Math 1: **Analysis**

FOR THE ELECTRONICS AND TELECOMMUNICATION STUDENTS

The course material selected and prepared by:
Andrzej Maćkiewicz
Technical University of Poznań

Abstract

Contents

Preface

These are lecture notes for a first course in calculus. The development of calculus in the seventeenth and eighteenth centuries was motivated by the need to understand physical phenomena such as the tides, the phases of the moon, the nature of light, and gravity.

Now, Calculus is one of the core topics studied at university level by students on many different types of degree programme. Alongside Linear Algebra, it provides the framework for mathematical modelling in many diverse areas. This text sets out to introduce and explain Calculus to students from electronics and telecommunication. It covers all the material that would be expected to be in most first-year university courses in the subject, together with some more advanced material that would normally be taught later.

Descriptions of the topics to be covered appear in the relevant chapters. However, it is useful to give a brief overview at this stage. We start by introducing the limit of a function of one variable and, in particular, how this can be used to define what it means to say that a function is continuous. We then introduce the Riemann integral and explain its relationship to differentiation via the Fundamental Theorem of Calculus. This leads on to a discussion of improper integrals and, in particular, some tests that we can use to determine whether such integrals are convergent or divergent.

More detailed map of dependences is presented in the Figure below.

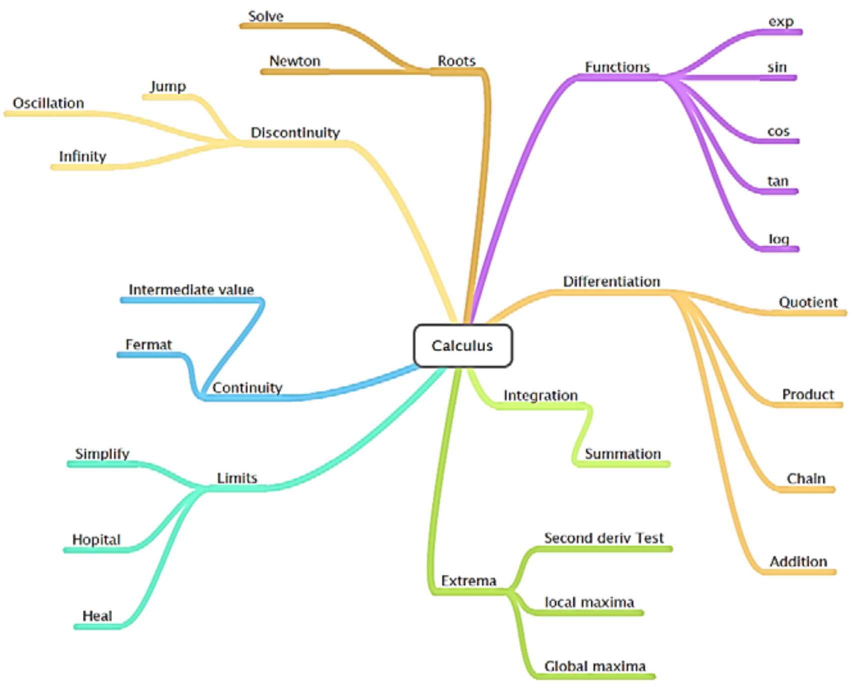
Graphing calculators and computers are playing an increasing role in the mathematics classroom. Without a doubt, graphing technology can enhance the understanding of calculus, so some instances where many of the graphs of surfaces were produced using *Mathematica, Matlab, GeoGebra, Graph* software and P. Seeburger's Dynamic Calculus Site (http://web.monroecc.edu/Dynamic Calculus/).

This text represents our best effort at distilling from my experience what it is that we think works best in helping students not only to do Calculus, but to understand it. We regard understanding as essential.

We have attempted to write a user-friendly, fairly interactive and helpful text, and we intend that it could be useful not only as a course text, but for self-study. These notes are quite informal, but they have been carefully read and criticized by the students, and their comments and suggestions have been incorporated. Although we've tried to be careful, there are undoubtedly some errors remaining. If you find any, please let us know.

Carefully designed exercises are provided in the supplementary volume of "Calculus , Problems , Solutions and Hints".

Andrzej Maćkiewicz
Poznań, September 2014



Map of the course. (Source: Oliver Knill)

1
Spotlight on calculus

Calculus is the mathematics that helps us deal with calculating information about changing quantities. It was developed to deal with such problems as finding the total distance travelled by an object with changing velocity, or finding the velocity of an object with changing position. Historically, calculus, with its origins in the 17th century, came first, and made rapid progress on the basis of informal intuition.

Sir Isaac Newton

Not until well through the 19th century was it possible to claim that the edifice was constructed on sound logical foundations. As for practicality, every university teacher knows that students are not ready for even a semi-rigorous course on analysis until they have acquired the intuitions and the sheer technical skills that come from a traditional calculus course. Therefore, for novice students, we present here calculus using more or less traditional approach. The goal of this course is for you to understand and appreciate the beautiful subject of calculus.

Gottfried Wilhelm von Leibniz

There are two main branches of calculus. The first is *differentiation* (or derivatives), which helps us to find a rate of change of one quantity compared to another. The second is *integration*, which is the reverse of differentiation. We may be given a rate of change and we need to work backwards to find the original relationship (or equation) between the two quantities.

Rene Descartes

1.1 The two basic concepts of calculus

The remarkable progress that has been made in science and technology during the last Centuries is due in large part to the development of mathematics. That branch of mathematics known as integral and differential calculus serves as a natural and powerful tool for attacking a variety of problems that arise in physics, astronomy, engineering, chemistry, geology, biology, and other fields including, rather recently, some of the social sciences. Calculus is concerned with comparing quantities which vary in a non-linear way. It is used extensively in science and engineering since many of the things we are studying (like velocity, acceleration, current in a circuit) do not behave in a simple, linear fashion. If quantities are continually changing, we need calculus to study what is going on.

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

Consider a curve C which lies above a horizontal base line such as that shown in Figure 1.1. We assume this curve has the property that every vertical line intersects it once at most. The shaded portion of the figure consists of those points which lie below the curve C, above the horizontal base, and between two parallel vertical segments joining C to the base. The first *fundamental problem of calculus* is this : *To assign a number which measures the area of this shaded region.*Consider next a line drawn tangent to the curve, as shown in Figure 1.1. The second fundamental problem may be stated as follows: *To assign a number which measures the steepness of this line.* Basically, calculus has to do with the precise formulation and solution of these two special problems. It enables us to *define* the concepts of area and tangent line and to *calculate* the area of a given region or the steepness of a given tangent line. *Integral calculus* deals with the problem of area and will be discussed in Chapter (xxx) . *Differential calculus* deals with the problem of tangents and will be introduced in Chapter (XXX). The birth of integral calculus occurred more than 2000 years ago when the Greeks attempted to determine areas by a process which they called the method *ofexhaustion.* The essential ideas of this method are very simple and can be described briefly as follows: Given a region whose area is to be determined, we inscribe in it a polygonal region

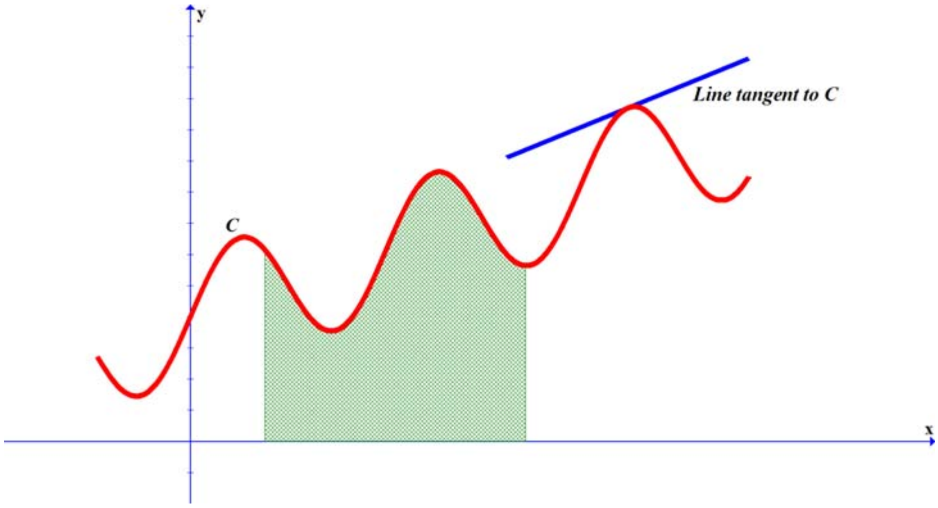


Fig. 1.1. Two specyfic problems: area of shaded region and the tangent line.

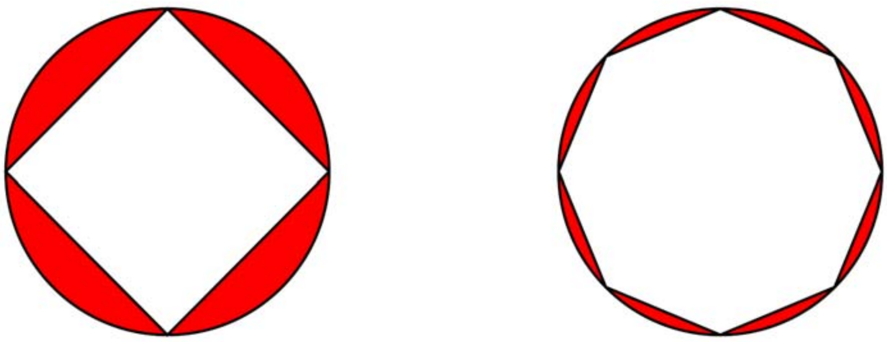


Fig. 1.2. Archimedes used the perimeters of inscribed polygons to approximate the circumference of the circle. For $n = 96$ the approximation method gives $\pi \approx 3.14103$ as the circumference of the unit circle.

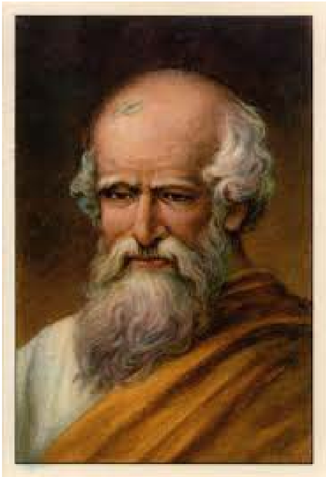


Fig. 1.3. Archimedes used the method of exhaustion to derive formulas for the areas of ellipses, parabolic segments, and sectors of a spiral. He is considered to have been the greatest applied mathematician of antiquity.

which approximates the given region and whose area we can easily compute. Then we choose another polygonal region which gives a better approximation, and we continue the process, taking polygons with more and more sides in an attempt to exhaust the given region. The method is illustrated for a circular region in Figure 1.2. It was used successfully by Archimedes (287 − 212 BC.– see Figure 1.3) to find exact formulas for the area of a circle and a few other special figures.The development of the method of exhaustion beyond the point to which Archimedes carried it had to wait nearly eighteen centuries until the use of algebraic symbols and techniques became a standard part of mathematics. The elementary algebra that is familiar to most high-school students today was completely unknown in Archimedes' time, and it would have been next to impossible to extend his method to any general class of regions without some convenient way of expressing rather lengthy calculations in a compact and simplified form.

A slow but revolutionary change in the development of mathematical notations began in the 16th Century A.D. The cumbersome system of Roman numerals was gradually displaced by the Hindu-Arabie characters used today, the symbols + and − were introduced for the first time, and the advantages of the decimal notation began to be recognized. During this same period, the brilliant successes of the Italian mathematicians Tartaglia, Cardano, and Ferrari in finding algebraic solutions of cubic and quartic equations stimulated a great deal of activity in mathematics and encouraged the growth and acceptance of a new and superior algebraic language. With the widespread introduction of well-chosen algebraic symbols, interest was revived in the ancient method of exhaustion and a large number of fragmentary results were discovered in the 16th Century by such pioneers as Cavalieri, Toricelli, Roberval, Fermat, Pascal, and Wallis.

Gradually the method of exhaustion was transformed into the subject now called integral calculus, a new and powerful discipline with a large variety of applications, not only to geometrical problems concerned with areas and volumes but also to problems in other sciences. This branch of mathematics, which retained some of the original features of the method of exhaustion, received its biggest impetus in the 17th Century after the development of an accurate clock. For scientists, it was very important to be able to predict the positions of the stars, to help in maritime navigation. The greatest challenge was to determine longitude when a ship was at sea. Whichever nation could send ships to the New World and successfully bring them back laden with goods, would become a rich country.

This led to significant developments in science and mathematics, and amongst the greatest of these was the calculus, largely due to the efforts of Isaac Newton (1642 – 1727) and Gottfried Leibniz (1646 – 1716). Newton and Leibniz built on the algebraic and geometric work of Rene Descartes, who developed the *Cartesian co-ordinate system*. There was a bitter dispute between the men over who developed calculus first. Because of this independent development, we have an unfortunate mix of notation and vocabulary that is used in calculus. From Leibniz we get the $\frac{dy}{dx}$ and $\int$ signs.

Calculus development continued well into the 19th Century before the subject was put on a firm mathematical basis by such men as Augustin-Louis Cauchy (1789 – 1857) and Bernhard Riemann (1826 – 1866). Further refinements and extensions of the theory are still being carried out in contemporary mathematics.

1.2 The method of exhaustion for the area of parabolic segment

Before we proceed to a systematic treatment of integral calculus, it will be instructive to apply the method of exhaustion directly to one of the special figures treated by Archimedes himself. The region in question is shown in Figure 1.4 and can be described as follows: If we choose an arbitrary point on the base of this figure and denote its distance from 0 by x, then the vertical distance from this point to the curve is x^2. In particular, if the length of the

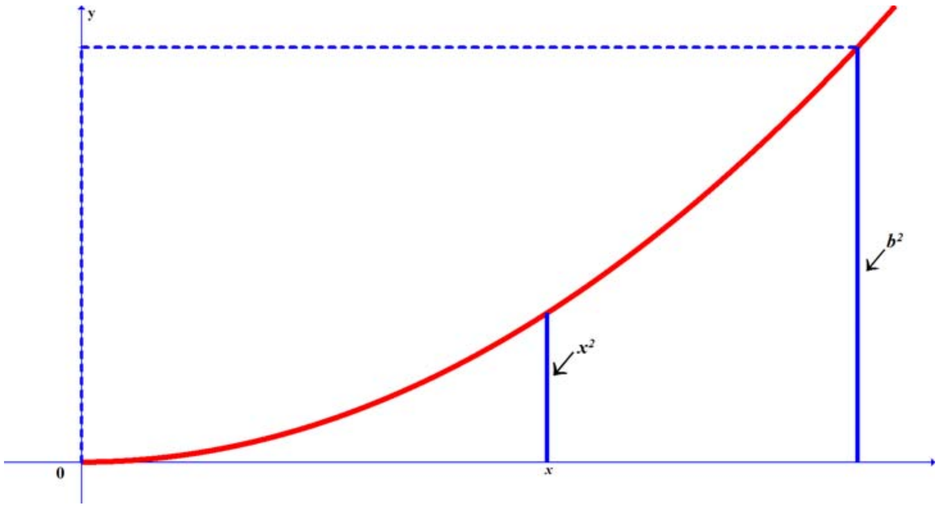


Fig. 1.4. A parabolic segment.

base itself is b, the altitude of the figure is b^2. The vertical distance from x to the curve is called the *ordinate* at x. The curve itself is an example of what is known as a parabola. The region bounded by it and the two line segments is called a *parabolic segment* (Figure 1.4). This figure may be enclosed in a rectangle of base b and altitude b^2, as shown in Figure 1.4. Examination of the figure suggests that the area of the parabolic segment is less than half the area of the rectangle. Archimedes made the surprising discovery that the area of the parabolic segment is exactly one-third that of the rectangle; that is to say, $A = b^3/3$, where A denotes the area of the parabolic segment. We shall show presently how to arrive at this result.

It should be pointed out that the parabolic segment in Figure 1.4 is not shown exactly as Archimedes drew it and the details that follow are not exactly the same as those used by him. Nevertheless, the essential ideas are those of Archimedes; what is presented here is the method of exhaustion in modern notation.

The method is simply this: We slice the figure into a number of strips and obtain two approximations to the region, one from below and one from above, by using two sets of rectangles as illustrated in Figure 1.5. (We use rectangles rather than arbitrary polygons to simplify the computations.) The area of the parabolic segment is larger than the total area of the inner rectangles but smaller than that of the outer rectangles.

If each strip is further subdivided to obtain a new approximation with a larger number of strips, the total area of the inner rectangles increases, whereas

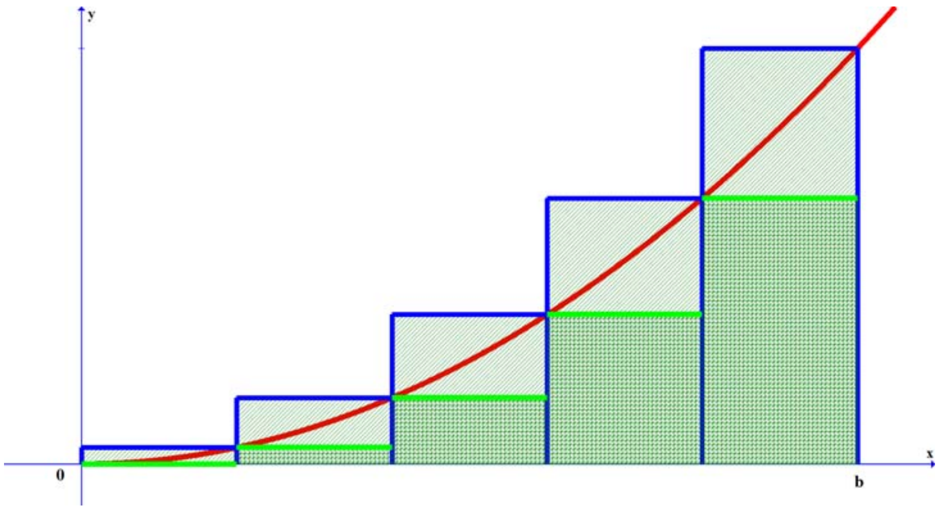


Fig. 1.5. Approximation from below and above.

the total area of the outer rectangles decreases. Archimedes realized that an approximation to the area within any desired degree of accuracy could be obtained by simply taking enough strips.

Let us carry out the actual computations that are required in this case. For the sake of simplicity, we subdivide the base into n equal parts, each of length b/n (see Figure 1.6). The points of subdivision correspond to the following values of x:

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

A typical point of subdivision corresponds to $x = kb/n$, where k takes the successive values $k = 0, 1, 2, 3, ..., n$. At each point kb/n we construct the outer rectangle of altitude $(kb/n)^2$ as illustrated in Figure 1.6. The area of this rectangle is the product of its base and altitude and is equal to

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

Let us denote by S_n, the sum of the areas of all the outer rectangles. Then since the kth rectangle has area $b^3\frac{k^2}{n^3}$, we obtain the formula

$$S_n = \frac{b^3}{n^3}\left(1^2 + 2^2 + ... + n^2\right). \tag{1.1}$$

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

$$s_n = \frac{b^3}{n^3}\left(1^2 + 2^2 + ... + (n-1)^2\right). \tag{1.2}$$

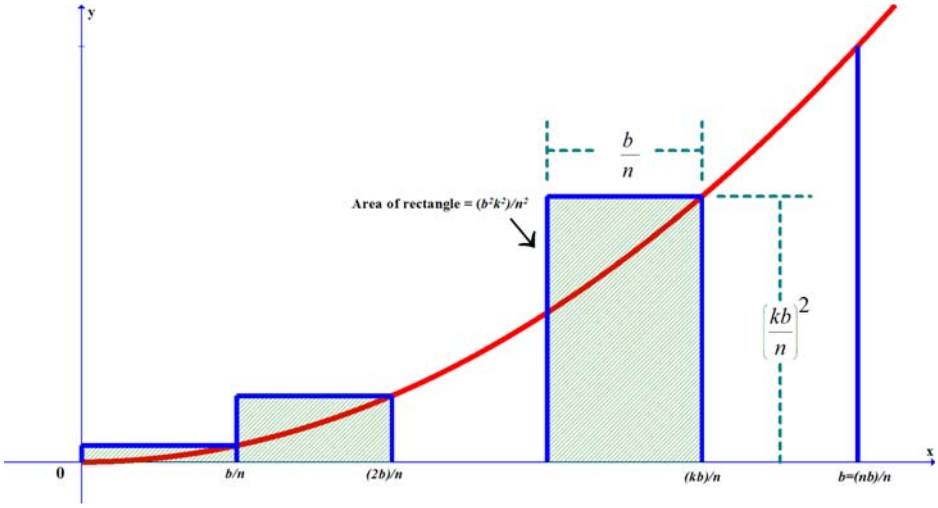


Fig. 1.6.

This brings us to a very important stage in the calculation. Notice that the factor multiplying $\frac{b^3}{n^3}$ in Equation (1.1) is the sum of the squares of the first n integers:

$$1^2 + 2^2 + ... + n^2$$

[The corresponding factor in Equation (1.2) is similar except that the sum has only $n - 1$ terms.] For a large value of n, the computation of this sum by direct addition of its terms is tedious and inconvenient. Fortunately there is an interesting identity which makes it possible to evaluate this sum in a simpler way, namely (see xxx) ,

$$1^2 + 2^2 + ... + n^2 = \frac{1}{6} n \left(2n + 1\right) \left(n + 1\right). \tag{1.3}$$

Here for our purposes, we do not need the exact expression given in the right-hand of (1.3). All we need are the two inequalities

$$1^2 + 2^2 + ... + (n - 1)^2 < \frac{n^3}{3} < 1^2 + 2^2 + ... + n^2 \tag{1.4}$$

which are valid for every integer $n \geq 1$. These inequalities can be deduced easily as consequences of (1.3). If we multiply both inequalities in (1.4) by $\frac{b^3}{n^3}$ and make use of (1.1) and (1.2) we obtain

$$s_n < \frac{b^3}{3} < S_n \tag{1.5}$$

for every n. The inequalities in (1.5) tell us that $\frac{b^3}{3}$ is a number which lies between s_n, and S_n, for every n. We will now prove that $\frac{b^3}{3}$ is the only number which has this property. In other words, we assert that if A is any number which satisfies the inequalities

$$s_n < A < S_n \tag{1.6}$$

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

To prove that $A = \frac{b^3}{3}$, we use the inequalities in (1.4) once more. Adding n^2 to both sides of the leftmost inequality in (1.4), we obtain

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

Multiplying this by $\frac{b^3}{n^3}$ and using (1.1), we find

$$S_n < \frac{b^3}{3} + \frac{b^3}{n}.$$

Similarly, by subtracting n^2 from both sides of the rightmost inequality in (1.4) and multiplying by $\frac{b^3}{n^3}$, we are led to the inequality

$$\frac{b^3}{3} - \frac{b^3}{n} < s_n.$$

Therefore, any number A satisfying (1.6) must also satisfy

$$\frac{b^3}{3} - \frac{b^3}{n} < A < \frac{b^3}{3} + \frac{b^3}{n} \tag{1.7}$$

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

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

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

Suppose the inequality $A > \frac{b^3}{3}$ were true. From the second inequality in (1.7) we obtain

$$A - \frac{b^3}{3} < \frac{b^3}{n} \tag{1.8}$$

for every integer $n \geq 1$. Since $A - \frac{b^3}{3}$ is positive, we may divide both sides of (1.8) by $A - \frac{b^3}{3}$ and then multiply by n to obtain the equivalent statement

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

for every n. The right hand side of this inequality has a constant value. This inequality is obviously false when $n \geq b^3/(A - \frac{b^3}{3})$. Hence the inequality $A > \frac{b^3}{3}$ leads to a contradiction. By a similar argument, we can show that the inequality $A < \frac{b^3}{3}$ also leads to a contradiction, and therefore we must have $A = \frac{b^3}{3}$, as asserted.

1.3 Calculus in Action

During this course you will see how calculus plays a fundamental role in all of science and engineering, as well as business and economics.

Example 1.1 *The volume of wine barrels (see Fig. 1.7) was one of the problems solved using the techniques of calculus. See a solution at Volumes by Integration (page* **??***).*

Fig. 1.7. Wine barrel.

Example 1.2 *Calculus is used to improve the efficiency of hard drives (see Fig. 1.8) and other computer components.*

Example 1.3 (Sustainable energy project in California) *A power tower produces electricity from sunlight by focusing thousands of sun-tracking mirrors, called heliostats, on a single receiver sitting on top of a tower (see Fig. 1.9). The receiver captures the thermal energy of the sun and stores it in tanks of molten salt (to the right of the tower) at temperatures greater than* 500 *degrees centigrade. When electricity is needed, the energy in the molten salt is used to create steam, which drives a conventional electricity-generating turbine (to the left of the tower).*

Fig. 1.8. Hard drive.

Fig. 1.9. A power tower in California.

1.4 Review exercises: Chapter 1

Exercise 1.1 *Modify the region in Figure 1.4 by assuming that the ordinate at each* x *is* $2x^2$ *instead of* x^2*. Draw the new figure. Check through the principal steps in the Section 1.2 and find what effect this has on the calculation of the area. Do the same if the ordinate at each* x *is*

b) $3x^2$,

c) $\frac{1}{4}x^2$,

d) $2x^2 + 1$,

e) $ax^2 + c$.

Exercise 1.2 *Modify the region in Figure 1.4 by assuming that the ordinate at each* x *is* x^3 *instead of* x^2*. Draw the new figure.*

a) *Use a construction similar to that illustrated in Figure 1.6 and show that the outer and inner sums* S_n*, and* s_n*, are given by*

$$s_n = \frac{b^4}{n^4}\left(1^3 + 2^3 + ... + (n-1)^3\right), \qquad S_n = \frac{b^4}{n^4}\left(1^3 + 2^3 + ... + n^3\right).$$

b) *Use the inequalities (which follow from the formula* $1^3 + 2^3 + ... + n^3 = \frac{n^2(n+1)^2}{4}$

$$1^3 + 2^3 + ... + (n-1)^3 < \frac{n^4}{4} < 1^3 + 2^3 + ... + n^3 \tag{1.9}$$

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

c) *What number takes the place of* $\frac{b^4}{n^4}$ *if the ordinate at each* x *is* $ax^3 + c$*?*

Exercise 1.3 *The inequalities (1.4) and (1.9) are special cases of the more general inequalities*

$$1^k + 2^k + ... + (n-1)^k < \frac{n^{k+1}}{k+1} < 1^k + 2^k + ... + n^k \tag{1.10}$$

that are valid for every integer $n \geq 1$ *and every integer* $k \geq 1$*. Assume the validity of (1.10) and generalize the results of Exercise 1.2.*

2

Logic and techniques of proof

2.1 Statements and Conditional Statements

Much of our work in mathematics deals with statements. In mathematics, a *statement* is a declarative sentence that is either true or false but not both. A statement is sometimes called a *proposition* . The key is that there must be no ambiguity. To be a statement, a sentence must be true or false, and it cannot be both. So a sentence such as "The sky is beautiful" is not a statement since whether the sentence is true or not is a matter of opinion. A question such as "Is it raining?" is not a statement because it is a question and is not declaring or asserting that something is true.

Some sentences that are mathematical in nature often are not statements because we may not know precisely what a variable represents. For example, the equation

$$2x + 5 = 10$$

is not a statement since we do not know what x represents. If we substitute a specific value for x (such as $x = 3$), then the resulting equation, $2 \cdot 3 + 5 = 10$ is a statement (which is a false statement).

How do we decide if a statement is true or false? In mathematics, we often establish that a statement is true by writing a mathematical proof. To establish that a statement is false, we often find a so-called counterexample. (These ideas will be explored later in this chapter.) So mathematicians must be able to discover and construct proofs. In addition, once the discovery has been made, the mathematician must be able to communicate this discovery to others who speak the language of mathematics.

One of the most frequently used types of statements in mathematics is the so-called conditional statement. Given statements A and B, a statement of the form "If A then B" is called a *conditional statement* or *implication*. It seems reasonable that the truth value (true or false) of the conditional statement "If A then B" depends on the truth values of A and B. The statement "If A then B" means that B must be true whenever P is true. The statement A is called the *hypothesis* of the conditional statement, and the statement B is called the *conclusion* **??**of the conditional statement. Since conditional statements

A	B	$A \Rightarrow B$
T	T	T
T	F	F
F	T	T
F	F	T

Table 2.1. Truth table for implication

are probably the most important type of statement in mathematics, we give a more formal definition.

Definition 2.1 *A conditional statement is a statement that can be written in the form "If A then B," where A and B are simpler statements. For this conditional statement, A is called the* hypothesis *and B is called the* conclusion.

Intuitively, "If A then B" means that B must be true whenever A is true. Because conditional statements are used so often, a symbolic shorthand notation is used to represent the conditional statement "If A then B." We will use the notation

$$A \Rightarrow B \tag{2.1}$$

to represent[1] "If A then B." When A and B are statements, it seems reasonable that the truth value (true or false) of the conditional statement $A \Rightarrow B$ depends on the truth values of A and B. There are four cases to consider:

A is true and B is true.	A is true and B is false.
A is false and B is true.	A is false and B is false.

The conditional statement $A \Rightarrow B$ means that B is true whenever A is true. It says nothing about the truth value of B when A is false. Using this as a guide, we define the conditional statement $A \Rightarrow B$ to be false only when A is true and B is false, that is, only when the hypothesis is true and the conclusion is false. In all other cases, $A \Rightarrow B$ is true. This is summarized in Table 2.1, which is called a truth table for the conditional statement $A \Rightarrow B$. (In Table 2.1, T stands for "true" and F stands for "false.")

The important thing to remember is that the conditional statement $A \Rightarrow B$ has its own truth value. It is either true or false (and not both). Its truth value depends on the truth values for A and B, but some find it a bit puzzling that the conditional statement is considered to be true when the hypothesis A is false. We will provide a justification for this through the use of an example.

[1] A symbol $\Rightarrow$ is called the *"forward implication arrow"*.

Example 2.2 *Suppose that I say*

"If it is not raining, then Daisy is riding her bike."

We can represent this conditional statement as $A \Rightarrow B$ where A is the statement, "It is not raining" and B is the statement, "Daisy is riding her bike."

Although it is not a perfect analogy, think of the statement $A \Rightarrow B$ as being false to mean that I lied and think of the statement $A \Rightarrow B$ as being true to mean that I did not lie. We will now check the truth value of $A \Rightarrow B$ based on the truth values of A and B.

1. Suppose that both A and B are true. That is, it is not raining and Daisy is riding her bike. In this case, it seems reasonable to say that I told the truth and that $A \Rightarrow B$ is true.

2. Suppose that A is true and B is false or that it is not raining and Daisy is not riding her bike. It would appear that by making the statement, "If it is not raining, then Daisy is riding her bike," that I have not told the truth. So in this case, the statement $A \Rightarrow B$ is false.

3. Now suppose that A is false and B is true or that it is raining and Daisy is riding her bike. Did I make a false statement by stating that if it is not raining, then Daisy is riding her bike? The key is that I did not make any statement about what would happen if it was raining, and so I did not tell a lie. So we consider the conditional statement, "If it is not raining, then Daisy is riding her bike," to be true in the case where it is raining and Daisy is riding her bike.

4. Finally, suppose that both A and B are false. That is, it is raining and Daisy is not riding her bike. As in the previous situation, since my statement was $A \Rightarrow B$, I made no claim about what would happen if it was raining, and so I did not tell a lie. So the statement $A \Rightarrow B$ cannot be false in this case and so we consider it to be true.

Hear are some examples of implication. Note that often a preliminary statement must be made, explaining what the symbols in the $A \Rightarrow B$ statement stand for.

Example 2.3

i) *A,B,C are the vertices of a triangle; a, b, c are the non-zero lengths of the opposite sides, respectively.*

$$ACB \quad \textit{is a right angle} \Rightarrow a^2 + b^2 = c^2 . \textit{(true)}$$

ii) *Let a be a real number.*

$$2a^6 + a^4 + 3a^2 = 0 \Rightarrow a = 0. \qquad \textit{(true)}$$

iii) *Let a, b, n be positive integers.*

$$n \quad \textit{divides} \quad ab \Rightarrow n \quad \textit{divides} \quad a \textit{ or } b. \qquad \textit{(false)}$$

We shall use where possible the arrow notation, since it allows the hypothesis and conclusion to stand out clearly. But with the if-then form one can avoid a preliminary sentence:

"If a is a real number such that $2a^6 + a^4 + 3a^2 = 0$, then $a = 0$."

However, the problem with all of this is that in ordinary mathematical writing, the hypothesis and conclusion may not be spelled out so clearly; it is you that has to extract them from the prose sentence. For instance, (ii) would probably appear in the form:

ii') 0 is the only real root of $2a^6 + a^4 + 3a^2 = 0$.

Thus, if a statement is given in the form $A \Rightarrow B$, some of the work has already been done for you.

Definition 2.4 *If we interchange hypothesis and conclusion in $A \Rightarrow B$, we get*

$$B \Rightarrow A \qquad (\textit{or} \quad A \Leftarrow B) \tag{2.2}$$

which is called the converse to the statement (2.1).

Example 2.5 *The converses to implications from example (2.3) are (omitting the preliminaries):*

i) $a^2 + b^2 = c^2 \Rightarrow ACB$ *is a right angle. (true)*

ii) $a = 0 \Rightarrow 2a^6 + a^4 + 3a^2 = 0$. *(true)*

iii) *n divides a or $b \Rightarrow n$ divides ab. (true)*

2.2 Equivalent statements

We can combine the two implication arrows into one double-ended arrow:

$$A \Leftrightarrow B \tag{2.3}$$

which is a true statement if both $A \Rightarrow B$ and $A \Leftarrow B$ are true. If this is so, we say A and B are *equivalent statements*. To give our examples one last time:

Example 2.6

i) $a^2 + b^2 = c^2 \Leftrightarrow ACB$ *is a right angle. (true)*

ii) $a = 0 \Leftrightarrow 2a^6 + a^4 + 3a^2 = 0$. *(true)*

iii) n *divides* a *or* $b \Leftrightarrow n$ *divides* ab. *(false)*

There are two verbal forms of $\Leftrightarrow$ which are in common use. We will mostly avoid them, but others do not, so you should know them. They are:

A if and only if B (abbreviated: A iff B)

A is a necessary and sufficient condition for B (abbreviated: *nasc*).

Occasionally these are separated into their component parts:

- $A \Rightarrow B$; A is a *sufficient* condition for B (if A is true, B follows);
- $B \Rightarrow A$; A is a *necessary* condition for B (i.e., B can't be true unless A is also true, since B implies A).

2.3 Stronger and weaker

If $A \Rightarrow B$ is true, but $B \Rightarrow A$ is false, we say: A is a *stronger statement* than B; B is *weaker* than A.

Example 2.7 *"ΔABC is equilateral" is stronger than "ΔABC is isosceles", since*

$$\Delta\ ABC \text{ is equilateral} \Rightarrow \Delta\ ABC \text{ is isosceles}$$

but

$$\Delta\ ABC \text{ is isosceles} \Rightarrow \Delta\ ABC \text{ is equilateral.}$$

is false.

The same terminology applies to entire "if-then" statements (theorems):

Example 2.8 *The if-then statement*

$$\Delta\ ABC \textit{ is equilateral} \Rightarrow \Delta\ ABC \textit{ has two equal angles} \tag{2.4}$$

can be made stronger in two different ways: make the hypothesis weaker:

$$\Delta\ ABC \textit{ is isosceles} \Rightarrow \Delta\ ABC \textit{ has two equal angles,} \tag{2.5}$$

or make the conclusion stronger:

$$\Delta\ ABC \textit{ is equilateral} \Rightarrow \Delta\ ABC \textit{ has three equal angles.} \tag{2.6}$$

Both (2.5) and (2.6) are stronger than (2.4) since they both imply (2.4): if you know (2.5) or (2.6) is true, then (2.4) follows, but not vice-versa.

Strengthen $A \Rightarrow B$ by making B stronger, or A weaker.

2.4 Contraposition and indirect proof

We turn now to discussing a style of mathematical proof which involves forming the negatives of statements.

Negation. In general, if A is a statement, we will use either not-A or $\sim A$ to denote its negation. Often the word "not" doesn't appear explicitly in the negation. Here are three examples (in the first, α is a positive integer).

A	$\sim A$
a is prime	a is composite or $a = 1$
$a > 2$	$a \leq 2$
$4a^2 + 2 = 36$	$4a^2 + 2 \neq 36$

Contraposition. In proving $A \Rightarrow B$, sometimes it is more convenient to use contraposition, i.e., prove the statement in its contrapositive form:

$$\sim B \Rightarrow \sim A \qquad \text{(contrapositive of } A \Rightarrow B\text{).}$$

This means exactly the same thing as $A \Rightarrow B$: if you prove one, you've proved the other. We will give a little argument for this later; however you will probably be even more convinced by looking at examples. Below, the original statement is on the left, the contrapositive is on the right; they say the same thing.

$$\begin{array}{ll} a \geq 0 \Rightarrow \sqrt{a} \quad \text{is real} & \sqrt{a} \quad \text{not real} \quad \Rightarrow \quad a < 0 \\ 4a^2 + 2 = 36 & 4a^2 + 2 \neq 36 \end{array}$$

Example 2.9 *Prove* $\left(2a^6 + a^4 + 3a^2 = 0\right) \Rightarrow a = 0.$

Solution: We use contraposition

$$\begin{aligned} a &\neq 0 \\ &\Rightarrow a^2 > 0, \quad a^4 > 0, \quad a^6 > 0; \\ &\Rightarrow 2a^6 + a^4 + 3a^2 > 0 \\ &\Rightarrow 2a^6 + a^4 + 3a^2 \neq 0. \end{aligned}$$

□

Indirect proof. This has the same style as contraposition but is more general. To give an indirect proof that a statement S is true, we assume it is not true and derive a contradiction, i.e., show some statement C is both true and false. C can be anything.

Example 2.10 *Prove that* $\sqrt{2}$ *is irrational.*

Solution: (Indirect proof). Suppose it were rational, that is,

$$\sqrt{2} = \frac{a}{b},$$

we may assume the fraction on the right is in lowest terms, i.e., a and b are integers with no common factor. (Call this last clause "statement C".) If we cross-multiply the above and square both sides, we get

$$2b^2 = a^2$$

the left side is even, so the right side is even, which means a itself is even (since the square of an odd number is easily seen to be odd). Thus we can write $a = 2a'$, where a' is an integer. If we substitute this into the above equation and divide both sides by 2, we get

$$b^2 = 2\left(a'\right)^2$$

by the same reasoning as before, b is even. Since we have shown both a and b are even, they have 2 as a common factor; but this contradicts the statement C [2]. □

The above (attributed to Pythagoras) is probably the oldest recorded indirect proof.

Remark 2.11 *In* 1999 *mathematicians Jack and Paul Abad, set forth on the arduous journey of generating the list of the* 100 *Greatest Theorems of Mathematics. In making the list, they used* 3 *criteria:*

- the place the theorem holds in literature,
- the quality of the proof,
- the unexpectedness of the result.

What did they come up with? Below is their top 5. For the complete list, see http://musingsonmath.com/2011/10/26/the-100-greatest-theorems-of-mathematics/.

1. The Irrationality of the Square Root of 2 by Pythagoras (500 B.C.)
2. Fundamental Theorem of Algebra by Karl Frederich Gauss (1799).
3. The Denumerability of the Rational Numbers by Georg Cantor (1867).
4. Pythagorean Theorem by Pythagoras (500 B.C.)
5. Prime Number Theorem by Jacques Hadamard and Charles-Jean de la Vallee Poussin - separately (1896).

The above Example (2.10) which occupies the top of this list, is a little atypical for us, in that almost always in this book the statement S to be proved will be an if-then statement $A \Rightarrow B$. To prove it indirectly, we have to derive a contradiction from the assumption that $A \Rightarrow B$ is false, i.e., that A does not imply B: in other words, A can be true, yet B be false. So we can now formulate

Indirect proof for if-then statements.

To prove $A \Rightarrow B$ indirectly, assume A true but B false, and derive a contradiction: C and $\sim C$ are both true. Our earlier proof by contraposition is just the special type of indirect proof where $C = A$. Namely, to prove $A \Rightarrow B$ by contraposition, we

[2] Note how the statement C to be contradicted just appears in the course of the proof; it's not part of the statement of the theorem.

a) assume A true and B false (i.e., $\sim B$ true);

b) prove $\sim B \Rightarrow \sim A$ (the contrapositive).

It follows that $\sim A$ is true, which contradicts our assumption that A is true.

To confuse you a little further, we illustrate the difference between the two styles of proof by giving two proofs of a simple proposition.

Proposition. $a^2 = 0 \Rightarrow a = 0$.

Proof by contraposition.

$$\begin{aligned} a &\neq 0 \Rightarrow a > 0 or a < 0; \\ &\Rightarrow a^2 > 0; \\ &\Rightarrow a^2 \neq 0. \qquad \square \end{aligned}$$

Indirect proof. Suppose the conclusion is false, that is, $a^2 = 0$, but $a \neq 0$. Since $a \neq 0$, we can divide both sides of the above equation by a; this gives

$$a = 0, \tag{2.7}$$

which contradicts our supposition that $a \neq 0$. $\square$

Why not just stop the proof at line (2.7) — it says $a = 0$ and isn't that what we were supposed to prove? This would be wrong; the last line of the proof is absolutely essential. We only got to line (2.7) by making a false supposition: that $a \neq 0$. Therefore (2.7) has no validity in itself; it is only a line in a bigger argument whose ultimate goal is to produce a contradiction.

The advantage of contraposition over the more general type of indirect proof is that since we know at the outset the statement A that is going to be contradicted, what has to to be proved ($\sim B \Rightarrow \sim A$) becomes a direct statement that we hope can be proved by a direct argument.

Example 2.12 *Formulate the negative statement without using "not":*

a) *In the plane, lines L and M are parallel.*

b) *Triangle ABC is isosceles.*

c) *There are infinitely many prime numbers.*

Solution:

a) L and M intersect in one point. (Why specify "one point"?)

b) Triangle ABC has sides of three different lengths.

c) There are only a finite number of primes. □

Example 2.13 *Write the converse and contrapositive, using* $\Rightarrow$*, and mark* T *or* F*:*
The square of an odd integer a *is odd.*

Solution:
Converse: a^2 odd $\Rightarrow a$ odd. (T)
Contrapositive: a^2 even $\Rightarrow a$ even. (T)

Example 2.14 *Prove the following by contraposition (the* a_i *are real numbers):*

a) *if* $a_1 a_2 < 0$*, exactly one of the* $a_i < 0$*..*

b) *if* $a_1 + ... + a_n = n$*, at least one* $a_i > 1$*.*

Solution:

a) (a) $a_1 \geq 0$ and $a_2 \geq 0 \Rightarrow a_1 a_2 \geq 0$; $a_1 < 0$ and $a_2 < 0 \Rightarrow a_1 a_2 > 0$.

b) All $a_i < 1 \Rightarrow a_1 + a_2 + ... + a_n < n$.

2.5 Counterexamples

Some statements in mathematics are particular, i.e., they assert that something is true for some definite numbers, or other objects. For example.

$$3^2 + 4^2 = 5^2; \qquad \Delta\ ABC \text{ is isosceles}; \qquad \text{there is a number} \geq\ 22.$$

Other statements are general; they assert something about a whole class of numbers or other objects. For example:

i) if a and b are numbers satisfying $a^2 = b^2$, then $a = b$;

ii) a triangle with three equal sides has three equal angles;

iii) every positive integer n is the sum of four squares of integers:

$$n = a_1^2 + a_2^2 + a_3^2 + a_4^2;$$

iv) if a, b, c are numbers satisfying $ab = ac$, then $b = c$.

These respectively assert that something is true about any numbers satisfying $a^2 = b^2$, all equilateral triangles, all positive integers, any numbers satisfying $ab = ac$.

As it happens, statements *ii*) and *iii*) are true — *iii*) is hard to prove — while *i*) and *iv*) are false. The problem we consider is:

How does one show a general statement like i) or iv) is false?

Since a general statement claims something is true for every member of some class of objects, to show it is false we only have to produce a single object in that class for which the general statement fails to hold. Such an object is called a *counterexample* to the general statement. For example a counterexample to *i*) would be the pair $a = 3$, $b = -3$. (What would be a counterexample to *iv*)?)

2.6 Mathematical induction

The principle of mathematical induction is especially useful in proving statements involving all positive integers $n \geq$ some n_0, when it is known for example that the statements $S(n)$ are valid for $n = n_0$; $n_0 + 1$; $n_0 + 2$ but it is *suspected* or *conjectured* that they hold for all positive integers $n \geq n_0$. The method of proof consists of the following steps:

1. Prove the statement $S(n)$ for $n = n_0$ (the basis step).

2. Assume the statement $S(n)$ true for $n = k$; where k is any positive integer $\geq n_0$.

3. From the assumption in 2 prove that the statement $S(n)$ must be true for $n = k + 1$ (the induction step). This is part of the proof establishing the induction and may be difficult or impossible.

4. Since the statement $S(n)$ is true for $n = n_0$ (from step 1) it must (from step 3) be true for $n = n_0 + 1$ and from this for $n = n_0 + 2$, and so on, and so $S(n)$ must be true for all positive integers $\geq n_0$.

A well-known illustration used to explain why the *Principle of Mathematical Induction* works is the unending line of dominoes represented by Figure 2.1. If the line actually contains infinitely many dominoes, it is clear that you could not knock down the entire line by knocking down only one domino at a time. However, suppose it were true that each domino would knock down the next one as it fell. Then you could knock them all down simply by pushing the first

Fig. 2.1. Dominoes in a row

one and starting a chain reaction. Mathematical induction works in the same way. If the truth of S_k implies the truth of S_{k+1} and if S_1 is true, the chain reaction proceeds as follows: S_1 implies S_2, S_2 implies S_3, S_3 implies S_4 and so on.

Example 2.15 *Use mathematical induction to prove the formula*

$$\sum_{i=1}^{n} i^2 = 1^2 + 2^2 + 3^2 + 4^2 + ... + n^2 = \frac{1}{6} n\,(2n+1)\,(n+1)$$

for all integers $n \geq 1$.

Proof.

1. When $n = 1$ the formula is valid because

$$\sum_{i=1}^{1} i^2 = 1^2 = \frac{1}{6} 1\,(2 \cdot 1 + 1)\,(1+1) = \frac{1 \cdot (2) \cdot (3)}{6}.$$

2. Assuming that

$$\sum_{i=1}^{k} i^2 = 1^2 + 2^2 + 3^2 + 4^2 + ... + k^2 = \frac{1}{6} k\,(2k+1)\,(k+1)$$

we must show that

$$\begin{aligned}\sum_{i=1}^{k+1} i^2 &= 1^2+2^2+3^2+4^2+...+k^2+(K+1)^2 \\ &= \frac{1}{6}(k+1)(2(k+1)+1)((k+1)+1) \\ &= \frac{1}{6}(k+1)(2k+3)(k+2).\end{aligned}$$

To do this we write the following :

$$\begin{aligned}\sum_{i=1}^{k+1} i^2 &= \sum_{i=1}^{k} i^2+(k+1)^2 \\ &= \frac{1}{6}k(2k+1)(k+1)+(k+1)^2 \quad \text{by assumption} \\ &= \frac{(k+1)\left(2k^2+7k+6\right)}{6} \\ & \frac{1}{6}(k+1)(2k+3)(k+2).\end{aligned}$$

Combining the results of parts 1) and 2), we can conclude by mathematical induction that the formula is valid for all integers $n \geq 1$.

■

The formula in Proposition 2.15 is one of a collection of useful summation formulas. This and other formulas dealing with the sums of various powers of the first positive integers are summarized below.

Sums of powers of Integers:

1.
$$\sum_{i=1}^{n} i = 1+2+3+4+...+n = \frac{1}{2}n(n+1) \tag{2.8}$$

2.
$$\sum_{i=1}^{n} i^2 = 1^2+2^2+3^2+4^2+...+n^2 = \frac{1}{6}n(2n+1)(n+1) \tag{2.9}$$

3.
$$\sum_{i=1}^{n} i^3 = 1^3+2^3+3^3+4^3+...+n^3 = \frac{1}{4}n^2(n+1)^2 \tag{2.10}$$

4.

$$\sum_{i=1}^{n} i^4 = 1^4 + 2^4 + 3^4 + 4^4 + ... + n^4 = \frac{1}{30} n (2n+1)(n+1)\left(3n^2 + 3n - 1\right) \tag{2.11}$$

5.

$$\sum_{i=1}^{n} i^5 = 1^5 + 2^5 + 3^5 + 4^5 + ... + n^5 = \frac{1}{12} n^2 (n+1)^2 \left(2n^2 + 2n - 1\right) \tag{2.12}$$

Each of these formulas for sums can be proven by mathematical induction (although other methods exist) . These problems are all related, and are all pretty mechanical. (See Exercise 2.15.) The following will illustrate the method of induction for many of the examples used in the rest of the book. Using mathematical induction we can, for example, easily *generalize triangle inequality.*

Example 2.16 (Proving an Inequality by Mathematical Induction) *Suppose that $a_1, a_2, ..., a_n$ are real numbers*[3]*. Prove by induction, that*

$$|a_1 + a_2 + \cdots + a_n| \leq |a_1| + |a_2| + \cdots + |a_n| .$$

Solution: Letting $S(n)$ denote $|a_1 + a_2 + \cdots + a_n| \leq |a_1| + |a_2| + \cdots + |a_n|$ we already know that both $S(1)$ and $S(2)$ are true. That is the fact that $|a_1| \leq |a_1|$ establishes the truth of $S(1)$, while the "triangle inequality" states that $|a_1 + a_2| \leq |a_1| + |a_2|$, which establishes the truth of $S(2)$. Let's assume $|a_1 + a_2 + \cdots + a_k| \leq |a_1| + |a_2| + \cdots + |a_k|$ is true. Then the required result follows as soon as we can prove that

$$|a_1 + a_2 + \cdots + a_{k+1}| \leq |a_1| + |a_2| + \cdots + |a_{k+1}| .$$

Well

$$\begin{aligned} |a_1 + a_2 + \cdots + a_{k+1}| &= |(a_1 + a_2 + \cdots + a_k) + a_{k+1}| \\ &\leq |a_1 + a_2 + \cdots + a_k| + |a_{k+1}| \\ &\leq (|a_1| + |a_2| + \cdots + |a_k|) + |a_{k+1}| \end{aligned}$$

(by the assumption that $S(k)$ is true). Since (by associative rule for addition)

$$(|a_1| + |a_2| + \cdots + |a_k|) + |a_{k+1}| = |a_1| + |a_2| + \cdots + |a_k| + |a_{k+1}|$$

[3] In fact, the same approach can be used in the complex case.

the required result follows.□

It is important to recognize that in order to prove a statement by induction, both parts of the Principle of Mathematical Induction are necessary.

Example 2.17 *Let S_n be the proposition that*

$$n! - 4^n > 0.$$

We will prove it by mathematical induction, but the initial index k_0 must be chosen with a special care here. If S_k is true, then

$$\begin{aligned}(k+1)! - 4^{k+1} &= k!\,(k+1) - 4^{k+1} \\ &> 4^k\,(k+1) - 4^{k+1} \\ &= 4^k(k-3)\end{aligned}$$

(by the induction assumption). Therefore, S_k implies S_{k+1} if $k > 3$. By trial and error, $k_0 = 9$ is the smallest integer such that S_{k_0} is true; hence, the basis step is $9! - 4^9 > 0$, and the proposition is true for $n \geq 9$ only.

We'll end this subsection by demonstrating one more use of this technique. This time we'll look at a formula for a product rather than a sum.

Proposition 2.18 *If $n \in \mathbb{Z}$ and $n \geq 2$, then*

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

Proof. (Using mathematical induction on n.)

Basis: When $n = 2$ the product has only one term, $1 - 1/2^2 = 3/4$. On the other hand, the formula is $\frac{2+1}{2\cdot 2} = 3/4$. Since these are equal, the basis is proved.

Inductive step: Let $k \geq 2$ be a particular but arbitrarily chosen integer such that

$$\prod_{j=2}^{k}(1 - \frac{1}{j^2}) = \frac{k+1}{2k}$$

Multiplying[4] both sides by the $k+1$-th term of the product gives

$$\left(1 - \frac{1}{(k+1)^2}\right)\prod_{j=2}^{k}(1 - \frac{1}{j^2}) = \left(1 - \frac{1}{(k+1)^2}\right)\frac{k+1}{2k}$$

[4] Please, notice that when you're doing the inductive step in a proof of a formula for a product, you don't add to both sides anymore, you *multiply*.

Thus

$$\begin{aligned}\prod_{j=2}^{k+1}(1-\frac{1}{j^2}) &= \left(1-\frac{1}{(k+1)^2}\right)\frac{k+1}{2k}\\ &= \frac{k+1}{2k}-\frac{k+1}{2k\,(k+1)^2}\\ &= \frac{k+1}{2k}-\frac{1}{2k\,(k+1)}\\ &= \frac{(k+1)^2-1}{2k\,(k+1)}\\ &= \frac{k^2+2k}{2k\,(k+1)}\\ &= \frac{k+2}{2\,(k+1)}\\ &= \frac{(k+1)+1}{2\,(k+1)}.\end{aligned}$$

■

Example 2.19 *For each nonnegative integer n, let x_n be a real number and suppose that*

$$|x_{n+1}-x_n| \le r\,|x_n-x_{n-1}| \quad \textit{if} \quad n \ge 1, \tag{2.13}$$

where r is a fixed positive number. By considering (2.13) for $n = 1, 2,$ and 3, we find that

$$\begin{aligned}|x_2-x_1| &\le r\,|x_1-x_0|\,,\\ |x_3-x_2| &\le r\,|x_2-x_1| \le r^2\,|x_1-x_0|\,,\\ |x_4-x_3| &\le r\,|x_3-x_2| \le r^3\,|x_1-x_0|\,.\end{aligned}$$

Therefore, we conjecture that

$$|x_n-x_{n-1}| \le r^{n-1}\,|x_1-x_0| \quad \textit{if} \quad n \ge 1. \tag{2.14}$$

This is trivial for $n = 1$. If it is true for some k, then (2.13) and (2.14) imply that

$$|x_{k+1}-x_k| \le r\left(r^{k-1}\,|x_1-x_0|\right),$$

so

$$|x_{k+1}-x_k| \le r^k\,|x_1-x_0|\,.$$

Hence (by induction) inequality (2.14) is true for every positive integer n. □

The interesting thing about the induction is that it does not just work for staff you can write down as a formula. It also works for things like geometric statement such as in the example below, where we generalize the result you know from elementary geometry, that the sum of the interior angles in a triangle is equal to π.

Example 2.20 *The sum of the interior angles of a plane n-gon with* $n \geq 3$ *vertices is* $(n-2)\pi$*.*

So we want to show, that any time we add a new corner it increases the sum of angles by π.

Solution:

- Base step, $n = 3$. The case $n = 3$ is a well-known result for triangles.
- Induction step. Idea: If we cut a corner off an $(n+1)$-gon, we get an n-gon. Let n be greater or equal to 3. We are done if cutting a corner off an $(n+1)$-gon reduces the sum of the interior angles by π. We always want to cut a corner in such a way that we go from one vertex to another, non–adjacent vertex (see Fig. 2.2). But first of all we have to find the right corner. If we just put the connection from A_{k+1} to A_{k-1}, then in case of the "convex corner" i.e. when the interior of a triangle $A_{k+1}A_kA_{k-1}$ is contained in the interior of our $(n+1)$-gon (see Fig. 2.4) then $\alpha_{k+1} + \alpha_k + \alpha_{k-1} = \pi$, and that is exactly of what is missing if we cut off that corner. It could also happen that we have another "concave" corner (see Fig. 2.3) which "cuts" our segment $A_{k+1}A_{k-1}$, but in this highly irregular case will consider (below) the triangle $A_{j+1}A_jA_{j-1}$ rather than $A_{k+1}A_kA_{k-1}$. In the case of the concave corner (see Fig. 2.5) we loose $2\pi - \alpha_k$ and wee gain $\pi - (\alpha_{k-1} + \alpha_{k+1})$, so totally we have lost π. So, in this case the result is also proven because of the n-gone assumption.

□

Example 2.21 (Towers of Hanoi) *Suppose you have three posts and a stack of n disks, initially placed on one post with the largest disk on the bottom and each disk above it is smaller than the disk below (see Fig. 2.6). A legal move involves taking the top disk from one post and moving it so that it becomes the top disk on another post, but every move must place a disk either on an empty post, or on top of a disk larger than itself. Show that for every n there is a sequence of moves that will terminate with all the disks on a post different from the original one. How many moves are required for an initial stack of n disks?*

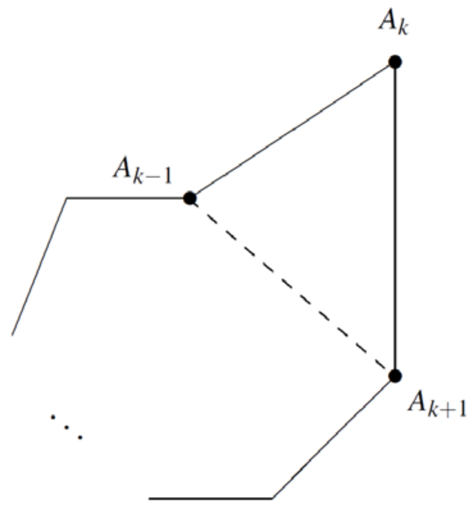


Fig. 2.2. Finding the right corner.

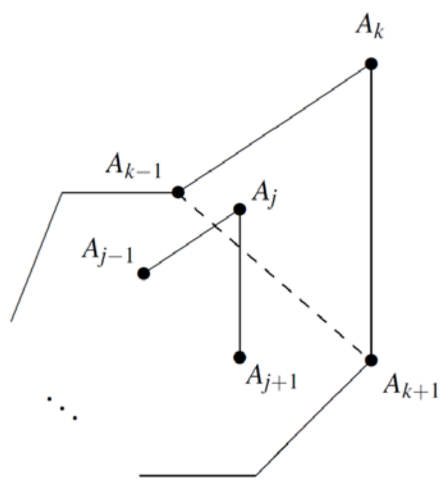


Fig. 2.3. The "concave" corner.

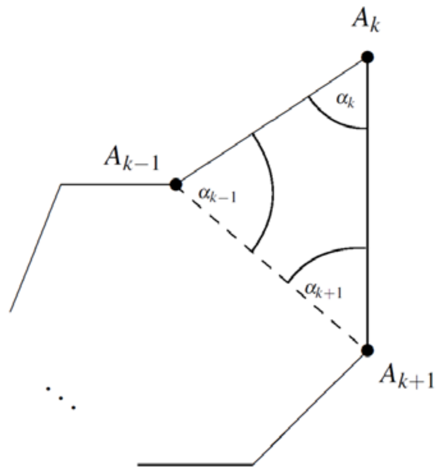


Fig. 2.4. Cutting a “convex corner”.

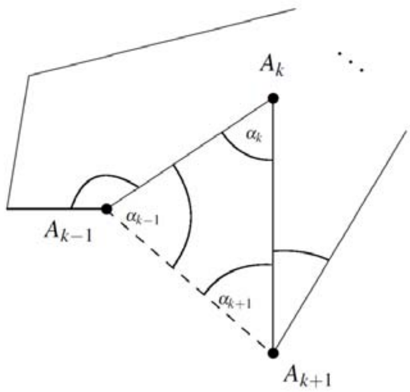


Fig. 2.5. Cutting a “concave corner”.

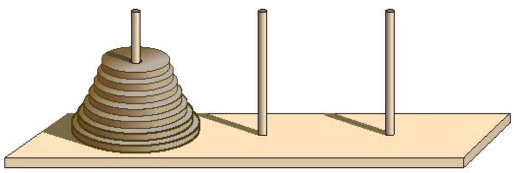


Fig. 2.6. The Towers of Hanoi

Solution: Again, this is an easy induction proof. If there is only one disk on a post, you can immediately move it to another post and you are done. If you know that it is possible to move k disks to another post, then if you initially have $k+1$ disks,move the top k of them to a different post, and you know that this is possible. Then you can move the largest disk on the bottom to the other empty post, followed by a movement of the k disks to that other post. This method, which can be shown to be the fastest possible, requires $2^k - 1$ steps to move k disks. This can also be shown by induction– if $k = 1$, it requires $2^1 - 1 = 1$ move. If it's true for stacks of size up to k disks, then to move $k+1$ requires $2^k - 1$ (to move the top k to a different post) then 1 (to move the bottom disk), and finally $2^k - 1$ (to move the k disks back on top of the moved bottom).The total for $k+1$ disks is thus $(2^k - 1) + 1 + (2^k - 1) = 2 \cdot 2^k - 1 = 2^{k+1} - 1$.

Inductive proofs not always are so easy and mechanical. Here is an example:

Example 2.22 *Prove by induction that for any positive integer n*

$$1 + \frac{1}{4} + ... + \frac{1}{n^2} \leq 2 - \frac{1}{n}.$$

Solution: Base case ($n = 1$):

$$1 \leq 2 - \frac{1}{1}.$$

Induction step: Assume that for some $1 \leq k$

$$1 + \frac{1}{4} + ... + \frac{1}{k^2} \leq 2 - \frac{1}{k}.$$

Then

$$1 + \frac{1}{4} + ... + \frac{1}{k^2} + \frac{1}{(k+1)^2} \leq 2 - \frac{1}{k} + \frac{1}{(k+1)^2},$$

so it will suffice to show

$$2 - \frac{1}{k} + \frac{1}{(k+1)^2} \leq 2 - \frac{1}{k+1}.$$

Equivalently, it suffices to show

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

But we have

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

Everything in sight is positive, so by clearing denominators, the desired inequality is equivalent to

$$k^2 + 2k = k(k+2) < (k+1)^2 = k^2 + 2k + 1$$

which is true! Thus we have all the ingredients of an induction proof, but we need to put things together in proper order, a task which we leave to the reader.□

Example 2.23 *Prove by induction that for any positive integer n*

$$1 + \frac{1}{2^r} + \frac{1}{3^r} + \frac{1}{4^r} + \frac{1}{5^r} + ... + \frac{1}{(2^n - 1)^r} \leq \frac{1 - \left(\frac{1}{2}\right)^{(r-1)n}}{1 - \left(\frac{1}{2}\right)^{(r-1)}} \qquad (r \neq 1).$$

Solution: In case $n = 1$ both sides of the inequality are 1 and so the result holds for $n = 1$. Assume that the results holds for some positive integer $k - 1$. Then

$$1 + \frac{1}{2^r} + \frac{1}{3^r} + \frac{1}{4^r} + \frac{1}{5^r} + ... + \frac{1}{(2^{k-1} - 1)^r} \leq \frac{1 - \left(\frac{1}{2}\right)^{(r-1)(k-1)}}{1 - \left(\frac{1}{2}\right)^{(r-1)}}$$

and so trying to derive the result for $n = k$ we deduce that

$$\begin{aligned}
& 1 + \frac{1}{2^r} + \frac{1}{3^r} + \frac{1}{4^r} + \frac{1}{5^r} + ... + \frac{1}{(2^k - 1)^r} \\
= & \left(1 + \frac{1}{2^r} + \frac{1}{3^r} + \frac{1}{4^r} + \frac{1}{5^r} + ... + \frac{1}{(2^{k-1} - 1)^r}\right) \\
& + \underbrace{\left(\frac{1}{(2^{k-1})^r} + \frac{1}{(2^{k-1} + 1)^r} .. + \frac{1}{(2^k - 1)^r}\right)}_{2^{k-1} \text{ terms each } \leq \frac{1}{(2^{k-1})^r}} \\
\leq & \frac{1 - \left(\frac{1}{2}\right)^{(r-1)(k-1)}}{1 - \left(\frac{1}{2}\right)^{(r-1)}} + \left(\frac{1}{2}\right)^{(r-1)(k-1)} \\
\leq & \frac{1 - \left(\frac{1}{2}\right)^{(r-1)k}}{1 - \left(\frac{1}{2}\right)^{(r-1)}}
\end{aligned}$$

Example 2.24 *Into how many regions do n straight lines "in general position" divide the plane? By "in general position" we mean that no two lines are parallel and no three lines are concurrent (see Fig. 2.7).*

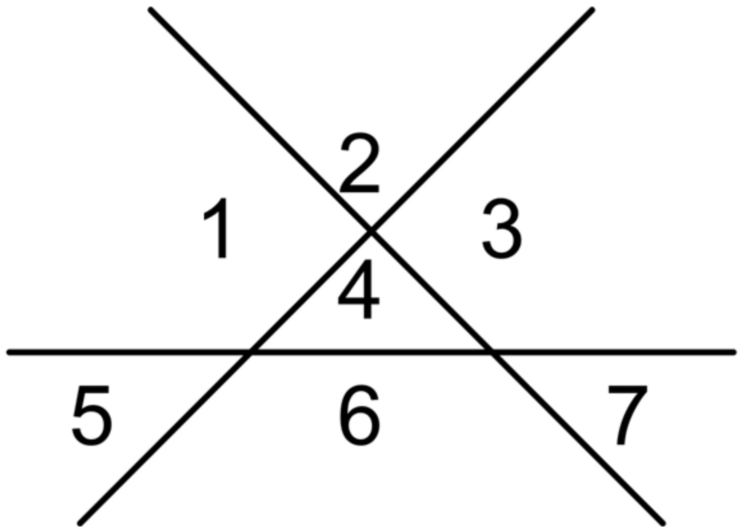


Fig. 2.7. Three straight lines "in general position"

Solution: With a bit of experimentation we can discover that the number we want is Reg(n), where the first few values are given by the table

n	1	2	3	4	5
Reg(n)	2	4	7	11	16

The relationship between n and Reg(n), of course, is not linear. If one suspects that the answer is a quadratic function of n, of the form

$$an^2 + bn + c,$$

then one can put $n = 1, 2, 3$ in succession and solve the equations

$$\begin{aligned} a + b + c &= 2 \\ 4a + 2b + c &= 4 \\ 9a + 3b + c &= 7 \end{aligned}$$

to find $a = b = 1/2$, $c = 1$.Certainly the formula

$$\mathrm{Reg}(n) = \frac{1}{2}\left(n^2 + n + 2\right)$$

is correct for $n = 1, 2, 3, 4, 5$. Suppose inductively that it is true for $1 \leq k$, so that k lines divide the plane into $\frac{1}{2}\left(k^2 + k + 2\right)$ regions. The $(k+1)$-th line intersects with each of the existing k lines in k points, and is divided into $k+1$

segments by these points. Each of these segments divides an existing region into two, and so the number of new regions created is $k+1$. Thus

$$\begin{aligned}
\operatorname{Reg}(k+1) &= \frac{1}{2}\left(k^2+k+2\right)+(k+1) \\
&= \frac{1}{2}\left(k^2+k+2+2k+2\right) \\
&= \frac{1}{2}\left[\left(k^2+2k+1\right)+(n+1)+2\right] \\
&= \frac{1}{2}\left[(k+1)^2+(k+1)+2\right]
\end{aligned}$$

and so the result is proved by induction. □

Example 2.25 (Arithmetic, Geometric, and Harmonic means) *Let*

$$A=\{a_1,a_2,...,a_n\}$$

be a set of positive numbers. We define the arithmetic, geometric, *and* harmonic means *($\mathcal{A}(A)$, $\mathcal{G}(A)$, and $\mathcal{H}(A)$, respectively) as follows:*

$$\begin{aligned}
\mathcal{A}(A) &= \frac{a_1+a_2+\cdots+a_n}{n}, \\
\mathcal{G}(A) &= \sqrt[n]{a_1a_2\cdots a_n}, \\
\mathcal{H}(A) &= \frac{n}{\frac{1}{a_1}+\frac{1}{a_2}+\cdots+\frac{1}{a_n}}.
\end{aligned}$$

Show that

$$\mathcal{H}(A)\le\mathcal{G}(A)\le\mathcal{A}(A). \tag{2.15}$$

The actual solution is proceeded below by a couple of hints.

`HINT`: Once you prove that $\mathcal{G}(A)\le\mathcal{A}(A)$ then you can find a relationship between the harmonic and geometric means that proves the inequality between those two means.

`HINT`: Prove that $\mathcal{G}(A)\le\mathcal{A}(A)$ if the set A has 2^n elements. Later, show it is true for an arbitrary number of elements.

Solution: We will first show that

$$\mathcal{G}(A)\le\mathcal{A}(A) \tag{2.16}$$

if A contains 2^n values. This can be done by induction. If $n = 1$, then Equation (2.16) amounts to:

$$a_1 \leq a_1$$

which is trivially true.
To see how the induction step works, just look at going from $n = 1$ to $n = 2$. We want to show that:

$$\sqrt{a_1 a_2} \leq \frac{a_1 + a_2}{2}.$$

Square both sides, so our problem is equivalent to showing that:

$$a_1 a_2 \leq \frac{a_1^2 + 2a_1 a_2 + a_2^2}{4}$$

or that

$$0 \leq \frac{a_1^2 - 2a_1 a_2 + a_2^2}{4} = \frac{(a_1 - a_2)^2}{4}.$$

This last result is clearly true, since the square of any number is positive. So in general, suppose it's true for sets of size $k = 2^n$ and we need to show it's true for sets of size $2k = 2^{n+1}$, or in other words show that:

$$\sqrt[2k]{a_1 a_2 \cdots a_{2k}} \leq \frac{a_1 + a_2 + \cdots + a_{2k}}{2k}. \tag{2.17}$$

Rewrite inequality 2.17 as

$$\sqrt{\sqrt[k]{a_1 a_2 \cdots a_k}\,\sqrt[k]{a_{k+1} a_{k+2} \cdots a_{2k}}} \leq \frac{a_1 + a_2 + \cdots + a_{2k}}{2k}.$$

If we let

$$\begin{aligned} a &= \sqrt[k]{a_1 a_2 \cdots a_k}, \\ b &= \sqrt[k]{a_{k+1} a_{k+2} \cdots a_{2k}}, \\ A &= \frac{a_1 + a_2 + \cdots + a_k}{k}, \\ B &= \frac{a_{k+1} + a_{k+2} + \cdots + a_{2k}}{k} \end{aligned}$$

and we know that $a \leq A$ and $b \leq B$ (because the induction hypothesis tells us so for $k = 2^n$) then we need to show that

$$\sqrt{ab} \leq \frac{A + B}{2}.$$

But we showed above that

$$\sqrt{AB} \leq (A+B)/2,$$

and we know that

$$\sqrt{ab} \leq \sqrt{AB}$$

so we are done. But of course, not all sets have a cardinality that is exactly a power of 2. Suppose we want to show that it's true for a set of cardinality m, where $2 \leq m < k = 2^n$. Our set $A = \{a_1, a_2, ..., a_m\}$ contains m elements. Let

$$u = \frac{a_1 + a_2 + \cdots + a_m}{m}. \tag{2.18}$$

If we add $k - m$ copies of u to the original members of the set A, we will have a new set $A^{'}$ with $k = 2^n$ members:

$$A^{'} = \{a_1, a_2, ..., a_m, u, u, ..., u\}.$$

Since we know that $\mathcal{G}(A^{'}) \leq \mathcal{A}(A^{'})$, we have:

$$\sqrt[k]{a_1 a_2 \cdots a_m u^{k-m}} \leq \left(\frac{a_1 + a_2 + \cdots + a_m + (k-m)u}{k} \right). \tag{2.19}$$

If we raise both sides of Equation (2.19) to the power k and do some algebra, we get:

$$\begin{aligned} & a_1 a_2 \cdots a_m u^{k-m} \\ \leq & \left(\frac{m}{k} \left(\frac{a_1 + a_2 + \cdots + a_m}{m} \right) + \frac{k-m}{k} \left(\frac{a_1 + a_2 + \cdots + a_m}{m} \right) \right)^k, \end{aligned}$$

which gives (see 2.18)

$$a_1 a_2 \cdots a_m \leq \left(\frac{m}{k} u + \frac{k-m}{k} u \right)^k u^{m-k},$$

and

$$a_1 a_2 \cdots a_m \leq u^k u^{m-k} = u^m = \left(\frac{a_1 + a_2 + \cdots + a_m}{m} \right)^m$$

This is exactly what we were trying to prove. Now to complete the problem, we need only show that $\mathcal{H}(A) \leq \mathcal{G}(A)$ for $1 \leq m$. To see this, consider the set

$$A^{''} = \{1/a_1, 1/a_2, ...1/a_m\}.$$

We know that the geometric mean is less than the arithmetic mean, so apply that fact to the set A'':

$$\sqrt[m]{\frac{1}{a_1 a_2 \cdots a_m}} = \frac{1}{\sqrt[m]{a_1 a_2 \cdots a_m}} \leq \left(\frac{\frac{1}{a_1} + \frac{1}{a_2} + \dots \frac{1}{a_m}}{m} \right).$$

If we invert both sides of the inequality (which will flip the direction), we have the desired result. □

The problem many beginners have with proof by induction is, of course, the apparent circularity: "How can you assume and use $S(k)$ in the proof, since that 's what you're trying to prove?" The answer is, it's $S(k+1)$ that you're trying to prove. The same problem appears in recursive definitions: in the briefest and most efficient form, the definition of $n!$ is

$$n! = n \cdot (n-1)! \quad 0! = 1.$$

The definition looks circular, but because the factorial on the right is for a smaller n, the definition makes sense.

In computer science, particularly, the idea of induction usually comes up in a form known as *recursion.* Recursion (sometimes known as "*divide and conquer*") is a method that breaks a large (hard) problem into parts that are smaller, and usually simpler to solve. If you can show that any problem can be subdivided into smaller ones, and that the smallest problems can be solved, you have a method to solve a problem of any size. Obviously, you can prove this *using induction.* Here's a simple example. Suppose you are given the coordinates of the vertices of a simple polygon (a polygon whose vertices are distinct and whose sides don't cross each other), and you would like to subdivide the polygon into triangles. If you can write a program that breaks any large polygon (any polygon with 4 or more sides) into two smaller polygons, then you know you can triangulate the entire thing. Divide your original (big) polygon into two smaller ones, and then repeatedly apply the process to the smaller ones you get. The concept of recursion is not unique to computer science–there are plenty of purely mathematical examples.

We can asks for a proof that the method of induction works; it is a good example of indirect proof.

Problem 2.26 *Prove the method of regular induction works: that is, if $S(k_0)$ is true and $S(k+1)$ is true whenever $S(k)$ is, for $k \geq k_0$, then $S(k)$ is true for all $k \geq n_0$.*

Proof. We will consider the set $S = \{k \geq k_0 : S(k) \text{ is false}\}$; which has a least element. We prove by indirect argument that S is empty, i.e., that $S(k)$ is true for all $k \geq k_0$. If S is non-empty, then it contains a smallest integer $m > k_0$, and $S(m)$ is false. Look at the number just before m:

$$\begin{array}{rcl} m-1 & \geq & n_0, \quad \text{since } m \geq k_0, \quad \text{and} \quad m \neq k_0 \quad \text{for} \quad S(k_0) \quad \text{is true,} \\ m-1 & \notin & S, \quad \text{since } m \text{ is the smallest number in } S. \end{array}$$

Therefore $S(m-1)$ is true; but since $S(k) => S(k+1)$ for $k \geq k_0$, it follows that $S(m)$ is true, contradiction.(Note: The self-evident fact we used:a non-empty set of positive integers has a smallest element is known as *the well-ordering property of* $\mathbb{N}$.). ■

2.7 Strong (complete) induction

When one uses in the proof of $S(n)$ not just the preceding value but lower values of n as well, the proof method is generally referred to as *strong* or *complete induction*; in this style of induction, often more than one value of n is needed for the basis step.

Example 2.27 *Prove that every integer $n > 2$ is the product of primes.*

Solution: Here is a case where you can only use strong induction, since there is no relation between the prime factorizations of n and $n+1$. If n is prime, we are done. If not, it factors into the product of two smaller positive integers, both ≥ 2 (since the factorization is not the trivial one $n = n \cdot 1$):

$$\begin{array}{rcl} n & = & n_1 \cdot n_2, \quad 2 \leq n_1, n_2 < n \\ & = & (p_1 p_2 \cdots p_k)\,(q_1 q_2 \cdots q_l) \end{array}$$

since by strong induction, we can assume the smaller numbers n_1 and n_2 factor into the product of primes. □

Example 2.28 *Let $a_1 = 2$, $a_2 = 0$, $a_3 = -14$, and*

$$a_{n+1} = 9a_n - 23a_{n-1} + 15a_{n-2}, \quad n \geq 3.$$

Show by induction, that $a_n = 3^{n-1} - 5^{n-1} + 2$, $n \geq 1$.

Solution: $S(1)$, $S(2)$ and $S(3)$ are true by inspection. Suppose that $S(k)$, $S(k-1)$ and $S(k-2)$ are true for some $k \geq 3$. Then

$$\begin{aligned} a_{k+1} &= 9\left(3^{k-1} - 5^{k-1} + 2\right) - 23\left(3^{k-2} - 5^{k-2} + 2\right) + 15(3^{k-3} - 5^{k-3} + 2) \\ &= \left(9 \cdot 3^{k-1} - 23 \cdot 3^{k-2} + 15 \cdot 3^{k-3}\right) \\ &\quad - \left(9 \cdot 5^{k-1} - 23 \cdot 5^{k-2} + 15 \cdot 5^{k-3}\right) + 2\,(9 - 23 + 15) \\ &= 3^{k-2}\,(27 - 23 + 5) - 5^{k-2}(45 - 23 + 3) + 2 \\ &= 3^k - 5^k + 2 \end{aligned}$$

which implies $S(k+1)$. Now we can use Theorem **??** □

Here we proved $S(k+1)$, using in the proof not just $S(k)$, but $S(k-1)$ and $S(k-2)$ as well.

2.8 Inductive definition

In addition to proof by induction, there is also *inductive definition* or as it is also called, *recursive definition*, in which the terms of a sequence $\{a_n\}$, $n \geq n_0$ are defined by expressing them in terms of lower values of n; as the basis, a starting value ano must also be given.

Example 2.29 *Let* $a_n = a_{n-1} + \frac{1}{n(n+1)}$*;* $a_0 = 0$*. Find a formula for* a_n*.*

Solution: We have

$$\begin{aligned} a_1 &= a_0 + 1/(1 \cdot 2) = \frac{1}{2}, \\ a_2 &= a_1 + 1/(2 \cdot 3) = \frac{2}{3}, \\ a_3 &= a_2 + 1/(3 \cdot 4) = \frac{3}{4} \end{aligned}$$

so we guess

$$a_n = \frac{n}{n+1}.$$

Taking this last statement as $S(n)$, we prove it by induction. It is true for a_0; as the induction step, we get for $k \geq 1$

$$\begin{aligned} a_k &= a_{k-1} + \frac{1}{k\,(k+1)} \\ &= \frac{k-1}{k} + \frac{1}{k\,(k+1)} \quad \text{using} \quad S(k-1) \\ &= \frac{k}{k+1} \quad \text{by algebra} \end{aligned}$$

which completes the proof by induction. □

Notice that here we proved $S(k)$, using $S(k-1)$.

3

The derivative

This chapter gives a complete definition of the derivative assuming a knowledge of high-school algebra, including inequalities, functions, and graphs. The next chapter will reformulate the definition in different language, and in the next we will prove that it is equivalent to the usual definition in terms of limits. The definition uses the concept of change of sign, so we begin with this.

3.1 Change of sign

A function is said to change sign when its graph crosses from one side of the x axis to the other. We can define this concept precisely as follows.

Definition 3.1 *Let f be a function and x_0 a real number. We say that f* changes sign from negative to positive at x_0 *if there is an open interval (a, b) containing x_0 such that f is defined on (a, b) (except possibly at x_0) and*

$$f(x) < 0 \quad \text{if} \quad a < x < x_0$$

and

$$f(x) > 0 \quad \text{if} \quad x_0 < x < b.$$

Similarly, we say that f changes sign from positive to negative at x_0 *if there is an open interval (a, b) containing x_0 such that f is defined on (a, b) (except possibly at x_0) and*

$$f(x) > 0 \quad \text{if} \quad a < x < x_0$$

and

$$f(x) < 0 \quad \text{if} \quad x_0 < x < b.$$

Notice that the interval (a, b) may have to be chosen small, since a function which changes sign from negative to positive may later change back from positive to negative (see Fig. 3.1).

Example 3.2 *Where does $f(x) = x^2 - 5x + 6$ change sign?*

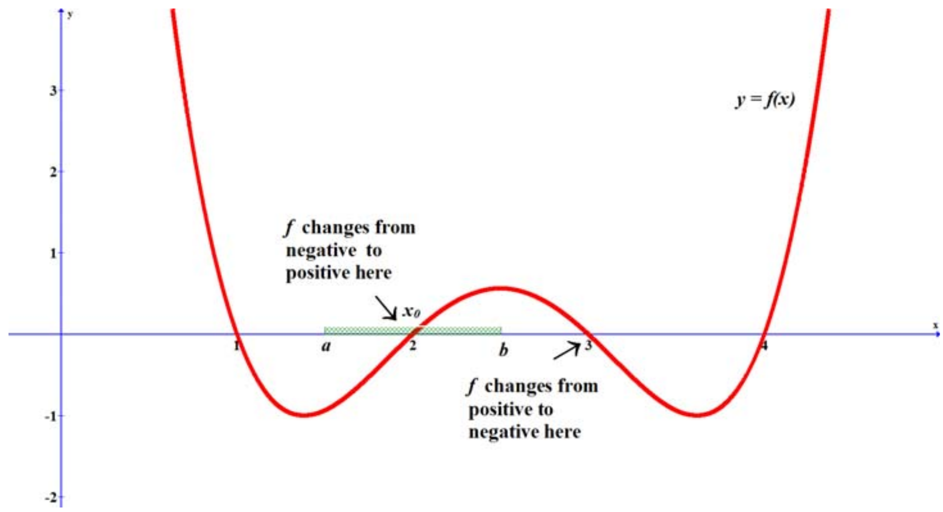


Fig. 3.1. Change of sign.

Solution: We factor f and write $f(x) = (x-3)(x-2)$. The function f changes sign whenever one of its factors does. This occurs at $x = 2$ and $x = 3$. The factors have opposite signs for x between 2 and 3, and the same sign otherwise, so f changes from positive to negative at $x = 2$ and from negative to positive at $x = 3$. (See Fig. 3.2). □

We can compare two functions, f and g, by looking at the sign changes of the difference $f(x) - g(x)$. The following example illustrates the idea.

Example 3.3 *Let* $f(x) = \frac{1}{2}x^3 - 1$ *and* $g(x) = x^2 - 1$.

a) Find the sign changes of $f(x) - g(x)$.

b) Sketch the graphs of f and g on the same set of axes.

c) Discuss the relation between the results of parts **a)** and **b)**.

Solution:

a) $f(x) - g(x) = \frac{1}{2}x^3 - 1 - (x^2 - 1) = \frac{1}{2}x^2(x-2)$. Since the factor x appears twice, there is no change of sign at $x = 0$ (x^2 is positive both for $x < 0$ and for $x > 0$). There is a change of sign from negative to positive at $x = 2$.

b) See Fig. 3.3

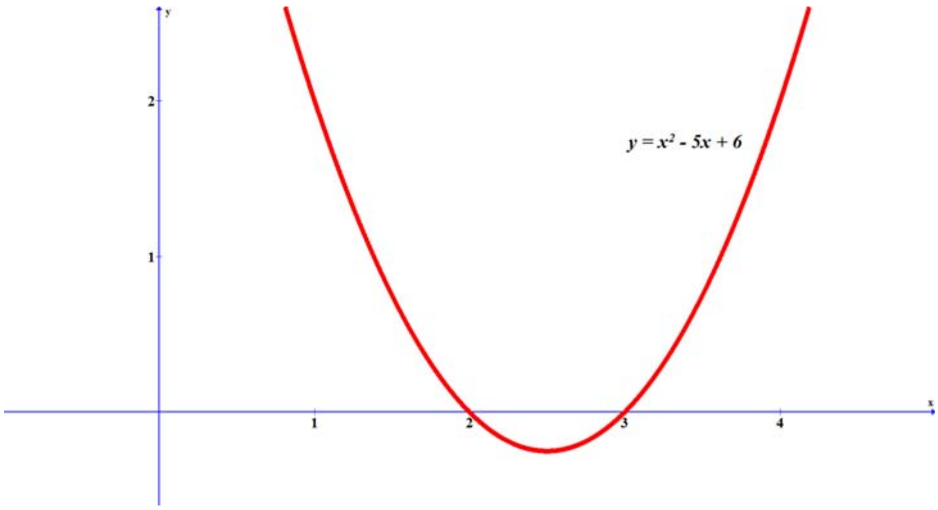


Fig. 3.2. Function $y = x^2 - 5x + 6$ changes sign at $x = 2$ and $x = 3$.

c) Since $f(x) - g(x)$ changes sign from negative to positive at $x = 2$, we can say: If x is near 2 and $x < 2$, then $f(x) - g(x) < 0$; that is, $f(x) < g(x)$. And, if x is near 2 and $x > 2$, then $f(x) - g(x) > 0$; that is,$f(x) > g(x)$. Thus the graph of f must cross the graph of g at $x = 2$, passing from below to above it as x passes 2 □.

Example 3.4 *If $f(x)$ is a polynomial and $f(x_0) = 0$, must f necessarily change* sign *at x_0?*

Solution: No. The polynomial $f(x) = x^2 - 2x + 1 = (x-1)^2$ (for example) has a root at 1, but it does not change sign there, since $(x-1)^2 > 0$ for all $x \neq 1$ □.

Example 3.5 *For which positive integers n does $f(x) = x^n$ change* sign *at zero?*

Solution: For n even, $x^n > 0$ for all $x \neq 0$, so there is no sign change. For n odd, x^n is negative for $x < 0$ and positive for $x > 0$, so there is a sign change from negative to positive at zero □.

Example 3.6 *If $r_1 \neq r_2$, describe the* sign *change at r_1 of*

$$f(x) = (x - r_1)(x - r_2).$$

Solution: The quadratic $f(x) = (x - r_1)(x - r_2)$ changes sign from positive to negative at the smaller root and from negative to positive at the larger root.

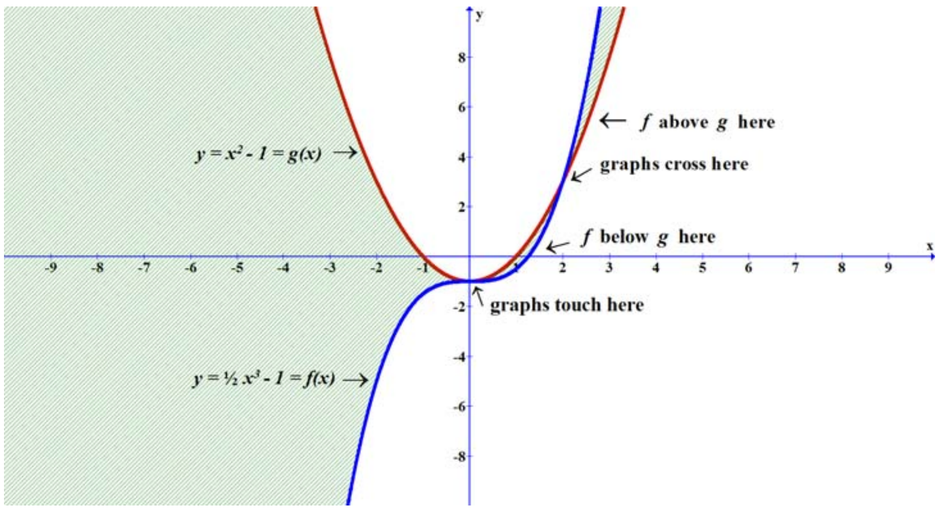


Fig. 3.3. $f - g$ changes sign when the graphs of f ans g cross.

Thus the sign change at r_1 is from negative to positive if $r_1 > r_2$ and from positive to negative if $r_1 < r_2$ $\square$.

