(s)$$

for all u satisfying

$$x \leqq u \leqq x + h.$$

(We are forced to use another letter, u, since x is already being used.)

The area under the curve between x and $x + h$ is bigger than the area of the small rectangle in the figure above, i.e. the rectangle having base h and height $f(t)$.

The area under the curve between x and $x + h$ is smaller than the area of the big rectangle, i.e. the rectangle having base h and height $f(s)$.

This gives us

$$h \cdot f(t) \leqq F(x + h) - F(x) \leqq h \cdot f(s).$$

Dividing by the positive number h yields

$$f(t) \leqq \frac{F(x + h) - F(x)}{h} \leqq f(s).$$

Since s, t are between x and $x + h$, as h approaches 0 both $f(s)$ and $f(t)$ approach $f(x)$. Hence the Newton quotient for F is squeezed between two numbers which approach $f(x)$. It must therefore approach $f(x)$ itself, and we have proved Theorem 3.1, when $h > 0$.

The proof is essentially the same as the proof which we used to get the derivative of $\log x$. The only difference in the present case is that we pick a maximum and a minimum without being able to give an explicit value for it, the way we could for the function $1/x$. Otherwise, there is no difference in the arguments.

If $x = b$, we look at negative values for h. The argument in that case is entirely similar to the one we have written down in detail, and we find again that the Newton quotient of F is squeezed between $f(s)$ and $f(t)$. We leave it as an exercise.

We now know that if $F(x)$ denotes the area under the graph of f between a and x then

$$F'(x) = f(x).$$

We can compute the area in practice by the following property:

Let G be any function on the interval $a \leqq x \leqq b$ such that

$$G'(x) = f(x).$$

Suppose $f(x) \geqq 0$ for all x. Then the area under the graph of f between $x = a$ and $x = b$ is equal to

$$G(b) - G(a).$$

Proof. Since $F'(x) = G'(x)$ for all x, the two functions F and G have the same derivative on the interval. Hence there is a constant C such that

$$F(x) = G(x) + C \qquad \text{for all } x.$$

Let $x = a$. We get

$$0 = F(a) = G(a) + C.$$

This shows that $C = -G(a)$. Hence letting $x = b$ yields

$$F(b) = G(b) - G(a).$$

Thus the area under the curve between a and b is $G(b) - G(a)$. This is very useful to know in practice, because we can usually guess the function G.

Example 1. Find the area under the curve $y = x^2$ between $x = 1$ and $x = 2$.

Let $f(x) = x^2$. If $G(x) = x^3/3$ then $G'(x) = f(x)$. Hence the area under the curve between 1 and 2 is

$$G(2) - G(1) = \frac{2^3}{3} - \frac{1^3}{3} = \frac{7}{3}.$$

Example 2. Find the area under one arch of the function $\sin x$.

We have to find the area under the curve between 0 and π. Let

$$G(x) = -\cos x.$$

Then $G'(x) = \sin x$. Hence the area is

$$\begin{aligned} G(\pi) - G(0) &= -\cos \pi - (-\cos 0) \\ &= -(-1) + 1 \\ &= 2. \end{aligned}$$

Note how remarkable this is. The arch of the sine curve going from 0 to π seems to be a very irrational curve, and yet the area turns out to be the integer 2!

IX, §3. EXERCISES

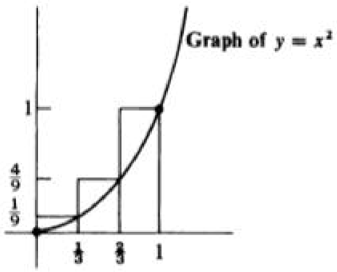
Find the area under the given curves between the given bounds.

1. $y = x^3$ between $x = 1$ and $x = 5$.
2. $y = x$ between $x = 0$ and $x = 2$.
3. $y = \cos x$, one arch.
4. $y = 1/x$ between $x = 1$ and $x = 2$.
5. $y = 1/x$ between $x = 1$ and $x = 3$.
6. $y = x^4$ between $x = -1$ and $x = 1$.
7. $y = e^x$ between $x = 0$ and $x = 1$.

IX, §4. UPPER AND LOWER SUMS

To show the existence of the integral, we use the idea of approximating our curves by constant functions.

Example. Consider the function $f(x) = x^2$. Suppose that we want to find the area between its graph and the x-axis, from $x = 0$ to $x = 1$. We cut up the interval $[0, 1]$ into smaller intervals, and approximate the function by constant functions, as shown on the next figure.



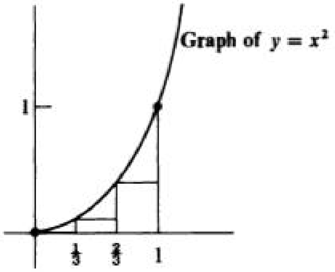
We have used three intervals, of length 1/3, and on each of these intervals, we take the constant function whose value is the square of the right end point of the interval. These values are, respectively,

$$f(1/3) = 1/9, \qquad f(2/3) = 4/9, \qquad f(1) = 1.$$

Thus we obtain three rectangles, lying above the curve $y = x^2$. Each rectangle has a base of length 1/3. The sum of the areas of these rectangles is equal to

$$\tfrac{1}{3}(\tfrac{1}{9} + \tfrac{4}{9} + 1) = \tfrac{14}{27}.$$

We could also have taken rectangles lying below the curve, using the values of f at the left end points of the intervals. The picture is as follows:



The heights of the three rectangles thus obtained are, respectively,

$$f(0) = 0, \qquad f(1/3) = 1/9, \qquad f(2/3) = 4/9.$$

The sum of their areas is equal to

$$\tfrac{1}{3}(0 + \tfrac{1}{9} + \tfrac{4}{9}) = \tfrac{5}{27}.$$

Thus we know that the area under the curve $y = x^2$ between $x = 0$ and $x = 1$ lies between 5/27 and 14/27. This is not a very good approximation to this area, but we can get a better approximation by using smaller intervals, say of lengths 1/4, or 1/5, or 1/6, or in general $1/n$. Let us write down the approximation with intervals of length $1/n$. The end points of the intervals will then be

$$0, \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \ldots, \frac{n-1}{n}, \frac{n}{n} = 1.$$

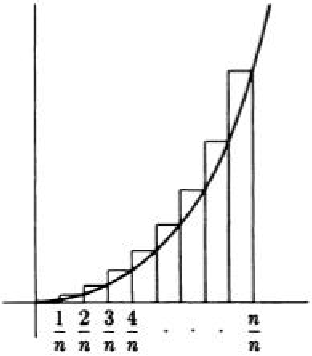
If we approximate the curve from above, then we get rectangles of heights respectively equal to

$$f\left(\frac{1}{n}\right) = \frac{1}{n^2}, \quad f\left(\frac{2}{n}\right) = \frac{2^2}{n^2}, \quad \ldots, \quad f\left(\frac{n}{n}\right) = \frac{n^2}{n^2}.$$

The general term for height of such a rectangle is

$$f\left(\frac{k}{n}\right) = \frac{k^2}{n^2} \qquad \text{for} \quad k = 1, 2, \ldots, n.$$

We have drawn these rectangles on the next figure.



We see pictorially that the approximation to the curve is already much better. The area of each rectangle is equal to

$$\frac{1}{n}\cdot\frac{k^2}{n^2}, \qquad k = 1,\ldots,n,$$

because it is equal to the base times the altitude. The sum of these areas is equal to

$$\frac{1}{n^3}\sum_{k=1}^{n} k^2 = \frac{1}{n^3}\,(1 + 2^2 + 3^2 + \cdots + n^2).$$

Such a sum is called an **upper sum**, because we took the maximum of the function x^2 over each interval $\left[\frac{k-1}{n}, \frac{k}{n}\right]$. When we take n larger and larger, it is plausible that such sums will approximate the area under the graph of x^2 between 0 and 1. In any case, this upper sum is bigger than the area.

We shall now write down in general the sums which approximate the area under a curve. Note that we can take rectangles lying above the curve or below the curve, thus giving rise to upper sums and lower sums.

Let a, b be two numbers, with $a \leqq b$. Let f be a continuous function in the interval $a \leqq x \leqq b$.

Definition. A partition of the interval $[a, b]$ is a sequence of numbers

$$a = x_0 \leqq x_1 \leqq x_2 \leqq \cdots \leqq x_n = b \quad \text{also written} \quad \{a = x_0,\ldots,x_n = b\}$$

between a and b, such that $x_i \leqq x_{i+1}$ $(i = 0, 1,\ldots,n-1)$. For instance, we could take just two numbers,

$$x_0 = a \qquad \text{and} \qquad x_1 = b.$$

This will be called the **trivial partition**.

A partition divides our interval in a lot of smaller intervals $[x_i, x_{i+1}]$.

Given any number between a and b, in addition to $x_0,\dots,x_n$, we can add it to the partition to get a new partition having one more small interval. If we add enough intermediate numbers to the partition, then the intervals can be made arbitrarily small.

Let f be a function defined on the interval

$$a \leqq x \leqq b$$

and continuous. If c_i is a point between x_i and x_{i+1}, then we form the sum

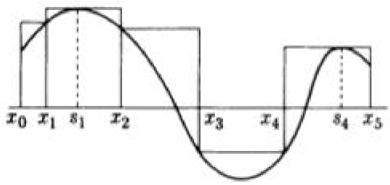
$$f(c_0)(x_1 - x_0) + f(c_1)(x_2 - x_1) + \cdots + f(c_{n-1})(x_n - x_{n-1}).$$

Such a sum will be called a **Riemann sum**. Each value $f(c_i)$ can be viewed as the height of a rectangle, and each $(x_{i+1} - x_i)$ can be viewed as the length of the base.

Let s_i be a point between x_i and x_{i+1} such that f has a **maximum** in this small interval $[x_i, x_{i+1}]$ at s_i. In other words,

$$f(x) \leqq f(s_i) \qquad \text{for} \quad x_i \leqq x \leqq x_{i+1}.$$

The rectangles then look like those in the next figure. In the figure, s_0 happens to be equal to x_1, $s_1 = s_1$ as shown, $s_2 = x_2$, $s_3 = x_4$, $s_4 = s_4$ as shown.



The main idea which we are going to carry out is that, as we make the intervals of our partitions smaller and smaller, the sum of the areas of the rectangles will approach a limit, and this limit can be used to define the area under the curve.

If P is the partition given by the numbers

$$x_0 \leqq x_1 \leqq x_2 \leqq \cdots \leqq x_n,$$

then the sum

$$f(s_0)(x_1 - x_0) + f(s_1)(x_2 - x_1) + \cdots + f(s_{n-1})(x_n - x_{n-1})$$

will be called the **upper sum** associated with the function f, and the partition P of the interval $[a, b]$. We shall denote it by the symbols

$$U_a^b(P, f) \qquad \text{or simply} \qquad U(P, f).$$

Observe, however, that when $f(x)$ becomes negative, the value $f(s_i)$ may be negative. Thus the corresponding rectangle gives a negative contribution

$$f(s_i)(x_{i+1} - x_i)$$

to the sum. Also, it is tiresome to write the sum by repeating each term, and so we shall use the abbreviation

$$\sum_{i=0}^{n-1} f(s_i)(x_{i+1} - x_i)$$

to mean the sum when i ranges from 0 to $n - 1$ of the terms $f(s_i)(x_{i+1} - x_i)$. Thus we have:

Definition. The **upper sum of** f **with respect to the partition** is

$$U_a^b(P, f) = \sum_{i=0}^{n-1} f(s_i)(x_{i+1} - x_i),$$

where $f(s_i)$ is the maximum of f on the interval $[x_i, x_{i+1}]$. Note that the **indices** i range from 0 to $n - 1$. Thus the sum is taken for $i = 0, \dots, n - 1$.