3.2 Estimating velocities

If the position of an object moving along a line changes linearly with time, the object is said to be in *uniform motion*, and the rate of change of position with respect to time is called the *velocity*. The velocity of a uniformly moving object is a fixed number, independent of time. Most of the motion we observe in nature is not uniform, but we still feel that there is a quantity which expresses the rate of movement at any instant of time. This quantity, which we may call the *instantaneous velocity*, will depend on the time.

To illustrate how instantaneous velocity might be estimated, let us suppose that we are looking down upon a car C which is moving along the middle lane of a three-lane, one-way road. Without assuming that the motion of the car is uniform,we wish to estimate the velocity v_o of the car at exactly 3 o'clock.

Suppose that we have the following information (see Fig. 3.4): A car which was moving uniformly with velocity 95 kilometers per hour was passed by car C at 3 0 'clock, and a car which was moving uniformly with velocity 100 kilometers per hour passed car C at 3 o'clock. We conclude that v_o was at least 95 kilometers per hour and at most 100 kilometers per hour. This estimate of the velocity could be improved if we were to compare car C with more "test cars."

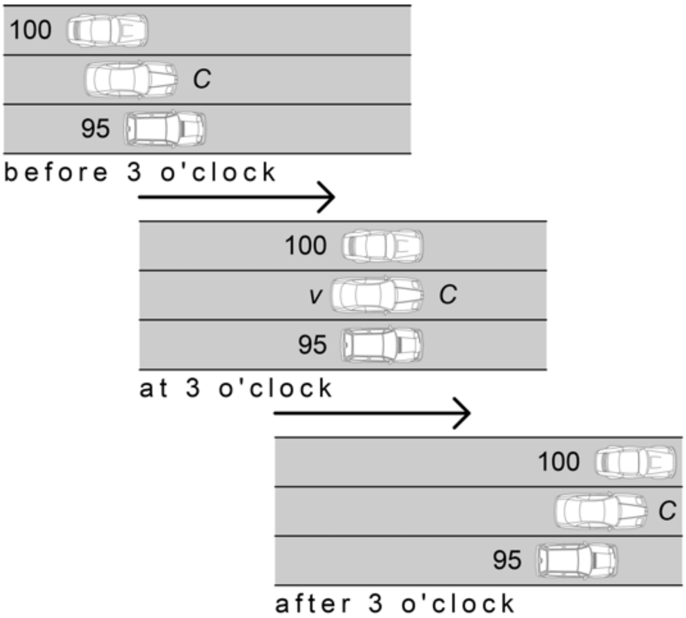


Fig. 3.4. The velocity of car C is between 95 and 100 kilometers per hour.

In general, let the variable y represent distance along a road (measured in kilometers from some reference point) and let x represent time (in hours from some reference moment). Suppose that the position of two cars traveling in. the positive direction is represented by functions $f_1(x)$ and $f_2(x)$. Then car 1 passes car 2 at time x_0 if the function $f_1(x) - f_2(x)$, which represents the "lead" of car 1 over car 2, changes sign from negative to positive at x_0. (See Fig. **??**) When this happens, we expect car 1 to have a higher velocity than car 2 at time x_0.

Since the graph representing uniform motion with velocity v is a straight line with slope v, we could estimate the velocity of a car whose motion is nonuniform by seeing how the graph of the function giving its position crosses straight lines with various slopes.

Example 3.7 *Suppose that a moving object is at position $y = f(x) = \frac{1}{2}x^2$ at time x. Show that its velocity at $x_0 = 1$ is at least $\frac{1}{2}$.*

Solution: We use a "test object" whose velocity is $v = \frac{1}{2}$ and whose position at time x is $\frac{1}{2}x$. When $x = x_0 = 1$, both objects are at $y = \frac{1}{2}$. When $0 < x < 1$, we have $x^2 < x$, so $\frac{1}{2}x^2 < \frac{1}{2}x$; when $x > 1$, we have $\frac{1}{2}x^2 > \frac{1}{2}x$. It follows that the difference $\frac{1}{2}x^2 - \frac{1}{2}x$ changes sign from negative to positive at 1, so the velocity of our moving object is at least $\frac{1}{2}$ (see Fig. **??**) $\square$.

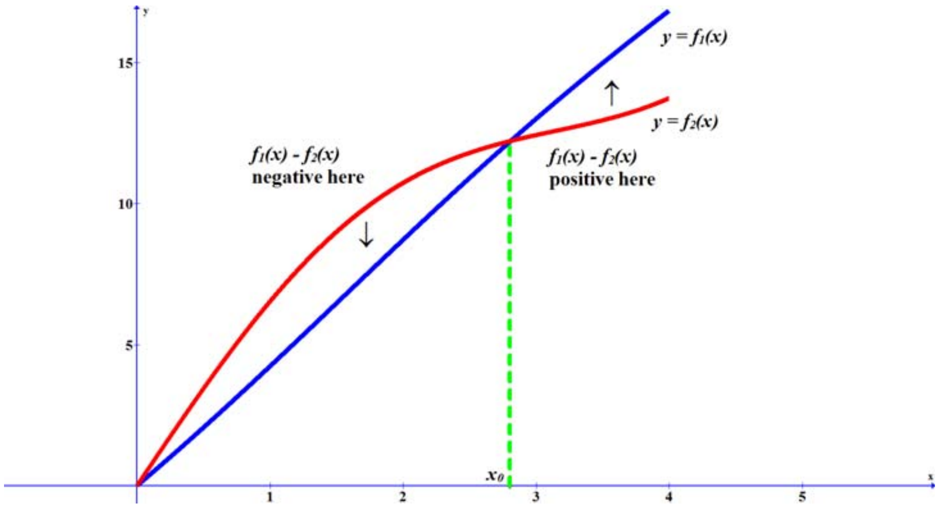


Fig. 3.5. $f_1\ - f_2$ changes sign from negative to positive at x_0.

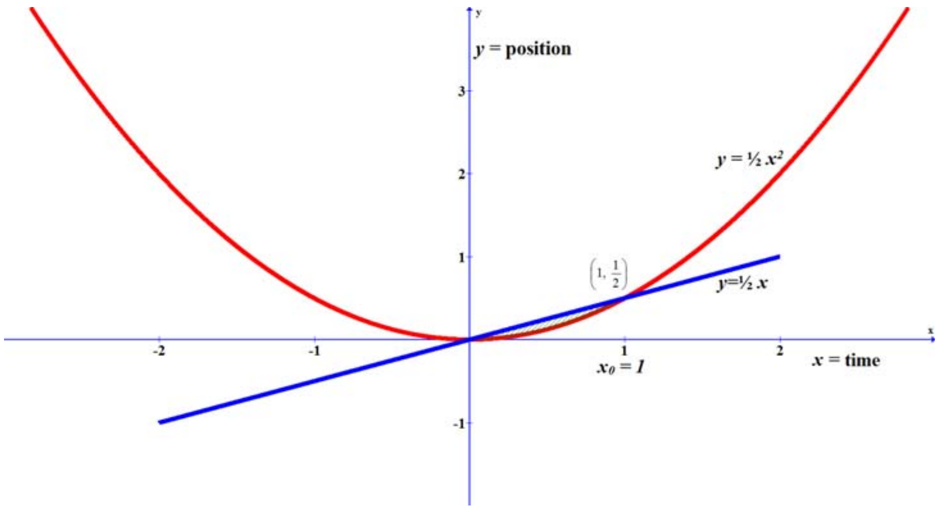


Fig. 3.6. The graph of $y = \frac{1}{2}x$ is above that of $y = \frac{1}{2}x^2$ when $0 < x < 1$ and is below when $x > 1$.

Example 3.8 *Show that the velocity at $x_0 = 1$ of the object in Example 3.7 is at most* 2

Solution: First we must find a motion $y = 2x + b$ which passes through $y = \frac{1}{2}$ when $x = 1$. We find $\frac{1}{2} = 2{\cdot}1 + b$, or b $=-\frac{3}{2}$. Now we look at the difference $\frac{1}{2}x^2 - (2x - \frac{3}{2}) = \frac{1}{2}x^2 - 2x + \frac{3}{2} = \frac{1}{2}(x^2 - 4x + 3) = \frac{1}{2}(x-3)(x-1)$. The factor $\frac{1}{2}(x-3)$ is negative near $x = 1$, so $\frac{1}{2}(x-3)(x-1)$ changes sign from positive to negative at 1. It follows that the "test" object with uniform velocity 2 passes our moving object, so its velocity is at most 2 □.

3.3 Definition of the derivative

While keeping the idea of motion and velocity in mind, we will continue our discussion simply in terms of functions and their graphs. Recall that the line through (x_0, y_0) with slope m has the equation $y - y_0 = m(x - x_0)$. Solving for y in terms of x, we find that this line is the graph of the linear function

$$l(x) = y_0 + m(x - x_0).$$

We can estimate the "slope" of a given function $f(x)$ at x_0 by comparing $f(x)$ and $l(x)$, i.e. by looking at the sign changes at x_0 of

$$f(x) - l(x) = f(x) - [f(x_0) + m(x - x_0)]$$

for various values of m. Here is a precise formulation.

Definition 3.9 *Let f be a function whose domain contains an open interval about x_0. We say that the number m_0 is the* derivative *of f at x_0, provided that:*

1. *For every $m < m_0$, the function*

$$f(x) - [f(x_0) + m(x - x_0)]$$

   changes sign from negative to positive at x_0.

2. *For every $m > m_0$, the function*

$$f(x) - [f(x_0) + m(x - x_0)]$$

   changes sign from positive to negative at x_0.

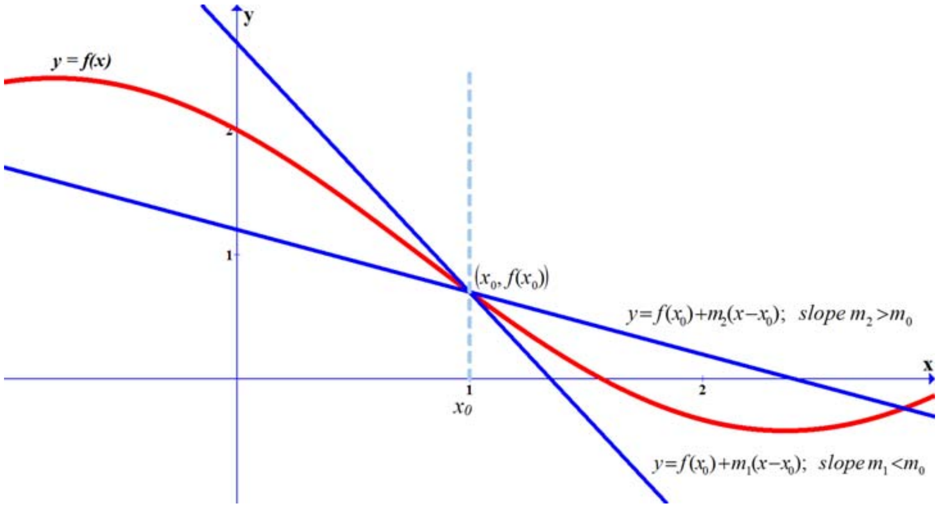


Fig. 3.7. Lines with slope different from m_0 cross the curve.

If such a number m_0 *exists, we say that* f is differentiable at x_0, *and we write* $f'(x_0) = m_0$. *If* f is *differentiable at each point of its domain, we just say that* f *is* differentiable. *The process of finding the derivative of a function is called* differentiation.

Geometrically, the definition says that lines through $(x_0, f(x_0))$ with slope less than $f'(x_0)$ cross the graph of f from above to below, while lines with slope greater than $f'(x_0)$ cross from below to above. (See Fig. 3.7.)

Given f and x_0, the number $f'(x_0)$ is uniquely determined if it exists. That is, at most one number satisfies the definition. Suppose that m_0 and $\tilde{m}_0$ both satisfied the definition, and $m_0 \neq \tilde{m}_0$; say $m_0 < \tilde{m}_0$. Let $m = (m_0 + \tilde{m}_0)/2$, so $m_0 < m < \tilde{m}_0$. By condition **1** for $\tilde{m}_0$,

$$f(x) - [f(x_0) + m(x - x_0)]$$

changes sign from negative to positive at x_0, and by condition **2** for m_0, it changes sign from positive to negative at x_0. But it can't do both! Therefore $m_0 = \tilde{m}_0$. The line through $(x_0, f(x_0))$, whose slope is exactly $f'(x_0)$ is pinched, together with the graph of f, between the "downcrossing" lines with slope less than $f'(x_0)$ and the "upcrossing" lines with slope greater than $f'(x_0)$. It is the line with slope $f'(x_0)$, then, which must be tangent to the graph of f at $(x_0, f(x_0))$.

We may take this as our definition of tangency. (See Fig. 3.8.)

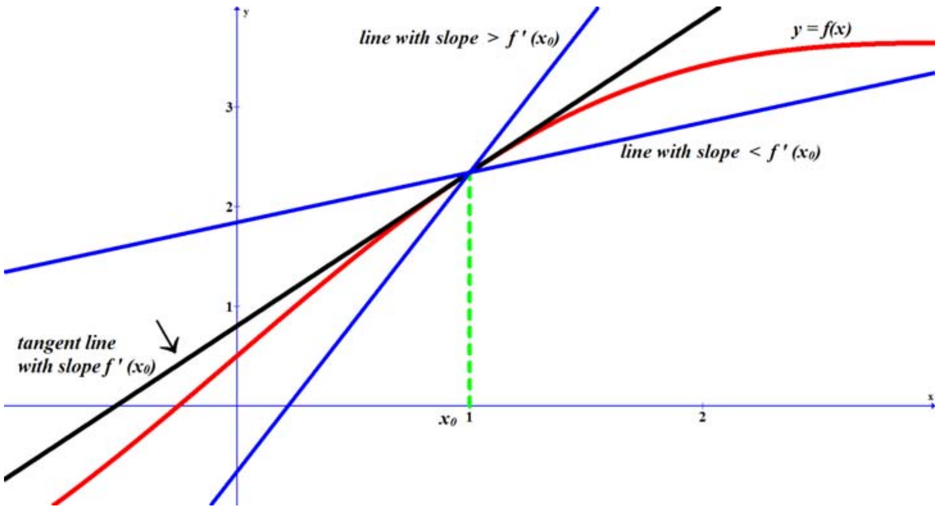


Fig. 3.8. The slope of the tangent line is the derivative.

Definition 3.10 *Suppose that the function f is differentiable at x_0. The line*

$$y = f(x_0) + f'(x_0)(x - x_0)$$

through $(x_0, f(x_0))$ with slope $f'(x_0)$ is called the tangent line *to the graph of f at $(x_0, f(x_0))$.*

Following this definition, we sometimes refer to $f'(x_0)$ as *the slope of the curve $y = f(x)$* at the point $(x_0, f(x_0))$. Note that the definitions do not say anything about how (or even whether) the tangent line itself crosses the graph of a function. (See Problem 3.4 at the end of this chapter.)

Recalling the discussion in which we estimated the velocity of a car by seeing which cars it passed, we can now give a mathematical definition of velocity.

Definition 3.11 *Let $y = f(x)$ represent the position at time x of a moving object. If f is differentiable at x_0, the number $f'(x_0)$ is called the* instantaneous velocity *of the object at the time x_0.*

More generally, we call $f'(x_0)$ the *rate of change* of y with respect to x at x_0.

Example 3.12 *Find the derivative of $f(x) = x^2$ at $x_0 = 3$. What is the equation of the tangent line to the parabola $y = x^2$ at the point $(3, 9)$?*

Solution: According to the definition of the derivative - with $f(x) = x^2$, $x_0 = 3$, and $f(x_0) = 9$ - we must investigate the sign change at 3, for various

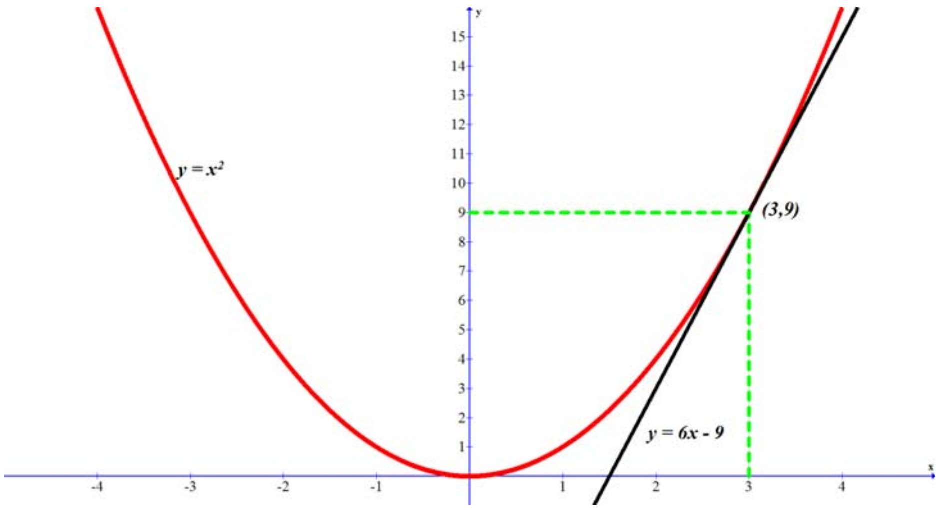


Fig. 3.9. The equation of the line tangent to $y = x^2$ at $x_0 = 3$ is $y = 6x - 9$.

values of m, of

$$\begin{aligned} f(x) - [f(x_0) + m(x - x_0)] &= x^2 - [9 + m(x - 3)] \\ &= x^2 - 9 - m(x - 3) \\ &= (x + 3)(x - 3) - m(x - 3) \\ &= (x - 3)(x + 3 - m). \end{aligned}$$

According to Example 3.6, with $r_1 = 3$ and $r_2. = m - 3$, the sign change is: From negative to positive if $m - 3 < 3$; that is, $m < 6$. From positive to negative if $3 < m - 3$; that is, $m > 6$. We see that the number $m_0 = 6$ fits the conditions in the definition of the derivative, so $f'(3) = 6$. The equation of the tangent line at $(3, 9)$ is therefore

$$y = 9 + 6(x - 3)$$

that is, $y = 6x - 9$. (See Fig. 3.9.) □

Example 3.13 *Let* $f(x) = x^3$. *What is* $f'(0)$? *What is the tangent line at* $(0, 0)$?

Solution: We must study the sign changes at $x_0 = 0$ of $x^3 - mx = x(x^2 - m)$. If $m < 0$, the factor $x^2 - m$ is everywhere positive and the product $x(x^2 - m)$ changes sign from negative to positive at $x_0 = 0$. If $m > 0$, then $(x^2 - m)$ is negative for x in $(-\sqrt{m}, \sqrt{m})$ so the sign change of $x(x^2 - m)$ at $x_0 = 0$ is from positive to negative. The number $m_0 = 0$ fits the definition of the derivative,

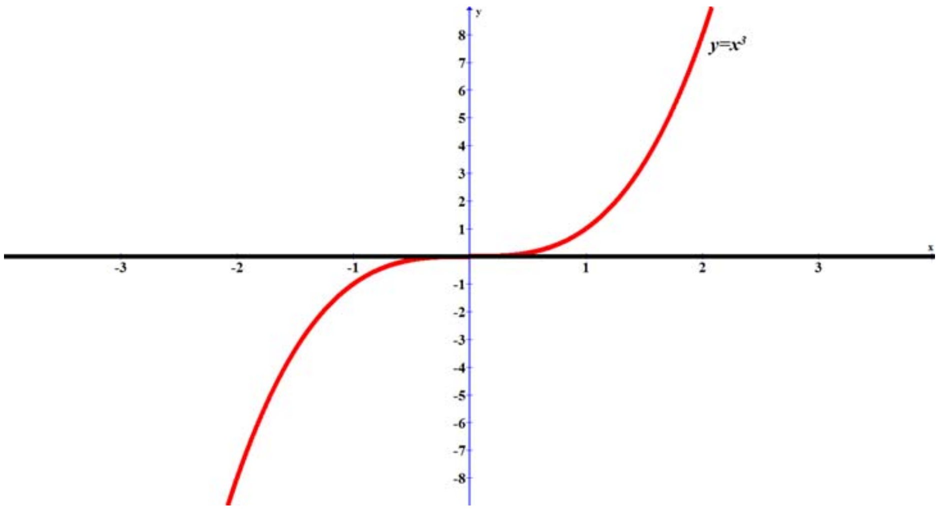


Fig. 3.10. The tangent line at $(0, 0)$ to $y = x^3$ is the x axis.

so the derivative at $x_0 = 0$ of $f(x) = x^3$ is zero. The tangent line at $(0, 0)$ has slope zero. so it is just the x axis (see Fig. 3.10). $\square$

Example 3.14 *Let f be a function for which we know that $f(3) = 2$ and $f'(3) = \sqrt[5]{8}$. Find the y intercept of the line which is tangent to the graph of f at $(3, 2)$.*

Solution: The equation of the tangent line at $(x_0, f(x_0))$ is

$$y = f(x_0) + f'(x_0)(x - x_0).$$

If $x_0 = 3,\ f(3) = 2$, and $f'(3) = \sqrt[5]{8}$, we get

$$y = 2 + \sqrt[5]{8}(x - 3) = \sqrt[5]{8}x + (2 - \sqrt[5]{8})$$

The y intercept is $2 - \sqrt[5]{8}$.

Example 3.15 *Let* $f(x) = |x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$ *(the* absolute value function *). Show by a geometric argument that f is not differentiable at zero.*

Solution: The graph of $f(x) = |x|$ is shown in Fig. 3.11. None of the lines through $(0, 0)$ with slopes between -1 and 1 cross the graph at $(0, 0)$, so there can be no m_0 satisfying the definition of the derivative. $\square$

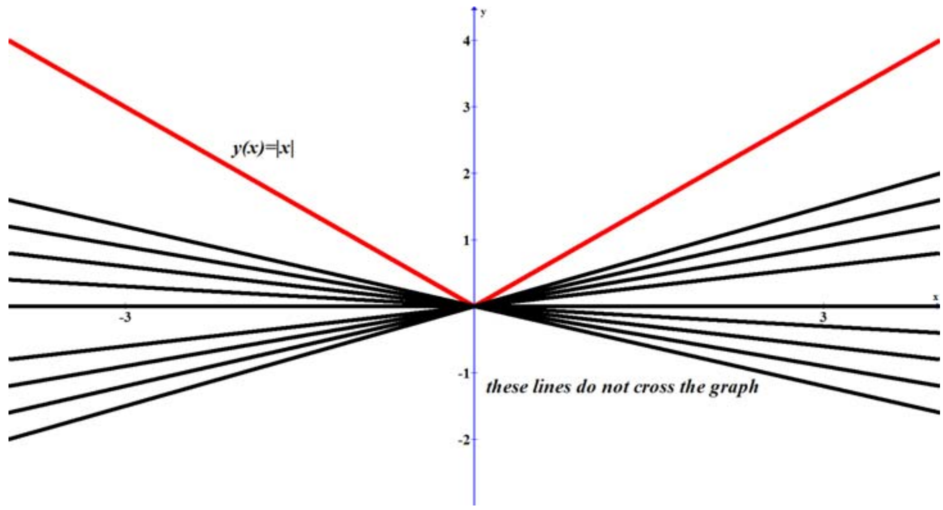


Fig. 3.11. The graph of $y = |x|$ has no tangent line at $(0, 0)$.

Example 3.16 *The position of a moving object at time x is x^2. What is the velocity of the object when $x = 3$?*

Solution: The velocity is the derivative of x^2 at $x = 3$. This derivative was calculated in Example 3.12, it is 6.□

3.4 The derivative as a function

The preceding examples show how derivatives may be calculated directly from the definition. Usually, we will not use this cumbersome method; instead, we will use *differentiation rules*. These rules, once derived, enable us to differentiate many functions quite simply. In this Chapter, we will content ourselves with deriving the rules for differentiating linear and quadratic functions. General rules will be introduced in the next chapters. The following theorem will enable us to find the tangent line to any parabola at any point.

Theorem 3.17 (Quadratic Function Rule.) *Let $f(x) = ax^2+bx+c$, where a, b, and c are constants, and let x_0 be any real number. Then f is differentiable at x_0, and $f'(x_0) = 2ax_0 + b$.*

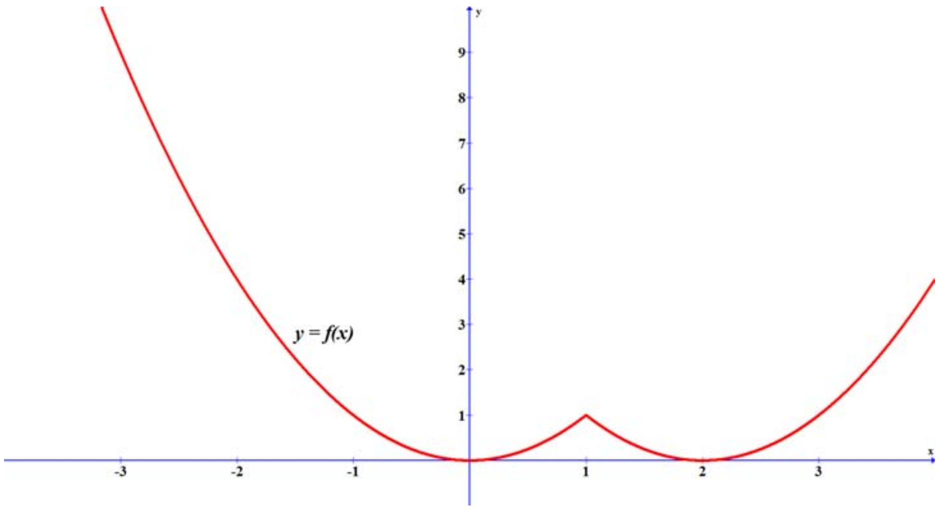


Fig. 3.12. Where is f differentiable ?

Proof. We must investigate the sign changes at x_0 of the function

$$\begin{aligned} & f(x) - [f(x_0) + m(x - x_0)] \\ = \; & ax^2 + bx + c - [ax_0^2 + bx_0 + c + m(x - x_0)] \\ = \; & a(x^2 - x_0^2) + b(x - x_0) - m(x - x_0) \\ = \; & (x - x_0)[a(x + x_0) + b - m] \end{aligned}$$

The factor $[a(x + x_0) + b - m]$ is a (possibly constant) linear function whose value at $x = x_0$ is $a(x_0 + x_0) + b - m = 2ax_0 + b - m$. If $m < 2ax_0 + b$, this factor is positive at $x = x_0$, and being a linear function it is also positive when x is near x_0. Thus the product of $[a(x + x_0) + b - m]$ with $(x - x_0)$ changes sign from negative to positive at x_0. If $m > 2ax_0 + b$, then the factor $[a(x + x_0) + b - m]$ is negative when x is near x_0, so its product with $(x - x_0)$ changes sign from positive to negative at x_0. Thus the number $m_0 = 2ax_0 + b$ satisfies the definition of the derivative, and so $f'(x_0) = 2ax_0 + b$. ■

Example 3.18 *Find the derivative at $x_0 = -2$ of $f(x) = 3x^2 + 2x - 1$.*

Solution: Applying the quadratic function rule with $a = 3, b = 2, c = -1$, and = -2, we find $f'(-2) = 2(3)(-2) + 2 = -10$.

We can use the quadratic function rule to obtain quickly a fact which may be known to you from analytic geometry. □

Example 3.19 *Suppose that $a \neq 0$. At which point does the parabola $y = ax^2 + bx + c$ have a horizontal tangent line?*

Solution: The slope of the tangent line through the point $(x_0, ax_0^2 + bx_0 + c)$ is $2ax_0 + b$. This line is horizontal when its slope is zero; that is, when $2ax_0 + b = 0$, or $x_0 = -b/2a$. The y value here is $a(-b/2a)^2 + b(-b/2a) + c = b^2/4a - b^2/2a + c = -(b^2/4a) + c$. The point $(-b/2a, -(b^2/4a) + c)$ is called the *vertex of the parabola* $y = ax^2 + bx + c$.□

In Theorem 3.17 we did not require that $a \neq 0$. When $a = 0$, the function $f(x) = ax^2 + bx + c$ is linear, so we have the following corollary:

Corollary 3.20 (Linear Function Rule.) *If $f(x) = bx + c$, and x_0 is any real number, then $f'(x_0) = b$. In particular, if $f(x) = c$, a constant function, then $f'(x_0) = 0$ for all x_0.*

For instance, if $f(x) = 3x + 4$, then $f'(x_0) = 3$ for any x_0;if $g(x) = 4$, then $g'(x_0) = 0$ for any x_0.

This corollary tells us that the rate of change of a linear function is just the slope of its graph. Note that it does not depend on x_0: the rate of change of a linear function is constant. For a general quadratic function, though, the derivative $f'(x_0)$ does depend upon the point x_0 at which the derivative is taken. In fact, we can consider f' as a new function; writing the letter x instead of x_0, we have $f'(x) = 2ax + b$.

Definition 3.21 *Let f be any function. We define the function f', with domain equal to the set of points at which f is differentiable, by setting $f'(x)$ equal to the derivative of f at x. The function $f'(x)$ is simply called the derivative of $f(x)$.*

Example 3.22 *What is the derivative of $f(x) = 3x^2 - 2x + 1$?*

Solution: By the quadratic function rule, $f'(x_0) = 2 \cdot 3x_0 - 2 = 6x_0 - 2$. Writing x instead of x_0, we find that the derivative of $f(x) = 3x^2 - 2x + 1$ is $f'(x) = 6x - 2$.□

Remark 3.23 *When we are dealing with functions given by specific formulas,we often omit the function names. For example, we could state the result of Example 3.22 as "the derivative of $3x^2 - 2x + 1$ is $6x - 2$."*

Since the derivative of a function f is another function f', we can go on to differentiate f' again. The result is yet another function, called the *second derivative* of f and denoted by f''.

Example 3.24 *Find the second derivative of $f(x) = 3x^2 - 2x + 1$.*

Solution: We must differentiate $f'(x) = 6x - 2$. This is a linear function; applying the formula for the derivative of a linear function, we find $f''(x) = 6$.

The second derivative of $3x^2-2x+1$ is thus the constant function whose value for every x is equal to 6.□

If $f(x)$ is the position of a moving object at time x, then $f'(x)$ is the velocity, so $f''(x)$ is the rate of change of velocity with respect to time. It is called "*the acceleration of the object*"

We end with a remark on notation. It is not necessary to represent functions by f and independent and dependent variables by x and y; as long as we say what we are doing, we can use any letters we wish.

Example 3.25 *Let $g(a) = 4a^2 + 3a - 2$. What is $g'(a)$? What is $g'(2)$?*

Solution: If $f(x) = 4x^2 + 3x - 2$, we know that $f'(x) = 8x + 3$. Using g instead of f and a instead of x, we have $g'(a) = 8a + 3$. Finally, $g'(2) = 8{\cdot}2 + 3 = 19$.□

Example 3.26 *Let $f(x) = 3x + 1$. What is $f'(8)$?*

Solution: $f'(x) = 3$ for all x, so $f'(8) = 3$.□

Example 3.27 *An apple falls from a tall tree toward the earth. After t seconds, it has fallen $4.9t^2$ meters. What is the velocity of the apple when $t = 3$? What is the acceleration?*

Solution: The velocity at time t is $f'(t)$, where $f(t) = 4.9t^2$. We have $f'(t) = 2(4.9)t = 9.8t$; at $t = 3$, this is 29.4 meters per second. The acceleration is $f'(r) = 9.8$ meters per second per second.□

Example 3.28 *Find the equation of the line tangent to the graph of $f(x) = 3x^2 + 4x + 2$ at the point where $x_0 = 1$.*

Solution: $f'(x) = 2 \cdot 3x + 4 = 6x + 4$, so $f'(1) = 10$. Also, $f(1) = 9$, so the equation of the tangent line is $y = 9 + 10(x - 1)$, or $y = 10x - 1$. □

Example 3.29 *For which functions $f(x) = ax^2 + bx + c$ is the second derivative equal to the zero function?*

Solution: Let $f(x) = ax^2+bx+c$. Then $f^{'}(x) = 2ax+b$ and the derivative of this is $f^{'}(x) = 2a$. Hence $f^{'}(x)$ is equal to zero when $a = 0$ - that is, when $f(x)$ is a linear function $bx + c$. □

3.5 Review exercises: Chapter 3

Exercise 3.1 *Find the* sign *changes of each of the following functions:*

$$\begin{array}{llcl} a) & f(x) & = & 2x-1, \\ b) & f(x) & = & x^2-1, \\ c) & f(x) & = & x^2, \\ d) & h(z) & = & z(z-l)(z-2). \end{array}$$

Exercise 3.2 *Describe the change of* sign *at $x = 0$ of the function $f(x) = mx$ for $m = -2, 0, 2$.*

Exercise 3.3 *Describe the change of* sign *at $x = 0$ of the function $f(x) = mx - x^2$ for $m = -l, -2, 0, 2, 1$.*

Exercise 3.4 *Sketch each of the following graphs together with its tangent line at $(0, 0)$; a) $y = x^2$ b) $y = x^3$ c) $y = -x^3$. Must a tangent line to a graph always lie on one side of the graph?*

Exercise 3.5 *Let $f(t)$ denote the angle of the sun above the horizon at time t. When does $f(t)$ change sign?*

Exercise 3.6 *Find the derivative of $f(x) = x^2$ at $x = 4$. What is the equation of the tangent line to the parabola $y = x^2$ at $(4, 16)$?*

Exercise 3.7 *If $f(x) = x^4$ what is $f'(0)$?*

Exercise 3.8 *The position at time x of a moving object is x^3. What is the velocity at $x = 0$?*

Exercise 3.9 *For which value of x_0 does the function in Fig. 3.12 fail to be differentiable?*

4

Limits and the foundations of calculus

We have developed some of the basic theorems in calculus without reference to limits. However limits are very important in mathematics and cannot be ignored. They are crucial for topics such as infinite series, improper integrals, and multivariable calculus. Historically, the concept of limit was difficult for mathematicians and scientists to figure out. The formal definition of limits was given by Augustin-Louis Cauchy (in 1800s). The Cauchy's ε-δ definition of limit is the standard used today. In this chapter we shall prove that our approach to calculus is equivalent to the usual approach via limits.

4.1 Continuity

Naively, we think of a curve as being continuous if we can draw it "without removing the pencil from the paper". Let (x_0, y_0) be a point on the curve, and draw the lines $y = c_1$ and $y = c_2$ with $c_1 < y_0 < c_2$. If the curve is continuous, at least a "piece" of the curve on each side of (x_0, y_0) should be between these lines. as in Fig. 4.1. Compare this with the behavior of the discontinuous curve in Fig. 4.2. The following definition is a precise formulation, for functions, of this idea.

Definition 4.1 *If x_0 is an element of the domain D of a function f we say that f is* continuous *at x_0 if:*

1. *For each $c_1 < f(x_0)$ there is an open interval I about x_0 such that, for those x in I which also lie in D, $c_1 < f(x)$,*
2. *For each $c_2 > f(x_0)$ there is an open interval J about x_0 such that, for those x in J which also lie in D, $f(x) < c_2$.*

If f is continuous at every point of its domain, we simply say that f is *continuous* or *f is continuous on D.*

The property by which continuity is defined might be called the "*principle of persistence of inequalities*" : f is continuous at x_0 when every strict inequality which is satisfied by $f(x_0)$ continues to be satisfied by $f(x)$ for x in some open

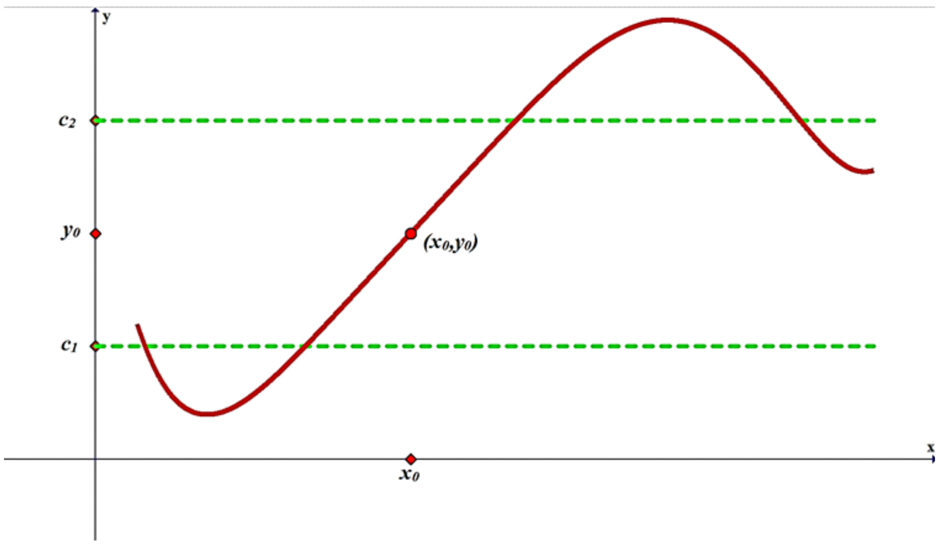


Fig. 4.1. A continuous curve.

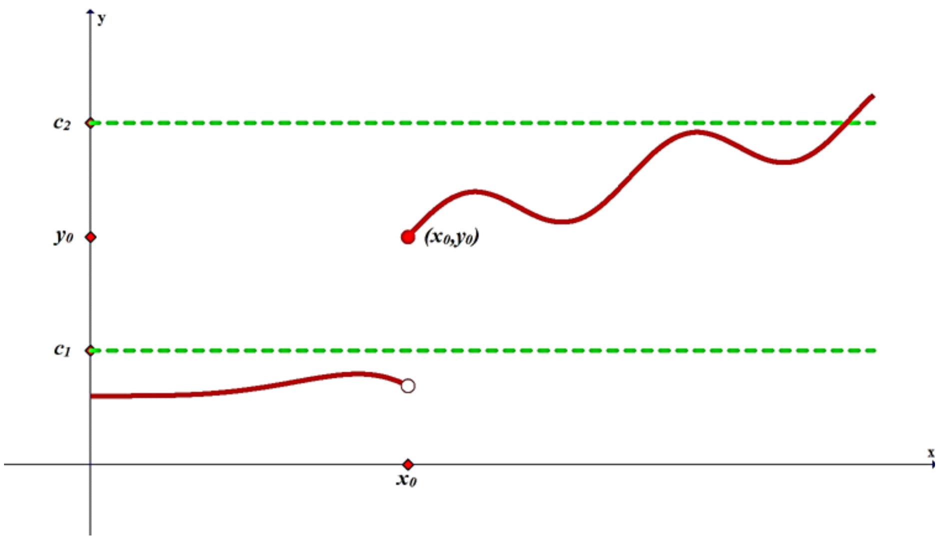


Fig. 4.2. A discontinuous curve.

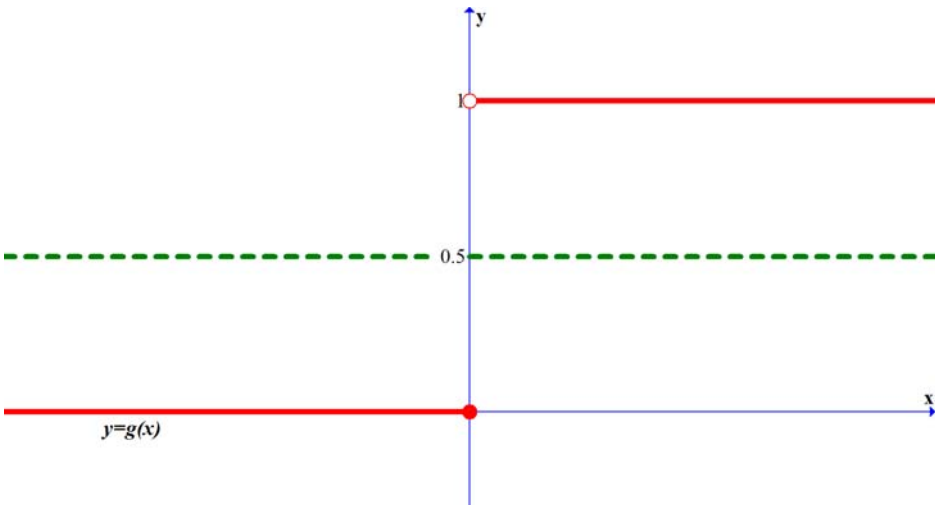


Fig. 4.3. This step function is discontinuous at 0.

interval about x_0. The intervals I and J in the definition may depend upon the value of c_1 and c_2.

Another way to paraphrase the definition of continuity is to say that $f(x)$ is close to $f(x_0)$ when x is close to x_0: The lines $y = c_1$ and $y = c_2$ in Fig. 4.1 and 4.2 provide a measure of closeness. The following example illustrates this idea.

Theorem 4.7, which appears later in this chapter, gives an easy way to verify that many functions are continuous. First, though, we try out the definition on a few simple cases in the following examples.

Example 4.2 ***Example 4.3*** *Let $g(x)$ be the step function defined by*

$$g(x) = \begin{cases} 0 & \textit{if} \quad x \le 0 \\ 1 & \textit{if} \quad x > 0. \end{cases}$$

Show that g is not continuous at $x_0 = 0$.

Solution: First sketch the graph of g (Fig. 4.3). Take $\frac{1}{2}$ for c_2. The inequality $g(0) < \frac{1}{2}$ is satisfied by g at $x_0 = 0$, since $g(0) = 0$, but no matter what open interval I we take about 0, there are positive numbers x in I for which $g(x) = 1$, which is greater than $\frac{1}{2}$. Since it is not possible to choose I such that condition 2 in the definition of continuity is satisfied, with $x_0 = 0$, $c = \frac{1}{2}$, it follows that g is not continuous at 0. □

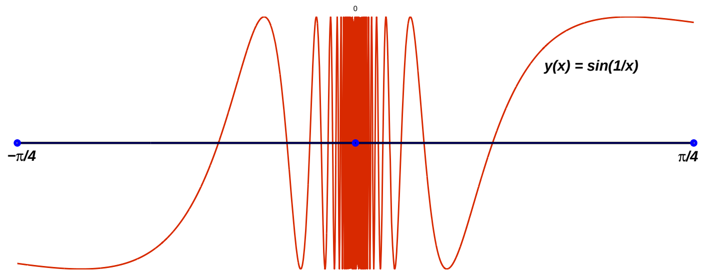


Fig. 4.4. The graph of $\sin\left(\frac{1}{x}\right)$ has infinitely many wiggles in the vicinity of the origin.

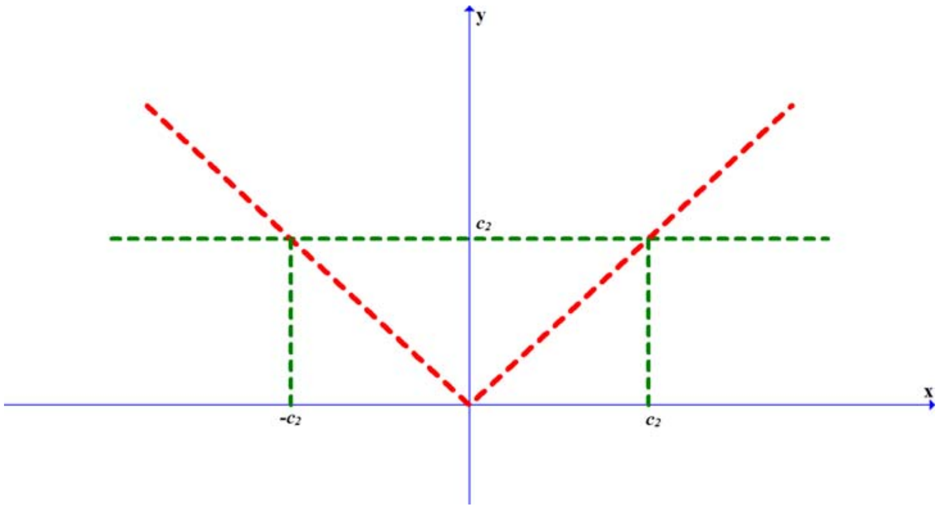


Fig. 4.5. The absolute value function is continuous at 0.

Example 4.4 *Let $f(x)$ be the absolute value function $f(x) = |x|$. Show that f is continuous at $x_0 = 0$.*

Solution: The graph of f is shown in Fig. 4.5. We must establish conditions 1 and 2 in the definition of continuity. First, we check condition 2. Let c_2 be such that $f(x_0) = f(0) = 0 < c_2$, i.e., $c_2 > 0$. We must find an open interval J about 0 such that $f(x) < c_2$ for all $x \in J$. From Fig. 4.5, we see that we should try $J = (-c_2, c_2)$. For $x \geq 0$ and $x \in J$, we have $f(x) = x < c_2$. For $x < 0$ and $x \in J$, we have $f(x) = -x$. Since $x > -c_2$, we have $-x < c_2$, i.e., $f(x) < c_2$. Thus, for all $x \in J$, $f(x) < c_2$. For condition 1, we have $c_1 < f(0) = 0$. We can take I to be any open interval about 0, even $(-\infty, \infty)$, since $c_1 < f(0) \leq f(x)$ for all real numbers x. Hence f is continuous at 0. Notice that the interval J

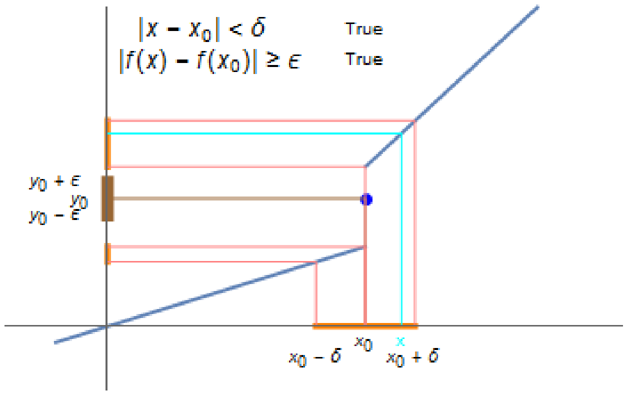


Fig. 4.6. A function with a jump discontinuity

has to be chosen smaller and smaller as $c_2 > 0$ is nearer and nearer to zero. (It is accidental to this example that the interval I can be chosen independently of c_1.) □

Example 4.5 *This is an example (see Figure 4.6) of a function with a jump discontinuity. A function f defined in some neighborhood of x_0 is discontinuous at x_0 if there exists $\varepsilon > 0$ such that for each $\delta > 0$, there exists an x such that $|x - x_0| < \delta$ and $|f(x) - f(x_0)| \geq \varepsilon$.*

Example 4.6 *Let f be continuous at x_0 and suppose that $f(x_0) \neq 0$. Show that $1/f(x)$ is defined on an open interval about x_0.*

Solution: If $f(x_0)$ is not zero, it is either positive or negative. Suppose first that $f(x_0) > 0$. In the definition of continuity, we may set $c_1 = 0$ in condition 1. We conclude that there is an open interval I about x_0 on which $0 < f(x)$, so $1/f(x)$ is defined on I. If $f(x_0) < 0$, we use condition 2 of the definition instead to conclude that $f(x) < 0$, and hence $1/f(x)$ is defined, for all x in some open interval J about x_0. □

4.2 Differentiability and continuity

If a function $f(x)$ is differentiable at $x = x_0$, then the graph of f has a tangent line at $(x_0, f(x_0))$. Our intuition suggests that if a curve is smooth

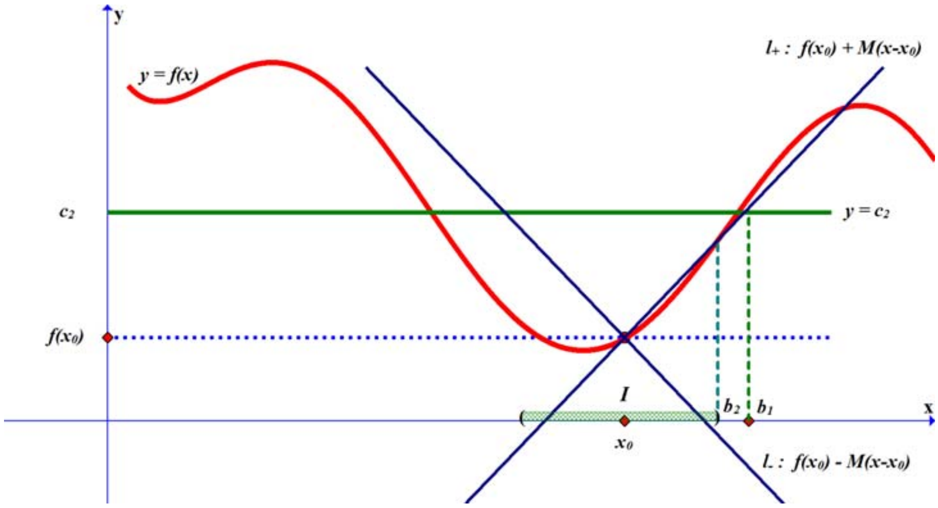


Fig. 4.7. The geometry needed for the proof of Theorem 4.7.

enough to have a tangent line then the curve should have no breaks-that is, a differentiable function is continuous. The following theorem says just that.

Theorem 4.7 *If the function f is differentiable at x_0, then f is continuous at x_0.*

Proof. We need to verify that conditions 1 and 2 of the definition of continuity hold, under the assumption that the definition of differentiability is met. We begin by verifying condition 2, so let c_2 be any number such that $f(x_0) < c_2$. We shall produce an open interval about x_0 such that $f(x_0) < c$ for all x in I. Choose a positive number M such that $-M < f'(x_0) < M$, and let l_- and l_+ be the lines through $(x_0, f(x_0))$ with slopes $-M$ and M. Referring to Fig. 4.7 we see that l_+ lies below the horizontal line $y = c_2$ for a certain distance to the right of x_0, and that the graph of f lies below l_+ for a certain distance to the right of x_0 because l_+ overtakes the graph of f at x_0. More precisely, the line $l_+ : y = f(x_0) + M(x - x_0)$ intersects $y = c_2$ at

$$b_1 = \frac{c_2 - f(x_0)}{M} + x_0 > x_0$$

and $f(x_0) + M(x - x_0) < c_2$ if $x < b_1$. (The reader should verify this.) Let (a_2, b_2) be an interval which works for l_+ overtaking the graph of f at x_0, so that

$$f(x) < f(x_0) + M(x - x_0)$$

for $x \in (x_0, b_2)$. If b is the smaller of b_1 and b_2 , then

$$f(x) < f(x_0) + M(x - x_0) < c_2 \quad \text{for} \quad x_0 < x < b. \tag{4.1}$$

Similarly, by using the line l_- to the left of x_0, we may find $a < x_0$ such that

$$f(x) < f(x_0) - M(x - x_0) < c_2 \quad \text{for} \quad a < x < x_0. \tag{4.2}$$

(The reader may wish to add the appropriate lines to Fig. 4.7.) Let $I = (a, b)$. Then inequalities (4.1) and (4.2), together with the assumption $f(x_0) < c_2$, imply that

$$f(x) < c_2 \quad \text{for} \quad x \in I,$$

so condition 2 of the definition of continuity is verified. Condition 1 is verified in an analogous manner. One begins with $c_1 < f(x_0)$ and uses the line l_+ to the left of x_0 and l_- to the right of x_0. We leave the details to the reader. ■

Example 4.8 *Show that the function $f(x) = x^2$ is continuous at $x_0 = 4$.*

Solution: We know that x^2 is differentiable everywhere. Theorem 4.7 implies that f is continuous at 4. □

This method is certainly much easier than attempting to verify directly the conditions in the definition of continuity. But be careful! As you will there exist nondifferentiable functions which are continuous!

4.3 Limits

Let f be a function defined on some open interval containing x_0, except possibly at x_0 itself, and let l be a real number. There are two definitions of the statement

$$\lim_{x \to x_0} f(x) = l$$

which is read "*the limit of $f(x)$ as x approaches x_0 is l.*"

Condition 4.9

(1) *Given any number $c_1 < l$, there is an interval (a_1, b_1) containing x_0 such that $c_1 < f(x)$ if $a_1 < x < b_1$ and $x \neq x_0$.*

(2) Given any number $c_2 > l$, there is an interval (a_2, b_2) containing x_0 such that $c_2 > f(x)$ if $a_2 < x < b_2$ and $x \neq x_0$.

Condition 4.10 *Given any positive number ε, there is a positive number δ such that $|f(x) - l| < \varepsilon$ whenever $|x - x_0| < \delta$ and and $x \neq x_0$.*

The first person to assign mathematically rigorous meanings to these two conditions was Augustin–Louis Cauchy (1789 − 1857). The ε-δ definition (Condition 4.10) of limit is the standard used today.

Depending upon circumstances, one or the other of these conditions may be easier to use[1]. The following theorem shows that they are interchangeable, so either one can be used as the definition of $\lim_{x\to x_0} f(x) = l$.

Theorem 4.11 *For any given f, x_0, and l, condition 4.10 holds if and only if condition 4.10 does.*

Proof. *Condition 4.9 implies condition* 4.10*.* Suppose that condition 4.9 holds, and let $\varepsilon > 0$ be given. To find an appropriate δ, we apply condition 4.9, with $c_1 = l - \varepsilon$ and $c_2 = 1 + \varepsilon$. By condition 4.9, there are intervals (a_1, b_1) and (a_2, b_2) containing x_0 such that

$$l - \varepsilon < f(x) \quad \text{whenever} \quad a_1 < x < b_1 \quad \text{and} \quad x \neq x_0,$$

$$l + \varepsilon > f(x) \quad \text{whenever} \quad a_2 < x < b_2 \quad \text{and} \quad x \neq x_0.$$

Now let δ be the smallest of the positive numbers $b_1 - x_0$, $x_0 - a_1$, $b_2 - x_2$, and $x_0 - a_2$ (see Fig. 4.8). Whenever

$$|x - x_0| < \delta \quad \text{and} \quad x \neq x_0,$$

we have

$$a_1 < x < b_1 \quad \text{and} \quad x \neq x_0 \tag{4.3}$$

so

$$l - \varepsilon < f(x)$$

and

$$a_2 < x < b_2 \quad \text{and} \quad x \neq x_0 \tag{4.4}$$

so

$$l + \varepsilon > f(x).$$

Statements (1) and (2) together say that $l - \varepsilon < f(x) < l + \varepsilon$, or equivalently $|f(x) - l| < \varepsilon$, which is what was required.

Condition 4.10 *implies condition* 4.9. Suppose that condition 4.10 holds, and let $c_1 < l$ and $c_2 > l$ be given. Let ε be the smaller of the two positive

[1] Throughout this text, the expression $\lim_{x\to x_0} f(x) = l$ implies two statements – the limit exists and the limit is equal to l.

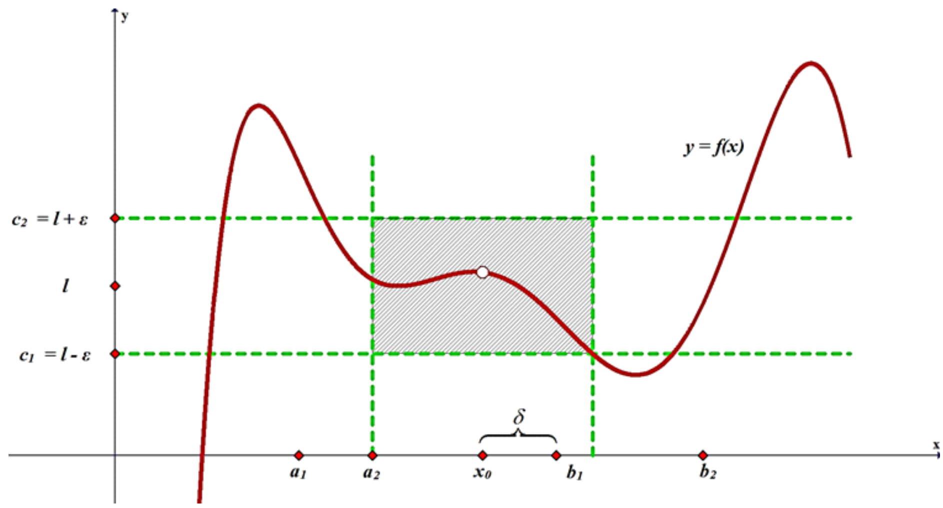


Fig. 4.8. When $|x - x_0| < \delta$ and $x \neq x_0$, $|f(x) - l| < \varepsilon$.

numbers $l - c_1$ and $c_2 - l$. By condition 4.10, there is a positive number δ such that

$$|f(x) - l| < \varepsilon \quad \text{whenever} \quad |x - x_0| < \delta \quad \text{and} \quad x \neq x_0.$$

Now we can verify parts (1) and (2) of condition 4.9, with $a_1 = b_1 = x_0 - \delta$ and $a_2 = b_2 = x_0 + \delta$. If $x_0 - \delta < x < x_0 + \delta$ and $x \neq x_0$, then $|x - x_0| < \delta$ and $x \neq x_0$, so we have $|f(x) - l| < \varepsilon$; that is, $l - \varepsilon < f(x) < l + \varepsilon$. But this implies that $c_1 < f(x)$ and $f(x) < c_2$ (see Fig. 4.9). ■

The following theorem shows that our definition of continuity can be phrased in terms of limits.

Theorem 4.12 *Let f be defined on an open interval containing x_0. Then f is continuous at x_0 if and only if*

1. *$f(x_0)$ is defined,*
2. *$\lim_{x \to x_0} f(x)$ exists*
3. *$\lim_{x \to x_0} f(x) = f(x_0)$.*

Proof. The definition of continuity (4.1) given on page 79 is exactly condition **1** for the statement $\lim_{x \to x_0} f(x) = f(x_0)$. ■

Corollary 4.13 *The function f is continuous at x_0 if and only if, for every positive number ε, there is a positive number δ such that*

$$|f(x) - f(x_0)| < \varepsilon$$

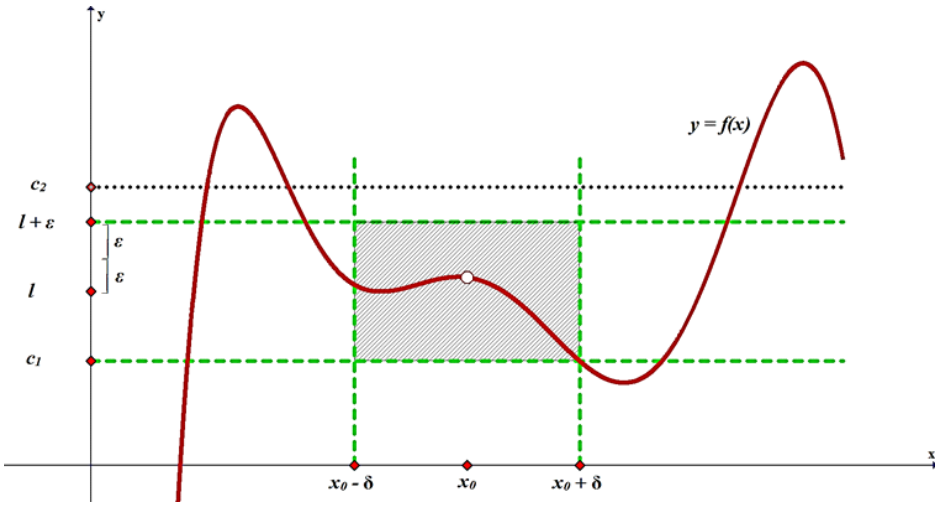


Fig. 4.9. When $x \in (x_0 - \delta, x_0 + \delta)$ and $x \neq x_0$, $c_1 < f(x) < c_2$.

whenever $|x - x_0| < \delta$.

Proof. We have simply replaced the statement $\lim_{x\to x_0} f(x) = f(x_0)$ by its condition **2** definition. (We do not need to require that $x \neq x_0$; if $x = x_0$, $|f(x) - f(x_0)| = 0$, which is certainly less than $\varepsilon > 0$.) ■

Definition 4.14 (Continuity on an open interval) *A function is* continuous on an open interval (a, b) *if it is continuous at each point in the interval. A function that is continuous on the entire real line* $(-\infty, \infty)$ *, is* everywhere continuous.

Notice that f doesn't actually have to be continuous at the endpoints $x = a$ or $x = b$. For example, if $f(x) = 1/x$, then f is continuous on the interval $(0, \infty)$ even though $f(0)$ isn't defined. This function is also continuous on $(-\infty; 0)$, but not on $(-2; 3)$, since 0 lies within that interval, and f isn't continuous there.

Example 4.15 *Using corollary 4.13 we will prove now, that* $f(x) = \cos(x)$ *is continuous for all* $x \in \mathbb{R}$*. Let we choose an arbitrary point* $x_0 \in \mathbb{R}$*. We need to prove that for all* $\varepsilon > 0$ *exists* $\delta > 0$ *so if* $|x - x_0| < \delta$ *then* $|\cos(x) - \cos(x_0)| < \varepsilon$*. But*

$$\begin{aligned} |\cos(x) - \cos(x_0)| &= \left| -2\sin\left(\frac{x + x_0}{2}\right)\sin\left(\frac{x - x_0}{2}\right)\right| \\ &\leq 2\left|\frac{x - x_0}{2}\right| * 1 = |x - x_0|, \end{aligned}$$

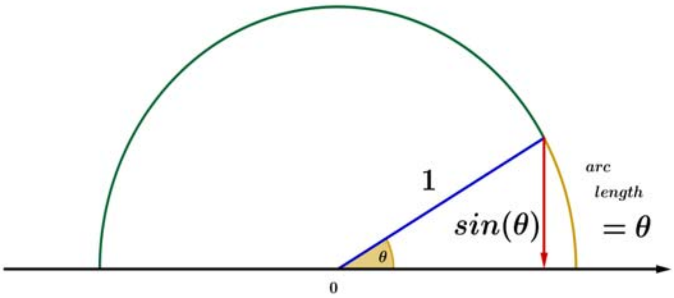


Fig. 4.10. A circle of radius 1 with an arc of angle θ.

as $|\sin(\theta)| \leq |\theta|$ *(see Fig. 4.10. Since the radius of the circle is* 1*,* $\sin(\theta) = \frac{|opposite|}{|hypotenuse|}$ *equals the length of the edge indicated.) So, if* $|x - x_0| < \varepsilon$ *then* $|\cos(x) - \cos(x_0)| < \varepsilon$ *and it is enough to choose* $\delta = \varepsilon$*. We proved that* $f(x) = \cos(x)$ *is continuous at some point* $x_0 \in \mathbb{R}$*. As* x_0 *was chosen arbitrarily, we conclude that our function is continuous for all* $x \in \mathbb{R}$*.* □

Remark 4.16 *In a similar fashion we can prove that* $f(x) = \sin(x)$ *is continuous for all* $x \in \mathbb{R}$*.*

Remark 4.17 *Continuity of* $\sin(x)$ *and* $\cos(x)$ *will follow directly from the differentiability of those functions.*

4.4 A little game

To familiar ourselves with the concept of a limit let us consider a simple game. Here's how the game works. Your move consists of picking an interval on the y-axis with l in the middle. You get to draw lines parallel to the x-axis through the endpoints of your interval. Here's an example of what your move might be (see Figure 4.11):Notice that the endpoints of the interval are labeled as $l - \varepsilon$ and $l + \varepsilon$. So both endpoints are a distance ε away from l.

Anyway, the point is, you can't tolerate any bit of the function being outside those two horizontal lines. My move, then, is to throw away some of the function by restricting the domain. I just have to make sure that the new domain is an interval with a at the center, and that every bit of the function remaining lies between your lines, except possibly at $x = a$ itself. Here's one way I could make my move, based on the move you just made (Figure 4.12):I could have taken away more and it would still have been fine–as long as what's left is

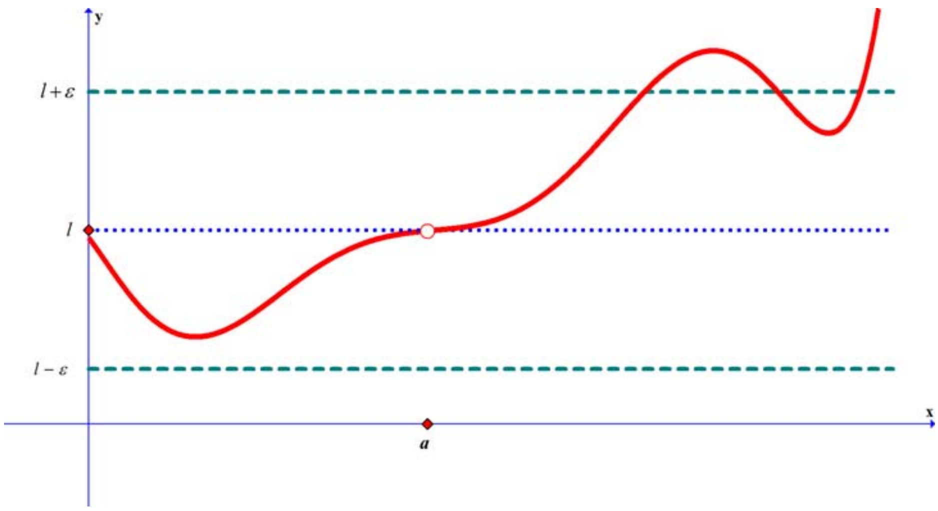


Fig. 4.11. Your first move in the game.

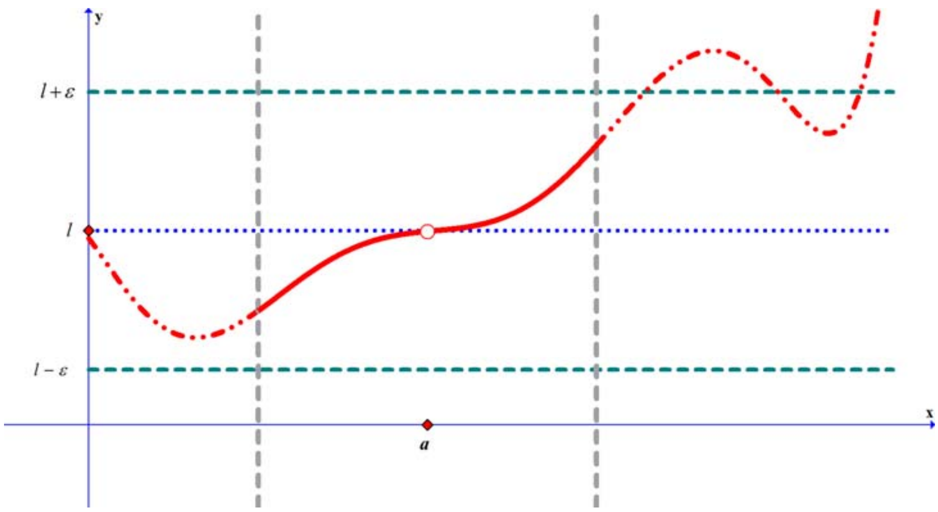


Fig. 4.12. My move in the game

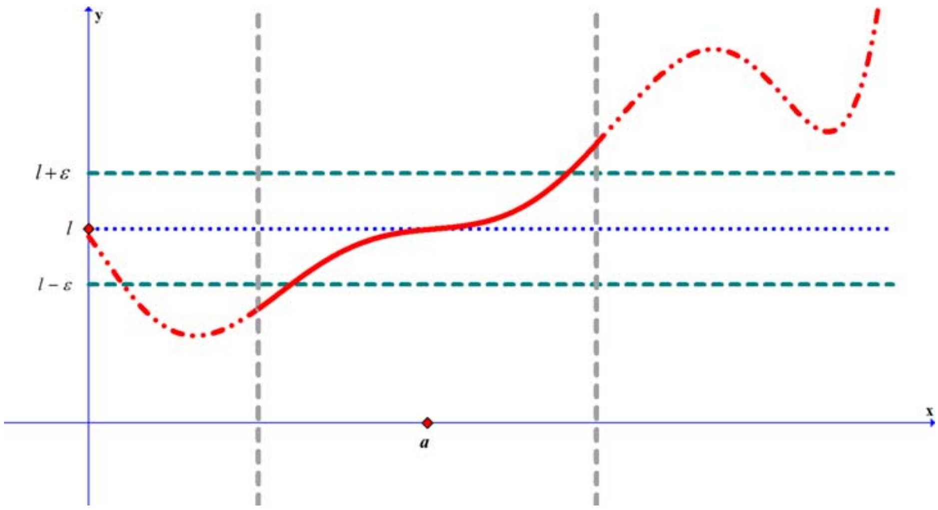


Fig. 4.13. The situation after your second move

between your lines. Now it's your move again. You have realized that my task is harder when your lines are closer together, so this time you pick a smaller value of ε. Here's the situation after your second move:Parts of the curve are outside the horizontal lines again, but I haven't had my second move yet. I'm going to throw away more of the function away from $x = a$, like this:So once again I was able to make a move to counter your move. When does the game stop? Hopefully, the answer is never! If I can always move, no matter how close together you make the lines, then it will indeed be true that $\lim_{x\to a} f(x) = l$. We will have zoomed in and in, you pushing your lines closer together, I responding by focusing only on the part of the function close enough to $x = a$. On the other hand, if I ever get stuck for a move, then it's not true that $\lim_{x\to a} f(x) = l$. The limit might be something else, or it may not exist, but it's definitely not l.

Example 4.18 *Prove that* $\lim_{x\to 3} x^2 = 9$.

Solution: We must show, that given any positive number ε we can find a positive number δ such that

$$\left|x^2 - 9\right| < \varepsilon \tag{4.5}$$

whenever x satisfies

$$0 < |x - 3| < \delta. \tag{4.6}$$

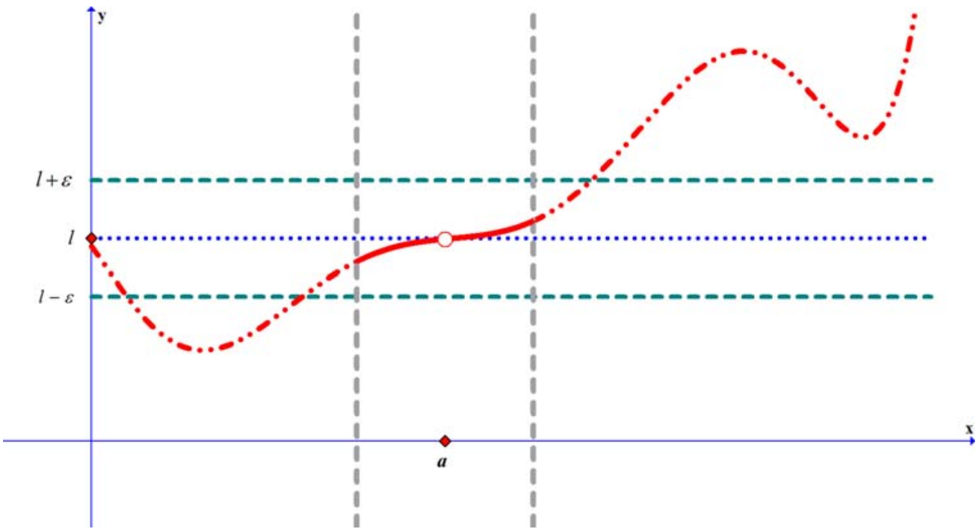


Fig. 4.14. Once again I was able to make a move to counter your move.

Because $|x-3|$ occurs in (4.6), it will be helpful to rewrite (4.5) so that $|x-3|$ appears as a factor on the left side. Therefore, we shall rewrite (4.5) as

$$|x+3|\,|x-3| < \varepsilon. \tag{4.7}$$

If we can somehow ensure that when x satisfies (4.6) the factor $|x+3|$ remains less than some positive constant, say,

$$|x+3| < k \tag{4.8}$$

then on choosing

$$\delta = \frac{\varepsilon}{k} \tag{4.9}$$

it will follow from (4.6) that

$$0 < |x-3| < \frac{\varepsilon}{k}$$

or

$$0 < k\,|x-3| < \varepsilon. \tag{4.10}$$

Form (4.8) and the right-hand inequality in (4.10) we shall then have

$$|x+3|\,|x-3| < k\,|x-3| < \varepsilon.$$

so that (4.7) will be satisfied, and the proof will be complete. We can assume in our discussion, that δ satisfies

$$0 < \delta \leq 1. \tag{4.11}$$

Assume, that x satisfies (4.6). From (4.11) and the right side of (4.6) we obtain

$$|x-3| < \delta \le 1$$

so that

$$|x-3| < 1$$

or equivalently

$$2 < x < 4$$

so

$$5 < x+3 < 7.$$

Therefore

$$|x+3| < 7.$$

Comparing the last inequality to (4.8) suggests $k = 7$; and from (4.9)

$$\delta = \frac{\varepsilon}{k} = \frac{\varepsilon}{7}.$$

In summary, given $\varepsilon > 0$, we choose

$$\delta = \min\{1, \frac{\varepsilon}{7}\}.$$

□

4.5 Making new limits from old ones

The Example **??** was pretty annoying. Just to show that $\lim_{x\to 3} x^2 = 9$, we had to do a lot of work. Luckily it turns out that once you know a couple of limits, you can put them together and get a whole bunch of new ones. For example, you can add, subtract, multiply, and divide limits within reason, and there's also the sandwich principle. Let's see why all this is true.

4.5.1 Sums and differences of limits

Suppose that we have two functions f and g, and we know that $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$. What should happen to $\lim_{x\to a} (f(x)+g(x))$? Intuitively, it should be equal to $L+M$. Let's prove this using the definition. So, we know that

$$\lim_{x\to a} f(x) = L \qquad \text{and} \qquad \lim_{x\to a} g(x) = M.$$

This means that if you pick $\varepsilon > 0$, I can ensure that $|f(x) - L| < \varepsilon$ by restricting x close enough to a. I can also ensure that $|g(x) - M| < \varepsilon$ if x

is close enough to a. The degrees of closeness that I need might be different for f and g, but it doesn't matter–I can just go close enough so that both inequalities work. Now, if $f(x) + g(x)$ is close to $L + M$, this means that the difference between these things should be small. So we'll need to worry about the quantity $|(f(x) + g(x)) - (L + M)|$. We'll write this as

$$|(f(x) - L) + (g(x) - M)|$$

We can then use the triangle inequality, which says that $|a + b| \leq |a| + |b|$ for any numbers a and b, as follows:

$$|(f(x) - L) + (g(x) - M)| \leq |f(x) - L| + |g(x) - M| < \varepsilon + \varepsilon = 2\varepsilon,$$

provided that x is close enough to a. This is almost good enough, except that you wanted a tolerance of ε, not 2ε! So I have to make my move again; this time I'll narrow my focus so that both $|f(x) - L|$ and $|g(x) - M|$ are less than $\varepsilon/2$ instead of ε. This is no problem, since I can deal with any positive number that you pick. Anyway, if you redo the above equation, you'll get ε on the right instead of 2ε, so we have proven that I can find a little window about $x = a$ such that

$$|(f(x) - L) + (g(x) - M)| \leq |f(x) - L| + |g(x) - M| < \varepsilon$$

whenever x is in my window. (You can use δ if you like to describe the window better, but that doesn't really get us anything extra.) So this proves the following

$$\text{if} \quad \lim_{x \to a} f(x) = L \quad \text{and} \quad \lim_{x \to a} g(x) = M, \quad \text{then} \quad \lim_{x \to a} (f(x) + g(x)) = L + M. \tag{4.12}$$

That is, the limit of the sum is the sum of the limits. Another way of writing this is

$$\lim_{x \to a} (f(x) + g(x)) = \lim_{x \to a} f(x) + \lim_{x \to a} g(x)$$

but here you have to be careful to check that both limits on the right exist and are finite. If either limit doesn't exist, the deal's off . Both limits have to be finite to guarantee that you can add them up. You might get lucky if they're not, but there's no guarantee.

How about $\lim_{x \to a} (f(x) - g(x))$? That should go to $L - M$, and it does:

$$\text{if} \quad \lim_{x \to a} f(x) = L \quad \text{and} \quad \lim_{x \to a} g(x) = M, \quad \text{then} \quad \lim_{x \to a} (f(x) - g(x)) = L - M.$$

The proof is almost identical to the one we just looked at, except that you need a slightly different form of the triangle inequality: $|a - b| \leq |a| + |b|$. Actually, this is just the triangle inequality applied to a and $-b$; indeed, $|a + (-b)| \leq |a| + |-b|$, but of course $|-b|$ is equal to $|b|$. I leave it to you to rewrite the above argument but change the plus signs between $f(x)$ and $g(x)$, and between L and M, into minus signs.

4.5.2 Products of limits

Now we once again assume that we have two functions f and g such that

$$\lim_{x \to a} f(x) = L \quad \text{and} \quad \lim_{x \to a} g(x) = M.$$

We want to show that

$$\lim_{x \to a} f(x)g(x) = LM. \tag{4.13}$$

That is, the limit of the product is the product of the limits. Another way of writing this is

$$\lim_{x \to a} f(x)g(x) = \lim_{x \to a} f(x) \times \lim_{x \to a} g(x)$$

again with the understanding that both limits on the right-hand side are already known to exist and be finite. To prove this, we need to show that the difference between $f(x)g(x)$ and the (hopeful) limit LM is small. Let's consider that difference $f(x)g(x) - LM$. The trick is to subtract $Lg(x)$ and add it back on again! That is,

$$f(x)g(x) - LM = f(x)g(x) - Lg(x) + Lg(x) - LM.$$

What does that get us? Let's take absolute values, then use the triangle inequality:

$$\begin{aligned} |f(x)g(x) - LM| &= |f(x)g(x) - Lg(x) + Lg(x) - LM| \\ &\leq |(f(x) - L)\, g(x)| + |L\,(g(x) - M)| \end{aligned}$$

We can tidy this up a little and write

$$|f(x)g(x) - LM| \leq |f(x) - L|\,|g(x)| + |L|\,|g(x) - M|\,.$$

Now it's time to play the game. You pick your positive number ε and then I get to work. I concentrate on an interval around $x = a$ so small that $|f(x) - L| < \varepsilon$ and $|g(x) - M| < \varepsilon$. In fact, if you pick $\varepsilon > 1$ (a pretty feeble move, if you ask me–you want ε to be small!) then I'm even going to insist that $|g(x) - M| < 1$ in that case. So we know in either case that $|g(x) - M| < 1$, which means that $M - 1 < g(x) < M + 1$ on my interval. In particular, we can see that

$|g(x)| \leq |M| + 1$. The whole point is that we have some nice inequalities on my interval:

$$|f(x) - L| < \varepsilon, \quad |g(x)| \leq |M| + 1, \quad \text{and} \quad |g(x) - M| < \varepsilon.$$

We can insert these into the inequality for $|f(x)g(x) - LM|$ above

$$\begin{aligned} |f(x)g(x) - LM| &\leq |f(x) - L|\,|g(x)| + |L|\,|g(x) - M| \\ &< \varepsilon(|M| + 1) + \varepsilon\,|L| = \varepsilon(|M| + |L| + 1) \end{aligned}$$

for x close enough to a. That's almost what I want! I was supposed to get ε on the right-hand side, but I got an extra factor of $(|M| + |L| + 1)$. This is no problem–you just have to allow me to make my move again, but this time I'll make sure that $|f(x) - L|$ is no more than $\varepsilon/(|M| + |L| + 1)$ and similarly for $|g(x) - M|$. Then when I replay all the steps, ε will be replaced by $\varepsilon/(|M| + |L| + 1)$, and at the very last step, the factor $(|M| + |L| + 1)$ will cancel out and we'll just get our ε! So we have proved the result.

By the way, it's worth noting a special case of the above. If c is constant, then

$$\lim_{x \to a} cf(x) = c \lim_{x \to a} f(x). \tag{4.14}$$

This is easy to see by setting $g(x) = c$ in our main formula above; I leave the details to you.

4.5.3 Quotients of limits

Here we want to show that if

$$\lim_{x \to a} f(x) = L \quad \text{and} \quad \lim_{x \to a} g(x) = M.$$

then we have

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{L}{M}. \tag{4.15}$$

So the limit of the quotient is the quotient of the limits. For this to work, we'd better have $M \neq 0$ or else we'll be dividing by 0. Another way of writing the above equation is

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \frac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)}$$

provided that both limits exist and are finite, and that the g-limit is nonzero.

Here's how the proof goes. We want $f(x)/g(x)$ to be close to LM, so we consider the difference. Then we'll need to take a common denominator, leaving us with

$$\frac{f(x)}{g(x)} - \frac{L}{M} = \frac{Mf(x) - Lg(x)}{Mg(x)}$$

Now we do a trick similar to the one we used in for products of limits: we'll subtract and add LM to the numerator, then factor. This gives us

$$\begin{aligned}\frac{f(x)}{g(x)} - \frac{L}{M} &= \frac{Mf(x) - LM + LM - Lg(x)}{Mg(x)} \\ &= \frac{M\left(f(x) - L\right)}{Mg(x)} + \frac{L\left(M - g(x)\right)}{Mg(x)} \\ &= \frac{f(x) - L}{g(x)} - \frac{L\left(g(x) - M\right)}{Mg(x)}.\end{aligned}$$

If we take absolute values and then use the triangle inequality in the form $|a - b| \leq |a| + |b|$, we get

$$\begin{aligned}\left|\frac{f(x)}{g(x)} - \frac{L}{M}\right| &= \left|\frac{f(x) - L}{g(x)} - \frac{L\left(g(x) - M\right)}{Mg(x)}\right| \\ &\leq \left|\frac{f(x) - L}{g(x)}\right| + \left|\frac{L\left(g(x) - M\right)}{Mg(x)}\right|.\end{aligned}$$

So you make your move by picking $\varepsilon > 0$, and then I narrow the window of interest around $x = a$ so that $|f(x) - L| < \varepsilon$ and $|g(x) - M| < \varepsilon$ in the little window. Now I need to be even trickier, though. You see, I know that $M - \varepsilon < g(x) < M + \varepsilon$ which means that $|g(x)| > |M| - \varepsilon$. All's well if this right-hand quantity $|M| - \varepsilon$ is positive, but if it's negative, it tells us nothing since we already knew that $|g(x)|$ can't be negative. So if your ε is small enough, then I don't worry, but if it's a little bigger, I need to narrow my window more so that $|g(x)| > |M|/2$ on the window. So altogether we have three inequalities which are true on the little interval:

$$|f(x) - L| < \varepsilon, \qquad |g(x)| > |M|/2, \quad \text{and} \quad |g(x) - M| < \varepsilon.$$

This middle inequality can be inverted to read

$$\frac{1}{|g(x)|} < \frac{2}{|M|}$$

Putting everything together, we have

$$\begin{aligned}\left|\frac{f(x)}{g(x)} - \frac{L}{M}\right| &< \frac{|f(x) - L|}{|g(x)|} + \frac{|L|\,|g(x) - M|}{|M|\,|g(x)|} \\ &< \varepsilon\frac{2}{|M|} + \varepsilon\frac{|L|}{|M|}\frac{2}{|M|} \\ &= \varepsilon\left(\frac{2}{|M|} + \frac{|L|}{|M|}\frac{2}{|M|}\right)\end{aligned}$$

Not quite what we wanted|we have an extra factor of $\left(\frac{2}{|M|} + \frac{|L|}{|M|}\frac{2}{|M|}\right)$, but we know how to handle this–I just make my move again, but instead of your ε, I use ε divided by this extra factor.

4.5.4 The limit of a composite function

In the future we will need another important fact which concerns limit of a composite functions. Precisely, let f and g are functions such that

$$\lim_{x\to a} g(x) = L \quad \text{and} \quad \lim_{x\to L} f(x) = f(L).$$

We will prove, that

$$\lim_{x\to a} f\left(g(x)\right) = f\left(\lim_{x\to a} g(x)\right) = f\left(L\right) \tag{4.16}$$

To do that, for a given $\varepsilon > 0$ we must find $\delta > 0$ such that

$$|f(g(x)) - f(L)| < \varepsilon \quad \text{whenever} \quad 0 < |x - a| < \delta.$$

Because the limit of $f(x)$ as $x \to L$ is $f(L)$ we know there exists $\delta_1 > 0$ such that

$$|f(u) - f(L)| < \varepsilon \quad \text{whenever} \quad 0 < |u - L| < \delta_1.$$

Moreover, because the limit of $g(x)$ as $x \to a$ is L you know there exists $\delta > 0$ such that

$$|g(x) - L| < \delta_1 \quad \text{whenever} \quad 0 < |x - a| < \delta.$$

Finally, letting $u = g(x)$, we have

$$|f(g(x)) - f(L)| < \varepsilon \quad \text{whenever} \quad 0 < |x - a| < \delta.$$

Remark 4.19 *Although limit properties are stated above for two functions only, the results hold for any finite number of functions (because of the mathematical induction). Moreover, the various properties can be used in combination to reformulate expressions involving limits.*

Example 4.20

$$\begin{aligned}
\lim_{x\to a}\left(f(x) - g(x) + 3h(x)\right) &= \lim_{x\to a} f(x) - \lim_{x\to a} g(x) + 3\lim_{x\to a} h(x),\\
\lim_{x\to a}\left(f(x)g(x)h(x)\right) &= \lim_{x\to a} f(x)\lim_{x\to a} g(x)\lim_{x\to a} h(x),\\
\lim_{x\to a}\left(f(x)\right)^4 &= \left(\lim_{x\to a}\left(f(x)\right)\right)^4, \qquad (4.17)\\
\lim_{x\to a} x^n &= \left(\lim_{x\to a} x\right)^n = a^n.
\end{aligned}$$

□

Example 4.21 *Find the limit*

$$\lim_{x \to 3} \frac{x^2 - 9}{x - 3}.$$

Solution: We cannot plug in the number 3, because that would yield the meaningless expression $\frac{0}{0}$. Suppose, that $x \neq 3$.Then

$$\frac{x^2 - 9}{x - 3} = \frac{(x+3)(x-3)}{(x-3)} = x + 3 \quad \text{as long as } x \neq 3.$$

Now take a closer look at this: here we have two functions. One of them is the original $\frac{x^2-9}{x-3}$, and the other is $x + 3$. These two functions agree everywhere (except $x = 3$) and they have the same limit, as we approach the point $x = 3$, because it is never important in limits what happens at the point we approach. Now we can easily find the solution $\lim_{x \to 3} \frac{x^2-9}{x-3} = \lim_{x \to 3} x + 3 = 6$. □

On the basis of that example we can formulate the following theorem:

Theorem 4.22 *Let a be a real number and let $f(x) = g(x)$ for all $x \neq a$ in an open interval containing a. If the limit $g(x)$ of as x approaches a exists, then the limit of $f(x)$ also exists and*

$$\lim_{x \to a} f(x) = \lim_{x \to a} g(x).$$

Proof. Let L be the limit of $g(x)$ as $x \to a$.Then for each $\varepsilon > 0$ there exists a $\delta > 0$ such that $f(x) = g(x)$ in the open intervals $(a - \delta, a)$ and $(a, a + \delta)$, and

$$|g(x) - L| < \varepsilon \quad \text{whenever} \quad 0 < |x - a| < \delta.$$

Because $f(x) = g(x)$ for all x for all x in the open interval other than $x = a$, it follows that

$$|f(x) - L| < \varepsilon \quad \text{whenever} \quad 0 < |x - a| < \delta.$$

So, the limit of $f(x)$ as $x \to a$ is also L. ■

4.5.5 Sandwich principle

The *sandwich principle* , also known as the *squeeze principle,* says that if a function f is sandwiched between two functions g and h that converge to the same limit L as $x \to \alpha$, then f also converges to L as $x \to \alpha$.

Here's a more precise statement of the principle. Suppose that for all x near a, we have $g(x) \leq f(x) \leq h(x)$. That is, $f(x)$ is sandwiched (or squeezed) between $g(x)$ and $h(x)$. Also, let's suppose that $\lim_{x \to a} g(x) = L$ and $\lim_{x \to a} h(x) =$

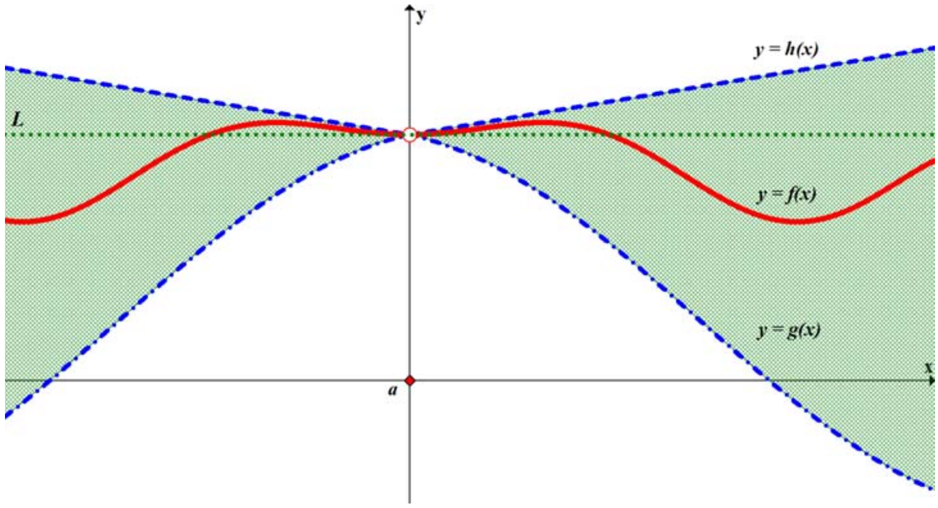


Fig. 4.15. Sandwich principle; the values of $f(x)$ are forced to tend to L in the limit as $x \to a$.

L. Then we can conclude that $\lim_{x \to a} f(x) = L$.; that is, all three functions have the same limit as $x \to a$. We use this principle, because sometimes it is easier to designate the limits (when $x \to a$) of the function $h(x)$ and $g(x)$ rather than $f(x)$. As usual, the picture tells the story (see Figure 4.15):

The function f, shown as a solid curve in the picture, is really squeezed between the other functions g and h in the vicinity a; the values of $f(x)$ are forced to tend to L in the limit as $x \to a$. (See next subsection for a proof of the sandwich principle.)

Example 4.23 *One of the most important trigonometry limits is the fact that*

$$\lim_{\theta \to 0} \frac{\sin(\theta)}{\theta} = 1. \tag{4.18}$$

The graph of $f(\theta) = \frac{\sin(\theta)}{\theta}$ *is shown on Figure 4.16, and as expected, this graph is not defined at* $a = 0$. *Two squeezing functions are:* $h(\theta) = 1$, *and* $g(\theta) = \cos^2(\theta)$. *Of course*

$$\lim_{\theta \to 0} 1 = 1$$

and

$$\lim_{\theta \to 0} \cos^2(\theta) = 1.$$

We now prove that $\cos^2(\theta) < \frac{\sin(\theta)}{\theta} < 1$ *for* $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$ *and* $\theta \neq 0$. *For this, consider the graph 4.17. You can move the blue point on the unit circle to*

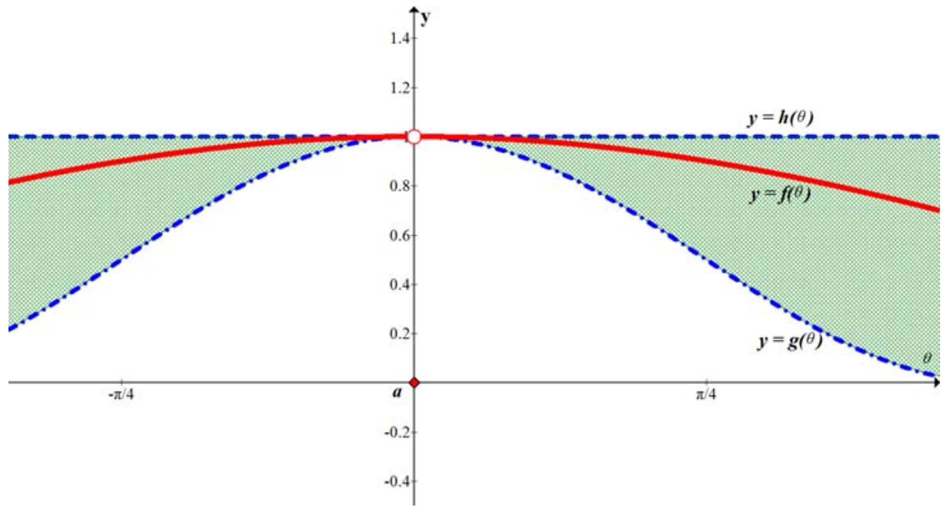


Fig. 4.16. Function $f(\theta) = \frac{\sin(\theta)}{\theta}$ is squished between $h(\theta) = 1$ and $g(\theta) = \cos^2(\theta)$ in the neighborhood of $\theta = 0$.

change the value of θ*. Observe that the* x*-value of the blue point* P *is* $\cos(\theta)$ *and the* y*-value of the blue point is* $\sin(\theta)$*. The area of the "small sector" (see Figure 4.18) is equal to* $\frac{1}{2}\theta r^2$ *where* r *is the radius of the circle The area of the triangle with vertices* 0, 1, *and* P *is equal to* $\frac{1}{2}\sin(\theta)$, *whereas the area of the "large sector" (see Figure 4.17) is equal to* $\theta/2$*. By comparing areas, it's clear that for* $0 < \theta < \pi/2$,

$$\frac{1}{2}\theta\cos^2(\theta) < \frac{1}{2}\sin(\theta) < \frac{1}{2}\theta.$$

Multiply through by 2,

$$\theta\cos^2(\theta) < \sin(\theta) < \theta.$$

Divide through by (positive) θ,

$$\cos^2(\theta) < \frac{\sin(\theta)}{\theta} < 1.$$

Finally, we consider the case where, $-\pi/2 < \theta < 0$*. We see that for positive* θ

$$\cos(-\theta) = \cos(\theta), \quad \textit{so} \quad \cos^2(-\theta) = \cos^2(\theta)$$

and

$$\frac{\sin(-\theta)}{-\theta} = \frac{-\sin(\theta)}{-\theta} = \frac{\sin(\theta)}{\theta}$$

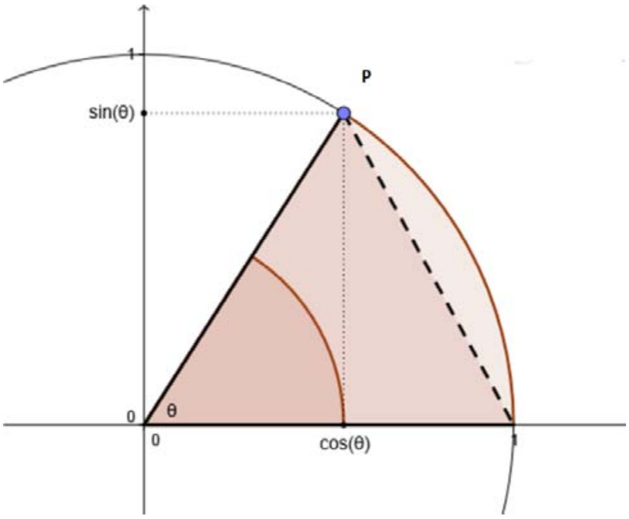


Fig. 4.17. The shaded area is the “large sector”.

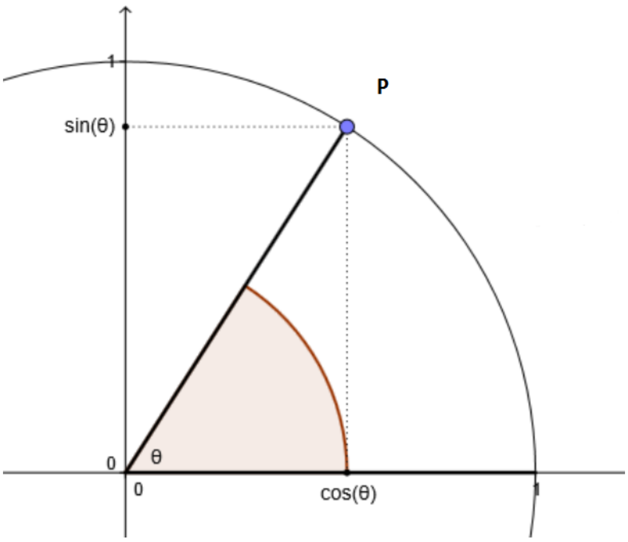


Fig. 4.18. The “small sector”.

so, the previous string of inequalities still holds. Thus, for $-\pi/2 < \theta < \pi/2$

$$\cos^2(\theta) < \frac{\sin(\theta)}{\theta} < 1$$

as needed. □

Example 4.24 *Using the limit (4.18), we can prove the related, important trigonometry limit*

$$\lim_{x \to 0} \frac{1 - \cos x}{x} = 0.$$

Notice, that you cannot just plug in the number 0 *because that would yield the meaningless expression* $\frac{0}{0}$. *Instead we can proceed as follows:*

$$\begin{aligned} & \lim_{x \to 0} \left(\frac{1 - \cos x}{x} \right) \left(\frac{1 + \cos x}{1 + \cos x} \right) \\ = & \lim_{x \to 0} \frac{1 - \cos^2 x}{x\,(1 + \cos x)} = \lim_{x \to 0} \frac{\sin^2 x}{x\,(1 + \cos x)} \\ = & \lim_{x \to 0} \left(\frac{\sin x}{x} \right) \left(\frac{\sin x}{(1 + \cos x)} \right) = 1 \cdot 0 = 0. \end{aligned}$$

□

Remark 4.25 *The* sinc *function (see Figure 4.19)*

$$f(x) = \begin{cases} \frac{\sin x}{x} & \textit{if} \quad x \neq 0 \\ 1 & \textit{if} \quad x = 0. \end{cases}$$

is an important function and appears in many applications like in the study of waves or signal processing (it is used in low pass filters). The name sinc *comes from its original latin name* sinus cardinalis.

4.5.6 The sandwich principle–proof

Now it's time to prove the sandwich principle. We start with functions $f, g,$ and h, such that $g(x) \leq f(x) \leq h(x)$ for all x close enough to a. We also know that

$$\lim_{x \to a} g(x) = L \quad \text{and} \quad \lim_{x \to a} h(x) = L.$$

Intuitively, f is squeezed between g and h more and more, so that in the limit as $x \to a$, we should have $f(x) \to L$ as well. That is, we need to prove that

$$\lim_{x \to a} f(x) = L$$

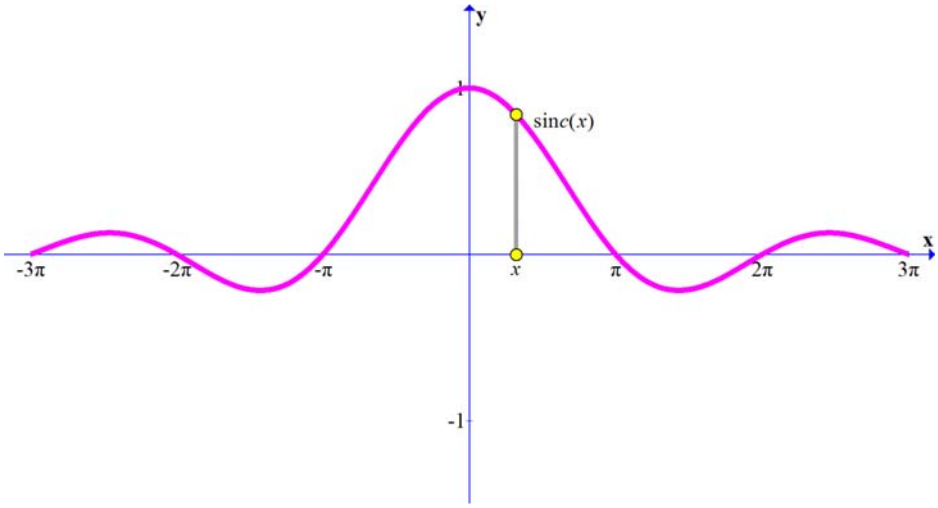


Fig. 4.19. The graph of the sinc function.

Well, you start off by picking your positive number ε, and then I can focus on an interval centered at a small enough so that $|g(x) - L| < \varepsilon$ and $|h(x) - L| < \varepsilon$ on the interval. I'm also going to need the inequality $g(x) \leq f(x) \leq h(x)$ to be true on the interval; since that inequality might only be true when x is very near to a, I may have to shrink my original interval.

Anyway, we know that $|h(x) - L| < \varepsilon$ when x is close enough to a; the inequality can be rewritten as

$$L - \varepsilon < h(x) < L + \varepsilon.$$

Actually, we only need the right-hand inequality, $h(x) < L + \varepsilon$; you see, on my little interval, we know that $f(x) \leq h(x)$, so we also have

$$f(x) \leq h(x) < L + \varepsilon.$$

Similarly, we know that

$$L - \varepsilon < g(x) < L + \varepsilon$$

when x is close enough to a; this time we throw away the right-hand inequality and use $g(x) \leq f(x)$ to get

$$L - \varepsilon < g(x) \leq f(x).$$

Putting all this together, we have shown that when x is close to a, we have

$$L - \varepsilon < f(x) < L + \varepsilon$$

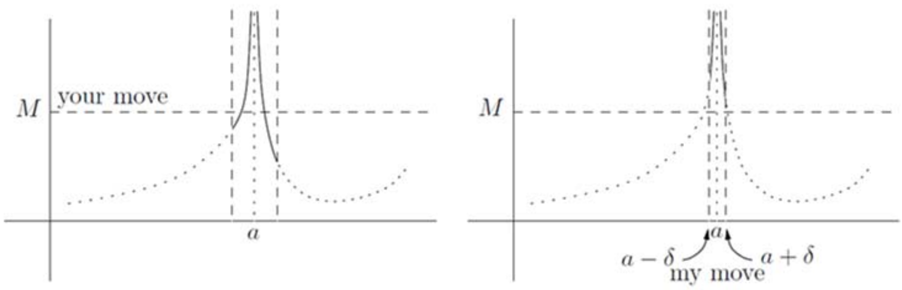


Fig. 4.20. A move you might make and then a possible response for me.

or simply

$$|f(x) - L| < \varepsilon.$$

That's what we need to show our limit–we've proved the sandwich principle!

4.6 Other varieties of limits

Now let's quickly look at the definitions of some other types of limits: *infinite limits*, and *right-hand limits*, and limits at $\pm\infty$.

4.6.1 Infinite limits

Our game isn't going to work if we want to use it to define a limit like this:

$$\lim_{x \to a} f(x) = \infty.$$

When you try to draw your two lines close to the limit, you'll be completely stuck, since the limit is supposed to be ∞ instead of some finite value L. So we have to modify the rules a little bit. My move won't change much, but yours will. Instead of picking a little number $\varepsilon > 0$ and then drawing two horizontal lines (at height $L - \varepsilon$ and $L + \varepsilon$), this time you'll pick a large number M and only draw in the line at height M. I still make my move by throwing away most of the function, except for a small bit around $x = a$; this time, though, I have to make sure that what's left is always above your line. For example, the following pictures show a move you might make and then a possible response for me (see Figure 4.20):

Now here's what happens if you make another move but with a larger value of M (Figure 4.21):

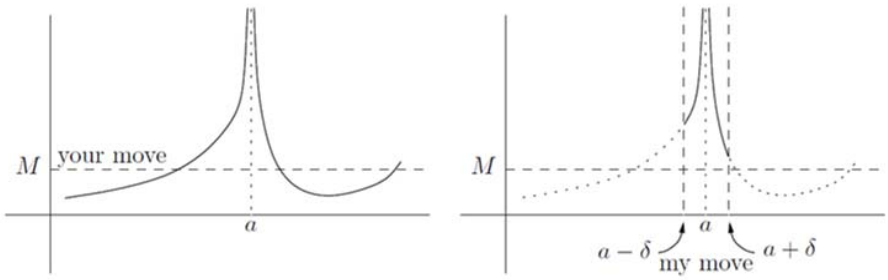


Fig. 4.21. Another move with a larger valueof M.

So the idea is that this time you raise your bar higher and higher; if I can always make a move in response, then the limit is indeed ∞. In symbols, I need to be able to ensure that $f(x) > M$ whenever x is close enough to a, no matter how big M is. So, we write

"$\lim_{x\to a} f(x) = \infty$" to indicate, that the limit fails to exist because $f(x)$ is increasing without bound, or (more precisely) for any choice of $M > 0$ there exists $\delta > 0$ such that $f(x) > M$ for all x satisfying $0 < \|x - a\| < \delta$.

It's very similar to the situation when the limit is some finite number L, except that the inequality $|f(x) - L| < \varepsilon$ is replaced by $f(x) > M$.

Example 4.26 *Suppose that we want to show that*

$$\lim_{x\to 0} \frac{1}{x^2} = \infty.$$

You start off by picking your number $M > 0$; then I have to make sure that $f(x) > M$ when x is close enough to 0. Well, suppose that I throw everything away except for x satisfying $|x| < 1/\sqrt{M}$. For such an x, we have $x^2 < 1/M$, so $1/x^2 > M$ (note that we have assumed that $x \neq 0$). That means that $f(x) > M$ in my interval, which means my move is valid. So for any M you pick, I can make a valid move, and we have proved that the limit is indeed ∞.
□

How about $-\infty$? Everything is just reversed. You still pick a large positive number M, but this time I need to make my move so that the function is always below the horizontal line of height $-M$. So here's what the definition looks like:

"$\lim_{x\to a} f(x) = -\infty$" means that for any choice of $M > 0$ you make, I can pick $\delta > 0$ such that: $f(x) < -M$ for all x satisfying $0 < \|x - a\| < \delta$.

4.7 Limits and continuous functions

We return now to the study of continuous functions, making use of these ideas about limits.

Theorem 4.27 (Positivity theorem for continuous functions) *Suppose, that $f(x)$ is continuous at x_0 and $f(x_0) > 0$ then there exists $\delta > 0$ such that $f(x) > 0$ whenever $|x - x_0| < \delta$.*

Proof. Choose an ε so that $0 < \varepsilon < f(x_0)$. Since f is continuous at x_0 there exists $\delta > 0$ such that

$$|f(x) - f(x_0)| < \varepsilon \quad \text{whenever} \quad |x - x_0| < \delta.$$

This when combined with the previous inequality, says

$$0 < f(x_0) - \varepsilon < f(x) < f(x_0) + \varepsilon \quad \text{for} \quad |x - x_0| < \delta.$$

The two left-hand inequalities prove the theorem. ■

As another example of the use of limits to establish results about continuity, the above results about limits of sums, products, and quotients imply almost immediately the corresponding results for continuous functions:

Theorem 4.28 (Algebraic operations on continuous functions) *If f and g are continuous at x_0, and a, b are constants, then the following functions are also continuous at x_0:*

$$af + bg, \qquad f \cdot g, \qquad \frac{f}{g} \quad (\textit{if} \;\; g(x_0) \neq 0).$$

Proof. Using the limit form of continuity (see Theorem 4.12), the three statements follow immediately from the corresponding statements from the Section 4.5

For the quotient statement, we must also verify that there exists $\delta > 0$ such that $g(x) \neq 0$ for $|x - x_0| < \delta$. But this follows from the Positivity Theorem 4.27. ■

Theorem 4.28 extends immediately to intervals, because continuity is a local property. *On an interval I, the sum, product, and quotient (where defined) of continuous functions is again continuous.* Continuity properties are stated above for two functions only, but the results hold for any finite number of functions (because of the mathematical induction).

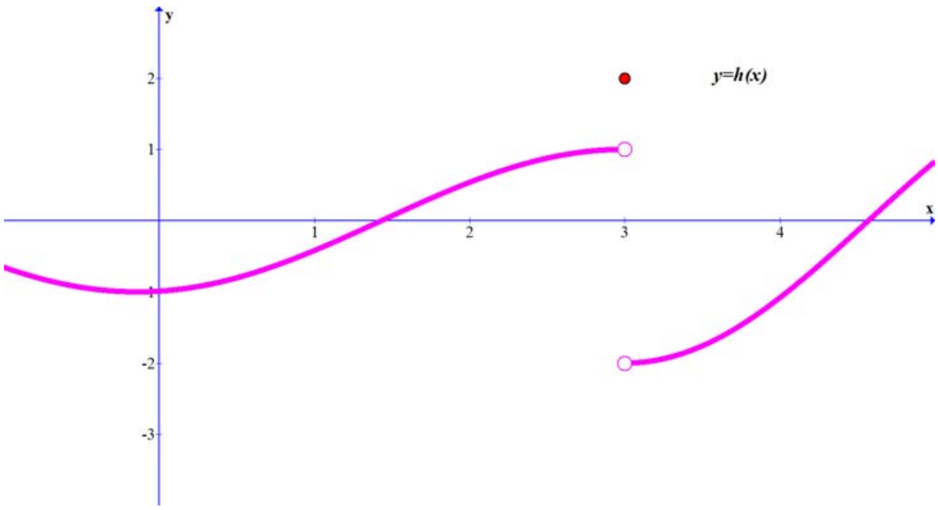


Fig. 4.22. At $x = 3$ one-sided limits exist, but are not equal.

4.8 One-sided limits and continuity on a closed interval

4.8.1 Left-hand and right-hand limits

To define a *right-hand limit* (denoted by $\lim_{x\to a^+} f(x)$), we play the same game, except this time before we start, we already throw away everything to the left of $x = a$. The effect is that instead of choosing an interval like $(a-\delta; a+\delta)$ when I make my move, now I just have to worry about $(a; a+\delta)$. Nothing to the left of a is relevant.

Similarly, for a *left-hand limit* (denoted by $\lim_{x\to a^-} f(x)$), only the values of x to the left of a matter. This means that my intervals look like $(a-\delta; a)$; I have thrown away everything to the right of $x = a$. This all means that you can take any of the above definitions in boxes and change the inequality $0 < |x-a| < \delta$ to $0 < a-x < \delta$ to get the right-hand limit. To get the left-hand limit, you change the inequality to $0 < a-x < \delta$ instead.

Example 4.29 *We've seen that limits describe the behavior of a function near a certain point. Think about how you would describe the behavior of* $h(x)$ *from Figure 4.22 near* $x = 3$*:Of course, the fact that* $h(3) = 2$ *is irrelevant as far as the limiting behavior is concerned. Now, what happens when you approach* $x = 3$ *from the left? Imagine that you're the hiker, climbing up and down the hill. The value of* $h(x)$ *tells you how high up you are when your horizontal position is at* x*. So, if you walk rightward from the left of the picture, then when your horizontal position is close to* 3*, your height is close to* 1*. Sure,*

there's a sheer drop when you get to $x = 3$, but we don't care about this for the moment. Everything to the right of $x = 3$, including $x = 3$ itself, is irrelevant. So we've just seen that the left-hand limit of $h(x)$ at $x = 3$ is equal to 1.
On the other hand, if you are walking leftward from the right-hand side of the picture, your height becomes close to -2 as your horizontal position gets close to $x = 3$. This means that the right-hand limit of $h(x)$ at $x = 3$ is equal to -2. Now everything to the left of $x = 3$ (including $x = 3$ itself) is irrelevant!
We can summarize our findings from above by writing

$$\lim_{x \to 3^-} h(x) = 1 \qquad \textit{and} \qquad \lim_{x \to 3^+} h(x) = 1,$$

whereas $\lim_{x \to 3} h(x)$ obviously does not exist.

As we have seen, limits don't always exist. But here's something important: the regular two-sided limit at $x = a$ exists **exactly when** both left-hand and right-hand limits at $x = a$ exist and **are equal** to each other! In that case, all three limits–two-sided, left-hand, and right-hand right-hand–are the same. In math-speak,

$$\lim_{x \to a^-} f(x) = L \qquad \textbf{and} \qquad \lim_{x \to a^+} f(x) = L$$

is the same thing as

$$\lim_{x \to a} f(x) = L.$$

If the left-hand and right-hand limits are not equal, as in the case of our function h from above, then the two-sided limit does not exist. We'd just write

$$\lim_{x \to a} h(x) \quad \text{does not exist.}$$

When the limit does not exist

We just saw that a two-sided limit doesn't exist when the corresponding left-hand and right-hand limits are different. Here's an even more dramatic example of this. Consider the graph of $f(x) = 1/x$ (Figure 4.23)What is $\lim_{x \to 0} f(x)$? It may be a bit much to expect the two-sided limit to exist here, so let's first try to find the right-hand limit, $\lim_{x \to 0^+} f(x)$. Looking at the graph, it seems as though $f(x)$ is very large when x is positive and close to 0. It doesn't really get close to any number in particular as x slides down to 0 from the right; it just gets larger and larger. How large? Larger than anything you can imagine! We say that the limit is infinity, and write

$$\lim_{x \to 0^+} \frac{1}{x} = \infty.$$

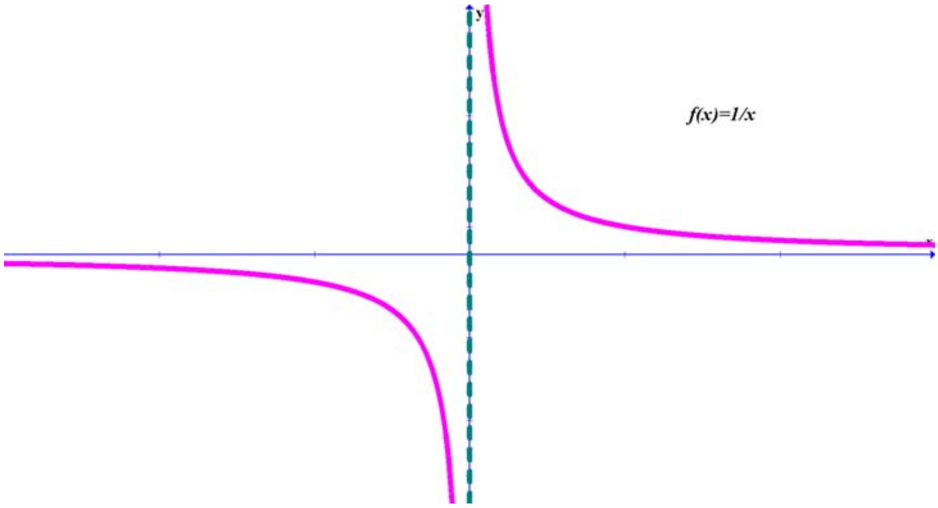


Fig. 4.23. Function $y = 1/x$.

Similarly, the left-hand limit here is $-\infty$, since $f(x)$ gets arbitrarily more and more negative as x slides upward to 0. That is,

$$\lim_{x\to 0^-} \frac{1}{x} = -\infty$$

The two-sided limit certainly doesn't exist, since the left-hand and right-hand limits are different. On the other hand, consider the function g defined by $g(x) = 1/x^2$. Both the left-hand and right-hand limits at $x = 0$ are ∞, so you can say that $\lim_{x\to 0} \frac{1}{x^2} = \infty$ as well. By the way, we now have a formal definition of the term *vertical asymptote*

Definition 4.30 *If $f(x)$ approaches infinity (or negative infinity) as x approaches c from the right or the left, then the line $x = c$ is a* vertical asymptote *of the graph of f.*

Vertical (dashed) line $x = 0$ on Figure (4.23) is a vertical asymptote of the graph of $f(x) = 1/x$.

Remark 4.31 *If the graph of a function f has a vertical asymptote at $x = c$ then f is not continuous at c.*

Remark 4.32 *It possible that even a left-hand or right-hand limit fails to exist. For example, let's meet the funky function y defined by $y(x) = sin(1/x)$ (see Figure 4.4). The function doesn't tend toward any one number as x goes to 0 from the right (or from the left). There is no vertical asymptote, of course.*

Theorem 4.33 *Let f and g be continuous on an open interval containing c. If $f(c) \neq 0$, $g(c) = 0$ and there exists an open interval containing c such that $g(x) \neq 0$ for all $x \neq c$ in the interval, then the graph of the function given by*

$$h(x) = \frac{f(x)}{g(x)}$$

has a vertical asymptote at $x = c$.

Proof. Consider the case for which $f(c) > 0$, and there exists $b > c$ such that $c < x < b$ implies $g(x) > 0$. Then for $M > 0$, choose δ_1 such that

$$0 < x - c < \delta_1, \quad \text{implies that} \quad \frac{f(c)}{2} < f(x) < \frac{3}{2} f(c),$$

and δ_2 such that

$$0 < x - c < \delta_2, \quad \text{implies that} \quad 0 < g(x) < \frac{f(c)}{2M}.$$

Now let be δ the smaller of δ_1 and δ_2. Then it follows that

$$0 < x - c < \delta, \quad \text{implies that} \quad \frac{f(x)}{g(x)} > \frac{f(c)}{2}\left(\frac{2M}{f(c)}\right) = M.$$

So, it follows that

$$\lim_{x \to c^+} \frac{f(x)}{g(x)} = \infty$$

and the line $x = c$ is a vertical asymptote of the graph of h. The proof in other cases is similar. ■

Theorem 4.33 requires that the value of the numerator at $x = c$ be nonzero. If both the numerator and the denominator are 0 at $x = c$ you obtain the *indeterminate form* 0/0 **??** and you cannot determine the limit behavior at $x = c$ without further investigation, as illustrated in Example 4.34.

$\frac{2}{x+1} + 1 = (x+3)(x-1) : (x-1)(x+3) = x^2 + 2x - 3$

Example 4.34 *Determine all vertical asymptotes of the graph of*

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

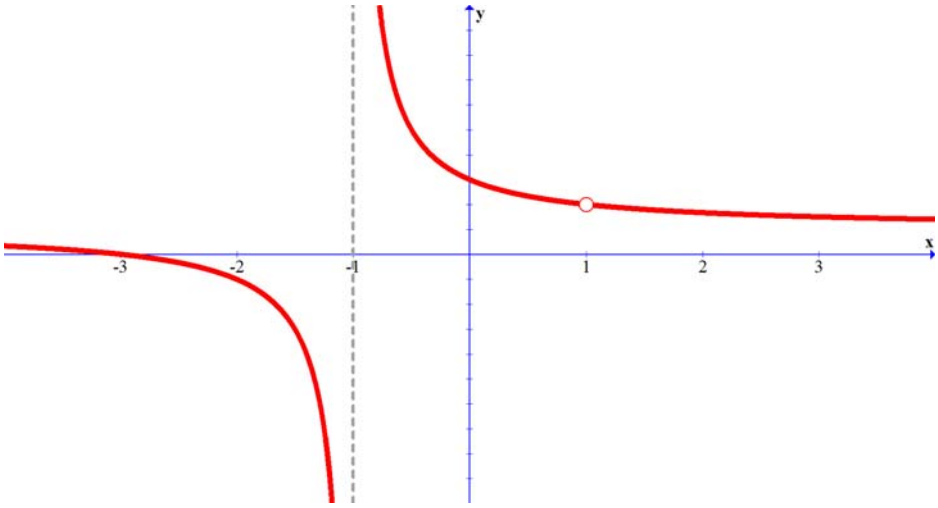


Fig. 4.24. Function $f(x)$ from example 4.34 has vertical asymptote at $x = -1$. It is undefined when $x = 1$.

Solution: Begin by factoring and simplifying the expression, as shown.

$$\begin{aligned} f(x) &= \frac{x^2+2x-3}{x^2-1} \\ &= \frac{(x+3)(x-1)}{(x+1)(x-1)} \\ &= \frac{(x+3)}{(x+1)} = 1 + \frac{2}{x+1} \quad \text{for} \quad x \neq 1. \end{aligned}$$

At all x-values other than 1 the graph of $f(x)$ coincides with the graph of $(x+3)/(x+1)$. So, you can apply Theorem 4.33 to to conclude that there is a vertical asymptote at $x = -1$ as shown in Figure 4.24. Note that $x = 1$ is *not* a vertical asymptote. □

4.8.2 Continuity on a Closed Interval

The concept of a one-sided limit allows you to extend the definition of continuity to closed intervals. Basically, a function is continuous on a closed interval if it is continuous in the interior of the interval and exhibits one-sided continuity at the endpoints. This is stated formally as follows.

Definition 4.35 *A function* $f(x)$ *is* continuous on the closed interval $[a, b]$ *if it is continuous on the open interval* (a, b) *and*

$$\lim_{x \to a^+} = f(a) \qquad \textit{and} \qquad \lim_{x \to b^-} = f(b)$$

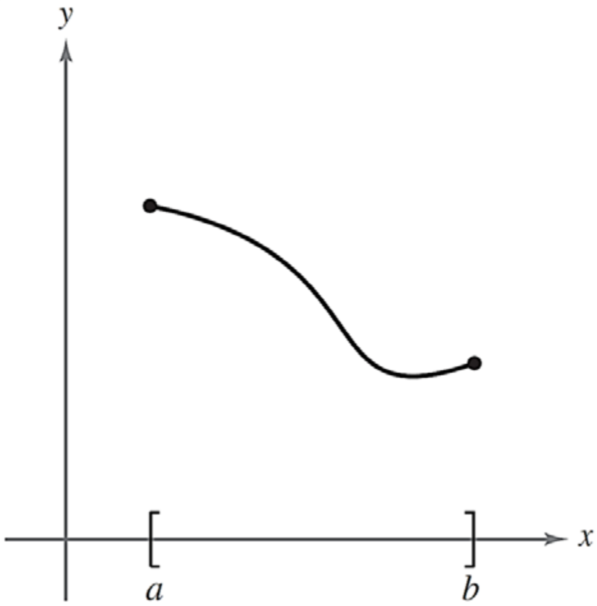


Fig. 4.25. Continuous function on a closed interval.

The function $f(x)$ is continuous from the right *at a and continuous from the left at b (see Figure 4.25).*

Example 4.36 *Show that for all integers $n \geq 2$, the function $f(x) = x^{\frac{1}{n}}$ is continuous on $[0, \infty)$.*

Solution: let $x, a \in (0, \infty)$. We use the identity[2]

$$p^n - q^n = (p-q)(p^{n-1} + p^{n-1}q + ... + pq^{n-2} + q^{n-1}).$$

Putting $p = x^{\frac{1}{n}}$ and $q = a^{\frac{1}{n}}$ gives

$$\begin{aligned} |x-a| &= \left|x^{\frac{1}{n}} - a^{\frac{1}{n}}\right| \left|x^{\frac{(n-1)}{n}} + x^{\frac{(n-2)}{n}} a^{\frac{1}{n}} + ... + a^{\frac{(n-1)}{n}}\right| \\ &> \left|x^{\frac{1}{n}} - a^{\frac{1}{n}}\right| a^{\frac{(n-1)}{n}} \end{aligned}$$

and so

$$\left|x^{\frac{1}{n}} - a^{\frac{1}{n}}\right| < \frac{|x-a|}{a^{\frac{(n-1)}{n}}}.$$

[2] It follows from the geometric sum formula.

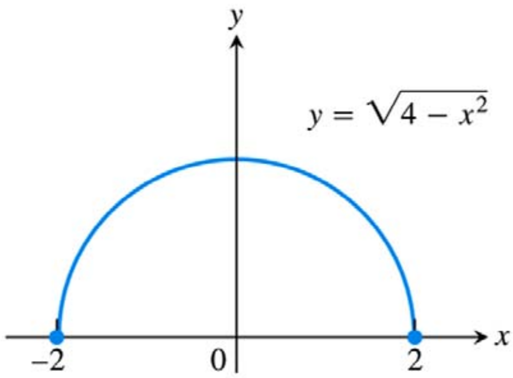


Fig. 4.26. The graph of the function of $f(x) = \sqrt{4 - x^2}$

It follows that we can arrange for $\left|x^{\frac{1}{n}} - a^{\frac{1}{n}}\right|$ to be less than any given $\varepsilon > 0$ by taking $|x - a|$ less than $\varepsilon a^{\frac{(n-1)}{n}}$. Thus the function is continuous at every a in $(0, \infty)$.

Continuity at 0 requires a separate argument, but is easily established, for we can make $x^{\frac{1}{n}}$ less than any given f by choosing x less than ε^n. □

Remark 4.37 *The result of this example is in fact a corollary of a general result on inverse functions, to be established later. If n is odd the natural domain of the function $f(x) = x^{\frac{1}{n}}$ is the whole of $\mathbb{R}$, and the function is continuous throughout its domain.*

Example 4.38 *Discuss the continuity of $f(x) = \sqrt{4 - x^2}$.*

Solution: The domain of f is the closed interval $[-2, 2]$. At all points in the open interval $(-2, 2)$ the continuity of follows from the Example 4.36. Moreover, because

$$\lim_{x \to -2^+} \sqrt{4 - x^2} = 0 = f(-2) \quad \text{(continuous from the right)}$$

and

$$\lim_{x \to 2^-} \sqrt{4 - x^2} = 0 = f(2) \quad \text{(continuous from the left)}$$

you can conclude that $f(x)$ is continuous on the closed interval $[-2, 2]$ as shown in Figure 4.26. □

4.8.3 Limits at ∞ and $-\infty$

There is one more type of limit that we need to investigate. We've concentrated on the behavior of a function near a point $x = a$. However sometimes it is important to understand how a function behaves when x gets really huge. Another way of saying this is that we are interested in the behavior of a function as its argument x goes to ∞. We'd like to write something like

$$\lim_{x \to \infty} f(x) = L$$

and mean that $f(x)$ gets really close, and stays close, to the value L when x is large.

The game has to change a little, of course, but we already know how. In fact we just have to adapt the methods from above. You'll

start by picking your little number $\varepsilon > 0$, establishing your tolerance interval $(L - \varepsilon, L + \varepsilon)$; then my move will be to throw away the function to the left of some vertical line $x = N$, so that all the function values to the right of the line lie in your tolerance interval. Then you pick a smaller ε, and I move the line rightward if I have to in order to lie within your new, smaller interval.

Here's what the first couple of moves for both of us might look like (see Figure 4.27) :

After your first move, my move ensures that all the function values to the right of the line $x = N$ lie in your tolerance interval. You respond by closing in the interval, but then I just move the line to the right until I can meet your new, more restrictive tolerance interval. Again, if I can always make a move in response to you, then the above limit is true.

More formally, my move consists of picking N such that $f(x)$ is in the interval $(L - \varepsilon, L + \varepsilon)$ whenever $x > N$ (so x is to the right of the vertical line $x = N$). Using absolute values, we can write this as follows:

"$\lim_{x \to \infty} f(x) = L$" means that for any choice of $\varepsilon > 0$ there exists N such that $\lvert f(x) - L \rvert < \varepsilon$ for all x satisfying $x > N$.

It's worth noting that any limit as $x \to \infty$ is necessarily a left-hand limit– there's nothing to the right of ∞! Anyway, there are still a couple of variations to look at. First, what does $\lim_{x \to \infty} f(x) = \infty$? You just have to adapt the previous definitions. In particular, you can take the above definition and change your move to picking $M > 0$, and now instead of requiring that $|f(x) - L| < \varepsilon$, this changes to $f(x) > M$. If instead you would like to show that $\lim_{x \to \infty} f(x) = -\infty$, you would change the inequality to $f(x) < -M$. Pretty straightforward. It's also pretty easy to define what

$$\lim_{x \to -\infty} f(x) = L, \qquad \lim_{x \to -\infty} f(x) = \infty. \qquad \lim_{x \to -\infty} f(x) = -\infty$$

mean. The only thing that changes from the respective case where $x \to \infty$ is that my vertical line will be at $x = -N$, and now the function values have to

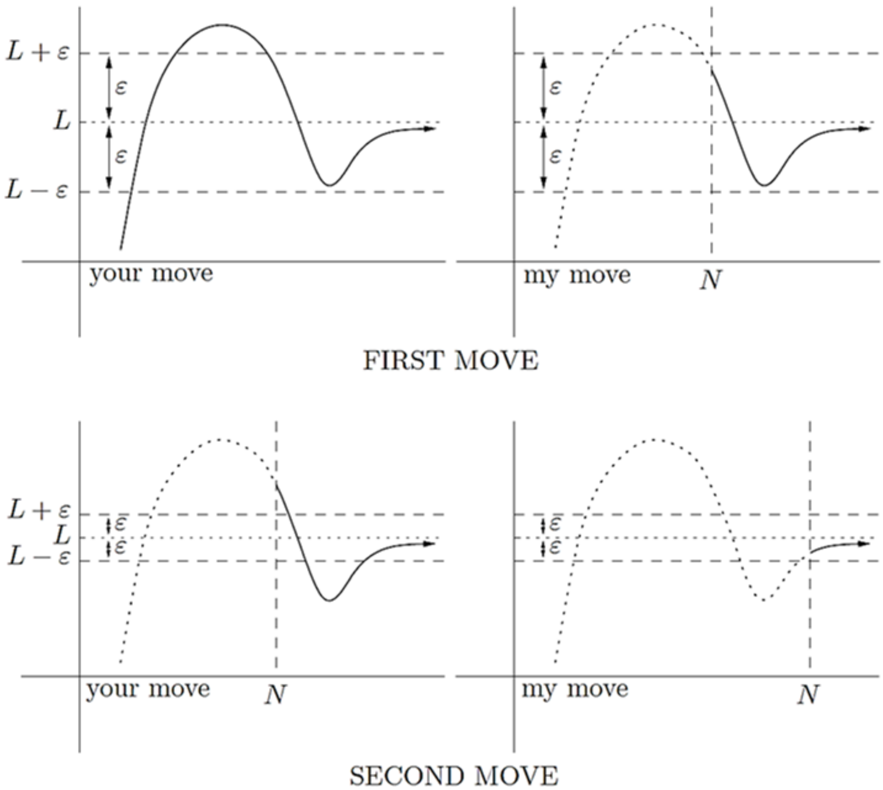


Fig. 4.27. First and second move of the game.

lie in your tolerance region to the left of the line instead of to the right. That is, you just change the inequality $x > N$ to $x < -N$ in all the definitions.

The important thing to realize is that writing "$\lim_{x\to\infty} f(x) = L$" indicates that the graph of f has a right-hand horizontal asymptote at $y = L$. There is a similar notion for when x heads toward $-\infty$: we write

$$\lim_{x\to-\infty} f(x) = M$$

which means that $f(x)$ gets extremely close, and stays close, to M when x gets more and more negative (or more precisely, $-x$ gets larger and larger). This of course corresponds to the graph of $y = f(x)$ having a left-hand horizontal asymptote. You can turn these definitions around if you like and say:

"f has a right-hand horizontal asymptote at $y = L$" means that $\lim_{x\to\infty} f(x) = L$.
"f has a left-hand horizontal asymptote at $y = M$" means that $\lim_{x\to-\infty} f(x) = M$.

Of course, something like $y = x^2$ doesn't have any horizontal asymptotes: the values of y just go up and up as x gets larger. In symbols, we can write this as $\lim_{x\to\infty} x^2 = \infty$.

It is obvious, that a given function $f(x)$ has $0, 1$ or maximally 2 horizontal asymptotes.

Example 4.39 *Looking at $f(x) = 1/x$ (see Figure 4.23), we observe that the x-axis is a right-hand horizontal asymptote of the curve because*

$$\lim_{x\to\infty} \frac{1}{x} = 0$$

and a left-hand horizontal asymptote because

$$\lim_{x\to-\infty} \frac{1}{x} = 0.$$

□

Example 4.40 *Find the horizontal asymptotes of the graph of*

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

Solution: We calculate the limits as $x \to \pm\infty$.
For $x \geq 0$:

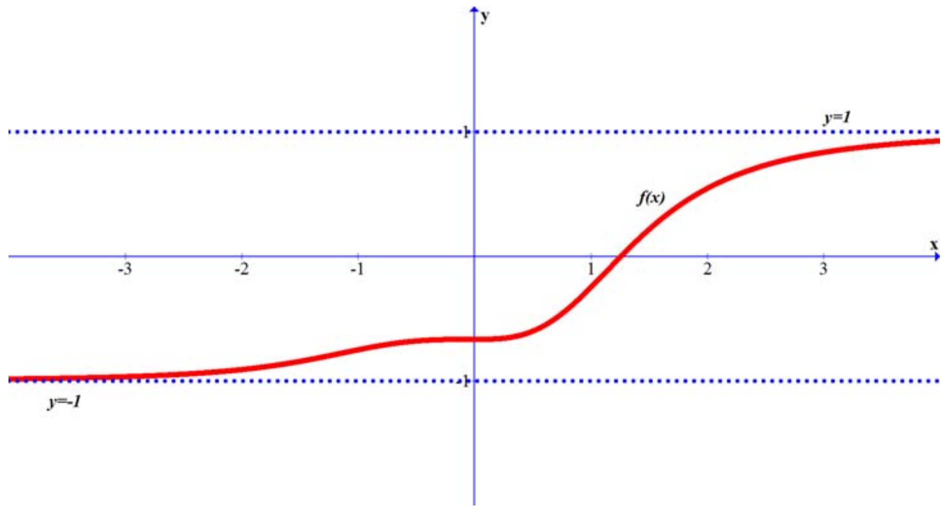


Fig. 4.28. Graph of the function $f(x) = \frac{x^3-2}{|x|^3+3}$ and its horizontal asymptotes.

$\lim_{x\to\infty} f(x) = \lim_{x\to\infty} \frac{x^3-2}{|x|^3+3} = \lim_{x\to\infty} \frac{x^3-2}{x^3+3} = \lim_{x\to\infty} \frac{1-\frac{2}{x^3}}{1+\frac{3}{x^3}} = 1.$

For $x \le 0$:

$\lim_{x\to-\infty} f(x) = \lim_{x\to-\infty} \frac{x^3-2}{|x|^3+3} = \lim_{x\to\infty} \frac{x^3-2}{-x^3+3} = \lim_{x\to\infty} \frac{1-\frac{2}{x^3}}{-1+\frac{3}{x^3}} = -1.$

The horizontal asymptotes are $y = -1$ and $y = 1$. The graph is displayed in Figure 4.28. □

Example 4.41 *Using the Sandwich Theorem, find the horizontal asymptote of the curve*

$$y(x) = 1 + \frac{\sin(x)}{x}.$$

Solution: We are interested in the behavior as $x \to \pm\infty$. Since

$$0 \le \left|\frac{\sin(x)}{x}\right| \le \left|\frac{1}{x}\right| \quad \text{for} \quad x \ne 0,$$

and $\lim_{x\to\pm\infty} \left|\frac{1}{x}\right| = 0$,we have $\lim_{x\to\pm\infty} \left|\frac{\sin(x)}{x}\right| = 0$ by the Sandwich Theorem. Hence,

$$\lim_{x\to\pm\infty} \left(1 + \frac{\sin(x)}{x}\right) = 1 + 0 = 1$$

and the line $y = 1$ is a left-hand and a righ-hand horizontal asymptote of the curve. This example illustrates that a curve may cross one of its horizontal asymptotes many times (see Figure 4.29).

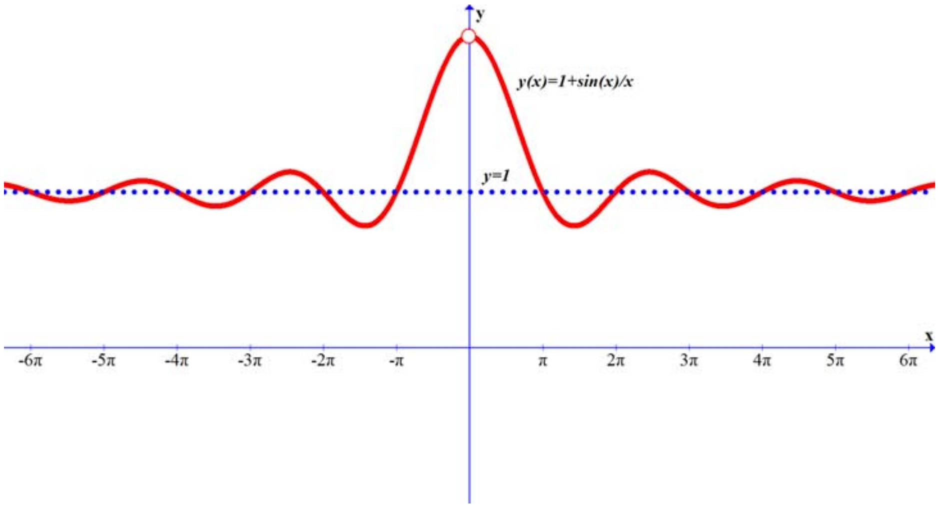


Fig. 4.29. A curve may cross one of its asymptotes infinitely often.

Example 4.42 *Find the horizontal asymptotes of the curve*

$$f(x) = \frac{x(1+x-|x|)}{\sqrt{x^2+1}}.$$

Solution: In this case only a righ-hand horizontal asymptote exists as

$$\lim_{x\to-\infty} f(x) = \infty, \qquad \lim_{x\to\infty} f(x) = 1$$

(see Figure 4.30.) □

Example 4.43 *Find the horizontal asymptotes of the graph of*

$$f(x) = x\cos(x).$$

Solution: As

$$\lim_{x\to-\infty} f(x) \quad \text{does not exist,}$$
$$\lim_{x\to\infty} f(x) \quad \text{does not exist}$$

there are no horizontal asymptotes. □

4.8.4 Examples of discontinuities

To understand continuity better, let's consider some ways that a function can fail to be continuous. Keep in mind that continuity at a point requires more

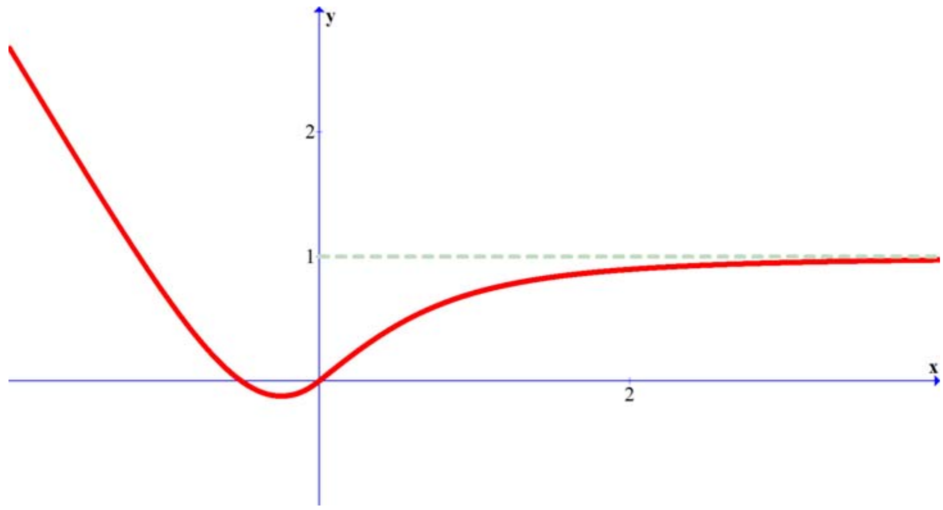


Fig. 4.30. Only a righ-hand horizontal asymptote exists here.

than just the existence of a limit. For $f(x)$ to be continuous at $x = c$, three conditions must hold

$$\textbf{1}.\ \lim_{x\to c} f(x)\ \text{ exists} \qquad \textbf{2}.\ f(c)\ \text{ exists} \qquad \textbf{3}.\ \text{They are equal.}$$

Otherwise, $f(x)$ is *discontinuous* at $x = c$.

If the first two conditions hold but the third fails, we say that f has a *removable discontinuity* at $x = c$. The function in Figure 4.31(A) has a removable discontinuity as $c = 2$ because $\lim_{x\to 2} f(x) = 5$ exists but is not equal to function value $f(2) = 10$.

Removable discontinuities are "mild" in the following sense: We can make f continuous at $x = c$ by redefining $f(c)$. In Figure 4.31(B), the value $f(2)$ has been redefined as $f(2) = 5$ and this makes f continuous at $x = 2$.

A "worse" type of discontinuity is a *jump discontinuity,* which occurs if the one-sided limits

$$\lim_{x\to c^-} f(x) \qquad \text{and} \qquad \lim_{x\to c^+} f(x)$$

exist but are not equal. Figure 4.32 shows two functions with jump discontinuities at $c = 2$. Unlike the removable case, we cannot make $f(x)$ continuous by redefining $f(c)$.

We say that $f(x)$ has an *infinite discontinuity* at $x = c$ if one or both of the one sided limits is infinite (even if $f(x)$ itself is not defined at $x = c$). Figure 4.33 illustrates three types of infinite discontinuities occurring at $x = 2$. Notice that $x = 2$ does not belong to the domain of the function in cases (A) and (B).

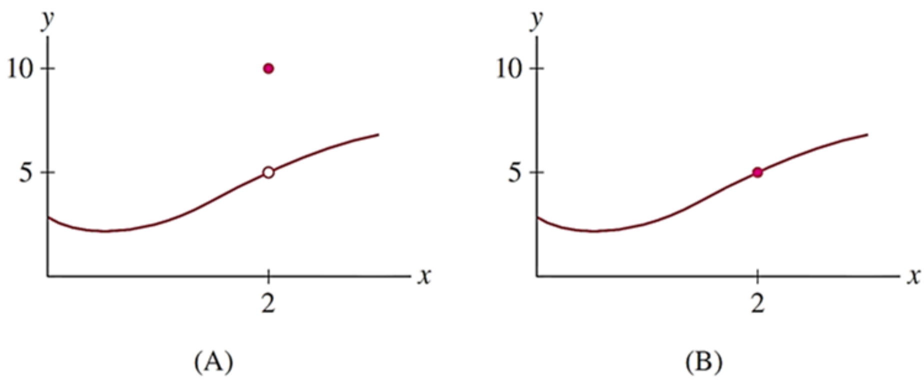


Fig. 4.31. redefining $f(2)$.

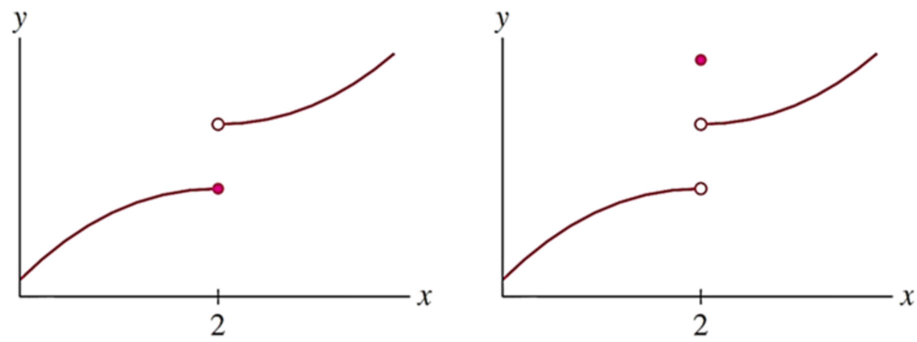


Fig. 4.32. Jump discontinuities.

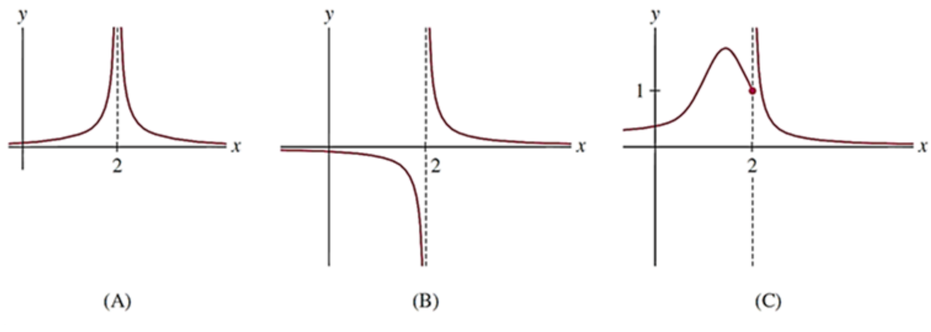


Fig. 4.33. Three types of infinite discontinuities occurring at $x = 2$.

We should mention that some functions have more "severe" types of discontinuity than those discussed above. For example, $f(x) = \sin(\frac{1}{x})$ oscillates infinitely often between $+1$ and -1 as $x \to 0$ (Figure 4.4 p. 74). Neither the left- nor the right-hand limits exist at $x = 0$, so this discontinuity is not a jump discontinuity.

4.9 The derivative as a limit of difference quotients

We recall the definition of the derivative given in Chapter 3

Definition 4.44 *Let f be a function whose domain contains an open interval about x_0. We say that the number m_0 is the derivative of f at x_0 provided that*

1. *For every $m < m_0$, the function*
$$f(x) - [f(x_0) + m(x - x_0)]$$
changes sign from negative to positive at x_0.

2. *For every $m > m_0$, the function*
$$f(x) - [f(x_0) + m(x - x_0)]$$
changes sign from positive to negative at x_0.

If such a number m_0 exists, we say that f is differentiable at x_0 and we write
$$m_0 = f'(x_0).$$

We will now prove that our definition of the derivative coincides with the definition found in most calculus books.

Theorem 4.45 *Let f be a function whose domain contains an open interval about x_0. Then f is differentiable at x_0 with derivative m_0 if and only if*
$$\lim_{\Delta x \to 0} \frac{[f(x_0 + \Delta x) - f(x_0)]}{\Delta x} \tag{4.19}$$
exists and equals m_0.

Proof. We will use the condition 1 form of the definition of limit. Suppose that
$$\lim_{\Delta x \to 0} \frac{[f(x_0 + \Delta x) - f(x_0)]}{\Delta x} = m_0.$$

To verify that

$$f'(x_0) = m_0$$

we must study the sign change at x_0 of

$$r(x) = f(x) - [f(x_0) + m(x - x_0)]$$

and see how it depends on m. First assume that $m < m_0$. Since the limit of difference quotients in (4.19) is m_0. there is an interval (a, b) containing zero such that

$$m < \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

whenever $a < \Delta x < b$, $\Delta x \neq 0$. Writing x for $x_0 + \Delta x$, we have

$$m < \frac{f(x) - f(x_0)}{\Delta x} \tag{4.20}$$

whenever $x_0 + a < x < x_0 + b$, $x \neq x_0$ - that is, whenever $x_0 + a < x < x_0 + b$, or $x_0 < x < x_0 + b$.

In case $x_0 + a < x < x_0$, we have $x - x_0 < 0$, and so equation (4.20) can be transformed to

$$\begin{aligned} m(x - x_0) &> f(x) - f(x_0) \\ 0 &> f(x) - [f(x_0) + m(x - x_0)]. \end{aligned}$$

When $x_0 < x < x + b$, we have $x - x_0 > 0$, so equation (4.20) becomes

$$\begin{aligned} m(x - x_0) &< f(x) - f(x_0) \\ 0 &< f(x) - [f(x_0) + m(x - x_0)]. \end{aligned}$$

In other words, $f(x) - [f(x_0) + m(x - x_0)]$ changes sign from negative to positive at x_0. Similarly, if $m > m_0$, we can use part 2XXX of the condition 1XXX from definition of limit to show that $f(x) - [f(x_0) + m(x - x_0)]$ changes sign from positive to negative at x_0. This completes the proof that $f'(x_0) = m_0$.

Next we show that if $f'(x_0) = m_0$, then

$$\lim_{\Delta x \to 0} \frac{[f(x_0 + \Delta x) - f(x_0)]}{\Delta x} = m_0.$$

This is mostly a matter of reversing the steps in the first half of the proof, with slightly different notation. Let $c_1 < m_0$. To find an interval (a, b) containing zero such that

$$c_1 < \frac{[f(x_0 + \Delta x) - f(x_0)]}{\Delta x} \tag{4.21}$$

whenever $a < \Delta x < b$, $\Delta x \neq 0$, we use the fact that

$$f(x) - [f(x_0) + c_1(x - x_0)]$$

changes sign from negative to positive at x_0. There is an interval (a_1, b_1) containing x_0 such that $f(x) - [f(x_0) + c_1(x - x_0)]$ is negative when $a_1 < x < x_0$ and positive when $x_0 < x < b_1$. Let $a = a_1 - x_0 < 0$ and $b = b_1 - x_0 > 0$. If $a < \Delta x < 0$, we have at $a_1 < x_0 + \Delta x < x_0$, and so

$$\begin{aligned}
0 &> f(x_0 + \Delta x) - [f(x_0) + c_1 \Delta x], \\
c_1 \Delta x &> f(x_0 + \Delta x) - f(x_0), \\
c_1 &< \frac{[f(x_0 + \Delta x) - f(x_0)]}{\Delta x} \text{ (since } \Delta x < 0)
\end{aligned}$$

which is just equation (4.21). If $0 < \Delta x < b$, we have $x_0 < x_0 + \Delta x < b_1$, and so

$$\begin{aligned}
0 &< f(x_0 + \Delta x) - [f(x_0) + c_1 \Delta x], \\
c_1 \Delta x &< f(x_0 + \Delta x) - f(x_0), \\
c_1 &< \frac{[f(x_0 + \Delta x) - f(x_0)]}{\Delta x}
\end{aligned}$$

which is equation (4.21) again.

Similarly, if $c_2 > m_0$, there is an interval (a, b) containing zero such that $c_2 > [f(x_0 + \Delta x) - f(x_0)]/\Delta x$ whenever $a < \Delta x < b$, $\Delta x \neq 0$. Thus we have shown that

$$\lim_{\Delta x \to 0} \frac{[f(x_0 + \Delta x) - f(x_0)]}{\Delta x} = m_0.$$

■

Combining Theorems (4.11) and (4.45), we can now give an ε-δ characterization of the derivative.

Corollary 4.46 *Let f be defined on an open interval containing x_0. Then f is differentiable at x_0 with derivative $f'(x_0)$ if and only if, for every positive number ε there is a positive number δ such that*

$$\left| \frac{[f(x_0 + \Delta x) - f(x_0)]}{\Delta x} - f'(x_0) \right| < \varepsilon$$

whenever $|\Delta x| < 0$ and $\Delta x \neq 0$.

Proof. We have just rephrased the statement

$$\lim_{\Delta x \to 0} \frac{[f(x_0 + \Delta x) - f(x_0)]}{\Delta x} = f'(x_0)$$

using the ε-δ definition of limit. ■

Example 4.47 *Find the derivative of $f(x) = x^3$.*

Solution: We use the limit definition as follows:

$$\begin{aligned}\lim_{\Delta x\to 0}\frac{[f(x+\Delta x)-f(x)]}{\Delta x} &= \lim_{\Delta x\to 0}\frac{(x+\Delta x)^3-x^3}{\Delta x}\\ &= \lim_{\Delta x\to 0}\frac{\left(x^3+3x^2\Delta x+3x(\Delta x)^2+(\Delta x)^3\right)-x^3}{\Delta x}\\ &= \lim_{\Delta x\to 0}\frac{\Delta x\left(3x^2+3x(\Delta x)+(\Delta x)^2\right)}{\Delta x}\\ &= \lim_{\Delta x\to 0}\left(3x^2+3x(\Delta x)+(\Delta x)^2\right)=3x^2.\end{aligned}$$

That is $f'(x) = 3x^2$. □

Example 4.48 (A graph with a sharp turn) *The function $f(x) = |x-2|$ is continuous for all real numbers. Is it differentiable at $c = 2$?*

Solution: We will use the tangent line formula at $c = 2$:

$$f'(c) = \lim_{x\to c}\frac{f(x)-f(c)}{x-c} = \lim_{x\to c}\frac{|x-2|-0}{x-2}.$$

By analyzing the 1-sided limits near 2, you see that this limit does not exist. Hence the function is not differentiable at 2. Geometrically, the graph has a sharp corner at 2 and is not smooth there. □

The definition of a tangent line to a curve does not cover the possibility of a vertical tangent line. For vertical tangent lines, you can use the following definition. If f is continuous at c and

$$\lim_{\Delta x\to 0}\frac{f(c+\Delta x)-f(x)}{\Delta x}=\infty \quad\text{or}\quad \lim_{\Delta x\to 0}\frac{f(c+\Delta x)-f(x)}{\Delta x}=-\infty$$

the vertical line $x = c$ passing through $(c, f(c))$ is a *vertical tangent line* to the graph of f (see Figure 4.34).

Example 4.49 (Graph with a vertical tangent line) *Is the function $f(x) = x^{\frac{1}{3}}$ differentiable at the point $(0,0)$?*

Solution:

$$\lim_{x\to 0}\frac{f(x)-f(0)}{x-0} = \lim_{x\to 0}\frac{x^{\frac{1}{3}}-0}{x} = \lim_{x\to 0}\frac{1}{x^{\frac{2}{3}}} = \infty.$$

Although the function is continuous at $(0,0)$, it is not differentiable there. In fact, the graph of f has a vertical tangent at this point. Recall in general that differentiability implies continuity, but the converse is false □

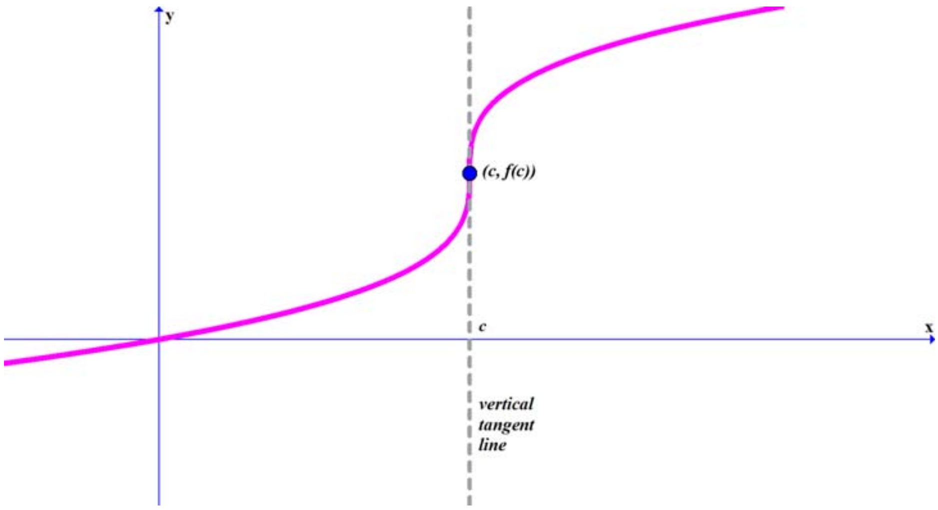


Fig. 4.34. The graph of f has a vertical tangent line at $(c, f(c))$.

4.10 Derivative Notation

For historical and practical reasons, several notations for the derivative are used. To see the origin of one notation, recall that the slope of the secant line $\vec{PQ}$ through two points $P(x, f(x))$ and $Q(x+h, f(x+h))$, on the curve $y = f(x)$ is

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

The quantity h is the change in the x-coordinates in moving from P to Q. A standard notation for change is the symbol Δ (uppercase Greek letter delta). So, we replace h by Δx to represent the change in x. Similarly, $f(x+h)-f(x)$ is the change in y, denoted Δy (Figure 4.35). Therefore, the slope of $\vec{PQ}$ is

$$m_{\text{sec}} = \frac{f(x+\Delta x)-f(x)}{\Delta x} = \frac{\Delta y}{\Delta x}.$$

By letting $\Delta x \to 0$, the slope of the tangent line at $(x, f(x))$ is

$$m_0 = f'(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x)-f(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \frac{dy}{dx}.$$

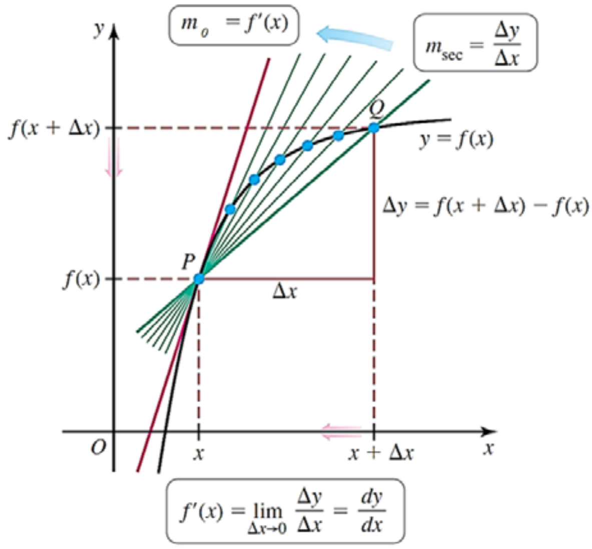


Fig. 4.35. As $\Delta x \to 0$, Q approaches P, the secant lines approach the tangent line and $m_{\text{sec}} \to m_0$.

The new notation[3] for the derivative is $\frac{dy}{dx}$; it reminds us that $f'(x)$ is the limit of $\frac{\Delta y}{\Delta x}$ as $\Delta x \to 0$. It is read *the derivative of y with respect to x* or dy/dx. It does not mean dy divided by dx, but it is a reminder of the limit of $\frac{\Delta y}{\Delta x}$.

In addition to the notation $f'(x)$ and $\frac{dy}{dx}$, other common ways of writing the derivative include

$$\frac{df}{dx}, \qquad \frac{d}{dx}(f(x)), \qquad D_x(f(x)), \qquad \text{and} \qquad y\prime(x).$$

Each of the following notations represents the derivative of f evaluated at a.

$$f'(a), \qquad y'(a), \qquad \frac{df}{dx}|_{x=a}, \qquad \text{and} \qquad \frac{dy}{dx}|_{x=a}.$$

4.11 Evaluating limits analytically

In Section 4.3, you learned that the limit of $f(x)$ as x approaches c does not depend on the value of f at c. It may happen, however, that the limit is precisely $f(c)$. In such cases, the limit can be evaluated by *direct substitution.*

[3]The derivative notation dy/dx was introduced by Gottfried Wilhelm von Leibniz (1646 – 1716), one of the coinventors of calculus. His notation is used today in its original form.

That is,

$$\lim_{x\to c} f(x) = f(c).$$

As we know such well-behaved functions are continuous at c.

Because polynomials are differentiable (so continuous) we can state the following (see Subsection 4.5.3):

Theorem 4.50 *If p is a polynomial function and c is a real number, then*

$$\lim_{x\to c} p(x) = p(c).$$

If r is a rational function given by $r(x) = p(x)/q(x)$ and c is a real number such that $q(c) \neq 0$, then

$$\lim_{x\to c} r(x) = r(c) = \frac{p(c)}{q(c)}.$$

Example 4.51 *Find the limit:*

$$\lim_{x\to 1} \frac{x^3 + 2x^2 - x + 6}{x + 1}.$$

Solution: Because the denominator is not 0 when you can apply Theorem 4.50 to obtain

$$\lim_{x\to 1} \frac{x^3 + 2x^2 - x + 6}{x + 1} = \frac{1^3 + 2 \cdot 1^2 - 1 + 6}{1 + 1} = 4.$$

□

Polynomial functions and rational functions are two of the three basic types of algebraic functions. The following theorem deals with the limit of the third type of algebraic function—one that involves a radical. A for a proof of this theorem follows from the differentiability of the functions of that kind (See xxx).

Theorem 4.52 *Let n be a positive integer. The following limit is valid for all c if n is odd, and is valid for $c > 0$ if n is even.*

$$\lim_{x\to c} \sqrt[n]{x} = \sqrt[n]{c}.$$

Having in mind Subsection **??** we can easily compute limits of some composite functions.

Example 4.53 a) *Because* $\lim_{x\to 0} \left(x^4 + 2\right) = 0^4 + 2 = 2$ *and* $\lim_{x\to 2} \sqrt{x} = \sqrt{2}$ *it follows that*

$$\lim_{x\to 0} \sqrt{x^4 + 2} = \sqrt{2}.$$

b) *Because* $\lim_{x\to3}\left(2x^2-10\right)=2\cdot3^2-10=8$ *and* $\lim_{x\to8}\sqrt[3]{x}=2$ *it follows that*

$$\lim_{x\to3}\sqrt[3]{2x^2-10}=2.$$

□

You have seen that the limits of many algebraic functions can be evaluated by direct substitution. The six basic trigonometric functions also exhibit this desirable quality, as shown in the next theorem (presented without proof, as it follows from the differentiability of these functions).

Theorem 4.54 *Let be a real number in the domain of the given trigonometric function. Then*

$$\begin{array}{llll} a) & \sin x=\sin c & b) & \lim_{x\to c}\cos x=\cos c \\ c) & \lim_{x\to c}\tan x=\tan c & d) & \lim_{x\to c}\cot x=\cot c \\ e) & \lim_{x\to c}\sec x=\sec c & f) & \lim_{x\to c}\csc x=\csc c \end{array} \tag{4.22}$$

Example 4.55 (Limits of trigonometric functions)

a) $\lim_{x\to0}\tan x=\tan(0)=0$,

b) $\lim_{x\to\pi/2}\left(x^2\sin x\right)=\left(\lim_{x\to\pi/2}x^2\right)\left(\lim_{x\to\pi/2}\sin x\right)=\frac{1}{4}\pi^2$,

c) $\lim_{x\to0}\left(\cos\left(x^2\right)\right)=1$.

4.12 A strategy for finding limits

On the previous pages, you studied several types of functions whose limits can be evaluated by direct substitution. This knowledge, together with the following theorem, can be used to develop a strategy for finding limits.

Theorem 4.56 *Let c be a real number and let $f(x)=g(x)$ for all $x\neq c$ in an open interval containing c. If the limit of $g(x)$ as approaches c exists, then the limit of $f(x)$ also exists and*

$$\lim_{x\to c}f(x)=\lim_{x\to c}g(x). \tag{4.23}$$

Proof. Let L be the limit of $g(x)$ as $x\to c$. Then, for each $\varepsilon>0$ there exists a $\delta>0$ such that $f(x)=g(x)$ in the open intervals $(c-\delta,c)$ and $(c,c+\delta)$ and

$$|g(x)-L|<\varepsilon \qquad \text{whenever} \quad 0<|x-c|<\delta.$$

Because $f(x) = g(x)$ for all x in the open interval other than $x = c$, it follows that $|f(x) - L| < \varepsilon$, whenever $0 < |x - c| < \delta$. So, the limit of $f(x)$ as $x \to c$ is also L. ■

	A STRATEGY FOR FINDING LIMITS
1.	Learn to recognize which limits can be evaluated by direct substitution.
2.	If the limit of $f(x)$ as x approaches c cannot be evaluated by direct substitution try to find a function $g(x)$ that agrees with f for all other than $x = c$. [Choose $g(x)$ such that the limit of $g(x)$ can be evaluated by direct substitution.]
3.	Apply Theorem 4.56 to conclude *analytically* that $\lim_{x\to c} f(x) = \lim_{x\to c} g(x) = g(c)$
4.	Use a graph or table to reinforce your conclusion.

Example 4.57 *Find the limit:*

$$\lim_{x\to 1} \frac{x^3 - 1}{x - 1}.$$

Solution: Although you are taking the limit of a rational function, you *cannot* apply Theorem 4.50 because the limit of the denominator is 0. Because the limit of the numerator is also 0, the numerator and denominator have a common factor of $(x - 1)$. So, for all $x \neq 1$ you can divide out this factor to obtain

$$f(x) = \frac{x^3 - 1}{x - 1} = \frac{(x - 1)\left(x + x^2 + 1\right)}{x - 1} = \left(x^2 + x + 1\right) = g(x) \quad \text{for} \quad x \neq 1.$$

So, for all x-values other than $x = 1$ the functions f and g agree, as shown in Figure 4.36-4.37. Because exists, you can apply Theorem 4.56 to conclude that f and g have the same limit at $x = 1$, and

$$\lim_{x\to 1} \frac{x^3 - 1}{x - 1} = \left(1^2 + 1 + 1\right) = 3. \quad \square$$

Example 4.58 (Rationalizing technique) *Find the limit:*

$$\lim_{x\to 0} \frac{\sqrt{x + 1} - 1}{x}.$$

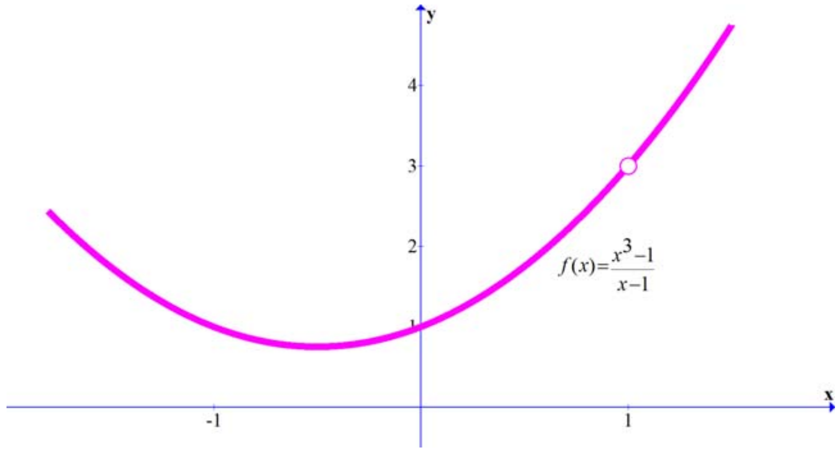


Fig. 4.36. The first function for example 4.57.

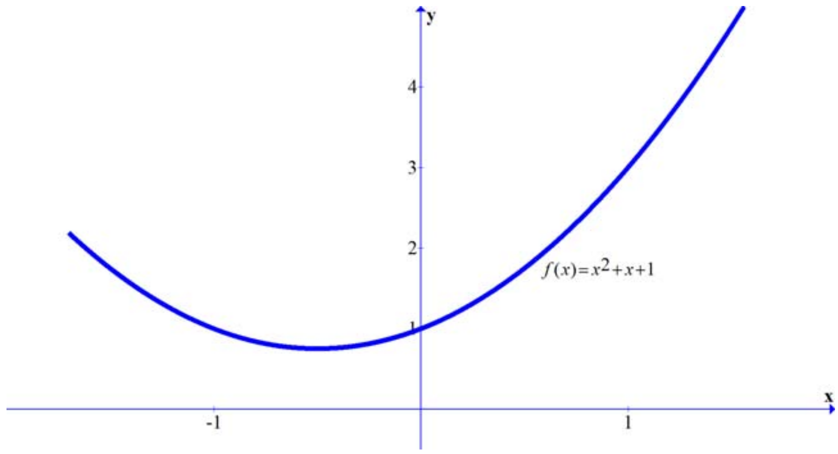


Fig. 4.37. The second function for example 4.57.

Solution: By direct substitution, you obtain the indeterminate form 0/0. In this case, you can rewrite the fraction by rationalizing the numerator.

$$\begin{aligned}\frac{\sqrt{x+1}-1}{x} &= \left(\frac{\sqrt{x+1}-1}{x}\right)\frac{\left(\sqrt{x+1}+1\right)}{\left(\sqrt{x+1}+1\right)}\\ &= \frac{1}{\sqrt{x+1}+1} \qquad \text{for} \quad x \neq 0.\end{aligned}$$

Now, using Theorem 4.56, you can evaluate the limit as shown.

$$\lim_{x\to 0}\frac{1}{\sqrt{x+1}+1} = \lim_{x\to 0}\frac{1}{\sqrt{x+1}+1} = \frac{1}{1+1} = \frac{1}{2}. \quad \square$$

As it was said (Section 4.5.5) when it is difficult to find the limit of a function directly, it is sometimes possible to obtain the limit indirectly by "squeezing" the function between the simpler functions whose limits are known.

Example 4.59 (Application of the Sandwich Principle)
Use the Sandwich Principle to evaluate the limit

$$\lim_{x\to 0} x^2 \sin^2 \frac{1}{x}.$$

Solution: If $x \neq 0$, we can write

$$0 \leq \sin^2 \frac{1}{x} \leq 1.$$

Multiplying through by x^2 yields

$$0 \leq x^2 \sin^2 \frac{1}{x} \leq x^2 \qquad \text{if} \quad x \neq 0.$$

But

$$\lim_{x\to 0} 0 = \lim_{x\to 0} x^2 = 0$$

so by the Sandwich Principle

$$\lim_{x\to 0} x^2 \sin^2 \frac{1}{x} = 0. \quad \square$$

Example 4.60 (A Limit involving a trigonometric function) *Find*

$$\lim_{x\to 0}\frac{\sin 3x}{\sin 5x}.$$

Solution: Direct substitution yields the indeterminate form To solve this problem, you can write

$$\lim_{x\to 0}\frac{\sin 3x}{\sin 5x} = \lim_{x\to 0}\frac{\frac{\sin 3x}{x}}{\frac{\sin 5x}{x}} = \lim_{x\to 0}\frac{3\frac{\sin 3x}{3x}}{5\frac{\sin 5x}{5x}} = \frac{3\lim_{x\to 0}\frac{\sin 3x}{3x}}{5\lim_{x\to 0}\frac{\sin 5x}{5x}} = \frac{3}{5}.$$

□

Example 4.61 (Divide by x/x) *Find*

$$\lim_{x\to\infty}\frac{(2x+1)}{(3x+1)}.$$

Solution:

$$\lim_{x\to\infty}\frac{(2x+1)}{(3x+1)} = \lim_{x\to\infty}\frac{\left(2+\frac{1}{x}\right)}{\left(3+\frac{1}{x}\right)} = \frac{2}{3}. \qquad \square$$

4.13 Review exercises: Chapter 4

<<<< * >>>>

Exercise 4.1 *Let $f(x)$ be the step function defined by*

$$f(x) = \begin{cases} = & -1 & \textit{if} & x < 0 \\ = & 2 & \textit{if} & x \geq 0 \end{cases}$$

Show that $f(x)$ is discontinuous at 0.

Exercise 4.2 *Show that, for any constants a and b, the linear function $f(x) = ax + b$ is continuous at $x_0 = 2$.*

Exercise 4.3 *Let $f(x)$ be the function defined by*

$$f(x) = \begin{cases} = & x^2+1 & \textit{if} & x < 1 \\ & ? & \textit{if} & 1 \leq x \leq 3 \\ = & x-6 & \textit{if} & 3 < x. \end{cases}$$

How can you define $f(x)$ on the interval $[1,3]$ in order to make $(-\infty,\infty)$? (A geometric argument will suffice.)

Exercise 4.4 *Let $f(x)$ be defined by $f(x) = (x^2-1)/(x-1)$ for $x \neq 1$. How should you define $f(1)$ to make the resulting function continuous? [`HINT`: Plot a graph of $f(x)$ for x near 1 by factoring the numerator.]*

Exercise 4.5 *Let $f(x)$ be defined by $f(x) = 1/x$ for $x \neq 0$. Is there any way to define $f(0)$ so that the resulting function will be continuous?*

Exercise 4.6 *Prove from the definition that the function $f(x) = x^2 + 1$ is continuous at 0.*

<<<< * >>>>

Exercise 4.7 *Prove that $f(x) = (x^2 - 1)/(x^3 + 3x)$ is continuous at $x = 1$.*

Exercise 4.8 *Is the converse of Theorem 4.7 true; i.e., is a function which is continuous at x_0 necessarily differentiable there? Prove or give an example.*

Exercise 4.9 *Prove that there is a number $\delta > 0$ such that $x^3 + 8x^2 + x < 1/1000$ if $0 < x < \delta$.*

Exercise 4.10 *Let $f(x)$ be. continuous at x_0 and A a constant. Prove that $f(x) + A$ is continuous at x_0.*

Exercise 4.11 *Why can't we ask whether the function $(x^3 - 1)/(x^2 - 1)$ is continuous at 1?*

Exercise 4.12 *Let*

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

Can you define $f(0)$ so that the resulting function is continuous at all x?

Exercise 4.13 *Find a function which is continuous on the whole real line, and which is differentiable for all x except $1, 2$, and 3. (A sketch will do.)*

Exercise 4.14 **a)** *Prove that if $f(x) < c_1$ for all x in J and $g(x) < c_2$ for all x in J, then $(f + g)(x) < c_1 + c_2$ for all x in J.*

b) *Prove that, if f and g are continuous at x_0, so is $f + g$.*

Exercise 4.15 *Let f be defined in an open interval about x_0. Suppose that $f(x) = f(x_0) + (x - x_0)g(x)$, where g is continuous at x_0. Prove that f is differentiable at x_0 and that $f'(x_0) = g(x_0)$. [HINT: Prove that $(x - x_0)(g(x) - g(x_0))$ vanishes rapidly at x_0.]*

<<<< * >>>>

Exercise 4.16 *Using the fact that the function $f(x) = \tan x$ is continuous, show that there is a number $\delta > 0$ such that $|\tan x - 1| < 0.001$ whenever $|x - \pi/4| < \delta$.*

Exercise 4.17 *Show that there is a positive number* δ *such that*

$$\left|\frac{x-4}{x+4}-\frac{1}{3}\right|<10^{-6}$$

whenever $|x-8|<\delta)$ *.*

Exercise 4.18 *Prove that limits are unique by using the definition.*

Exercise 4.19 *Which of the following functions are continuous at* 0*?*

a) $f(x)=x\sin\frac{1}{x},\ x\neq 0,\ f(0)=0,$

b) $f(x)=\frac{1}{x}\sin\frac{1}{x},\ x\neq 0,\ f(0)=0,$

c) $f(x)=x^2/\sin x,\ x\neq 0,\ f(0)=0.$

<<<< Sec. 4.9 >>>>

Exercise 4.20 *If* f *is differentiable at* x_0*, what is*

$$\lim_{x\to x_0}[f(x)-f(x_0)]/(x-x_0)?$$

Exercise 4.21 *Let* f *be defined near* x_0*, and define the function* $g(\Delta x)$ *by*

$$g(\Delta x)=\begin{cases}\frac{f(x_0+\Delta x)-f(x_0)}{\Delta x} & \Delta x\neq 0\\ m_0 & \Delta x=0\end{cases}$$

where m_0 *is some number. Show that* $f'(x_0)=m_0$ *if and only if* g *is continuous at* 0*.*

Exercise 4.22 *Find* $\lim_{x\to 2}(x^2+4x+3-15)/(x-2)$*.*

Exercise 4.23

a) *Suppose that* $f'(x_0)=g'(x_0)\neq 0$*. Find* $\lim_{x\to x_0}\frac{f(x)-f(x_0)}{g(x)-g(x_0)},$

b) *Find* $\lim_{x\to 1}\frac{2x^3-2}{3x^2-3},$

c) *Find* $\lim_{x\to 1}\frac{x^n-1}{x^m-1}.$

Exercise 4.24 *Evaluate* $\lim_{x\to 0} \dfrac{\sqrt{1-x^2}-1}{x}$

a) *By recognizing the limit to be a derivative.*

b) *By rationalizing.*

Exercise 4.25 *Evaluate the following limit by recognizing the limit to be a derivative:*

$$\lim_{x\to\frac{\pi}{4}} \frac{\sin(x)-\left(\sqrt{2}/2\right)}{x-(\pi/4)}.$$

<<<< * >>>>

Exercise 4.26 *Find*

$$\lim_{x\to -4}(x+3)^{2014}.$$

Answer: 1.

Exercise 4.27 *Find*

$$\lim_{x\to 1} \frac{\left(x^3-3x^2+5x-3\right)}{(x-1)}.$$

Answer: 2.

Exercise 4.28 *Find*

$$\lim_{u\to\sqrt{3}} \frac{\left(u-\sqrt{3}\right)}{(u^2-3)}.$$

Answer: $\frac{1}{6}\sqrt{3}$.

Exercise 4.29 *Find*

$$\lim_{x\to 3} \frac{(x-1)}{(x+1)}.$$

Answer: $\frac{1}{2}$.

Exercise 4.30 *Find*

$$\lim_{x\to 2} \frac{\left(x^2-5x+6\right)}{(x^2-6x+8)}$$

Answer: $\frac{1}{2}$.

Exercise 4.31 *Find*

$$\lim_{h\to 0} \frac{\left(125+75h+15\,(h)^2+(h)^3-125\right)}{h}$$

Answer: 75.

Exercise 4.32 *Find*

$$\lim_{x\to 0}\frac{\sqrt{2+x}-\sqrt{2}}{x}.$$

Answer: $\frac{1}{4}\sqrt{2}$.

Exercise 4.33 *Find*

$$\lim_{x\to 0}\frac{\frac{1}{x+4}-\frac{1}{4}}{x}.$$

Answer: $-\frac{1}{16}$.

Exercise 4.34 *Find*

$$\lim_{x\to 0}\frac{\tan^2 x}{x}.$$

Answer: 0.

Exercise 4.35 *Find*

$$\lim_{\theta\to 0}\frac{\cos\theta\tan\theta}{\theta}.$$

Answer: 1.

Exercise 4.36 *Find*

$$\lim_{x\to\pi/4}\frac{1-\tan x}{\sin x-\cos x}.$$

Answer: $-\sqrt{2}$.

Exercise 4.37 *Find*

$$\lim_{x\to 0}\frac{\sin x}{x^{1/3}}.$$

Answer: 0.