By definition, we let

$$\max_{[x_i, x_{i+1}]} f = f(s_i) = \text{maximum of } f \text{ on the interval} \quad x_i \leqq x \leqq x_{i+1}.$$

Then the sum could also be written with the notation

$$\boxed{U_a^b(P, f) = \sum_{i=0}^{n-1} \left(\max_{[x_i, x_{i+1}]} f\right)(x_{i+1} - x_i).}$$

Instead of taking a maximum s_i in the interval $[x_i, x_{i+1}]$ we could have taken a minimum. Let t_i be a point in this interval, such that

$$f(t_i) \leqq f(x) \qquad \text{for} \quad x_i \leqq x \leqq x_{i+1}.$$

We call the sum

$$f(t_0)(x_1 - x_0) + f(t_1)(x_2 - x_1) + \cdots + f(t_{n-1})(x_n - x_{n-1})$$

the **lower sum** associated with the function f, and the partition P of the interval $[a, b]$. The lower sum will be denoted by

$$L_a^b(P, f) \qquad \text{or simply} \qquad L(P, f).$$

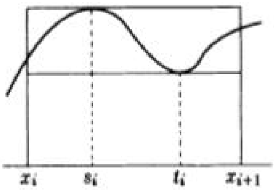
Thus we have the

Definition. The **lower sum of f with respect to the partition** is the sum

$$L_a^b(P, f) = \sum_{i=0}^{n-1} f(t_i)(x_{i+1} - x_i),$$

where $f(t_i)$ is the minimum of f on the interval $[x_i, x_{i+1}]$.

On the next figure, we have drawn a typical term of the sum.



We can rewrite this lower sum using a notation similar to that used in the upper sum. Namely, we let

$$\begin{aligned}\min_{[x_i, x_{i+1}]} f &= \text{minimum of } f \text{ on the interval } [x_i, x_{i+1}] \\ &= f(t_i).\end{aligned}$$

Then

$$\boxed{L_a^b(P, f) = \sum_{i=0}^{n-1} \left(\min_{[x_i, x_{i+1}]} f \right)(x_{i+1} - x_i).}$$

For all numbers x in the interval $[x_i, x_{i+1}]$ we have

$$f(t_i) \leqq f(x) \leqq f(s_i).$$

Since $x_{i+1} - x_i$ is $\geqq 0$, it follows that each term of the lower sum is less than or equal to each term of the upper sum. Therefore

$$L_a^b(P, f) \leqq U_a^b(P, f).$$

Furthermore, any Riemann sum taken with points c_i (which are not necessarily maxima or minima) is between the lower and upper sum.

Example. Let $f(x) = x^2$ and let the interval be $[0, 1]$. Write out the upper and lower sums for the partition consisting of $\{0, \frac{1}{2}, 1\}$.

The minimum of the function in the interval $[0, \frac{1}{2}]$ is at 0, and $f(0) = 0$. The minimum of the function in the interval $[\frac{1}{2}, 1]$ is at $\frac{1}{2}$ and $f(\frac{1}{2}) = \frac{1}{4}$. Hence the lower sum is

$$L_0^1(P, f) = f(0)(\tfrac{1}{2} - 0) + f(\tfrac{1}{2})(1 - \tfrac{1}{2}) = \tfrac{1}{4} \cdot \tfrac{1}{2} = \tfrac{1}{8}.$$

The maximum of the function in the interval $[0, \frac{1}{2}]$ is at $\frac{1}{2}$ and the maximum of the function in the interval $[\frac{1}{2}, 1]$ is at 1. Thus the upper sum is

$$U_0^1(P, f) = f(\tfrac{1}{2})(\tfrac{1}{2} - 0) + f(1)(1 - \tfrac{1}{2}) = \tfrac{1}{8} + \tfrac{1}{2} = \tfrac{5}{8}.$$

We have given a numerical value for the upper and lower sums. Unless we wanted to compare them explicitly, we could have left the answer in the shape that it has on the left-hand side of these equalities.

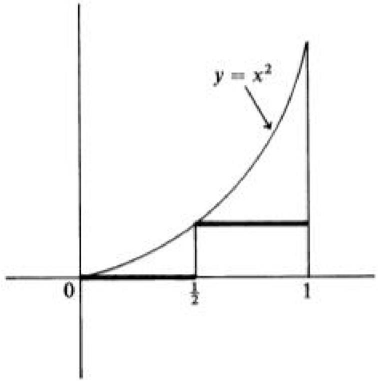


figure for the lower sum

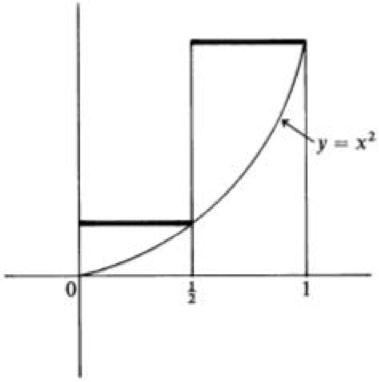


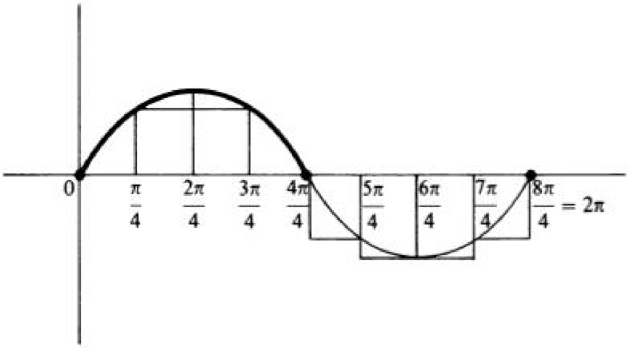
figure for the upper sum

Example. We shall write down the lower sums in another special case when the function is positive and negative in the interval.

Let $f(x) = \sin x$, let the interval be $[0, 2\pi]$, and let the partition be

$$P = \left\{0, \frac{\pi}{4}, \frac{2\pi}{4}, \frac{3\pi}{4}, \frac{4\pi}{4}, \frac{5\pi}{4}, \frac{6\pi}{4}, \frac{7\pi}{4}, \frac{8\pi}{4}\right\}.$$

We illustrate this on the next figure.



Each small interval of the partition has length $\pi/4$. Let us write down the lower sum:

$$\begin{aligned} L(P, f) &= 0\frac{\pi}{4} + \left(\sin\frac{\pi}{4}\right)\frac{\pi}{4} + \left(\sin\frac{3\pi}{4}\right)\frac{\pi}{4} + 0\frac{\pi}{4} \\ &\quad + \left(\sin\frac{5\pi}{4}\right)\frac{\pi}{4} + \left(\sin\frac{6\pi}{4}\right)\frac{\pi}{4} + \left(\sin\frac{6\pi}{4}\right)\frac{\pi}{4} + \left(\sin\frac{7\pi}{4}\right)\frac{\pi}{4} \\ &= \frac{1}{\sqrt{2}}\frac{\pi}{4} + \frac{1}{\sqrt{2}}\frac{\pi}{4} - \frac{1}{\sqrt{2}}\frac{\pi}{4} - \frac{\pi}{4} - \frac{\pi}{4} - \frac{1}{\sqrt{2}}\frac{\pi}{4}. \end{aligned}$$

Observe that the first term of the lower sum is 0 because the minimum of the function $\sin x$ on the interval $[0, \pi/4]$ is equal to 0. Similarly, the fourth term is also 0, and so can be omitted since $0 + A = A$ for all numbers A.

Also observe that:

$$\begin{aligned} \text{minimum of } \sin x \text{ on the interval } [5\pi/4, 6\pi/4] &= \sin 6\pi/4 \\ &= -1. \end{aligned}$$

With negative numbers, we have for instance

$$-1 = \sin 6\pi/4 < -\frac{1}{\sqrt{2}} = \sin 5\pi/4.$$

Thus the lower sum $L(P, f)$ contains positive terms and negative terms.

The negative terms represent minus the area of certain rectangles, as shown on the figure.

What happens to our sums when we add a new point to a partition? We shall see that the lower sum increases and the upper sum decreases.

Theorem 4.1. *Let f be a continuous function on the interval $[a, b]$. Let*

$$P = (x_0, \dots, x_n)$$

be a partition of $[a, b]$. Let $\bar{x}$ be any number in the interval, and let Q be the partition obtained from P by adding $\bar{x}$ to $(x_0, \dots, x_n)$. Then

$$L_a^b(P, f) \leqq L_a^b(Q, f) \leqq U_a^b(Q, f) \leqq U_a^b(P, f).$$

Proof. Let us look at the lower sums, for example. Suppose that our number $\bar{x}$ is between x_j and x_{j+1}:

$$x_j \leqq \bar{x} \leqq x_{j+1}.$$

When we form the lower sum for P, it will be the same as the lower sum for Q except that the term

$$f(t_j)(x_{j+1} - x_j)$$

will now be replaced by two terms. If u is a minimum for f in the interval between x_j and $\bar{x}$, and v is a minimum for f in the interval between $\bar{x}$ and x_{j+1}, then these two terms are

$$f(u)(\bar{x} - x_j) + f(v)(x_{j+1} - \bar{x}).$$

We can write $f(t_j)(x_{j+1} - x_j)$ in the form

$$f(t_j)(x_{j+1} - x_j) = f(t_j)(\bar{x} - x_j) + f(t_j)(x_{j+1} - \bar{x}).$$

Since $f(t_j) \leqq f(u)$ and $f(t_j) \leqq f(v)$ (because t_j was a minimum in the whole interval between x_j and x_{j+1}), it follows that

$$f(t_j)(x_{j+1} - x_j) \leqq f(u)(\bar{x} - x_j) + f(v)(x_{j+1} - \bar{x}).$$

Thus when we replace the term in the sum for P by the two terms in the sum for Q, the value of the contribution of these two terms increases. Since all other terms are the same, our assertion is proved.

The assertion concerning the fact that the upper sum decreases is left as an exercise. The proof is very similar.

As a consequence of our theorem, we obtain:

Corollary 4.2. *Every lower sum is less than or equal to every upper sum.*

Proof. Let P and Q be two partitions. If we add to P all the points of Q and add to Q all the points of P, we obtain a partition R such that every point of P is a point of R and every point of Q is a point of R. Thus R is obtained by adding points to P and to Q. Consequently, we have the inequalities

$$L_a^b(P, f) \leqq L_a^b(R, f) \leqq U_a^b(R, f) \leqq U_a^b(Q, f).$$

This proves our assertion.

It is now a very natural question to ask whether **there is a unique number between the lower sums and the upper sums.** The answer is yes.

Theorem 4.3. *Let f be a continuous function on $[a, b]$. There exists a unique number which is greater than or equal to every lower sum and less than or equal to every upper sum.*

Definition. The definite integral of f between a and b

$$\boxed{\int_a^b f}$$

is the unique number which is greater than or equal to every lower sum, and less than or equal to every upper sum.

We shall also use notation like

$$\boxed{\int_a^b f(x)\,dx \qquad \text{or} \qquad \int_a^b f(t)\,dt.}$$

It does not matter which letter we use, but it should be the same letter in both occurrences, i.e. in $f(x)\,dx$ or $f(t)\,dt$, or $f(u)\,du$, etc.

We shall not give the details of the proof of Theorem 4.3, which are tedious. The technique involved will not be used anywhere else in the course.

There is another statement which it is illuminating to know. Let P be a partition,

$$x_0 \leqq x_1 \leqq \cdots \leqq x_n.$$

The maximum length of the intervals $[x_i, x_{i+1}]$ is called the **size** of the partition. For example, if we cut up the interval $[0, 1]$ into n small intervals of the same length $1/n$, then the size of this partition is $1/n$.