Exercise 4.38 *Find*

$$\lim_{x\to 0}\frac{\sec x-1}{x^2}.$$

Answer: $\frac{1}{2}$

Exercise 4.39 *Find*

$$\lim_{x\to 4}\frac{4-x}{5-\sqrt{x^2+9}}.$$

Answer: $\frac{5}{4}$.

Exercise 4.40 *Find*

$$\lim_{x\to\infty}\sqrt[3]{\frac{3x+5}{6x-8}}.$$

Answer: $\sqrt[3]{\frac{1}{2}}\approx 0.793\,7$.

Exercise 4.41 *Find*

$$\lim_{x\to-\infty} \frac{\sqrt{x^2+2}}{3x-6}.$$

`Answer:` $-\frac{1}{3}$.

Exercise 4.42 *Find*

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

`Answer:` -4.

Exercise 4.43 *Find* $\lim_{x\to 3} f(x)$ *for*

$$f(x) = \begin{cases} x^2-5, & x \le 3 \\ \sqrt{x+13}, & x > 3. \end{cases}$$

Exercise 4.44 *Using the Sandwich Theorem show that*

a) $\lim_{x\to 0} x \sin \frac{1}{x} = 0$,

b) $\lim_{x\to 0} x^2 \cos \frac{1}{x^2} = 0$.

Exercise 4.45 *A prototype function for studying limits is the* sinc *function (see p. 95). Answer following questions:*

1. *Does the function* $\dfrac{\cos x}{x}$ *have a limit at* $x \to 0$*?*

   `Answer: Undefined.`

2. *Does the function* $\dfrac{\sin(x^2)}{x^2}$ *have a limit at* $x \to 0$*?*

   `Answer: 1.`

3. *Does the function* $\dfrac{\sin(x^2)}{x}$ *have a limit at* $x \to 0$*?*

   `Answer: 0.`

4. *Does the function* $\dfrac{\sin(x)}{x^2}$ *have a limit at* $x \to 0$*?*

   `Answer: Undefined.`

5. *Does the function* $\dfrac{x}{\sin x}$ *have a limit at* $x \to 0$*?*

   `Answer: 1.`

6. *Does the function* $\dfrac{\sin(x)}{|x|}$ *have a limit at* $x \to 0$*?*

   `Answer: Undefined.`

Exercise 4.46 *For the following functions, determine the points, where f is not continuous.*

a) $\text{sinc}(x) + 1/\cos(x)$,

b) $\sin(\tan(x))$,

c) $f(x) = \cot(2 - x)$,

d) $\text{sign}(x)/x$,

e) $\dfrac{x^2 + 5x + x^4}{x - 3}$,

f) $\dfrac{x}{\sin(x)}$.

State which kind of discontinuity appears.

Solution:

a) continuous everywhere except at $x = \pi/2 - \pi, \pi/2, \pi/2 + \pi, ...$ Note that $\text{sinc}(x)$ is considered continuous since we defined it to be 1 at $x = 0$ by the fundamental theoremof trigonometry.

b) same answer as in a). It is the $\tan(x)$ function which is discontinuous at $\pi/2 + k\pi$.

c) $x = 2 + k\pi$.

d) This function has a discontinuity at $x = 0$. There is no way, we can fix the pole discontinuity there.

e) This function has a pole at $x = 3$. There is no way we can fix the discontinuity at this point.

f) This function is discontinuous at $k\pi$. We can fix this discontinuity.

5

Benefits of continuity

Knowing that a function is continuous brings some benefits. We're going to look at two such benefits and state two key theorems about continuous functions *on a closed interval*. The first is called the *intermediate value theorem*, or IVT for short. The second is usually called *max-min* theorem (or *extreme value theorem*). Both of them are very important not only for theoretical but also for applied mathematics.

5.1 The intermediate value theorem (IVT)

A function which is continuous on an interval does not "skip" any values, and thus its graph is an "unbroken curve." There are no "holes" in it and no "jumps." This idea is expressed coherently by the intermediate-value theorem.

Figure 5.1 shows the graph of a function that is continuous on the closed interval $[a,\ b]$. The figure suggests that if we draw any horizontal line $y = k$, where k is between $f(a)$and $f(b)$, then that line will cross the curve $y = f(x)$ at least once over the interval $[a,\ b]$. Stated in numerical terms, if f is continuous on $[a,\ b]$, then the function f must take on every value k between $f(a)$ and $f(b)$ at least once as x varies from a to b. For example, the polynomial $p(x) = x^5 - x + 3$ has a value of 3 at $x = 1$ and a value of 33 at $x = 2$. Thus, it follows from the continuity of p that the equation $x^5 - x + 3 = k$ has at least one solution in the interval $[1,\ 2]$ for every value of k between 3 and 33. This idea is stated more precisely in the following theorem.

Theorem 5.1 (Intermediate-value theorem) *If f is continuous on a closed interval $[a,\ b]$ and k is any number between $f(a)$ and $f(b)$, inclusive, then there is at least one number x in the interval $[a,\ b]$ such that $f(x) = k$.*

Although this theorem is intuitively obvious, its proof depends on a mathematically precise development of the real number system, which is beyond the scope of this text.

It's a small step from the intermediate-value theorem to the following observation:

"*continuous functions map intervals onto intervals.*"

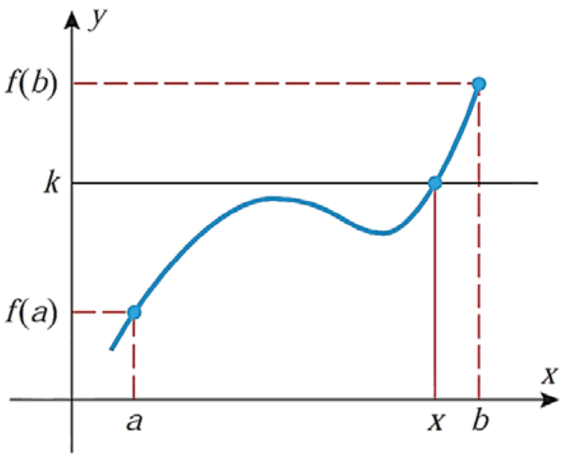


Fig. 5.1. The intermediate value theorem says, that there exists at least one number x between a and b at which $f(x) = k$.

5.2 Bisection method

Definition 5.2 *The function $f(x)$ has a root at $x = r$ if $f(r) = 0$.*

A variety of problems can be reduced to solving an equation $f(x) = 0$ for its roots. Sometimes it is possible to solve for the roots exactly using algebra, but often this is not possible and one must settle for decimal approximations of the roots. One procedure for approximating roots is based on the following consequence of the intermediate-value theorem.

Theorem 5.3 *If f is continuous on $[a, b]$, and if $f(a)$ and $f(b)$ are nonzero and have opposite signs, then there is at least one solution of the equation $f(x) = 0$ in the interval (a, b).*

Proof. Since $f(a)$ and $f(b)$ have opposite signs, 0 is between $f(a)$ and $f(b)$. Thus, by the intermediate-value theorem there is at least one number x in the interval $[a, b]$ such that $f(x) = 0$. However, $f(a)$ and $f(b)$ are nonzero, so x must lie in the interval (a, b), which completes the proof. ■

If f is a continuous function, then there will be at least one root: an r between a and b for which $f(r) = 0$. This fact is summarized in the following corollary of the intermediate value theorem 5.1:

Theorem 5.4 *Let f be a continuous function on $[a, b]$, satisfying $f(a)f(b) < 0$. Then f has a root between a and b, that is, there exists a number r satisfying $a < r < b$ and $f(r) = 0$.*

Using Theorem 5.4 many times consecutively we can construct the following procedure.

Bisection method: Assume f is continuous. Suppose that $f(a)$ and $f(b)$ have opposite signs, so that f has a zero in $(a,\ b)$. Then f has a zero in $[a,\ c]$ or $[c,\ b]$, where $c = (a+b)/2$ is the midpoint of $[a,\ b]$. To determine which, compute $f(c)$. A zero lies in $(a,\ c)$ if $f(a)$ and $f(c)$ have opposite signs and in $(c,\ b)$ if $f(c)$ and $f(b)$ have opposite signs. Continuing the process, we can locate a zero with arbitrary accuracy. This process can be presented as a pseudocode in the following form:

Algorithm 5.5 (Bisection method) *Given initial interval* $[a,\ b]$ *such that* f *is a continuous function on* $[a,\ b]$*, satisfying* $f(a)f(b) < 0$

while $(b-a)/2 > TOL$
  $c = (a+b)/2$
  if $f(c) = 0$*, stop,* ***end***
  if $f(a)f(c) < 0$
    $b = c$
  else
    $a = c$
  end
end

Example 5.6 *Show that* $f(x) = \cos x$ *has a zero in* $(0.5,\ 2)$*. Then locate the zero more accurately using the bisection method.*

Solution: Using a calculator, we find that $f(0.5)$ and $f(2)$ have opposite signs: $f(0.5) \approx 0.877\,58 > 0$, $f(2) \approx -0.416\,15 < 0$.Theorem 5.3 guarantees that $f(x) = 0$ has a solution in $(0.5,\ 2)$ (Figure 5.2).We can locate a zero more accurately by dividing $[0.5,\ 2]$ into two intervals $[0.5,\ 1.25]$ and $[1.25,\ 2]$. One of these must contain a zero of $f(x)$. To determine which, we evaluate $f(x)$ at the midpoint $c = 1.25$. A calculator gives $f(1.25) \approx 0.315\,32 > 0$, and since $f(0.5) > 0$, we see that $f(x)$ takes on opposite signs at the endpoints of $[1.25,\ 2]$. Therefore, $(1.25,\ 2)$ must contain a zero. We discard $[0.5,\ 1.25]$ because both $f(0.5)$ and $f(1.25)$ are positive. (Figure 5.3).

The Bisection Method consists of continuing this process until we narrow down the location of the zero to the desired accuracy. In the Figure 5.4, the process is carried out five times:

We conclude that $f(x)$ has a zero r satisfying $1.53125 < r < 1.625$. The bisection method can be continued to locate the root with any degree of accuracy. The exact solution for this problem is equal to $\pi/2 \approx 1.5708$ □

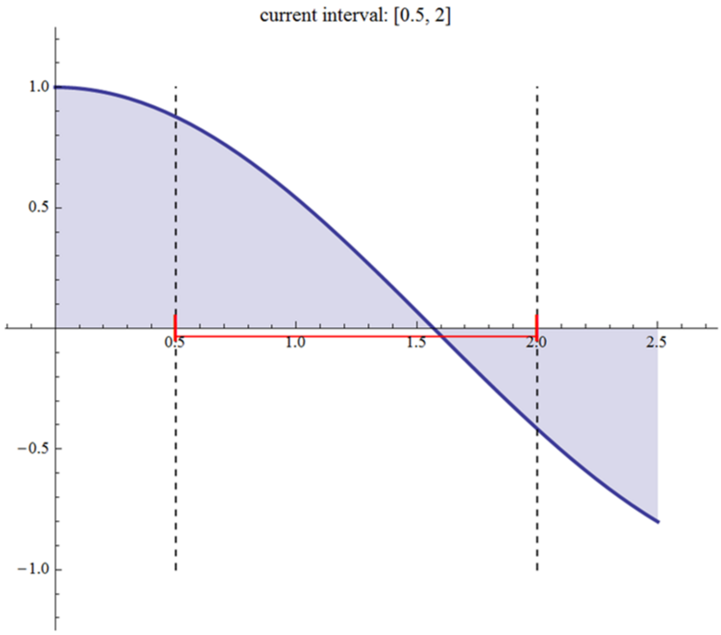


Fig. 5.2. Iteration step 1.

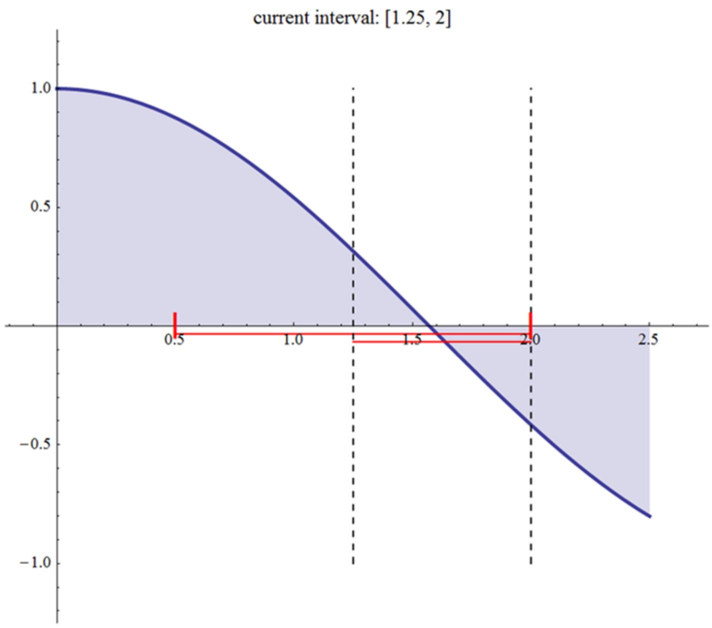


Fig. 5.3. Iteration step 2.

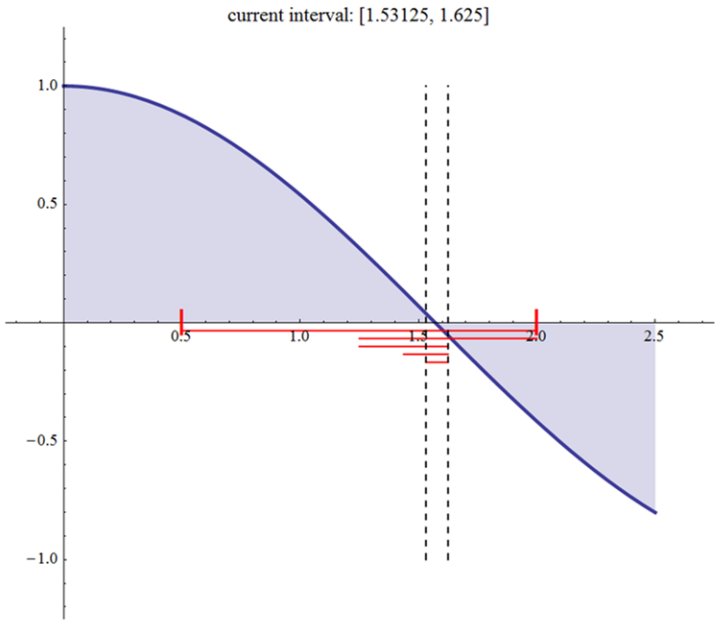


Fig. 5.4. Iteration step 5.

5.3 Max-min theorem

Let's move on to the second benefit of knowing that a function is continuous. Suppose we have a function f which we know is continuous on the closed interval $[a,\ b]$. (It's very important that the interval *is closed* at both ends.) That means that we put our pen down at the point $(a,\ f(a))$ and draw a curve that ends up at $(b,\ f(b))$ without taking our pen off the paper. The question is, how high can we go? In other words, is there any limit to how high up this curve could go? The answer is yes: there must be a highest point, although the curve could reach that height multiple times.

In symbols, let's say that the function f defined on the interval $[a,\ b]$ has a *maximum*[1] at $x = c$ if $f(c)$ is the highest value of f on the whole interval $[a,\ b]$. That is, $f(c) \geq f(x)$ for all x in the interval. The idea that I've been driving at is that a continuous function on $[a,\ b]$ has a *maximum* in the interval $[a,\ b]$. The same is true for the limbo question, "how low can you go?" We'll say that f has a *minimum*[2] at $x = c$ if $f(c)$ is the lowest value of f on the

[1] Also called "*absolute maximum*" or "*global maximum*".
[2] Also called "*absolute minimum*" or "*global minimum*"

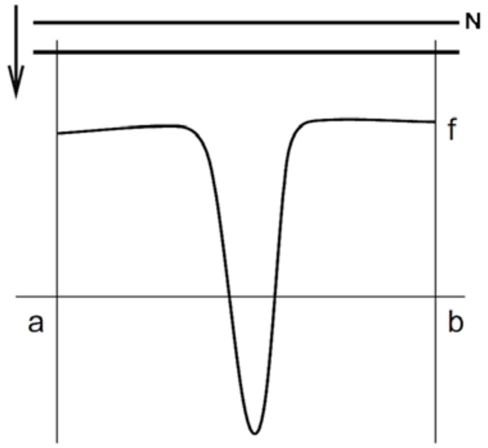


Fig. 5.5. A horizontal line moves down, seeking the minimum of a continuous function. The max-min guarantees that there is a last value where you can stop the moving line, keeping it in contact with the graph.

whole interval; that is, that $f(c) \leq f(x)$ for all x in $[a,\ b]$. Once again, any continuous function on the interval $[a,\ b]$ has a minimum in that interval.

Alternatively, let us look at the graph of f and imagine a line parallel to the x-axis slid vertically upward until it just touches the graph of f at some last point of intersection, which is the maximum. Similarly, slide a line parallel to the x-axis vertically downward. The last point of intersection with the graph of f is the minimum value of f (Fig. 5.5).

These facts form a theorem, sometimes known as the *max-min theorem* (or *extreme value theorem*), which can be stated as follows:

Theorem 5.7 (Max-min) *If f is a continuous function on a closed interval $[a,\ b]$, then f takes on both a maximum value and a minimum value at some points in $[a,\ b]$.*

One consequence of the extreme value theorem is that every function that is continuous on a closed interval is bounded. Although the extreme value theorem does not tell us how or where to find the bounds, it is still very useful. Figure 5.6 presents an example of continuous function on $[a,\ b]$ and their *maxima* and *minima* (these are the plurals of maximum and minimum, respectively.)

Why does the function f need to be continuous? And why can't it be an open interval, like $(a,\ b)$? The following diagrams show some potential problems:

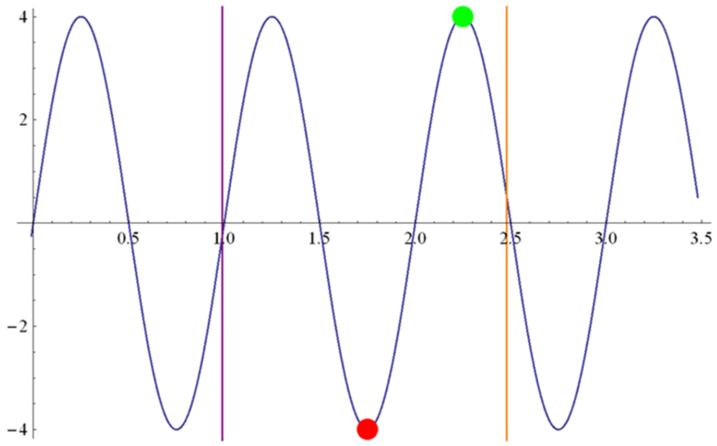


Fig. 5.6. If the function f is continuous over the closed interval $[a, b]$, then there is at least one maximum value (green) and one minimum value (red) of f in that interval (here a is the purple endpoint, and b is the orange endpoint.)

In the first (Figure 5.7), the function f has an asymptote in the middle of the interval $[a, \ b]$, which certainly creates a discontinuity. The function has no maximum value–it just keeps going up and up on the left side of the asymptote. Similarly, it has no minimum value either, since it just plummets way down on the right side of the asymptote.

The second diagram (Figure 5.8) involves a more subtle situation. Here the function is only continuous on the open interval $(a, \ b)$. It clearly has a minimum at $x = c$, but what is the maximum of this function? You might think that it occurs at $x = b$, but think again. The function isn't even defined at $x = b$! So it can't have a maximum there. If the function has a maximum, it must be somewhere near b. In fact, you'd like it to be the number less than b which is closest to b. Unfortunately, there is no such number! Whatever you think the closest number is, you can always take the average of this number and b to get an even closer number. So there is no maximum; this illustrates that the interval of continuity has to be closed in order to guarantee that the max-min theorem works.

Of course, the conclusion of the theorem could still be true even if the interval isn't closed. For example, the function in the third diagram (Figure 5.9) is only continuous on the open interval $(a, \ b)$, but it still has a maximum at $x = c$ and a minimum at $x = d$. This was just a lucky accident: you can only use the theorem to guarantee the existence of a maximum and minimum

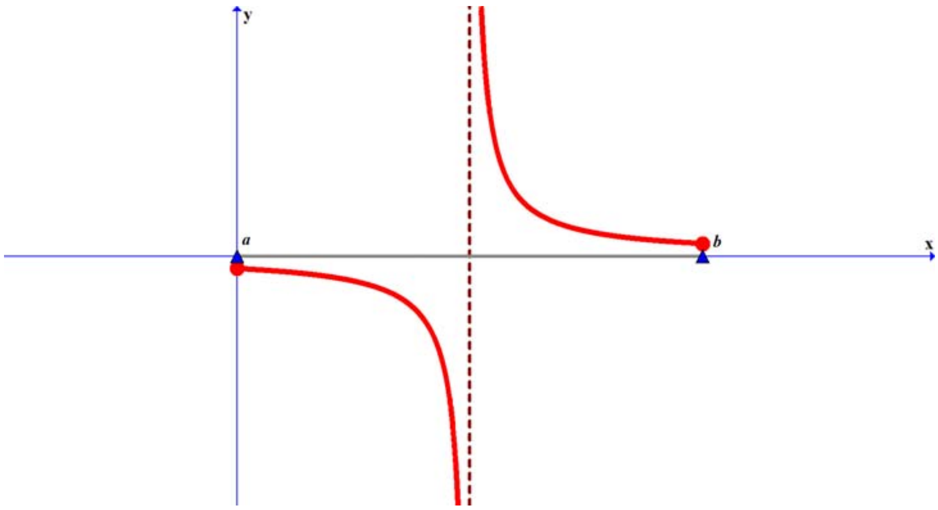


Fig. 5.7. The function has no maximum value–it just keeps going up and up on the left side of the asymptote.

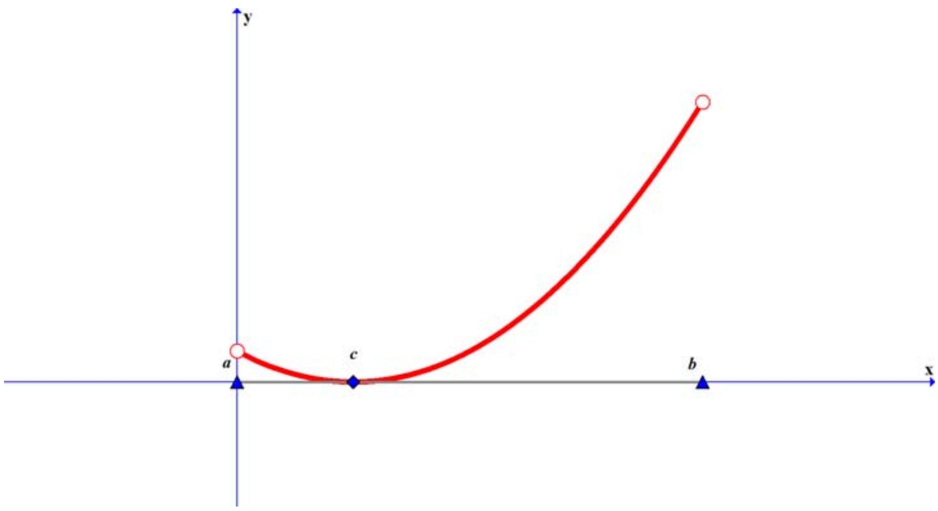


Fig. 5.8. There is no maximum in an interval $[a; b]$.

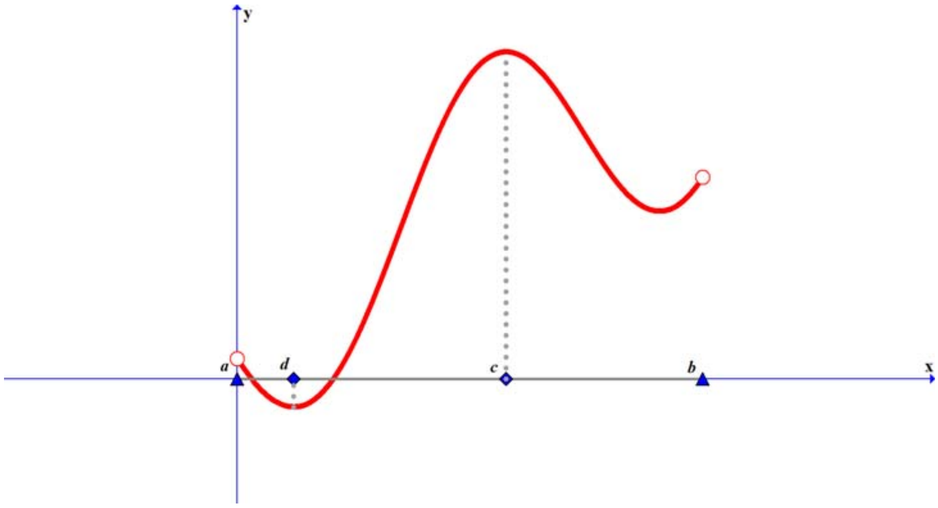


Fig. 5.9. A lucky accident: extrema exist in an open interval (a, b).

in an interval $[a,\ b]$ if you know the function is continuous on the entire closed interval.

Now let's prove the max-min Theorem, using a the method of successive bisection.

Proof. The first thing we want to show is that you can plonk down some horizontal line at $y = N$, say, such that the function values $f(x)$ all lie below that line. If you couldn't do that, then the function would somehow grow bigger and bigger somewhere inside $[a,\ b]$, and it wouldn't have a maximum. So, let's suppose you can't draw such a line (Figure 5.5). Then for every positive integer N, there's some point x_N in $[a,\ b]$ such that $f(x_N)$ is above the line $y = N$. That is, we have found some points x_N such that $f(x_N) > N$ for every N. Let's mark them on the x-axis with an X.

Now, where are these marked points? There are infinitely many. So if we chop the interval $[a,\ b]$ in half to get two new closed intervals, one of them muststill have infinitely many points from X. Perhaps they both do, but they can't both have finitely many marked points or else the total would be finite.

Let's focus on the half of the original interval that has infinitely many marked points; if they both do, choose your favorite one (it doesn't matter). Now repeat the exercise with the new, smaller interval: chop it in half. One of the halves must have infinitely many marked points. Continue doing this for as long as you like, and you will get a collection of intervals which get smaller and smaller, all nested inside each other, and each of which has infinitely many marked points.

Intuitively, there has to be some real number which is inside every single one of these intervals[3]. Let's call the number q. What is $f(q)$? We can use thecontinuity of f to get some idea of what it should be. Indeed, we know that

$$\lim_{x \to q} f(x) = f(q).$$

So if you pick your ε to be 1, for example, then I should be able to find an interval $(q - \delta,\ q + \delta)$ so that $|f(x) - f(q)| < 1$ for all x in the interval. The problem is that the interval $(q - \delta,\ q + \delta)$ contains infinitely many marked points! This is because eventually one of the little nested intervals that we chose will lie within $(q - \delta,\ q + \delta)$, no matter how small is. This is a real problem: we are supposed to have all these marked points inside our interval $(q-\delta,\ q+\delta)$, but when you take f of any of them, you get a number between $f(q) - 1$ and $f(q) + 1$. So, no matter what $f(q)$ is, we're going to get in trouble: some of the marked points are going to have function values which are much bigger than $f(q) + 1$. The whole thing is out of control. So we were wrong about not being able to draw in a line like $y = N$ which had the whole function beneath it!

We're still not done. We have this line $y = N$ which lies above the graph of $y = f(x)$ on $[a,\ b]$, but now we need to move it down until it hits the graph in order to find the maximum. So, let's pick N as small as possible so that $f(x) \leq N$ for all x in $[a,\ b]$. (We have used completeness once again.) Now we need to show that $N = f(c)$ for some c. To do this, we're going to repeat the same trick as we did above with marked points, except this time they'll be circled. Pick a positive integer n; we must be able to find some number c_n in $[a,\ b]$ such that $f(c_n) > N - 1/n$. If not, then we should have drawn our line at $y = N - 1/n$ (or even lower) instead of $y = N$. So there is such a c_n, and there's one for every positive integer n. Circle all of these points. There are infinitely many of them, and when you apply f to them, the resulting values get closer and closer|arbitrarily close, in fact–to N. (None of the values canbe bigger than N because $f(x) \leq N$ for all x!) Now all we have to do is keep bisecting the interval $[a,\ b]$ over and over again, such that each little interval has infinitely many circled points in it. As before, there is a number c in all the intervals. This number is really surrounded by a fog of circled points. What is $f(c)$? It can't be more than N, but maybe it can be less than N. Let's suppose that $f(c) = M$, where $M < N$, and let's set $\varepsilon = (N - M)/2$. Since f is continuous, we really need

$$\lim_{x \to c} f(x) = f(c) = M.$$

[3]Again, one needs to use the completeness property of the real line to show this. Actually, there has to be exactly one such number|can you see why?

You have your ε, and so I need to find an interval $(c-\delta,\ c+\delta)$ so that $f(x)$ lies in $(M-\varepsilon,\ M+\varepsilon)$ for x in my interval. The problem is that $M+\varepsilon = N-\varepsilon$, and also that there are infinitely many circled points lying in $(c-\delta,\ c+\delta)$, no matter how I choose $\delta > 0$. Some of them might have function values lying in $(M-\varepsilon,\ M+\varepsilon)$, but since the function values get closer to N, most of them won't. So I can't make my move. The only way out is that $f(c) = N$ after all. This means that c is a maximum, and we're done! To get the minimum version of the theorem, just reapply the theorem to $g(x) = -f(x)$. After all, if c is a maximum for g, then it is a minimum for f. ■

This theorem shows that if f is continuous on $[a,\ b]$, then it takes on both a maximum value and a minimum value at some points in $[a,\ b]$. Hence, by the intermediate-value theorem (as it was said), the range of f is the closed interval between its minimum and its value maximum value.

5.4 Review exercises: Chapter

6
How to solve differentiation problems

Calculating the derivative of

$$f(x) = \frac{x^2 + x + 1}{x^3 - 1} \quad \text{or} \quad f(x) = \left(x^2 + x + 1\right)\left(x^3 - 1\right)$$

by forming the appropriate *difference quotient*

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

and then taking the limit as h tends to 0 is somewhat laborious. Here we derive some general formulas that enable us to calculate such derivatives quite quickly and easily.

6.1 General formulas for finding derivatives

6.1.1 Derivatives of sums and scalar multiples

It's easy to deal with a constant multiple of a function: you just multiply by the constant after you differentiate. For example, we know the derivative of x^2 is $2x$; so the derivative of $7x^2$ is 7 times $2x$, or $14x$.

Theorem 6.1 (Constant multiple rule) *If f is differentiable at x and c is any real number, then cf is also differentiable at x and*

$$(cf)'(x) = cf'(x) \tag{6.1}$$

Proof.

$$\begin{aligned} (cf)'(x) &= \lim_{h \to 0} \frac{cf(x+h) - cf(x)}{h} \\ &= \lim_{h \to 0} c\frac{f(x+h) - f(x)}{h} \\ &= c \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \\ &= cf'(x) \end{aligned}$$

As a constant factor can be moved through a limit sign. ■

Theorem 6.2 (Sum and difference rules) *If f and g are differentiable at x, then so are $f+g$ and $f-g$ and*

$$(f+g)'(x) = f'(x)+g'(x) \tag{6.2}$$
$$(f-g)'(x) = f'(x)-g'(x) \tag{6.3}$$

Proof.

$$\begin{aligned}(f+g)'(x) &= \lim_{h\to 0}\frac{(f(x+h)+g(x+h))-(f(x)+g(x))}{h}\\ &= \lim_{h\to 0}\frac{(f(x+h)-f(x))+(g(x+h)-g(x))}{h}\\ &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}+\lim_{h\to 0}\frac{g(x+h)-g(x)}{h}\\ &= f'(x)+g'(x),\end{aligned}$$

because limit of a sum is the sum of the limits. Formula (6.3) can be proved in a similar manner or, alternatively, by writing $f(x)-g(x)$ as $f(x)+(-1)g(x)$ and then applying Formulas (6.2) and (6.1). ∎

You may find it useful to put these formulas into words. According to Theorem 6.2,

"*the derivative of a sum is the sum of the derivatives*",

"*the derivative of a difference is the difference of the derivatives*",

and (Theorem 6.1)

"*the derivative of a scalar multiple is the scalar multiple of the derivative.*"

These results can be extended by induction to any finite collection of functions: if $f_1, f_2, ..., f_n$ are differentiable at x, and $\alpha_1, \alpha_2, ..., \alpha_n$ are numbers, then the linear combination $\alpha_1 f_1 + \alpha_2 f_2 + ...\alpha_n f_n$ is differentiable at x and

$$(\alpha_1 f_1 + \alpha_2 f_2 + ...\alpha_n f_n)'(x) = \alpha_1 f_1'(x) + \alpha_2 f_2'(x) + ... + \alpha_n f_n'(x). \tag{6.4}$$

So,

"the derivative of a linear combination is the linear combination of the derivatives."

6.1.2 Derivatives of products and quotiens

Suppose we know the derivatives of $f(x)$ and $g(x)$ and want to calculate the derivative of the product, $f(x)g(x)$. We start by looking at an example. Let $f(x) = x$ and $g(x) = x^2$. Then

$$f(x)g(x) = xx^2 = x^3,$$

so the derivative of the product is $3x^2$. Notice that the derivative of the product is *not equal* to the product of the derivatives, since $f'(x) = 1$ and $g'(x) = 2x$, so $f'(x)g'(x) = (1)(2x) = 2x$. In general, we have the following rule:

Theorem 6.3 (The product rule) *If f and g are differentiable at x, then so is their product, and*

$$(f{\cdot}g)'(x) = f(x)g'(x) + g(x)f'(x). \tag{6.5}$$

In words:

"the derivative of a product is the first function times the derivative of the second plus the second function times the derivative of the first."

Proof. We form the difference quotient

$$\begin{aligned} & \frac{(f{\cdot}g)(x+h) - (f{\cdot}g)(x)}{h} \\ = & \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\ = & \frac{f(x+h)g(x+h) - f(x+h)g(x) + f(x+h)g(x) - f(x)g(x)}{h} \end{aligned}$$

and rewrite it as

$$f(x+h)\left[\frac{g(x+h)-g(x)}{h}\right] + g(x)\left[\frac{f(x+h)-f(x)}{h}\right].$$

(Here we have added and subtracted $f(x+h)g(x)$ in the numerator and then regrouped the terms so as to display the difference quotients for f and g.) Since f is differentiable at x, we know that f is continuous at x (Theorem 4.7) and thus

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

Since

$$\lim_{h\to 0}\frac{g(x+h)-g(x)}{h} = g'(x) \qquad \text{and} \qquad \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} = f'(x),$$

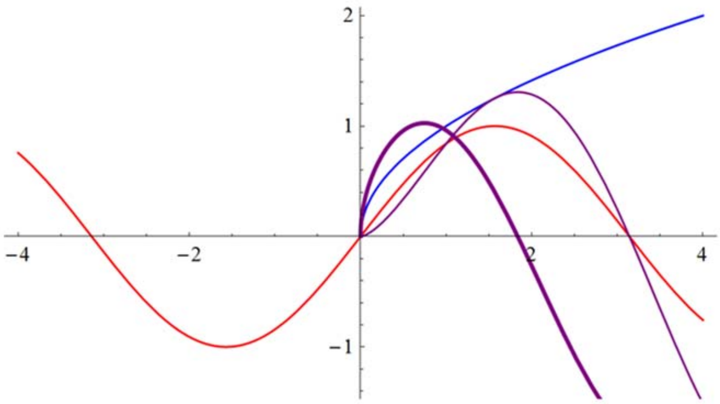


Fig. 6.1. If f and g are both differentiable at x, then the derivative of their product at x is given by the product rule. On the graph, $f(x) = \sqrt{x}$ is blue, $g(x) = \sin(x)$ is red, $f(x)g(x)$ is purple, and the derivative of $f(x)g(x)$ ie. $\frac{\sin x}{2\sqrt{x}} + \sqrt{x}\cos x$ is thick purple.

we obtain

$$\begin{aligned} & \lim_{h\to 0} \frac{(f\cdot g)(x+h) - (f\cdot g)(x)}{h} \\ = & \lim_{h\to 0} f(x+h) \lim_{h\to 0} \frac{g(x+h)-g(x)}{h} + g(x) \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \\ = & f(x)g'(x) + f'(x)g(x). \end{aligned}$$

■

Using the product rule, it is not hard to show (by induction) that

$$\boxed{\begin{array}{l} \text{for each positive integer } n \\ \quad p(x) = x^n \text{ has derivative } p'(x) = nx^{n-1}. \end{array}} \tag{6.6}$$

In particular,

$$\begin{array}{lll} p(x) = x & \text{has derivative} & p'(x) = 1 = 1\cdot x^0,^{1} \\ p(x) = x^2 & \text{has derivative} & p'(x) = 2x, \\ p(x) = x^3 & \text{has derivative} & p'(x) = 3x^2, \\ p(x) = x^4 & \text{has derivative} & p'(x) = 4x^3, \end{array}$$

and so on.

[1] In this setting we are following the convention that x^0 is identically 1 even though in itself 0^0 is meaningless.

Remark 6.4 *Formula (6.6) can be obtained without induction. From the difference quotient*

$$\frac{p(x+h)-p(x)}{h} = \frac{(x+h)^n - x^n}{h}$$

apply the formula

$$a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + \cdots + ab^{n-2} + b^{n-1}),$$

and you'll see that the difference quotient becomes the sum of n terms, each of which tends to x^{n-1} as h tends to zero.

The formula for differentiating polynomials follows from (6.4) and (6.6):

$$\boxed{\begin{array}{l} P(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0, \\ \text{then} \quad P'(x) = n a_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + 2a_2 x + a_1. \end{array}} \tag{6.7}$$

Example 6.5 *For example,*

$$P(x) = 12x^3 - 6x^2 - 2x - 1 \quad \textit{has derivative} \quad P'(x) = 36x^2 - 12x - 2$$

and

$$Q(x) = \frac{1}{4}x^4 - 2x^2 + x + 5 \quad \textit{has derivative} \quad Q'(x) = x^3 - 4x + 1.$$

□

Example 6.6 *Differentiate $F(x) = (x^3 - 2x + 3)(4x^2 + 1)$ and find $F'(-1)$.*

Solution: We have a product $F(x) = f(x)g(x)$ with

$$f(x) = x^3 - 2x + 3 \quad \text{and} \quad g(x) = 4x^2 + 1.$$

The product rule gives

$$\begin{aligned} F'(x) &= f(x)g'(x) + f'(x)g(x) \\ &= (x^3 - 2x + 3)(8x) + (3x^2 - 2)(4x^2 + 1) \\ &= 20x^4 - 21x^2 + 24x - 2 \end{aligned}$$

Setting $x = -1$, we have

$$F'(-1) = -27.$$

□

We come now to reciprocals.

Theorem 6.7 (The reciprocal rule) *If g is differentiable at x and $g(x) \neq 0$, then $1/g$ is differentiable at x and*

$$\left(\frac{1}{g}\right)'(x) = \frac{g'(x)}{(g(x))^2}.$$

Proof. Since g is differentiable at x, g is continuous at x. Since $g(x) \neq 0$, we know that $1/g$ is continuous at x and thus that

$$\lim_{h \to 0} \frac{1}{g(x+h)} = \frac{1}{g(x)}.$$

For h different from 0 and sufficiently small, $g(x+h) \neq 0$. The continuity of g at x and the fact that $g(x) \neq 0$ guarantee this (see Example 4.6). The difference quotient for $1/g$ can be written

$$\begin{aligned}
\frac{1}{h}\left(\frac{1}{g(x+h)} - \frac{1}{g(x)}\right) &= \frac{1}{h}\left(\frac{g(x) - g(x+h)}{g(x)g(x+h)}\right) \\
&= \left(\frac{g(x) - g(x+h)}{h}\right)\frac{1}{g(x)g(x+h)}.
\end{aligned}$$

As h tends to zero, the right-hand side (and thus the left) tends to

$$\frac{g'(x)}{(g(x))^2}.$$

■

Using the reciprocal rule, we can show that Formula (6.6) also holds for negative integers:

$$\boxed{\begin{array}{l}\text{for each negative integer } n \\ \qquad p(x) = x^n \text{ has derivative } p'(x) = nx^{n-1}.\end{array}} \tag{6.8}$$

This formula holds at all x except, of course, at $x = 0$, where no negative power is even defined. In particular, for $x \neq 0$,

$$\begin{array}{lll}
p(x) = x^{-1} & \text{has derivative} & p'(x) = (-1)x^{-2} = -x^{-2}, \\
p(x) = x^{-2} & \text{has derivative} & p'(x) = -2x^{-3}, \\
p(x) = x^{-3} & \text{has derivative} & p'(x) = -3x^{-4}, \\
p(x) = x^{-4} & \text{has derivative} & p'(x) = -4x^{-5},
\end{array}$$

and so on.

Proof. (of 6.8) Note that

$$p(x) = \frac{1}{g(x)} \quad \text{where} \quad g(x) = x^{-n} \quad \text{and } -n \text{ is a positive integer.}$$

The rule for reciprocals gives

$$\begin{aligned} p'(x) &= \frac{g'(x)}{(g(x))^2} = \frac{(-nx^{-n-1})}{x^{-2n}} \\ &= (nx^{-n-1})x^{2n} = nx^{n-1}. \end{aligned}$$

■

Example 6.8 *Differentiate* $f(x) = \frac{4}{x^3} + \frac{1}{x}$*, and find* $f'(1/2)$.

Solution: To apply (6.8), we write

$$f(x) = 4x^{-3} + x^{-1}$$

Differentiation gives

$$f'(x) = -12x^{-4} - x^{-2}$$

Back in fractional notation,

$$f'(x) = \frac{-12}{x^4} - \frac{1}{x^2}.$$

Setting $x = \frac{1}{2}$, we have

$$f'(\frac{1}{2}) = -196$$

□

Example 6.9 *Find the area of the triangle formed from the coordinate axes and the tangent line to the curve* $y = \frac{1}{x}$ *at the point* $(x_0, y_0) = (x_0, \frac{1}{x_0})$*, where* $x_0 > 0$.

Solution: First we write the equation for the tangent line at the point(x_0, y_0) using the equation for a line,

$$y - y_0 = f'(x_0)(x - x_0).$$

Plug in $y_0 = f(x_0) = \frac{1}{x_0}$, $f'(x_0) = -\frac{1}{x_0^2}$ to get:

$$y - \frac{1}{x_0} = -\frac{1}{x_0^2}(x - x_0)$$

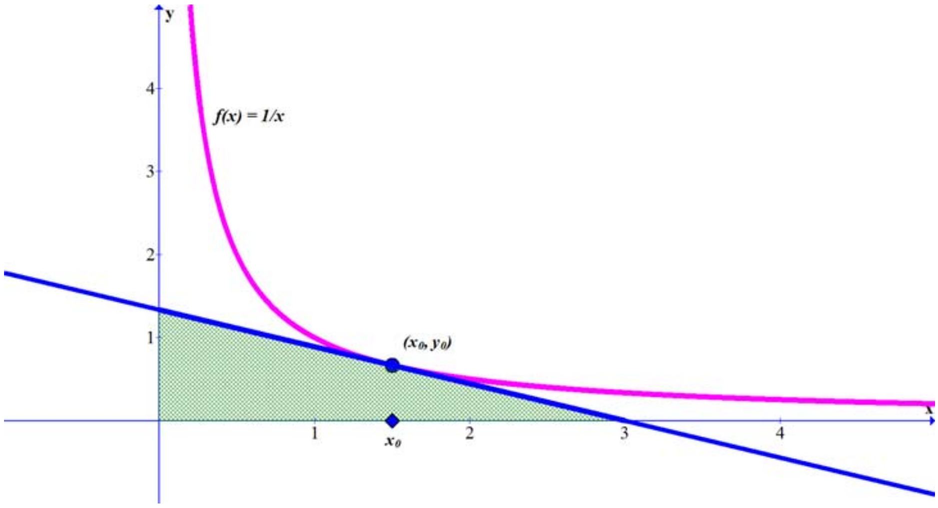


Fig. 6.2. The area of the triangle is always 2, no matter where on the graph we draw the tangent line.

Next calculate the x-intercept of this tangent line. The x-intercept is where $y = 0$. Plug $y = 0$ into the equation for this tangent line to get:

$$\begin{aligned} 0 - \frac{1}{x_0} &= -\frac{1}{x_0^2}(x - x_0) \\ -\frac{1}{x_0} &= \frac{1}{x_0} - \frac{x}{x_0^2} \\ \frac{x}{x_0^2} &= \frac{2}{x_0} \\ x &= 2x_0 \end{aligned}$$

So, the x-intercept of this tangent line is at $x = 2x_0$. Next we claim that the y-intercept is at $y = 2y_0$. Since $y = \frac{1}{x}$ and $x = \frac{1}{y}$ are identical equations, the graph is symmetric when x and y are exchanged. By symmetry, then, the y-intercept is at $y = 2y_0$. If you don't trust reasoning with symmetry, you may follow the same chain of algebraic reasoning that we used in finding the x-intercept. (Remember, the y-intercept is where $x = 0$.)

Finally,

$$\text{Area} = \frac{1}{2}(2y_0)(2x_0) = 2x_0y_0 = 2 \quad \text{(see Figure 6.2).}$$

Curiously, the area of the triangle is always 2, no matter where on the graph we draw the tangent line. □

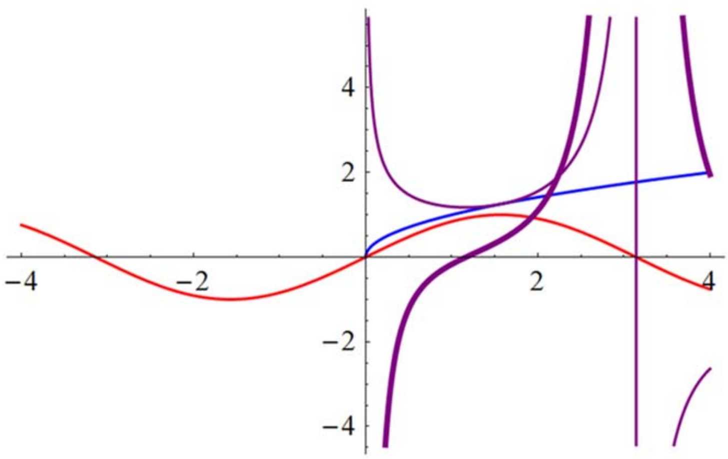


Fig. 6.3. If f and g are both differentiable at x and $g(x) \neq 0$, then the derivative of their quotient at x is given by the quotient rule. On the graph, $f(x) = \sqrt{x}$ is blue, $g(x)$ is red, $f(x)/g(x)$ is purple, and the derivative of $f(x)/g(x)$ i.e. $\frac{1}{2\sqrt{x}\sin^2 x}(\sin x - 2x \cos x)$ is thick purple.

Finally we come to quotients in general. Suppose we want to differentiate a function of the form $Q(x) = f(x)/g(x)$. (Of course, we have to avoid points where $g(x) = 0$.) We want a formula for Q' in terms of f' and g'. Notice that the derivative of the quotient is *not equal* to the quotient of the derivatives (give an example!). Instead, we have the following rule:

Theorem 6.10 (The quotient rule) *If f and g are differentiable at x and $g(x) = 0$, then the quotient f/g is differentiable at x and*

$$\left(\frac{f}{g}\right)'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{(g(x))^2}. \qquad (6.9)$$

"*The derivative of a quotient is the denominator times the derivative of the numerator minus the numerator times the derivative of the denominator, all divided by the square of the denominator.*"

Since $f/g = f \cdot (1/g)$, the quotient rule can be obtained from the product and reciprocal rules. The proof of the quotient rule is left to you as an exercise. Finally, note that the reciprocal rule is just a special case of the quotient rule. (Take $f(x) = 1$.)

$\frac{d}{dx}\left(\frac{\sqrt{x}}{\sin x}\right) = \frac{1}{2\sqrt{x}\sin^2 x}(\sin x - 2x \cos x)$

Remark 6.11 *From the quotient rule you can see that all* rational functions *(quotients of polynomials) are differentiable wherever they are defined.*

Example 6.12 *Differentiate* $F(x) = \frac{6x^2-1}{x^4+5x+1}$.

Solution: Here we are dealing with a quotient $F(x) = f(x)/g(x)$. The quotient rule, gives

$$\begin{aligned} F'(x) &= \frac{(x^4+5x+1)(12x) - \left(6x^2-1\right)\left(4x^3+5\right))}{(x^4+5x+1)^2} \\ &= \frac{\left(-12x^5+4x^3+30x^2+12x+5\right)}{(x^4+5x+1)^2} \end{aligned}$$

□

Example 6.13 *Find the points on the graph of*

$$f(x) = \frac{4x}{x^2+4}$$

where the tangent line is horizontal.

Solution: The rational function f is well defined for each x, so it is differentiable everywhere. The quotient rule gives

$$f'(x) = \frac{(x^2+4)(4) - 4x(2x)}{(x^2+4)^2} = \frac{16-4x^2}{(x^2+4)^2}$$

The tangent line is horizontal only at the points $(x, f(x))$ where $f'(x) = 0$. Therefore, we set $f'(x) = 0$ and solve for x:

$$\frac{16-4x^2}{(x^2+4)^2} \quad \text{iff} \quad 16-4x^2 = 0 \quad \text{iff} \quad x = \pm 2.$$

The tangent line is horizontal at the points where $x = -2$ or $x = 2$. These are the points $(-2, f(-2)) = (-2, -1)$ and $(2, f(2)) = (2, 1)$. See Figure 6.4.

6.2 Derivatives of higher order

Higher derivatives are derivatives of derivatives. For instance, if $g = f'$, then $h = g'$ is the second derivative of f. We write $h = (f')' = f''$. Different

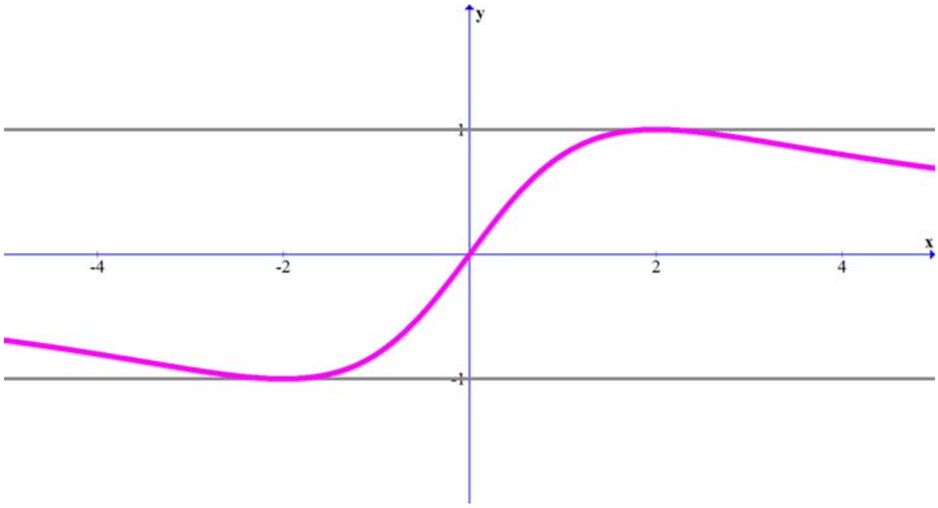


Fig. 6.4. The graph of $f(x) = \frac{4x}{x^2+4}$ with two horizontal tangent lines.

notations used:

$f'(x)$	Df	$\frac{df}{dx}$
$f''(x)$	$D^2 f$	$\frac{d^2 f}{dx^2}$
$f'''(x)$	$D^3 f$	$\frac{d^3 f}{dx^3}$
$f^{(n)}(x)$	$D^n f$	$\frac{d^{(n)} f}{dx^n}$

Example 6.14 *If* $f(x) = 3x^4 - 2x^3 + x^2 - 4x + 2,$ *then*

$$\begin{aligned} f'(x) &= 12x^3 - 6x^2 + 2x - 4, \\ f''(x) &= 36x^2 - 12x + 2 \\ f'''(x) &= 72x - 12 \\ f^{(4)}(x) &= 72 \\ f^{(5)}(x) &= 0 \\ f^{(n)}(x) &= 0 \quad n \geq 5 \quad \square \end{aligned}$$

We will discuss the significance of second derivatives and those of higher order in later sections.

6.3 Derivatives of trigonometric functions

From variations in market trends and ocean temperatures to daily fluctuations in tides and hormone levels, change is often cyclical or periodic. Trigonometric functions are well suited for describing such cyclical behavior. In this section, we investigate the derivatives of trigonometric functions and their many uses. Beware, that the results stated in this section assume that angles are measured in radians. Our principal goal is to determine derivative formulas for $\sin x$ and $\cos x$. In order to do this, we use two special trigonometric limits (found earlier)

$$\lim_{x\to 0} \frac{\sin x}{x} = 1, \qquad \lim_{x\to 0} \frac{\cos x - 1}{x} = 0. \tag{6.10}$$

6.3.1 Derivatives of sine and cosine functions

With the trigonometric limits 6.10, the derivative of the sine function can be found. We start with the definition of the derivative,

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

with $f(x) = \sin x$, and then appeal to the sine addition identity

$$\sin(x+h) = \sin x \cos h + \cos x \sin h.$$

The derivative is

$f'(x) = \lim_{h\to 0} \dfrac{\sin(x+h) - \sin(x)}{h}$	Definition of derivative.
$= \lim_{h\to 0} \dfrac{\sin x \cos h + \cos x \sin h - \sin(x)}{h}$	Sine addition identity.
$= \lim_{h\to 0} \dfrac{\sin x(\cos h - 1) + \cos x \sin h}{h}$	Factor $\sin x$.
$= \lim_{h\to 0} \dfrac{\sin x(\cos h - 1)}{h} + \lim_{h\to 0} \dfrac{\cos x \sin h}{h}$	Relation 4.12.
$= \sin x \lim_{h\to 0} \dfrac{(\cos h - 1)}{h} + \cos x \lim_{h\to 0} \dfrac{\sin h}{h}$	Both $\sin x$ and $\cos x$ are independent of h.
$= (\sin x)\,(0) + \cos x\,(1)$	Limits from 6.10.
$= \cos x$	Simplify.

We have proved the important result that

$$\boxed{\frac{d}{dx} \sin x = \cos x.} \tag{6.11}$$

The fact that

$$\boxed{\frac{d}{dx} \cos x = -\sin x} \tag{6.12}$$

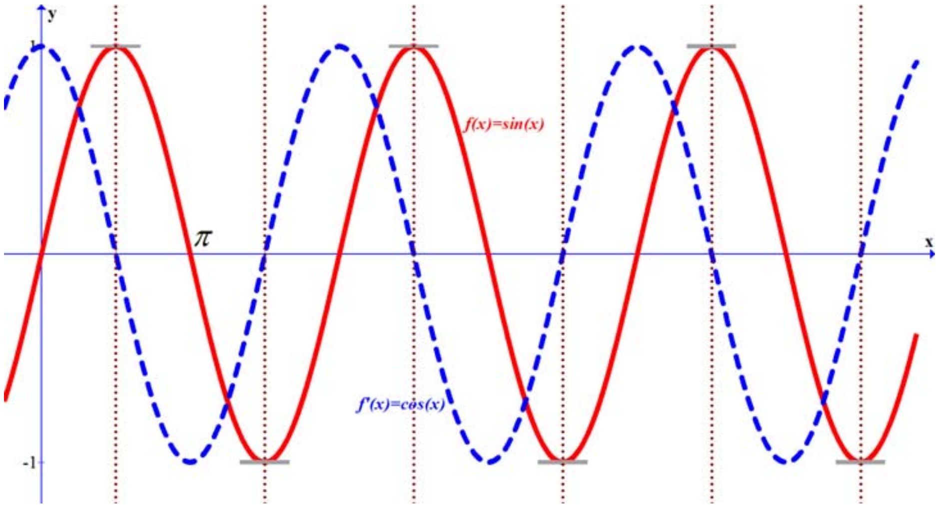


Fig. 6.5. The horizontal tangent lines on the graph of $f(x) = \sin x$ occur at the zeros of $f(x) = \cos x$.

is proved in a similar way using a cosine addition identity

$$\cos(x + h) = \cos x \cos h - \sin x \sin h.$$

From a geometric point of view, these derivative formulas make sense. Because $f(x) = \sin x$ is a periodic function, we expect its derivative to be periodic. Observe that the horizontal tangent lines on the graph of $f(x) = \sin x$ occur at the zeros of $f(x) = \cos x$.

Similarly, the horizontal tangent lines on the graph of $f(x) = \cos x$ occur at the zeros of $f(x) = -\sin x$. (Figure 6.5).

Example 6.15 (Derivatives involving trigonometric functions) *Calculate $f'(x)$ for the following functions.*

$$\textbf{a}.\ f(x){=}\sin x - x^2 \cos x \qquad \textbf{b}.\ f(x) = \frac{1 + \sin x}{1 - \sin x}.$$

Solution:

a. $f'(x) = \cos x - (2x \cos x - x^2 \sin x) = \cos x + x^2 \sin x - 2x \cos x.$

b. $f'(x) = \dfrac{(1 - \sin x)\cos x - (1 + \sin x)(-\cos x)}{(1 - \sin x)^2} = \frac{2\cos x}{(1-\sin x)^2}$

□

6.3.2 Derivatives of other trigonometric functions

The derivatives of $\tan x$, $\cot x$, $\sec x$, and $\csc x$ are obtained using the derivatives of $\sin x$ and $\cos x$ together with the quotient rule and trigonometric identities. Recall that

$$\tan x = \frac{\sin x}{\cos x}, \quad \cot x = \frac{\cos x}{\sin x}, \quad \sec x = \frac{1}{\cos x} \quad \text{and} \quad \csc x = \frac{1}{\sin x}.$$

Example 6.16 (Derivative of the tangent function) *Calculate* $\frac{d}{dx}(\tan x)$.

Solution: Using the identity $\tan x = \frac{\sin x}{\cos x}$ and the quotient rule, we have

$$\begin{aligned}
&\frac{d}{dx}(\tan x) \\
&= \frac{d}{dx}\left(\frac{\sin x}{\cos x}\right) \\
&= \frac{\cos x \cos x - \sin x(-\sin x)}{(\cos x)^2} && \text{Quotient rule.} \\
&= \frac{\cos^2 x + \sin^2 x}{\cos^2 x} && \text{Simplify numerator.} \\
&= \frac{1}{\cos^2 x} = \sec^2 x && \cos^2 x + \sin^2 x = 1.
\end{aligned}$$

Therefore, $\frac{d}{dx}(\tan x) = \sec^2 x$. □

The derivatives of $\cot x$, $\sec x$, and $\csc x$ are given in Theorem 6.17. The formulas for $\frac{d}{dx}(\cot x)$, $\frac{d}{dx}(\sec x)$, $\frac{d}{dx}(\csc x)$ can be determined using the quotient rule.

Theorem 6.17 (Derivatives of the trigonometric functions)

$\frac{d}{dx}(\sin x) = \cos x$	$\frac{d}{dx}(\cos x) = -\sin x$
$\frac{d}{dx}(\tan x) = \sec^2 x = \frac{1}{\cos^2 x}$	$\frac{d}{dx}(\cot x) = -\csc^2 x = \frac{-1}{\sin^2 x}$
$\frac{d}{dx}(\sec x) = \sec x \tan x$	$\frac{d}{dx}(\csc x) = -\csc x \cot x$

Higher-order derivatives of the sine and cosine functions are important in many applications. A few higher-order derivatives of $y = \sin x$ reveal a pattern.

$$\frac{dy}{dx} = \cos x \qquad \frac{d^2y}{dx^2} = \frac{d}{dx}(\cos x) = -\sin x$$

$$\frac{d^3y}{dx^3} = \frac{d}{dx}(-\sin x) = -\cos x \qquad \frac{d^4y}{dx^4} = \frac{d}{dx}(-\cos x) = \sin x$$

We see that the higher-order derivatives of $\sin x$ cycle back periodically to $\pm\sin x$. In general, it can be shown that

$$\frac{d^{(2n)}y}{dx^{(2n)}} = (-1)^n \sin x,$$

with a similar result for $\cos x$. This cyclic behavior in the derivatives of $\sin x$ and $\cos x$ does not occur with the other trigonometric functions.

6.4 Chain rule

Composition is an important way of constructing new functions. The composition of f and g is the function $f \circ g$ defined by $(f \circ g)(x) = f(g(x))$. The domain of $f \circ g$ is the set of values of x in the domain of g such that $g(x)$ lies in the domain of f .

Example 6.18 *Compute the composite functions $f \circ g$ and $g \circ f$ and discuss their domains, where*

$$f(x) = \sqrt{x}, \quad g(x) = 1 - x.$$

Solution: We have

$$(f \circ g)(x) = f(g(x)) = f(1 - x) = \sqrt{1 - x}.$$

The square root $\sqrt{1 - x}$ is defined if $1 - x \geq 0$ or $x \leq 1$, so the domain of $f \circ g$ is $\{x : x \leq 1\}$. On the other hand,

$$(g \circ f)(x) = g(f(x)) = g(\sqrt{x}) = 1 - \sqrt{x}.$$

The domain of $g \circ f$ is $\{x : x \geq 0\}$. □

Remark 6.19 *Example 6.18 shows that the composition of functions is not commutative: The functions $f \circ g$ and $g \circ f$ may be (and usually are) different.*

Now suppose we know, that the functions f and g are differentiable. What about differentiability of $f \circ g$? How do we know that the composition of differentiable functions is differentiable? What assumptions do we need to compute $(f \circ g)'(x)$? If we know the derivatives of f and g, how can we use this information to find the derivative of the composition $f \circ g$? The following theorem provides the definitive answer.

Theorem 6.20 (The chain-rule theorem) *If g is differentiable at x and f is differentiable at $g(x)$, then the composition $f \circ g$ is differentiable at x and*

$$(f \circ g)'(x) = f'(g(x))g'(x). \tag{6.13}$$

"the derivative of a composition $f \circ g$ at x is the derivative of f at $g(x)$ times the derivative of g at x."

Example 6.21 *Suppose $h(x) = (x^2+1)^{99}$ and you want to find $h'(x)$. It would be ridiculous to multiply it out–you'd have to multiply $x^2 + 1$ by itself 99 times and it would take days. It would also be crazy to use or binomial theorem or the product rule (since you'd need to use it too many times). Instead, let's view h as the composition of two functions f and g, where $g(x) = x^2 + 1$ and $f(x) = x^{99}$. Indeed, if you take your x and hit it with g, you end up with x^2+1. If you now hit that with f, you get $(x^2 + 1)^{99}$, which is just $h(x)$. So we have written $h(x)$ as $f(g(x))$. Now, we have $f(x) = x^{99}$, so $f'(x) = 99x^{98}$. We also have $g(x) = x^2 + 1$, so $g'(x) = 2x$. There's our second factor: just $2x$. How about the first one? Well, we take $f'(x)$, but instead of x, we put in $x^2 + 1$ (since that's what $g(x)$ is). That is, $f'(g(x)) = f'(x^2+1) = 99(x^2+1)^{98}$. Now we multiply our two factors together to get*

$$h'(x) = f'(g(x))g'(x) = 99(x^2 + 1)^{98}(2x) = 198x(x^2 + 1)^{98}$$

□

Example 6.22 *Find $f'(x)$ if*

$$f(x) = \frac{1}{2x^4 - x^2 + 8}$$

Solution: First write

$$f(x) = \left(2x^4 - x^2 + 8\right)^{-1}$$

so the *inside function* is $2x^4 - x^2 + 8$ and the outside function is x^{-1}. Thus (according to 6.13)

$$f'(x) = -\left(2x^4 - x^2 + 8\right)^{-2}(8x^3 - 2x) = -\frac{8x^3 - 2x}{(2x^4 - x^2 + 8)^2}$$

□

Next we will prove the chain rule (Theorem 6.20), but first we need a preliminary result.

Theorem 6.23 *If f is differentiable at x and if $y = f(x)$, then*

$$\Delta y = f'(x)\Delta x + \varepsilon \Delta x$$

where $\Delta y = f(x + \Delta x) - f(x)$, $\varepsilon \to 0$ as $\Delta x \to 0$ and $\varepsilon = 0$ if $\Delta x = 0$.

Proof. Define

$$\varepsilon = \begin{cases} \frac{f(x+\Delta x)-f(x)}{\Delta x} & \text{if} \quad \Delta x \neq 0 \\ 0 & \text{if} \quad \Delta x = 0 \end{cases} \tag{6.14}$$

If $\Delta x \neq 0$, it follows from (6.14) that

$$\varepsilon \Delta x = [f(x + \Delta x) - f(x)] - f'(x)\Delta x \tag{6.15}$$

But,

$$\Delta y = f(x + \Delta x) - f(x)$$

so (6.15) can be written as

$$\varepsilon \Delta x = \Delta y - f'(x)\Delta x. \tag{6.16}$$

If $\Delta x = 0$, then (6.16) still holds (why?), so (6.16) is valid for all values of Δx. It remains to show that $\varepsilon \to 0$ as $\Delta x \to 0$. But this follows from the assumption that f is differentiable at x, since

$$\lim_{\Delta x \to 0} \varepsilon = \lim_{\Delta x \to 0} \left[\frac{f(x + \Delta x) - f(x)}{\Delta x} - f'(x)\right] = f'(x) - f'(x) = 0.$$

■

We are now ready to prove the chain rule.

Proof. Since g is differentiable at x and $u = g(x)$, it follows from Theorem 6.23 that

$$\Delta u = g'(x)\Delta x + \varepsilon_1 \Delta x \tag{6.17}$$

where $\varepsilon_1 \to 0$ as $\Delta x \to 0$. And since $y = f(u)$ is differentiable at $u = g(x)$, it follows from Theorem 6.23 that

$$\Delta y = f'(u)\Delta u + \varepsilon_2 \Delta u \tag{6.18}$$

where $\varepsilon_2 \to 0$ as $\Delta u \to 0$.

Factoring out the Δu in (6.18) and then substituting (6.17) yields

$$\Delta y = \left[f'(u) + \varepsilon_2\right] \left[g'(x)\Delta x + \varepsilon_1 \Delta x\right]$$

or

$$\Delta y = \left[f'(u) + \varepsilon_2\right] \left[g'(x) + \varepsilon_1\right] \Delta x$$

or if $\Delta x \neq 0$,

$$\frac{\Delta y}{\Delta x} = \left[f'(u) + \varepsilon_2\right] \left[g'(x) + \varepsilon_1\right] \tag{6.19}$$

But (6.17) implies that $\Delta u \to 0$ as $\Delta x \to 0$, and hence $\varepsilon_1 \to 0$ and $\varepsilon_2 \to 0$ as $\Delta x \to 0$. Thus, from (6.19)

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = f'(u)g'(x)$$

or

$$\frac{dy}{dx} = f'(u)g'(x) = \frac{dy}{du} \cdot \frac{du}{dx}$$

■

The chain rule can actually be invoked multiple times all at once. For example, let

$$y = ((x^3 - 10x)^9 + 22)^8$$

What is dy/dx? Simply let $u = x^3 - 10x$, and $v = u^9 + 22$, so that $y = v^8$. Then use a longer form of the chain rule:

$$\frac{dy}{dx} = \frac{dy}{dv}\frac{dv}{du}\frac{du}{dx}.$$

You can't get this wrong if you think about it: y is a function of v, which is a function of u, which is a function of x. So there's only one way the formula could possibly look! Anyway, we have

$$y = v^8 \qquad v = u^9 + 22 \qquad u = x^3 - 10x$$

$$\frac{dy}{dv} = 8v^7 \qquad \frac{dv}{du} = 9u^8 \qquad \frac{du}{dx} = 3x^3 - 10$$

Plugging everything in, we have

$$\frac{dy}{dx} = \frac{dy}{dv}\frac{dv}{du}\frac{du}{dx} = \left(8v^7\right)\left(9u^8\right)\left(3x^3 - 10\right)$$

We're close, but we need to get rid of the u and v terms. First, replace v by $u^9 + 22$

$$\begin{aligned} \frac{dy}{dx} &= \left(8v^7\right)\left(9u^8\right)\left(3x^3 - 10\right) \\ &= \left(8\left(u^9 + 22\right)^7\right)\left(9u^8\right)\left(3x^3 - 10\right). \end{aligned}$$

Now replace u by $x^3 - 10x$ and group the factors of 8 and 9 together to get the actual answer:

$$\begin{aligned} \frac{dy}{dx} &= \left(8\left(u^9 + 22\right)^7\right)\left(9u^8\right)\left(3x^3 - 10\right) \\ &= \left(8\left(\left(x^3 - 10x\right)^9 + 22\right)^7\right)\left(9\left(x^3 - 10x\right)^8\right)\left(3x^3 - 10\right) \\ &= 72\left(\left(x^3 - 10x\right)^9 + 22\right)^7\left(x^3 - 10x\right)^8\left(3x^3 - 10\right) \end{aligned}$$

The name “chain rule” is appropriate because the desired derivative is obtained by a two-link “chain” of simpler derivatives. But there are different explonations. One very well known professor said: "You know, people often wonder where the name chain rule comes from. I was just wondering about that myself. So is it because it chains you down? Is it like a chain fence? I decided what it is. It's because by using it, you burst the chains of differentiation, and you can differentiate many more functions using it. So when you want to think of the chain rule, just think of that chain there. It lets you burst free."[2]

6.5 Implicit differentiation; rational powers

6.5.1 Functions defined explicitly and implicitly

An equation of the form $y = f(x)$ is said to define y explicitly as a function of x because the variable y appears alone on one side of the equation and does

[2] Citation from: David Jerison, 18.01 Single Variable Calculus, Fall 2007. (Massachusetts Institute of Technology: MIT OpenCourseWare). http://ocw.mit.edu (accessed 07 11, 2011). License: Creative Commons Attribution-Noncommercial-Share Alike.

not appear at all on the other side. However, sometimes functions are defined by equations in which y is not alone on one side; for example, the equation

$$yx + y + 1 = x \tag{6.20}$$

is not of the form $y = f(x)$, but it still defines y as a function of x since it can be rewritten as

$$y = \frac{x-1}{x+1}$$

Thus, we say that (6.20) defines y *implicitly* as a function of x, the function being

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

An equation in x and y can implicitly define more than one function of x. This can occur when the graph of the equation fails the vertical line test, (XXX dotychczas niezdefiniowane VLT) so it is not the graph of a function of x. For example, if we solve the equation of the circle

$$x^2 + y^2 = 1 \tag{6.21}$$

for y in terms of x, we obtain $y = \pm\sqrt{1-x^2}$, so we have found two functions that are defined implicitly by (6.21), namely

$$f_1(x) = +\sqrt{1-x^2} \quad \text{and} \quad f_2(x) = -\sqrt{1-x^2}.$$

The graphs of these functions are the upper and lower semicircles of the circle $x^2 + y^2 = 1$. This leads us to the following definition.

Definition 6.24 *We will say that a given equation in x and y defines the function f* implicitly *if the graph of $y = f(x)$ coincides with a portion of the graph of the equation.*

Although it was a trivial matter in the last example to solve the equation $x^2 + y^2 = 1$ for y in terms of x, it is difficult or impossible to do this for some equations. For example the problem with $y^5 + xy = 3$ is that it can't be solved for y. Galois proved[3] that there is no solution formula for fifth-degree equations. The function $y(x)$ cannot be given explicitly in this case. Thus, even though an equation may define one or more functions of x, it may not be possible or practical to find explicit formulas for those functions.

[3] That was before he went to the famous duel, and met his end. Fourth-degree equations do have a solution formula, but it is practically never used.

6.5.2 *Implicit differentiation*

Up to this point we have been differentiating functions defined explicitly in terms of an independent variable. We can also differentiate functions not explicitly given in terms of an independent variable. Suppose we know that y is a differentiable function of x and satisfies a particular equation in x and y. If we find it difficult to obtain the derivative of y, either because the calculations are burdensome or because we are unable to express y explicitly in terms of x, we may still be able to obtain dy/dx by a process called *implicit differentiation.* This process is based on differentiating both sides of the equation satisfied by x and y. To illustrate this, let us consider the simple equation

$$xy = 1. \tag{6.22}$$

One way to find dy/dx is to rewrite this equation as

$$y = \frac{1}{x} \tag{6.23}$$

from which it follows that

$$\frac{dy}{dx} = -\frac{1}{x^2}. \tag{6.24}$$

Another way to obtain this derivative is to differentiate both sides of (6.22) *before* solving for y in terms of x, treating y as a (temporarily unspecified) differentiable function of x. With this approach we obtain

$$\frac{d}{dx}[xy] = \frac{d}{dx}[1]$$

$$x\frac{d}{dx}[y] + y\frac{d}{dx}[x] = 0$$

$$x\frac{dy}{dx} + y = 0$$

$$\frac{dy}{dx} = -\frac{y}{x}$$

If we now substitute (6.23) into the last expression, we obtain

$$\frac{dy}{dx} = -\frac{1}{x^2}$$

which agrees with Equation (6.24). □

Now we can establish general strategy for implicit differentiation:

1. Differentiate both sides of the equation *with respect to x.*

2. Collect all terms involving dy/dx on the left side of the equation and move all other terms to the right side of the equation.

3. Factor dy/dx out of the left side of the equation.

4. Solve for dy/dx

Example 6.25 *Assume that y is a differentiable function of x which satisfies the given equation*

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

Use implicit differentiation to express dy/dx in terms of x and y.

Solution: We have

$$\underbrace{2x^2\frac{dy}{dx} + 4xy}_{\text{(by the product rule)}} - \underbrace{3y^2\frac{dy}{dx}}_{\text{(by the chain rule)}} = 1 + 2\tfrac{dy}{dx}$$

so,

$$(2x^2 - 3y^2 - 2)\frac{dy}{dx} = 1 - 4xy.$$

Therefore

$$\frac{dy}{dx} = \frac{1 - 4xy}{2x^2 - 3y^2 - 2}$$

□

Remark 6.26 *The graph of an equation does not always define a function because there may be more than one y-value for a given value of x. Implicit differentiation works because the graph is generally made up of several pieces called* branches*, each of which does define a function (a proof of this fact relies on the Implicit Function Theorem from advanced calculus). For example, the branches of the unit circle $x^2 + y^2 = 1$ are the graphs of the functions $y = \sqrt{1 - x^2}$ and $y = -\sqrt{1 - x^2}$. In most examples, the branches are differentiable except at certain exceptional points where the tangent line may be vertical.*

Example 6.27 *Figure 6.6 shows the curve $2x^3 + 2y^3 = 9xy$ and the tangent line at the point $(1, 2)$. What is the slope of the tangent line at that point?*

Solution: We want dy/dx where $x = 1$ and $y = 2$. We proceed by implicit differentiation:

$$6x^2 + 6y^2\frac{dy}{dx} = 9x\frac{dy}{dx} + 9y$$

$$2x^2 + 2y^2\frac{dy}{dx} = 3x\frac{dy}{dx} + 3y$$

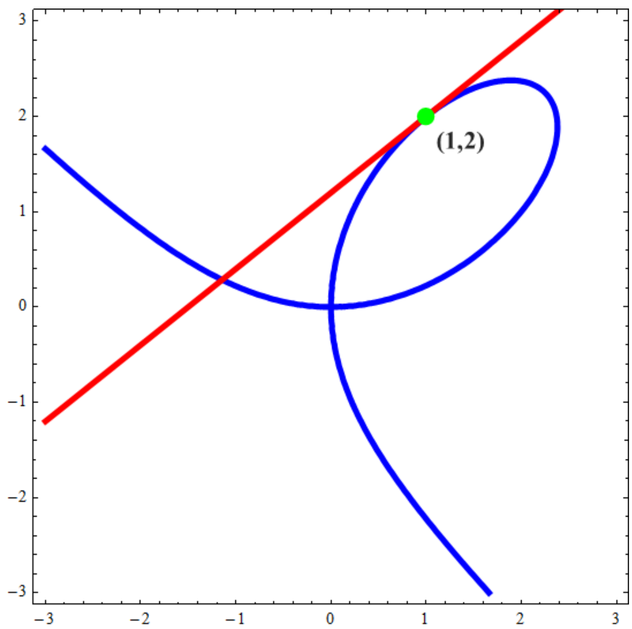


Fig. 6.6. The graph of $2x^3 + 2y^3 = 9xy$ and the tangent line (in red) at $(1, 2)$.

Setting $x = 1$ and $y = 2$, we have

$$2 + 8\frac{dy}{dx} = 3\frac{dy}{dx} + 6$$

$$5\frac{dy}{dx} = 4$$

$$\frac{dy}{dx} = \frac{4}{5}.$$

The slope of the tangent line at the point $(1, 2)$ is $4/5$. □

Example 6.28 *Find the slope of the graph of* $x^3y^2 = xy^3 + 6$ *(see Figure 6.7) at the point* $(2, 1)$.

Solution: Having the equation

$$x^3y^2 = xy^3 + 6$$

we can differentiate both sides of it with respect to x

$$\frac{d}{dx}(x^3y^2) = \frac{d}{dx}\left(xy^3\right) + 0.$$

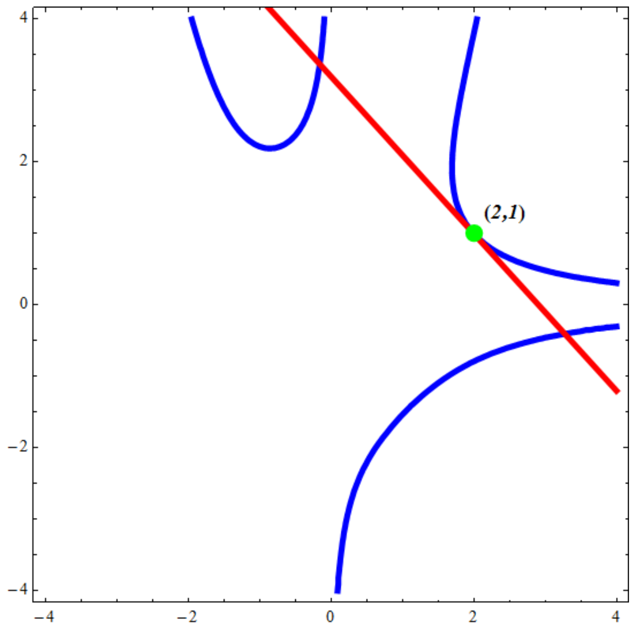


Fig. 6.7. The graph of $x^3y^2 = xy^3 + 6$ and the tangent line (in red) at $(2,1)$.

Using the product rule and the chain rule we get

$$3x^2y^2 + x^3(2y)\frac{dy}{dx} = (1)\,y^3 + x\left(3y^2\right)\frac{dy}{dx}.$$

Now, we want to isolate dy/dx, so we have

$$2x^3y\frac{dy}{dx} - 3xy^2\frac{dy}{dx} = y^3 - 3x^2y^2$$

or, after factoring

$$xy\left(2x^2 - 3y\right)\frac{dy}{dx} = y^2\left(y - 3x^2\right)$$

which gives

$$\frac{dy}{dx} = \frac{y^2\left(y - 3x^2\right)}{xy\left(2x^2 - 3y\right)} = \frac{y\left(y - 3x^2\right)}{x\left(2x^2 - 3y\right)}$$

We want to know the slope at $(2,1)$ so we evaluate

$$\frac{dy}{dx}|_{(2,1)} = \frac{1\left(1 - 3 \cdot 2^2\right)}{2\left(2 \cdot 2^2 - 3 \cdot 1\right)} = -\frac{11}{10}. \qquad \square$$

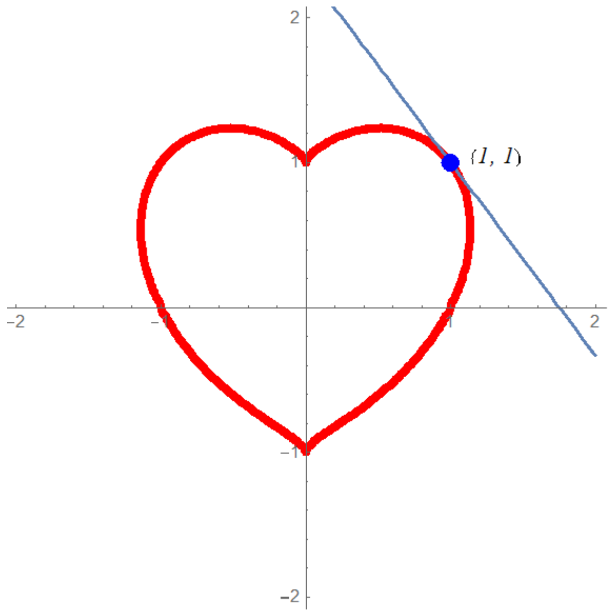


Fig. 6.8. The set of points satisfying the Valentine equation looks like a heart.

Example 6.29 (Application of the chain rule) *The* Valentine equation

$$(x^2 + y^2 - 1)^3 - x^2 y^3 = 0$$

relates x with y, but we can not write the curve as a graph of a function $y = g(x)$. Extracting y or x is difficult. The set of points satisfying the equation looks like a heart. You can check that $(1, 1)$ satisfies the Valentine equation. Near it, the curve looks like the graph of a function $g(x)$. Lets fill that in and look at the function $g(x)$

$$f(x) = (x^2 + (g(x))^2 - 1)^3 - x^2 \left(g(x)\right)^3 .$$

The key is that $f(x)$ is actually zero and if we take the derivative, then we get zero too. Using the chain rule, we can take the derivative

$$f'(x) = 3(x^2+(g(x))^2-1)(2x+2g(x)g'(x)) - 2x \left(g(x)\right)^3 - x^2 3 \left(g(x)\right)^2 g'(x) = 0.$$

We can now solve solve for g'

$$g'(x) = -\frac{3(x^2 + (g(x))^2 - 1)2x - 2x \left(g(x)\right)^3}{3(x^2 + (g(x))^2 - 1)2g(x) - 3x^2 \left(g(x)\right)^2}.$$

Filling in $x = 1$, $g(x) = 1$, we see this is $-4/3$. We have computed the slope of g without knowing g. The tangent line to our cure at $(1,1)$ is

$$y(x) = -\frac{4}{3}x + \frac{7}{3}$$

(see Figure 6.8). □

6.5.3 Rational Powers

You have seen that the formula

$$\frac{d}{dx}\left(x^n\right) = nx^{n-1}$$

holds for all real x if n is a positive integer and for all $x \neq 0$ if n is a negative integer. For $x \neq 0$, we can stretch the formula to $n = 0$ by writing

$$\frac{d}{dx}\left(x^0\right) = \frac{d}{dx}(1) = 0 = 0x^{-1}.$$

The formula can then be extended to all rational exponents p/q:

$$\frac{d}{dx}\left(x^{p/q}\right) = \frac{p}{q}x^{p/q-1} \tag{6.25}$$

and is called "*the rational power rule*". The formula applies to all $x \neq 0$ where $x^{p/q}$ is defined. To justify it we operate under the assumption that the function $y = x^{1/q}$ is differentiable at all x where $x^{1/q}$ is defined.

Proof. From $y = x^{1/q}$ we get

$$y^q = x.$$

Implicit differentiation with respect to x gives

$$qy^{q-1}\frac{dy}{dx} = 1$$

and therefore

$$\frac{dy}{dx} = \frac{1}{q}y^{1-q} = \frac{1}{q}x^{(1-q)/q} = \frac{1}{q}x^{(1/q)-1}.$$

The function $y = x^{p/q}$ is a composite function:

$$y = x^{p/q} = (x^{1/q})^p.$$

Applying the chain rule, we have

$$\frac{dy}{dx} = p(x^{1/q})^{p-1}\frac{d}{dx}(x^{1/q}) = px^{(p-1)/q}\frac{1}{q}x^{(1/q)-1} = \frac{p}{q}x^{\frac{p}{q}-1}$$

as asserted. ■

Here are some simple examples:

$$\frac{d}{dx}\left(x^{2/3}\right) = \frac{2}{3\sqrt[3]{x}}, \qquad \frac{d}{dx}\left(x^{5/2}\right) = \frac{5}{2}x^{\frac{3}{2}}, \qquad \frac{d}{dx}\left(x^{-7/9}\right) = -\frac{7}{9x^{\frac{16}{9}}}.$$

If u is a differentiable function of x, then, by the chain rule

$$\frac{d}{dx}(u^{p/q}) = \frac{p}{q}u^{(p/q)-1}\frac{du}{dx}. \qquad (6.26)$$

The verification of this is left to you. The result holds on every open x-interval where $u^{(p/q)-1}$ is defined.

Example 6.30 *For all real x,*

$$\frac{d}{dx}[(1+x^2)^{1/5}] = \frac{1}{5}(1+x^2)^{-4/5}(2x) = \frac{2}{5}\frac{x}{(x^2+1)^{\frac{4}{5}}}.$$

□

Example 6.31 *Find and simplify $f'(t)$ if $f(t) = \frac{t}{(t^2+1)^{1/3}}$.*

Solution: Using the quotient rule we have

$$f'(t) = \frac{(1)\,(t^2+1)^{1/3} - t\frac{1}{3}(t^2+1)^{-2/3}2t}{(t^2+1)^{2/3}}.$$

To simplify the result, we can multiply the nominator and the denumerator by $3(1+t^2)^{2/3}$ to get

$$\begin{aligned} f'(t) &= \frac{(1)\,(t^2+1)^{1/3} - t\frac{1}{3}(t^2+1)^{-2/3}2t}{(t^2+1)^{2/3}} \cdot \frac{3(1+t^2)^{2/3}}{3(1+t^2)^{2/3}} \\ &= \frac{3(1+t^2) - 2t^2}{(t^2+1)^{1/3}} \\ &= \frac{\left(t^2+3\right)}{3\,(t^2+1)^{\frac{4}{3}}}. \end{aligned}$$

□

Example 6.32 *Find $\frac{dy}{dx}$ if $2x^{5/2} + 7y^{2/7} = 9xy$, and compute the slope of the graph at $(1,1)$.*

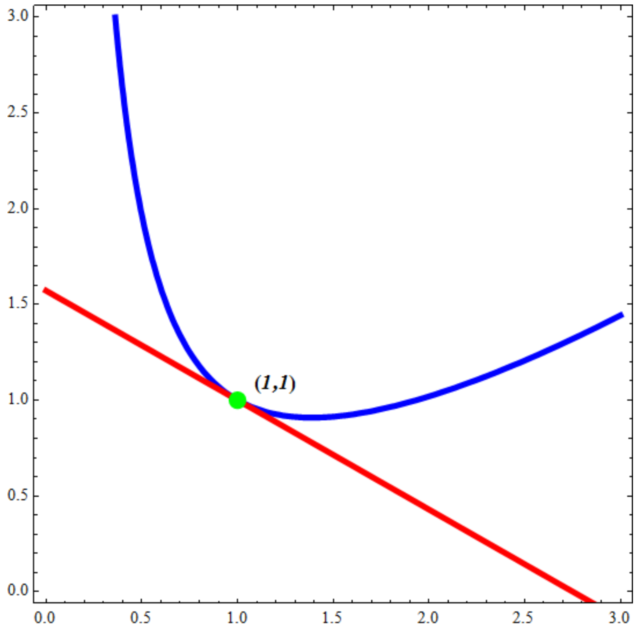


Fig. 6.9. The graph of $2x^{5/2} + 7y^{2/7} = 9xy$ and the tangent line (in red) at $(1, 1)$.

Solution: Figure 6.9 presents the graph of

$$2x^{5/2} + 7y^{2/7} = 9xy \tag{6.27}$$

and the tangent line (in red) at the indicated point. Implicit differentiation with respect to x, the rational power rule (6.25) and the product rule give

$$5x^{3/2} + 2y^{-5/7}\frac{dy}{dx} = 9\left((1)\,y + x\frac{dy}{dx}\right)$$

Now, we want to isolate dy/dx, so we have

$$2y^{-5/7}\frac{dy}{dx} - 9x\frac{dy}{dx} = 9y - 5x^{3/2}.$$

After factoring we get

$$\left(2y^{-5/7} - 9x\right)\frac{dy}{dx} = 9y - 5x^{3/2}$$

or

$$\frac{dy}{dx} = \frac{9y - 5x^{3/2}}{2y^{-5/7} - 9x} = y^{\frac{5}{7}}\frac{5x^{\frac{3}{2}} - 9y}{9xy^{\frac{5}{7}} - 2}.$$

To obtain the required slope, we evaluate

$$\frac{dy}{dx}|_{(1,1)} = \frac{5-9}{9-2} = -\frac{4}{7}$$

□

6.6 Review exercises: Chapter

7

The derivative and graphs

We have seen how to differentiate functions from several different families. Now we can use this knowledge to help us sketch graphs of functions in general. Although graphing utilities are useful for determining the general shape of a graph, many problems require more precision than graphing utilities are capable of producing. The purpose of this section is to develop mathematical tools that can be used to determine the exact shape of a graph and the precise locations of its key features. We'll see how the derivative helps us understand the maxima and minima of functions, and how the second derivative helps us to understand the so-called concavity of functions.

7.1 Extrema of functions

If we say that $x = a$ is an *extremum* of a function f, this means that f has a maximum or minimum at $x = a$. (The plural of "extremum" is "extrema.") We've already looked a little bit at maxima and minima in Section 5.3 of Chapter 5; I strongly suggest taking a peek back at that before you read on. In any event, we need to go a little deeper and distinguish between two types of extrema: global and local.

7.1.1 Global and local extrema

The basic idea of a maximum is that it occurs when the function value is highest. Think about where the maximum of the following function on its domain $[0, 7]$ should be (Figure 7.1):

Certainly the maximum value that this function gets to is 3, which occurs when $x = 0$, so it's true that the function has a maximum at $x = 0$. On the other hand, imagine the graph is a hill (in cross-section) and you're climbing up it. Suppose you start at the point $(2, -1)$ and walk up the hill to the right. Eventually you reach the peak at $(5, 2)$, and then you start going back down again. It sure feels as if the peak is some sort of maximum–it's the top of the mountain, at height 2, even though there's a neighboring peak to the left that's taller. If the high ground near $x = 0$ were covered in fog, you couldn't

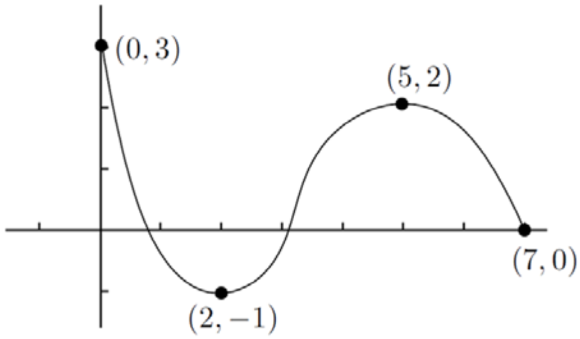


Fig. 7.1. Where the maximum of the following function on its domain $[0; 7]$ should be?

even see it when you climbed the peak at $(5, 2)$, so you'd really feel as if you were at a maximum. In fact, if we restrict the domain to $[2, 7]$, then the point $x = 5$ is *actually* a maximum.

We need a way of clarifying the situation. Let's say that a *global maximum* (or *absolute maximum*) occurs at $x = a$ if $f(a)$ is the highest value of f on the **entire** domain of f. In symbols, we want $f(a) \geq f(x)$ for any value $x$in the domain of f. This is exactly the same definition we used before when we looked at maxima in general; we're simply being more precise and saying "global maxima" instead of just "maxima."

As we noted before, there could be multiple global maxima; for example, $\cos(x)$ has a maximum value of 1, but this occurs for infinitely many values of x on $(-\infty, \infty)$. (These values are all the integer multiples of 2π, as you can see from the graph of $y = \cos(x)$.)

How about that other type of maximum? Let's say that a *local maximum* (or relative maximum) occurs at $x = a$ if $f(a)$ is the highest value of f on *some small open interval containing* a. You can think of this as throwing away most of the domain, just concentrating on values of x close to a (which belong to the domain of f), then insisting that the function is at its maximum out of only those values.

Let's see how this works in the case of our above graph. We see that $x = 5$ is a local maximum, since $(5, 2)$ is the highest point around if you only concentrate on the function near $x = 5$. For example, if you cover up the part of the graph to the left of $x = 3$, then the point $(5, 2)$ is the highest point remaining. On the other hand, $x = 5$ isn't a global maximum, since the point $(0, 3)$ is higher up. This means that $x = 0$ is a global maximum. It's also a local maximum;

in fact, it's pretty obvious that **every global maximum is also a local maximum**.

In the same way, we can define *global* and *local minima*. In the above graph, you can see that $x = 2$ is a global minimum (with value -1), since the height is at its lowest. On the other hand, $x = 7$ is actually a local minimum (with value 0). Indeed, if you just look at the function to the right of $x = 5$, you can see that the lowest height occurs at the endpoint $x = 7$.

7.1.2 The Extreme Value Theorem (EVT)

In Chapter 5, we looked at the max-min theorem. This says that a continuous function on a closed interval $[a, b]$ must have a global maximum somewhere in the interval and also a global minimum somewhere in the interval.

We also saw that if the function isn't continuous, or even if it is continuous but the domain isn't a closed interval, then there might not be a global maximum or minimum. For example, the function f given by $f(x) = 1/x$ on the domain $[-1, 1]\backslash\{0\}$ doesn't have a global maximum or minimum on that domain. (Draw it and see why!)

The problem with the max-min theorem is that it doesn't tell you anything about where these global maxima and minima are. That's where the derivative comes in. Let's say that $x = c$ is a *critical point* for the function f if either $f'(c) = 0$ or if $f'(c)$ does not exist. Then we have this nice result:

Theorem 7.1 *Suppose that f is defined on (a, b) and c is in (a, b). If c is a local maximum or minimum of f, then c must be a critical point for f. That is, either $f'(c) = 0$ or if $f'(c)$ does not exist.*

Proof. Let's first suppose that $x = c$ is a local minimum for f. If $f'(c)$ does not exist, then it's a critical point, which is exactly what we were hoping for. On the other hand, if $f'(c)$ exists, then

$$f'(c) = \lim_{h\to 0} \frac{f(c + h) - f(c)}{h}$$

Since c is a local minimum, we know that $f(c + h) \geq f(c)$ when $c + h$ is very close to c. Of course, $c + h$ is close to c exactly when h is close to 0. For such h, the numerator $f(c + h) - f(c)$ in the above fraction must be nonnegative.

When $h > 0$, the quantity

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

is positive (or 0), but when $h < 0$, the quantity is negative (or 0). So the right-hand limit is positive (or 0), but when $h < 0$, the quantity is negative

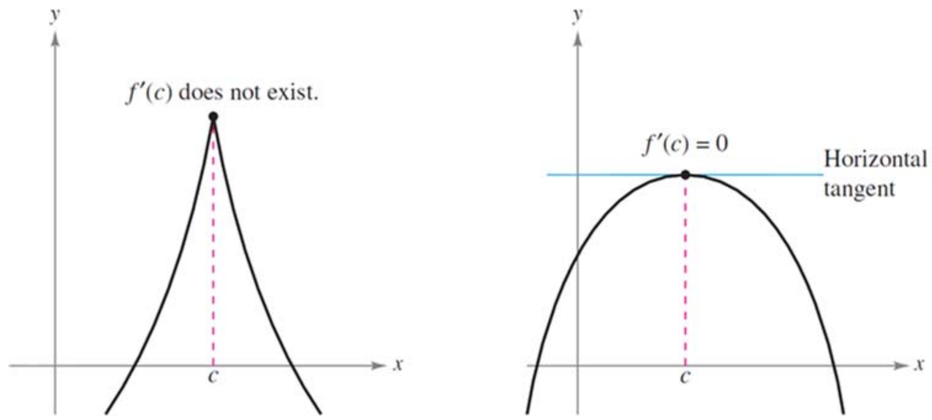


Fig. 7.2. c is a critical point of f.

(or 0). So the right-hand limit

$$f'(c) = \lim_{h \to 0^+} \frac{f(c+h) - f(c)}{h}$$

must be greater than or equal to 0, while the same left-hand limit is less than or equal to 0. Since the two-sided limit exists, the left-hand and right-hand limits are equal; the only possibility is that they are both 0. This shows that $f'(c) = 0$, so $x = c$ is once again a critical point for f.

If $x = c$ is a local maximum we can repeat the argument. ■

So local maxima and minima in an open interval *occur only at critical points.* Figure 7.2 illustrates the two types of critical numbers. Notice in the definition that the critical number has to be in the domain of f but does not have to be in the domain of f'. It's *not true* that a critical point must be a local maximum or minimum! For example, if $f(x) = x^3$, then $f'(x) = 3x^2$, and you can see that $f'(0) = 0$. This means that $x = 0$ is a critical point for f. On the other hand, $x = 0$ is neither a local maximum nor a local minimum, as you can see by drawing the graph of $y = x^3$.

The above theorem applies to open intervals. How about when the domain of your function is a closed interval $[a, b]$? Then the endpoints a and b might be local maxima and minima; they aren't covered by the theorem. So in the case of a closed interval, local maxima and minima can occur only at critical points or at the endpoints of the interval.

Guidelines for finding extrema on a closed interval:

To find the extrema of a continuous function on a closed interval:
1. Find the critical numbers of in (a, b).
2. Evaluate at each critical number in (a, b).
3. Evaluate at each endpoint of $[a, b]$.
4. The least of these values is the minimum. The greatest is the maximum.

Example 7.2 *Find the extrema of* $f(x) = 3x^4 - 4x^3$ *on the interval* $[-1, 2]$.

Solution: Begin by differentiating the function.

$$f'(x) = 12x^3 - 12x^2$$

To find the critical numbers of f you must find all x-values for which $f'(x) = 0$ and all x-values for which $f'(x)$ does not exist. Set $f'(x)$ equal to 0 and factor

$$\begin{aligned} f'(x) &= 12x^3 - 12x^2 = 0 \\ & 12x^2 (x - 1) = 0. \end{aligned}$$

Now you can easily find that the critical numbers are: $x = 0$ and $x = 1$. Because $f'(x)$ is defined for all x from the interval $[-1, 2]$, you can conclude that these are the only critical numbers of f. By evaluating f at these two critical numbers and at the endpoints of $[-1, 2]$ you can determine that the maximum is $f(2) = 16$ and the minimum is $f(1) = -1$ as shown in the table.

Left endpoint	Critical number	Critical number	Right endpoint
$f(-1) = 7$	$f(0) = 0$	$f(1) = -1$ Minimum	$f(2) = 16$ Maximum

The graph of is shown in Figure 7.3. In Figure 7.3, note that the critical number $x = 0$ does not yield a relative minimum or a relative maximum.
□

Example 7.3 *Find the absolute maximum and minimum values of* $f(x) = 2x - 3x^{2/3}$ *on the interval* $[-1, 3]$.

Solution: We calculate the derivative to find the critical numbers. Note the difficult factoring steps.

$$\begin{aligned} f(x) &= 2x - 3x^{2/3} \\ f'(x) &= 2 - \frac{2}{\sqrt[3]{x}} = 2\left(\frac{\sqrt[3]{x} - 1}{\sqrt[3]{x}}\right). \end{aligned}$$

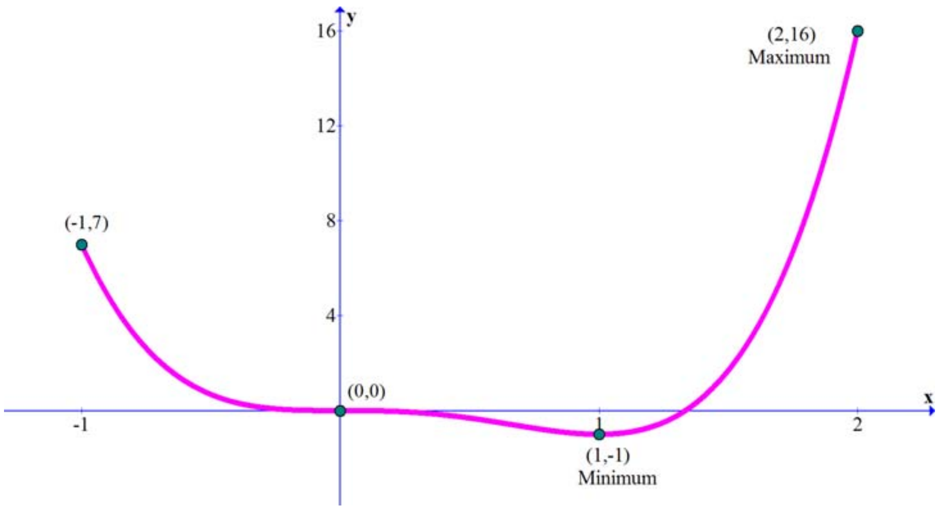


Fig. 7.3. The graph of $f(x) = 3x^4 - 4x^3$ on the interval $[-1, 2]$.

The critical numbers are 1 (where the derivative is 0) and 0 (where the derivative is undefined). We now evaluate the function at these 2 points and the endpoints.

Left endpoint	Critical number	Critical number	Right endpoint
$f(-1) = -5$ Minimum	$f(0) = 0$ Maximum	$f(1) = -1$	$f(3) = 6 - 3\sqrt[3]{9} \approx -0.24025$

The absolute maximum is 0, and the absolute minimum is -5 . Notice on the graph that there is a relative maximum at $(3, 6-3\sqrt[3]{9})$ and a relative minimum at $(1, -1)$. The graph has a sharp corner, or *cusp*, at $(0, 0)$. □

If the domain isn't bounded, then the situation is a little more complicated. For example, consider the two functions f and g, both with domain $[0, \infty)$, whose graphs look like as it is shown in Figure 7.5. In both cases, $x = 2$ is obviously a critical point, while the endpoints are 0 and ∞. Wait a second, ∞ isn't really an endpoint, since it doesn't really exist! Let's add it to the list anyway, so that the list is $0, 2$, and ∞; note that the same list works for both f and g. Let's take a look at f first. We see that $f(0) = 0$, $f(2) = 3$, while $f(\infty)$ only makes sense if you think of it as

$$\lim_{x\to\infty} f(x).$$

This limit is 1, since $y = 1$ is a horizontal asymptote for f. The highest of these function values is 3, which occurs at $x = 2$, so $x = 2$ is a global maximum for

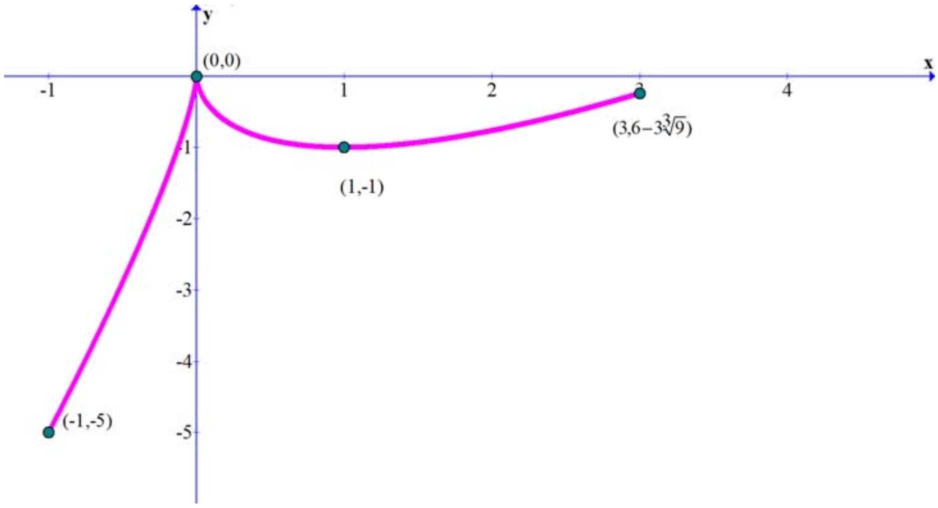


Fig. 7.4. The graph of $f(x) = 2x - 3x^{2/3}$ on the interval $[-1, 3]$.

f. The lowest function value is at $x = 0$, so $x = 0$ is a global minimum for f. The right-hand "endpoint" at 1 doesn't even come into it.

How about g? Well, this time $g(0) = 2$, $g(2) = 3$, and the right-hand endpoint is covered by the observation that

$$\lim_{x\to\infty} g(x) = 1.$$

The highest value is still 3, which occurs at $x = 2$, so $x = 2$ is also a global maximum for g. How about the lowest value? Well, that value, which is 1, occurs as $x \to \infty$. Does this mean that 1 is a global minimum for g? Of course not, because ∞ isn't even a number; the function g has no global minimum.[1] At $x = 0$ the function $g(x)$ has a local minimum only.

7.2 Mean Value Theorem (MVT)

The Mean Value Theorem is a cornerstone in the theoretical framework of calculus. Several critical theorems rely on the Mean Value Theorem; the theorem also appears in practical applications. We begin with a preliminary result known as Rolle's theorem.

Consider a function f that is continuous on a closed interval $[a, b]$ and differentiable on the open interval (a, b). Furthermore, assume f has the special

[1] On the other hand, g does have a global *infimum*. This concept is a little beyond our scope, though. Check out a book on real analysis if you want to learn more.

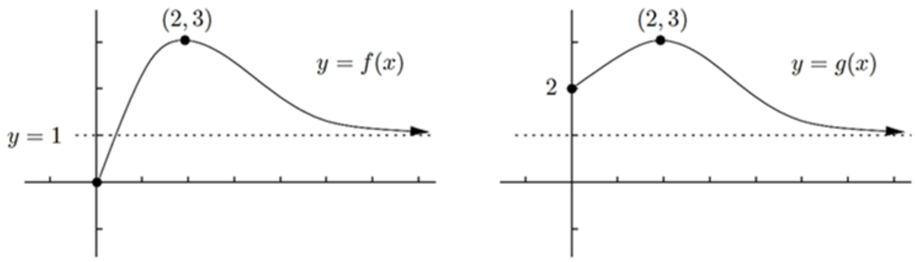


Fig. 7.5. Functions f and g, both with domain $[0;\infty)$.

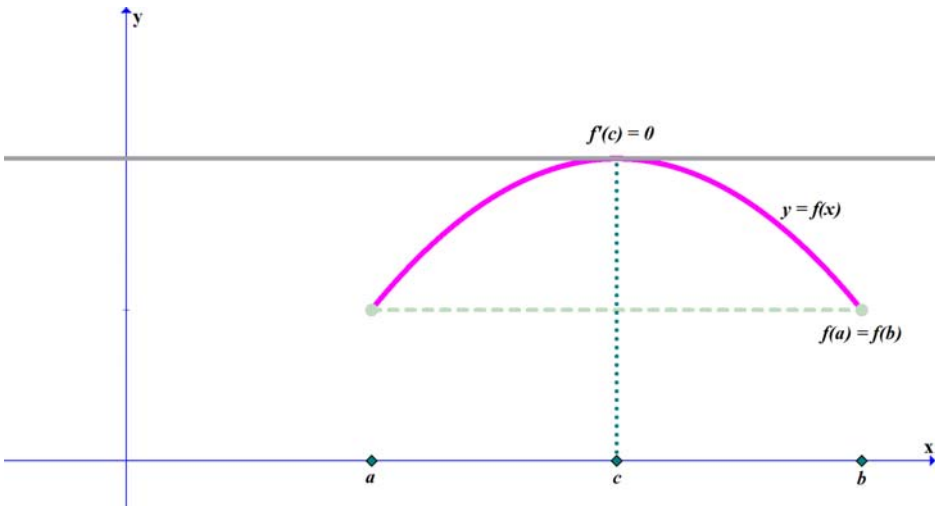


Fig. 7.6. Rolle's theorem

property that $f(a) = f(b)$ (Figure 7.6). The statement of Rolle's Theorem is not surprising: It says that somewhere between a and b, there is at least one point at which f has a horizontal tangent line.

Theorem 7.4 (Rolle) *Let f be continuous on a closed interval $[a, b]$ and differentiable on (a, b) with $f(a) = f(b)$. There is at least one point c in (a, b) such that $f'(c) = 0$.*

Proof. The function f satisfies the conditions of max-min Theorem (Section 5.3 of Chapter 5) and thus attains its absolute maximum and minimum values on $[a, b]$. Those values are attained either at an endpoint or at an interior point c.

Case 1: First suppose that f attains both its absolute maximum and minimum values at the endpoints. Because $f(a) = f(b)$, the maximum and minimum values are equal, and it follows that f is a constant function on $[a, b]$. Therefore, $f'(x) = 0$ for all x in (a, b),and the conclusion of the theorem holds.

Case 2: Assume at least one of the absolute extreme values of f does not occur at an endpoint. Then, f must attain an absolute extreme value at an interior point of (a, b); therefore,f must have either a local maximum or a local minimum at a point c in (a, b). We know from Theorem 7.1 that at a local extremum the derivative is zero. Thus, $f'(c) = 0$ for at least one point c of (a, b), and again the conclusion of the theorem holds. ■

Why does Rolle's Theorem require continuity? A function that is not continuous on $[a, b]$ may have identical values at both endpoints and still not have a horizontal tangent line at any point on the interval. Similarly, a function that is continuous on $[a, b]$ but not differentiable at a point of (a, b) may also fail to have a horizontal tangent line. Sketch examples and see Figure 7.8.

Exercise 7.1 *In order to verify Rolle's theorem find an interval I on which Rolle's theorem applies to $f(x) = x^3 - 6x^2 + 8x$. Then find all the points c in I at which $f'(c) = 0$.*

Solution: Because f is a polynomial, it is everywhere continuous and differentiable. We need an interval $[a, b]$ with the property that $f(a) = f(b)$. Noting that

$$f(x) = x(x - 2)(x - 4),$$

we choose the interval $[0, 4]$, because $f(0) = f(4) = 0$ (other intervals are possible). The goal is to find points c in the interval $(0, 4)$ at which $f'(c) = 0$, which amounts to the familiar task of finding the critical points of f. The critical points satisfy

$$f'(x) = 3x^2 - 12x + 8 = 0.$$

Using the quadratic formula, the roots are

$$x_{1/2} = 2 \pm \frac{2}{3}\sqrt{3} \qquad \text{or} \qquad x_1 \approx 3.1547 \quad \text{and} \quad x_2 \approx 0.84530.$$

As shown in Figure 7.7, the graph of f has two points at which the tangent line is horizontal. □

Rolle's theorem is not just a stepping stone toward the mean-value theorem. It is in itself a useful tool.

Example 7.5 *Use Rolle's theorem to show that $p(x) = 2x^3+5x-1$ has exactly one real zero.*

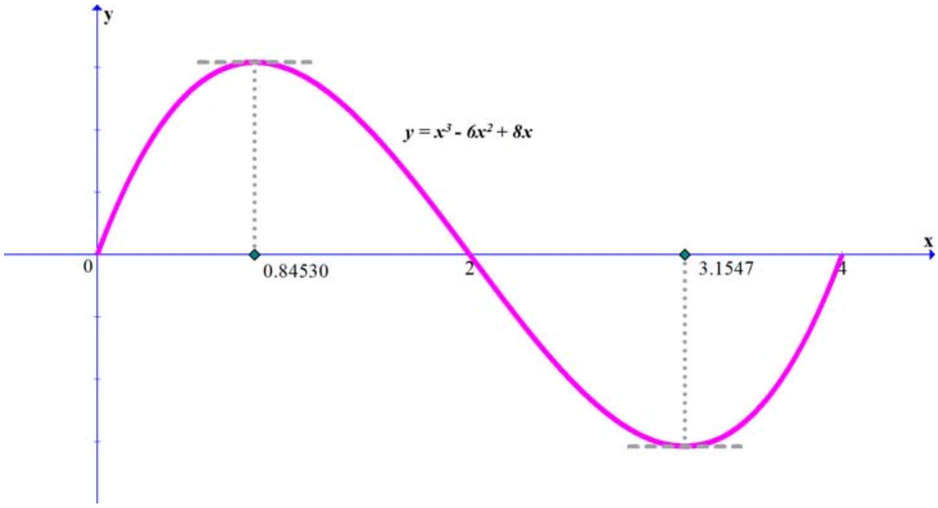


Fig. 7.7. $f'(c) = 0$ at two points in $(0, 4)$.

Solution: Since p is a cubic, we know that p has at least one real zero. Suppose that p has more than one real zero. In particular, suppose that $p(a) = p(b) = 0$ where a and b are real numbers and $a \neq b$. Without loss of generality, we can assume that $a < b$. Since every polynomial is everywhere differentiable, p is differentiable on (a, b) and continuous on $[a, b]$. Thus, by Rolle's theorem, there is a number c in (a, b) for which $p'(c) = 0$. But

$$p'(x) = 6x^2 + 5 \geq 5 \qquad \text{for all } x,$$

and $p'(c)$ cannot be 0. The assumption that p has more than one real zero has led to a contradiction. We can conclude therefore that p has only one real zero. □

Remark 7.6 *If f fails to be differentiable at even one number in the interval, then the conclusion of the Rolle's theorem may be false (see Figure 7.8)*

Now we can consider extremely important generalization of the Rolle's theorem: The Mean Value Theorem. It is easily understood with the aid of a picture. Figure 7.9 shows a function f differentiable on (a, b) with a secant line passing through $(a, f(a))$ and $(b, f(b))$; the slope of the secant line is the average rate of change of f over $[a, b]$. The Mean Value Theorem claims that there exists a point c in (a, b) at which the slope of the tangent line at c is equal to the slope of the secant line. In other words, we can find a point on the graph of f where the tangent line is parallel to the secant line.

We are now ready to state and give a proof of the Mean-value theorem.

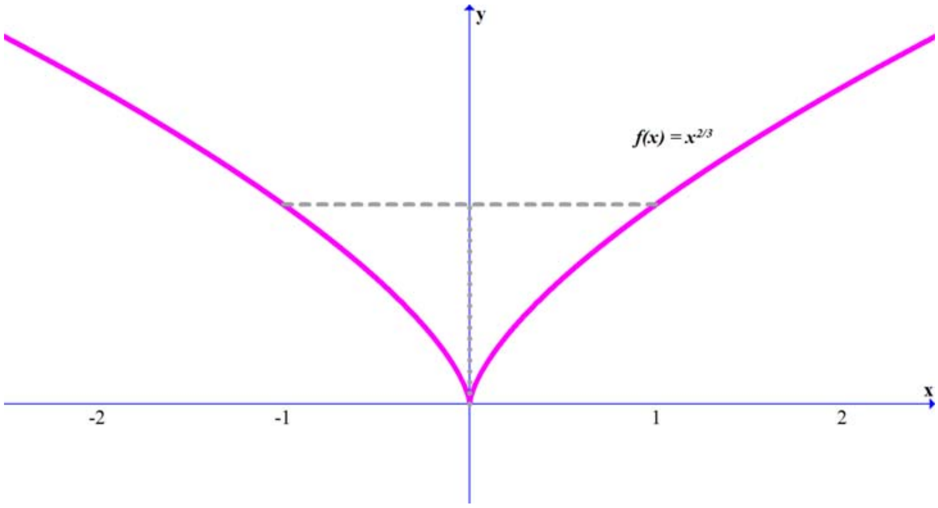


Fig. 7.8. Conclusion of the Rolle's theorem is false for $f(x) = x^{2/3}$ (with $f'(x) = \dfrac{2}{3\sqrt[3]{x}}$) on $[-1, 1]$.

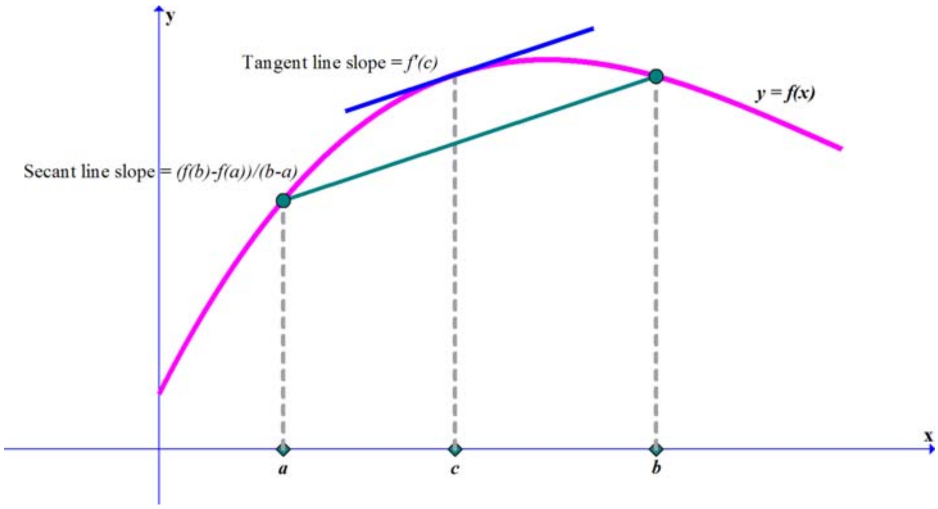


Fig. 7.9. $\dfrac{f(b) - f(a)}{b - a} = f'(c)$.

Theorem 7.7 (Mean-value theorem) *If f is continuous on the closed interval $[a, b]$ and differentiable on (a, b), then there is at least one point c in (a, b) such that*

$$\frac{f(b) - f(a)}{b - a} = f'(c).$$

Proof. The equation of the secant through $(a, f(a))$ and $(b, f(b))$ is

$$y - f(a) = \frac{f(b) - f(a)}{b - a}(x - a)$$

which we can rewrite as

$$y = \frac{f(b) - f(a)}{b - a}(x - a) + f(a).$$

Let

$$g(x) = f(x) - \left[\frac{f(b) - f(a)}{b - a}(x - a) + f(a)\right].$$

Note that $g(a) = g(b) = 0$. Also, g is continuous on $[a, b]$ and differentiable on (a, b) since f is. So, by Rolle's theorem there exists c in (a, b) such that $g'(c) = 0$. But

$$g'(x) = f'(x) - \frac{f(b) - f(a)}{b - a},$$

so

$$g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a} = 0.$$

Therefore

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

and the proof is complete. ■

The main significance of the Mean Value Theorem is that it enables us to obtain information about a function from information about its derivative. Our immediate use of this principle is to prove the basic facts concerning increasing and decreasing functions.

Remark 7.8 *Rolle's theorem is a special case of MVT, but the Mean Value Theorem is also (as we have seen) a consequence of Rolle's theorem. If two mathematical statements are each consequences of each other, they are called* ***equivalent.*** *Thus Rolle's theorem is equivalent to the Mean Value Theorem.*

Now let's look at an example of how to use the MVT theorem.

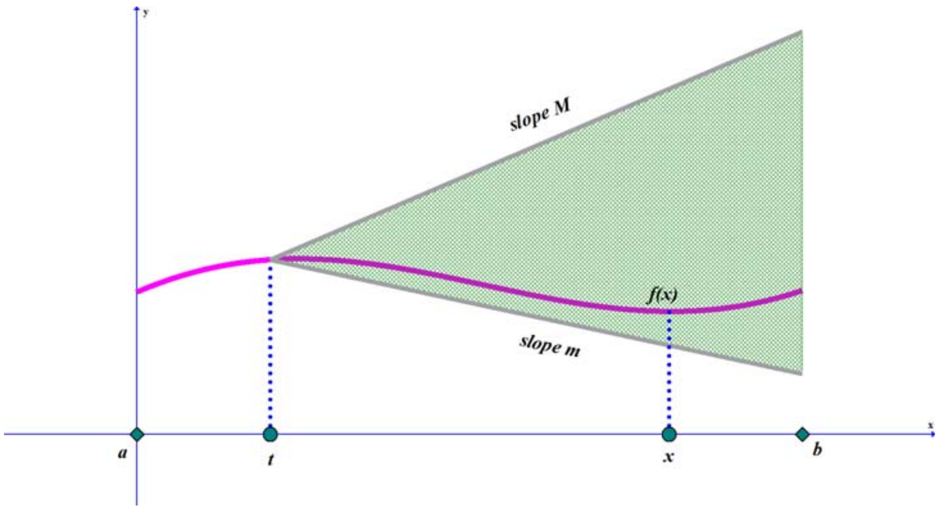


Fig. 7.10. Starting from t the graph of f must lie between two stright lines with the slopes m and M.

Example 7.9 *Let f be continuous on $[a, b]$ and differentiable on (a, b) with $m \le f'(x) \le M$ for all x in (a, b). Show that*

$$m(x - t) \le f(x) - f(t) \le M(x - t) \quad \text{if} \quad a \le t \le x \le b.$$

Solution: Let $a \le t \le x \le b$. Then for come c between t and x,

$$f'(c) = \frac{f(x) - f(t)}{x - t}.$$

So

$$m \le f'(c) = \frac{f(x) - f(t)}{x - t} \le M,$$

and therefore

$$m(x - t) \le f(x) - f(t) \le M(x - t).$$

You can interpret this inequality graphically fixing t and drawing lines with the slope m and M through the point $(t, f(t))$. The result says, that starting from t the graph of f must lie between these two lines (Figure 7.10). □
Now, let's use the Mean Value Theorem to show two useful facts about derivatives:

1. Suppose that a function f has derivative $f'(x) = 0$ for every x in some interval (a, b). It is intuitively obvious that the function should be constant on the whole interval. How do we prove it? First, fix some

special number S in the interval, and then pick any other number x in the interval. We know from the Mean Value Theorem that there's some number c between S and x such that

$$f'(c) = \frac{f(x) - f(S)}{x - S}.$$

Now we have assumed that f' is always equal to 0, the quantity $f'(c)$ must be 0. So the above equation says that

$$\frac{f(x) - f(S)}{x - S} = 0,$$

which means that $f(x) = f(S)$. If we now let $C = f(S)$, we have shown that $f(x) = C$ for all x in the interval (a, b), so f is constant! In summary,

$$\boxed{\text{if } f'(x) = 0 \text{ for all } x \text{ in } (a, b), \text{ then } f \text{ is constant on } (a, b).}$$

2. Suppose that two differentiable functions have exactly the same derivative. Are they the same function? Not necessarily. They could differ by a constant; for example, $f(x) = x^2$ and $g(x) = x^2 + 1$ have the same derivative, $2x$, but f and g are clearly not the same function. Is there any other way that two functions could have the same derivative everywhere? The answer is no. Differing by a constant is the only way:

$$\boxed{\text{if } f'(x) = g'(x) \text{ for all } x, \text{ then } f(x) = g(x) + C \text{ for some constant } C.}$$

It turns out to be quite easy to show this using #1 above. Suppose that $f'(x) = g'(x)$ for all x. Now set $h(x) = f(x) - g(x)$. Then we can differentiate to get $h'(x) = f'(x) - g'(x) = 0$ for all x, so h is constant. That is, $h(x) = C$ for some constant C. This means that $f(x) - g(x) = C$, or $f(x) = g(x) + C$. The functions f and g do indeed differ by a constant. This fact will be very useful when we look at integration in a few chapters' time.

7.3 Increasing and decreasing functions

In this section you will learn how derivatives can be used to relative extrema as either relative minima or relative maxima. First, it is important to define increasing and decreasing functions.

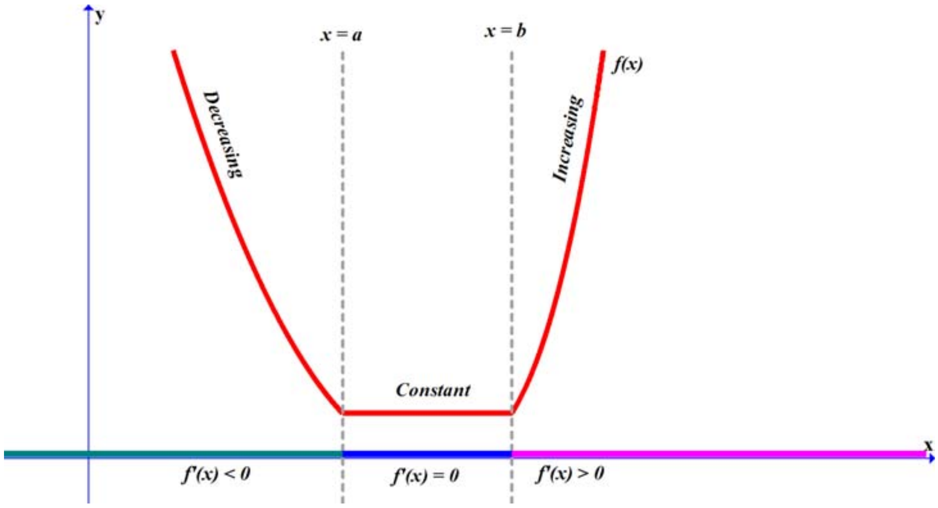


Fig. 7.11. The derivative is related to the slope of a function.

Definition 7.10 (of increasing and decreasing functions) *A function f is* increasing *on an interval if for any two numbers x_1 and x_2 in the interval, $x_1 < x_2$ implies $f(x_1) < f(x_2)$.*
A function f is decreasing *on an interval if for any two numbers x_1 and x_2 in the interval, $x_1 < x_2$ implies $f(x_1) > f(x_2)$.*

A function is increasing if, *as x moves to the right*, its graph moves up, and is decreasing if its graph moves down. Increasing function "preserves inequalities", whereas decreasing function "reverses" inequalities. For example, the function in Figure 7.11 is decreasing on the interval $(-\infty, a)$, is constant on the interval (a, b) and is increasing on the interval (b, ∞). As shown in Theorem 7.11 below, a positive derivative implies that the function is increasing; a negative derivative implies that the function is decreasing; and a zero derivative on an entire interval implies that the function is constant on that interval.

Theorem 7.11 (Test for increasing and decreasing functions) *Let f be a function that is continuous on the closed $[a, b]$ interval and differentiable on the open interval (a, b).*

1. *If $f'(x) > 0$ for all x in (a, b) then f is increasing on $[a, b]$.*
2. *If $f'(x) < 0$ for all x in (a, b) then f is decreasing on $[a, b]$.*
3. *If $f'(x) = 0$ for all x in (a, b) then f is constant on $[a, b]$.*

Proof. To prove the first case, assume that $f'(x) > 0$ for all x in the interval (a, b) and let $x_1 < x_2$ be any two points in the interval. By the Mean Value Theorem, you know that there exists a number c such that $x_1 < c < x_2$ and

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = f'(c).$$

Because and you know that $f'(c) > 0$ and $x_2 - x_1 > 0$, you know that

$$f(x_2) - f(x_1) > 0$$

which implies that $f(x_1) < f(x_2)$. So, f is increasing on the interval. The second case has a similar proof, and the third case was considered on page 188. ■

Remark 7.12 *The conclusions in the first two cases of Theorem 7.11 are valid even if $f'(x) = 0$ at a finite number of x-values in (a, b) (see Example 7.15).*

To carry out the test for increasing and decreasing functions, we make a useful observation: $f'(x)$ can change sign at a critical point, but it cannot change sign on the interval between two consecutive critical points (one can prove this is true even if $f'(x)$ is not assumed to be continuous). So we can determine the sign of $f'(x)$ on an interval between consecutive critical points by evaluating $f'(x)$ at an any test point x_0 inside the interval. The sign of $f'(x_0)$ is the sign of $f'(x)$ on the entire interval.

Example 7.13 *Let $f(x) = 3x^4 - 4x^3 - 12x^2 + 10$. Find the intervals on which f is increasing and the intervals on which f is decreasing.*

Solution: Note that $f(x)$ is differentiable on the entire real number line. First compute $f'(x)$ and simplify

$$\begin{aligned} f'(x) &= 12x^3 - 12x^2 - 24 \\ &= 12x\left(x^2 - x - 2\right) \\ &= 12x\,(x+1)\,(x-2) \end{aligned}$$

To determine the critical numbers of f, set f' equal to zero.

$$f'(x) = 0 \qquad \text{at} \quad x = 0, -1, 2.$$

Because there are no points for which $f'(x)$ does not exist, you can conclude that $x = 0, -1, 2$ are the only critical numbers. The table summarizes the

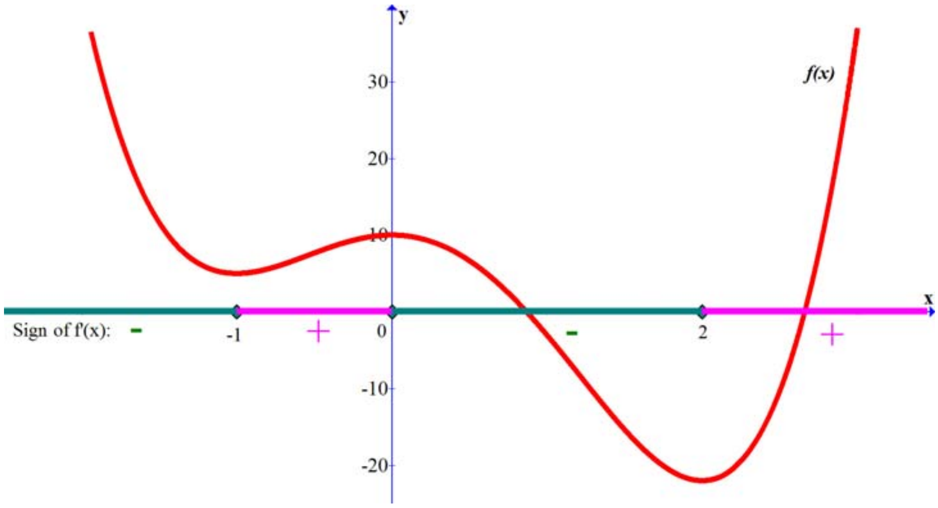


Fig. 7.12. The graph of the function $f(x) = 3x^4 - 4x^3 - 12x^2 + 10$.

testing of the four intervals determined by these three critical numbers.

Interval	$(-\infty, -1)$	$(-1, 0)$	$(0, 2)$	$(2, \infty)$
Test Value	-2	-0.5	1	3
Sign of $f'(x)$	$-$	$+$	$-$	$+$
Conclusion	Decreasing	Increasing	Decreasing	Increasing

So, f is increasing on the intervals $[-1, 0]$ and $[2, \infty)$ and decreasing on the intervals $(-\infty, -1]$ and $[0, 2]$ as shown in Figure 7.12.

Example 7.14 *Let* $f(x) = x^{1/3}(2 - x)$. *Find the intervals on which* f *is increasing and the intervals on which* f *is decreasing.*

Solution: First we compute $f'(x)$ and simplify

$$f(x) = x^{1/3}(2 - x) = 2x^{1/3} - x^{4/3}$$

$$\begin{aligned} f'(x) &= \frac{2}{3}x^{-2/3} - \frac{4}{3}x^{1/3} \\ &= \frac{2(1 - 2x)}{3x^{2/3}}. \end{aligned}$$

$$f'(x) = 0 \quad \text{at} \quad x = \frac{1}{2}, \quad \text{and is undefined at } x = 0.$$

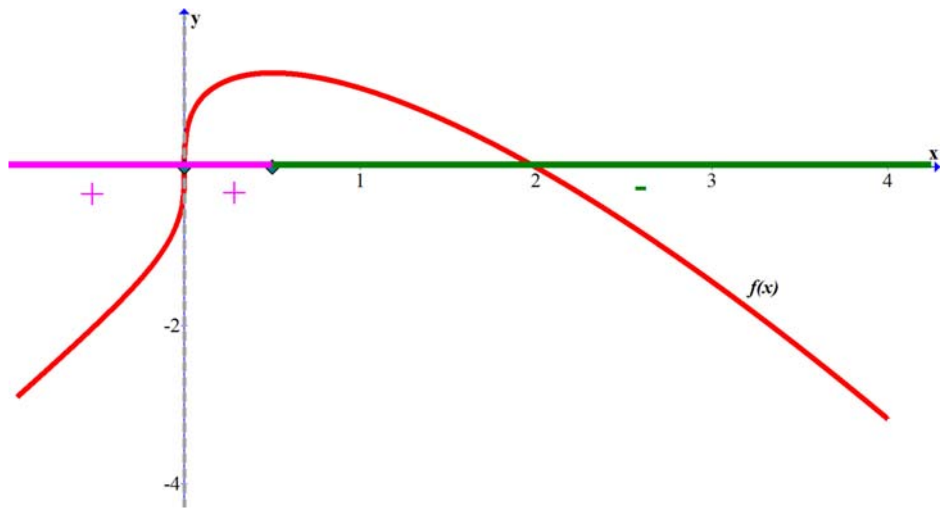


Fig. 7.13. The graph of the function $f(x) = x^{1/3}(2 - x)$.

Now you can conclude that $x = 0$, and $x = 1/2$ are the only critical numbers. The table summarizes the testing of the three intervals determined by these two critical numbers.

Interval	$(-\infty, 0)$	$(0, \frac{1}{2})$	$(\frac{1}{2}, \infty)$
Test Value	-2	0.25	1
Sign of $f'(x)$	$+$	$+$	$-$
Conclusion	Increasing	Increasing	Decreasing

Continuous function $f(x)$ has the vertical asymptote at $x = 0$, as you can see on Figure 7.13. □

Example 7.15 *Let* $f(x) = 6x^5 - 15x^4 + 10x^3$. *Find the intervals on which* f *is increasing and the intervals on which* f *is decreasing.*

Solution: Note that $f(x)$, as a polynomial, is differentiable on the entire real number line. We have

$$\begin{aligned} f'(x) &= 30x^4 - 60x^3 + 30x^2 \\ &= 30x^2\left(x^2 - 2x + 1\right) \\ &= 30x^2(x-1)^2. \end{aligned}$$

So

$$f'(x) = 0 \quad \text{at} \quad x = 0 \text{ and } x = 1.$$

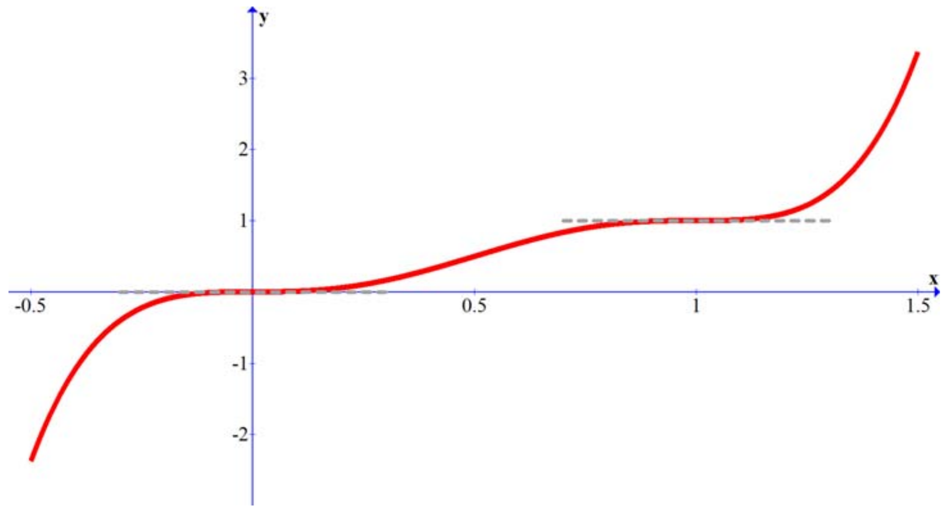


Fig. 7.14. The graph of the function $f(x) = 6x^5 - 15x^4 + 10x^3$.

Now we can conclude that $x = 0$, and $x = 1$ are the only critical numbers. The table summarizes the testing of the three intervals determined by these two critical numbers.

Interval	$(-\infty, 0)$	$(0, 1)$	$(1, \infty)$
Test Value	-2	0.5	2
Sign of $f'(x)$	$+$	$+$	$+$
Conclusion	Increasing	Increasing	Increasing

So, f is increasing on $(-\infty, \infty)$ as it is shown in Figure 7.14, where two horizontal lines are also drawn. This example illustrates the Remark 7.12. □

The guidelines below summarize the steps followed in the examples above.

Guidelines for finding intervals on which a function is increasing or decreasing:

Let f be continuous on the interval (a, b). To find the open intervals on which f is increasing or decreasing, use the following steps.

1. Locate the critical numbers of f in (a, b) and use these numbers to determine test intervals.

2. Determine the sign of $f'(x)$ at one test value in each of the intervals.

3. Use Theorem 7.11 to determine whether is increasing or decreasing on each interval.

These guidelines are also valid if the interval (a, b) is replaced by an interval of the form $(-\infty, b)$, (a, ∞), or $(-\infty, \infty)$.

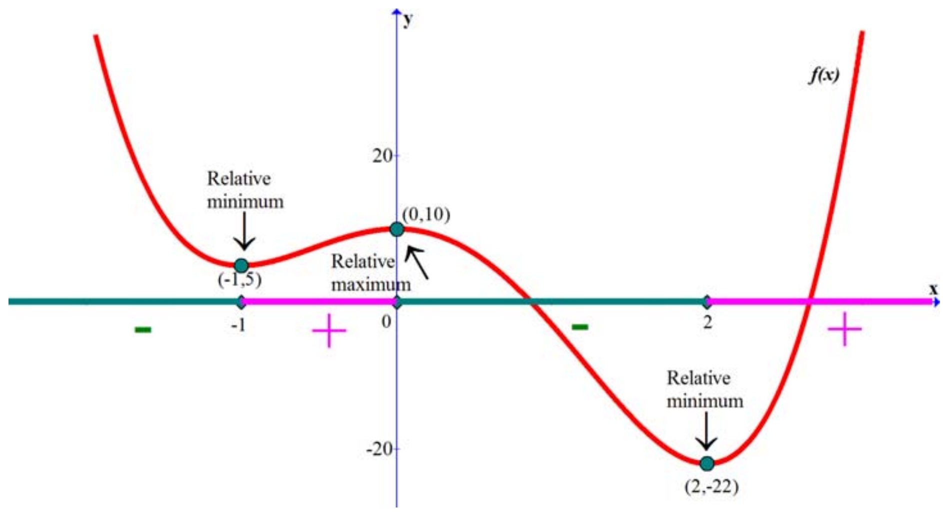


Fig. 7.15. Relative extrema of $f(x) = 3x^4 - 4x^3 - 12x^2 + 10$.

A function is *strictly monotonic* on an interval if it is either increasing on the entire interval or decreasing on the entire interval. For instance, the function $f(x) = x^3$ is strictly monotonic on the entire real number line because it is increasing on the entire real number line.

7.4 The First Derivative Test

There is a useful test for determining whether a critical point is a min or max (or neither) based on the sign change of the derivative $f'(x)$. For instance, in Figure 7.12 (from Example 7.13), the function $f(x) = 3x^4 - 4x^3 - 12x^2 + 10$ has a relative maximum at the point $(0, 0)$ because is increasing immediately to the left of and decreasing immediately to the right of $x = 0$. Similarly, it has a relative minimum at the point $(-1, 5)$ because is decreasing immediately to the left of and increasing immediately to the right of $x = -1$ (see Figure 7.15). The following theorem, called the First Derivative Test, makes this more explicit.

Theorem 7.16 (First Derivative Test) *Let c be a critical number of a function f that is continuous on an open interval I containing c. If f is differentiable on the interval, except possibly at c then $f(c)$ can be classified as follows.*

1. *If $f'(x)$ changes from negative to positive at c, then f has a* relative minimum *at $(c, f(c))$.*

2. *If $f'(x)$ changes from positive to negative at c, then f has a* relative maximum *at $(c, f(c))$.*

3. *If $f'(x)$ is positive on both sides of c or negative on both sides of c, then $f(c)$ is neither a relative minimum nor relative maximum (see Example 7.15).*

Proof. Assume that $f'(x)$ changes from negative to positive at c. Then there exist a and b in such that

$$f'(x) < 0 \quad \text{for all } x \text{ in } (a, c)$$

and

$$f'(x) > 0 \quad \text{for all } x \text{ in } (c, b).$$

By Theorem 7.11, f is decreasing on $[a, c]$ and increasing on $[c, b]$. So, is a minimum of on the open interval (a, b) and, consequently, a relative minimum of f. This proves the first case of the theorem. The second case can be proved in a similar way. ■

Example 7.17 *Use the first derivative test to find the relative extrema of $f(x) = \frac{1}{2}x - \sin x$ on the interval $(0, 2\pi)$.*

Solution: Note that $f(x)$ is continuous on the interval $(0, 2\pi)$. We have $f'(x) = \frac{1}{2} - \cos x$. The critical numbers are the solutions to the trigonometric equation $\cos x = \frac{1}{2}$ on the interval $(0, 2\pi)$, which are $\pi/3$ and $5\pi/3$. Because there are no points for which $f'(x)$ does not exist, you can conclude that $x = \pi/3$ and $x = 5\pi/3$ are the only critical numbers. The table summarizes the testing of the three intervals determined by these two critical numbers.

Interval	$(0, \pi/3)$	$(\pi/3, 5\pi/3)$	$(5\pi/3, 2\pi)$
Test Value	$x = \pi/4$	$x = \pi$	$x = 7\pi/4$
Sign of $f'(x)$	$f'(\pi/4) < 0$	$f'(\pi) > 0$	$f'(7\pi/4) < 0$
Conclusion	Decreasing	Increasing	Decreasing

By applying the First Derivative Test, you can conclude that has a relative minimum at the point where

$$x = \frac{\pi}{3}$$

and a relative maximum at the point where

$$x = \frac{5\pi}{3}$$

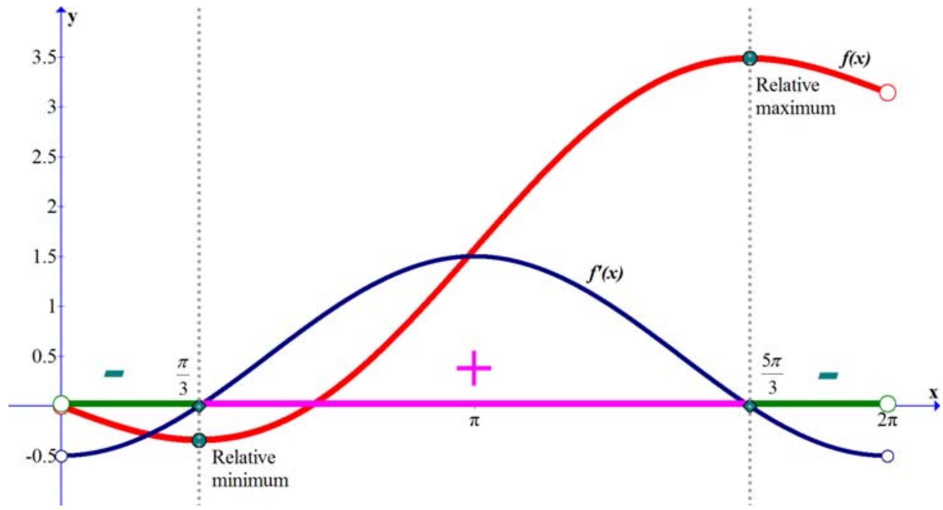


Fig. 7.16. A relative minimum occurs where f changes from decreasing to increasing, and a relative maximum occurs where f changes from increasing to decreasing.

as shown in Figure 7.16. □

Example 7.18 *Analyze the critical points and the increase/decrease behavior of* $f(x) = \cos^2 x + \sin x$ *in* $(0, \pi)$.

Solution: First, find the critical points:

$$f'(x) = -2\cos x \sin x + \cos x = (\cos x)(1 - 2\sin x) = 0$$

implies

$$\cos x = 0 \quad \text{or} \quad \sin x = \frac{1}{2}.$$

The critical points are $\pi/6$, $\pi/2$, and $5\pi/6$. They divide $(0, \pi)$ into four intervals:

$$\left(0, \frac{\pi}{6}\right), \quad \left(\frac{\pi}{6}, \frac{\pi}{2}\right), \quad \left(\frac{\pi}{2}, \frac{5\pi}{6}\right), \quad \left(\frac{5\pi}{6}, \pi\right).$$

We determine the sign of f' by evaluating f' at a test point inside each interval. Since

$$\frac{\pi}{6} \approx 0.52, \quad \frac{\pi}{2} \approx 1.57, \quad \frac{5\pi}{6} \approx 2.62 \quad \text{and} \quad \pi \approx 3.14,$$

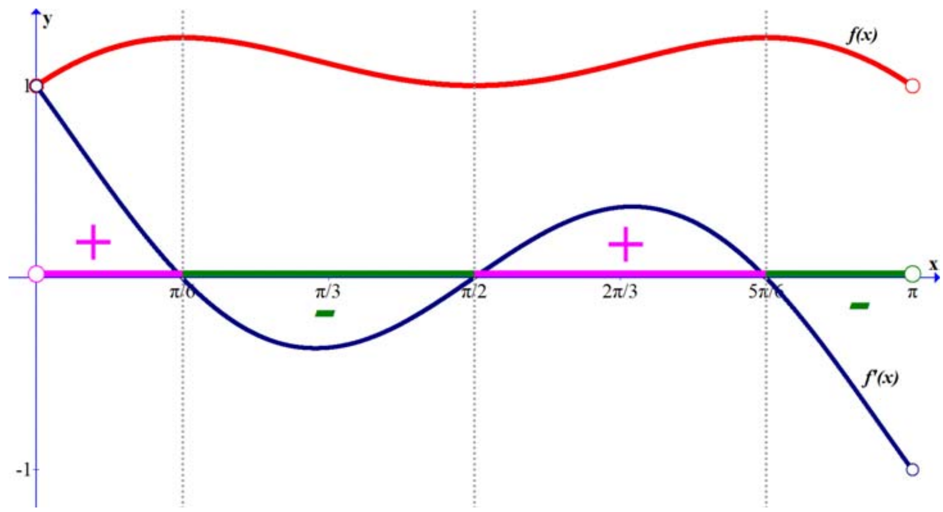


Fig. 7.17. Graph of $f(x) = \cos^2 x + \sin x$ and its derivative.

we can use the test points shown in the following table:

Interval	$\left(0, \frac{\pi}{6}\right)$	$\left(\frac{\pi}{6}, \frac{\pi}{2}\right)$	$\left(\frac{\pi}{2}, \frac{5\pi}{6}\right)$	$\left(\frac{5\pi}{6}, \pi\right)$
Test Value	$f'(0.5) \approx 0.04$	$f'(1) \approx -0.37$	$f'(2) \approx 0.34$	$f'(3) \approx -0.71$
Sign of $f'(x)$	$+$	$-$	$+$	$-$
Conclusion	Increasing	Decreasing	Increasing	Decreasing

Now apply the First Derivative Test:

- Local max at $c = \frac{\pi}{6}$ and $c = \frac{5\pi}{6}$ because f' changes from $+$ to $-$.
- Local min at $c = \frac{\pi}{2}$ because f' changes from $-$ to $+$.

The behavior of $f(x)$ and $f'(x)$ is reflected in the graphs in Figure 7.17. □

Note that in Examples 7.17 and 7.18 the given functions are differentiable on the entire domain. For such functions, the only critical numbers are those for which $f'(x) = 0$. Example 7.19 concerns a function that has two types of critical numbers–those for which it is and those for which it is not differentiable.

Example 7.19 *The function* $f(x) = 2x^{5/3} + 5x^{2/3}$ *is defined and continuous for all real* x*. The derivative of* f *is given by*

$$f'(x) = \frac{10}{3\sqrt[3]{x}} + \frac{10}{3}x^{\frac{2}{3}} = \frac{10}{3\sqrt[3]{x}}(x+1), \quad x \neq 0.$$

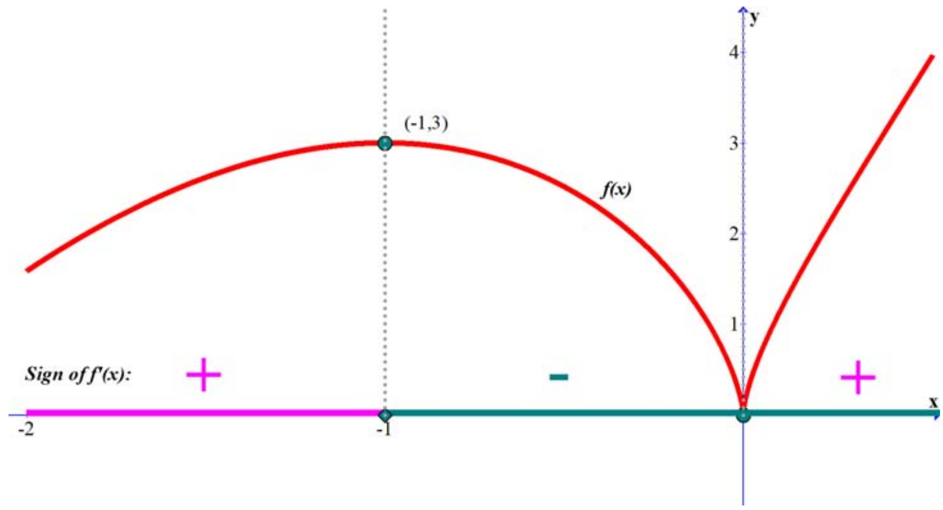


Fig. 7.18. Function $f(x) = 2x^{5/3} + 5x^{2/3}$.

Since $f'(-1) = 0$ and $f'(0)$ does not exist, the critical points are -1 and 0. The sign of f' is recorded in the table below.

Interval	$(-\infty, -1)$	$(-1, 0)$	$(0, \infty)$
Test Value	$x = -2$	$x = -0.5$	$x = 1$
Sign of $f'(x)$	$f'(-2) = \frac{5}{3}2^{\frac{2}{3}} > 0$	$f'(-0.5) \approx -2.1 < 0$	$f'(1) = \frac{20}{3} > 0$
Conclusion	*Increasing*	*Decreasing*	*Increasing*

In this case $f(-1) = 3$ is a local maximum and $f(0) = 0$ is a local minimum. The graph appears in Figure 7.18. □

Remark 7.20 *Note that the first-derivative test can be used at c only if f is continuous at c. The function*

$$f(x) = \begin{cases} 1 + 2x, & x \leq 1 \\ 5 - x, & x > 1 \end{cases}$$

(Figure 7.19) has no derivative at $x = 1$. Therefore 1 is a critical point. While it is true that $f'(x) > 0$ for $x < 1$ and $f(x) < 0$ for $x > 1$, it does not follow that $f(1)$ is a local maximum. The function f is discontinuous at $x = 1$ and the first-derivative test does not apply.

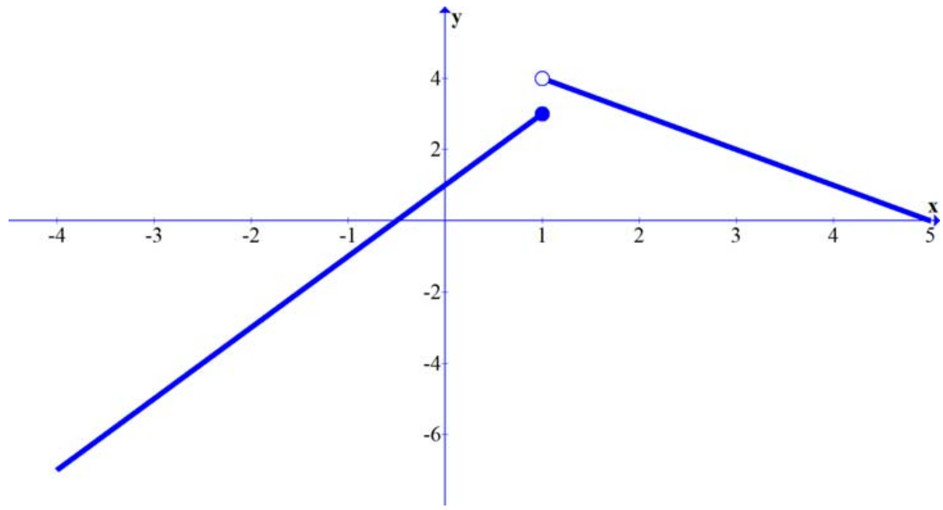


Fig. 7.19. The graph of $f(x) = \begin{cases} 1+2x, & x \leq 1 \\ 5-x, & x > 1 \end{cases}$

7.5 Concavity and the Second Derivative Test

You have already seen that locating the intervals in which a function increases or decreases helps to describe its graph. In this section, you will see how locating the intervals in which f' increases or decreases can be used to determine where the graph of f is curving upward or curving downward.

Definition 7.21 *A differentiable function f on some interval I is said to be* concave up *if f' is increasing and* concave down *if f' is decreasing. If f' is constant, then the function* has no concavity*. Points where a function changes concavity are called* inflection point*.*

To visualize the idea of concavity using the first derivative, consider the tangent line at a point. Recall that the slope of the tangent line is precisely the derivative. As you move along an interval, if the slope of the line is increasing, then f' is increasing and so the function is concave up. Similarly, if the slope of the line is decreasing, then f' is decreasing and so the function is concave down. In Figure 7.20, the tangent line at $P = (x, f(x))$ is drawn in red. The tangent line at $(x + .15, f(x + .15))$ is denoted by a dashed blue line.

If f' never changes sign twice in an interval .15 units wide or smaller, as is the case in example considered by this figure, then whenever the blue line has a larger slope than the red line, the derivative is increasing from x to $x + .15$ and the function is concave up on that interval. Likewise, whenever the blue

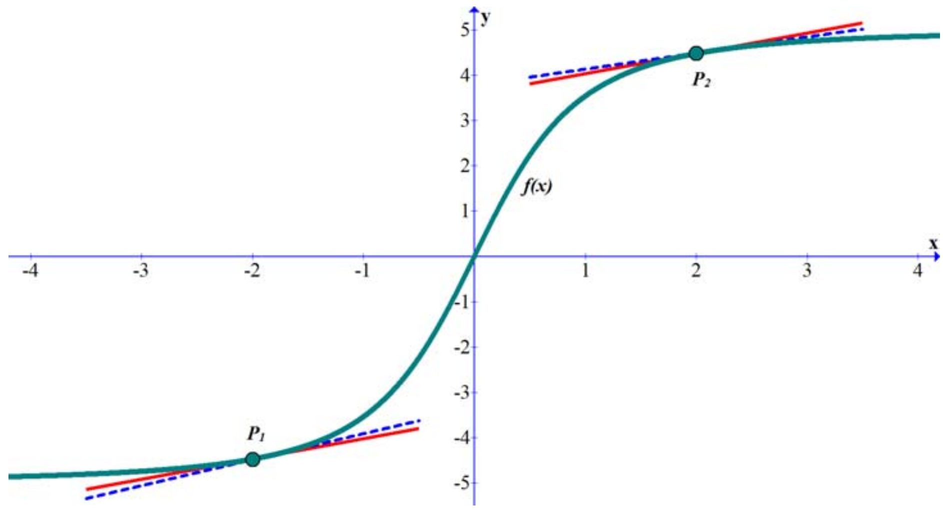


Fig. 7.20. Function $f(x) = \dfrac{5x}{\sqrt{1+x^2}}$ is concave up around $P_1 = (-2,\ -4.47214)$ and concave down around $P_2 = (2,\ 4.47214)$. The red line is the tangent to the curve at $P_{1/2}$ and the dashed blue line is the tangent to the curve a little to the right of $P_{1/2}$.

line has a smaller slope than the red line, the derivative is decreasing from x to $x + .15$ and the function is concave down on that interval.

In practice, we use the second derivative test to check concavity. The Second Derivative Test says that a function f is concave up when $f'' > 0$ and concave down when $f'' < 0$. This follows directly from the definition as the f is concave up when f' is increasing, and f' is increasing when its derivative f'' is positive. Similarly f is concave down when f' is decreasing, which occurs when $f'' < 0$. To apply this test, locate the x-values at which $f''(x) = 0$ or $f''(x)$ does not exist. Second, use these x-values to determine test intervals. Finally, test the sign of in each of the test intervals.

There is a convenient test for relative extrema that is based on the following geometric observation:

A function f has a relative maximum at a stationary point if the graph of f is concave down on an open interval containing that point, and it has a relative minimum if it is concave up. Now we can state this as a theorem.

Theorem 7.22 (Second Derivative Test for extrema) *Suppose that f is twice differentiable at the point x_0*

a. *If $f'(x_0) = 0$ and $f''(x0) > 0$, then f has a relative minimum at x_0.*

b. *If $f'(x_0) = 0$ and $f''(x0) < 0$, then f has a relative maximum at x_0.*

c. *If $f'(x_0) = 0$ and $f''(x0) = 0$, then the test is inconclusive; that is, f may have a relative maximum, a relative minimum, or neither at x_0.*

Proof. We will prove parts (a) and (c) and leave part (b) as an exercise.

a. We are given that $f'(x_0) = 0$ and $f''(x_0) > 0$, and we want to show that f has a relative minimum at x_0. Expressing $f''(x_0)$ as a limit and using the two given conditions we obtain

$$f''(x_0) = \lim_{x \to \infty} \frac{f'(x) - f'(x_0)}{x - x_0} = \lim_{x \to \infty} \frac{f'(x)}{x - x_0} > 0.$$

This implies that for x sufficiently close to but different from x_0 we have

$$\frac{f'(x)}{x - x_0} > 0. \tag{7.1}$$

Thus, there is an open interval extending left from x_0 and an open interval extending right from x_0 on which (7.1) holds. On the open interval extending left the denominator in (7.1) is negative, so $f'(x) < 0$, and on the open interval extending right the denominator is positive, so $f'(x) > 0$. It now follows from part (1) of the first derivative test (Theorem 7.16) that f has a relative minimum at x_0.

c. To prove this part of the theorem we need only provide functions for which $f'(x_0) = 0$ and $f''(x_0) = 0$ at some point x_0, but with one having a relative minimum at x_0, one having a relative maximum at x_0, and one having neither at x_0. We leave it as an exercise for you to show that three such functions are $f(x) = x^4$ (relative minimum at $x = 0$), $f(x) = -x^4$ (relative maximum at $x = 0$), and $f(x) = x^3$ (neither a relative maximum nor a relative minimum at x_0).

■

Example 7.23 *Find the relative extrema of $f(x) = 3x^5 - 5x^3$.*

Solution: We have

$$f'(x) = 15x^4 - 15x^2 = 15x^2 (x - 1) (x + 1)$$

$$f''(x) = 60x^3 - 30x = 30x \left(2x^2 - 1\right)$$

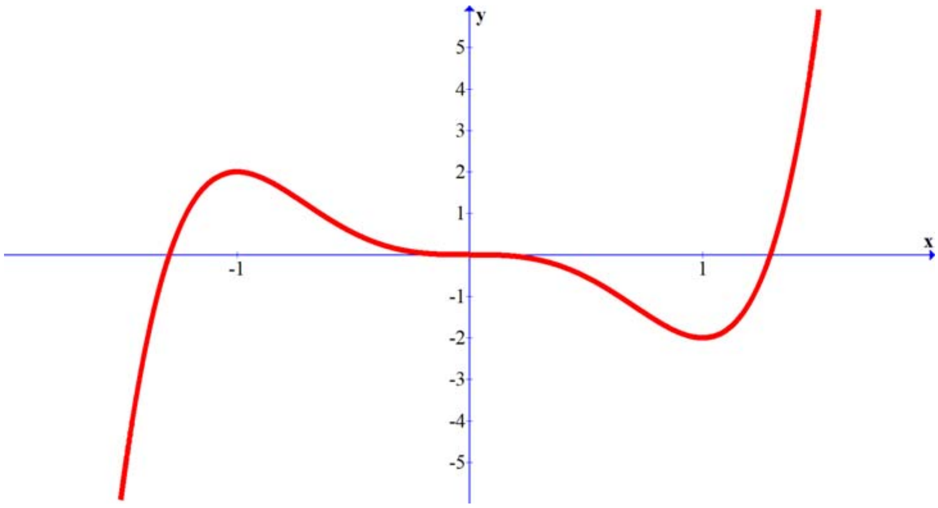


Fig. 7.21. The graph of $f(x) = 3x^5 - 5x^3$.

Solving $f'(x) = 0$ yields the stationary points $x = 0,\ x = -1$, and $x = 1$. As shown in the following table, we can conclude from the second derivative test that f has a relative maximum at $x = -1$ and a relative minimum at $x = 1$.

Stationary point	$30x\left(2x^2 - 1\right)$	$f''(x)$	Second derivative test
$x = -1$	-30	$-$	f has a relative maximum
$x = \ \ 0$	0	0	Inconclusive
$x = \ \ 1$	30	$+$	f has a relative minimum

The test is inconclusive at $x = 0$, so we will try the first derivative test at that point. A sign analysis of f' is given in the following table:

Interval	$(-1, 0)$	$(0, -1)$
Test Value	$x = -0.5$	$x = 0.5$
Sign of $f'(x)$	$f'(-0.5) < 0$	$f'(0.5) < 0$
Conclusion	Decreasing	Decreasing

Since there is no sign change in f' at $x = 0$, there is neither a relative maximum nor a relative minimum at that point. It is a point of inflection. All of this is consistent with the graph of f shown in Figure 7.21. □

Remark 7.24 *The second derivative test for extrema is often easier to apply than the first derivative test. However, the first derivative test can be used at any critical point of a continuous function, while the second derivative test applies only at stationary points where the second derivative exists.*

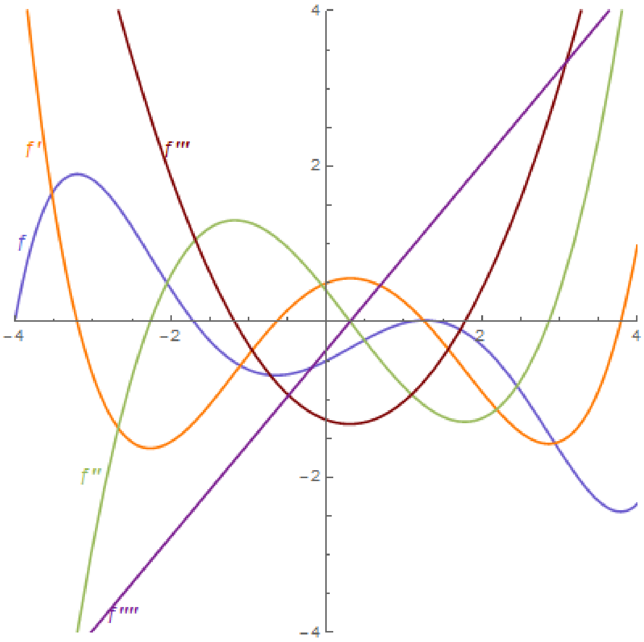


Fig. 7.22. The graph of $f(x) = 0.01x^5 - 0.015x^4 - 0.21x^3 + 0.2x^2 + 0.5x - 0.5$, and up to four of its derivatives.

Example 7.25 *This Example displays (Figure) the graph of the fifth-degree polynomial,* $f(x) = 0.01x^5 - 0.015x^4 - 0.21x^3 + 0.2x^2 + 0.5x - 0.5$, *and up to four of its derivatives. As you move across, note how when the function (blue curve) goes down, the first derivative (orange) is below the x axis, and when the function has a maximum or minimum the first derivative crosses the x axis. Check how the second derivative (green) shows the concavity of the function. Why is the fourth derivative a straight line?*

7.6 Review exercises: Chapter

8

Sketching graphs

In many problems, the properties of interest in the graph of a function are:

- x-intercepts and y-intercepts,
- Symmetry,
- Domain and range,
- Continuity,
- Vertical asymptotes,
- Differentiability,
- Relative extrema,
- Concavity,
- Points of inflection,
- Horizontal asymptotes,
- Infinite limits at infinity.
- Periodicity.
- Cusps.

Some of these properties may not be relevant in certain cases; for example, asymptotes are characteristic of rational functions but not of polynomials, and periodicity is characteristic of trigonometric functions but not of polynomial or rational functions. Thus, when analyzing the graph of a function f, it helps to know something about the general properties of the family to which it belongs.

In a given problem you will usually have a definite objective for your analysis of a graph. For example, you may be interested in showing all of the important characteristics of the function, you may only be interested in the behavior of the graph as $x \to \infty$ or as $x \to -\infty$, or you may be interested in some specific feature such as a particular inflection point. Thus, your objectives in the problem will dictate those characteristics on which you want to focus.

Moreover, when you are sketching the graph of a function, either by hand or with a graphing utility, remember that normally you cannot show the entire graph. The decision as to which part of the graph you choose to show is often crucial.

Here we will show how you can use calculus to prepare graphs (on example of rational functions.)

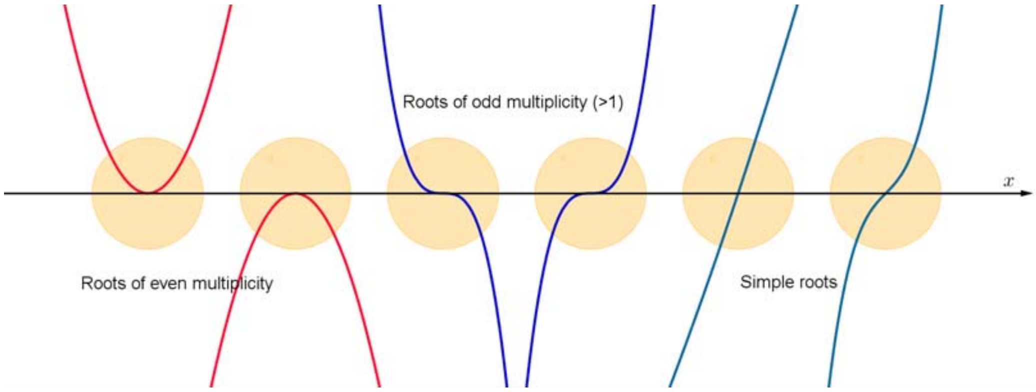


Fig. 8.1. Roots of polynomials with different multiplicities.

8.1 Geometric implications of multiplicity

Our goal in this section is to outline a general procedure that can be used to analyze and graph polynomials in the vicinity of its roots. For example, it would be nice to know what property of the polynomial produce the inflection point and horizontal tangent for the rational functions.

Recall that a root $x = r$ of a polynomial $p(x)$ has multiplicity m if $(x-r)^m$ divides $p(x)$ but $(x-r)^{m+1}$ does not. A root of multiplicity 1 is called a *simple root*. Figure 8.1 and the following theorem show that the behavior of a polynomial in the vicinity of a real root is determined by the multiplicity of that root (we omit the proof).

Theorem 8.1 (The geometric implications of multiplicity)
Suppose that $p(x)$ is a polynomial with a root of multiplicity m at $x = r$.

a) *If m is even, then the graph of $y = p(x)$ is tangent to the x-axis at $x = r$, does not cross the x-axis there, and does not have an inflection point there.*

b) *If m is odd and greater than 1, then the graph is tangent to the x-axis at $x = r$, crosses the x-axis there, and also has an inflection point there.*

c) *If $m = 1$ (so that the root is simple), then the graph is not tangent to the x-axis at $x = r$, crosses the x-axis there, and may or may not have an inflection point there.*

8.2 Graphing rational functions

Recall that a rational function is a function of the form $f(x) = P(x)/Q(x)$ in which $P(x)$ and $Q(x)$ are polynomials. Graphs of rational functions are more complicated than those of polynomials because of the possibility of asymptotes and discontinuities. If $P(x)$ and $Q(x)$ have no common factors, then the information obtained in the following steps will usually be sufficient to obtain an accurate sketch of the graph of a rational function.

Graphing a Rational Function $f(x) = P(x)/Q(x)$ (if $P(x)$ and $Q(x)$ have no Common Factors)

1. (**symmetries**). Determine whether there is symmetry about the y-axis or the origin.
2. (**x- and y-intercepts**). Find the x- and y-intercepts.
3. (**vertical asymptotes**). Find the values of x for which $Q(x) = 0$. The graph has a vertical asymptote at each such value.
4. (**sign of f(x)**). The only places where $f(x)$ can change sign are at the x-intercepts or vertical asymptotes. Mark the points on the x-axis at which these occur and calculate a sample value of $f(x)$ in each of the open intervals determined by these points. This will tell you whether $f(x)$ is positive or negative over that interval.
5. (**end behavior**). Determine the end behavior of the graph by computing the limits of $f(x)$ as $x \to \infty$ and as $x \to -\infty$. If either limit has a finite value L, then the line $y = L$ is a horizontal asymptote.
6. (**derivatives**). Find $f'(x)$ and $f''(x)$.
7. (**conclusions and graph**). Analyze the sign changes of $f'(x)$ and $f''(x)$ to determine the intervals where $f(x)$ is increasing, decreasing, concave up, and concave down. Determine the locations of all stationary (i.e. critical) points, relative extrema, and inflection points. Use the sign analysis of $f(x)$ to determine the behavior of the graph in the vicinity of the vertical asymptotes. Sketch a graph of f that exhibits these conclusions.

Example 8.2 *Sketch a graph of the function*

$$y(x) = \frac{2x^2 - 8}{x^2 - 16} \tag{8.1}$$

and identify the locations of the intercepts, relative extrema, inflection points, and asymptotes.

Table 8.1. Sign analysis of $y(x) = \frac{2x^2-8}{x^2-16}$

Interval	**Test point**	**Value of y**	**Sign of y**
$(-\infty, -4)$	-5	$14/3$	$+$
$(-4, -2)$	-3	$-10/7$	$-$
$(-2, 2)$	0	$1/2$	$+$
$(2, 4)$	3	$-10/7$	$-$
$(4, +\infty)$	5	$14/3$	$+$

Solution: The numerator and denominator have no common factors, so we will use the procedure just outlined.

1. *Symmetries:* Replacing x by $-x$ does not change the equation, so the graph is symmetric about the y-axis.

2. *x- and y-intercepts*: Setting $y = 0$ yields the x-intercepts $x = -2$ and $x = 2$. Setting $x = 0$ yields the y-intercept $y = 1/2$

3. *Vertical asymptotes*: We observed above that the numerator and denominator of y have no common factors, so the graph has vertical asymptotes at the points where the denominator of y is zero, namely, at $x = -4$ and $x = 4$.

4. *Sign of y* : The set of points where x-intercepts or vertical asymptotes occur is $\{-4, -2, 2, 4\}$. These points divide the x-axis into the open intervals

$$(-\infty, -4), \quad (-4, -2), \quad (-2, 2), \quad (2, 4), \quad (4, +\infty)$$

We can find the sign of y on each interval by choosing an arbitrary test point in the interval and evaluating $y = f(x)$ at the test point (Table 8.1). This analysis is summarized on the first line of Table 8.2.

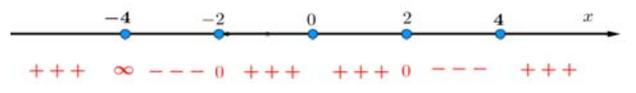
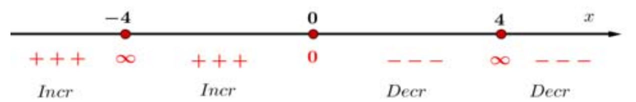
5. *End behavior*: The limits

$$\lim_{x\to\infty} \frac{2x^2 - 8}{x^2 - 16} = \lim_{x\to\infty} \frac{2 - \frac{8}{x^2}}{1 - \frac{16}{x^2}} = 2$$

$$\lim_{x\to-\infty} \frac{2x^2 - 8}{x^2 - 16} = \lim_{x\to-\infty} \frac{2 - \frac{8}{x^2}}{1 - \frac{16}{x^2}} = 2$$

yield the horizontal asymptote $y = 2$.

Table 8.2. Graphs for sign analysis of $y(x), y'(x)$ and $y''(x)$

a)		sign of $y(x)$
b)		sign of $y'(x)$
c)		sign of $y''(x)$

6. *Derivatives*:

$$y'(x) = 4\frac{x}{x^2-16} - 2\frac{x}{(x^2-16)^2}\left(2x^2-8\right) = -48\frac{x}{(x^2-16)^2}$$

$$y''(x) = \frac{48}{(x^2-16)^3}\left(3x^2+16\right) \quad \text{(verify)}$$

7. *Conclusions and graph*:

- The sign analysis of y in Table 8.2 reveals the behavior of the graph in the vicinity of the vertical asymptotes: The graph increases without bound as $x \to -4^-$ and decreases without bound as $x \to -4^+$; and the graph decreases without bound as $x \to 4^-$ and increases without bound as $x \to 4^+$ (Table 8.2).
- The sign analysis of $y'(x)$ in Table 8.2 shows that the graph is increasing to the left of $x = 0$ and is decreasing to the right of $x = 0$. Thus, there is a relative maximum at the stationary point $x = 0$. There are no relative minima.
- The sign analysis of $y''(x)$ in Table 8.2 shows that the graph is concave up to the left of $x = -4$, is concave down between $x = -4$ and $x = 4$, and is concave up to the right of $x = 4$. There are no inflection points.

The graph is shown in Figure 8.2

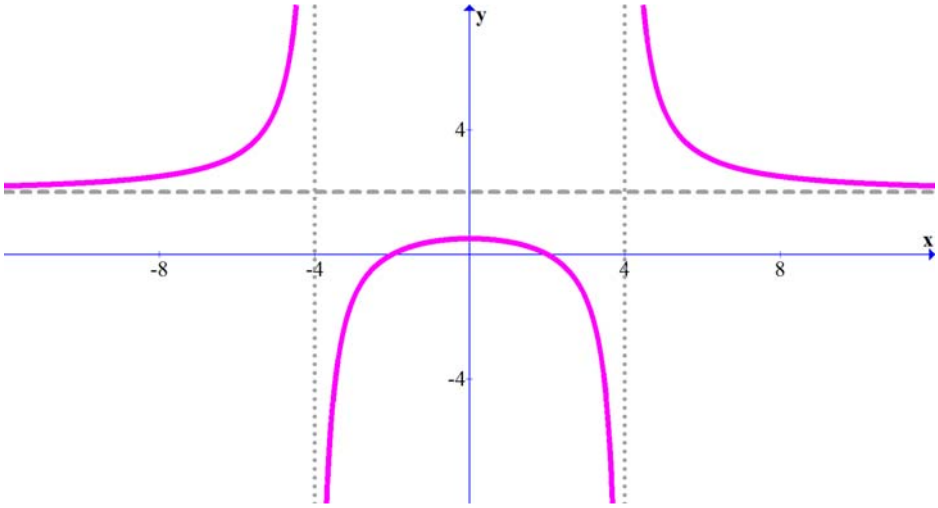


Fig. 8.2. The graph of the function $y(x) = \frac{2x^2-8}{x^2-16}$.

Remark 8.3 *The procedure we stated for graphing a rational function$P(x)/Q(x)$ applies only if the polynomials $P(x)$ and $Q(x)$ have no common factors. How would you find the graph if those polynomials have common factors?*

Example 8.4 *Sketch a graph of*

$$y(x) = \frac{x^2 - 1}{x^3} \tag{8.2}$$

and identify the locations of all asymptotes, intercepts, relative extrema, and inflection points.

Solution: The numerator and denominator have no common factors, so we will use the procedure outlined previously.

1. *Symmetries:* Replacing x by $-x$ and y by $-y$ yields an equation that simplifies to the original equation, so the graph is symmetric about the origin.

2. *x- and y-intercepts*: Setting $y = 0$ yields the x-intercepts $x = -1$ and $x = 1$. Setting $x = 0$ leads to a division by zero, so there is no y-intercept.

3. *Vertical asymptotes*: Setting $x^3 = 0$ yields the solution $x = 0$. This is not a root of $x^2 - 1$, so $x = 0$ is a vertical asymptote.

Table 8.3. Sign analysis of $y(x) = \frac{x^2-1}{x^3}$

Interval	**Test point**	**Value of y**	**Sign of y**
$(-\infty,-1)$	-2	$-3/8$	$-$
$(-1,0)$	$-\frac{1}{2}$	6	$+$
$(0,1)$	$\frac{1}{2}$	-6	$-$
$(1,+\infty)$	2	$3/8$	$+$

4. *Sign of y* : The set of points where x-intercepts or vertical asymptotes occur is $\{-1,0,1\}$. These points divide the x-axis into the open intervals

$$(-\infty,-1), \quad (-1,0), \quad (0,1), \quad (1,+\infty)$$

We can find the sign of y on each interval by choosing an arbitrary test point in the interval and evaluating $y = f(x)$ at the test point (Table 8.1). This analysis is summarized on the first line of Table 8.3

5. *End behavior*: The limits

$$\lim_{x\to\infty} \frac{x^2-1}{x^3} = \lim_{x\to\infty}\left(\frac{1}{x} - \frac{1}{x^3}\right) = 0$$

$$\lim_{x\to-\infty} \frac{2x^2-8}{x^2-16} = \lim_{x\to-\infty}\left(\frac{1}{x} - \frac{1}{x^3}\right) = 0$$

yield the horizontal asymptote $y = 0$.