Theorem 4.4. *Let f be a continuous function on $[a, b]$. Then the lower sums $L_a^b(P, f)$ and the upper sums $U_a^b(P, f)$ come arbitrarily close to the integral*

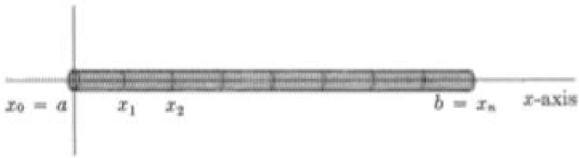
$$\int_a^b f$$

if the size of the partition P is sufficiently small.

Again, we shall not prove this theorem. But intuitively, it tells us that the upper and lower sums are good approximations to the integral when the size of the partition is taken sufficiently small.

Example. We give a physical example illustrating the application of upper and lower sums, relating density to mass.

Consider an interval $[a, b]$ with $0 \leqq a < b$. We think of this interval as a rod, and let f be a continuous positive function defined on this interval. We interpret f as a density on the rod, so that $f(x)$ is the density at x.



Given

$$a \leqq c \leqq d \leqq b,$$

we denote by $M_c^d(f)$ the mass of the rod between c and d, corresponding to the given density f. We wish to determine a mathematical notion to represent $M_c^d(f)$. If f is a constant density, with constant value $K \geqq 0$ on $[c, d]$, then the mass $M_c^d(f)$ should be $K(d - c)$. On the other hand, if g is another density such that

$$f(x) \leqq g(x),$$

then certainly we should have $M_c^d(f) \leqq M_c^d(g)$. In particular, if k, K are constants $\geqq 0$ such that

$$k \leqq f(x) \leqq K$$

for x in the interval $[c, d]$, then the mass should satisfy

$$k(d-c) \leqq M_c^d(f) \leqq K(d-c).$$

Finally, the mass should be additive, that is the mass of two disjoint pieces should be the sum of the mass of the pieces. In particular,

$$M_a^c(f) + M_c^d(f) = M_a^d(f).$$

We shall now see that the mass of the rod is given by the integral of the density, that is

$$\boxed{M_a^b(f) = \int_a^b f(x)\,dx.}$$

Let P be a partition of the interval $[a, b]$:

$$a = x_0 \leqq x_1 \leqq x_2 \leqq \cdots \leqq x_n = b.$$

Let $f(t_i)$ be the minimum for f on the small interval $[x_i, x_{i+1}]$, and let $f(s_i)$ be the maximum for f on this same small interval. Then the mass of each piece of the rod between x_i and x_{i+1} satisfies the inequality

$$f(t_i)(x_{i+1} - x_i) \leqq M_{x_i}^{x_{i+1}}(f) \leqq f(s_i)(x_{i+1} - x_i).$$

Adding these together, we find

$$\sum_{i=0}^{n-1} f(t_i)(x_{i+1} - x_i) \leqq M_a^b(f) \leqq \sum_{i=0}^{n-1} f(s_i)(x_{i+1} - x_i).$$

The expressions on the left and right are lower and upper sums for the integral, respectively. Since the integral is the unique number between the lower sums and upper sum, it follows that

$$M_a^b(f) = \int_a^b f(x)\,dx,$$

as we wanted to show.

IX, §4. EXERCISES

Write out the lower and upper sums for the following functions and intervals. Use a partition such that the length of each small interval is (a) $\frac{1}{2}$, (b) $\frac{1}{3}$, (c) $\frac{1}{4}$, (d) $1/n$.

1. $f(x) = x^2$ in the interval $[1, 2]$.
2. $f(x) = 1/x$ in the interval $[1, 3]$.

3. $f(x) = x$ in the interval $[0, 2]$.
4. $f(x) = x^2$ in the interval $[0, 2]$.
5. Let $f(x) = 1/x$ and let the interval be $[1, 2]$. Let n be a positive integer. Write out the upper and lower sum, using the partition such that the length of each small interval is $1/n$.
6. Using the definition of a definite integral, prove that

$$\frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{n+n} \leqq \log 2 \leqq \frac{1}{n} + \frac{1}{n+1} + \cdots + \frac{1}{2n-1}.$$

7. Let $f(x) = \log x$. Let n be a positive integer. Write out the upper and lower sums, using the partition of the interval between 1 and n consisting of the integers from 1 to n, i.e. the partition $(1, 2, \ldots, n)$.

IX, §5. THE FUNDAMENTAL THEOREM

The integral satisfies two basic properties which are very similar to those satisfied by area. We state them explicitly.

Property 1. *If M, m are two numbers such that*

$$m \leqq f(x) \leqq M$$

for all x in the interval $[b, c]$, then

$$m(c - b) \leqq \int_b^c f \leqq M(c - b).$$

Property 2. *We have*

$$\int_a^b f + \int_b^c f = \int_a^c f.$$

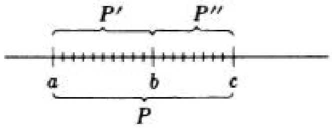
We shall not give the details of the proofs of these properties but we make some comments which we hope make them clear.

For Property 1, suppose we want to verify the inequality on the left-hand side. We may take the trivial partition of the interval $[b, c]$ consisting just of this interval. Then a lower sum is certainly $\geqq m(c - b)$. Since the lower sums increase when we take a finer partition, and since the lower sums are at most equal to the integral, we see that the left-hand inequality

$$m(c - b) \leqq \int_b^c f$$

is true. The right-hand inequality of Property 1 is proved in the same way.

For Property 2, suppose that $a \leqq b \leqq c$. Let P be a partition of sufficiently small size, such that the lower sum $L_a^c(P, f)$ approximates the integral $\int_a^c f$ very closely. The point b may not be in this partition. We may take a finer partition by inserting this point b, as shown on the figure.



Then P together with b form partitions P' and P'' of the intervals $[a, b]$ and $[b, c]$. If P has sufficiently small size, then P' and P'' have small size, and the lower sums

$$L_a^b(P', f) \qquad \text{and} \qquad L_b^c(P'', f)$$

give good approximations to the integrals

$$\int_a^b f \qquad \text{and} \qquad \int_b^c f,$$

respectively. But we have

$$L_a^c(P', P'', f) = L_a^b(P', f) + L_b^c(P'', f).$$

Since $L_a^c(P', P'', f)$ is an approximation of the integral

$$\int_a^c f,$$

one can see by passing to a limit that

$$\int_a^c f = \int_a^b f + \int_b^c f.$$

Property 2 was formulated when $a < b < c$. We now want to formulate it when a, b, c are taken in any order. For this, suppose a, b are numbers in an interval where f is continuous, and $b < a$. We **define**

$$\boxed{\int_a^b f = -\int_b^a f.}$$

Then we have **Property 2 in general**:

Let a, b, c be three numbers in an interval where f is continuous. Then

$$\int_a^c f = \int_a^b f + \int_b^c f.$$

Proof. We have to distinguish cases. Suppose for instance that $b < a < c$. Then by the original property, for this ordering we get

$$\int_b^c = \int_b^a + \int_a^c = -\int_a^b + \int_a^c \qquad \text{by definition.}$$

Adding $\int_a^b$ to both sides proves the desired relation. All other cases can be proved similarly.

Theorem 5.1. *Let f be continuous on an interval $[a, b]$. Let*

$$F(x) = \int_a^x f.$$

Then F is differentiable and its derivative is

$$F'(x) = f(x).$$

Proof. We have to form the Newton quotient

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{h}\left(\int_a^{x+h} f - \int_a^x f\right),$$

and see if it approaches a limit as $h \to 0$. (If $x = a$, then it is to be understood that $h > 0$, and if $x = b$, then $h < 0$. If $a < x < b$, then h may be positive or negative. The proof then shows that f is right differentiable at a and left differentiable at b.)

Assume for the moment that $h > 0$. By Property 2, applied to the numbers a, x, $x + h$ we conclude that our Newton quotient is equal to

$$\frac{1}{h}\left(\int_a^x f + \int_x^{x+h} f - \int_a^x f\right) = \frac{1}{h}\int_x^{x+h} f.$$

This reduces our investigation of the Newton quotient to the interval between x and $x + h$.

Let s be a point between x and $x + h$ such that f reaches a maximum in this small interval $[x, x + h]$ and let t be a point in this interval such that f reaches a minimum. We let

$$m = f(t) \qquad \text{and} \qquad M = f(s)$$

and apply Property 1 to the interval $[x, x+h]$. We obtain

$$f(t)(x+h-x) \leqq \int_x^{x+h} f \leqq f(s)(x+h-x),$$

which we can rewrite as

$$f(t) \cdot h \leqq \int_x^{x+h} f \leqq f(s) \cdot h.$$

Dividing by the positive number h preserves the inequalities, and yields

$$f(t) \leqq \frac{\int_x^{x+h} f}{h} \leqq f(s).$$

Since s, t lie between x and $x+h$, we must have (by continuity)

$$\lim_{h \to 0} f(s) = f(x) \qquad \text{and} \qquad \lim_{h \to 0} f(t) = f(x).$$

Thus our Newton quotient is squeezed between two numbers which approach $f(x)$. It must therefore approach $f(x)$, and our theorem is proved when $h > 0$.

The argument when $h < 0$ is entirely similar. We omit it.

IX, §5. EXERCISES

1. Using Theorem 5.1 prove that if f is continuous on an open interval containing 0, then

$$\lim_{h \to 0} \frac{1}{h} \int_0^h f = f(0).$$

[*Hint*: Can you interpret the left-hand side as the limit of a Newton quotient?]

2. Let f be continuous on the interval $[a, b]$. Prove that there exists some number c in the interval such that

$$f(c)(b-a) = \int_a^b f(t)\, dt.$$

[*Hint*: Apply the mean value theorem to $\int_a^x f(t)\, dt = F(x)$.]

CHAPTER X

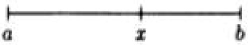
Properties of the Integral

This is a short chapter. It shows how the integral combines with addition and inequalities. There is no good formula for the integral of a product. The closest thing is integration by parts, which is postponed to the next chapter.

Connecting the integral with the derivative is what allows us to compute integrals. The fact that two functions having the same derivative differ by a constant is again exploited to the hilt.

X, §1. FURTHER CONNECTION WITH THE DERIVATIVE

Let f be a continuous function on some interval. Let a, b be two points of the interval such that $a < b$, and let F be a function which is differentiable on the interval and whose derivative is f.



By the fundamental theorem, the functions

$$F(x) \qquad \text{and} \qquad \int_a^x f$$

have the same derivative. Hence there is a constant C such that

$$\int_a^x f = F(x) + C \qquad \text{for } all \ x \text{ in the interval.}$$

What is this constant? If we put $x = a$, we get

$$0 = \int_a^a f = F(a) + C,$$

whence $C = -F(a)$. We also have

$$\int_a^b f = F(b) + C.$$

From this we obtain: *If* $dF/dx = f(x)$, *then*

$$\boxed{\int_a^b f = F(b) - F(a).}$$

This is extremely useful in practice, because we can usually guess the function F, and once we have guessed it, we can then compute the integral by means of this relation.

Furthermore, it is also practical to use the notation

$$F(x)\Big|_a^b = F(b) - F(a).$$

Remark. The argument which we gave to compute C shows that the value $F(b) - F(a)$ does not depend on the choice of function F such that $F'(x) = f(x)$. But you may want to see this another way. Suppose that $G'(x) = f(x)$ also, for all x in the interval. Then there is a constant C such that

$$G(x) = F(x) + C \qquad \text{for all } x \text{ in the interval.}$$

Then

$$\begin{aligned} G(b) - G(a) &= F(b) + C - [F(a) + C] \\ &= F(b) - F(a) \qquad \text{because } C \text{ cancels.} \end{aligned}$$

Finally, we shall usually call the **indefinite integral** such as

$$\int \sin x\,dx, \qquad \text{or} \qquad \int \frac{1}{1+x^2}\,dx$$

simply an **integral**, since the context makes clear what is meant. When we deal with a **definite integral**

$$\int_a^b$$

the numbers a and b are sometimes called the **lower limit** and **upper limit** respectively.