6. *Derivatives*:

$$y'(x) = -\frac{1}{x^4}\left(x^2-3\right) = -\frac{1}{x^4}\left(x-\sqrt{3}\right)\left(x+\sqrt{3}\right)$$

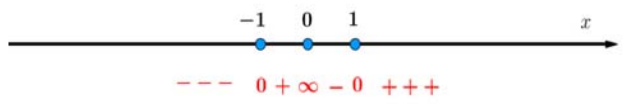
$$y''(x) = \frac{2}{x^5}\left(x^2-6\right) = \frac{2}{x^5}\left(x-\sqrt{6}\right)\left(x+\sqrt{6}\right) \quad \text{(verify)}$$

7. *Conclusions and graph*:

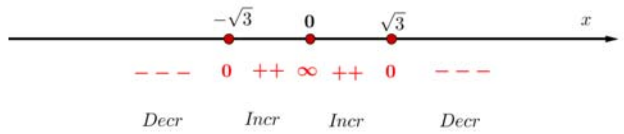
   - The sign analysis of y in Table 8.4 reveals the behavior of the graph in the vicinity of the vertical asymptote $x = 0$. The graph increases without bound as $x \to 0^-$ and decreases without bound as $x \to 0^+$; (see Figure 8.3).
   - The sign analysis of $y'(x)$ in Table 8.4 shows that there is a relative minimum at $x = -\sqrt{3}$ and a relative maximum at $x = \sqrt{3}$.
   - The sign analysis of $y''(x)$ in Table 8.4 shows that the graph shows that the graph changes concavity at the vertical asymptote $x = 0$ and that there are inflection points at $x = -\sqrt{6}$ and $x = \sqrt{6}$.

The graph is shown in Figure 8.3

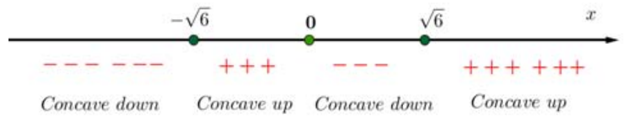
Table 8.4. Graphs for sign analysis of $y(x), y'(x)$ and $y''(x)$

a) 

sign of $y(x)$

b) 

sign of $y'(x)$

c) 

sign of $y''(x)$

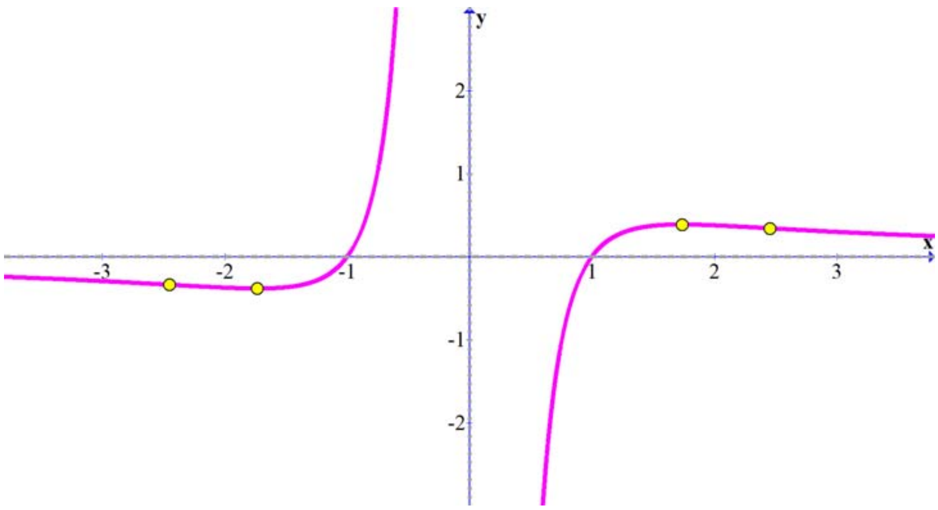


Fig. 8.3. The graph of the function $y(x) = \frac{x^2-1}{x^3}$.

8.3 Rational functions with oblique or curvilinear asymptotes

In the rational functions of Examples 8.2 and 8.4, the degree of the numerator did not exceed the degree of the denominator, and the asymptotes were either vertical or horizontal. If the numerator of a rational function has greater degree than the denominator, then other kinds of "asymptotes" are possible. For example, consider the rational functions

$$f(x) = \frac{x^2+1}{x} \qquad \text{and} \qquad g(x) = \frac{x^3-x^2-8}{x-1} \tag{8.3}$$

By division we can rewrite these as

$$f(x) = x + \frac{1}{x} \qquad \text{and} \qquad g(x) = x^2 - \frac{8}{x-1}.$$

Since the second terms both approach 0 as $x \to +\infty$ or as $x \to -\infty$, it follows that

$$(f(x) - x) \to 0 \quad \text{as} \quad x \to +\infty \quad \text{or as} \quad x \to -\infty$$
$$(g(x) - x^2) \to 0 \quad \text{as} \quad x \to +\infty \quad \text{or as} \quad x \to -\infty.$$

Geometrically, this means that the graph of $y = f(x)$ eventually gets closer and closer to the line $y = x$ as $x \to +\infty$ or as $x \to -\infty$. The line $y = x$ is called an *oblique* or *slant asymptote* of f. Similarly, the graph of $y = g(x)$ eventually gets closer and closer to the parabola $y = x2$ as $x \to +\infty$ or as $x \to -\infty$.. The parabola is called a *curvilinear asymptote* of g. The graphs of the functions in (8.3) are shown in Figures 8.4 and 8.5.

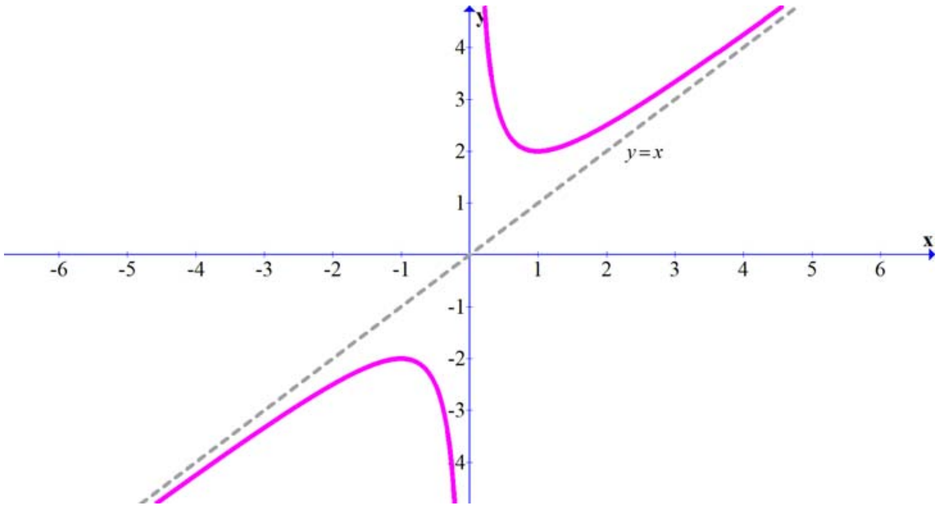


Fig. 8.4. The graph of the function $y(x) = \frac{x^2-1}{x^3}$.

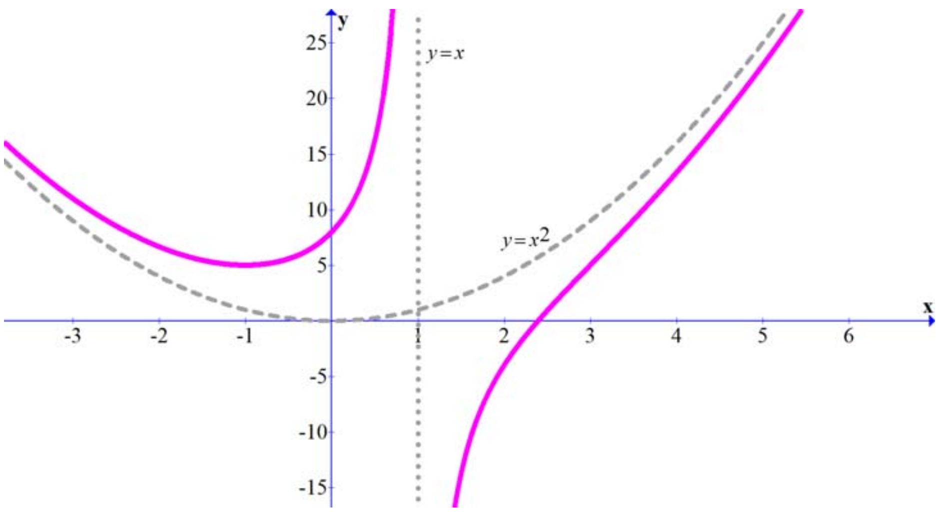


Fig. 8.5. The graph of the function $g(x) = \frac{x^3-x^2-8}{x-1}$.

9

Optimization and linearization

We're now going to look at two practical applications of calculus: optimization and linearization. Believe it or not, these techniques are used every day by engineers, economists, and doctors, for example. Basically, optimization involves finding the best situation possible, whether that be the cheapest way to build a bridge without it falling down or something as mundane as finding the fastest driving route to a specific destination.

Generally such a problem will have the following mathematical form: Find the largest (or smallest) value of $f(x)$ when $a \leq x \leq b$. Sometimes a or b are infinite, but frequently the real world imposes some constraint on the values that x may have. Such a problem differs in two ways from the local maximum and minimum problems we encountered when graphing functions: We are interested only in the function between a and b, and we want to know the largest or smallest value that $f(x)$ takes on, not merely values that are the largest or smallest in a small subinterval. That is, we seek not a local maximum or minimum but a *global maximum* or *minimum*, sometimes also called an absolute maximum or minimum.

Any global maximum or minimum must of course be a local maximum or minimum. If we find all possible local extrema, then the global maximum, if it exists, must be the largest of the local maxima and the global minimum, if it exists, must be the smallest of the local minima. We already know where local extrema can occur: only at those points at which $f'(x)$ is zero or undefined. Actually, there are two additional points at which a maximum or minimum can occur if the endpoints a and b are not infinite, namely, at a and b.

On the other hand, linearization is a useful technique for finding approximate values of hard-to calculate quantities. It can also be used to find approximate values of zeroes of functions; this is called Newton's method.

9.1 Three types of optimization problems

We shall first discuss the theory, and then look at some serious examples. Optimization problems are essentially always "word problems", which are also called "modeling problems", since we are taking a real-world problem and modeling it by a pure mathematics problem.

There are three types of optimization problems that we deal with. In all of them, we assume that we have a continuous function f, whose domain is an interval I, and that f has a finite number of critical points, where f' is zero or undefined, on the interior of the interval I.

Optimization Case 1: The interval I is a closed, bounded interval $[a, b]$.

In this case, we have a theorem which tells us how to proceed.

Theorem 9.1 *The function f attains a global maximum and a global minimum value on I. These extreme values are attained at critical points of f (on the interval I); thus, they are attained either when $x = a$, $x = b$, or when x is in the open interval (a, b) and $f'(x) = 0$ or $f'(x)$ is undefined. Therefore, we can find the global extreme values of f by making a table of x and $f(x)$ values, where the x values are a, b, and all of the x-coordinates in (a, b) where $f'(x) = 0$ or $f'(x)$ is undefined. Then, we simply look at the $f(x)$ values and select the largest and smallest ones.*

Proof. Proof. This follows immediately from the Extreme Value Theorem, and other theorems of the previous Chapter. ■

Remark 9.2 *You need to be careful when looking at critical points where $f'(x) = 0$. Frequently, we have a function F that is defined and differentiable on a larger set than the interval $[a, b]$, but physical constraints require that x be in the closed interval $[a, b]$; so, our function f is actually the function F, restricted to the interval $[a, b]$. This function f is technically a different function from F, even though the rule specifying them is the same. The danger is that the formula you derive for the derivative will be $F'(x)$, which maybe defined and equal to zero outside of the interval $[a, b]$. The points where $F'(x) = 0$ that are outside $[a, b]$ are not critical points of f, and do not belong in the table of x and $f(x)$ values that you use to find the maximum and minimum values of f on $[a, b]$.*

Example 9.3 *Find the extreme values of $f(x) = 2x^3 - 3x^2 - 12x$ on the interval $[0, 3]$.*

Solution: We find

$$f'(x) = 6x^2 - 6x - 12 = 6(x^2 - x - 2) = 6(x - 2)(x + 1).$$

When does $f'(x) = 0$? When $x = 2$ and when $x = -1$. However, $x = -1$ is outside the interval under consideration, and so we ignore it. We find one critical point, $x = 2$, of f, inside the open interval $(0; 3)$, and we also have to check the endpoints of the closed interval.

Thus, we make the small table

x	0	0	3
$f(x)$	2	-20	-9

Therefore, the maximum value of f on the interval $[0, 3]$ is 0, which occurs at the single x-coordinate $x = 0$, and the minimum value of f on the interval $[0, 3]$ is -20, which occurs at the single x-coordinate $x = 2$. □

Optimization Case 2: The interval I is not closed and bounded, but the First or Second Derivative Test applies.

Example 9.4 *Suppose that we need to construct an aluminum can that holds* 500 *cubic centimeters (*$0,5$ *liters). The can is required to have the shape of a right circular cylinder. We assume that the aluminum has some uniform small thickness, so that it is reasonable to describe the amount of aluminum using area units. Suppose that the cost per square centimeter of aluminum for the sides of the can is some constant* $c > 0$ *dollars, while the top and bottom of the can use aluminum that costs* $2c$ *dollars per square centimeter. Find the dimensions (radius and height) of the can that minimizes the cost.*

Solution: Let r denote the radius of the can, and h the height, both in centimeters. The volume of the can is the area of the base times the height, i.e., $\pi r^2 h$; this is required to equal 500. Therefore, we have

$$\pi r^2 h = 500.$$

This type of equation is known as a *constraint*, because it implies that r and h may not vary independently; the allowable $(r; h)$ pairs are *constrained* by the equation. We are trying to minimize T, the total cost of the can in dollars. The cost of the sides of the can is the area of the sides times the cost per unit area, i.e., $2\pi rh \cdot c$. The cost of the top and bottom of the can is the area of the top and bottom times the cost per unit area, i.e., $2\pi r^2 \cdot 2c$. Thus, the total cost of the can is

$$T = 2\pi rhc + 4\pi r^2 c.$$

We need a function of one variable to apply our methods; that is, we need to write T as a function of either r or h, but not both. As we mentioned above, r and h are not allowed to vary independently. So, at this point, we solve the constraint equation for h, and write T as a function of the single variable r. We find

$$h = \frac{500}{\pi r^2}$$

and

$$T(r) = 4c\left(\pi r^2 + \frac{250}{r}\right)$$

The domain of this function is all $r > 0$.

We calculate the derivative

$$T'(r) = -4c\left(\frac{250}{r^2} - 2\pi r\right),$$

and see that it equals 0 when $\left(\frac{250}{r^2} - 2\pi r\right) = 0$, i.e., when $2\pi r = \dfrac{250}{r^2}$; this means that

$$r^3 = \frac{125}{\pi},$$

i.e., $r = \dfrac{5}{\sqrt[3]{\pi}}$. Hence, T has a single critical point at $r = \dfrac{5}{\sqrt[3]{\pi}}$ on the interval $(0, \infty)$. The 2nd derivative of T is

$$T''(r) = 4c\left(2\pi + \frac{500}{r^3}\right)$$

which is positive for $r > 0$. Thus, the Second Derivative Test tells us that T attains a global minimum value at $r = \dfrac{5}{\sqrt[3]{\pi}}$.

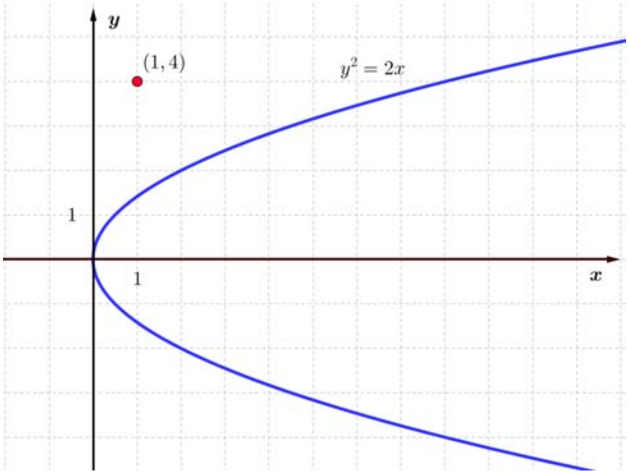
To find the corresponding height, we go back to

$$h = \frac{500}{\pi r^2} = \frac{500}{\pi\left(\frac{5}{\sqrt[3]{\pi}}\right)^2} = \frac{20}{\sqrt[3]{\pi}} = 4r.$$

Finally, we conclude that the dimensions of the can that minimize the total cost of the can are $r = \dfrac{5}{\sqrt[3]{\pi}}$ centimeters $h = \frac{20}{\sqrt[3]{\pi}}$ centimeters, which means that the height of the can should be twice the diameter of the can. □

Optimization Case 3: We are told, or physical reasons imply, that an extreme value is attained, and there is only one critical point.

Example 9.5 *Find the point on the parabola* $y^2 = 2x$ *that is closest to the point* $(1, 4)$.



Solution: The distance between the point (x, y) and the point $(1, 4)$ is

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

But if (x, y) lies on the parabola, then $x = y^2/2$, so the expression for d becomes

$$d = \sqrt{(y^2/2-1)^2 + (y-4)^2}.$$

Instead of minimizing d, we minimize its square:

$$d^2 = f(y) = (y^2/2-1)^2 + (y-4)^2.$$

(You should convince yourself that the minimum of d occurs at the same point as the minimum of d^2, but d^2 is easier to work with.) Differentiating, we obtain

$$f'(y) = y^3 - 8,$$

so $f'(y) = 0$ when $y = 2$. Because of the geometric nature of the problem, it's obvious that there is a closest point but not a farthest point. The corresponding value of x is $x = y^2/2 = 2$. Thus, the point on $y^2 = 2x$ closest $(1, 4)$ to is $(2, 2)$. □

9.2 Exemplary optimization problems

In all examples presented below, the challenge is to find an efficient way to carry out a task, where "efficient" could mean least expensive, most profitable, least time consuming, or, as you will see, many other measures.

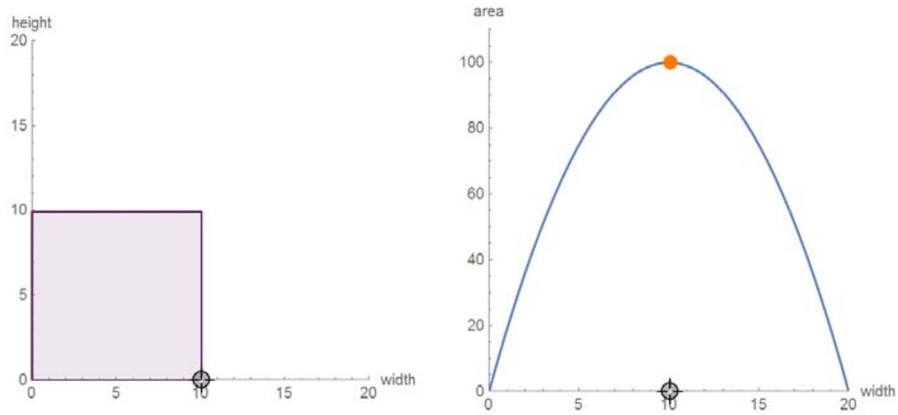


Fig. 9.1. Optimal rectangle with a perimeter of 40 and a maximal surface area

Example 9.6 *Find the rectangle with a perimeter of* 40 *and a maximal possible area.*

Solution: For a rectangle with a perimeter of 40, the height h is always 20 minus the width w. This allows you to reduce the formula for the area, $Area = wh$, to $Area = (20 - w)w = 20w - w^2$. Elementary calculus shows that this formula is at a maximum when $w = 10$ (see Figure 9.1).

The condition that $h = (20 - w)$ is called a *constraint*. It tells us to consider only (nonnegative) values of h and w satisfying this equation. It allows us to reduce the *primary equation* $Area = wh$ to one with a single independent variable. For the physical reasons we can suppose that (additionally) $w \leq 20$. The quantity that we wish to maximize (or minimize in other cases) is called the *objective function*; in this case, the objective function is the product $Area = wh$. □

Example 9.7 *Suppose you want to reach a point A that is located across the sand from a nearby road (see Figure 9.2). Suppose that the road is straight, and b is the distance from A to the closest point C on the road. Let v be your speed on the road, and let w, which is less than v, be your speed on the sand. Right now you are at the point D, which is a distance a from C. At what point B should you turn off the road and head across the sand in order to minimize your travel time to A?*

Solution: Let x be the distance short of C where you turn off, i.e., the distance from B to C. We want to minimize the total travel time. Recall that when traveling at constant velocity, time is distance divided by velocity. You

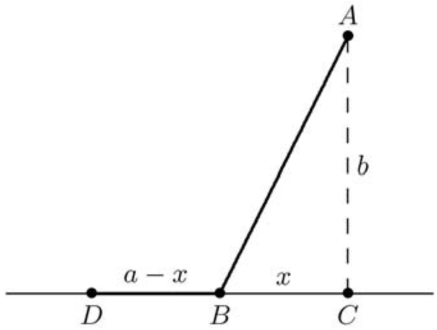


Fig. 9.2. Minimizing travel time.

travel the distance $\overline{DB}$ at speed v, and then the distance $\overline{BA}$ at speed w. Since $\overline{DB} = a - x$ and, by the Pythagorean theorem, $\overline{BA} = \sqrt{x^2 + b^2}$, the total time for the trip is

$$f(x) = \frac{a - x}{v} + \frac{\sqrt{x^2 + b^2}}{w}.$$

We want to find the minimum value of f when x is between 0 and a. As usual we set $f'(x) = 0$ and solve for x:

$$0 = f'(x) = -\frac{1}{v} + \frac{x}{w\sqrt{x^2 + b^2}}$$

$$w\sqrt{x^2 + b^2} = vx$$

$$w^2\left(x^2 + b^2\right) = v^2x^2$$

Notice that a does not appear in the last expression, but a is not irrelevant, since we are interested only in critical values that are in $[0, a]$, and $wb/\sqrt{v^2 - w^2}$ is either in this interval or not. If it is, we can use the second derivative to test it:

$$f''(x) = \frac{b^2}{(x^2 + b^2)^{3/2}w},$$

If the critical value is not in $(0, a)$ it is greater or equal to a. In this case the minimum must occur at one of the endpoints. We can compute$f(0) = \frac{a}{v} + \frac{b}{w}$,

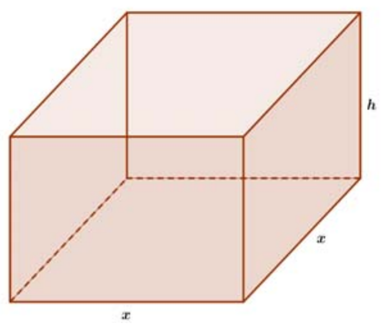
$$f(a) = \frac{\sqrt{a^2 + b^2}}{w}$$

but it is difficult to determine which of these is smaller by direct comparison. If, as is likely in practice, we know the values of v, w, a, and b, then it is easy to determine this. With a little cleverness, however, we can determine the minimum in general. We have seen that $f''(x)$ is always positive, so the derivative $f'(x)$ is always increasing. We know that at $wb/\sqrt{v^2 - w^2}$ the derivative is zero, so for values of x less than that critical value, the derivative is negative. This means that $f(0) > f(a)$, so the minimum occurs when $x = a$. So the upshot is this: If you start farther away from C than $wb/\sqrt{v^2 - w^2}$ then you always want to cut across the sand when you are a distance $wb/\sqrt{v^2 - w^2}$ from point C. If you start closer than this to C, you should cut directly across the sand. With two examples providing some insight, we present a procedure for solving optimization problems. These guidelines provide a general framework, but the details may vary depending upon the problem. □

Guidelines for solving applied minimum and maximum problems:

1. Identify all given quantities and all quantities to be determined.
2. If possible, make a sketch.
3. Write a primary equation for the quantity to be maximized or minimized.
4. Reduce the primary equation to one with a single independent variable. This may involve the use of secondary equations relating the independent variables of the primary equation.
5. Determine the domain of the primary equation.
6. Use calculus to determine the desired maximum or minimum value.
7. Use calculus to verify the answer.

Example 9.8 *A manufacturer wants to design an open box with a square base and a surface area of* 108 *square inches. What dimensions will produce a box with maximum volume?*



Solution: Draw the box and label the length of the square base x and the height h. The volume is given by the primary equation $V = hx^2$ The surface area of the open box consists of the bottom and 4 sides: $x^2 + 4xh = 108$. We can use this constraint to eliminate the variable in the volume formula as follows:

$$4xh = 108 - x^2,$$

$$h = \frac{108 - x^2}{4x},$$

$$V(x) = hx^2 = \left(\frac{108 - x^2}{4x}\right)x^2 = -\frac{1}{4}x\left(x^2 - 108\right) = 27x - \frac{1}{4}x^3.$$

The domain of this function is $0 \leq x \leq \sqrt{108}$. We now use our calculus skills to find the maximum value of V.

$$V'(x) = 27 - \frac{3}{4}x^2 = 0$$

$$3x^2 = 108$$

$$x = 6.$$

The critical number is $x = 6$. Since $V(0) = V(\sqrt{108}) = 0$, $V(6) = 108$ is the maximum value on the interval. For $x = 6$ you obtain $h = 3$. The maximum volume is 108 cubic inches, and the dimensions are $6 \times 6 \times 3$. Note that you could have used the first or second derivative tests to verify that 6 gave a maximum. □

Example 9.9 *A window in the shape of rectangle capped by a semicircle is to have perimeter p. Choose the radius of the semicircular part so that the window admits the most light.*

Solution: We take the point of view that the window which admits the most light is the one with maximum area. As in Figure above,we let x be the radius of the semicircular part and y be the height of the rectangular part. We want to express the area

$$A = \frac{1}{2}\pi r^2 + 2xy$$

as a function of x alone. To do this, we must express y in terms of x.

Since the perimeter is p, we have

$$p = 2x + 2y + \pi x$$

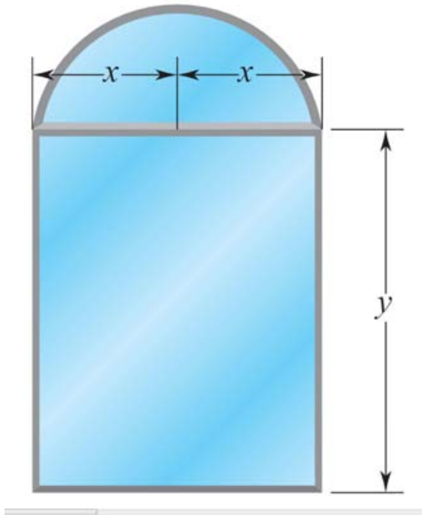


Fig. 9.3.

and thus

$$y = \frac{1}{2}\left[p - (2+\pi)x\right].$$

Since x and y represent lengths, these variables must be nonnegative. For both x and y to be nonnegative, we must have $0 \le x \le p/(2+\pi)$. The area can now be expressed in terms of x alone:

$$\begin{aligned} A(x) &= \frac{1}{2}\pi r^2 + 2xy \\ &= \frac{1}{2}\pi r^2 + 2x \cdot \frac{1}{2}\left[p - (2+\pi)x\right] \\ &= px - \left(2 + \frac{\pi}{2}\right)x^2. \end{aligned}$$

We want to maximize the function

$$A(x) = px - \left(2 + \frac{\pi}{2}\right)x^2, \qquad 0 \le x \le p/(2+\pi).$$

The derivative

$$A'(x) = p - 2x\left(\frac{1}{2}\pi + 2\right) = p - 4x - \pi x$$

is 0 only at $x = p/(4+\pi)$. Since $A(0) = A[p/(2+\pi)] = 0$, and since $A(x) > 0$ for $0 < x < p/(4+\pi)$ and $A(x) < 0$ for $p/(4+\pi) < x < p/(2+\pi)$, the function

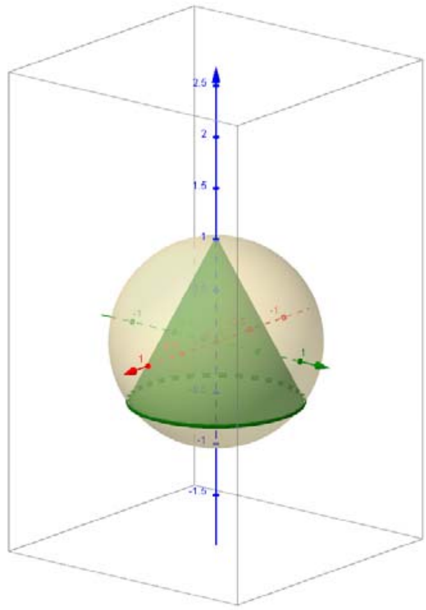


Fig. 9.4. Cone in a sphere ($R = 1$).

A is maximized by setting $x = p/(4+\pi)$. For the window to have maximum area, the radius of the semicircular part must be $p/(4+\pi)$. □

Optimization problems are difficult because they require formulas from precalculus, such as areas, volumes, and trigonometric relationships. Here is an example.

Example 9.10 *If you fit the largest possible (in volume) cone inside a sphere, what fraction of the volume of the sphere is occupied by the cone? (Here by "cone" we mean a right circular cone, i.e., a cone for which the base is perpendicular to the axis of symmetry, and for which the cross-section cut perpendicular to the axis of symmetry at any point is a circle- see Figure 9.4)*

Solution: Let R be the radius of the sphere, and let r and h be the base radius and height of the cone inside the sphere. What we want to maximize is the volume of the cone: $\pi r^2 h/3$. Here R is a fixed value, but r and h can vary. Namely, we could choose r to be as large as possible—equal to R—by taking the height equal to R; or we could make the cone's height h larger at the expense of making r a little less than R. See the cross-section depicted in

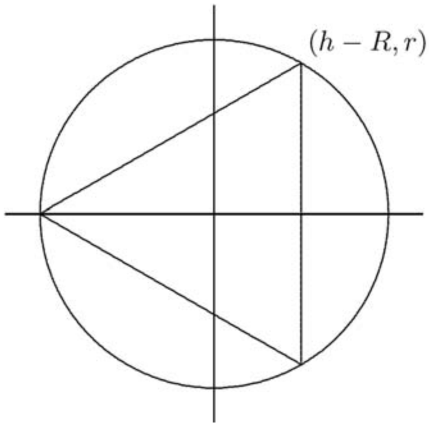


Fig. 9.5. Cone in a sphere (cross-section)..

Figure 9.5. We have situated the picture in a convenient way relative to the x and y axes, namely, with the center of the sphere at the origin and the vertex of the cone at the far left on the x-axis.Notice that the objective function we want to maximize, $\pi r^2 h/3$, depends on two variables. This is frequently the case, but often the two variables are related in some way so that "really" there is only one variable. So our next step is to find the relationship and use it to solve for one of the variables in terms of the other, so as to have a function of only one variable to maximize. In this problem, the condition is apparent in the figure: the upper corner of the triangle, whose coordinates are $(h-R,r)$, must be on the circle of radius R. That is,

$$(h-R)^2 + r^2 = R^2$$

We can solve for h in terms of r or for r in terms of h. Either involves taking a square root, but we notice that the volume function contains r^2, not r by itself, so it is easiest to solve for r^2 directly: $r^2 = R^2 - (h-R)^2$. Then we substitute the result into $\pi r^2 h/3$:

$$\begin{aligned} V(h) &= \pi(R^2 - (h-R)^2)h/3 \\ &= \frac{2}{3}\pi R h^2 - \frac{1}{3}\pi h^3. \end{aligned}$$

We want to maximize $V(h)$ when h is between 0 and $2R$. Now we solve

$$0 = f'(h) = -\pi h^2 + (4/3)\pi h R,$$

getting $h = 0$ or $h = 4R/3$. We compute $V(0) = V(2R) = 0$ and $V(4R/3) = (32/81)\pi R^3$. The maximum is the latter; since the volume of the sphere is

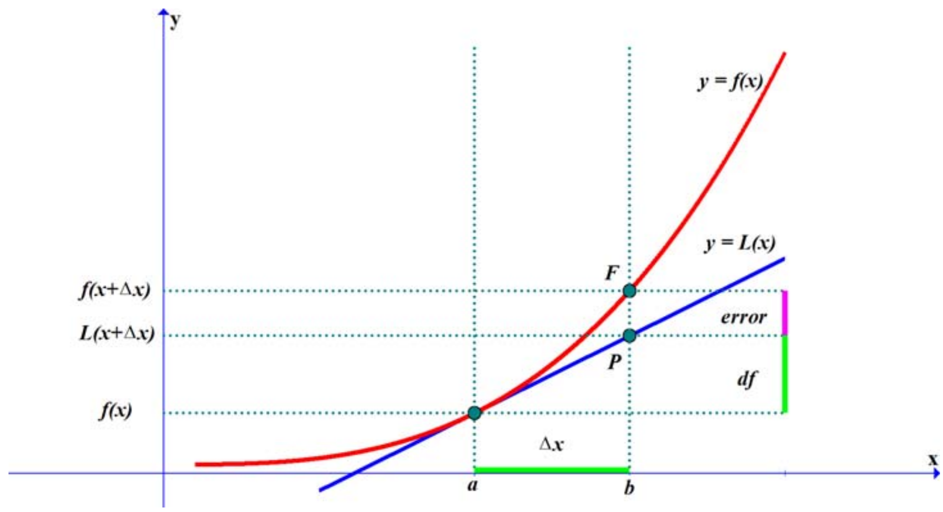


Fig. 9.6.

$(4/3)\pi R^3$, the fraction of the sphere occupied by the cone is

$$\frac{(32/81)\pi R^3}{(4/3)\pi R^3} = \frac{8}{27} \approx 30\%. \qquad \square$$

9.3 Linearization

The graph 9.6 shows the curve $y = f(x)$ and the linearization $y = L(x)$, which is the tangent line to the curve at $x = a$. We want to estimate the value of $f(a + \Delta x)$. That's the height of the point F in the above picture. As an approximate value, we're actually using $L(a + \Delta x)$, which is the height of P in the picture. The difference between the two quantities is labeled "error".

Here's the basic strategy for estimating, or approximating, a nasty number:

1. Write down the main formula

$$\boxed{f(x) \approx L(x) = f(a) + f'(a)(x - a)}$$

2. Choose a function f, and a number x such that the nasty number is equal to $f(x)$. Also, choose a close to x such that $f(a)$ can easily be computed.

3. Differentiate f to find f'.

4. In the above formula, replace f and f' by the actual functions, and a by the actual number you've chosen.

5. Finally, plug in the value of x from step 2 above. Also note that the differential df is the quantity $f'(a)(x-a)$.

Remark 9.11 *We refer to $x = a$ as the* center of the linearization*. Notice that $y = L(x)$ is the equation of the tangent line to the graph of $f(x)$ at $x = a$ (see Figure 9.6).*

Example 9.12 *How would you estimate* $\sin(11\pi/30)$ *using linearization?*

Solution: Start with the standard formula

$$f(x) \approx L(x) = f(a) + f'(a)(x-a).$$

We need to take the sine of something, so let's set $f(x) = \sin(x)$. We are interested in what happens when $x = 11\pi/30$. Now, we need some number a which is close to $11\pi/30$, such that $f(a)$ is nice. Of course, $f(a)$ is just $\sin(a)$. What number close to $11\pi/30$ has a manageable sine? How about $10\pi/30$? After all, that's just $\pi/3$, and we certainly understand $\sin(\pi/3)$. So, set $a = \pi/3$.

We've completed the first two steps. Moving on to the third step, we find that $f'(x) = \cos(x)$, so the linearization formula becomes

$$f(x) \approx L(x) = \sin\left(\frac{\pi}{3}\right) + \cos\left(\frac{\pi}{3}\right)(x - \frac{\pi}{3}).$$

Since $f(x) = \sin(x)$, this simplifies to

$$\sin(x) \approx \frac{1}{2}x - \frac{1}{6}\pi + \frac{1}{2}\sqrt{3}.$$

Finally, put $x = 11\pi/30$ to get

$$\sin(\frac{11}{30}\pi) \approx \frac{11}{60}\pi - \frac{1}{6}\pi + \frac{1}{2}\sqrt{3} = \frac{1}{60}\pi + \frac{1}{2}\sqrt{3} = 0.91839,$$

whereas the exact value $\sin(\frac{11}{30}\pi) = 0.91355$.

Remark 9.13 *Sometimes we compute the* relative percentage error*, which is often more important than the error itself. By definition,*

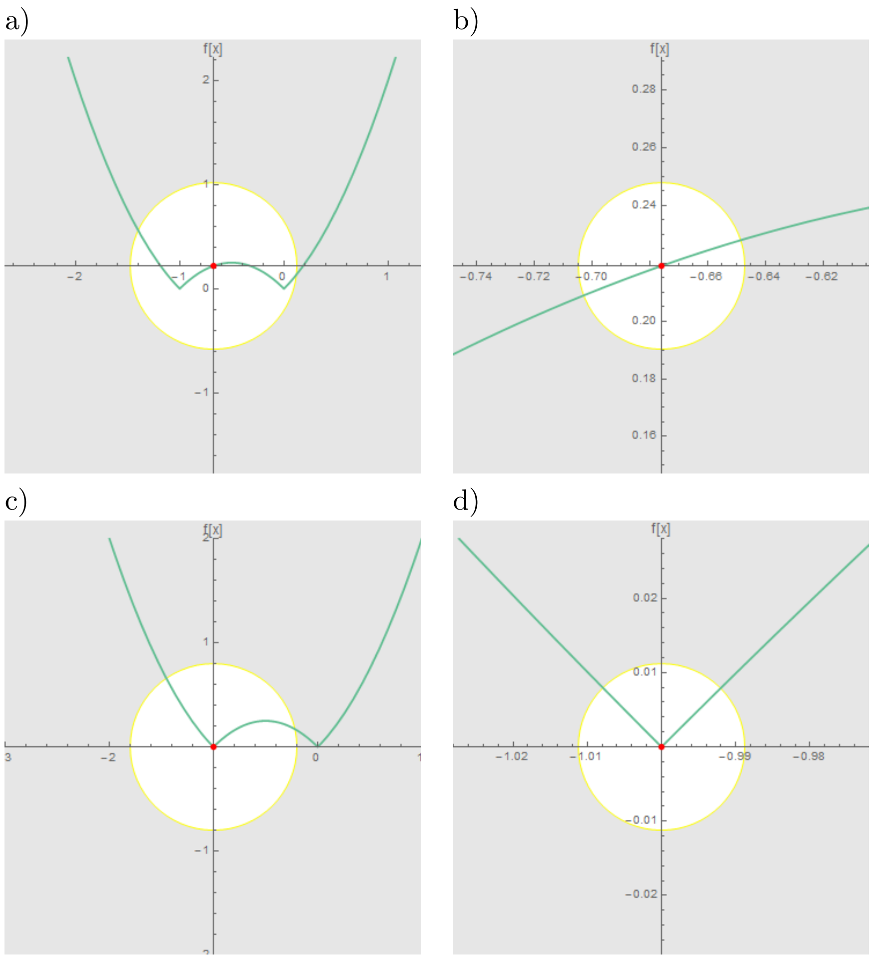
$$\textit{Relative percentage error} = \left|\frac{\textit{error}}{\textit{actual value}}\right| \times 100\%. \qquad (9.1)$$

In the example 9.12

$$\textit{Relative percentage error} = \left|\frac{0.91355 - 0.91839}{0.91355}\right| \times 100\% = 0.5298\%$$

The next example illustrates the fact, that the linearization can be successful only when the function f is differentiable.

Example 9.14 *The "differentiation microscope" lets you graphically explore if a function is differentiable at a given point* x_0*. If you "zoom in" at any point* x_0 *and you see that in a very close neighborhood of* x_0 *the function looks like a linear function, then the function* $f(x)$ *is differentiable at* x_0 *(images a-b) and, otherwise not (images c-d). Can you find points where* $f(x) = |x^2 + x|$ *is differentiable and points where otherwise it is not?*



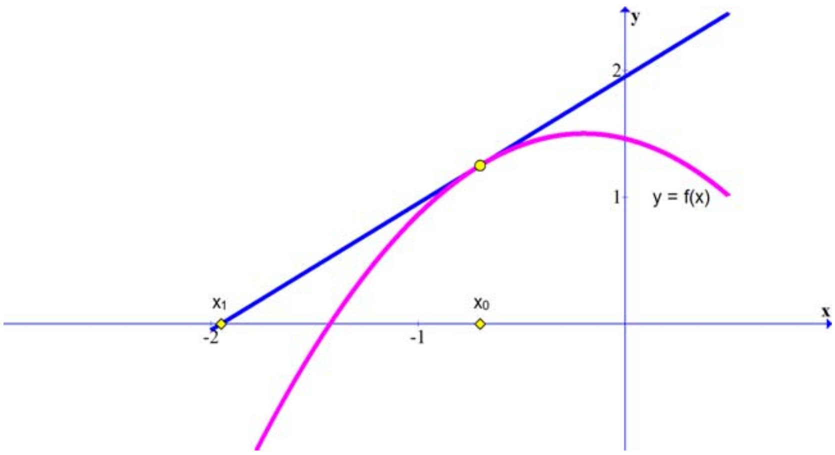


Fig. 9.7. **One step of Newton's Method.** Starting with x_0, the tangent line to the curve $y = f(x)$ is drawn. The intersection point with the x-axis is x_1, the next approximation to the root.

9.4 Newton's method

Suppose you have a differentiable function $f(x)$, and you want to find as accurately as possible where it crosses the x-axis; in other words, you want to solve $f(x) = 0$. We were able to find the roots of functions using a "divide and conquer" (i.e. bisection) technique: start with an interval $[a, b]$ for which $f(a) < 0$ and $f(b) > 0$. If $f((a + b)/2)$ is positive, then use the interval $[a, (a + b)/2]$ otherwise$[(a + b)/2, b]$. After n steps, we are $(b - a)/2n$ close to the root.

If the function f is differentiable, we can do much better. Suppose you know of no way to find an exact solution by any algebraic procedure, but you are able to use an approximation, provided it can be made quite close to the true value. Newton's method is a way to find a solution to the equation to as many decimal places as you want. It is what is called an "*iterative procedure*," meaning that it can be repeated again and again to get an answer of greater and greater accuracy. Iterative procedures like Newton's method are well suited to programming for a computer. Newton's method uses the fact that the tangent line to a curve is a good approximation to the curve near the point of tangency. To find a root of $f(x) = 0$, a starting point x_0 is given, and the tangent line to the function f at x_0 is drawn. The tangent line will approximately follow the function down to the x-axis toward the root. The intersection point of the line with the x-axis is an approximate root, but probably not exact if f curves. Therefore, this step is iterated.

From the geometric picture, we can develop an algebraic formula for Newton's Method. The tangent line at x_0 has slope given by the derivative $f'(x_0)$. One point on the tangent line is $(x_0, f(x_0))$. The point-slope formula for the equation of a line is

$$y - f(x_0) = f'(x_0)(x - x_0),$$

so that looking for the intersection point of the tangent line with the x-axis is the same as substituting $y = 0$ in the line:

$$f'(x_0)(x - x_0) = 0 - f(x_0)$$

$$x - x_0 = -\frac{f(x_0)}{f'(x_0)}$$

$$x = x_0 - \frac{f(x_0)}{f'(x_0)} = T(x_0) \quad \text{where} \quad T(x) = x - \frac{f(x)}{f'(x)}.$$

Solving for x gives an approximation for the root, which we call $x_1 = T(x_0)$. Next, the entire process is repeated, beginning with x_1, to produce $x_2 = T(x_1)$, and so on, yielding the following iterative formula:

$$x_0 = \text{starting point}$$

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} := T(x_i) \qquad \text{for} \quad i = 0, 1, 2, \ldots$$

So, Newton's method is the process of applying map T again and again until we are sufficiently close to the root. If p is a root such that $f'(p) \neq 0$, and x_0 is close enough to p, then $x_1 = T(x_0)$, $x_2 = T(T(x_0))$ converges to the root p. It is an extremely fast method to find the root of a function.

1. If $f(x) = ax + b$, we reach the root in one step.

2. If $f(x) = x^2$ then $T(x) = x - x^2/(2x) = x/2$. We get quite fast to the root 0 but not as fast as the method promises. Indeed, the root 0 is also a critical point of f. This slows us down.

3. If $f(x) = x^2$ then $T(x) = x - x^2/(2x) = x/2$. We get quite fast to the root 0 but not as fast as the method promises. Indeed, the root 0 is also a critical point of f. This slows us down.

4. The Newton method converges extremely fast to a root $f(p) = 0$ if $f'(p) \neq 0$ if we start sufficiently close to the root.

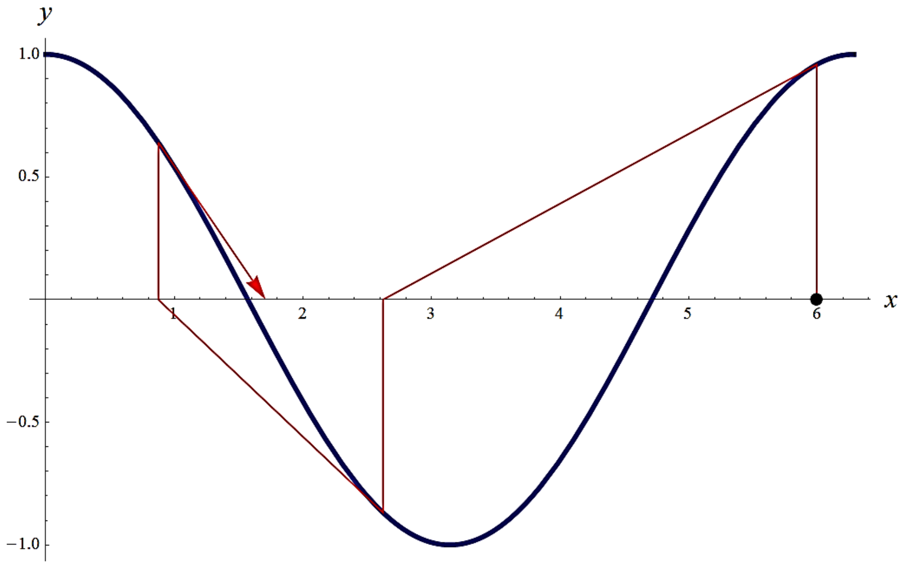


Fig. 9.8. When $x_0 = 6$ the first three iterations of the Newton Method for $f(x) = \cos(x)$ are 6.0, 2.62161, 0.875009, 1.71012.

Figure 9.8 illustrates how the Newton method works when $f(x) = \cos(x)$ and $x_0 = 6.0$.

Newton used this method to find the roots of polynomials. It is amazingly fast: Starting 0.1 close to the point, we have after one step 0.01 after 2 steps 0.0001 after 3 steps 0.00000001 and after 4 steps 0.0000000000000001.

In 10 steps we can get a $2^{10} = 1024$ digits accuracy. Having a fast method to compute roots is useful. For example in computer graphics, where things can not be fast enough. Also in number theory, when working with integers having thousands of digits the Newton method can help. There is much theoretical use of the method. It goes so far as to explain stability of planetary motion or stability of plasma in fusion reactors.

If we have several roots, and we start at some point, to which root will the Newton method converge? Does it at all converge? This is an interesting question. It is also historically intriguing because it is one of the first cases, where "chaos" was observed at the end of the 19'th century.

Example 9.15 *Find the Newton map in the case* $f(x) = x^4 - 1$. *Solution* $T(x) = x - \dfrac{(x^4-1)}{(4x^3)}$. *If we look for roots in the complex like for* $f(x) = x^4 - 1$ *which has* 4 *roots in the complex plane, the "basin of attraction" of each of the roots is a complicated set which we call the Newton fractal. See Figure 9.9*

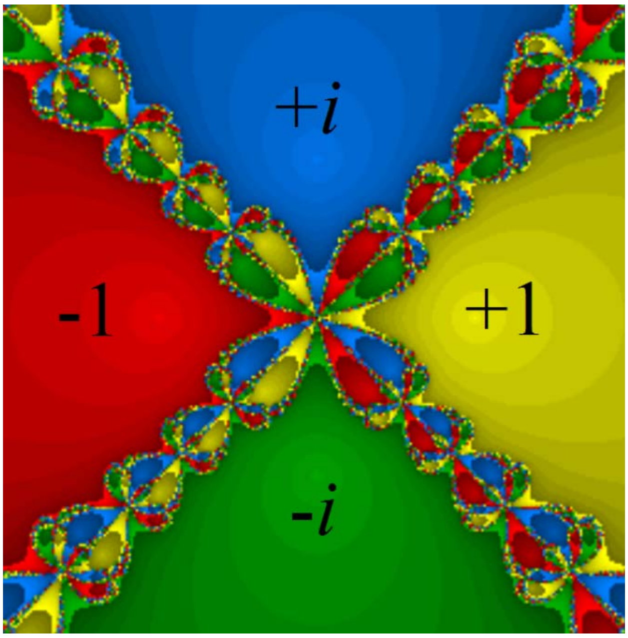


Fig. 9.9. Newton fractal on $z^4 - 1$.

9.4.1 Using Newton's Method to Estimate π

For those of us who enjoy mathematical challenges, estimating irrational numbers like π, e, and $\sqrt{2}$ to a large number - no, to a very large number of decimal places - is an exercise of enjoyment. Let's see if we can crash the computer! To estimate π, we will use Newton's method to estimate the zero of the function $f(x) = \tan(x)$, for $\pi/2 < x < 3\pi/2$. As you know already, the exact answer is π.

We set up Newton's iteration scheme thus. We have set the precision at 50 digits for the estimate of the zero.

Using the graph of the function, we specify a starting value of $x_0 = 3$ for the iterations. We iterate until two consecutive calculated values of x are equal to the precision specified by digits. The symbol n is used to count the number of iterations required to achieve the specified precision.

iteration	approximate root
0	3.0
1	3.142546543074277805295635410533913493226092284901 8
2	3.141592653300476815449885771719913096643593169213 6
3	3.141592653589793238462643383287575197443209871201 7
4	3.141592653589793238462643383279502884197169399375 1
5	3.141592653589793238462643383279502884197169399375 1

All digits are correct!

9.4.2 Using Newton's Method to Estimate $\sqrt{2}$

Lets compute $\sqrt{2}$ to 12 digits accuracy. We want to find a root $f(x) = x^2 - 2$. The Newton
map is $T(x) = x - (x^2 - 2)/(2x) = \frac{1}{2}\left(x + \frac{2}{x}\right)$. Lets start with $x_0 = 1.0$

$$\begin{aligned} T(1) &= 3/2 \\ T(3/2) &= 17/12 \\ T(17/12) &= 577/408 \\ T(577/408) &= 665857/470832 \end{aligned}$$

- This is already $1.6 \cdot 10^{-12}$ close to the real root! 12 digits, by hand!

10
Integrals

We are now at a critical point in the calculus story. Many would argue that this chapter is the cornerstone of calculus because it explains the relationship between the two processes of calculus: differentiation and integration. We begin by explaining why finding the area of regions bounded by the graphs of functions is such an important problem in calculus. Then you will see how antiderivatives lead to definite integrals, which are used to solve this problem. But there is more to the story. You will also see the remarkable connection between derivatives and integrals, which is expressed in the Fundamental Theorem of Calculus. In this chapter, we develop key properties of definite integrals, investigate a few of their many applications, and present the first of several powerful techniques for evaluating definite integrals.

10.1 Antiderivatives and Basic Integration Rules

Suppose you were asked to find a function F whose derivative is $f(x) = 4x^3$. From your knowledge of derivatives, you would probably say that

$$F(x) = x^4 \quad \text{because} \quad \frac{d}{dx}\left[x^4\right] = 4x^3.$$

The function F is an antiderivative of f.

Definition 10.1 (of antiderivative) *A function F is an antiderivative of f on an interval I if for $F'(x) = f(x)$ all x in I.*

Note that F is called *an* antiderivative of f rather than *the* antiderivative of f. To see why, observe that

$$F_1(x) = x^4, \quad F_2(x) = x^4 - 7, \quad \text{and} \quad F_3(x) = x^4 + 11$$

are all antiderivatives of $f(x) = 4x^3$. In fact, for any constant C, the function given by $F(x) = x^4 + C$ is an antiderivative of f.

Theorem 10.2 (Representation of antiderivatives) *If F is an antiderivative of f on an interval I, then G is an antiderivative of f on the interval I*

if and only if G is of the form $G(x) = F(x) + C$, for all x in I where C is a constant.

Proof. The proof of Theorem 10.2 in one direction is straightforward. That is, if $G(x) = F(x) + C$, $F'(x) = f(x)$, and C is a constant, then

$$G'(x) = \frac{d}{dx}[F(x) + C] = F'(x) + 0 = f(x).$$

To prove this theorem in the other direction, assume that G is an antiderivative of f. Define a function H such that

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

For any two points a and b $(a < b)$ in the interval, H is continuous on and differentiable on $[a, b]$. By the Mean value theorem,

$$H'(c) = \frac{H(b) - H(a)}{b - a}$$

for some c in (a, b). However $H'(c) = 0$, so $H(a) = H(b)$. Because a and b are arbitrary points in the interval, you know that H is a constant function C. So, $G(x) - F(x) = C$ and it follows that $G(x) = F(x) + C$. ■

The expression $\int f(x)dx$ is read as the *antiderivative f of with respect to x.* So, the differential dx serves to identify x as the variable of integration. The term *indefinite integral* is a synonym for antiderivative.

Notation:

$$\int f(x)dx = G(x)dx + C$$

1. f is the *integrand*.
2. dx indicates that the variable of integration is x .
3. G is an antiderivative (or integral) of f .
4. C is the constant of integration.
5. The expression is read as “the antiderivative of f with respect to x .”

Basic Integration Rules Based on Derivative Formulas:

1. $\int 0dx = C.$
2. $\int kdx = kx + C.$

3. $\int (f(x) \pm g(x))\, dx = \int f(x)dx \pm \int g(x)dx.$

4. $\int x^n dx = \frac{x^{n+1}}{n+1} + C, \quad n \neq 1.$

5. $\int \cos x dx = \sin x + C.$

6. $\int \sin x dx = -\cos x + C.$

7. $\int \sec^2 x dx = \tan x + C.$

Summary: Antidifferentiation (or integration) is the inverse of differentiation. The antiderivative of $f(x) = 4x^3$ is $G(x) = x^4 + C$ because the derivative of G is f.

Remark 10.3

- *The power rule* $\int x^n dx = \frac{x^{n+1}}{n+1} + C$ *is not valid for* $n = -1$. *We do not yet know how to find the antiderivative* $\int \frac{1}{x} dx$. *We will return to this question when we study logarithms.*
- *Finding antiderivatives is more difficult than calculating derivatives, but remember that you can always check your answer to an integral question by differentiating the result.*
- *Some functions do not have antiderivatives among the functions that we use in elementary calculus. For example, you cannot solve* $\int e^{-x^2} dx$. *That is, there is no function* G *(among the functions in elementary calculus) whose derivative is* e^{-x^2}.

Example 10.4 (Finding antiderivatives.) *We will find simple indefinite integrals:*

$\int (2x - 6x^4 + 5)dx = -\frac{6}{5}x^5 + x^2 + 5x + C.$

$\int \left(x^3 + \cos x + 4\right) dx = 4x + \sin x + \frac{1}{4}x^4 + C.$

$\int (\sin x + \cos x + 2)\, dx = 2x - \cos x + \sin x + C.$

$\int \frac{x+1}{\sqrt{x}} dx = \int \left(\frac{x}{\sqrt{x}} + \frac{1}{\sqrt{x}} \right) dx = \int \left(x^{1/2} + x^{-1/2}\right) dx = \frac{2}{3}\sqrt{x}\,(x+3) + C.$

$\int \frac{t^2+1}{t^2} dt = \frac{1}{t}\left(t^2 - 1\right) + C.$

$\int (y-3)\sqrt{y} dy = \frac{2}{5} y^{\frac{3}{2}} (y - 5) + C.$

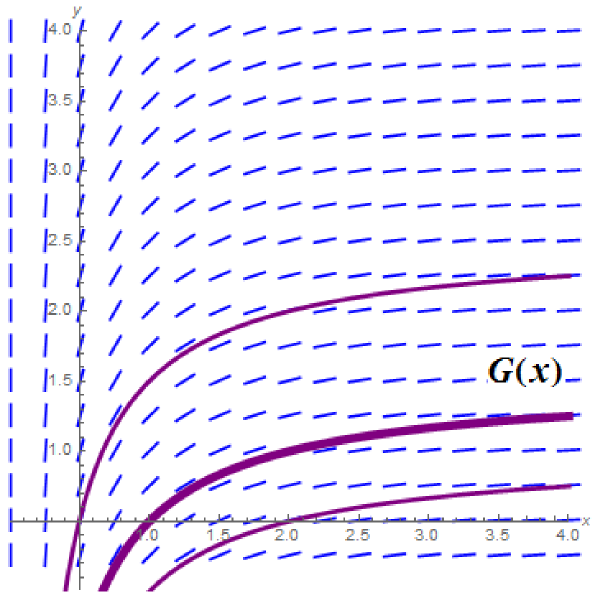


Fig. 10.1. Slope field and several solutions of the differential equation 10.1.

You can check that these are correct by differentiating the answers. □

Example 10.5 (Solving a simple differential equation) *Solve the differential equation*

$$G'(x) = \frac{1}{x^2}, \quad x > 0 \tag{10.1}$$

that satisfies the initial condition $G(1) = 0$.

Solution:

$$G(x) = \int \frac{1}{x^2} dx = \int x^{-2} dx = \frac{x^{-1}}{-1} + C = -\frac{1}{x} + C.$$

We can determine the constant of integration using the initial condition.

$$G(1) = -\frac{1}{1} + C = 0.$$

Thus, $C = 1$, and the particular solution to the differential equation is $G(x) = -\frac{1}{x} + 1$.. Notice that the general solution of the differential equation has a constant of integration and represents a family of curves in the plane. The particular solution is one of these curves (see Figure 10.1).

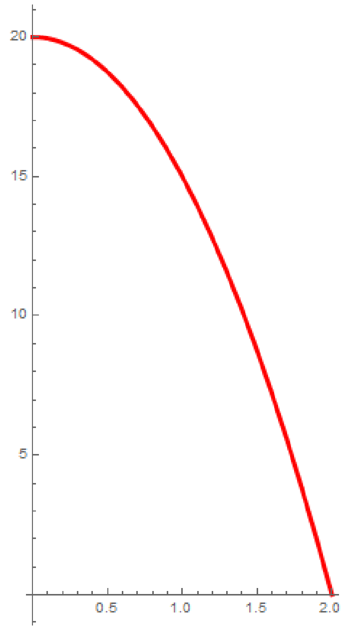


Fig. 10.2. Visible trajectory $g(x) = 20 - 5x^2$.

Example 10.6 *Galileo measured free fall, a motion with constant acceleration. Assume $s(t)$ is the height of the ball at time t. Assume the ball has zero velocity initially and is located at height $s(0) = 20$. We know that the velocity is $v(t)$ is the derivative of $s(t)$ and the acceleration $a(t)$ is constant equal to -10. So, $v(t) = -10t + C$ is the antiderivative of a. By looking at v at time $t = 0$ we see that$C = v(0)$ is the initial velocity and so zero. We know now $v(t) = -10t$. We need now to compute the anti derivative of $v(t)$. This is $s(t) = -10t^2/2 + C$. Comparing $t = 0$ shows $C = 20$. Now $s(t) = 20 - 5t^2$. The graph of s is a parabola. If we give the ball an additional horizontal velocity, such that time $t$is equal to x then $s(x) = 20 - 5x^2$ is the visible trajectory (see Figure 10.2). We see that jumping from 20 meters leads to a fall which lasts 2 seconds.*

10.2 The area problem and the definite integral

In Figure 10.3 you can see a region Ω bounded above by the graph of a nonnegative continuous function f, bounded below by the x-axis, bounded on the left by the line $x = a$, and bounded on the right by the line $x = b$. The question before us is this: What number, if any, should be called the area of

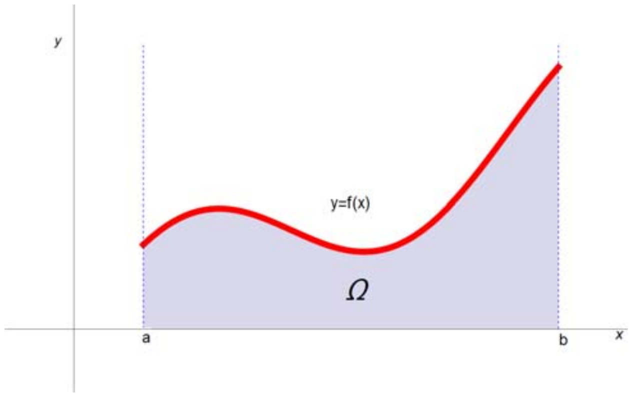


Fig. 10.3. What number should be called the area of Ω?.

Ω? Calculus provides the answer, and we begin by looking at a technique for estimating the area of Ω by a summation process called Riemann sums.

Definition 10.7 (of the area of a region in the plane) *Let f be continuous and nonnegative on the interval $[a,b]$. Partition the interval into n equal subintervals of length $\Delta x = \dfrac{b-a}{n}$, $x_0 = a$, $x_1 = a + \Delta x, ..., x_n = a + n\Delta x = b$. The* area of the region *bounded by f, the x-axis, and the vertical lines $x = a$ and $x = b$ is*

$$A = \lim_{n\to\infty} \sum_{i=1}^{n} f(c_i)\Delta x, \quad x_{i-1} \le c_i \le x_i, \tag{10.2}$$

provided this limit exists and is equal independently on the choice of c_i, $i = 1, ..., n$ $(x_{i-1} \le c_i \le x_i)$. The expression $\sum_{i=1}^{n} f(c_i)\Delta x$ is called a Riemann sum[1]*. The endpoints $x_0, x_1, x_2, ..., x_{n-1}, x_n$ of the subintervals are called* grid points *, and they create* regular partition *of the interval $[a,b]$. The sum $\sum_{i=1}^{n} f(c_i)\Delta x$ is called*

- *a* left Riemann sum *if c_i is the left endpoint of $[x_{i-1}, x_i]$;*
- *a* right Riemann sum *if c_i is the right endpoint of $[x_{i-1}, x_i]$;*
- *a* midpoint Riemann sum *if c_i is the midpoint of $[x_{i-1}, x_i]$*

In the first example, we find the area under a parabola.

Example 10.8 *Calculate the area A under the parabola $f(x) = x^2$ and above the x-axis, where $0 \le x \le 2$ (the parabolic region S illustrated in Figure 10.4)*

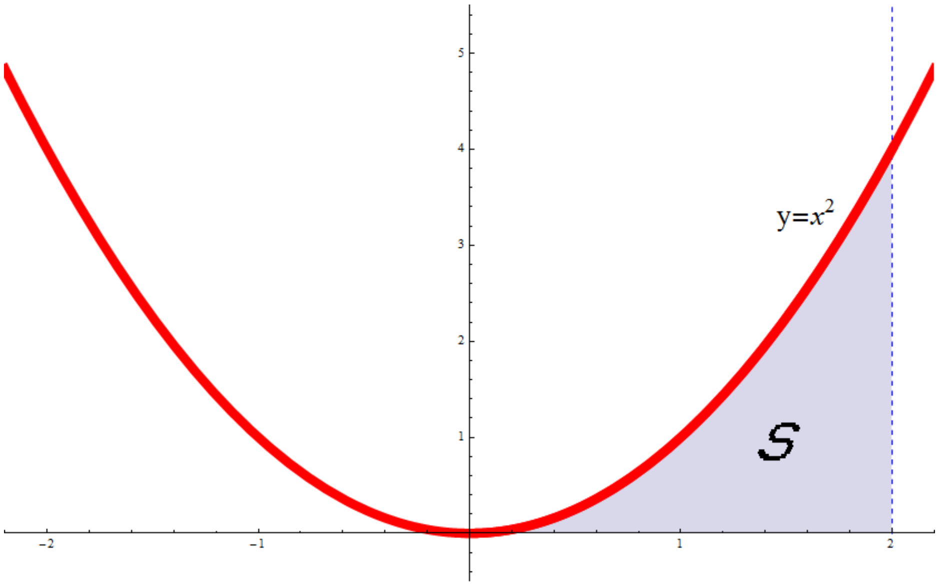


Fig. 10.4. This isn't a triangle or a rectangle.

Solution: We want to chop up our interval $[0, 2]$ into n pieces, each the same length. Since the total length is 2, and we're using n pieces, each piece must have length $\Delta x = 2/n$ units. The first piece goes from 0 to $2/n$; the second piece goes from $2/n$ to $4/n$; and so on. In this case, the equidistant partition

$$a = x_0 < x_1 < x_2 < ... < x_{n-1} < x_n = b$$

specializes to

$$0 = \frac{0}{n} < \frac{2}{n} < \frac{4}{n} < ... < \frac{2(n-1)}{n} < \frac{2n}{n} = 2.$$

The *mesh* Δx of this partition is $2/n$, since every smaller interval has width $2/n$. It's also pretty clear that the formula for a general x_i in this partition is $2i/n$. Now, we need to choose our numbers c_i . For example, c_0 could be anywhere in the interval $[0, 2/n]$, c_1 could be anywhere inside $[2/n; 4/n]$, and so on. We'll make life simple by always choosing the right endpoint of each smaller interval, so that $c_i = x_i = 2i/n$. That is,

$$c_i = \frac{2i}{n} \quad \text{is our choice for the smaller interval } [x_{i-1}, x_i] = \left[\frac{2(i-1)}{n}, \frac{2i}{n}\right]$$

This will lead to the rectangles shown in Figure 10.5.So we're actually dealing

[1]It will be generalized later.

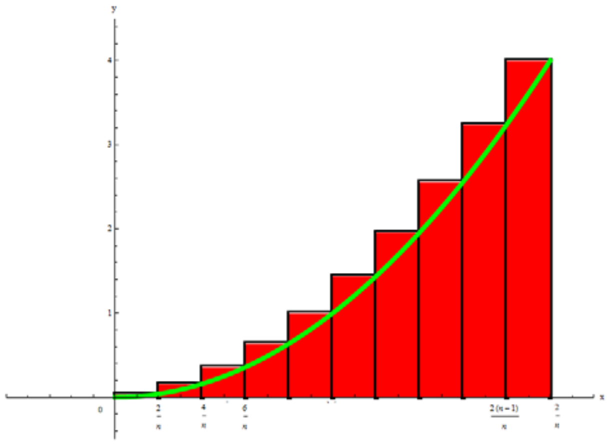


Fig. 10.5. Upper sum area ($n = 10$).

with an *upper sum* here–all the rectangles lie above the curve. Now, we're finally ready to use the formula. Consider the sum

$$\sum_{i=1}^{n} f(c_i)(x_i - x_{i-1}).$$

We know that $f(x) = x^2$, $c_i = 2i/n$, $x_i = 2i/n$ as well, and $x_{i-1} = 2(i-1)/n$. So the sum becomes

$$R(n) = \sum_{i=1}^{n} \left(\frac{2i}{n}\right)^2 \left(\frac{2i}{n} - \frac{2(i-1)}{n}\right) = \frac{8}{n^3} \sum_{i=1}^{n} i^2$$

which gives

$$R(n) = \frac{4}{3n^2}\,(2n+1)\,(n+1)\,.$$

Instead of using the rectangles in Figure 10.5, we could use the smaller rectangles whose heights are the values of f at the left endpoints of the subintervals. (The leftmost rectangle has collapsed because its height is 0.) The sum of the areas of these approximating rectangles is

$$L(n) = \frac{8}{n^3} \sum_{i=1}^{n} (i-1)^2$$

which leads to

$$L(n) = \frac{4}{3n^2}\,(2n-1)\,(n-1)\,.$$

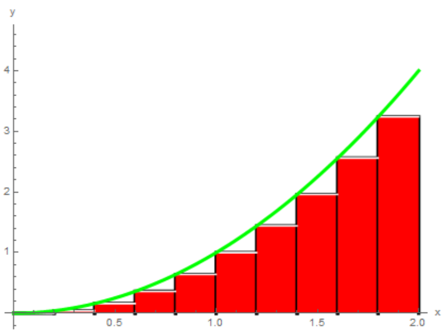
It is easy to check, that as n increases values of $R(n)$ decreases, but values of $L(n)$ increases. Tables 10.3 and 10.4 shows what happens when we divide the region S into more and more strips of equal width. By computing the sum of the areas of the smaller rectangles and the sum of the areas of the larger rectangles , we obtain better lower and upper estimates for area A:

$$L(n) < A < R(m) \quad \text{for} \quad m, n \geq 1$$

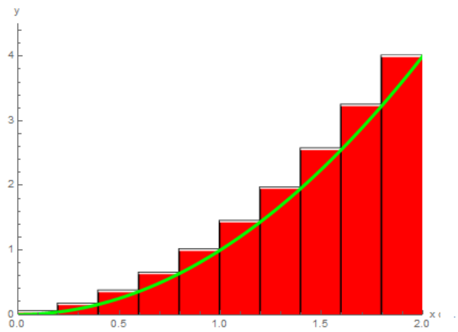
So one possible answer to the question is to say that the true area of S lies somewhere between 2.6663 and 2.6671.

n	$L(n)$	$R(n)$
10	2.2800	3.0800
25	2.5088	2.8288
50	2.5872	2.7472
100	2.6268	2.7068
1000	2.6627	2.6707
10000	2.6663	2.6671

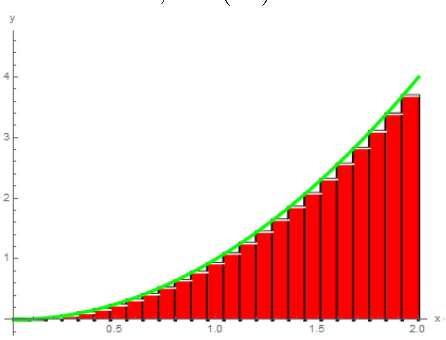
(10.3)



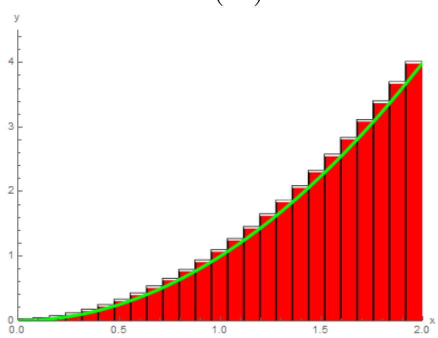
$n = 10, \quad L(10) = 2.2800$



$n = 10 \quad R(10) = 3.0800$



$n = 25, 3 \quad L(25) = 2.5088$



$n = 25 \quad R(25) = 2.8288$

(10.4)

The sum $L(n)$ as well as $R(n)$ is only an approximation to the area we're looking for. Since the mesh Δx of the partition is $2/n$, we can force the mesh to go to 0 by letting $n \to \infty$. The rectangles become smaller and smaller, but there are more and more of them which hug the curve $y = x^2$ better and better. So we have

$$\lim_{n\to\infty} \frac{4}{3n^2}(2n-1)(n-1) \le A \le \lim_{n\to\infty} \frac{4}{3n^2}(2n+1)(n+1).$$

All that's left is to find the last limit. You can show that (independently on the choice of c_i) the limit of $\sum_{i=1}^{n} f(c_i)\Delta x$ is 8/3, so we have finally shown that

$$A = \frac{8}{3}. \qquad \square$$

There is one final note. Though the use of a Riemann sum to estimate a definite integral is a very direct approach, it is not very efficient. More efficient methods of estimating definite integrals have been developed, and these are studied in the branch of mathematics called Numerical Analysis.

10.3 Net area

We introduced Riemann sums in Definition 10.7 as a way to approximate the area of a region bounded by a curve $y = f(x)$ and the x-axis on an interval $[a,b]$. In that discussion, we assumed f to be nonnegative on the interval. Our next task is to discover the geometric meaning of Riemann sums when f is negative on some or all of $[a,b]$. Once this matter is settled, we can proceed to the main event of this chapter, which is to define the definite integral. With definite integrals, the approximations given by Riemann sums become exact. So, how do we interpret Riemann sums when f is negative at some or all points of $[a,b]$? The answer follows directly from the Riemann sum definition.

Example 10.9 *Evaluate and interpret the midpoint Riemann sums for $f(x) = 1 - x^2$ on the interval $[a,b] = [1,3]$ with $n = 8$ equally spaced subintervals (see Figure 10.6).*

Solution: The length of each subinterval is $\Delta x = \dfrac{(b-a)}{n} = \dfrac{(3-1)}{8} = \dfrac{1}{4}$. So the grid points are $x_i = a + i\Delta x$, $i = 0, ..., n$. To compute the midpoint Riemann sum, we evaluate f at the midpoints of the subintervals, which are

$$c_1 = 1.1250, \quad c_2 = 1.3750, \quad c_3 = 1.6250, \quad c_4 = 1.8750,$$

$$c_5 = 2.1250, \quad c_6 = 2.3750, \quad c_7 = 2.6250, \quad c_8 = 2.8750.$$

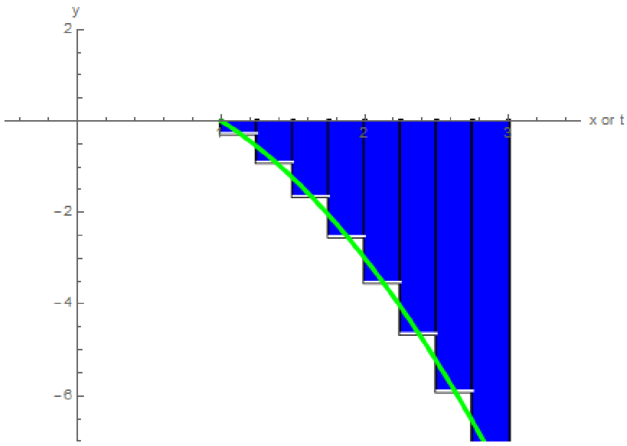


Fig. 10.6. The midpoint Riemann sum for $n = 8$ and $f(x) = 1 - x^2$ on $[1, 3]$ is -6.6563.

The resulting midpoint Riemann sum is

$$\sum_{i=1}^{n} f(c_i)\Delta x = \frac{1}{4}\sum_{i=1}^{8} f(c_i) = -6.6563.$$

All values of $f(c_i)$ are negative, so the Riemann sum is also negative. Because area is always a nonnegative quantity, this Riemann sum does not approximate an area. Notice, however, that the values of $f(c_i)$ are the negative of the heights of the corresponding rectangles (Figure 10.6). Therefore, the Riemann sum is an approximation to the negative of the area of the region bounded by the curve. □

In the more general case that f is positive on only part of $[a, b]$, we get positive contributions to the sum where f is positive and negative contributions to the sum where f is negative. In this case, Riemann sums approximate the area of the regions that lie above the x-axis minus the area of the regions that lie below the x-axis (Figure 10.7). This difference between the positive and negative contributions is called the *net area*; it can be positive, negative, or zero (see Figure 10.7).

Definition 10.10 (Net area) *Consider the region Ω bounded by the graph of a continuous function f and the x-axis between $x = a$ and $x = b$. The net area of Ω is the sum of the areas of the parts of Ω that lie above the x-axis minus the sum of the areas of the parts of Ω that lie below the x-axis on $[a, b]$.*

What we have learned is this: If $f(x) < 0$ on the interval under discussion, then the integral of f will be a negative number. If we want to calculate *positive*

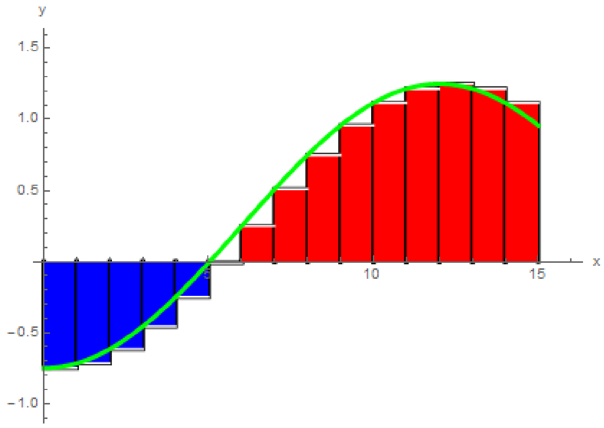


Fig. 10.7. Riemann sum approximates the area of the regions that lie above the x-axis minus the area of the regions that lie below the x-axis

area then we must interject a minus sign (see Figure 10.8). Let us nail down our understanding of these ideas by considering an example.

Example 10.11 *Use the areas shown in the figure* **??** *to find*

a) $\int_a^b f(x)dx$

b) $\int_b^c f(x)dx$

c) $\int_a^c f(x)dx$

d) $\int_a^d f(x)dx$

e) $\int_a^d |f(x)|\, dx$

Solution:

a) $\int_a^b f(x)dx = 7.28$, (net area)

b) $\int_b^c f(x)dx = -1.65$, (net area)

c) $\int_a^c f(x)dx = 7.28 + (-1.65) = 5.63$ (net area)

d) $\int_a^d f(x)dx = 7.28 + (-1.65) + 6.26 = 11.89$ (net area)

e) $\int_a^d |f(x)|\, dx = 7.28 - (-1.65) + 6.26 = 15.19$ (positive area)

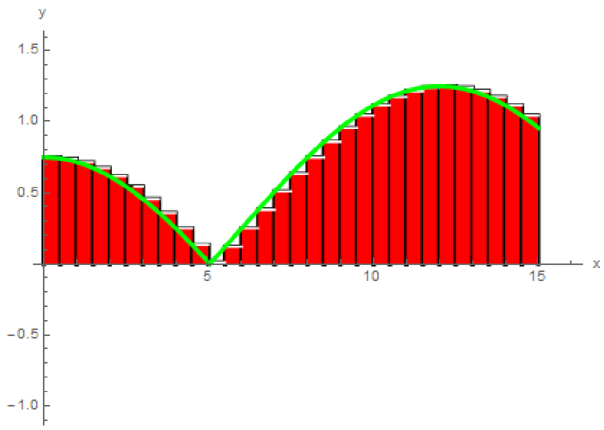


Fig. 10.8. Positive area for the function from Figure 10.7.

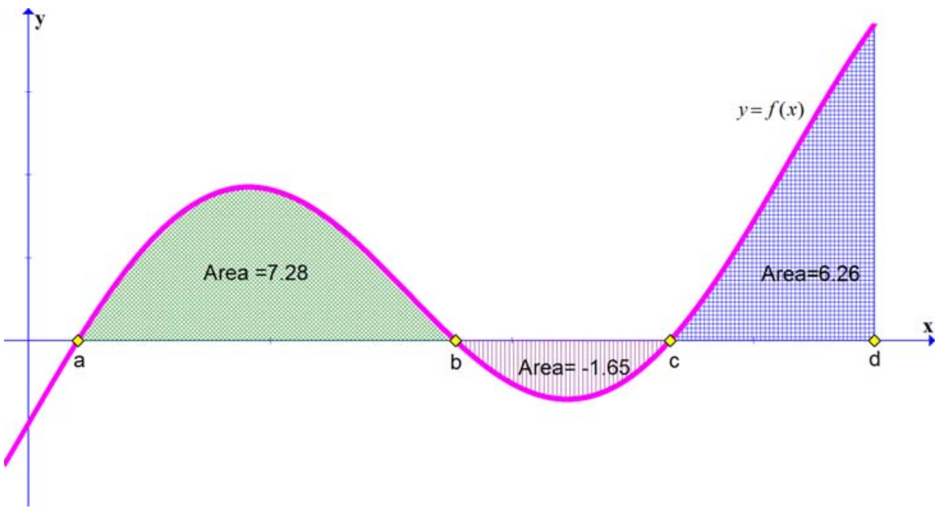


Fig. 10.9. Net areas for Example 10.11.

10.4 The definite integral

Riemann sums for f on $[a, b]$ give *approximations* to the net area of the region bounded by the graph of f and the x-axis between $x = a$ and $x = b$, where $a < b$. How can we make these approximations exact? If f is continuous on $[a, b]$, it is reasonable to expect the Riemann sum approximations to approach the exact value of the net area as the number of subintervals $n \to \infty$ and as the length of the subintervals $\Delta x \to 0$ (Figure 10.4). In terms of limits, we write

$$\textit{net area} = \lim_{n\to\infty} \sum_{i=1}^{n} f(c_i)\Delta x. \tag{10.5}$$

The Riemann sums we have used so far involve regular partitions in which the subintervals have the same length Δx. We now introduce partitions of $[a, b]$ in which the lengths of the subintervals are not necessarily equal. A *general partition* of $[a, b]$ consists of the n subintervals

$$[x_0, x_1],\ [x_1, x_2],\ [x_2, x_3], ..., [x_{n-1}, x_n]$$

where $x_0 = a$ and $x_n = b$. The length of the i-th subinterval is $\Delta x_i = x_i - x_{i-1}$, for $i = 1, ..., n$. We let c_i be any point in the subinterval $[x_{i-1}, x_i]$.

This general partition is used to define the general Riemann sum.

Definition 10.12 (General Riemann sum) *Suppose*

$$[x_0, x_1],\ [x_1, x_2],\ [x_2, x_3], ..., [x_{n-1}, x_n]$$

are subintervals of $[a, b]$ *with*

$$a = x_0 < x_1 < x_2 < ... < x_{n-1} < x_n = b.$$

Let Δx_i *be the length of the subinterval* $[x_{i-1}, x_i]$ *and let* c_i *be any point in* $[x_{i-1}, x_i]$, *for* $i = 1, 2, ..., n$. *If* f *is defined on* $[a, b]$, *the sum*

$$\sum_{i=0}^{n} f(c_i)\Delta x_i = f(c_1)\Delta x_1 + f(c_2)\Delta x_2 + ... + f(c_n)\Delta x_n \tag{10.6}$$

is called a general Riemann sum for f *on* $[a, b]$ *(see Figure 10.10).*

Now consider the limit of $\sum_{i=0}^{n} f(c_i)\Delta x_i$ as $n \to \infty$ and as all the $\Delta x_i \to 0$. We let Δ denote the largest value of Δx_i; that is,

$$\Delta = \max\{\Delta x_1, \Delta x_2, ..., \Delta x_n\}.$$

Observe that if $\Delta \to 0$, then $\Delta x_i \to 0$, for $i = 1, 2, ..., n$. In order for the limit $\lim_{\Delta\to 0} \sum_{i=0}^{n} f(c_i)\Delta x_i$ to exists, it must have the same value over all general partitions of $[a, b]$ and for all choices of c_i on a partition.

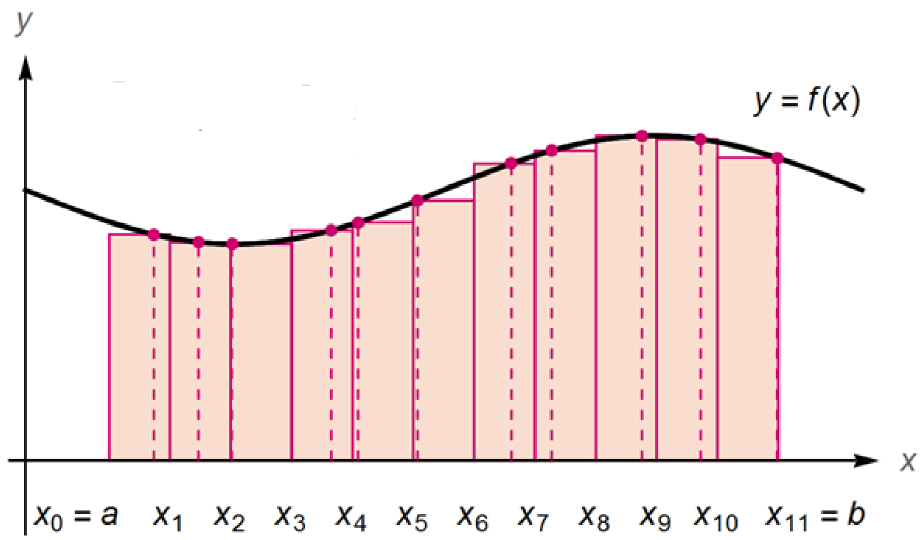


Fig. 10.10. Riemann Sum. Each c_i can be any point in the k-th subinterval.

Definition 10.13 *A function f defined on $[a, b]$ is* ***integrable*** *on $[a, b]$ if*

$$\lim_{\Delta\to 0}\sum_{i=0}^{n} f(c_i)\Delta x_i \tag{10.7}$$

exists and is unique over all general partitions of $[a, b]$ and all choices of c_i on a partition. This limit is the ***definite integral*** *of f from a to b, which we write*

$$\int_a^b f(x)dx = \lim_{\Delta\to 0}\sum_{i=0}^{n} f(c_i)\Delta x_i. \tag{10.8}$$

The symbol

$$\int_a^b f(x)dx$$

which is read as "the integral from a to b of f of x dee x" or sometimes as "the integral from $a$to b of f of x with respect to x." When you find the value of the integral, you have *evaluated the integral.*

The component parts in the integral symbol also have names:

- $f(x)$ is the *integrand* ,
- x is the *variable of integration,*

- a is the *lower limit of integration,*
- b is the *upper limit of integration,*
- $\int$ is the *integral sign.*

The variable of integration is a *dummy variable* that is completely internal to the integral. It does not matter what the variable of integration is called, as long as it does not conflict with other variables that are in use. Therefore, the integrals

$$\int_a^b f(x)dx, \quad \int_a^b f(t)dt, \quad \int_a^b f(s)ds, \quad \int_a^b f(\omega)d\omega$$

all have the same meaning.

The strategy of slicing a region into smaller parts, summing the results from the parts, and taking a limit is used repeatedly in calculus and its applications. We call this strategy the *slice-and-sum method.* It often results in a Riemann sum whose limit is a definite integral.

10.5 Integrable and nonintegrable functions

Not every function defined over the closed interval $[a, b]$ is integrable there, even if the function is bounded. That is, the Riemann sums for some functions may not converge to the same limiting value, or to any value at all. A full development of exactly which functions defined over $[a, b]$ are integrable requires advanced mathematical analysis, but fortunately most functions that commonly occur in applications are integrable. In particular, every continuous function over $[a, b]$ is integrable over this interval, and so is every function having no more than a finite number of jump discontinuities on $[a, b]$. (The latter are called *piecewise-continuous.* The following theorem, which is proved in more advanced courses, establishes these results.

Theorem 10.14 (Integrability of continuous functions) *If a function f is continuous over the interval $[a, b]$, or if f has at most finitely many jump discontinuities there (i.e. there is a positive number M such that $-M \leq f(x) \leq M$ for all x in $[a, b]$), then the definite integral $\int_a^b f(x)dx$ exists and f is integrable over $[a, b]$. If f is not bounded on $[a, b]$, then f is not integrable on $[a, b]$.*

The idea behind Theorem 10.14 for continuous functions is given in Example 10.8. Briefly, when f is continuous we can choose each c_i so that $f(c_i)$ gives the maximum value of f on the subinterval $[x_{i-1}, x_i]$, resulting in an *upper Riemann sum* . Likewise, we can choose c_i to give the minimum value of f on $[x_{i-1}, x_i]$ to obtain a *lower sum.* The upper and lower sums can be shown to converge to the same limiting value as $\Delta = \max\{\Delta x_1, \Delta x_2, ..., \Delta x_n\}$ (called the *norm or diameter of the partition*) tends to zero. Moreover, every Riemann sum is trapped between the values of the upper and lower sums, so every Riemann sum converges to the same limit as well. Therefore, the definite integral exists, and the continuous function f is integrable over $[a, b]$.

For integrability to fail, a function needs to be sufficiently discontinuous that the region between its graph and the x-axis cannot be approximated well by increasingly thin rectangles. The next example shows a function that is not integrable over a closed interval.

Example 10.15 *The function*

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ \\ 0 & \text{if } x \text{ is irrational} \end{cases}$$

has no Riemann integral over $[0, 1]$*. Underlying this is the fact that between any two numbers there is both a rational number and an irrational number. Thus the function jumps up and down too erratically over* $[0, 1]$ *to allow the region beneath its graph and above the* x*-axis to be approximated by rectangles, no matter how thin they are. We show, in fact, that upper sum approximations and lower sum approximations converge to different limiting values. If we pick a partition of* $[0, 1]$ *and choose* c_i *to be the point giving the maximum value for* f *on* $[x_{i-1}, x_i]$ *then the corresponding Riemann sum is*

$$R = \sum_{i=0}^{n} f(c_i)\Delta x_i = \sum_{i=0}^{n} 1 \cdot \Delta x_i = 1$$

since each subinterval $[x_{i-1}, x_i]$ *contains a rational number* c_i *where* $f(c_i) = 1$*. Note that the length of the intervals in the partition sum to* 1, $\sum_{i=0}^{n} \Delta x_i = 1$*. So each such Riemann sum equals* 1*, and a limit of Riemann sums using these choices equals* 1*. On the other hand, if we pick* c_i*. to be the point giving the minimum value for* f *on* $[x_{i-1}, x_i]$*, then the Riemann sum is*

$$L = \sum_{i=0}^{n} f(c_i)\Delta x_i = \sum_{i=0}^{n} 0 \cdot \Delta x_i = 0$$

since each subinterval $[x_{i-1}, x_i]$ *contains an irrational number* c_i*. where* $f(c_i) = 0$*. The limit of Riemann sums using these choices equals zero. Since the limit depends on the choices of* c_i*, the function* f *is not integrable.* □

Theorem 10.14 says nothing about how to calculate definite integrals. A method of calculation will be developed in the next chapter, through a connection to the process of taking antiderivatives.

The following example shows that it is not necessary to have subintervals of equal width in the general Riemann sums. This is a key feature of the development of definite integrals.

Example 10.16 *Consider the region bounded by the graph of $f(x) = \sqrt{x}$ and the x-axis for $0 \leq x \leq 1$. As $f(x)$ is continuous it is integrable. So, in order to find area of our region we can (taking into account that $f(x)$ is nonnegative) evaluate the limit*

$$\lim_{n\to\infty} \sum_{i=0}^{n} f(c_i)\Delta x_i$$

where c_i is the right endpoint of the partition given by $c_i = \dfrac{i^2}{n^2}$ and Δx_i is the width of the i-th interval. Notice, that the width of the i-th interval is given by

$$\Delta x_i = \frac{i^2}{n^2} - \frac{(i-1)^2}{n^2} = \frac{2i-1}{n^2}$$

and tends to zero if $n \to \infty$. So, the limit is

$$\begin{aligned}
\lim_{n\to\infty} \sum_{i=0}^{n} f(c_i)\Delta x_i &= \lim_{n\to\infty} \sum_{i=0}^{n} \sqrt{\frac{i^2}{n^2}} \left(\frac{2i-1}{n^2}\right) \\
&= \lim_{n\to\infty} \frac{1}{n^3} \sum_{i=0}^{n} \left(2i^2 - i\right) \\
&= \lim_{n\to\infty} \frac{1}{n^3} \cdot \frac{1}{6} n\,(4n-1)\,(n+1) \\
&= \frac{2}{3}. \quad \square
\end{aligned}$$

Remark 10.17 *Sometimes there is an easier way. Then we can find definite integrals without using messy Riemann sums explicitly (see next chapter entitled The "Fundamental Theorem of Calculus") or we can use geometric arguments.*

Example 10.18 *Let us find*

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{\sqrt{n^2 - i^2}}{n^2}. \tag{10.9}$$

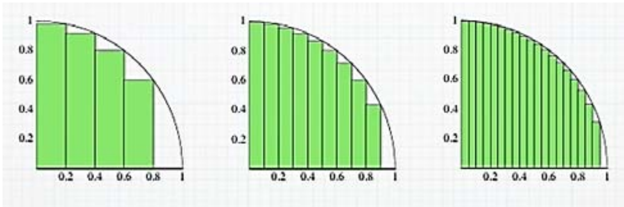


Fig. 10.11.

Solution: Our aim is to interpret that limit as an area. Let us begin by looking at a few numbers

$$\sum_{i=1}^{2}\frac{\sqrt{2^2-i^2}}{2^2} = \frac{1}{4}\sqrt{3},$$
$$\sum_{i=1}^{5}\frac{\sqrt{5^2-i^2}}{5^2} = \frac{1}{25}\sqrt{21}+\frac{1}{25}\sqrt{24}+\frac{7}{25} = 0.659\,26$$
$$\sum_{i=1}^{10}\frac{\sqrt{10^2-i^2}}{10^2} \approx 0.726\,13$$
$$\sum_{i=1}^{1000}\frac{\sqrt{1000^2-i^2}}{1000^2} \approx 0.784\,89$$

It appears that when n gets larger and larger, that sums approach a number which is close to 0.8. From a geometric point of view, in we are computing 10.9 the Riemann sums which approximate (better and better when n increases) the integral

$$\int_0^1 \sqrt{1-x^2}dx = \frac{\pi}{4} \approx 0.785\,40,$$

(it is the area of the quater-circle with a center at the origin and the radius one). $\square$

10.6 Properties of definite integrals

In defining $\int_a^b f(x)dx$ as a limit of sums $\sum_{i=0}^n f(c_i)\Delta x_i$ we moved from left to right across the interval $[a,b]$. What would happen if we instead move right to left, starting with $x_0 = b$ and ending at $x = a$? Each Δx_i. in the Riemann sum would change its sign, with $x_i - x_{i-1}$ now negative instead of positive. With the same choices of c_i. in each subinterval, the sign of any Riemann sum would change, as would the sign of the limit, the integral $\int_a^b f(x)dx$. Since we have not previously given a meaning to integrating backward, we are led to define

$$\int_b^a f(x)dx = -\int_a^b f(x)dx.$$

Although we have only defined the integral over an interval $[a, b]$ when $a < b$, it is convenient to have a definition for the integral over $[a, b]$ when $a = b$, that is, for the integral over an interval of zero width. Since $a = b$ gives $\Delta x = 0$, whenever $f(a)$ exists we define

$$\int_a^a f(x)dx = 0$$

Theorem 10.19 states basic properties of integrals, given as rules that they satisfy, including the two just discussed. These rules become very useful io the process of computing integrals. We will refer to them repeatedly to simplify our calculations.

Theorem 10.19 *When f and g are integrable over the interval $[a, b]$, the definite integral satisfies the following rules:*

1. *Order of integration:*

$$\int_a^b f(x)dx = -\int_b^a f(x)dx \qquad \text{(by definition)}. \tag{10.10}$$

2. *Zero width interval*:

$$\int_a^a f(x)dx = 0 \qquad \text{(by definition)}. \tag{10.11}$$

3. *Constant multiple*:

$$\int_a^b kf(x)dx = k\int_a^b f(x)dx \qquad \text{(for any constant } k\text{)}. \tag{10.12}$$

4. *Sum and difference*:

$$\int_a^b (f(x) \pm g(x))\, dx = \int_a^b f(x)dx \pm \int_a^b g(x)dx. \tag{10.13}$$

5. *Interval additivity (see Fig. 10.12):*

$$\int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx. \tag{10.14}$$

6. *Max-min inequality*:

   If f has maximum value $\max f$ and minimum value $\min f$ on $[a, b]$ then

$$\min f \cdot (b - a) \le \int_a^b f(x)dx \le \max f \cdot (b - a). \qquad (10.15)$$

7. *Domination*:

$$f(x) \;\ge\; g(x) \quad \text{on } [a, b] \Rightarrow \int_a^b f(x)dx \ge \int_a^b g(x)dx \qquad (10.16)$$

$$f(x) \;\ge\; 0 \quad \text{on } [a, b] \Rightarrow \int_a^b f(x)dx \ge 0. \qquad (10.17)$$

While Rules 1 and 2 are definitions, Rules 3 to 7 from above must be proved. The following is a proof of Rule 6. Similar proofs can be given to verify the other properties.

Proof. Rule 6 says that the integral of f over $[a, b]$ is never smaller than the minimum value of f times the length of the interval and never larger than the maximum value of f times the length of the interval. The reason is that for every partition of $[a, b]$ and for every choice of the points c

i,

$$\begin{aligned}
\min\, f \cdot (b - a) &= \min f \cdot \textstyle\sum_{i=1}^n \Delta x_i && (\text{as } \textstyle\sum_{i=1}^n \Delta x_i = (b - a)) \\
&= \textstyle\sum_{i=1}^n \min f \cdot \Delta x_i && (\text{by constant multiple rule}) \\
&\le \textstyle\sum_{i=1}^n f(c_i) \cdot \Delta x_i && (\min f \le f(c_i)) \\
&\le \textstyle\sum_{i=1}^n \max f \cdot \Delta x_i && (f(c_i) \le \max f) \\
&= \max f \textstyle\sum_{i=1}^n \Delta x_i && (\text{by constant multiple rule}) \\
&= \max f \cdot (b - a)
\end{aligned}$$

In short, all Riemann sums for f on $[a, b]$ satisfy the inequality

$$\min\ f \cdot (b - a) \le \sum_{i=1}^n f(c_i) \cdot \Delta x_i \le \max\ f \cdot (b - a).$$

Hence their limit, the integral, does too. ∎

Remark 10.20 *In the Interval additivity property the point c does not need to to be a number in the middle (see Fig. 10.12) where the situation is explained graphically.*

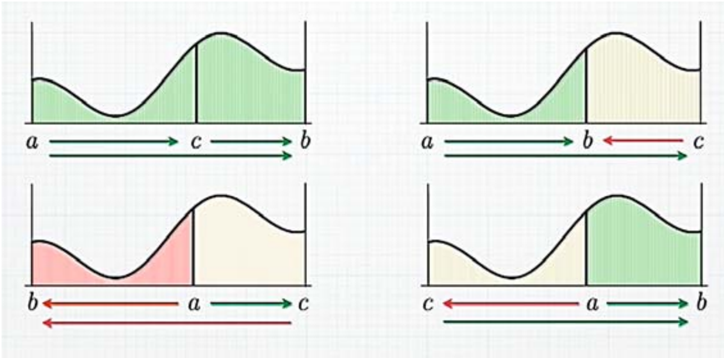


Fig. 10.12. Possible locations of the point c in the Interval additivity property.

Example 10.21 *To illustrate some of the rules, we suppose that*

$$\int_0^2 f(x)dx = 4, \quad \int_2^5 f(x)dx = -2 \quad \int_0^2 h(x)dx = 6.$$

Then

1. $\int_2^0 f(x)dx = -\int_0^2 f(x)dx = -4$
2. $\int_0^2 (4f(x) + 2h(x))\, dx = 4\int_0^2 f(x)dx + 2\int_0^2 h(x)dx = 4(4) + 2(6) = 28$
3. $\int_0^5 f(x)dx = \int_0^2 f(x)dx + \int_2^5 f(x)dx = 4 + (-2) = 2$

Notice, that the domination property 10.16 can be generalized in Theorem (A proof of this theorem is left as an exercise).

Definition 10.22 (Extension) *If f is integrable on $[a, b]$ and $g(x) = f(x)$ for all but finitely many x in $[a, b]$, then (by definition)*

$$\int_a^b f(x)dx = \int_a^b g(x)dx. \tag{10.18}$$

Example 10.23 *Let us consider*

$$\int_{-1}^1 \frac{\sin x}{x} dx. \tag{10.19}$$

Here the integrand function $g(x) = \dfrac{\sin x}{x}$ is not defined at $x = 0$, but we know that

$$\lim_{x\to 0} \frac{\sin x}{x} = 1.$$

This discontinuity is removable, so we can understood the integral 10.19 as

$$\int_{-1}^{1} f(x)dx$$

where $f(x)$ *is a continuous function*

$$f(x) = \begin{cases} \frac{\sin x}{x} & \text{if} \quad x \neq 0 \\ 1 & \text{if} \quad x = 0. \end{cases}$$

Without proof we will accept the following theorem:

Theorem 10.24 *If* $f(x)$ *is bounded and has finite number of discontinuities* $t_1 < t_2 < ... < t_k$ *in* $[a, b]$ *then* $f(x)$ *is integrable on* $[a, b]$ *and*

$$\int_a^b f(x)dx = \int_a^{t_1} f(x)dx + \int_{t_1}^{t_2} f(x)dx + ... + \int_{t_k}^{b} f(x)dx. \tag{10.20}$$

Example 10.25 *Let*

$$f(x) = \begin{cases} 1 - \sqrt{1 - x^2} & \text{if} \quad 0 \leq x < 1, \\ x - 2 & \text{if} \quad 1 \leq x < 2 \\ 1 & \text{if} \quad 2 < x. \end{cases}$$

Find

$$\int_0^3 f(x)dx.$$

Solution: According to Figure 10.13 we can find the value of that definite integral using geometric arguments

$$\begin{aligned} \int_0^3 f(x)dx &= \int_0^1 f(x)dx + \int_1^2 f(x)dx + \int_2^3 f(x)dx \\ &= \int_0^1 \left(1 - \sqrt{1 - x^2}\right) dx + \int_1^2 (x - 2)\, dx + \int_2^3 (1)\, dx \\ &= 1 - \frac{\pi}{4} - \frac{1}{2} + 1 = \frac{3}{2} - \frac{1}{4}\pi. \end{aligned}$$

The value of the integral $\int_0^1 \left(1 - \sqrt{1 - x^2}\right) dx$ is easy to find if you know the formula for a circle area (here with radius 1). We will be able to find this value analytically, after discussing "trigonometric substitutions". □

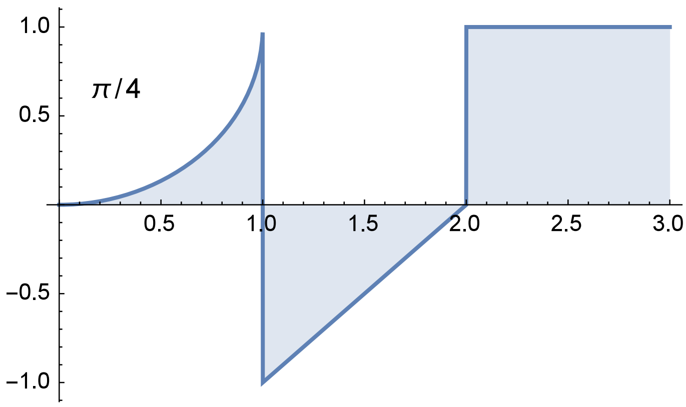


Fig. 10.13. f is bounded and has finite number of discontinuities.

Exercise 10.1 *The* ***triangle inequality for definite integrals*** *says, that if $a < b$, then*

$$\left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)|\, dx.$$

Show, that it can be deduced from the definition of the definite integral.

11

The fundamental theorems of calculus

In the previous chapter we computed integrals from the definition as a limit of Riemann sums and we saw that this procedure is sometimes long and difficult. Sir Isaac Newton discovered a much simpler method for evaluating integrals and a few years later Leibniz made the same discovery. They realized that they could calculate $\int_a^b f(x)dx$ if they happened to know an antiderivative F of f. This establishes a connection between integral calculus and differential calculus. The Fundamental Theorems of Calculus relate the integral to the derivative, and we will see in this chapter that it greatly simplifies the solution of many problems. We will see why

the processes of integration and differentiation are inverses to one another.

11.1 The First Fundamental Theorem of Calculus

Theorem 11.1 (The First Fundamental Theorem of Calculus) *If a function f is continuous on the closed interval $[a,b]$ and F is an antiderivative of f on the interval $[a,b]$ then*

$$\int_a^b f(x)dx = F(b) - F(a). \tag{11.1}$$

This theorem states that if we know an antiderivative F of f, then we can evaluate $\int_a^b f(x)dx$ simply by subtracting the values of F at the endpoints of the interval $[a,b]$. It is very surprising that $\int_a^b f(x)dx$, which was defined by a complicated procedure involving all of the values of $f(x)$ for between a and b, can be found by knowing the values of $F(x)$ at only two points, a and b.

Proof. The key to the proof is in writing the difference $F(b) - F(a)$ in a convenient form. Let P_Δ be any partition of $[a,b]$

$$a = x_0 < x_1 < x_2 < ... < x_{n-1} < x_n = b$$

with diameter Δ, where $\Delta x_i = x_i - x_{i-1}$, for $i = 1, ..., n$ and

$$\Delta = \max\{\Delta x_1, \Delta x_2, ..., \Delta x_n\}.$$

By pairwise subtraction and addition of like terms, you can write

$$F(b) - F(a) = F(x_n) - F(x_{n-1}) - ... - F(x_1) + F(x_1) - F(x_0) = \sum_{i=1}^{n} [F(x_i) - F(x_{i-1})].$$

By the Mean Value Theorem, you know that there exists a number c_i in the i-th subinterval such that

$$F'(c_i) = \frac{F(x_i) - F(x_{i-1})}{x_i - x_{i-1}}.$$

Because $F'(c_i) = f(c_i)$, we obtain

$$F(b) - F(a) = \sum_{i=1}^{n} f(c_i)\Delta x_i.$$

This important equation tells you that by repeatedly applying the Mean Value Theorem, you can always find a collection of c_i's such that the *constant* $F(b) - F(a)$ is a Riemann sum of f on $[a, b]$ for any partition. Theorem 10.14 guarantees that the limit of Riemann sums over the partition with $\Delta \to 0$ exists. So, taking the limit (as $\Delta \to 0$) produces

$$F(b) - F(a) = \lim_{\max \Delta x_i \to 0} \sum_{i=1}^{n} f(c_i)\Delta x_i = \int_a^b f(x)dx.$$

■

Provided you can find an antiderivative of you now have a way to evaluate a definite integral without having to use the limit of a sum.

Remark 11.2 *When applying the Fundamental Theorem of Calculus, the following notation is convenient:*

$$\int_a^b f(x)dx = F(x)|_a^b = F(b) - F(a).$$

For instance, to evaluate $\int_1^2 x^2 dx$*, you can write*

$$\int_1^2 x^2 dx = \frac{x^3}{3}|_1^2 = \frac{2^3}{3} - \frac{1^3}{3} = \frac{7}{3}.$$

Remark 11.3 *It is not necessary to include a constant of integration in the antiderivative because*

$$\begin{aligned}\int_a^b f(x)dx &= (F(x)+C)\,|_a^b \\ &= (F(b)+C)-(F(a)+C) \\ &= F(b)-F(a).\end{aligned}$$

Remark 11.4 *Integrating a function of x over an interval and integrating the same function of t over the same interval of integration produce the same value for the integral. For example,*

$$\int_1^2 x^2 dx = \frac{x^3}{3}|_1^2 = \frac{7}{3}, \qquad \int_1^2 t^2 dt = \frac{t^3}{3}|_1^2 = \frac{7}{3}.$$

Because the variable of integration in a definite integral plays no role in the end result, it is often referred to as a ***dummy variable****. Whenever you find it convenient to change the letter used for the variable of integration in a definite integral, you can do so without changing the value of the integral. Whenever you find it convenient to change the letter used for the variable of integration in a definite integral, you can do so without changing the value of the integral.*

Example 11.5 *Evaluate the definite integral:*

1. $\int_1^2 x^3 dx$,
2. $\int_0^\pi \cos x dx$,
3. $\int_0^{\pi/4} \sec^2 x dx$
4. $\int_0^4 |x-3|\, dx$

Solution:

1. The function $G = \dfrac{x^4}{4}$ is an antiderivative of $f(x) = x^3$. We now evaluate G at the 2 endpoints and subtract.

$$\int_1^2 x^3 dx = \frac{x^4}{4}|_1^2 = \frac{2^4}{4} - \frac{1^4}{4} = \frac{15}{4}.$$

2. $\int_0^\pi \cos x dx = \sin x|_0^\pi = \sin\pi - \sin 0 = 0.$
3. $\int_0^{\pi/4} \sec^2 x dx = \tan x|_0^{\pi/4} = \tan\dfrac{\pi}{4} - \tan 0 = 1.$

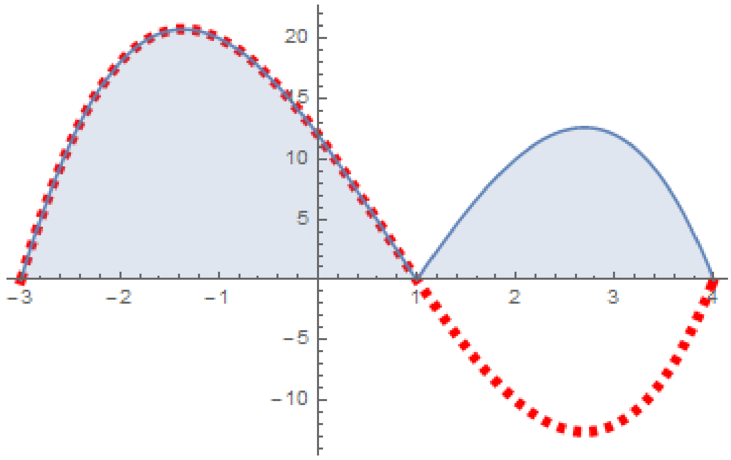


Fig. 11.1. How to calculate a positive area.

4. If you are dealing with absolute value functions, you might have to split up the interval of integration. In this case

$$\begin{aligned}\int_0^4 |x-3|\,dx &= \int_0^3 |x-3|\,dx + \int_3^4 |x-3|\,dx \\ &= \int_0^3 (3-x)\,dx + \int_3^4 (x-3)\,dx \\ &= \frac{9}{2} + \frac{1}{2} = 5 \qquad \square\end{aligned}$$

Example 11.6 *Calculate the (positive) area, between the graph of* $f(x) = x^3 - 2x^2 - 11x + 12$ *and the* x*-axis, between* $x = -3$ *and* $x = 4$*. The graph of* $f(x)$ *is shown in Figure 11.1 (red dashed line).*

Solution: It is easy to check, that f is nonnegative on $[-3, 1]$ and nonpositive on $[1, 4]$. From the discussion preceding this example, we know then how to find the positive (shaded) area

$$\begin{aligned} Area &= \int_{-3}^{1} f(x)dx - \int_{1}^{4} f(x)dx \\ &= \left(\frac{1}{4}x^4 - \frac{2}{3}x^3 - \frac{11}{2}x^2 + 12x\right)\Big|_{-3}^{1} \\ &\quad - \left(\frac{1}{4}x^4 - \frac{2}{3}x^3 - \frac{11}{2}x^2 + 12x\right)\Big|_{1}^{4} \\ &= \frac{160}{3} + \frac{99}{4} \\ &= \frac{937}{12} \qquad \square \end{aligned}$$

Example 11.7 *Compute* $\int_1^2 \frac{(s+5)^2}{s^4} ds$

Solution: The integrand may be broken apart:

$$\frac{(s+5)^2}{s^4} = \frac{s^2 + 10s + 25}{s^4} = \frac{1}{s^2} + \frac{10}{s^3} + \frac{25}{s^4}.$$

We can find an antiderivative term by term, by the power rule:

$$\begin{aligned} &\int_1^2 \left(\frac{1}{s^2} + \frac{10}{s^3} + \frac{25}{s^4}\right) ds \\ &= \left(-\frac{1}{s} - \frac{10}{2s^2} - \frac{25}{3s^3}\right)\Big|_1^2 \\ &= \frac{277}{24} \approx 11.542. \end{aligned}$$

□

11.2 Average Value of a Function

It is easy to calculate the average value of finitely many numbers $y_1, y_2, ..., y_n$:

$$y_{ave} = \frac{y_1 + y_2 + ... + y_n}{n}.$$

But how do we compute the average temperature during a day if infinitely many temperature readings are possible? Figure 11.2 shows the graph of a temperature function $T(t)$, where t is measured in hours and T in ^oC, and a guess at the average temperature, T_{ave}.

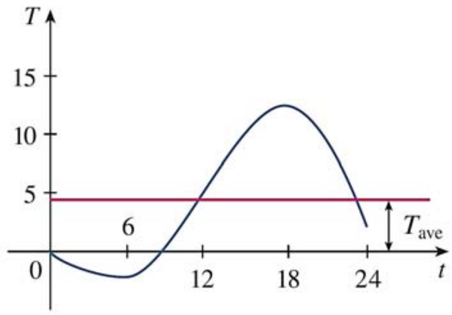


Fig. 11.2. A guess at the average temperature, T_{ave}.

In general, let's try to compute the average value of a function $y = f(x)$, $a \leq x \leq b$. We start by dividing the interval $[a, b]$ into n equal subintervals, each with length $\Delta x = (b - a)/n$. Then we choose points $c_1, c_2, ..., c_n$ in successive subintervals and calculate the average of the numbers $f(c_1), f(c_2), ..., f(c_n)$:

$$\frac{f(c_1) + f(c_2) + ... + f(c_n)}{n}.$$

Since $\Delta x = (b - a)/n$ we can write $n = (b - a)/\Delta x$ and the average value becomes

$$\begin{aligned}
\frac{f(c_1) + f(c_2) + ... + f(c_n)}{\frac{(b-a)}{\Delta x}} &= \frac{1}{(b-a)} \left[f(c_1)\Delta x + f(c_2) + \Delta x ... + f(c_n)\Delta x \right] \\
&= \frac{1}{(b-a)} \sum_{i=1}^{n} f(c_i)\Delta x.
\end{aligned}$$

If we let n increase, we would be computing the average value of a large number of closely spaced values. (For example, we would be averaging temperature readings taken every minute or even every second.) If $f(x)$ is integrable, then the limiting value is

$$\lim_{n \to \infty} \frac{1}{(b-a)} \sum_{i=1}^{n} f(c_i)\Delta x = \frac{1}{(b-a)} \int_a^b f(x)dx$$

by the definition of a definite integral.

Therefore, we define the *average value* of f on the interval $[a, b]$ as

$$\boxed{f_{ave} = \frac{1}{(b-a)} \int_a^b f(x)dx} \tag{11.2}$$

Example 11.8 *Find the average value of the function* $f(x) = 1 + x^3$ *on the interval* $[-1, 2]$.

Solution: With $a = -1$ and $b = 2$ we have

$$\begin{aligned} f_{ave} &= \frac{1}{(b-a)}\int_a^b fx)dx = \frac{1}{2-(-1)}\int_{-1}^2 (1+x^3)dx \\ &= \frac{1}{3}\left(x + \frac{x^4}{4}\right)|_{-1}^2 = \frac{9}{4}. \qquad \square \end{aligned}$$

If $T(t)$ is the temperature at time t, we might wonder if there is a specific time when the temperature is the same as the average temperature. For the temperature function graphed in Figure 11.2, we see that there are two such times—just before noon and just before midnight. In general, is there a number c at which the value of a function f is exactly equal to the average value of the function, that is, $f(c) = f_{ave}$? The following theorem says that this is true for continuous functions.

Theorem 11.9 (The Mean Value Theorem for Integrals) *If* f *is continuous on the closed interval* $[a, b]$ *then there exists a number* c *in the closed interval* $[a, b]$ *such that*

$$\int_a^b f(x)dx = f(c)(b-a). \tag{11.3}$$

Proof.

Case 1: If is constant on the interval $[a, b]$ the theorem is clearly valid because c can be any point in $[a, b]$.

Case 2: If f is not constant on $[a, b]$ then, by the Extreme Value Theorem, you can choose $f(m)$ and $f(M)$ to be the minimum and maximum values of f on $[a, b]$.Because $f(m) \leq f(x) \leq f(M)$ for all in you can apply the domination property 10.16 to write the following.

$$\int_a^b f(m)dx \leq \int_a^b f(x)dx \leq \int_a^b f(M)dx$$

$$f(m)(b-a) \leq \int_a^b f(x)dx \leq f(M)(b-a)$$

$$f(m) \leq \frac{1}{(b-a)}\int_a^b f(x)dx \leq f(M)$$

From the third inequality, you can apply the Intermediate Value Theorem to conclude that there exists some c in $[a, b]$ such that

$$f(c) = \frac{1}{(b-a)} \int_a^b f(x)dx \qquad \text{or} \qquad f(c)(b-a) = \int_a^b f(x)dx.$$

■

Remark 11.10 *Notice that Theorem 11.9 does not specify how to determine c. It merely guarantees the existence of at least one number c in the interval $[a, b]$.*

Since $f(x) = x^2$ is continuous on the interval $[1, 4]$, the Mean-Value Theorem for Integrals guarantees that there is a point x^* in $[1, 4]$ such that

$$\int_1^4 x^2 dx = f(x^*)(4-1) = 3(x^*)^2.$$

But

$$\int_1^4 x^2 dx = 21$$

so that

$$3(x^*)^2 = 21 \quad \text{or} \quad x^* = \pm\sqrt{7}.$$

Thus, $x^* = \sqrt{7} \approx 2.65$ is the point in the interval $[1, 4]$ whose existence is guaranteed by the Mean-Value Theorem for Integrals. □

11.3 Second fundamental theorem of calculus

Earlier you saw that the definite integral of f on the interval was defined using the constant b as the upper limit of integration and x as the variable of integration. However, a slightly different situation may arise in which the variable x is used in the upper limit of integration. To avoid the confusion of using x in two different ways, t is temporarily used as the variable of integration. (Remember that the definite integral is *not* a function of its variable of integration.)

Example 11.11 (The definite integral as a function) *Evaluate the function*

$$F(x) = \int_0^x \sin t dt$$

at $x = 0$, $x = \pi/6$, $x = \pi/4$, $x = \pi/3$ *and* $x = \pi/2$.

Solution: You could evaluate five different definite integrals, one for each of the given upper limits. However, it is much simpler to fix x (as a constant) temporarily to obtain

$$\int_0^x \sin t dt = 1 - \cos x = F(x).$$

Now, using you can obtain the results shown in the following table:

$$\begin{array}{l} F(0) = 0 \\ F(\frac{\pi}{6}) = 1 - \frac{1}{2}\sqrt{3} \approx 0.13397 \\ F(\frac{\pi}{4}) = 1 - \frac{1}{2}\sqrt{2} \approx 0.29289 \\ F(\frac{\pi}{3}) = \frac{1}{2} \\ F(\frac{\pi}{2}) = 1 \end{array}$$

You can think of the function $F(x)$ as *accumulating* the area under the curve $f(t) = \sin t$ from $t = 0$ to $t = x$.For $x = 0$ the area is 0 and $F(0) = 0$. For $x = \frac{\pi}{2}$, $F(\frac{\pi}{2}) = 1$ gives the accumulated area under the sine curve on the entire interval $[0, \frac{\pi}{2}]$. This interpretation of an integral as an *accumulation function* is used often in applications of integration.

In Example 11.11, note that the derivative of F is the original integrand (with only the variable changed). That is,

$$\frac{d}{dx}[F(x)] = \frac{d}{dx}[1 - \cos x] = \frac{d}{dx}\left[\int_0^x \sin t dt\right] = \sin x.$$

This result is generalized in the following theorem, called the Second Fundamental Theorem of Calculus.

Theorem 11.12 (Second Fundamental Theorem of Calculus) *If f is continuous on an open interval I containing a then, for every x in the interval*

$$\frac{d}{dx}\left[\int_a^x f(t)dt\right] = f(x). \tag{11.4}$$

Proof. Begin by defining F as

$$F(x) = \int_a^x f(t)dt.$$

Then, by the definition of the derivative, you can write

$$\begin{aligned} F'(x) &= \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{1}{\Delta x}\left[\int_a^{x+\Delta x} f(t)dt - \int_a^x f(t)dt\right] \\ &= \lim_{\Delta x \to 0} \frac{1}{\Delta x}\left[\int_x^{x+\Delta x} f(t)dt\right]. \end{aligned}$$

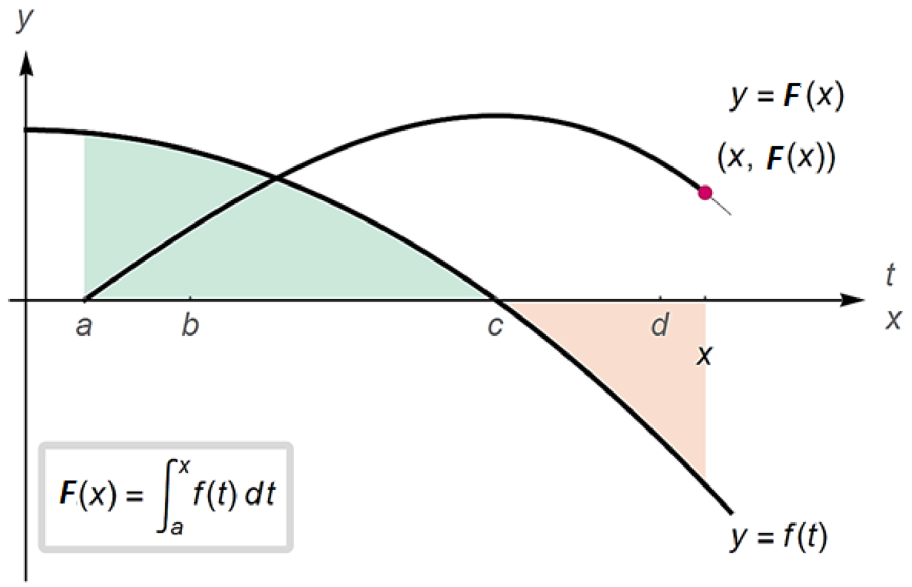


Fig. 11.3.

From the Mean Value Theorem for Integrals (assuming $\Delta x > 0$) you know there exists a number c in the interval $[x, x + \Delta x]$ such that the integral in the expression above is equal to $f(c)\Delta x$. Moreover, because $x \leq c \leq x + \Delta x$ it follows that $c \to x$ as $\Delta x \to 0$. So, you obtain

$$\begin{aligned} F'(x) &= \lim_{\Delta x \to 0} \left[\frac{1}{\Delta x} f(c)\Delta x \right] \\ &= \lim_{\Delta x \to 0} f(c) \\ &= f(x). \end{aligned}$$

A similar argument can be made for $\Delta x < 0$. ■

Example 11.13 (Using the Second Fundamental Theorem of Calculus)

$$\frac{d}{dx}\left[\int_1^x 5t^4 dt\right] = 5x^4$$
$$\frac{d}{dx}\left[\int_1^x \sqrt{t^2+1} dt\right] = \sqrt{x^2+1}$$
$$\frac{d}{dx}\left[\int_x^1 e^{-t^2} dt\right] = \frac{d}{dx}\left[-\int_1^x e^{-t^2} dt\right] = -e^{-x^2}$$
$$\frac{d}{dx}\left[\int_{-e}^x 2^{\cos(t^2+1)} dt\right] = 2^{\cos(x^2+1)}$$

11.4 Variation: tricky limits of integration

Consider

$$\frac{d}{dx}\int_0^{x^2} \cos t dt.$$

Because the right-hand limit of integration is x^2, not x, we can't just use the Second Fundamental Theorem directly. We're going to need the chain rule as well. Start 0 by letting y be the quantity we want to differentiate

$$y = \int_0^{x^2} \cos t dt.$$

We want to find dy/dx. Since y is really a function of x^2, not x directly, we should let $u = x^2$. This means that

$$y = \int_0^{u} \cos t dt.$$

The chain rule says that

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

while the First Fundamental Theorem says that

$$\frac{dy}{du} = \int_0^{u} \cos t dt = \sin u.$$

Also, since $u = x^2$, we have $du/dx = 2x$. Altogether,

$$\frac{dy}{dx} = 2x \sin u.$$

Now all we have to do is replace u by x^2 to see that

$$\frac{dy}{dx} = 2x \sin x^2.$$

In summary,

$$\frac{d}{dx}\int_0^{x^2} \cos t dt = 2x \sin x^2$$

Not so bad when you break it down into little pieces. Let's look at one more example of this sort of problem: what is

$$\frac{d}{dx}\int_x^{x+2} (4t+1)\, dt \text{ ?}$$

Now there are functions of x in both the left-hand and right-hand limits of integration. The way to handle this is to split the integral into two pieces at some number. It actually doesn't matter where you split it, as long as it is at a constant (where the function is defined). So, pick your favorite number–say 0 and split the integral there:

$$\begin{aligned}\int_x^{x+2} (4t+1)\,dt &= \int_x^0 (4t+1)\,dt + \int_0^{x+2} (4t+1)\,dt \\ &= -\int_0^x (4t+1)\,dt + \int_0^{x+2} (4t+1)\,dt\end{aligned}$$

We've reduced the problem to two easier derivatives. Now it is easy to check, that

$$\frac{d}{dx}\int_x^{x+2} (4t+1)\,dt = -(4x+1) + 4(x+2) + 1 = 8$$

11.5 Net Change Theorem

The First Fundamental Theorem of Calculus (Theorem 11.1) states that if f is continuous on the closed interval $[a,b]$ and F is an antiderivative of f on $[a,b]$ then

$$\int_a^b f(x)dx = F(b) - F(a).$$

But because $F'(x) = f(x)$ this statement can be rewritten as

$$\int_a^b F'(x)dx = F(b) - F(a).$$

where the quantity $F(b) - F(a)$ represents the *net change* of F on the interval $[a,b]$.

Theorem 11.14 (Net Change Theorem) *The definite integral of the rate of change of a quantity $F'(x)$ gives the* total change, *or* net change, *in that quantity on the interval* $[a,b]$.

$$\int_a^b F'(x)dx = F(b) - F(a).$$

Popular way to illustrate the Net Change Theorem is to examine the velocity of a particle moving along a straight line where $s(t)$ is the position at time t. Then its velocity is $v(t) = s'(t)$ and

$$\int_a^b v(t)dt = s(b) - s(a).$$

This definite integral represents the net change in position, or displacement, of the particle.

When calculating the *total* distance traveled by the particle, you must consider the intervals where $v(t) \leq 0$ and the intervals where $v(t) \geq 0$. When $v(t) \leq 0$ the particle moves to the left, and when $v(t) \geq 0$ the particle moves to the right. To calculate the total distance traveled, integrate the absolute value of velocity $|v(t)|$. So, the displacement of a particle and the total distance traveled by a particle over can be written as

Displacement on $[a, b] = \int_a^b v(t)dt$

Total distance traveled on $[a, b] = \int_a^b |v(t)|\, dt.$

11.6 The area between two curves

Frequently it is useful to find the area between two curves. See Fig.11.4 . Following the model that we have set up earlier, we first note that the intersected region has left endpoint at $x = a$ and right endpoint at $x = b$. Moreover $f(x) \geq g(x)$ for all $x \in [a, b]$.We partition the interval $[a, b]$ as usual. Call the partition

$$P = \{x_0, x_1, x_2, ..., x_{n-1}, x_n\}$$

Then, we erect rectangles over the intervals determined by the partition. Notice that the upper curve, over the interval $[a, b]$, is $y = f(x)$ and the lower curve is $y = g(x)$. Assuming $c_i \in [x_{i-1}, x_i]$ the sum of the areas of the rectangles is therefore

$$\sum_{i=1}^{n} (f(c_i) - g(c_i))\, \Delta x_i.$$

But of course this is a Riemann sum for the integral

$$\int_a^b (f(x) - g(x))\, dx.$$

We declare this integral to be the area determined by the two curves. So, we can obtain the searched area of by integrating the *vertical separation* $f(x) - g(x)$ from $x = a$ to $x = b$.

Example 11.15 *Find the area between the curves* $y = x^2 - 2$ *(lower) and* $y = -(x-1)^2 + 3$ *(upper).*

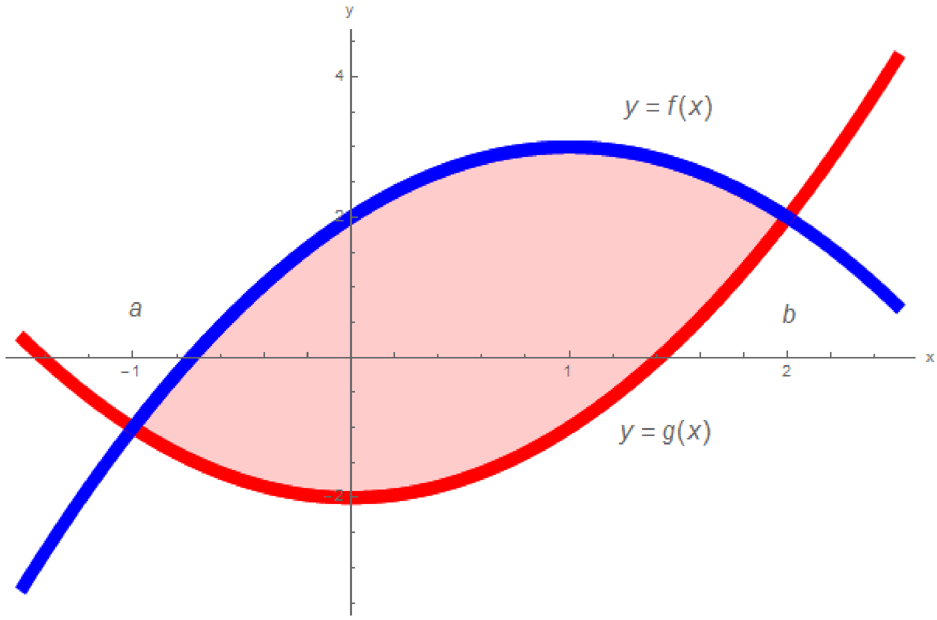


Fig. 11.4. Area between two curves.

Solution: We set the two equations equal and solve to find that the curves intersect at $x = -1$ and $x = 2$. The situation is shown in Fig. 11.4. Notice that $y = -(x-1)^2 + 3$ is the upper curve and $y = x^2 - 2$ is the lower curve. Thus the desired area is

$$\begin{aligned} Area &= \int_{-1}^{2} \left(-(x-1)^2 + 3\right) - \left(x^2 - 2\right) dx \\ &= \int_{-1}^{2} \left(-2x^2 + 2x + 4\right) dx = 9. \qquad \square \end{aligned}$$

Example 11.16 (Curves that intersect at more than 2 points) *Find the area of the region between the graphs* $f(x) = 3x^3 - x^2 - 10x$ *and* $g(x) = -x^2 + 2x$ *(see Figure 11.5).*

Solution: Set the equations equal to each other to find the points of intersection:

$$3x^3 - x^2 - 10x = -x^2 + 2x$$

$$3x^3 - 12x = 0$$

$$3x\,(x-2)\,(x+2) = 0$$

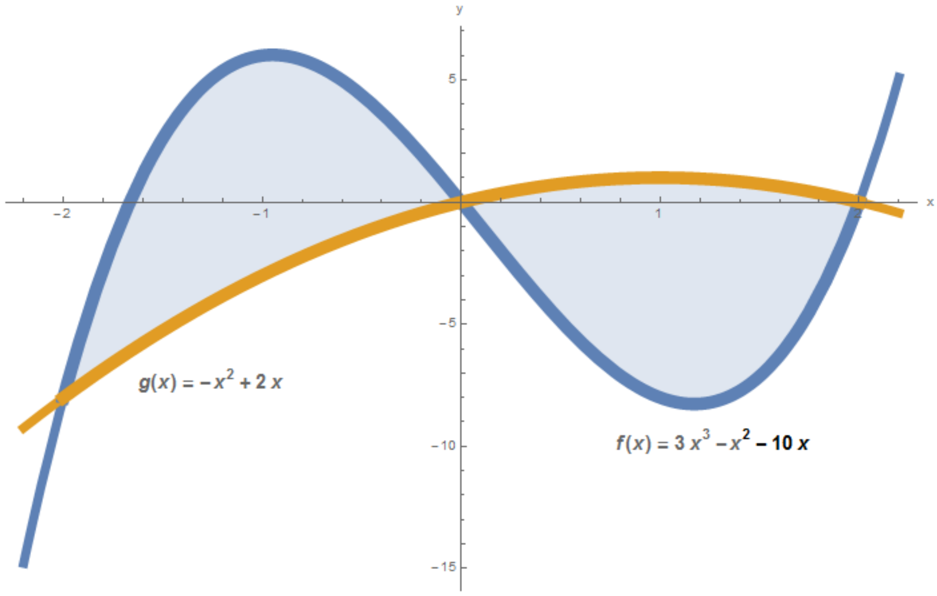


Fig. 11.5. Curves that intersect at more than 2 points.

The curves intersect at 3 points: $(0,0)$, $(2,0)$, and $(-2,-8)$. On the interval $-2 \leq x \leq 0$, the graph of f is above that of g, whereas on the interval $0 \leq x \leq 2$, the graph of g is above that of f. Hence, the area is given by the 2 integrals shown below.

$$Area = \int_{-2}^{0} (f(x) - g(x))\, dx + \int_{0}^{2} (g(x) - f(x))\, dx$$

Although these integrals can be time consuming, they are easy to evaluate. We obtain

$$Area = 12 + 12 = 24 \qquad \square$$

Notice, that we can interchange the roles played by x and y. In Figure 11.6 you see a region Ω, the boundaries of which are given not in terms of x but in terms of y. Here we set the representative rectangles horizontally and calculate the area of the region as the limit of sums of the form

$$[F(d_1) - G(d_1)]\, \Delta y_1 + [F(d_2) - G(d_2)]\, \Delta y_2 + ... + [F(d_n) - G(d_n)]\, \Delta y_n$$

where $y_{i-1} \leq d_i \leq y_i$. These are Riemann sums for the integral of $F - G$. The area formula now reads

$$Area = \int_{c}^{d} [F(y) - G(y)]\, dy \tag{11.5}$$

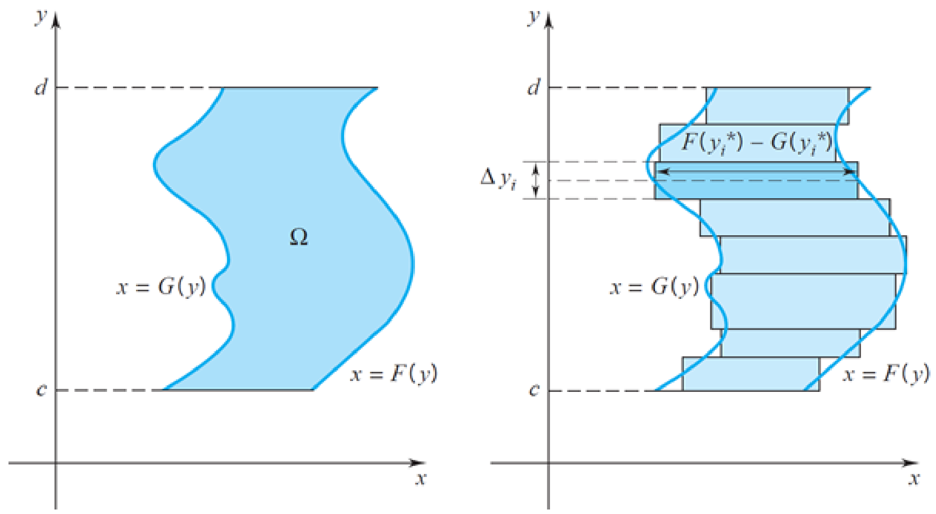


Fig. 11.6. Areas obtained by integration with respect to y.

In this case we are integrating with respect to y the horizontal separation $F(y) - G(y)$ from $y = c$ to $y = d$.

Example 11.17 *Calculate the area of the region bounded by the curves* $x = y^2$ *and* $x - y = 2$ *first*

a) *by integrating with respect to* x *and then (b) by integrating with respect to* y*,*

and then

b) *by integrating with respect to* y*.*

Solution: Simple algebra shows that the two curves intersect at the points $(1, -1)$ and $(4, 2)$.

a) To obtain the area of the region by integration with respect to x, we set the representative rectangles vertically and express the bounding curves as functions of x. Solving $x = y^2$ for y we get $y = \pm\sqrt{x}$; $y = \sqrt{x}$ is the upper half of the parabola and $y = -\sqrt{x}$. The equation of the line can be written $y = x - 2$. (See Figure 11.7.) The upper boundary of the region is the curve $y = \sqrt{x}$. However, the lower boundary consists of two parts: $y = -\sqrt{x}$ from $x = 0$ to $x = 1$, and $y = x - 2$ from $x = 1$ to $x = 4$.

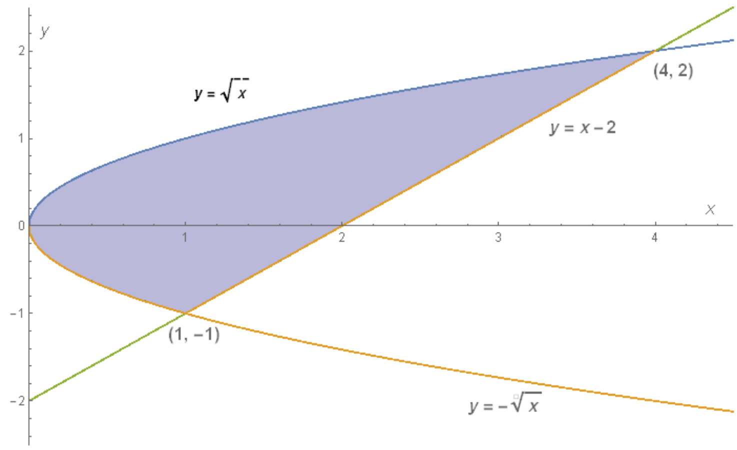


Fig. 11.7.

Thus, we use two integrals:

$$\begin{aligned} Area &= \int_0^1 \left(\sqrt{x} - (-\sqrt{x})\right) dx + \int_1^4 \left(\sqrt{x} - (x-2)\right) dx \\ &= 2\int_0^1 \sqrt{x}dx + \int_1^4 \left(\sqrt{x} - x + 2\right) dx \\ &= \frac{4}{3}x^{\frac{3}{2}}|_0^1 + \left(2x - \frac{1}{2}x^2 + \frac{2}{3}x^{\frac{3}{2}}\right)|_1^4 \\ &= \frac{4}{3} + \frac{19}{6} = \frac{9}{2}. \end{aligned}$$

b) To obtain the area by integration with respect to y, we set the representative rectangles horizontally. (See Figure 11.8) The right boundary is the line $x = y + 2$ and the left boundary is the curve $x = y^2$. Since y ranges from -1 to 2,

$$\begin{aligned} Area &= \int_{-1}^2 \left((y+2) - y^2\right) dy \\ &= \left(-\frac{1}{3}y^3 + \frac{1}{2}y^2 + 2y\right)|_{-1}^2 \\ &= \frac{9}{2}. \end{aligned}$$

In this instance integration with respect to y was the more efficient route to take. □

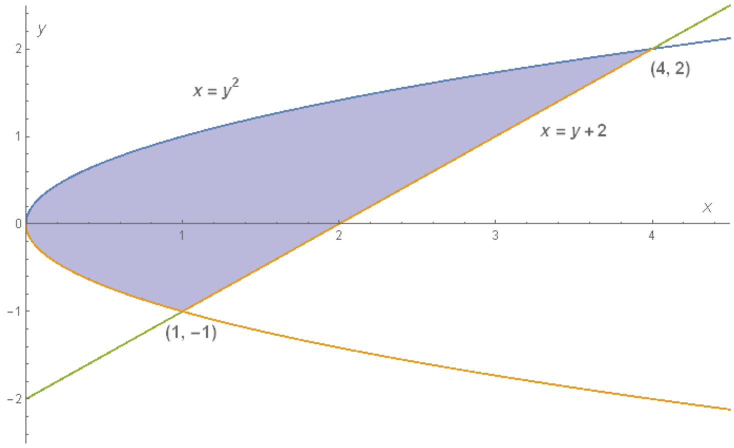


Fig. 11.8.

11.7 Review exercises: Chapter 11

Exercise 11.1 *Check the following integral*

$$\int \frac{3t^2}{\left(1+t^3\right)^2}dt = -\frac{1}{t^3+1} + c$$

and evaluate

$$\int_0^1 \frac{3t^2}{\left(1+t^3\right)^2}dt.$$

`Answer:` $\frac{1}{2}$

Exercise 11.2 *Calculate the derivative of*

$$\frac{x^3}{x^2+1}$$

and find

$$\int_0^1 \frac{\left(3x^2+x^4\right)}{\left(1+x^2\right)^2}dx.$$

`Answer:` $\frac{1}{2}$

Exercise 11.3 *Differentiate*

$$\frac{x}{1+x} \qquad \text{and} \qquad -\frac{1}{1+x}.$$

Next find

$$\int_3^2 \frac{1}{(1+x)^2} dx$$

in two ways.

`Answer:` : $-\frac{1}{12}$

Exercise 11.4 *If*

$$\int_0^1 f(x)dx = 3, \qquad \int_1^2 f(x)dx = 4, \quad \textit{and} \quad \int_2^3 f(x)dx = -8$$

calculate the following quantities using the properties of integration.

a) $\int_0^2 f(x)dx$,

b) $\int_0^1 3f(x)dx$,

c) $\int_0^3 8f(x)dx$,

d) $\int_1^3 10f(x)dx$.

Exercise 11.5 *Calculate the definite integrals:*

a) $\int_{-2}^3 \left(x^4 + 5x^3 + 2x + 1\right) dx$, *Answer:* $\frac{585}{4}$

b) $\int_0^2 x^6 dx$, *Answer:* $\frac{128}{7}$

c) $\int_1^2 \frac{x^2 + 2x + 2}{x^4} dx$, *Answer:* $\frac{11}{6}$

d) $\int_2^3 \frac{dt}{t^2}$, *Answer:* $\frac{1}{6}$

e) $\int_0^{476} 0 ds$, *Answer:* 0

f) $\int_2^{-2} t^4 dt$, *Answer:* $-\frac{64}{5}$

g) $\int_1^8 \frac{1+\theta^2}{\theta^4} d\theta$, *Answer:* $\frac{1855}{1536}$

h) $\int_2^{-1} \left(1 - x^2\right)^3 dx$, *Answer:* $\frac{162}{35}$

i) $\int_8^4 (x^2 - 1)dx$, *Answer:* $-\frac{436}{3}$

j) $\int_9^{10} \dfrac{x+1}{x^3} dx$, *Answer:* $\frac{199}{16\,200}$

k) $\int_{-3}^{3} \frac{x^3-1}{x-1} dx$, *Answer:* 24

Exercise 11.6 *Verify the formula*

$$\frac{d}{dx}\int_a^x f(s)ds = f(x)$$

for the following functions:

a) $f(x) = x^3 - 1$,

b) $f(x) = x^4 - x^2 + 1$.

Exercise 11.7 *Let*

$$F(t) = \int_3^t \frac{1}{\left[(4-s)^2+8\right]^3} ds.$$

Find $F'(4)$.

Exercise 11.8 *Evaluate the following derivatives:*

a) $\dfrac{d}{dt}\int_0^t \dfrac{3}{(x^4-x^2+1)} dx$,

b) $\dfrac{d}{dt}\int_0^t \dfrac{1}{(x^4+1)} dx$,

c) $\dfrac{d}{dt}\int_0^t x^2\left(x^4+1\right)^3 dx$,

d) $\dfrac{d}{dt}\int_0^t \dfrac{u^4}{(u^2+1)} du$.

Exercise 11.9 *Suppose that*

$$f(t) = \begin{cases} t^2, & 0 \le t \le 1, \\ 1, & 1 \le t \le 5, \\ (t-6)^2 & 5 \le t \le 6. \end{cases}$$

a) *Draw a graph off on the interval* $[0, 6]$,

b) *Find* $\int_0^6 f(t)dt$,

c) *Find* $\int_0^6 f(x)dx$,

d) *Let* $F(t) = \int_0^t f(s)ds$. *Find the formula for* $F(t)$ *in* $[0, 6]$ *and draw a graph of* F.

e) *Find* $F'(t)$ *for* t *in* $(0, 6)$.

Exercise 11.10 *Find all values of* x^* *in the stated interval that satisfy Equation (11.3) in the Mean-Value Theorem for Integrals, and explain what these numbers represent.*

a) $f(x) = \sqrt{x}, \quad 0 \leq x \leq 3$,

b) $f(x) = x^2 + x, \quad -12 \leq x \leq 0$,

c) $f(x) = \sin x, \quad -\pi \leq x \leq \pi$,

d) $f(x) = 1/x^2, \quad 1 \leq x \leq 3$.

12
Natural logarithm

There are two types of functions: *polynomial* and *transcendental.* A polynomial of degree k is a function of the form $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_kx^k$. Such a polynomial has precisely k roots (not necessarily different), and there are algorithms that enable us to solve for those roots. For most purposes, polynomials are the most accessible and easy-to-understand functions. But there are other functions that are important in mathematics and physics. These are the *transcendental functions.* Among this more sophisticated type of functions are sine, cosine, the other trigonometric functions, and also the *logarithm* and the *exponential.* The present chapter is devoted to the study of the natural logarithm function from calculus point of view. As each new type of functions is introduced, you will study its properties, its derivative, and its antiderivative. We will also familiar ourself with the irrational (like π) number e. Although written references to the number π go back more than 4000 years, mathematicians first became aware of the special role played by e in the seventeenth century. The notation e was introduced by Leonhard Euler, who discovered many fundamental properties of this important number.

12.1 Calculus gives birth to new function

Recall that the general power rule

$$\int x^n dx = \frac{x^{n+1}}{n+1}$$

has an important disclaimer–it doesn't apply when $n = -1$. Consequently, you have not yet found an antiderivative for the function $f(x) = 1/x$. In this section, you will use the second fundamental theorem of calculus to define such a function. This antiderivative is a function that you have not encountered previously in the text. It is neither algebraic nor trigonometric, but falls into a new class of functions called *logarithmic functions.* This particular function is the *natural logarithmic function.*

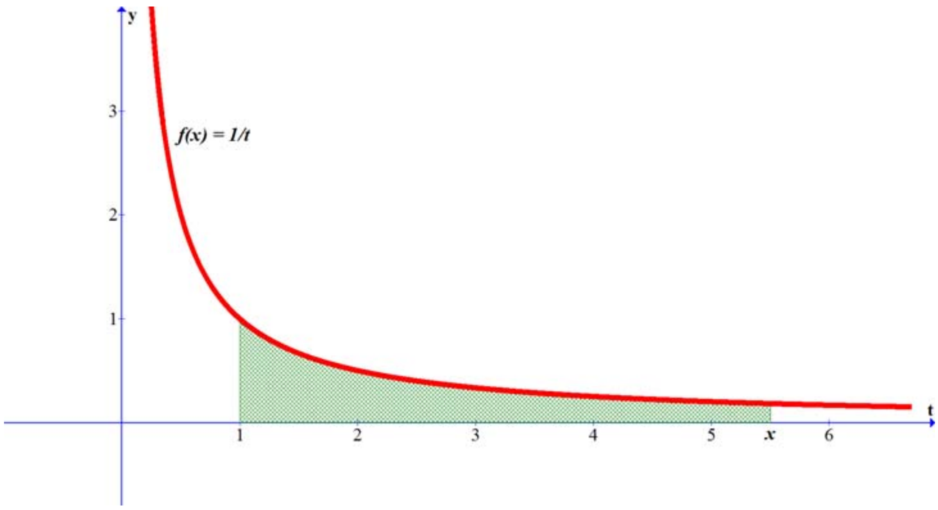


Fig. 12.1. For $x > 1$ he area under the curve $1/t$ from 1 out to x equals $\ln x$.

Definition 12.1 (Natural logarithmic function) *The* natural logarithmic function *is defined by*

$$\ln x = \int_1^x \frac{1}{t} dt, \quad x > 0. \tag{12.1}$$

The domain of the natural logarithmic function is the set of all positive real numbers.

For $x > 1$ you can think about that function as being the area under the curve $1/t$ from 1 out to x (see Figure). For $0 < x < 1$ the value of $\ln x$ is the negative of the actual area between the graph and the x-axis. This is so because the limits of integration, x and 1, occur in reverse order:

$$\ln x = \int_1^x \frac{1}{t} dt, \quad \text{with} \quad x < 1.$$

If you plug in 1 in place of x in 12.1, you will get $\ln 1 = 0$.

To sketch the graph of $\ln x$ you can think of the natural logarithmic function as an antiderivative given by the differential equation

$$\frac{dy}{dx} = \frac{1}{x}.$$

Figure 12.2 is a computer-generated graph, called a *slope* (or *direction*) *field*, showing small line segments of slope $1/x$. The graph of $y = \ln x$ is the solution that passes through the point $(1, 0)$.

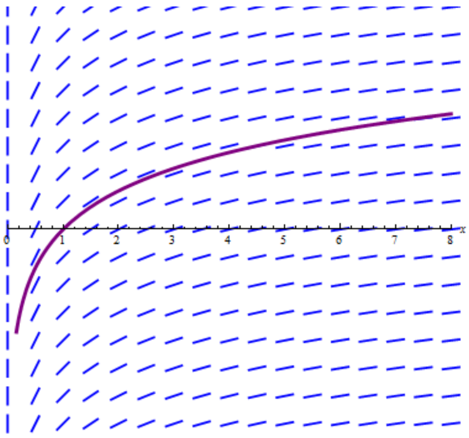


Fig. 12.2. Natural logarithmic function.

Using the definition of the natural logarithmic function, you can prove several important properties involving operations with natural logarithms. If you are already familiar with logarithms, you will recognize that these properties are characteristic of all logarithms.

Theorem 12.2 (Logaritmic properties) *If a and b are positive numbers and n is rational, then the following properties are true.*

1. $\ln(1) = 0$
2. $\ln(ab) = \ln(a) + \ln(b)$
3. $\ln(a^n) = n \ln a$
4. $\ln\left(\frac{a}{b}\right) = \ln a - \ln b$

Proof. The first property has already been discussed. The proof of the second property follows from the fact that two antiderivatives of the same function differ at most by a constant. From the second fundamental theorem of calculus and the definition of the natural logarithmic function, you know that

$$\frac{d}{dx}[\ln x] = \frac{d}{dx}\left[\int_1^x \frac{1}{t} dt\right] = \frac{1}{x}.$$

So, consider the two derivatives

$$\frac{d}{dx}[\ln(ax)] = \frac{a}{ax} = \frac{1}{x},$$

and

$$\frac{d}{dx}[\ln(a) + \ln(x)] = 0 + \frac{1}{x} = \frac{1}{x}.$$

Because $\ln(ax)$ and $\ln(a) + \ln(x)$ are both antiderivatives of $\frac{1}{x}$ they must differ at most by a constant.

$$\ln(ax) = \ln(a) + \ln(x) + C.$$

By letting $x = 1$, you can see that $C = 0$. The third property can be proved similarly by comparing the derivatives of $\ln(x^n)$ and $n \ln x$. Finally, using the second and third properties, you can prove the fourth property.

$$\ln\left(\frac{a}{b}\right) = \ln\left(ab^{-1}\right) = \ln a + \ln\left(b^{-1}\right) = \ln a - \ln b.$$

■

Example 12.3 shows how logarithmic properties can be used to expand logarithmic expressions.

Example 12.3 *Expand the expressions*

a.

$$\begin{aligned} \ln\left(\frac{a^3b^2}{c^{-4}d}\right) &= \ln\left(a^3b^2\right) - \ln\left(c^{-4}d\right) \\ &= \left[\ln a^3 + \ln b^2\right] - \left[\ln c^{-4} + \ln d\right] \\ &= [3\ln a + 2\ln b] - [-4\ln c + \ln d] \\ &= 3\ln a + 2\ln b + 4\ln c - \ln d. \end{aligned}$$

b.

$$\begin{aligned} \ln\frac{6x^2}{11} &= \ln\left[6x^2\right] - \ln 11 \\ &= \ln 6 + \ln x^2 - \ln 11 \\ &= 2\ln x + \ln 6 - \ln 11. \end{aligned}$$

c.

$$\begin{aligned} \ln\sqrt{6x+1} &= \ln(6x+1)^{1/2} \\ &= \frac{1}{2}\ln(6x+1) \end{aligned}$$

d.

$$\begin{aligned}
\ln \frac{(x^2+1)^2}{x\sqrt[3]{x^2+3}} &= \ln(x^2+1)^2 - \ln\left(x\sqrt[3]{x^2+3}\right) \\
&= 2\ln(x^2+1) - \left[\ln x + \ln\left(x^2+3\right)^{1/3}\right] \\
&= 2\ln(x^2+1) - \ln x - \frac{1}{3}\ln\left(x^2+3\right). \quad \square
\end{aligned}$$

When using the properties of logarithms to rewrite logarithmic functions, you must check to see whether the domain of the rewritten function is the same as the domain of the original. For instance, the domain of $f(x) = \ln x^2$ is all real numbers except $x = 0$ and the domain of $g(x) = 2\ln x$ is all positive real numbers.

From the second fundamental theorem of calculus it follows, that we can easily find the derivative of this new function. Namely

$$\frac{d}{dx}[\ln x] = [\ln x]' = \frac{1}{x}.$$

Observe, that this derivative is positive for all $x > 0$, so this function is increasing (but not very fast when x gets larger). The second derivative $[\ln x]''$ is equal to $-1/x^2$ (negative) so, our function is concave down. The following theorem lists some basic properties of the natural logarithmic function.

Theorem 12.4 (Properties of the natural logarithmic function) *The natural logarithmic function has the following properties.*

1. *The domain is* $(0, \infty)$ *and the range is* $(-\infty, \infty)$
2. *The function is continuous, increasing, and one-to-one.*
3. *The graph is concave downward.*

Proof. The domain of $f(x) = \ln x$ is $(0, \infty)$ by definition. Moreover, the function is continuous *because it is differentiable.* It is increasing because its derivative

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

is positive for $x > 0$. It is concave downward because its second derivative

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

is negative for $x > 0$.

We know that $f(x)$ is increasing on its entire domain $(0, \infty)$ and therefore is strictly monotonic. Choose x_1 and x_2 in the domain of f such that $x_1 \neq x_2$. Because is strictly monotonic, it follows that either

$$f(x_1) < f(x_2) \qquad \text{or} \qquad f(x_1) > f(x_2).$$

In either case, $f(x_1) \neq f(x_2)$. So, $f(x) = \ln x$ is one-to-one.

The following limits imply that its range is the entire real line

$$\lim_{x \to 0^+} \ln x = -\infty \qquad \text{and} \qquad \lim_{x \to \infty} \ln x = \infty.$$

To verify the limits, begin by showing that from the mean value theorem for integrals, you can write

$$\ln 2 = \int_1^2 \frac{1}{t} dt = \frac{1}{c}(2-1)$$

where c is in $[1, 2]$. This implies that

$$\begin{aligned} 1 &\leq c \leq 2 \\ 1 &\geq \frac{1}{c} \geq \frac{1}{2} \\ 1 &\geq \ln 2 \geq \frac{1}{2}. \end{aligned}$$

Now, let N be any positive (large) number. Because $\ln x$ is increasing, it follows that if $x > 2^{2N}$, then

$$\ln x > \ln 2^{2N} = 2N \ln 2.$$

However, because $\ln 2 \geq \frac{1}{2}$

$$\ln x > 2N \ln 2 \geq 2N(\frac{1}{2}) = N.$$

This verifies the second limit. To verify the first limit, let $z = 1/x$. Then $z \to \infty$ as $x \to 0^+$, and you can write

$$\begin{aligned} \lim_{x \to 0^+} \ln x &= \lim_{x \to 0^+} \left(-\ln \frac{1}{x} \right) \\ &= \lim_{z \to \infty} (-\ln z) \\ &= -\lim_{z \to \infty} (\ln z) \\ &= -\infty. \end{aligned}$$

■

We will state next limit as a theorem. It says, that $\ln x$ grows slower than any positive power as $x \to \infty$. This statement is the most interesting when r is small.

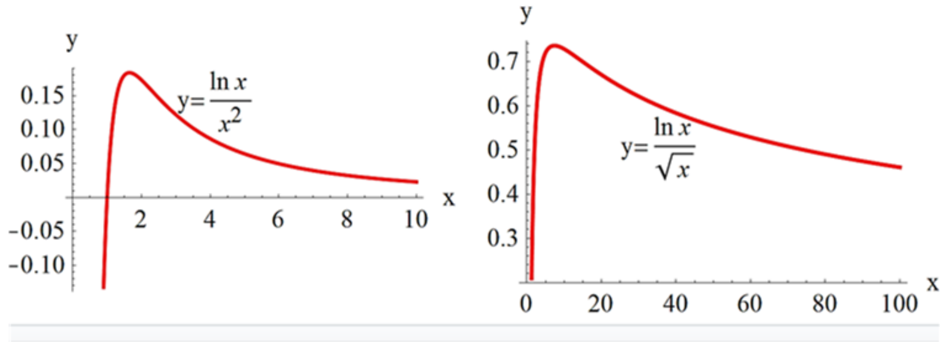


Fig. 12.3. The second function tends to zero much slower than the first when $x \to \infty$.

Theorem 12.5

$$\lim_{x\to\infty} \frac{\ln x}{x^r} = 0 \quad \text{for any } r > 0.$$

Proof. Choose p such that $0 < p < 1$ and $p > 1 - r$. (Notice, that if $r > 1$, then the second inequality is automatically true if the first is true.) By the definition of $\ln x$ and because $t > t^p$ for $x > 1$,

$$\begin{aligned} \ln x &= \int_1^x \frac{1}{t} dt < \int_1^x \frac{1}{t^p} dt = \frac{1}{1-p} t^{1-p}|_1^x \\ &= \frac{1}{1-p}\left(x^{1-p} - 1\right) \end{aligned}$$

and so

$$0 < \frac{\ln x}{x^r} < \frac{1}{1-p}\left(x^{1-p-r} - x^{-r}\right) \to 0 \quad \text{for any } r > 0$$

as both exponents $1 - p - r$ and $-r$ are negative (see Figure 12.3). ■

From the limit we have just proved another important limit follows.

Corollary 12.6

$$\lim_{x\to 0^+} x^r \ln x = 0 \quad \text{for any } r > 0.$$

Proof.

$$\lim_{x\to 0^+} x^r \ln x = \lim_{x\to\infty} \left(\frac{1}{x}\right)^r \ln\left(\frac{1}{x}\right) = -\lim_{x\to\infty} \frac{\ln x}{x^r} = 0.$$

See Figure 12.4. ■

Example 12.7 *Find an equation of the tangent line to the graph of* $f(x) = 3x^2 - \ln x$ *at the point* $(1, 3)$.

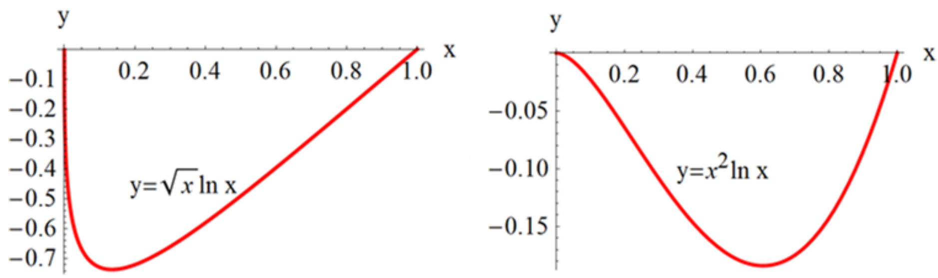


Fig. 12.4. The second function tends to zero much faster than the first when $x \to 0^+$ (compare the slopes)

Solution: Knowing that

$$f'(x) = 6x - \frac{1}{x}$$

we can find the slope of the tangent line

$$f'(1) = 6 - 1 = 5.$$

As this line passes through the point $(1, 3)$ we can use point-slope formula

$$y - 3 = 5(x - 1)$$

and finally (see Figure 12.5)

$$y = 5x - 2. \qquad \square$$

Example 12.8 *The offset logarithmic integral is defined as*

$$\text{Li}(x) = \int_2^x \frac{dt}{\ln t}$$

It is a specific anti-derivative. It is a good approximation of the number of prime numbers less than x. The graph below illustrates this. The second stair graph shows the number $\pi(x)$ of primes below x. For example, $\pi(x) = 4$ because 2, 3, 5, 7 *are the only primes below it. The function* Li(x) *is not an elementary function.*

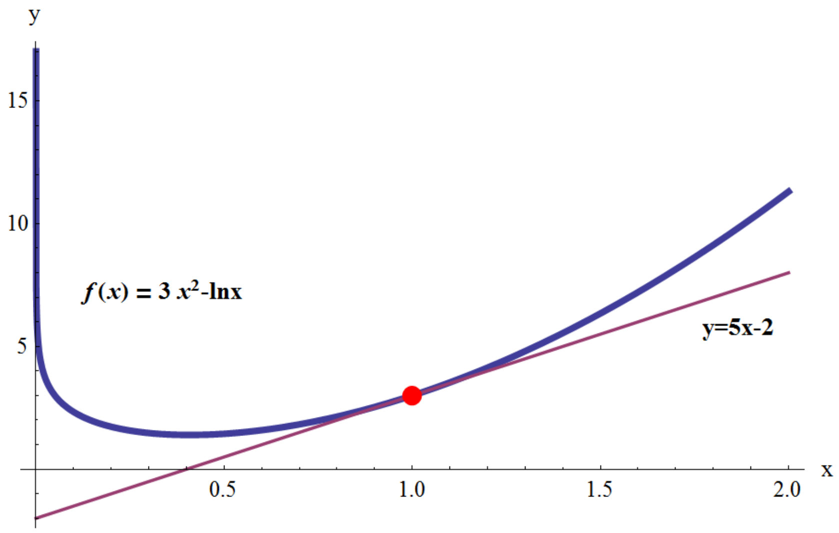


Fig. 12.5. The tangent line from Example 12.7

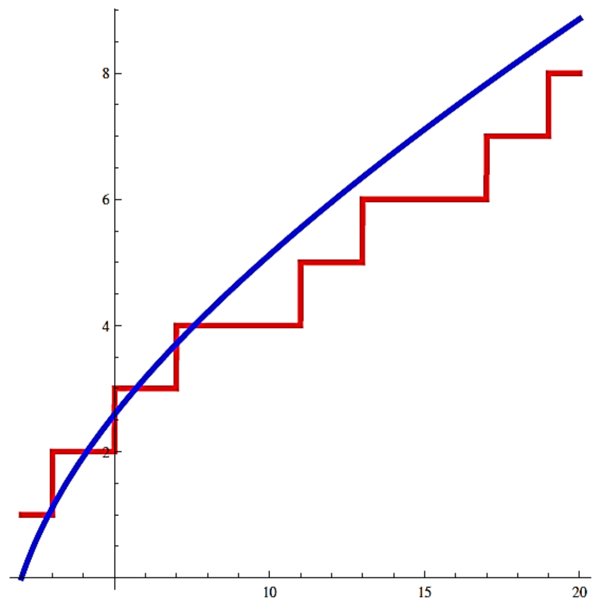


Fig. 12.6. Li(x) versus $\pi(x)$

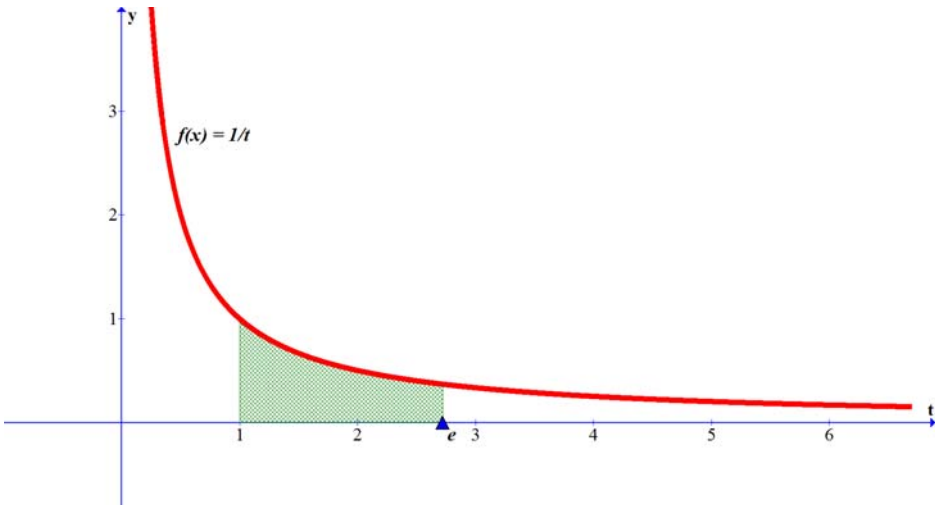


Fig. 12.7. $\ln e = \int_1^e \frac{1}{t} dy = 1$.

12.2 The number e

It is likely that you have studied logarithms in an algebra course. There, without the benefit of calculus, logarithms would have been defined in terms of a base number. For example, *common logarithms* have a base of 10 and therefore $\log_{10} 10 = 1$.

The *base for the natural logarithm* is defined using the fact that the natural logarithmic function is continuous, is one-to-one, and has a range of $(-\infty, \infty)$. Lets calculate approximate value of $\ln 2$ (using the trapezoid rule to approximate $\int_1^2 \frac{1}{t} dt$ for example). However you do it, $\ln 2$ turns out to be approximately equal to 0.6932. Similarly $\ln 3 \approx 1.0986$. Now, from the intermediate value theorem it follows, that between 2 and 3 there exists a special number (call it e) such that (see Figure 12.7).

$$\ln e = \int_1^e \frac{1}{t} dt = 1. \tag{12.2}$$

Definition 12.9 (of e) *The letter e denotes the positive real number such that*

$$\boxed{\ln e = \int_1^e \tfrac{1}{t} dt = 1.}$$

Once you know that $\ln e = 1$ you can use logarithmic properties to evaluate the natural logarithms of several other numbers. For example, by using the property

$$\ln e^n = n \ln e = n$$

you can evaluate $\ln e^n$ for various values of n.

12.3 The derivative of the natural logarithmic function

The derivative of the natural logarithmic function is given in Theorem 12.10. The first part of the theorem follows from the definition of the natural logarithmic function as an antiderivative. The second part of the theorem is simply the chain rule version of the first part.

Theorem 12.10 (Derivative of the natural logarithmic function) *Let u be a differentiable function of x. Then*

$$\textbf{1.}\ \frac{d}{dx}[\ln x] = \frac{1}{x}, \quad x > 0 \qquad \textbf{2.}\ \frac{d}{dx}[\ln u] = \frac{1}{u}\frac{du}{dx} = \frac{u'}{u} \quad u > 0.$$

Example 12.11 (Differentiation of logarithmic functions) *Find the derivatives*

a.

$$\frac{d}{dx}[\ln(3x)] = \frac{1}{3x}\frac{d}{dx}[3x] = \frac{1}{x}.$$

Another way to solve that problem:

$$\frac{d}{dx}[\ln(3x)] = \frac{d}{dx}[\ln 3 + \ln(x)] = \frac{1}{x}$$

b.

$$\frac{d}{dx}\ln(x^4 - x) = \frac{1}{x^4 - x} \cdot \frac{d}{dx}(x^4 - x) = \frac{4x^3 - 1}{x^4 - x}$$

c.

$$\begin{aligned} \frac{d}{dx}[(\ln x)\cdot(\cot x)] &= \left[\frac{d}{dx}\ln x\right]\cdot(\cot x) + (\ln x)\frac{d}{dx}[\cot x] \\ & \frac{1}{x}\cot x - (\ln x)\left(\cot^2 x + 1\right) \end{aligned}$$

d.

$$\frac{d}{dx}[x \ln x] = x \cdot \left(\frac{1}{x}\right) + \ln x \cdot (1) = 1 + \ln x$$

e.

$$\frac{d}{dx}\left[\ln\left(x^4\right)\right] = \frac{d}{dx}\left[4\ln x\right] = \frac{4}{x}$$

Compare it with

$$\frac{d}{dx}\left(\ln x\right)^4 = \frac{4}{x}\ln^3 x.$$

f. Logarithm can help us to find complicated derivatives. Look for examples:

$$\frac{d}{dx}\left(\ln\sqrt{x+2}\right) = \frac{d}{dx}\left(\frac{1}{2}\ln\left(x+2\right)\right) = \frac{1}{2}\frac{1}{(x+2)} = \frac{1}{2\left(x+2\right)}$$

g. Compute

$$\frac{d}{dx}\left(x\sqrt{x^2+1}\right) \quad \text{at} \quad x = 1.$$

We can calculate this derivative using our previous skills, but the new method seems to be more convenient. Let us denote

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

and apply the logarithm to both sizes

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

Using implicit differentiation and the chain rule we get

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

So

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

Now we can plug in our point $x = 1$, obtaining

$$y'(1) = \sqrt{2}\left(1 + \frac{1}{2}\right) = \frac{3}{2}\sqrt{2}$$

h. Find dy/dx if

$$y = \frac{(x^2+1)\sqrt{(x+3)}}{x-1}, \qquad x > 1.$$

We take the natural logarithm of both sides and simplify the result with the properties of logarithms:

$$\begin{aligned} \ln y &= \ln\left(\frac{(x^2+1)\sqrt{(x+3)}}{x-1}\right) \\ &= \ln(x^2+1)+\ln\sqrt{(x+3)}-\ln(x-1) \\ &= \ln(x^2+1)+\frac{1}{2}\ln(x+3)-\ln(x-1)\,. \end{aligned}$$

We then take derivatives of both sides with respect to x:

$$\frac{1}{y}\frac{dy}{dx}=\frac{1}{x^2+1}\cdot(2x)+\frac{1}{2}\cdot\frac{1}{x+3}-\frac{1}{x-1}.$$

Next we solve for dy/dx:

$$\frac{dy}{dx}=y\left(\frac{2x}{x^2+1}+\frac{1}{2\,(x+3)}-\frac{1}{x-1}\right).$$

Finally, we substitute for y from the original equation:

$$\frac{dy}{dx}=\frac{(x^2+1)\sqrt{(x+3)}}{x-1}\cdot\left(\frac{2x}{x^2+1}+\frac{1}{2\,(x+3)}-\frac{1}{x-1}\right) \qquad \square$$

The method used above is called *logarithmic differentiation*

Products, quotients, and powers of functions are usually differentiated using the derivative rules of the same name (perhaps combined with the Chain Rule). There are times, however, when the direct computation of a derivative is very tedious. Consider the function

$$f(x)=\frac{(x^2-2)^3\sqrt{x^2+1}}{x^2+4}$$

We would need the Quotient, Product, and Chain Rules just to compute $f'(x)$, and simplifying the result would require additional work. The properties of logarithms are useful for differentiating such functions.

Example 12.12 *Let* $f(x)=\frac{(x^1+1)^3\sqrt{x^2+1}}{x^2+4}$ *and compute* $f'(x)$.

Solution: We begin by taking the natural logarithm of both sides and simplifying the result:

$$\begin{aligned} \ln(f(x)) &= \ln\left[\frac{(x^2-2)^3\sqrt{x^2+1}}{x^2+4}\right] \\ &= 3\ln(x^2-2)+\frac{1}{2}\ln(x^2+1)-\ln(x^2+4) \end{aligned}$$

We now differentiate both sides using the Chain Rule; specifically the derivative of the left side is $\frac{d}{dx}\left[\ln(f(x))\right] = \frac{f'(x)}{f(x)}$. Therefore,

$$\frac{f'(x)}{f(x)} = 6\frac{x}{x^2-2} + \frac{x}{x^2+1} - 2\frac{x}{x^2+4}.$$

Solving for $f'(x)$, we have

$$f'(x) = f(x)\left[6\frac{x}{x^2-2} + \frac{x}{x^2+1} - 2\frac{x}{x^2+4}\right].$$

Finally, we replace $f(x)$ with the original function:

$$\begin{aligned} f'(x) &= \frac{(x^2-2)^3\sqrt{x^2+1}}{x^2+4}\left(6\frac{x}{x^2-2} + \frac{x}{x^2+1} - 2\frac{x}{x^2+4}\right) \\ &= \frac{x}{\sqrt{x^2+1}}\frac{\left(x^2-2\right)^2}{\left(x^2+4\right)^2}\left(5x^4+34x^2+20\right) \qquad \square \end{aligned}$$

Logarithmic differentiation also provides an alternative method for finding derivatives of functions of the form $g(x)^{h(x)}$.

12.4 Derivative involving absolute value

Since we have only defined the function $\ln x$ when $x > 0$, the graph is only sketched in Fig. 12.2 to the right of the y-axis. However it certainly makes sense to discuss the function $\ln|x|$ when $x \neq 0$ (Figure 12.8.)

The following theorem states that you can differentiate functions of the form $y = \ln|u|$ as if the absolute value notation was not present.

Theorem 12.13 *If u is a differentiable function of x such $u \neq 0$ that then*

$$\frac{d}{dx}\ln|u| = \frac{u'}{u}. \tag{12.3}$$

Proof. If $u > 0$ then and the result follows from Theorem 12.10. If $u < 0$ then $|u| = -u$, and we have

$$\frac{d}{dx}\ln|u| = \frac{d}{dx}\ln(-u) = \frac{-u'}{-u} = \frac{u'}{u}.$$

■

Example 12.14 *Find the derivative of $f(x) = \ln|\sin x|$.*

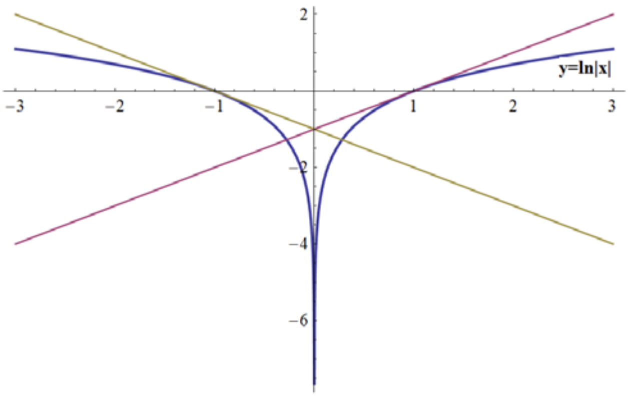


Fig. 12.8. The graph of the function $\ln|x|$ and two its tangent lines at $(-1, 0)$ and $(1, 0)$.

Solution:

$$f'(x) = \frac{\cos x}{\sin x} = \cot x, \quad u = \sin x > 0 \quad \text{on} \quad (0, \pi). \qquad \square$$

Equation (12.3) leads to the following integral formula.

$$\boxed{\begin{array}{c}\text{If } u \text{ is a differentiable function that is never zero,}\\ \int \frac{1}{u} du = \ln|u| + C.\end{array}} \tag{12.4}$$

It says that integrals of a certain form lead to logarithms and allows us (for example) to integrate certain trigonometric functions. We will use those formulas for developing some integration techniques.

Example 12.15 *The integral of $2/x$ is*

$$\int \frac{2}{x} dx = 2 \int \frac{1}{x} dx = 2 \ln|x|^2 + C = 2 \ln x^2 + C.$$

Note, that the absolute value sign has disappeared, because x^2 is always positive for nonzero arguments. $\square$

Example 12.16 *Calculate $\int \frac{1}{5x+2} dx$ using Formula 12.4.*

$$\int \frac{1}{5x+2} dx = \frac{1}{5} \int \frac{5}{5x+2} dx = \frac{1}{5} \ln|5x+2| + C$$

or (equivalently)

$$\int \frac{1}{5x+2} dx = \frac{1}{5} \int \frac{1}{x + \frac{2}{5}} dx = \frac{1}{5} \ln\left(\left|x + \frac{2}{5}\right|\right) + C.$$

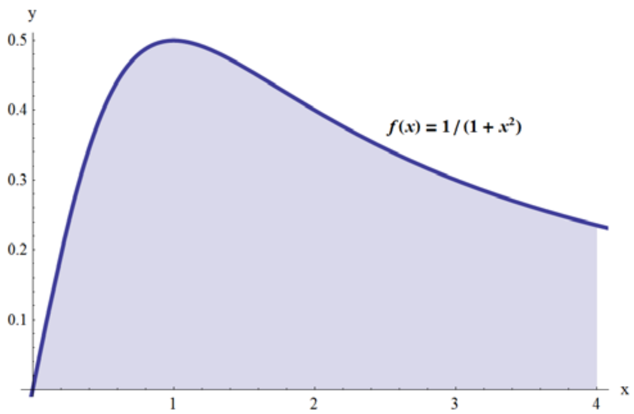


Fig. 12.9. The region described in the Example 12.18.