Example. We want to find the integral

$$\int_0^\pi \sin x\, dx.$$

Here we have $f(x) = \sin x$, and the indefinite integral is

$$\int \sin x\, dx = F(x) = -\cos x.$$

Hence

$$\int_0^\pi \sin x\, dx = -\cos x \Big|_0^\pi = -\cos \pi - (-\cos 0) = 2.$$

Example. Suppose we want to find

$$\int_1^3 x^2\, dx.$$

Let $F(x) = x^3/3$. Then $F'(x) = x^2$. Hence

$$\int_1^3 x^2\, dx = \frac{x^3}{3}\Big|_1^3 = \frac{27}{3} - \frac{1}{3} = \frac{26}{3}.$$

Example. Let us find

$$\int_0^1 \frac{1}{1+x^2}\, dx.$$

Since $d \arctan x/dx = 1/1 + x^2$, we have the indefinite integral

$$\int \frac{1}{1+x^2}\, dx = \arctan x.$$

Hence

$$\int_0^1 \frac{1}{1+x^2}\,dx = \arctan x \Big|_0^1$$

$$= \arctan 1 - \arctan 0$$

$$= \pi/4.$$

Example. Prove the inequality

$$\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \leqq \log n.$$

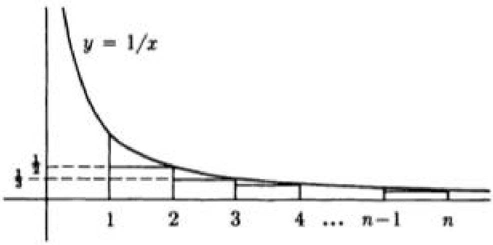
To do this, we try to identify the left-hand side with a lower sum, and the right-hand side with a corresponding integral. We have the indefinite integral

$$\log x = \int \frac{1}{x}\,dx.$$

We let $f(x) = 1/x$.

We let the interval $[a, b]$ be $[1, n]$, that is all x with $1 \leqq x \leqq n$.

We let the partition $P = \{1, 2, \ldots, n\}$ consist of the positive integers from 1 to n.



Then

$$L(P, f) = \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$$

because the length of the base of each rectangle is equal to 1. The value of the integral is

$$\int_1^n \frac{1}{x}\,dx = \log x \Big|_1^n = \log n - \log 1 = \log n.$$

Thus we obtain the desired inequality, because a lower sum is $\leqq$ the integral.

In working out and proving similar inequalities, you should give:

The function $f(x)$;
The interval $[a, b]$ and the value of the definite integral

$$\int_a^b f(x)\,dx;$$

The partition P of the interval $[a, b]$.

You should then identify the sum with a lower sum (or upper sum, as the case may be) with respect to the above data, thus obtaining a comparison with the integral of the desired type.

Example. By a similar method, one can give an inequality having considerable practical interest. Let $n!$ denote the product of the first n integers. Thus

$$n! = 1 \cdot 2 \cdot 3 \cdots n.$$

We have the first few:

$$\begin{aligned}
1! &= 1,\\
2! &= 1 \cdot 2 = 2,\\
3! &= 1 \cdot 2 \cdot 3 = 6,\\
4! &= 1 \cdot 2 \cdot 3 \cdot 4 = 6 \cdot 4 = 24,\\
5! &= 24 \cdot 5 = 120.
\end{aligned}$$

Exercise 10 will show you how to prove the inequality

$$\boxed{(n-1)! \leqq n^n e^{-n} e \leqq n!}$$

It is fun to work it out, so we don't do it here in the text.

The principle of these examples applies to comparing sums of functions with integrals, and the functions may be decreasing, as, for instance, the functions

$$\frac{1}{x}, \quad \frac{1}{x^{1/2}}, \quad \frac{1}{x^4}, \quad \text{etc.,}$$

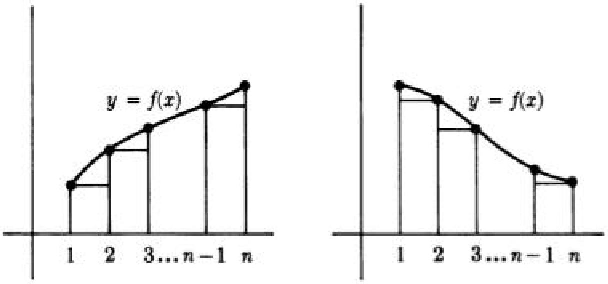
or they may be increasing, as for instance the functions

$$x, \quad x^2, \quad x^4, \quad x^{1/3}.$$

The graphs may look like this, say on the interval $[1, n]$ where n is a positive integer, and the partition

$$P = \{1, \dots, n\}$$

consists of the positive integers from 1 to n.



Increasing function Decreasing function

In such a case, the base of each rectangle has length 1. Hence we obtain inequalities, for f increasing:

$$f(1) + f(2) + \cdots + f(n-1) \leqq \int_1^n f(x)\,dx \leqq f(2) + \cdots + f(n),$$

and for f decreasing:

$$f(2) + f(3) + \cdots + f(n) \leqq \int_1^n f(x)\,dx \leqq f(1) + \cdots + f(n-1).$$

X, §1. EXERCISES

Find the following integrals:

1. $\displaystyle\int_1^2 x^5\,dx$
2. $\displaystyle\int_{-1}^1 x^{1/3}\,dx$
3. $\displaystyle\int_{-\pi}^{\pi} \sin x\,dx$
4. $\displaystyle\int_0^{\pi} \cos x\,dx$
5. Prove the following inequalities:

(a) $\dfrac{1}{2} + \dfrac{1}{3} + \cdots + \dfrac{1}{n} \leqq \log n \leqq \dfrac{1}{1} + \dfrac{1}{2} + \dfrac{1}{3} + \cdots + \dfrac{1}{n-1}$

(b) $\dfrac{1}{2^{1/2}} + \dfrac{1}{3^{1/2}} + \cdots + \dfrac{1}{n^{1/2}} \leqq 2(\sqrt{n} - 1)$

(c) $2(\sqrt{n} - 1) \leqq 1 + \dfrac{1}{2^{1/2}} + \dfrac{1}{3^{1/2}} + \cdots + \dfrac{1}{(n-1)^{1/2}}$

6. Prove the following inequalities:

(a) $1^2 + 2^2 + \cdots + (n-1)^2 \leqq \dfrac{n^3}{3} \leqq 1^2 + 2^2 + \cdots + n^2$

(b) $1^3 + 2^3 + \cdots + (n-1)^3 \leqq \dfrac{n^4}{4} \leqq 1^3 + 2^3 + \cdots + n^3$

(c) $1^{1/4} + 2^{1/4} + \cdots + (n-1)^{1/4} \leqq \frac{4}{5}n^{5/4} \leqq 1^{1/4} + 2^{1/4} + \cdots + n^{1/4}$

7. Give similar inequalities as in Exercise 6, for the sums:

(a) $\displaystyle\sum_{k=1}^{n-1} k^4$ (b) $\displaystyle\sum_{k=1}^{n-1} k^{1/3}$ (c) $\displaystyle\sum_{k=1}^{n} k^5$ (d) $\displaystyle\sum_{k=1}^{n} \frac{1}{k^4}$

8. Prove the inequalities

$$\frac{1}{n}\sum_{k=1}^{n}\frac{n^2}{n^2+k^2} \leqq \frac{\pi}{4} \leqq \frac{1}{n}\sum_{k=0}^{n-1}\frac{n^2}{n^2+k^2}.$$

$\left[\textit{Hint}: \text{Write } \dfrac{n^2}{n^2+k^2} = \dfrac{1}{1+k^2/n^2}.\right]$ What is the interval? What is the partition?

9. Prove the inequality

$$\frac{1}{n}\left[\left(\frac{1}{n}\right)^2 + \left(\frac{2}{n}\right)^2 + \cdots + \left(\frac{n-1}{n}\right)^2\right] \leqq \frac{1}{3}.$$

10. For this exercise, verify first that if we let

$$F(x) = x\log x - x$$

then $F'(x) = \log x$.

(a) Evaluate the integral

$$\int_1^n \log x\,dx.$$

(b) Compare this integral with the upper and lower sum associated with the partition $P = \{1, 2, \ldots, n\}$ of the interval $[1, n]$.

(c) In part (b), you will have found certain inequalities of the form

$$A \leqq B \leqq C.$$

Using the fact that

$$e^A \leqq e^B \leqq e^C,$$

prove the following inequality:

$$\boxed{(n-1)! \leqq n^n e^{-n} e \leqq n!.}$$

Here we denote by $n!$ the product of the first n integers.

X, §2. SUMS

Let $f(x)$ and $g(x)$ be two functions defined over some interval, and let $F(x)$ and $G(x)$ be (indefinite) integrals for f and g, respectively. This means $F'(x) = f(x)$ and $G'(x) = g(x)$. Since the derivative of a sum is the sum of the derivatives, we see that $F + G$ is an integral for $f + g$; in other words,

$$(F + G)'(x) = F'(x) + G'(x),$$

and therefore

$$\boxed{\int [f(x) + g(x)]\,dx = \int f(x)\,dx + \int g(x)\,dx.}$$

Similarly, let c be a number. The derivative of $cF(x)$ is $cf(x)$. Hence

$$\boxed{\int cf(x)\,dx = c \int f(x)\,dx.}$$

A constant can be taken in and out of an integral.

Example. Find the integral of $\sin x + 3x^4$.

We have

$$\int (\sin x + 3x^4)\,dx = \int \sin x\,dx + \int 3x^4\,dx$$

$$= -\cos x + 3x^5/5.$$

Any formula involving the indefinite integral yields a formula for the definite integral. Using the same notation as above, suppose we have to find

$$\int_a^b [f(x) + g(x)]\,dx.$$

We know that it is

$$[F(x) + G(x)]\Big|_a^b$$

which is equal to

$$F(b) + G(b) - F(a) - G(a).$$

Thus we get the formula

$$\int_a^b [f(x) + g(x)]\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx.$$

Similarly, for any constant c,

$$\int_a^b cf(x)\,dx = c\int_a^b f(x)\,dx.$$

Example. Find the integral

$$\int_0^\pi [\sin x + 3x^4]\,dx.$$

This (definite) integral is equal to

$$\begin{aligned} -\cos x + 3x^5/5\Big|_0^\pi &= -\cos\pi + 3\pi^5/5 - (-\cos 0 + 0) \\ &= 1 + 3\pi^5/5 + 1 \\ &= 2 + 3\pi^5/5. \end{aligned}$$

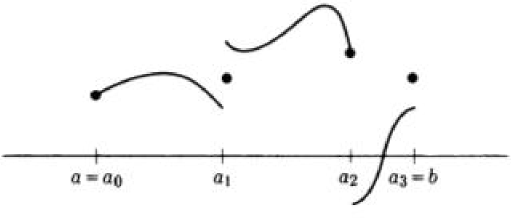
In some applications, one meets a slightly wider class of functions than continuous ones. Let f be a function defined on an interval $[a, b]$. We shall say that f is **piecewise continuous** on $[a, b]$ if there exist numbers

$$a = a_0 < a_1 < \cdots < a_n = b$$

and on each interval $[a_{i-1}, a_i]$ there is a continuous function f_i such that $f(x) = f_i(x)$ for $a_{i-1} < x < a_i$. If this is the case, then we define the integral of f from a to b to be the sum

$$\int_a^b f = \int_{a_0}^{a_1} f_1 + \int_{a_1}^{a_2} f_2 + \cdots + \int_{a_{n-1}}^{a_n} f_n.$$

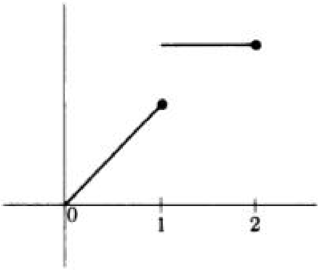
A piecewise continuous function may look like this:



Example. Let f be the function defined on the interval $[0, 2]$ by the conditions:

$$f(x) = x \quad \text{if} \quad 0 \leqq x \leqq 1,$$
$$f(x) = 2 \quad \text{if} \quad 1 < x \leqq 2.$$

The graph of f looks like this:



To find the integral of f between 0 and 2, we have

$$\int_0^2 f = \int_0^1 x\,dx + \int_1^2 2\,dx = \frac{x^2}{2}\Big|_0^1 + 2x\Big|_1^2$$
$$= \tfrac{1}{2} + (4 - 2) = \tfrac{5}{2}.$$

We can also find

$$\int_0^x f(t)\,dt \quad \text{for} \quad 0 \leqq x \leqq 2.$$

If $0 \leqq x \leqq 1$:

$$\int_0^x f(t)\,dt = \int_0^x t\,dt = \frac{x^2}{2}.$$

If $1 \leqq x \leqq 2$:

$$\int_0^x f(t)\,dt = \int_0^1 f(t)\,dt + \int_1^x f(t)\,dt$$
$$= \tfrac{1}{2} + \int_1^x 2\,dt = \tfrac{1}{2} + 2x - 2.$$

Example. Let $f(x)$ be defined for $0 \leqq x \leqq \pi$ by the formulas:

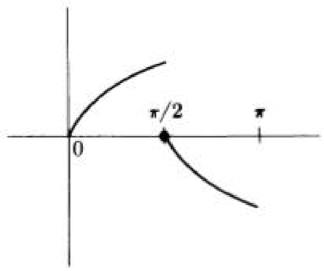
$$f(x) = \sin x \quad \text{if} \quad 0 \leqq x < \pi/2,$$
$$f(x) = \cos x \quad \text{if} \quad \pi/2 \leqq x \leqq \pi.$$

Then the integral of f from 0 to π is given by:

$$\int_0^\pi f = \int_0^{\pi/2} \sin x\, dx + \int_{\pi/2}^\pi \cos x\, dx$$

$$= -\cos x \Big|_0^{\pi/2} + \sin x \Big|_{\pi/2}^{\pi} = 0.$$

The graph of f looks like this:



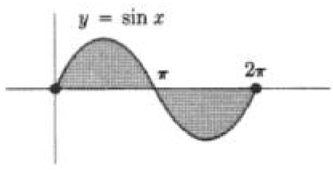
> *The integral of a function represents the area between the graph of the function and the x-axis only when the function is positive. If the function is negative, then this area is represented by minus the integral.*

Example. The function $\sin x$ is **negative** on the interval $[\pi, 2\pi]$. The **area** between the curve $y = \sin x$ and the x-axis over this interval is given by **minus the integral**:

$$-\int_\pi^{2\pi} \sin x\, dx = -(-\cos x)\Big|_\pi^{2\pi} = 2.$$

The **area** between the graph of $\sin x$ and the x-axis between 0 and 2π is equal to twice the area of one of the loops, and is therefore equal to 4. On the other hand,

$$\int_0^{2\pi} \sin x\, dx = -\cos x \Big|_0^{2\pi} = 0.$$

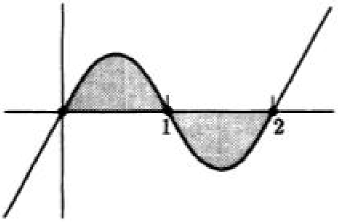


Example. Find the area between the curve

$$y = f(x) = x(x-1)(x-2)$$

and the x-axis,

The curve looks like this:



There are two portions between the curve and the x-axis, corresponding to the intervals $[0, 1]$ and $[1, 2]$. However, the function is negative between $x = 1$ and $x = 2$, so that to find the sum of the areas of the two regions, we have to take the absolute value of the integral over the second one. We therefore compute these areas separately.

First we expand the product giving $f(x)$, and get

$$f(x) = x^3 - 3x^2 + 2x.$$

The first integral is equal to

$$\begin{aligned}\int_0^1 f(x)\,dx &= \frac{x^4}{4} - 3\frac{x^3}{3} + 2\frac{x^2}{2}\bigg|_0^1 \\ &= \tfrac{1}{4} - 1 + 1 = \tfrac{1}{4}.\end{aligned}$$

The second integral is equal to

$$\begin{aligned}\int_1^2 f(x)\,dx &= \frac{x^4}{4} - x^3 + x^2\bigg|_1^2 \\ &= \tfrac{16}{4} - 8 + 4 - (\tfrac{1}{4} - 1 + 1) = -\tfrac{1}{4}.\end{aligned}$$

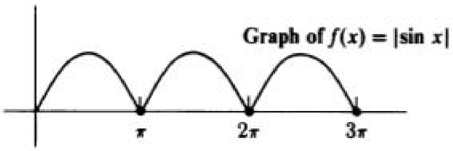
Hence the **area** of the two regions is equal to

$$\tfrac{1}{4} + \tfrac{1}{4} = \tfrac{1}{2}.$$

Example. Find the integral

$$\int_0^{3\pi} |\sin x|\,dx.$$

Note that because of the absolute value sign, the graph of the function $|\sin x|$ looks like this:



$$\text{If} \quad \sin x \geqq 0 \quad \text{on an interval, then} \quad |\sin x| = \sin x.$$

$$\text{If} \quad \sin x \leqq 0 \quad \text{on an interval, then} \quad |\sin x| = -\sin x.$$

Hence

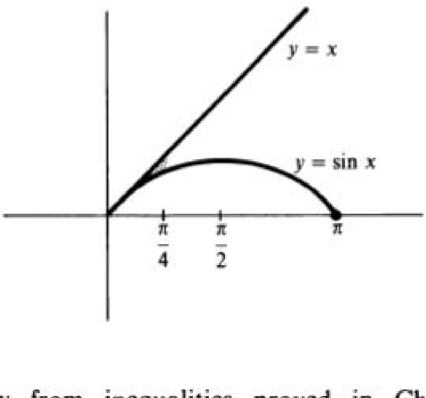
$$\int_0^{3\pi} |\sin x|\, dx = \int_0^{\pi} \sin x\, dx - \int_{\pi}^{2\pi} \sin x\, dx + \int_{2\pi}^{3\pi} \sin x\, dx$$

$$= 2 + 2 + 2 = 6.$$

Warning. Of course, this comes out three times the area under one arch of the graph, because of symmetries. But if you try to use symmetries in such integrals, be sure to *prove* that they are valid.

Example. Find the area between the curves $y = x$ and $y = \sin x$ for $0 \leqq x \leqq \pi/4$.

The graphs are as follows.



You should know from inequalities proved in Chapter V, §2 that $\sin x \leqq x$ for $0 \leqq x$. The area between the two curves between $x = 0$ and $x = \pi/4$ is the difference of the areas under the bigger curve and the smaller one, that is:

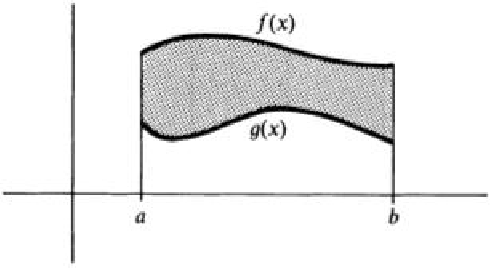
$$\int_0^{\pi/4} (x - \sin x)\,dx = \int_0^{\pi/4} x\,dx - \int_0^{\pi/4} \sin x\,dx$$

$$= \frac{x^2}{2}\bigg|_0^{\pi/4} + \cos x\bigg|_0^{\pi/4}$$

$$= \frac{\pi^2}{32} + \frac{1}{\sqrt{2}} - 1.$$

In general, if $f(x)$ and $g(x)$ are two continuous functions such that $f(x) \geqq g(x)$ on an interval $[a, b]$, then the area between the two curves, from a to b, is

$$\boxed{\int_a^b (f(x) - g(x))\,dx.}$$



X, §2. EXERCISES

Find the following integrals:

1. $\int 4x^3\,dx$
2. $\int (3x^4 - x^5)\,dx$
3. $\int (2\sin x + 3\cos x)\,dx$
4. $\int (3x^{2/3} + 5\cos x)\,dx$
5. $\int \left(5e^x + \frac{1}{x}\right)dx$
6. $\int_{-\pi}^{\pi} (\sin x + \cos x)\,dx$
7. $\int_{-1}^{1} 2x^5\,dx$
8. $\int_{-1}^{2} e^x\,dx$
9. $\int_{-1}^{3} 4x^2\,dx$
10. Find the area between the curves $y = x$ and $y = x^2$ from 0 to their first point of intersection for $x > 0$.

11. Find the area between the curves $y = x$ and $y = x^3$.

12. Find the area between the curves $y = x^2$ and $y = x^3$.

13. Find the area between the curve $y = (x - 1)(x - 2)(x - 3)$ and the x-axis. (Sketch the curve.)

14. Find the area between the curve $y = (x + 1)(x - 1)(x + 2)$ and the x-axis.

15. Find the area between the curves $y = \sin x$, $y = \cos x$, the y-axis, and the first point where these curves intersect for $x > 0$.

In each of the following problems 16 through 25:
(a) Sketch the graph of the function $f(x)$.
(b) Find the integral of the function over the given interval.

16. On $[-1, 1]$, $f(x) = x$ if $-1 \leqq x < 0$ and $f(x) = 5$ if $0 \leqq x \leqq 1$.

17. On $[-1, 1]$, $f(x) = x^2$ if $-1 \leqq x \leqq 0$ and $f(x) = -x$ if $0 < x \leqq 1$.

18. On $[-1, 1]$, $f(x) = x - 1$ if $-1 \leqq x < 0$ and $f(x) = x + 1$ if $0 \leqq x \leqq 1$.

19. On $[-\pi, \pi]$, $f(x) = \sin x$ if $-\pi \leqq x \leqq 0$, and $f(x) = x$ if $0 < x \leqq \pi$.

20. On $[-\pi, \pi]$, $f(x) = |\sin x|$.

21. On $[-\pi, \pi]$, $f(x) = |\cos x|$.

22. On $[-1, 1]$, $f(x) = |x|$.

23. On $[-\pi, \pi]$, $f(x) = \sin x + |\cos x|$.

24. On $[-\pi, \pi]$, $f(x) = x - |x|$.

25. On $[-\pi, \pi]$, $f(x) = \sin x + |\sin x|$.

26. Find the value of the integrals (a) $\int_0^{7\pi} |\sin x|\,dx$, (b) $\int_0^{7\pi} |\cos x|\,dx$. (c) For any positive integer n, $\int_0^{n\pi} |\sin x|\,dx$.

X, §3. INEQUALITIES

You may wait to read this section until it is used to estimate remainder terms in Taylor's formula.

Theorem 3.1. *Let a, b be two numbers, with $a \leqq b$. Let f, g be two continuous functions on the interval $[a, b]$ and assume that $f(x) \leqq g(x)$ for all x in the interval. Then*

$$\int_a^b f(x)\,dx \leqq \int_a^b g(x)\,dx.$$

Proof. Since $g(x) - f(x) \geqq 0$, we can use the basic Property 1 of Chapter IX, §5 (with $m = 0$) to conclude that

$$\int_a^b (g - f) \geqq 0.$$

But

$$\int_a^b (g - f) = \int_a^b g - \int_a^b f.$$

Transposing the second integral on the right in our inequality, we obtain

$$\int_a^b g \geqq \int_a^b f,$$

as desired.

Theorem 3.1 will be used mostly when $g(x) = |f(x)|$. Since a negative number is always $\leqq$ a positive number, we know that

$$f(x) \leqq |f(x)|$$

and

$$-f(x) \leqq |f(x)|.$$

Theorem 3.2. *Let a, b be two numbers, with $a \leqq b$. Let f be a continuous function on the interval $[a, b]$. Then*

$$\left| \int_a^b f(x)\,dx \right| \leqq \int_a^b |f(x)|\,dx.$$

Proof. We simply let $g(x) = |f(x)|$ in the preceding theorem. Then

$$f(x) \leqq |f(x)|$$

and also $-f(x) \leqq |f(x)|$. The absolute value of the integral on the left is equal to

$$\int_a^b f(x)\,dx \qquad \text{or} \qquad -\int_a^b f(x)\,dx.$$

We can apply Theorem 3.1 either to $f(x)$ or $-f(x)$ to get Theorem 3.2.

We make one other application of Theorem 3.2.

Theorem 3.3. *Let a, b be two numbers and f a continuous function on the closed interval between a and b. (We do not necessarily assume that $a < b$.) Let M be a number such that $|f(x)| \leqq M$ for all x in the interval. Then*

$$\left| \int_a^b f(x)\,dx \right| \leqq M|b - a|.$$

Proof. If $a \leqq b$, we can use Theorem 3.2 to get

$$\left|\int_a^b f(x)\,dx\right| \leqq \int_a^b M\,dx = M\int_a^b dx = M(b-a).$$

If $b < a$, then

$$\int_a^b f = -\int_b^a f.$$

Taking the absolute value gives us the estimate $M(a-b)$. Since

$$a - b = |b - a|$$

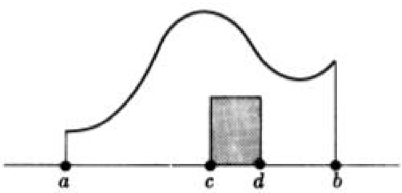
in case $b < a$, we have proved our theorem.

Theorem 3.4. *Let f be a continuous function on the interval $[a, b]$ with $a < b$. Assume that $f(x) \geqq 0$ for every x in this interval, and $f(x) > 0$ for some x in this interval. Then*

$$\int_a^b f(x)\,dx > 0.$$

Proof. Let c be a number of the interval such that $f(c) > 0$, and suppose for simplicity that $c \neq b$.

The geometric idea behind the proof is quite simple, in terms of area. Since the function f is assumed $\geqq 0$ everywhere, and > 0 at the point c, then it is greater than some fixed positive number [taken to be $f(c)/2$, say] in some interval near c. This means that we can insert a small rectangle of height > 0 between the curve $y = f(x)$ and the x-axis. Then the area under the curve is at least equal to the area of this rectangle, which is > 0.



This "proof" can be phrased in terms of the formal properties of the integral as follows. Since f is continuous, there exists some number d close to c in the interval, with $c < d \leqq b$ such that $f(x)$ is close to $f(c)$ for all x satisfying

$$c \leqq x \leqq d.$$

In particular, we have

$$f(x) \geqq \frac{f(c)}{2}, \qquad c \leqq x \leqq d.$$

Then

$$\begin{aligned}\int_a^b f(x)\,dx &= \int_a^c f + \int_c^d f + \int_d^b f \\ &\geqq \int_c^d f(x)\,dx \geqq \int_c^d \frac{f(c)}{2}\,dx \\ &\geqq \frac{f(c)}{2}(d-c) > 0.\end{aligned}$$

This proves our theorem if $c \neq b$. If $c = b$, we take $d < c$ and argue similarly.

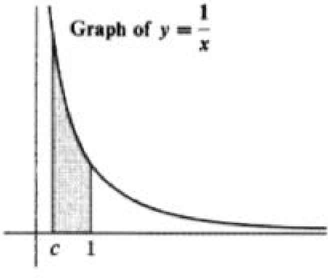
Theorem 3.4 will not be used in the rest of this book except in a couple of exercises, but it is important in subsequent applications.

X, §4. IMPROPER INTEGRALS

Example 1. We start with an example. Let $0 < c < 1$. We look at the integral

$$\int_c^1 \frac{1}{x}\,dx = \log x \Big|_c^1 = \log 1 - \log c = -\log c.$$

The figure illustrates this integral.



We have shaded the portion of the area under the graph lying between c and 1. As c approaches 0, we see that the area becomes arbitrarily large, because

$$-\log c \to \infty \qquad \text{as} \quad c \to 0.$$

Example 2. However, it is remarkable that an entirely different situation will occur when we consider the area under the curve $1/\sqrt{x} = x^{-1/2}$. We take $x > 0$, of course. Let $0 < c < 1$. We compute the integral:

$$\int_c^1 \frac{1}{x^{1/2}}\,dx = \int_c^1 x^{-1/2}\,dx = 2x^{1/2}\Big|_c^1$$

$$= 2 - 2c^{1/2}.$$

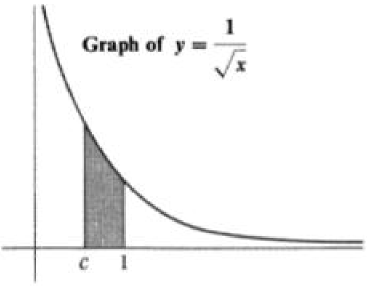
Then

$$\text{as} \quad c \to 0, \qquad 2c^{1/2} \to 0$$

and therefore

$$\int_c^1 x^{-1/2}\,dx \to 2 \qquad \text{as} \quad c \to 0.$$

We can illustrate the graph of $1/\sqrt{x}$ on the following figure. Note that at first sight, it does not differ so much from that of the preceding example, but the computation of the area shows the existence of a fundamental difference.



Both in Example 1 and Example 2 we are looking at an infinite chimney, as $c \to 0$. But in Example 1 the area becomes arbitrarily large, whereas in Example 2, the area approaches the limit 2.

Definition. In Example 2, we say that the integral

$$\int_0^1 x^{-1/2}\,dx$$

exists, or **converges** even though the function $x^{-1/2}$ is not defined at 0 and is not continuous in the **closed** interval $[0, 1]$.

In general, suppose we have two numbers a, b with, say, $a < b$. Let f be a continuous function in the interval $a < x \leqq b$. This means that for every number c with $a < c < b$, the function f is continuous on the interval

$$c \leqq x \leqq b.$$

Let F be any function such that $F'(x) = f(x)$.

We can then evaluate the integral as usual:

$$\int_c^b f(x)\,dx = F(b) - F(c).$$

Definition. If the limit

$$\lim_{c \to a} F(c)$$

exists, then we say that the **improper integral**

$$\int_a^b f(x)\,dx$$

exists, and then we define

$$\int_a^b f(x)\,dx = \lim_{c \to a} \int_c^b f(x)\,dx = F(b) - \lim_{c \to a} F(c).$$

We make similar definitions when we deal with an interval $a \leqq x < b$ and a function f which is continuous on this interval. If the limit

$$\lim_{c \to b} \int_a^c f(x)\,dx$$

exists, then we say that the **improper integral exists**, and it is equal to this limit.

Example 3. Show that the improper integral

$$\int_0^1 \frac{1}{x^2}\,dx$$

does not exist.

Let $0 < c < 1$. We first evaluate the integral:

$$\int_c^1 x^{-2}\,dx = \left.\frac{x^{-1}}{-1}\right|_c^1 = -1 - \left(-\frac{1}{c}\right) = -1 + \frac{1}{c}.$$

But $1/c \to \infty$ as $c \to 0$, and hence the improper integral does not exist.

Example 4. Determine whether the integral

$$\int_1^3 \frac{1}{x-1}\,dx$$

exists, and if it does, find its value.

Let $1 < c < 3$. Then the function $1/(x-1)$ is not continuous on the interval $[1, 3]$ but is continuous on the interval $[c, 3]$. Furthermore

$$\int_c^3 \frac{1}{x-1}\,dx = \log(x-1)\Big|_c^3 = \log 2 - \log(c-1).$$

But

$$-\log(c-1) \to \infty \qquad \text{as} \quad c \to 1 \quad \text{and} \quad c > 1.$$

Hence the integral does not exist.

There is another type of improper integral, dealing with large values.

Let a be a number and f a continuous function defined for $x \geqq a$. Consider the integral

$$\int_a^B f(x)\,dx$$

for some number $B > a$. If $F(x)$ is any indefinite integral of f, then our integral is equal to $F(B) - F(a)$. If it approaches a limit as B becomes very large, then we **define**

$$\boxed{\int_a^\infty f(x)\,dx \quad \text{or} \quad \int_a^\infty f = \lim_{B\to\infty} \int_a^B f(x)\,dx,}$$

and say that the **improper integral converges**, or **exists**.

Thus

$$\int_a^\infty f \qquad \text{exists if} \qquad \lim_{B\to\infty} \int_a^B f \quad \text{exists,}$$

and is equal to the limit. Otherwise, we say that the improper integral **does not converge**, or **does not exist**.

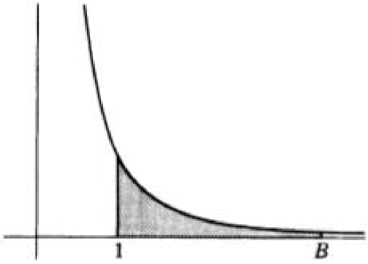
Example 5. Determine whether the improper integral $\int_1^\infty 1/x\,dx$ exists, and if it does, finds its value.

Let B be a number > 1. Then

$$\int_1^B \frac{1}{x}\,dx = \log B - \log 1 = \log B.$$

As B becomes large, so does $\log B$, and hence the improper integral does not exist.

Let us look at the function $1/x^2$. Its graph looks like that in the next figure. At first sight, there seems to be no difference between this function and $1/x$, except that $1/x^2 < 1/x$ when $x > 1$. However, intuitively speaking, we shall find that $1/x^2$ approaches 0 sufficiently faster than $1/x$ to guarantee that the area under the curve between 1 and B approaches a limit as B becomes large.



Example 6. Determine whether the improper integral

$$\int_1^\infty \frac{1}{x^2}\,dx$$

exists, and if it does, find its value.

Let B be a number > 1. Then

$$\int_1^B \frac{1}{x^2}\,dx = \frac{-1}{x}\bigg|_1^B = -\frac{1}{B} + 1.$$

As B becomes large, $1/B$ approaches 0. Hence the limit as B becomes large exists and is equal to 1, which is the value of our integral. We thus have by definition

$$\int_1^\infty \frac{1}{x^2}\,dx = \lim_{B\to\infty}\left(-\frac{1}{B} + 1\right) = 1.$$

X, §4. EXERCISES

Determine whether the following improper integrals exist or not.

1. $\displaystyle\int_2^\infty \frac{1}{x^{3/2}}\,dx$
2. $\displaystyle\int_1^\infty \frac{1}{x^{2/3}}\,dx$
3. $\displaystyle\int_0^\infty \frac{1}{1+x^2}\,dx$
4. $\displaystyle\int_0^5 \frac{1}{5-x}\,dx$
5. $\displaystyle\int_2^3 \frac{1}{x-2}\,dx$
6. $\displaystyle\int_1^4 \frac{1}{x-1}\,dx$
7. $\displaystyle\int_0^2 \frac{1}{x-2}\,dx$
8. $\displaystyle\int_2^3 \frac{1}{(x-2)^2}\,dx$
9. $\displaystyle\int_2^3 \frac{1}{(x-2)^{3/2}}\,dx$
10. $\displaystyle\int_1^4 \frac{1}{(x-1)^{2/3}}\,dx$

In the preceding exercises, you evaluated improper integrals of the form

$$\int_a^b \frac{1}{x^s}\,dx$$

in special cases. **The next two exercises are important because they tell you in general when such integrals exist or not.**

11. (a) Let s be a number < 1. Show that the improper integral

$$\int_0^1 \frac{1}{x^s}\,dx$$

exists.

(b) If $s > 1$, show that the integral does not exist.

(c) Does the integral exist when $s = 1$?

12. (a) If $s > 1$ show that the following integral exists.

$$\int_1^\infty \frac{1}{x^s}\,dx$$

(b) *If* $s < 1$ show that the integral does not exist.

Determine whether the following integrals exist, and if so find their values.

13. $\displaystyle\int_1^\infty e^{-x}\,dx$
14. $\displaystyle\int_1^\infty e^{x}\,dx$

15. Let B be a number > 2. Find the area under the curve $y = e^{-2x}$ between 2 and B. Does this area approach a limit when B becomes very large? If so, what limit?

CHAPTER XI

Techniques of Integration

The purpose of this chapter is to teach you certain basic tricks to find indefinite integrals. It is of course easier to look up integral tables, but you should have a minimum of training in standard techniques.

XI, §1. SUBSTITUTION

We shall formulate the analogue of the chain rule for integration.

Suppose that we have a function $u(x)$ and another function f such that $f(u(x))$ is defined. (All these functions are supposed to be defined over suitable intervals.) We wish to evaluate an integral having the form

$$\int f(u) \frac{du}{dx} dx,$$

where u is a function of x. We shall first work out examples to learn the mechanics for finding the answer.

Example 1. Find $\int (x^2 + 1)^3(2x)\, dx$.

Put $u = x^2 + 1$. Then $du/dx = 2x$ and our integral is in the form

$$\int f(u) \frac{du}{dx} dx,$$

the function f being $f(u) = u^3$. We abbreviate $(du/dx)\,dx$ by du, **as if we could cancel** dx. Then we can write the integral as

$$\int f(u)\,du = \int u^3\,du = \frac{u^4}{4} = \frac{(x^2+1)^4}{4}.$$

We can check this by differentiating the expression on the right, using the chain rule. We get

$$\frac{d}{dx}\frac{(x^2+1)^4}{4} = \frac{4}{4}(x^2+1)^3 2x = (x^2+1)^3(2x),$$

as desired.

Example 2. Find $\int \sin(2x)(2)\,dx$.

Put $u = 2x$. Then $du/dx = 2$. Hence our integral is in the form

$$\int \sin u \frac{du}{dx}\,dx = \int \sin u\,du = -\cos u = -\cos(2x).$$

Observe that

$$\int \sin(2x)\,dx \neq -\cos(2x).$$

If we differentiate $-\cos(2x)$, we get $\sin(2x)\cdot 2$.

The integral in Example 2 could also be written

$$\int 2\sin(2x)\,dx.$$

It does not matter, of course, where we place the 2.

Example 3. Find $\int \cos(3x)\,dx$.

Let $u = 3x$. Then $du/dx = 3$. There is no extra 3 in our integral. However, we can take a constant in and out of an integral. Our integral is equal to

$$\frac{1}{3}\int 3\cos(3x)\,dx = \frac{1}{3}\int \cos u\,du = \frac{1}{3}\sin u = \frac{1}{3}\sin(3x).$$

It is convenient to use a purely formal notation which allows us to make a substitution $u = g(x)$, as in the previous examples. Thus instead of writing

$$\frac{du}{dx} = 2x$$

in Example 1, we would write $du = 2x\,dx$. Similarly, in Example 2, we would write $du = 2\,dx$, and in Example 3 we would write $du = 3\,dx$. We do not attribute any meaning to this. It is merely a device of a type used in programming a computing machine. A machine does not think. One simply adjusts certain electric circuits so that the machine performs a certain operation and comes out with the right answer. The fact that writing

$$du = \frac{du}{dx}\,dx$$

makes us come out with the right answer will be proved in a moment.

Example 4. Find

$$\int (x^3 + x)^9(3x^2 + 1)\,dx.$$

Let

$$u = x^3 + x.$$

Then

$$du = (3x^2 + 1)\,dx.$$

Hence our integral is of type $\int f(u)\,du$ and is equal to

$$\int u^9\,du = \frac{u^{10}}{10} = \frac{(x^3 + x)^{10}}{10}.$$

We now show how the above procedure, which can be checked in each case by differentiation, actually must give the right answer in all cases.

We suppose therefore that we wish to evaluate an integral of the form

$$\int f(u)\frac{du}{dx}\,dx,$$

where u is a function of x. Let F be a function such that

$$F'(u) = f(u).$$

Thus F is an indefinite integral

$$\boxed{F(u) = \int f(u)\,du.}$$

If we use the chain rule, we get

$$\frac{dF(u(x))}{dx} = \frac{dF}{du}\frac{du}{dx} = f(u)\frac{du}{dx}.$$

Thus we have proved that

$$\boxed{\int f(u)\frac{du}{dx}\,dx = F(u(x))}$$

as desired.

We should also observe that the formula for integration by substitution applies to the definite integral. We can state this formally as follows.

Let g be a differentiable function on the interval $[a, b]$, whose derivative is continuous. Let f be a continuous function on an interval containing the values of g. Then

$$\boxed{\int_a^b f(g(x))\frac{dg}{dx}\,dx = \int_{g(a)}^{g(b)} f(u)\,du.}$$

The proof is immediate. If F is an indefinite integral for f, then $F(g(x))$ is an indefinite integral for

$$f(g(x))\frac{dg}{dx}$$

by the chain rule. Hence the left-hand side of our formula is equal to

$$F(g(b)) - F(g(a)),$$

which is also the value of the right-hand side.

Example 5. Suppose that we consider the integral

$$\int_0^1 (x^3 + x)^9(3x^2 + 1)\,dx,$$

with $u = x^3 + x$. When $x = 0$, $u = 0$, and when $x = 1$, $u = 2$. Thus our definite integral is equal to

$$\int_0^2 u^9\,du = \frac{u^{10}}{10}\bigg|_0^2 = \frac{2^{10}}{10}.$$

Example 6. Evaluate

$$\int_0^{\sqrt{\pi}} x \sin x^2\,dx.$$

We let $u = x^2$, $du = 2x\,dx$. When $x = 0$, $u = 0$. When $x = \sqrt{\pi}$, $u = \pi$. Thus our integral is equal to

$$\tfrac{1}{2}\int_0^{\pi} \sin u\,du = \tfrac{1}{2}(-\cos u)\bigg|_0^{\pi} = \tfrac{1}{2}(-\cos \pi + \cos 0) = 1.$$

Example 7. Evaluate

$$I = \int x^5\sqrt{1 - x^2}\,dx.$$

We let $u = 1 - x^2$ so $du = -2x\,dx$. Then $x^2 = 1 - u$, $x^4 = (1 - u)^2$, and

$$\begin{aligned} I = -\tfrac{1}{2}\int (1 - u)^2 u^{1/2}\,du &= -\tfrac{1}{2}\int (1 - 2u + u^2)u^{1/2}\,du \\ &= -\tfrac{1}{2}\int (u^{1/2} - 2u^{3/2} + u^{5/2})\,du \\ &= -\tfrac{1}{2}\left(\frac{u^{3/2}}{3/2} - 2\frac{u^{5/2}}{5/2} + \frac{u^{7/2}}{7/2}\right). \end{aligned}$$

Substituting $u = 1 - x^2$ gives the answer in terms of x.

XI, §1. EXERCISES

Find the following integrals.

1. $\int xe^{x^2}\,dx$
2. $\int x^3 e^{-x^4}\,dx$
3. $\int x^2(1 + x^3)\,dx$
4. $\int \frac{\log x}{x}\,dx$

5. $\displaystyle\int \frac{1}{x(\log x)^n}\,dx$ (n = integer)

6. $\displaystyle\int \frac{2x+1}{x^2+x+1}\,dx$

7. $\displaystyle\int \frac{x}{x+1}\,dx$

8. $\displaystyle\int \sin x \cos x\,dx$

9. $\displaystyle\int \sin^2 x \cos x\,dx$

10. $\displaystyle\int_0^\pi \sin^5 x \cos x\,dx$

11. $\displaystyle\int_0^\pi \cos^4 x \sin x\,dx$

12. $\displaystyle\int \frac{\sin x}{1+\cos^2 x}\,dx$

13. $\displaystyle\int \frac{\arctan x}{1+x^2}\,dx$

14. $\displaystyle\int_0^1 x^3\sqrt{1-x^2}\,dx$

15. $\displaystyle\int_0^{\pi/2} x\sin(2x^2)\,dx$

16. (a) $\displaystyle\int \sin 2x\,dx$ (b) $\displaystyle\int \cos 2x\,dx$

(c) $\displaystyle\int \sin 3x\,dx$ (d) $\displaystyle\int \cos 3x\,dx$

(e) $\displaystyle\int e^{4x}\,dx$ (f) $\displaystyle\int e^{5x}\,dx$ (g) $\displaystyle\int e^{-5x}\,dx$

In the next problems, you will use limits from the theory of the exponential function in Chapter VIII, §5, namely

$$\boxed{\lim_{B\to\infty} \frac{B}{e^B} = 0.}$$

17. Find the area under the curve $y = xe^{-x^2}$ between 0 and a number $B > 0$. Does this area approach a limit as B becomes very large? If so, what limit?

18. Find the area under the curve $y = x^2e^{-x^3}$ between 0 and a number $B > 0$. Does this area approach a limit as B becomes very large? If so, what limit?

XI, §1. SUPPLEMENTARY EXERCISES

Find the following integrals.

1. $\displaystyle\int \frac{x^3}{x^4+2}\,dx$

2. $\displaystyle\int \sqrt{3x+1}\,dx$

3. $\displaystyle\int \sin^4 x \cos x\,dx$

4. $\displaystyle\int \frac{e^x - e^{-x}}{e^x + e^{-x}}\,dx$

5. $\displaystyle\int \frac{x}{\sqrt{x^2-1}}\,dx$

6. $\displaystyle\int x^3\sqrt{x^4+1}\,dx$

7. $\displaystyle\int \frac{x}{(3x^2+5)^2}\,dx$

8. $\displaystyle\int (x^2+3)^4 x^3\,dx$

9. $\displaystyle\int \frac{\cos x}{\sin^3 x}\,dx$

10. $\displaystyle\int e^x\sqrt{e^x+1}\,dx$

11. $\displaystyle\int (x^3+1)^{7/5}x^5\,dx$

12. $\displaystyle\int \frac{x}{(x^2-4)^{3/2}}\,dx$

13. $\displaystyle\int \sin 3x\,dx$

14. $\displaystyle\int \cos 4x\,dx$

15. $\displaystyle\int e^x \sin e^x\,dx$

16. $\displaystyle\int x^2\sqrt{x^3+1}\,dx$

17. $\displaystyle\int \frac{1}{x\log x}\,dx$

18. $\displaystyle\int \frac{\sin x}{1+\cos^2 x}\,dx$

19. $\displaystyle\int \frac{e^x}{e^x+1}\,dx$

20. $\displaystyle\int \frac{(\log x)^4}{x}\,dx$

21. $\displaystyle\int_0^{\pi/2} \sin^3 x\cos x\,dx$

22. $\displaystyle\int_{-3}^{-1} \frac{1}{(x-1)^2}\,dx$

23. $\displaystyle\int_0^1 \sqrt{2-x}\,dx$

24. $\displaystyle\int_0^{\pi} \sin^2 x\cos x\,dx$

25. $\displaystyle\int_0^{\pi/2} \frac{\cos x}{1+\sin^2 x}\,dx$

26. $\displaystyle\int_0^{2\pi} \frac{\sin x}{1+\cos^2 x}\,dx$

27. $\displaystyle\int_0^{1/2} \frac{\arcsin x}{\sqrt{1-x^2}}\,dx$

28. $\displaystyle\int_0^1 \frac{\arctan x}{1+x^2}\,dx$

29. $\displaystyle\int_0^1 \frac{1+e^{2x}}{e^x}\,dx$

30. $\displaystyle\int_0^{\pi/2} x\sin x^2\,dx$

XI, §2. INTEGRATION BY PARTS

If f, g are two differentiable functions of x, then

$$\frac{d(fg)}{dx} = f(x)\frac{dg}{dx} + g(x)\frac{df}{dx}.$$

Hence

$$f(x)\frac{dg}{dx} = \frac{d(fg)}{dx} - g(x)\frac{df}{dx}$$

Using the formula for the integral of a sum, which is the sum of the integrals, and the fact that

$$\int \frac{d(fg)}{dx}\,dx = f(x)g(x),$$

we obtain

$$\int f(x)\frac{dg}{dx}\,dx = f(x)g(x) - \int g(x)\frac{df}{dx}\,dx,$$

which is called the formula for **integrating by parts**.

If we let $u = f(x)$ and $v = g(x)$, then the formula can be abbreviated in our shorthand notation as follows:

$$\boxed{\int u\,dv = uv - \int v\,du.}$$

Example 1. Find the integral $\int \log x\,dx$.

Let $u = \log x$ and $dv = dx$. Then $du = (1/x)\,dx$ and $v = x$. Hence our integral is in the form $\int u\,dv$ and is equal to

$$uv - \int v\,du = x\log x - \int 1\,dx = x\log x - x.$$

Example 2. Find the integral $\int xe^x\,dx$.

Let $u = x$ and $dv = e^x\,dx$. Then $du = dx$ and $v = e^x$. So

$$\int xe^x\,dx = \int u\,dv = xe^x - \int e^x\,dx = xe^x - e^x.$$

Observe how du/dx is a simpler function than u itself, whereas v and dv/dx are the same, in the present case. A similar procedure works for integrals of the form $\int x^n e^x\,dx$ when n is a positive integer, putting $u = x^n$, $du = nx^{n-1}\,dx$. See Exercises 7 and 8.

The two preceding examples illustrate a general fact: Passing from the function u to du/dx in the process of integrating by parts will work provided that going up from dv to v does not make this side of the procedure too much worse. In the first example, with $u = \log x$, then $du/dx = 1/x$ is a simpler function (a power of x), while going up from $dv = dx$ to $v = x$ still only contributes powers of x to the procedure.

Next we give an example where we have to integrate by parts twice before getting an answer.

Example 3. Find $\int e^x \sin x\, dx$.

Let $u = e^x$ and $dv = \sin x\, dx$. Then

$$du = e^x\, dx \qquad \text{and} \qquad v = -\cos x.$$

If we call our integral I, then

$$\begin{aligned} I &= -e^x \cos x - \int -e^x \cos x\, dx \\ &= -e^x \cos x + \int e^x \cos x\, dx. \end{aligned}$$

This looks as if we were going around in circles. Don't lose heart. Rather, repeat the same procedure on $e^x \cos x$. Let

$$t = e^x \qquad \text{and} \qquad dz = \cos x\, dx.$$

Then,

$$dt = e^x\, dx \qquad \text{and} \qquad z = \sin x.$$

The second integral becomes

$$\int t\, dz = tz - \int z\, dt = e^x \sin x - \int e^x \sin x\, dx.$$

We have come back to our original integral

$$\int e^x \sin x\, dx$$

but with a minus sign! Thus

$$I = -e^x \cos x + e^x \sin x - I.$$

Hence

$$2I = e^x \sin x - e^x \cos x,$$

and dividing by 2 gives us the value

$$I = \frac{e^x \sin x - e^x \cos x}{2}.$$

We give an example where we first make a substitution *before* integrating by parts.

Example 4. Find the integral $\int e^{-\sqrt{x}}\,dx$.

We let $x = u^2$ so that $dx = 2u\,du$. Then

$$\int e^{-\sqrt{x}}\,dx = \int e^{-u}2u\,du = 2\int ue^{-u}\,du.$$

This can now be integrated by parts with $u = u$ and $dv = e^{-u}\,du$. Then $v = -e^{-u}$, and you can do the rest.

Remark. It can be shown, although not easily, that no procedure will allow you to express the integral

$$\int e^{-x^2}\,dx$$

in terms of the standard functions: powers of x, trigonometric functions, exponential and log function, sums, products, or composites of these.

XI, §2. EXERCISES

Find the following integrals.

1. $\int \arcsin x\,dx$
2. $\int \arctan x\,dx$
3. $\int e^{2x}\sin 3x\,dx$
4. $\int e^{-4x}\cos 2x\,dx$
5. $\int (\log x)^2\,dx$
6. $\int (\log x)^3\,dx$
7. $\int x^2e^x\,dx$
8. $\int x^2e^{-x}\,dx$
9. $\int x\sin x\,dx$
10. $\int x\cos x\,dx$
11. $\int x^2\sin x\,dx$
12. $\int x^2\cos x\,dx$
13. $\int x^3\cos x^2\,dx$
14. $\int x^5\sqrt{1-x^2}\,dx$

[*Hint*: In Exercise 13, make first the substitution $u = x^2$, $du = 2x\,dx$. In Exercise 14, let $u = 1 - x^2$, $x^2 = 1 - u$.]

15. $\int x^2\log x\,dx$
16. $\int x^3\log x\,dx$

17. $\int x^2(\log x)^2\,dx$

18. $\int x^3 e^{-x^2}\,dx$

19. $\int \frac{x^7}{(1-x^4)^2}\,dx$

20. $\int_{-\pi}^{\pi} x^2 \cos x\,dx$

In the next improper integrals, we shall use limits from the theory of exponentials and logarithms, Chapter VIII, §5. These limits are as follows. If n is a positive integer, then

$$\boxed{\lim_{B\to\infty} \frac{B^n}{e^B} = 0.}$$

Also

$$\boxed{\lim_{a\to 0} a \log a = 0.}$$

The first limit states that the exponential function becomes large faster than any polynomial. The second limit can be deduced from the first, namely write $a = e^{-B}$. Then $a \to 0$ if and only if $B \to \infty$. But $\log a = -B$. So

$$a \log a = \frac{-B}{e^B},$$

whence we see that the first limit implies the second by taking $n = 1$.

21. Let B be a number > 0. Find the area under the curve $y = xe^{-x}$ between 0 and B. Does this area approach a limit as B becomes very large? If yes, what limit?

22. Does the improper integral $\int_1^\infty x^2 e^{-x}\,dx$ exist? If yes, what is its value?

23. Does the improper integral $\int_1^\infty x^3 e^{-x}\,dx$ exist? If yes, what is its value?

24. Let B be a number > 2. Find the area under the curve

$$y = \frac{1}{x(\log x)^2}$$

between 2 and B. Does this area approach a limit as B becomes very large? If so, what limit?

25. Does the improper integral

$$\int_3^\infty \frac{1}{x(\log x)^4}\,dx$$

exist? If yes, what is its value?

26. Does the improper integral
$$\int_0^2 \log x \, dx$$
exist? If yes, what is its value?

XI, §2. SUPPLEMENTARY EXERCISES

Find the following integrals, using substitutions and integration by parts, as needed.

1. $\displaystyle\int x \arctan x \, dx$

2. (a) $\displaystyle\int \sqrt{1-x^2}\, dx$ [*Hint*: Use $u = \sqrt{1-x^2}$ and $dv = dx$.]

   (b) $\displaystyle\int x \arcsin x \, dx$

3. $\displaystyle\int x \arccos x \, dx$

4. $\displaystyle\int x^3 e^{2x} \, dx$

5. $\displaystyle\int_{-1}^{0} \arcsin x \, dx$

6. $\displaystyle\int_1^2 x^3 \log x \, dx$

7. $\displaystyle\int_0^{1/2} x \arcsin 2x \, dx$

8. $\displaystyle\int_1^2 \sqrt{x} \log x \, dx$

9. $\displaystyle\int_{-1}^{1} xe^x \, dx$

10. $\displaystyle\int_0^1 x^3\sqrt{1-x^2}\, dx$

11. $\displaystyle\int xe^{-\sqrt{x}} \, dx$

12. $\displaystyle\int \sqrt{x} e^{-\sqrt{x}} \, dx$

Prove the formulas, where m, n are positive integers.

13. $\displaystyle\int (\log x)^n \, dx = x(\log x)^n - n \int (\log x)^{n-1} \, dx$

14. $\displaystyle\int x^n e^x \, dx = x^n e^x - n \int x^{n-1} e^x \, dx$

15. $\displaystyle\int x^m (\log x)^n \, dx = \frac{x^{m+1}(\log x)^n}{m+1} - \frac{n}{m+1} \int x^m (\log x)^{n-1} \, dx$

*16. Let $n! = n(n-1)\cdots 1$ be the product of the first n integers. Show that
$$\int_0^\infty x^n e^{-x} \, dx = n!.$$
[*Hint*: First find the indefinite integral $\int x^n e^{-x}\, dx$ in terms of $\int x^{n-1} e^{-x}\, dx$ as in Exercise 14. Then evaluate between 0 and B, and let B become large. If I_n denotes the desired integral, you should find $I_n = nI_{n-1}$. You can then continue stepwise, until you are reduced to evaluating $I_0 = \int_0^\infty e^{-x}\, dx$.]

XI, §3. TRIGONOMETRIC INTEGRALS

We shall investigate integrals involving sine and cosine. It will be useful to have the following formulas:

$$\boxed{\sin^2 x = \frac{1 - \cos 2x}{2}, \qquad \cos^2 x = \frac{1 + \cos 2x}{2}.}$$

These are easily proved, using

$$\cos 2x = \cos^2 x - \sin^2 x \qquad \text{and} \qquad \sin^2 x + \cos^2 x = 1.$$

Example. We use the first boxed formula and get:

$$\int \sin^2 x\, dx = \int \frac{1}{2}\, dx - \frac{1}{2}\int \cos 2x\, dx = \frac{x}{2} - \frac{1}{4}\sin 2x.$$

The next two examples deal with *odd* powers of sine or cosine. A method can be used in this case, but cannot be used when there are only *even* powers.

Example. We wish to find

$$\int \sin^3 x\, dx.$$

We replace $\sin^2 x$ by $1 - \cos^2 x$, so that

$$\begin{aligned}\int \sin^3 x\, dx &= \int (\sin x)(1 - \cos^2 x)\, dx \\ &= \int \sin x\, dx - \int (\cos^2 x)(\sin x)\, dx.\end{aligned}$$

The second of these last integrals can be evaluated by the substitution

$$u = \cos x, \qquad du = -\sin x\, dx.$$

We therefore find that

$$\int \sin^3 x\, dx = -\cos x + \frac{\cos^3 x}{3}.$$

For low powers of the sine and cosine, the above means are the easiest. Especially if we have to integrate an *odd* power of the sine or cosine, we can use a similar method.

Example. Find $I = \int \sin^5 x \cos^2 x \, dx$.

We replace

$$\sin^4 x = (\sin^2 x)^2 = (1 - \cos^2 x)^2,$$

so that

$$\int \sin^5 x \cos^2 x \, dx = \int (1 - \cos^2 x)^2 \cos^2 x \sin x \, dx.$$

Now we put

$$u = \cos x \quad \text{and} \quad du = -\sin x \, dx.$$

Then

$$I = -\int (1 - u^2)^2 u^2 \, du = -\int (1 - 2u^2 + u^4) u^2 \, du$$

$$= -\int (u^2 - 2u^4 + u^6) \, du$$

$$= -\left(\frac{u^3}{3} - 2\frac{u^5}{5} + \frac{u^7}{7}\right).$$

If you want the answer in terms of x, you substitute back $u = \cos x$ in this last expression.

In the two preceding examples, we had an *odd* power of the sine. Then by using the identity $\sin^2 x = 1 - \cos^2 x$ and the substitution

$$u = \cos x, \qquad du = -\sin x \, dx,$$

we transform the integral into a sum of integrals of the form

$$\int u^n \, du,$$

where n is a positive integer.

This method works only when there is an odd power of sine or cosine. In general, one has to use another method to integrate arbitrary powers.

There is a general way in which one can integrate $\sin^n x$ for any positive integer n: integrating by parts. Let us take first an example.

Example. Find the integral $\int \sin^3 x \, dx$.