Example 12.17 *The integrals of* $\tan x$ *and* $\cot x$*.*

$$\begin{aligned}\int \tan x dx &= \int \frac{\sin x}{\cos x} dx \\ &= \int \frac{-du}{u} \qquad u = \cos x > 0 \quad \textit{on} \quad \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \\ &= -\ln|u| + C = -\ln|\cos x| + C \\ &= \ln \frac{1}{|\cos x|} + C = \ln|\sec x| + C.\end{aligned}$$

Similarly,

$$\begin{aligned}\int \cot x dx &= \int \frac{\cos x}{\sin x} dx \\ &= \int \frac{du}{u} \qquad u = \sin x \\ &= \ln|u| + C = \ln|\sin x| + C \\ &= -\ln|\csc x| + C.\end{aligned}$$

Example 12.18 *Find the area of the region bounded by the curve* $y = \dfrac{x}{1+x^2}$*, the* x *axis and the line* $x = 4$ *(see Figure 12.9).*

Solution: The area is equal to

$$\begin{aligned}\int_0^4 \frac{x}{1+x^2} dx &= \frac{1}{2} \int_0^4 \frac{2x}{1+x^2} dx = \frac{1}{2} \ln(1+x^2)|_1^4 \\ &= \frac{1}{2}(\ln 17 - \ln 1) = \frac{1}{2} \ln 17 = 1.4166. \qquad \square\end{aligned}$$

Example 12.19 *In a similar manner (as in the previous example) we can calculate many complicated integrals. For example:*

a.

$$\int \frac{1}{x\ln x}dx = \int \left(\frac{1}{\ln x}\right)\left(\frac{1}{x}\right) dx = \ln|\ln x| + C. \qquad \square$$

b.

$$\begin{aligned}\int \frac{x^2+x+2}{x^2+2}dx &= \int \left(\frac{x^2+2}{x^2+2} + \frac{x}{x^2+2}\right) dx \\ &= \int \left(\frac{x^2+2}{x^2+2}\right) dx + \int \left(\frac{x}{x^2+2}\right) dx \\ &= \int (1)\, dx + \frac{1}{2}\int \left(\frac{2x}{x^2+2}\right) dx \\ &= x + \frac{1}{2}\ln(x^2+2) + C.\end{aligned}$$

13

The exponential function

An *exponential function with base* a is a function of the form $f(x) = a^x$, where $a > 0$ and $a \neq 1$. Some examples are 3^x, $(1.4)^x$, and 10^x . We exclude the case $a = 1$ because $f(x) = 1^x$ is a constant function. All exponential functions have domain $(-\infty, \infty)$ and range $(0, \infty)$. So an exponential function never assumes the value 0. Note that an exponential function has a constant base and variable exponent. Thus, functions such as $f(x) = x^2$ and $f(x) = x^\pi$ would not be classified as exponential functions, since they have a variable base and a constant exponent. The present chapter is devoted to the study of exponential functions defined in a modern, calculus like way via natural exponential function.

13.1 The natural exponential function

The function $f(x) = \ln(x)$ is increasing on its entire domain, and therefore it has an inverse function f^{-1}. The domain of f^{-1} is the set of all reals, and the range is the set of positive reals, as shown in Figure 13.1. So, for any real number x

$$f(f^{-1}(x)) = \ln(f^{-1}(x)) = x.$$

Definition 13.1 (of the natural exponential function) *The inverse function of the natural logarithmic function* $f(x) = \ln x$ *is called the* natural exponential function *and is denoted by*

$$f^{-1}(x) = e^x.$$

That is,

$$y = e^x \quad \textit{if and only if} \quad x = \ln y.$$

The natural exponential function is the most important function in the whole calculus. We will meet it many times later and we must study its properties very carefully. First notice, that the familiar rules for operating with rational exponents can be extended to the natural exponential function, as shown in the following theorem.

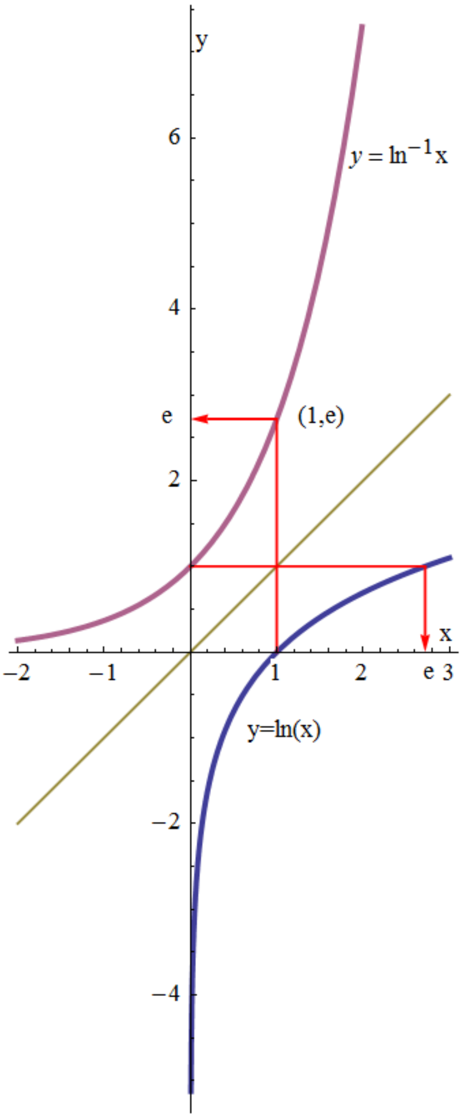


Fig. 13.1. The graphs of $y = \ln x$ and $y = \ln^{-1} x = e^x$. The inverse function of the natural logarithmic function is the natural exponential function.

Theorem 13.2 (Operations with the natural exponential functions) *Let a and b be any real numbers. Then*

1. $e^a e^b = e^{a+b}$,

2. $\dfrac{e^a}{e^b} = e^{a-b}$.

Proof. To prove property 1, you can write

$$\begin{aligned} \ln\left(e^a e^b\right) &= \ln\left(e^a\right) + \ln\left(e^b\right) \\ &= a + b \\ &= \ln\left(e^{a+b}\right). \end{aligned}$$

Because the natural logarithmic function is one-to-one, you can conclude that

$$e^a e^b = e^{a+b}.$$

The proof of the other property is similar. ■

As we know, an inverse function f^{-1} shares many properties with f. So, the natural exponential function inherits the following properties from the natural logarithmic function.

Properties of the natural exponential function:

1. The domain of $f(x) = e^x$ is $(-\infty, \infty)$, and the range is $(0, \infty)$.

2. The function $f(x) = e^x$ is continuous, increasing, and one-to-one on its entire domain.

3. The graph of $f(x) = e^x$ is concave upward on its entire domain.

4. $\lim_{x\to-\infty} e^x = 0$ and $\lim_{x\to\infty} e^x = \infty$.

13.2 Derivatives of the natural exponential function

One of the most intriguing (and useful) characteristics of the natural exponential function is that it is its *own derivative*. In other words, it is a solution to the differential equation $y' = y$. This result is stated in the next theorem.

Theorem 13.3 (Derivatives of the natural exponential function) *Let u be a differentiable function of x. Then*

$$\boxed{\begin{array}{ll} \textbf{1.} & \dfrac{d}{dx}\left(e^{x}\right)=e^{x}, \\ \textbf{2.} & \dfrac{d}{dx}\left(e^{u}\right)=e^{u}\dfrac{du}{dx}. \end{array}} \tag{13.1}$$

Proof. To prove Property 1, use the fact that $\ln e^{x}=x$, and differentiate each side of the equation. So, from definition of the exponential function

$$\ln e^{x}=x.$$

Now differentiate each side with respect to x

$$\frac{d}{dx}\left[\ln e^{x}\right]=\frac{d}{dx}[x]$$

obtaining (by chain rule) the equality

$$\frac{1}{e^{x}}\frac{d}{dx}\left[e^{x}\right]=1,$$

which (after rearrangement) gives the final relation

$$\frac{d}{dx}\left[e^{x}\right]=e^{x}.$$

■

You can interpret this theorem geometrically by saying that the slope of the graph of $f(x)=e^{x}$ at any point (x,e^{x}) is equal to the y-coordinate of the point.

Example 13.4 (Differentiating exponential functions)

a. $\dfrac{d}{dx}\left[e^{2x^2+1}\right]=e^{2x^2+1}\dfrac{du}{dx}=4xe^{2x^2+1}$ $\qquad u=2x^2+1$

b. $\dfrac{d}{dx}\left[e^{-4/x}\right]=e^{-4/x}\dfrac{du}{dx}=\dfrac{4}{x^2}e^{-\frac{4}{x}}$ $\qquad u=4/x$

c. $\dfrac{d}{dx}\left[e^{x}\ln x\right]=\dfrac{e^{x}}{x}+e^{x}\ln x=e^{x}\left(\dfrac{1}{x}+\ln x\right)$ $\qquad$ *by the product rule* □

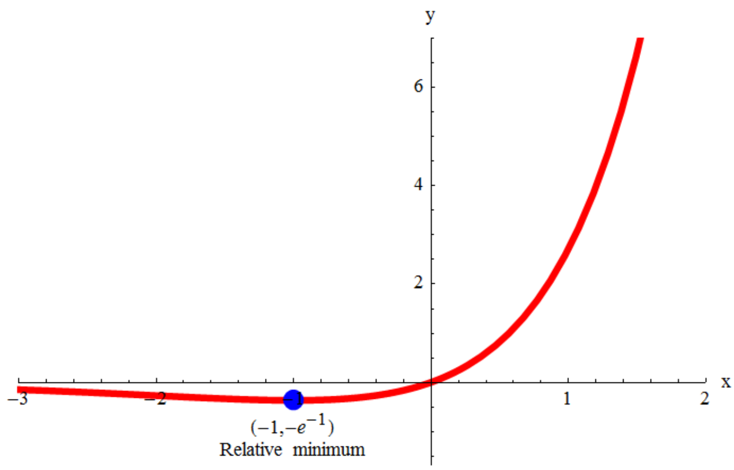


Fig. 13.2. Unique relative minimum of $f(x) = xe^x$.

Example 13.5 (Locating relative extrema) *Find the relative extrema of* $f(x) = xe^x$.

Solution: The derivative of is given by $f'(x) = e^x + xe^x = e^x(1+x)$. Because e^x is never 0, the derivative is 0 only when $x = 1$. Moreover, by the first derivative test, you can determine that this corresponds to a relative minimum, as shown in Figure 13.2. Because the derivative $f'(x) = e^x(1+x)$ is defined for all there are no other critical points. □

13.3 Integrals of exponential functions

Each differentiation formula in Theorem 13.3 has a corresponding integration formula.

Theorem 13.6 (Integration rules for exponential functions) *Let* u *be a*
differentiable function of x*. Then*

$$\begin{array}{|ll|} \textbf{1.} & \int e^x dx = e^x + C, \\ \textbf{2.} & \int e^u du = e^u + C. \end{array} \tag{13.2}$$

Example 13.7 *Find* $\int xe^{-x^2} dx$.

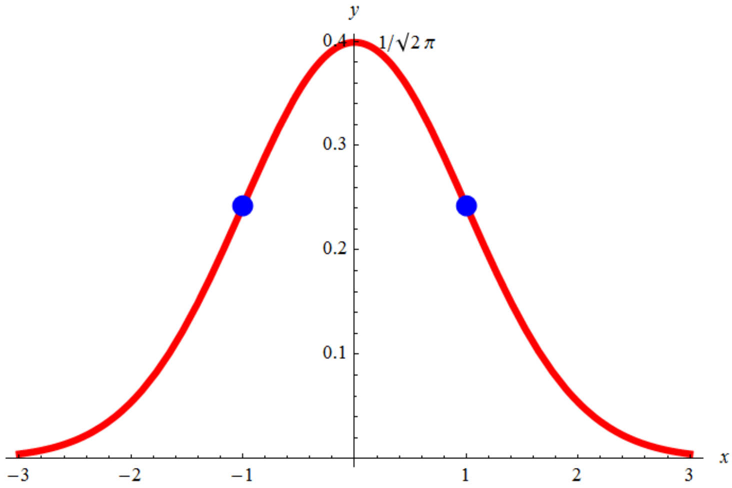


Fig. 13.3. The graph of the standard normal probability density function with two points of inflection.

Solution:

$$\begin{aligned}\int xe^{-x^2}dx &= \int e^{-x^2}(xdx)\\ &= \int e^u\left(-\frac{du}{2}\right) \qquad \text{where} \quad u=-x^2\\ &= -\frac{1}{2}\int e^u du\\ &= -\frac{1}{2}e^u + C\\ &= -\frac{1}{2}e^{-x^2}+C. \qquad \square\end{aligned}$$

Example 13.8 *It is interesting, that the innocent looking function*

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

called standard normal probability density function *does not have an anti-derivative function expressed in terms of elementary functions.Let us show, that this important function (see Figure 13.3) has points of inflection where* $x = \pm 1$*. To locate possible points of inflection, we will find the x-values for which the second derivative is* 0*.*

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

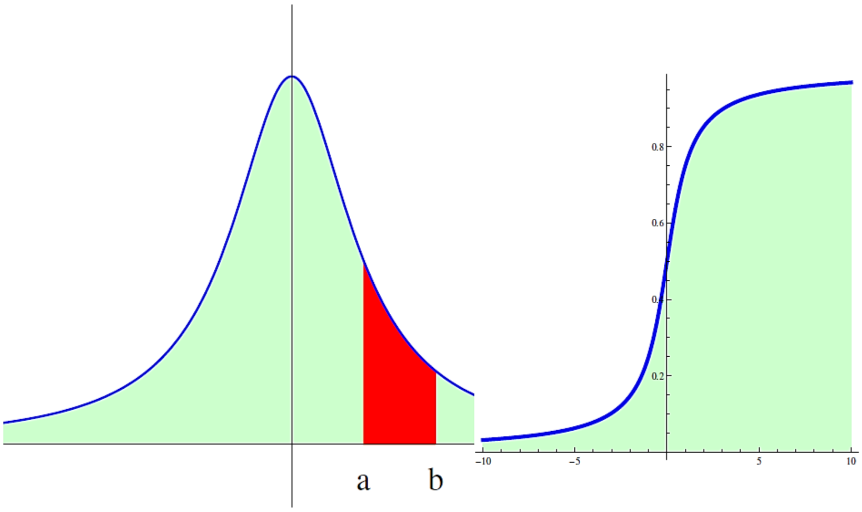


Fig. 13.4. Standard normal probability density function versus standard normal cumulative distribution function

$$f'(x) = -\frac{1}{2}\frac{\sqrt{2}}{\sqrt{\pi}}xe^{-\frac{1}{2}x^2}$$

$$\begin{aligned} f''(x) &= \frac{1}{2}\frac{\sqrt{2}}{\sqrt{\pi}}x^2e^{-\frac{1}{2}x^2} - \frac{1}{2}\frac{\sqrt{2}}{\sqrt{\pi}}e^{-\frac{1}{2}x^2} \\ &= \frac{1}{2}\frac{\sqrt{2}}{\sqrt{\pi}}e^{-\frac{1}{2}x^2}\left(x^2-1\right) \end{aligned}$$

So, $f''(x) = 0$ when $x = \pm 1$, and you can apply the standard techniques of Chapter 7 to conclude that these values yield the two points of inflection shown in Figure 13.3. It is interesting, that area under that bell-shaped curve (bounded by x-axis from below) can be found (using multivariable calculus), is finite and equal to 1. We call the antiderivative of f the standard normal cumulative distribution function $F(x)$.

Example 13.9 *Find $\int 4xe^{2x^2+1}dx$.*

Solution:

$$\begin{aligned} \int 4xe^{2x^2+1}dx &= \int e^u du \qquad \text{where} \quad u = 2x^2+1 \\ &= e^u + C \\ &= e^{2x^2+1} + C. \end{aligned}$$

Compare the result with Example 13.4a. □

Example 13.10 *Find* $\int_0^1 \frac{e^x}{1+e^x}dx$. *(This is an area problem because the integrated function is nonnegative.)*

Solution: Substituting $u = 1 + e^x$ we get $\int_0^1 \frac{e^x}{1+e^x}dx = \ln(1+e^x)\,|_0^1 = \ln(e+1) - \ln 2 \approx 0.620\,11$. $\square$

Example 13.11 *Sketch the graph of* $f(x) = x^2 e^{-x}$.

Solution: First, let us consider the first derivative $f'(x)$

$$f'(x) = 2xe^{-x} + x^2e^{-x}(-1) = 2xe^{-x} - x^2e^{-x} = -e^{-x}x(x-2).$$

This gives us two critical numbers: 0 and 2. At the point 0 this derivative changes sign (from negative to positive) so, we have local minimum $f(0) = 0$. Similarly we can conclude that the function has a local maximum at $x = 2$ where $f(2) = 4/e^2$, as $f'(x)$ changes sign from positive to negative there.

Examining the second derivative

$$f''(x) = 2e^{-x} - 4xe^{-x} + x^2e^{-x} = e^{-x}\left(x^2 - 4x + 2\right)$$

we deduce, that the graph of f has two inflection points where

$$x^2 - 4x + 2 = 0$$

i.e. at $x = 2 - \sqrt{2}$ and $x = 2 + \sqrt{2}$. Additionally, f is concave up on intervals $(-\infty, 2-\sqrt{2})$ and $(2+\sqrt{2}, \infty)$, and concave down on $(2-\sqrt{2}, 2+\sqrt{2})$. Finally, function f has a horizontal asymptote as

$$\lim_{x\to\infty} x^2e^{-x} = 0.$$

See Figure 13.5 $\square$.

13.4 One important limit

The base e may seem strange at first. Bit, it comes up everywhere. Later you will learn to appreciate just how *natural* it is.

Example 13.12 *Use logs to evaluate* $\lim_{x\to\infty}(1+\frac{1}{x})^x$.

Solution: Because the exponent x changes, it is better to find the limit of the logarithm

$$\lim_{x\to\infty} \ln\left[\left(1+\frac{1}{x}\right)^x\right].$$

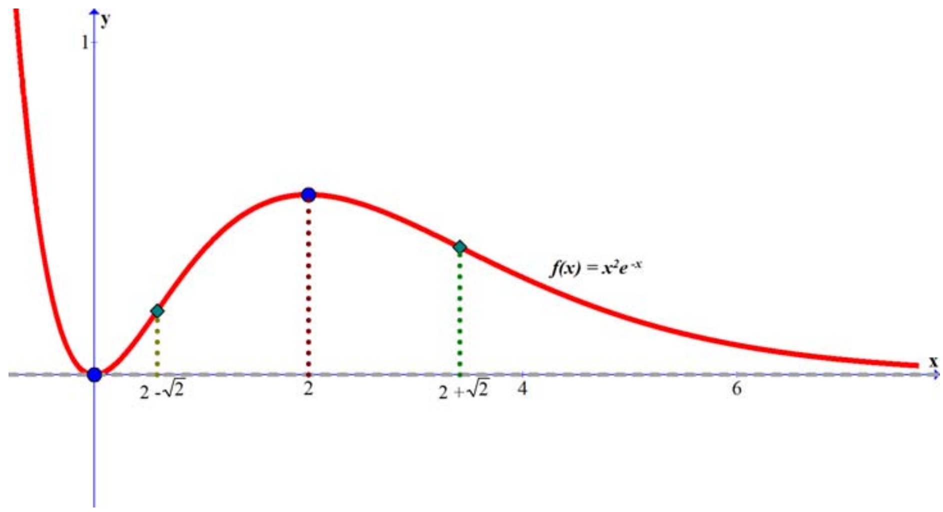


Fig. 13.5. The graph of the function $f(x) = x^2e^{-x}$.

We know that

$$\ln\left[\left(1+\frac{1}{x}\right)^x\right] = x\ln\left(1+\frac{1}{x}\right).$$

This expression has two competing parts, which balance: $x \to \infty$ while $\ln(1+\frac{1}{x}) \to 0$. But

$$\ln\left[\left(1+\frac{1}{x}\right)^x\right] = x\ln\left(1+\frac{1}{x}\right) = \frac{\ln\left(1+\frac{1}{x}\right)}{\frac{1}{x}} = \frac{\ln(1+h)}{h} \quad \text{with} \quad h = \frac{1}{x}.$$

Next, because $\ln 1 = 0$

$$\ln\left[\left(1+\frac{1}{x}\right)^x\right] = \frac{\ln(1+h) - \ln 1}{h}.$$

Take the limit: $h = \frac{1}{x} \to 0$ as $x \to \infty$, so that

$$\lim_{h\to\infty} \frac{\ln(1+h)-\ln 1}{h} = \frac{d}{dx}\ln(x)\mid_{x=1} = 1.$$

In all

$$\lim_{x\to\infty} \ln\left[\left(1+\frac{1}{x}\right)^x\right] = 1.$$

We have just found, that $a_x = \ln\left[\left(1+\frac{1}{x}\right)^x\right] \to 1$ as $x \to \infty$. If $b_x = \left(1+\frac{1}{x}\right)^x$ then $b_x = e^{a_x} \to e^1$ as $x \to \infty$. In other words, we have evaluated the limit we wanted:

$$\lim_{x\to\infty}(1+\frac{1}{x})^x = e. \tag{13.3}$$

□

Remark 13.13 *We never figured out what the exact numerical value of* e *was. Formula tells us that, for large values of* n*, the expression*

$$\left(1+\frac{1}{n}\right)^n$$

gives a good approximation to the value of e*. Use your calculator or computer to check that the following calculations are correct*

$$\begin{array}{lll} n=1 & \left(1+\frac{1}{n}\right)^n = & 2 \\ n=10 & \left(1+\frac{1}{n}\right)^n = & 2.59374246013 \\ n=50 & \left(1+\frac{1}{n}\right)^n = & 2.69158802907 \\ n=100 & \left(1+\frac{1}{n}\right)^n = & 2.70481382942 \\ n=1000 & \left(1+\frac{1}{n}\right)^n = & 2.71692393224 \\ n=10000000 & \left(1+\frac{1}{n}\right)^n = & 2.71828169254 \end{array}.$$

With the use of a sufficiently large value of n*, together with estimates for the error term*

$$\left|e-\left(1+\frac{1}{n}\right)^n\right|$$

it can be determined that

$$e = 2.71828182846$$

to eleven place decimal accuracy. Like the number π*, the number* e *is an irrational number.* □

Logs are used in all sciences and even if finance–look at the example below.

Example 13.14 (Euler's very special number) *Here is a function that you may have seen before if you studied interest on an investment that compounds more and more frequently during a year. Let's suppose that you invest* $100 *in an account that pays interest at an annual percentage rate of* 5% *and you leave the earned interest in the account to compound. If the interest is compounded* n *times per year, the amount of principal in the account at the end of one year will be* $100(1+.05/n)^n$*.*
Here are some examples of the amount in the account (in dollars) at the end of one year if the interest is compounded once, twice,..., up to 12 *times a year*

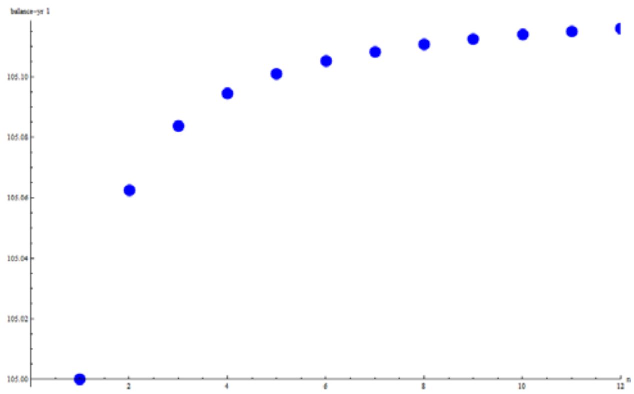


Fig. 13.6. The amount in the account (in dollars) at the end of one year if the interest is compounded n times a year.

(monthly).

$n =$ *interest pd/yr*	*account balance*
1	105.
2	105.063
3	105.084
4	105.095
5	105.101
6	105.105
7	105.108
8	105.111
9	105.113
10	105.114
11	105.115
12	105.116

Look at the Figure 13.6. What do you think happens to the function below, as n gets larger and larger, in other words, as you approach a situation where the interest compounds continuously at every instant in time during the year? To find out, let's evaluate the limit of the principal function as n goes to infinity.

$$\lim_{n\to\infty} 100\left(1+\frac{5}{100n}\right)^n = 100e^{\frac{1}{20}} \approx 105.13$$

Are you surprised? Why is this happening? The base is Euler's number e. Why is this happening? □

13.5 Arbitrary powers. The function $f(x) = x^r$

The elementary notion of exponent applies only to rational numbers. Expressions such as

$$10^4, \quad 2^{2/3}, \quad 6^{-1/5}, \quad \pi^{1/2,}$$

make sense, but so far we have attached no meaning to expressions such as

$$10^e, \quad 2^\pi, \quad 6^{-\sqrt{2}}, \quad \pi^e.$$

The extension of our sense of exponent to allow for irrational exponents is conveniently done by making use of the logarithm function and the exponential function. The heart of the matter is to observe that for $x > 0$ and p/q rational,

$$x^{p/q} = e^{(p/q)\ln x}.$$

We define x^z for irrational z by setting

$$x^z = e^{z \ln x}.$$

We can now state that

$$\boxed{\text{if } x > 0, \quad \text{then} \quad x^r = e^{r\ln x} \quad \text{for all real numbers } r.} \tag{13.4}$$

In particular

$$10^e = e^{e\ln 10}, \quad 2^\pi = e^{\pi \ln 2}, \quad 6^{-\sqrt{2}} = e^{-\sqrt{2}\ln 6}, \quad \pi^e = e^{e\ln\pi}.$$

With this extended sense of exponent, the usual laws of exponents still hold:

$$x^{r+s} = x^r x^s, \qquad x^{r-s} = \tfrac{x^r}{x^s}, \qquad (x^r)^s = x^{rs} \tag{13.5}$$

Proof.

$$x^{r+s} = e^{(r+s)\ln x} = e^{r\ln x} \cdot e^{s \ln x} = x^r x^s.$$

$$x^{r-s} = e^{(r-s)\ln x} = e^{r\ln x} \cdot e^{-s\ln x} = \frac{e^{r\ln x}}{e^{s\ln x}} = \frac{x^r}{x^s}.$$

$$(x^r)^s = e^{s \ln x^r} = e^{rs\ln x} = x^{rs}.$$

■

The differentiation of arbitrary powers follows the pattern established for rational powers; namely, for each real number r and each $x > 0$

$$\boxed{\frac{d}{dx}(x^r) = rx^{r-1}.} \tag{13.6}$$

Proof.

$$\frac{d}{dx}\left(x^r\right) = \frac{d}{dx}\left(e^{r\ln x}\right) = e^{r\ln x}\frac{d}{dx}\left(r\ln x\right) = x^r\frac{r}{x} = rx^{r-1}.$$

■

Thus

$$\frac{d}{dx}\left(x^{\sqrt{3}}\right) = \sqrt{3}x^{\sqrt{3}-1} \qquad \text{and} \qquad \frac{d}{dx}\left(x^{\pi}\right) = \pi x^{\pi-1}.$$

As usual, we differentiate compositions by the chain rule. Thus

$$\frac{d}{dx}\left(\left(x^3+2\right)^{\sqrt{2}}\right) = \sqrt{2}\left(x^3+2\right)^{\sqrt{2}-1}\frac{d}{dx}\left(x^3+2\right) = 3\sqrt{2}x^2\left(x^3+2\right)^{\sqrt{2}-1}$$

Logarithmic differentiation is a useful technique for dealing with derivatives of things like $f(x)^{g(x)}$, where both the base and the exponent are functions of x. After all, how would you find

$$\frac{d}{dx}\left(x^{\sin(x)}\right)$$

with what we have seen already? It doesn't fit any of the rules. Still, we have these nice log rules which cut exponents down to size. If we let $y = x^{\sin(x)}$, then

$$\ln(y) = \ln(x^{\sin(x)}) = \sin(x)\ln(x)$$

Now let's differentiate both sides (implicitly) with respect to x:

$$\frac{d}{dx}\left(\ln(y)\right) = \frac{d}{dx}\left(\sin(x)\ln(x)\right)$$

Let's look at the right-hand side first. This is just a function of x and requires the product rule; you should check that the derivative works out to be $\cos(x)\ln(x) + \sin(x)/x$. Now let's look at the left-hand side. To differentiate $\ln(y)$ with respect to x (not y!), we should use the chain rule. Set $u = \ln(y)$, so that $du/dy = 1/y$. We need to find du/dx; by the chain rule,

$$\frac{du}{dx} = \frac{du}{dy}\frac{dy}{dx} = \frac{1}{y}\frac{dy}{dx}.$$

So, implicitly differentiating the equation $\ln(y) = \sin(x)\ln(x)$ produces

$$\frac{1}{y}\frac{dy}{dx} = \cos(x)\ln(x) + \frac{\sin(x)}{x}.$$

Now we just have to multiply both sides by y and then replace y by $x^{\sin(x)}$:

$$\frac{dy}{dx} = \left(\cos(x)\ln(x) + \frac{\sin(x)}{x}\right)y = \left(\cos(x)\ln(x) + \frac{\sin(x)}{x}\right)x^{\sin(x)}.$$

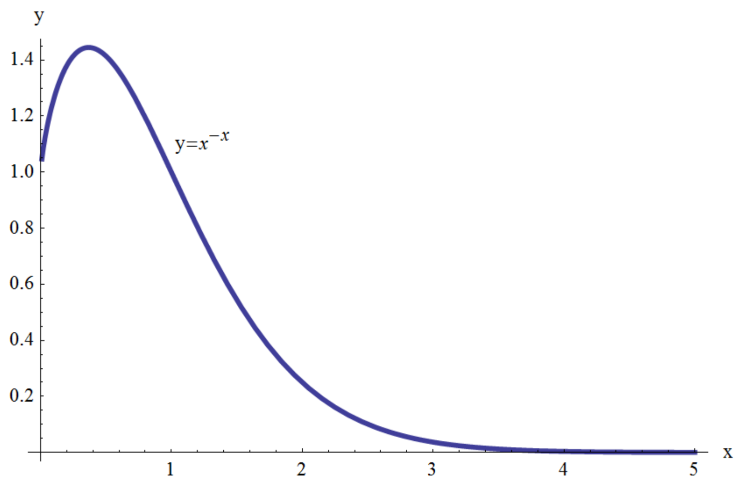


Fig. 13.7. The graph of $f(x) = x^{-x}$

That's the answer we're looking for. (By the way, there is another way we could have done this problem. Instead of using the variable y, we could just have used our formula $A = e^{\ln(A)}$ to write

$$x^{\sin(x)} = e^{\ln(x^{\sin(x)})} = e^{\sin(x)\ln(x)}$$

Now we can differentiate the right-hand side of this with respect to x by using the product and chain rules. When you've finished, you should replace $e^{\sin(x)\ln(x)}$ by $x^{\sin(x)}$ and check that you get the same answer as the original one above.)

Example 13.15 *Find the derivative of*

$$f(x) = x^{-x} \quad at \quad x > 0. \tag{13.7}$$

One way to find this derivative is to observe that

$$x^{-x} = e^{-x\ln x}$$

and then differentiate:

$$\frac{d}{dx}\left(x^{-x}\right) = \frac{d}{dx}\left(e^{-x\ln x}\right) = e^{-x\ln x}\frac{d}{dx}(-x\ln x) = -\frac{1}{x^x}(\ln x + 1)\,.$$

Another way to find this derivative is to take the natural logarithm of both sides in

$$y = x^{-x}$$

obtaining

$$\ln x = \ln x^{-x} = -x \ln x.$$

We then take derivatives of both sides with respect to x:

$$\frac{1}{y}\frac{dy}{dx} = -((1)\ln x + x(1/x)) = -(\ln x + 1).$$

Next we solve for dy/dx:

$$\frac{dy}{dx} = -(\ln x + 1)\, y.$$

Finally, we substitute for y from the original equation:

$$\frac{dy}{dx} = -(\ln x + 1)\, x^{-x}. \qquad \square$$

Each derivative formula gives rise to a companion integral formula. The integral version of (13.6) takes the form

$$\boxed{\int x^r dx = \frac{x^{r+1}}{r+1} + C, \qquad \text{for} \quad r \neq -1.} \tag{13.8}$$

Note the exclusion of $r = -1$. What is the integral if $r = -1$?

Example 13.16 *Find* $\int \frac{x^2}{(2x^3+1)^\pi} dx$.

Solution: Set

$$u = 2x^3 + 1, \qquad du = 6x^2 dx.$$

Then

$$\begin{aligned}
\int \frac{x^2}{(2x^3+1)^\pi} dx &= \frac{1}{6}\int u^{-\pi} du = \frac{1}{6}\left(\frac{u^{1-\pi}}{1-\pi}\right) + C \\
&= \frac{1}{6}\left(\frac{(2x^3+1)^{1-\pi}}{1-\pi}\right) + C \\
&= -\frac{1}{3\pi - 3}\frac{x^3 + \frac{1}{2}}{(2x^3+1)^\pi} + C. \qquad \square
\end{aligned}$$

13.6 Bases other than e

The **base** of the natural exponential function is e.This "natural" base can be used to assign a meaning to a general base a.

Definition 13.17 (Definition of exponential function to base a) *If a is a positive real number ($a \neq 1$) and x is any real number, then the* exponential function to the base *a is denoted by a^x and is defined by*

$$a^x = e^{(\ln a)x}$$

If $a = 1$, then $y = 1^x = 1$ is a constant function.

These functions obey the usual laws of exponents. For instance, here are some familiar properties.

1. $a^0 = 1$
2. $a^x a^y = a^{x+y}$
3. $\dfrac{a^x}{a^y} = a^{x-y}$
4. $(a^x)^y = a^{xy}$

The proof is left to the reader as an exercise.

When modeling the half-life of a radioactive sample, it is convenient to use as the base of the exponential model. (Half-life is the number of years required for half of the atoms in a sample of radioactive material to decay.)

Example 13.18 *The half-life of carbon-14 is about 5715 years. A sample contains 1 gram of carbon-14. How much will be present in 10,000 years?*

Solution: Let $t = 0$ represent the present time and let y represent the amount (in grams) of carbon-14 in the sample. Using a base of $\frac{1}{2}$, you can model by the equation

$$y = \left(\frac{1}{2}\right)^{t/5715}.$$

Notice that when $t = 5715$ the amount is reduced to half of the original amount.

$$y = \left(\frac{1}{2}\right)^{5715/5715} = \frac{1}{2} \quad \text{gram.}$$

When $t = 11,430$ the amount is reduced to a quarter of the original amount, and so on. To find the amount of carbon-14 after 10,000 years, substitute 10,000 for t

$$y = \left(\frac{1}{2}\right)^{10000/5715} \approx 0.30 \quad \text{gram.} \qquad \square$$

Logarithmic functions to bases other than can be defined in much the same way as exponential functions to other bases are defined.

Definition 13.19 (Definition of logarithmic function to base a) *If a is a positive real number ($a \neq 1$) and x is any positive real number, then the logarithmic function to the base a is denoted by $\log_a x$ and is defined as*

$$\boxed{\log_a x = \frac{1}{\ln a} \ln x.} \tag{13.9}$$

Remark 13.20 *In precalculus, you learned that $\log_a x$ is the value to which a must be raised to produce x.This agrees with the definition given here because*

$$a^{\log_a x} = a^{(1/\ln a)\ln x} = e^{(1/\ln a)\ln a \ln x} = e^{\ln x} = x.$$

Logarithmic functions to the base a have properties similar to those of the natural logarithmic function given in Theorem 12.2 (Assume x and y are positive numbers and n is rational.)

1. $\log_a 1 = 0$
2. $\log_a xy = \log_a x + \log_a y$
3. $\log_a x^n = n \log_a x$
4. $\log_a \dfrac{x}{y} = \log_a x - \log_a y$

From the definitions of the exponential and logarithmic functions to the base a, it follows that $f(x) = a^x$ and $g(x) = \log_a x$ are inverse functions of each other (see Figure 13.8, where $a = 2$.)

Theorem 13.21 (Properties of inverse functions)

1. $y = a^x$ *if and only if* $x = \log_a$

2. $a^{\log_a x} = x$, *for* $x > 0$

3. $\log_a a^x = x$, *for all* x

The logarithmic function to the base 10 is called the *common logarithmic function.* So, for common logarithms, $y = 10^x$ if and only if $x = \log_{10} y$.

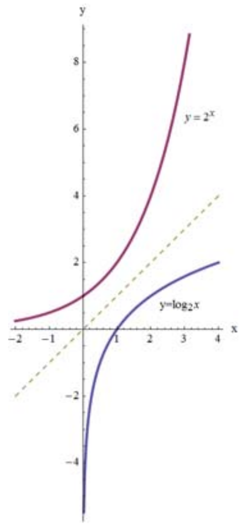


Fig. 13.8. The graph of 2^x and its inverse, $\log_2 x$.

13.7 Differentiation and Integration

To differentiate exponential and logarithmic functions to other bases, you have three options:

1. use the definitions of a^x and $\log_a x$ and differentiate using the rules for the natural exponential and logarithmic functions,
2. use logarithmic differentiation, or
3. use the following differentiation rules for bases other than e.

Theorem 13.22 (Derivatives for bases other than e) *Let a be a positive real number $(a \neq 1)$ and let u be a differentiable function of x. Then*

$$\textbf{1.}\ \frac{d}{dx}[a^x] = (\ln a)\, a^x \qquad \textbf{2.}\ \frac{d}{dx}[u^x] = (\ln u)\, u^x \frac{du}{dx}$$
$$\textbf{3.}\ \frac{d}{dx}[\log_a x] = \frac{1}{(\ln a)\, x} \qquad \textbf{4.}\ \frac{d}{dx}[\log_a u] = \frac{1}{(\ln a)\, u}\frac{du}{dx} \tag{13.10}$$

Proof. By definition, $a^x = e^{x(\ln a)}$. So, you can prove the first rule by letting $u = x(\ln a)$ and differentiating with base to e obtain

$$\frac{d}{dx}[a^x] = \frac{d}{dx}\left[e^{x(\ln a)}\right] = e^{x(\ln a)}(\ln a) = (\ln a)\, a^x.$$

To prove the third rule, you can write

$$\frac{d}{dx}\left[\log_a x\right] = \frac{d}{dx}\left[\frac{\ln x}{\ln a}\right] = \frac{1}{\ln a}\left(\frac{1}{x}\right) = \frac{1}{(\ln a)\, x}.$$

The second and fourth rules are simply the chain rule versions of the first and third rules. ■

From the theorem 13.22 it follows why the function $\ln x$ is called "natural logarithm". It comes out in the formulas 13.10 even if you are currently using bases different than e.

Example 13.23 *Calculate*

$$\textbf{a.}\ \frac{d}{dx}(\log_5 |x|) \qquad \textbf{b.}\ \frac{d}{dx}\left(\log_2\left(3x^2+1\right)\right)$$

Solution:

a.

$$\frac{d}{dx}(\log_5 |x|) = \frac{d}{dx}\left[\frac{\ln |x|}{\ln 5}\right] = \frac{1}{5}\frac{d}{dx}\left[\ln |x|\right] = \frac{1}{5}\left(\frac{1}{x}\right) = \frac{1}{5x}.$$

b.

$$\begin{aligned}\frac{d}{dx}\left(\log_2\left(3x^2+1\right)\right) &= \frac{d}{dx}\left[\frac{\ln\left(3x^2+1\right)}{\ln 2}\right] = \frac{1}{\ln 2}\frac{1}{(3x^2+1)}\frac{d}{dx}\left(3x^2+1\right)\\ &= \frac{1}{\ln 2}\frac{1}{(3x^2+1)}(6x) = \frac{6x}{\ln 2\,(3x^2+1)} \qquad \square\end{aligned}$$

Occasionally, an integrand involves an exponential function to a base other than e. When this occurs, there are two options:

1. convert to base e using the formula $a^x = e^{x(\ln a)}$ and then integrate, or
2. integrate directly, using the integration formula

$$\boxed{\int a^x dx = \left(\frac{1}{\ln a}\right) a^x + C}$$

(which follows from Theorem 13.22).

Example 13.24 *Find* $\int 3^x dx$.

Solution:

$$\int 3^x dx = \frac{3^x}{\ln 3} + C. \qquad \square$$

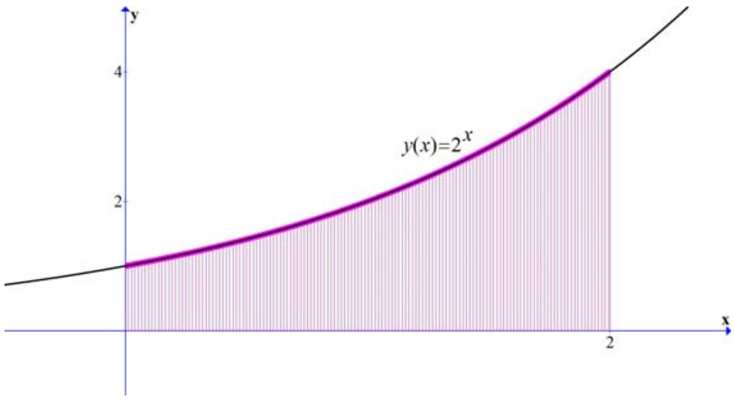


Fig. 13.9.

Example 13.25 *Find area under the curve $y = 2^x$ between $x = 0$ and $x = 2$ (see Fig. 13.9).*

Solution:

$$Area = \int_0^2 2^x dx = \frac{1}{\ln 2}\left(2^2 - 2^0\right) = \frac{3}{\ln 2} \approx: 4.328\,1. \quad \square$$

When the power rule $\frac{d}{dx}[x^n] = nx^{n-1}$, was introduced before, the exponent was required to be a rational number. Now the rule is extended to cover any real value of n. Try to prove this theorem using logarithmic differentiation.

Theorem 13.26 (The power rule for real exponents) *Let n be any real number and let u be a differentiable function of x*

$$\boxed{\begin{array}{ll} \textbf{1.} & \frac{d}{dx}[x^n] = nx^{n-1} \\ \textbf{2.} & \frac{d}{dx}[u^n] = nu^{n-1}\frac{du}{dx}. \end{array}} \tag{13.11}$$

The next example compares the derivatives of four types of functions. Each function uses a different differentiation formula, depending on whether the base and the exponent are constants or variables.

Example 13.27 (Comparing Variables and Constants)

a) $\frac{d}{dx}[e^e] = 0$, *Constant rule*

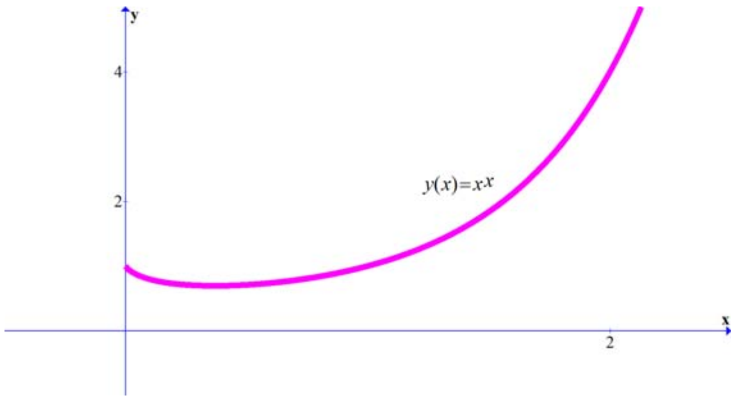


Fig. 13.10. The graph of the function x^x.

b) $\dfrac{d}{dx}\left[e^x\right] = e^x,$ *Exponential rule*

c) $\dfrac{d}{dx}\left[x^e\right] = ex^{e-1},$ *Power rule*

d) $\dfrac{d}{dx}\left[x^x\right] = x^x\left(\ln x + 1\right),$ *Logarithmic differentiation.*

14

Inverse functions and inverse trig functions

In everyday language the term “inversion” conveys the idea of a reversal. For example, in meteorology a temperature inversion is a reversal in the usual temperature properties of air layers, and in music a melodic inversion reverses an ascending interval to the corresponding descending interval. In mathematics the term inverse is used to describe functions that reverse one another in the sense that each *undoes* the effect of the other. In this section we discuss this fundamental mathematical idea. Next we will show how the derivative of a one-to-one function can be used to obtain the derivative of its inverse function. There is a remarkable special case of the chain rule. This will provide the tools we need to obtain derivative formulas for exponential functions from the derivative formulas for logarithmic functions and to obtain derivative formulas for inverse trigonometric functions from the derivative formulas for trigonometric functions.

14.1 Inverse functions

14.1.1 Introduction

The idea of solving an equation $y = f(x)$ for x as a function of y, say $x = g(y)$, is one of the most important ideas in mathematics. Sometimes, solving an equation is a simple process; for example, using basic algebra the equation

$$y = x^3 + 1 \qquad \boxed{y = f(x)} \tag{14.1}$$

can be solved for x as a function of y:

$$x = \sqrt[3]{y-1} \qquad \boxed{x = g(y)} \tag{14.2}$$

The first equation is better for computing y if x is known (Figure 14.1), and the second is better for computing x if y is known (Figure 14.2).Our primary interest in this section is to identify relationships that may exist between the functions f and g when an equation $y = f(x)$ is expressed as $x = g(y)$, or conversely. For example, consider the functions $f(x) = x^3 + 1$ and $g(y) =$

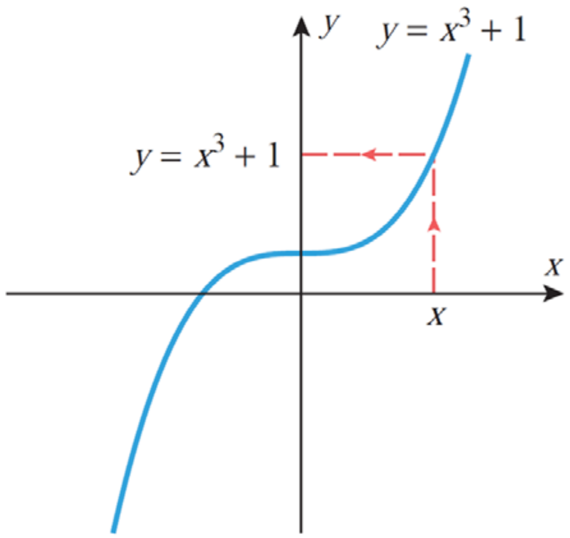


Fig. 14.1. Here is better for computing y if x is known.

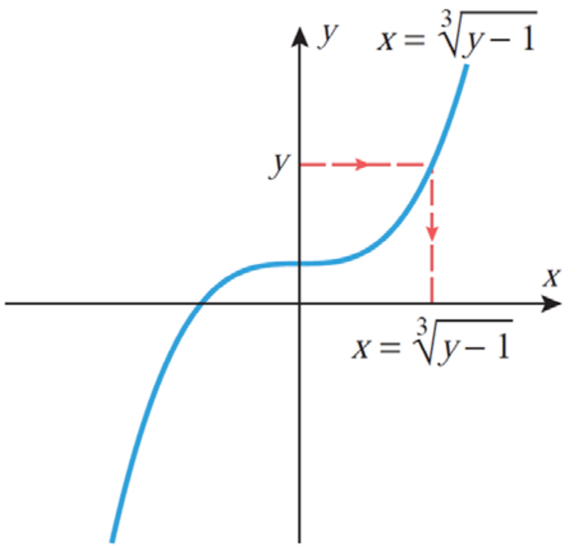


Fig. 14.2. Here is better for computing x if y is known.

$\sqrt[3]{y-1}$ discussed above. When these functions are composed in either order, they cancel out the effect of one another in the sense that

$$\begin{aligned} g(f(x)) &= \sqrt[3]{f(x)-1} = \sqrt[3]{(x^3+1)-1} = x \\ f(g(y)) &= [g(y)]^3+1 = \left[\sqrt[3]{y-1}\right]^3+1 = y. \end{aligned}$$

Pairs of functions with these two properties are so important that there is special terminology for them.

Definition 14.1 *If the functions f and g satisfy the two conditions*

$$\begin{aligned} &g(f(x)) = x \textit{ for every } x \textit{ in the domain of } f \\ &f(g(y)) = y \textit{ for every } y \textit{ in the domain of } g \end{aligned} \tag{14.3}$$

then we say that f is an inverse of g *and g* is an inverse of f *or that f and g are* inverse functions*.*

We will call these the *cancellation equations* for f and f^{-1}.

It can be shown (do it!) that if a function f has an inverse, then that inverse *is unique.* Thus, if a function f has an inverse, then we are entitled to talk about "the" inverse of f , in which case we denote it by the symbol f^{-1}.

Remark 14.2 *If f is a function, then the -1 in the symbol f^{-1} always denotes an inverse and never an exponent. That is,*

$$f^{-1}(x) \quad \textit{never means} \quad \frac{1}{f(x)}.$$

Example 14.3 *$y = g(x) = \frac{5}{9}(x-32)$ and $x = f(y) = \frac{9}{5}y + 32$ are inverse functions (for temperature). Here x is degrees Fahrenheit and y is degrees Celsius. From $x = 32$ (freezing in Fahrenheit) you find $y = 0$ (freezing in Celsius). The inverse function takes $y = 0$ back to $x = 32$.*

14.1.2 Changing the independent variable

The formulas in (14.3) use x as the independent variable for f and y as the independent variable for f^{-1}. Although it is often convenient to use different independent variables for f and f^{-1}, there will be occasions on which it is desirable to use the same independent variable for both. For example, if we want to graph the functions f and f^{-1} together in the same xy-coordinate system, then we would want to use x as the independent variable and y as the dependent variable for both functions. (We have just changed letters !) Thus, to graph the functions $f(x) = x^3 + 1$ and $f^{-1}(y) = \sqrt[3]{y-1}$ of Example from page 321 in the same xy-coordinate system, we would change the independent variable y to x, use y as the dependent variable for both functions, and graph the equations

$$y = x^3 + 1 \quad \text{and} \quad y = \sqrt[3]{x-1}.$$

We will talk more about graphs of inverse functions later in this chapter , but for reference we give the following reformulation of the cancellation equations in (14.3) using x as the independent variable for both f and f^{-1}

$$\boxed{\begin{array}{ll} f^{-1}(f(x)) = x & \text{for every } x \text{ in the domain of } f \\ f(f^{-1}(x)) = x & \text{for every } x \text{ in the domain of } f^{-1}. \end{array}} \tag{14.4}$$

The equations in (14.4) imply the following relationships between the domains and ranges of f and f^{-1} :

$$\boxed{\begin{array}{l} \text{domain of } f^{-1} = \text{range of } f \\ \text{range of } f^{-1} = \text{domain of } f. \end{array}} \tag{14.5}$$

Example 14.4 *Confirm that the inverse of $f(x) = x^3$ is $f^{-1}(x) = \sqrt[3]{x}$.*

Solution:

$$\begin{aligned} f^{-1}(f(x)) &= f^{-1}(x^3) = \sqrt[3]{x^3} = x \\ f(f^{-1}(x)) &= f(\sqrt[3]{x}) = \left(\sqrt[3]{x}\right)^3 = x. \end{aligned}$$

□

14.1.3 Existence of inverse functions

At the beginning of this section we observed that solving $y = f(x) = x^3 + 1$ for x as a function of y produces $x = f^{-1}(y) = \sqrt[3]{y-1}$. The following theorem shows that this is not accidental.

Theorem 14.5 *If an equation $y = f(x)$ can be solved for x as a function of y, say $x = g(y)$, then f has an inverse and that inverse is $g(y) = f^{-1}(y)$.*

Proof. Substituting $y = f(x)$ into $x = g(y)$ yields$x = g(f(x))$, which confirms the first equation in Definition 14.1, and substituting $x = g(y)$ into $y = f(x)$ yields $y = f(g(y))$, which confirms the second equation in Definition 14.1. ■

The procedure we used above for finding the inverse of a function f was based on solving the equation $y = f(x)$ for x as a function of y. This procedure can fail for two reasons—the function f may not have an inverse, or it may have an inverse but the equation $y = f(x)$ cannot be solved explicitly for x as a function of y. Thus, it is important to establish conditions that ensure the existence of an inverse, even if it cannot be found explicitly.

If a function f has an inverse, then it must assign distinct outputs to distinct inputs. For example, the function $f(x) = x^2$ cannot have an inverse because it assigns the same value to $x = 2$ and $x = -2$, namely, $f(2) = f(-2) = 4$. Thus, if $f(x) = x^2$ were to have an inverse, then the equation $f(2) = 4$ would imply that $f^{-1}(4) = 2$, and the equation $f(-2) = 4$ would imply that $f^{-1}(4) = -2$. But this is impossible because $f^{-1}(4)$ cannot have two different values . Another way to see that $f(x) = x^2$ has no inverse is to attempt to find the inverse by solving the equation $y = x^2$ for x as a function of y. We run into trouble immediately because the resulting equation $x = \pm\sqrt{y}$ does not express x as a single function of y.

A function that assigns distinct outputs to distinct inputs is said to be *one-to-one* or *invertible*, so we know from the preceding discussion that if a function f has an inverse, then it must be one-to-one. The converse is also true, thereby establishing the following theorem.

Theorem 14.6 *A function has an inverse if and only if it is one-to-one.*

Stated algebraically, a function f is one-to-one if and only if $f(x_1) = f(x_2)$ whenever $x_1 = x_2$; stated geometrically, a function f is one-to-one if and only if the graph of $y = f(x)$ is cut at most once by any horizontal line (Figure 14.4). The latter statement together with Theorem 14.6 provides the following geometric test for determining whether a function has an inverse.

Theorem 14.7 (The horizontal line test) *A function has an inverse function if and only if its graph is cut at most once by any horizontal line.*

Example 14.8 *Use the horizontal line test to show that $f(x) = x^2$ has no inverse but that $f(x) = x^3$ does.*

Solution: Figure 14.3 shows a horizontal line that cuts the graph of $y = x^2$ more than once, so $f(x) = x^2$ is not invertible. Figure 14.4 shows that the graph of $y = x^3$ is cut at most once by any horizontal line, so $f(x) = x^3$ is invertible. (Recall from Example 14.4 that the inverse of $f(x) = x^3$ is $f^{-1}(x) = x^{1/3}$).

14.1.4 Graphs of inverse functions

Our next objective is to explore the relationship between the graphs of f and f^{-1}. For this purpose, it will be desirable to use x as the independent variable for both functions so we can compare the graphs of $y = f(x)$ and $y = f^{-1}(x)$.

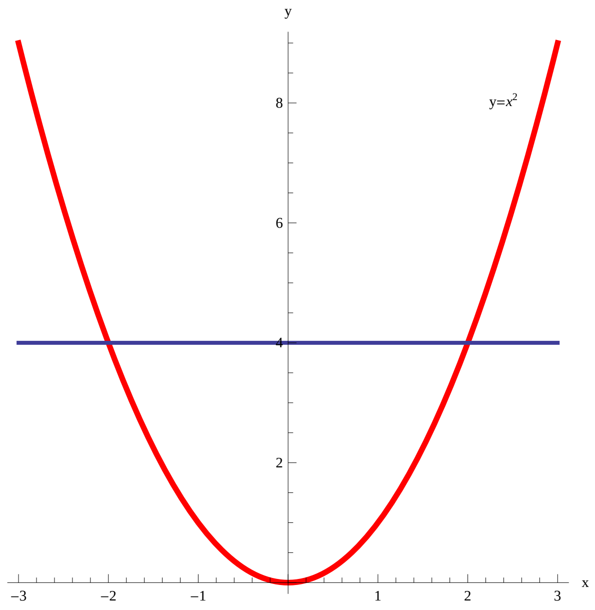


Fig. 14.3. Horizontal line test for $f(x) = x^2$.

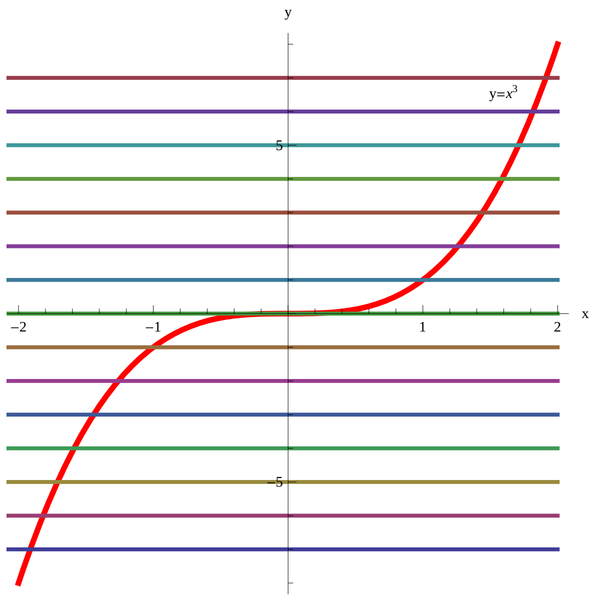


Fig. 14.4. Horizontal line test for $f(x) = x^3$.

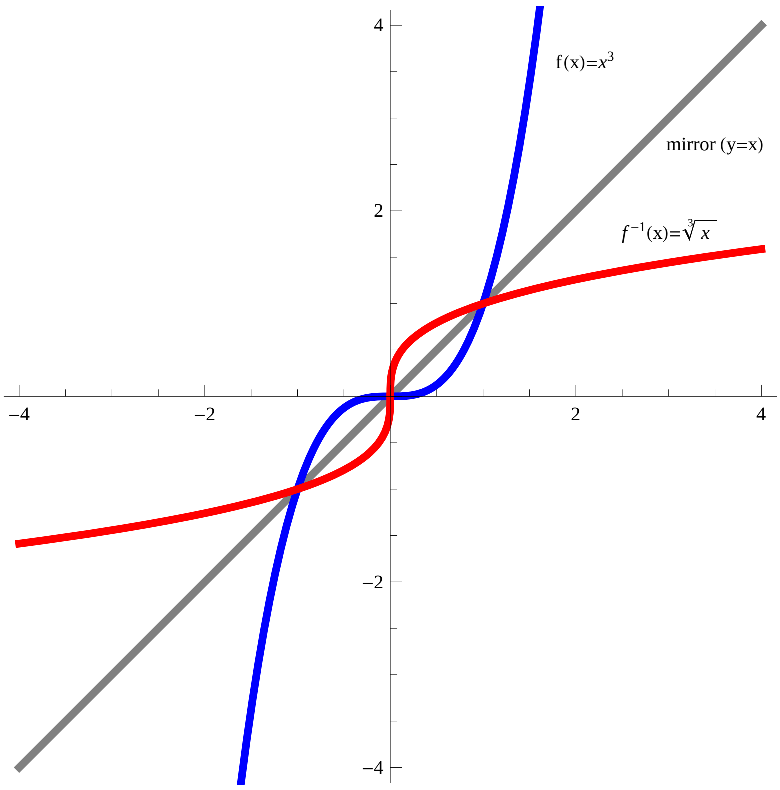


Fig. 14.5. Function $f(x) = x^3$ and its inverse.

If (a, b) is a point on the graph $y = f(x)$, then $b = f(a)$. This is equivalent to the statement that $a = f^{-1}(b)$, which means that (b, a) is a point on the graph of $y = f^{-1}(x)$. In short, reversing the coordinates of a point on the graph of f produces a point on the graph of f^{-1}. Similarly, reversing the coordinates of a point on the graph of f^{-1} produces a point on the graph of f (verify). However, the geometric effect of reversing the coordinates of a point is to reflect that point about the line $y = x$, and hence the graphs of $y = f(x)$ and $y = f^{-1}(x)$ are reflections of one another about this line (Figure 14.5). In summary, we have the following result.

Theorem 14.9 *If f has an inverse, then the graphs of $y = f(x)$ and $y = f^{-1}(x)$ are reflections of one another about the line $y = x$; that is, each graph is the mirror image of the other with respect to that line.*

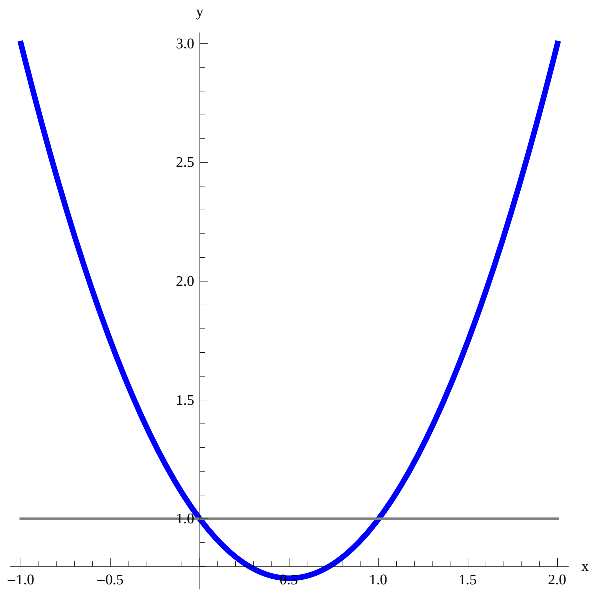


Fig. 14.6. Function $f(x) = x^2 - x + 1$ defined on $(-\infty, \infty)$ fails the horizontal line test.

Example 14.10 *Determine if the function $f(x) = x^2 - x + 1$ has an inverse. If it exists, then find the inverse.*

Solution: We note that $f(0) = f(1) = 1$. Thus, f is not one-to-one. We can also plot the graph (Figure 14.6) of f and note that it fails the horizontal line test.

However, observe that if we restrict the domain of f to an interval where f is either increasing or decreasing, say $[0.5, \infty)$, then

its inverse exists (see plot below–Figure 14.7).It can be easily checked, that

$$f^{-1}(x) = \frac{1}{2}(1 + \sqrt{-3 + 4x})$$

is having range $[0.5, \infty)$, which agrees with the restricted domain of f. Notice, that the domain of $f^{-1}(x)$ is equal to $[3/4, \infty)$, which agrees with the range of (restricted) f. Lastly (Figure 14.8), a plot of the graphs of $f(x)$ and $f^{-1}(x)$ (in blue and red respectively) shows their expected symmetry about the diagonal line $y = x$ (dotted line). The plot below (Figure 14.9) illustrates how the slopes of the two tangent lines, that of f at $(2, 3)$ and that of f^{-1} at $(3, 2)$ (both in green), are reciprocal. As you will see, this is not an accident. □

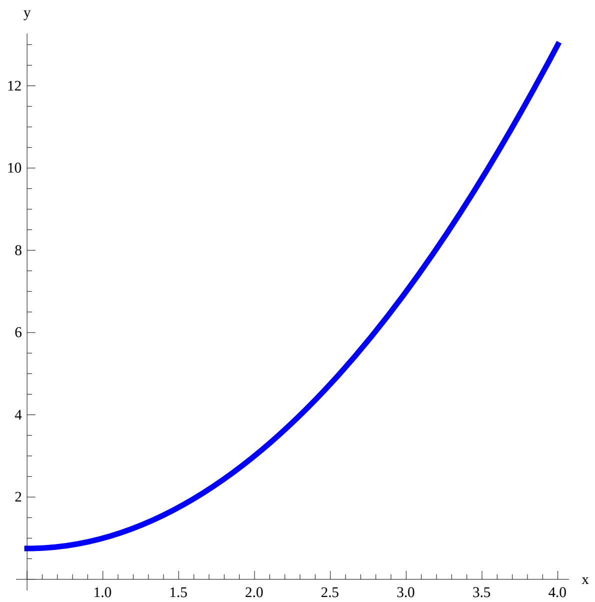


Fig. 14.7. Function f restricted to an interval $[0.5, \infty)$ where it is increasing.

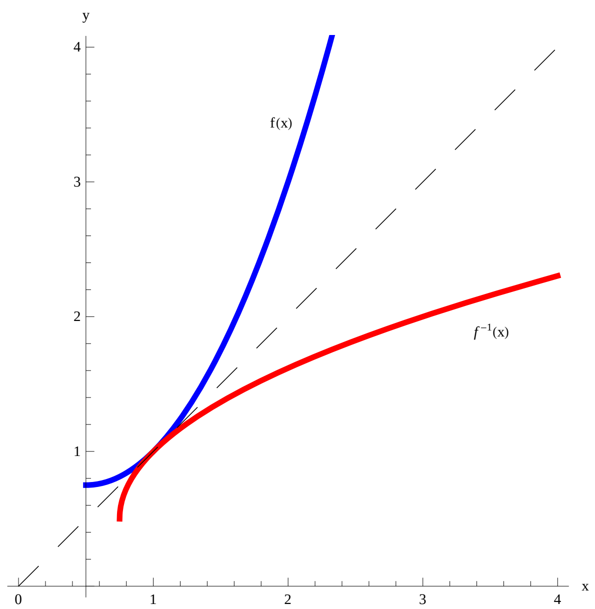


Fig. 14.8. Functions $f(x)$ and $f^{-1}(x)$.

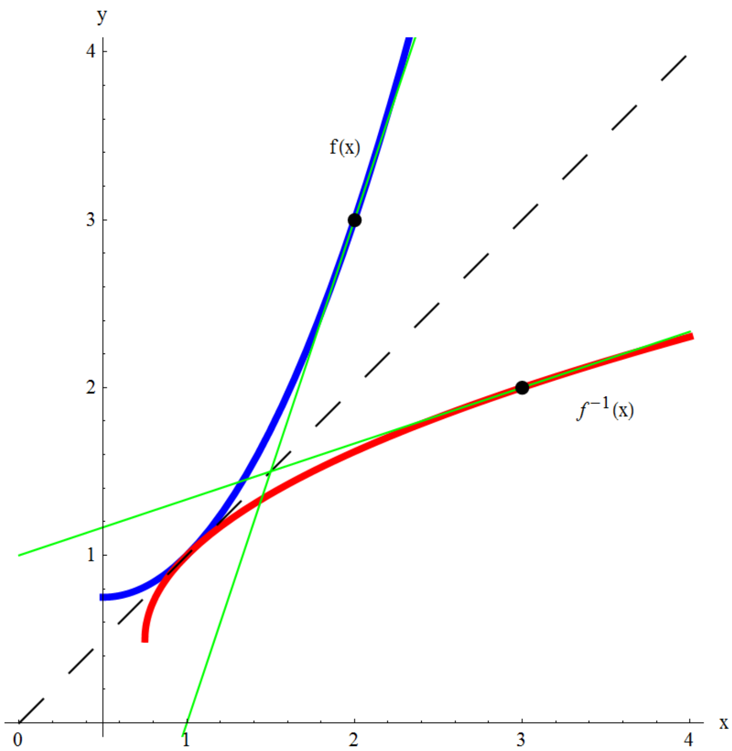


Fig. 14.9. Slopes of the two tangent lines, that of f at $(2, 3)$ and that of f^{-1} at $(3, 2)$ are reciprocal.

14.2 The derivative and inverse functions

In the previous sections we introduced the basics of inverse functions. Now we're going to explore interesting (and useful) connections between derivatives and inverse functions.

14.2.1 Using the derivative to show that an inverse exists

Suppose that you have a differentiable function f whose derivative is always positive. Then the function must be *increasing.* If our function f is always increasing, then it must satisfy the horizontal line test. No horizontal line could possibly hit the graph of $y = f(x)$ twice. Since the horizontal line test is satisfied by f, we know that f has an inverse. This has given us a nice strategy for showing that a function has an inverse: show that its derivative is *always positive* on its domain.

Example 14.11 *Suppose that*

$$f(x) = \frac{1}{3}x^3 - 2x^2 + 5x - 7.$$

on the domain $\mathbb{R}$ (the whole real line). Does f has an inverse? It would be a real mess to switch x and y in the equation $y = \frac{1}{3}x^3 - 2x^2 + 5x - 7$ and then try to solve for y. (Try it and see!) A much better way to show that f has an inverse is to find the derivative. We get

$$f'(x) = x^2 - 4x + 5 = (x-2)^2 + 1$$

That is $f'(x) > 0$ for all x.This means that f is increasing. In particular, f satisfies the horizontal line test, so it has an inverse.

There are some variations. For example, if $f'(x) < 0$ for all x, then the graph $y = f(x)$ is decreasing. The horizontal line test still works, though| the graph is just going down and down, so it can't come back up and hit the same horizontal line twice. Another variation is that the derivative might be 0 for an instant but positive everywhere else. This is OK as long as the derivative doesn't stay at 0 for a long time. Here's a summary of the situation:

Theorem 14.12 (Derivatives and inverse functions) *If f is differentiable on its domain (a, b) and any of the following are true:*

1. *$f'(x) > 0$ for all x in (a,b);*
2. *$f'(x) < 0$ for all x in (a,b);*

3. *$f'(x) \geq 0$ for all x in (a,b) and $f'(x) = 0$ for only a finite number of x;*

4. *$f'(x) \leq 0$ for all x in (a,b) and $f'(x) = 0$ for only a finite number of x;*

then f has an inverse. If instead the domain is of the form $[a,b]$, or $[a,b)$, or $(a,b]$, and f is continuous on the whole domain, then f still has an inverse if any of the above four conditions are true.

Example 14.13 *Suppose $g(x) = \cos(x)$ on the domain $(0,\pi)$. Does g have an inverse? Well, $g'(x) = -\sin(x)$. We know that $\sin(x) > 0$ on the interval $(0,\pi)$ –just look at its graph if you don't believe this. Since $g'(x) = -\sin(x)$, we see that $g'(x) < 0$ for all x in $(0,\pi)$. This means that g has an inverse. In fact, we know that g has an inverse on all of $[0,\pi]$, since g is continuous there. The idea is that $g(0) = 1$, so g starts out at height 1; then, since $g'(x) < 0$ when $0 < x < \pi$, we know that g immediately gets lower than 1. Since $g(\pi) = -1$, the values of $g(x)$ go down to -1 without ever hitting the same value twice. So g has an inverse on all of $[0,\pi]$. We'll come back to this particular function in section 14.3 below.*

Example 14.14 *Finally, let $h(x) = x^3$ on all of $\mathbb{R}$. We know that $h'(x) = 3x^2$, which can't be negative. So $h'(x) \geq 0$ for all x. Luckily, $h'(x) = 0$ only when $x = 0$, so there's just one little point where $h'(x) = 0$. That's OK, so h still has an inverse; in fact (as we know), $h^{-1}(x) = \sqrt[3]{x}$.*

So the methods of this section won't work, in general, when your function has discontinuities or vertical asymptotes.

14.2.2 Finding the derivative of an inverse function

If you know that a function f has an inverse, which we'll call f^{-1} as usual, then what's the derivative of that inverse? Here's how you find it. Start

off with the equation $y = f^{-1}(x)$. You can rewrite this as $f(y) = x$. Now differentiate implicitly with respect to x to get

$$\frac{d}{dx}(f(y)) = \frac{d}{dx}(x).$$

The right-hand side is easy: it's just 1. To find the left-hand side, we use implicit differentiation. If we set $u = f(y)$, then by the chain rule (noting that $du/dy = f'(y)$), we have

$$\frac{d}{dx}(f(y)) = \frac{d}{dx}(u) = \frac{du}{dy}\frac{dy}{dx} = f'(y)\frac{dy}{dx}.$$

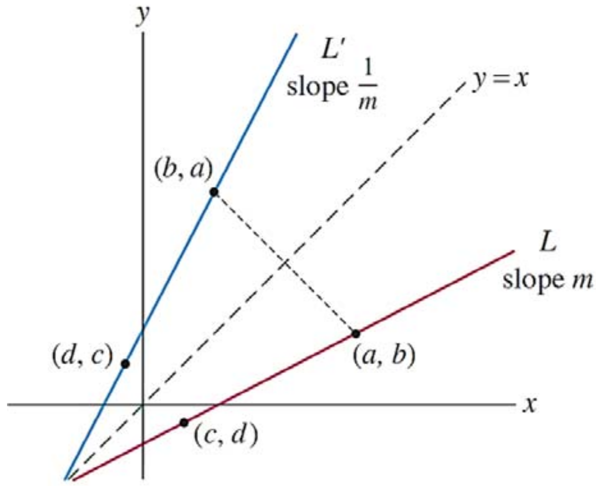


Fig. 14.10. If L has slope m, then its reflection $L\prime$ has slope $1/m$.

Now divide both sides by $f^{-1}(x)$ to get the following principle:

$$\boxed{\text{if} \quad y = f^{-1}(x), \quad \text{then} \quad \frac{dy}{dx} = \frac{1}{f'(y)}}$$

If you want to express everything in terms of x, then you have to replace y by $f^{-1}(x)$ to get

$$\boxed{\frac{d}{dx}(f^{-1}(x)) = \frac{1}{f'(f^{-1}(x))}} \tag{14.6}$$

In words, this means that if $f'(f^{-1}(x)) \neq 0$ the derivative of the inverse is basically the reciprocal of the derivative of the original function, except that you have to evaluate this latter derivative at $f^{-1}(x)$ instead of x.

We can explain the result graphically as follows. Recall that the graph of the inverse $f^{-1}(x)$ is obtained by reflecting the graph of $f(x)$ through the line $y = x$. Now, consider a line L of slope m and let L' be its reflection through $y = x$ as in Figure 14.10. Then the slope of L' is $1/m$. Indeed, if (a, b) and (c, d) are any two different points on L, then (b, a) and (d, c) lie on L' and

$$\underbrace{\text{Slope of} \quad L = \frac{b-d}{a-c}, \qquad \text{Slope of} \quad L' = \frac{a-c}{b-d}.}_{\text{Reciprocal slopes}}$$

Figure 14.11 tells the rest of the story. Let $g(x) = f^{-1}(x)$. The reflection of the tangent line to $y = f(x)$ at $x = a$ is the tangent line to $y = g(x)$ at $x = b$

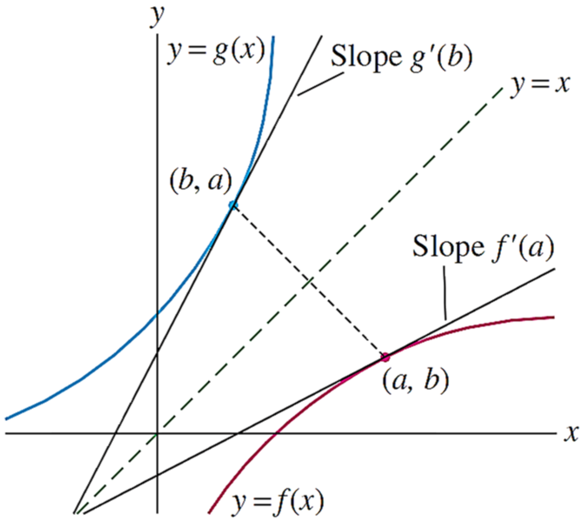


Fig. 14.11. The tangent line to the inverse $y = g(x)$ is the reflection of the tangent line to $y = f(x)$.

[where $b = f(a)$ and $a = g(b)$]. These tangent lines have reciprocal slopes, and thus $g'(b) = 1/f'(a) = 1/f'(g(b))$ as claimed in 14.6.

Example 14.15 *Suppose that $h(x) = x^3$ as in Section 14.1.1 (Example 14.4), above. We saw there $y = f(x)$ that h has an inverse, and we even have a way to write it: $h^{-1}(x) = x^{1/3}$. Of course, we could just use the rule for differentiating x^a with respect to x (a is a rational number), but let's try the above method. We know that $h'(x) = 3x^2$; if $y = h^{-1}(x)$, then*

$$\frac{dy}{dx} = \frac{1}{h'(y)} = \frac{1}{3y^2}.$$

Now we can solve the equation $x = y^3$ for y to get $y = x^{1/3}$, and substitute into the above equation to get

$$\frac{dy}{dx} = \frac{1}{3\left(x^{1/3}\right)^2} = \frac{1}{3x^{2/3}}.$$

We could just have differentiated $y = x^{1/3}$ and gotten the same answer without nearly so much work. Nevertheless it's nice to know that it all works out . □

Before we move on to another example, let's just note that the derivative of the inverse function doesn't exist when $x = 0$, since the denominator $3x^{2/3}$

vanishes. So even though the original function is differentiable everywhere, the inverse isn't differentiable everywhere: its derivative doesn't exist at $x = 0$. This is true in general, not just for the function h from above. If you have any function which has an inverse, and it has slope 0 at the point (x,y), the inverse function will have infinite slope at the point (y,x), and a vertical tangent line there.

Example 14.16 *Calculate $g'(1)$, where $g(x)$ is the inverse of $f(x) = x + e^x$.*

Solution: We have $f'(x) = 1 + e^x$, thus $f(x)$ is strictly increasing on its domain $(-\infty, \infty)$, and

$$g'(1) = \frac{1}{f'(g(1))} = \frac{1}{f'(c)} = \frac{1}{1+e^c} \qquad \text{where} \quad c = g(1).$$

We do not have to solve for $g(x)$ (which cannot be done explicitly in this case) if we can compute $c = g(1)$ directly. By definition of the inverse, $f(c) = 1$ and we see by inspection that $f(0) = 0 + e^0 = 1$. It must be the only solution because the inverse exists. Therefore, $c = 0$ and the formula above yields

$$g'(1) = \frac{1}{1+e^0} = \frac{1}{2} \qquad \square$$

Example 14.17 *Set*

$$f(x) = \frac{1}{3}x^3 - 2x^2 + 5x - 7.$$

We saw in Example 14.11 above that f has an inverse on all of $\mathbb{R}$. If we set $y = f^{-1}(x)$, then what is dy/dx in general? What is its value when $x = -7$? To do the first part, all you have to do is to see that $f'(x) = x^2 - 4x + 5$, so

$$\frac{dy}{dx} = \frac{1}{f'(y)} = \frac{1}{y^2 - 4y + 5}.$$

Note that it's important to replace x by y here. Anyway, now we can solve the second part. By the definition of f, we have

$$\frac{1}{3}y^3 - 2y^2 + 5y - 7 = -7$$

Now clearly $y = 0$ is a solution to this equation, and it must be the only solution because the inverse exists. So, when -7, we have $y = 0$, and

$$\frac{dy}{dx} = \frac{1}{(0)^2 - 4\,(0) + 5} = \frac{1}{5}.$$

More formally, one can write $\left(f^{-1}\right)'(-7) = \frac{1}{5}$. $\square$

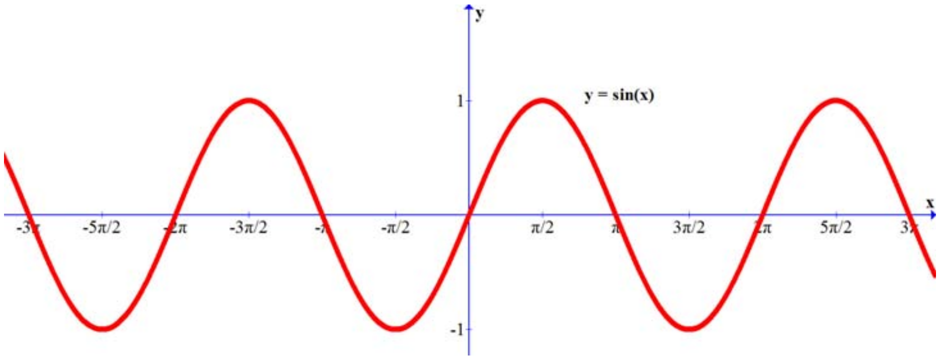


Fig. 14.12. The graph of $y = \sin(x)$

14.3 Inverse trigonometric functions

Now it's time to investigate the inverse trig functions. We'll see how to define them, what their graphs look like, and how to differentiate them. Let's look at them one at a time, beginning with inverse sine.

14.3.1 Inverse sine

Let's start by looking at the graph of $y = \sin(x)$ once again:Does the sine function have an inverse? You can see from the above graph that the horizontal line test fails pretty miserably. In fact, every horizontal line of height between -1 and 1 intersects the graph infinitely many times, which is a lot more than the zero or one time we can tolerate. So, using the tactic described in Section 14.1.4, we throw away as little of the domain as possible in order to pass the horizontal line test. There are many options, but the sensible one is to restrict the domain to the interval $[-\pi/2,\pi/2]$. Here's the effect of this:

Clearly we can't go to the right of $\pi/2$ or else we'll start repeating the values immediately to the left of $\pi/2$ as the curve dips back down. A similar thing happens at $-\pi/2$. So, we're stuck with our interval.

OK, if $f(x) = sin(x)$ with domain $[-\pi/2,\pi/2]$, then it satisfies the horizontal line test, so it has an inverse f^{-1}. We'll write $f^{-1}(x)$ as $\sin^{-1}(x)$ or $\arcsin(x)$.

Remark 14.18 *Beware: The first of these notations is a little confusing at first, since* $\sin^{-1}(x)$ *does not mean the same thing as* $(sin(x))^{-1}$*, even though* $\sin^2(x) = (sin(x))^2$ *and* $sin^3(x) = (sin(x))^3$.

So, what is the domain of the inverse sine function? Well, since the range of $f(x) = \sin(x)$ is $[-1,1]$, the domain of the inverse function is $[-1,1]$. And since

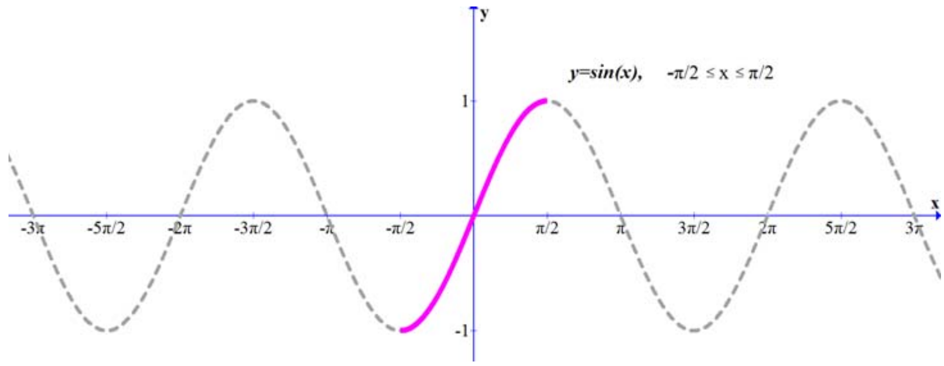


Fig. 14.13. Function $\sin(x)$ on the restricted domain $[-\pi/2, \pi/2]$ passes the horizontal line test.

the domain of our function f is $[-\pi/2,\pi/2]$ (since that's how we restricted the domain), the range of the inverse is $[-\pi/2,\pi/2]$.

How about the graph of $y = \sin^{-1}(x)$? We just have to take the restricted graph of $y = \sin(x)$ and reflect it in the mirror line $y = x$; it looks like this (see Figure 14.14):Note that since $\sin(x)$ is an odd function of x, so is $\sin^{-1}(x)$. This is consistent with the above graphs.

Now let's differentiate the inverse sine function. Set $y = \sin^{-1}(x)$; we want to find dy/dx. The snazziest way to do this is to write $x = sin(y)$ and then differentiate both sides implicitly with respect to x:

$$\frac{d}{dx}(x) = \frac{d}{dx}\left(\sin(y)\right).$$

The left-hand side is just 1, but the right-hand side needs the chain rule. You should check that you get $\cos(y)\dfrac{dy}{dx}$. So we have

$$1 = \cos(y)\frac{dy}{dx}$$

which simplifies to

$$\frac{dy}{dx} = \frac{1}{\cos(y)}.$$

Actually, we could have written this down immediately using the Formula 14.6 from above. Now, we really want the derivative in terms of x, not y. No problem–we know that $\sin(y) = x$, so it shouldn't be too hard to find $\cos(y)$. In fact, $\cos^2(y) + \sin^2(y) = 1$, which means that $\cos^2(y) + x^2 = 1$. This leads to the equation $\cos(y) = \pm\sqrt{1 - x^2}$, so we have

$$\frac{dy}{dx} = \pm\frac{1}{\sqrt{1 - x^2}}.$$

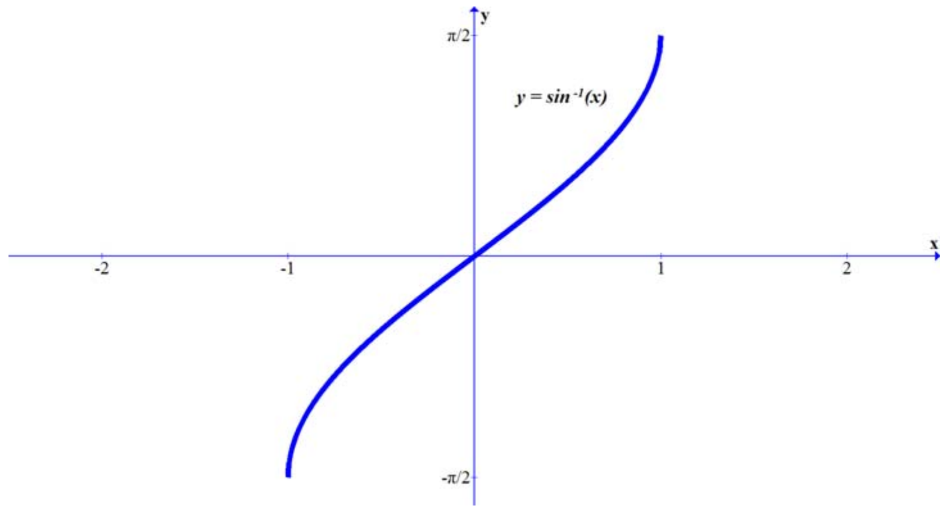


Fig. 14.14. The graph of $y = \sin^{-1}(x)$

But which is it? Plus or minus? If you look at the graph of $y = \sin^{-1}(x)$ above, you can see that the slope is always positive. This means that we have to take the positive square root:

$$\boxed{\frac{d}{dx}\sin^{-1}(x) = \frac{1}{\sqrt{1-x^2}} \qquad \text{for} \quad -1 < x < 1.} \tag{14.7}$$

Note that $\sin^{-1}(x)$ is not differentiable, even in the one-sided sense, at the endpoints $x = 1$ and $x = -1$, since the denominator$\sqrt{1-x^2}$ is 0 in both these cases.

In addition to the derivative formula and the above graph, here's a summary of the important facts about the inverse sine function

$$\boxed{\sin^{-1} \text{ is odd; } \quad \text{it has domain } [-1{,}1] \quad \text{and range } [-\pi/1, \pi/2].} \tag{14.8}$$

Now that you have a new derivative formula, you should become comfortable using the product, quotient, and chain rules in association with it.

Example 14.19 *Calculate* $f'(\frac{1}{2})$*, where* $f(x) = \arcsin\left(x^2\right)$*.*

Solution: By the chain rule

$$\begin{aligned} \frac{d}{dx}\arcsin\left(x^2\right) &= \frac{d}{dx}\sin^{-1}\left(x^2\right) \\ &= \frac{1}{\sqrt{1-x^4}}\frac{d}{dx}\left(x^2\right) \\ &= \frac{2x}{\sqrt{1-x^4}}. \end{aligned}$$

$$f'\left(\frac{1}{2}\right) = \frac{2\left(\frac{1}{2}\right)}{\sqrt{1-\left(\frac{1}{2}\right)^4}} = \frac{1}{\sqrt{\frac{15}{16}}} = \frac{4}{\sqrt{15}} \approx 1.0328. \qquad \square$$

Example 14.20 *Calculate* $d/dx\left(\arcsin\left(4x^2\right)\right)$.

Solution:

$$\frac{d}{dx}\left(\arcsin\left(4x^2\right)\right) = \underbrace{\frac{1}{\sqrt{1-\left(4x^2\right)^2}} \cdot \frac{d}{dx}\left(4x^2\right)}_{\text{chain rule}} = \frac{8x}{\sqrt{1-16x^4}}. \qquad \square$$

Remark 14.21 *We continue with the convention that if the domain of a function* f *is not specified explicitly, then it is understood to be the maximal set of real numbers* x *for which* $f(x)$ *is a real number. In this case, the domain is the set of real numbers* x *for which* $-1 \le 4x^2 \le 1$*. This is the interval* $[-1/2, 1/2]$.

The integral counterpart of (14.7) reads

$$\boxed{\int \frac{1}{\sqrt{1-x^2}}dx = \arcsin x + C.} \tag{14.9}$$

Example 14.22 *Show that for* $a > 0$

$$\int \frac{1}{\sqrt{a^2-x^2}}dx = \arcsin\frac{x}{a} + C.$$

Solution: We change variables so that a^2 is replaced by 1 and we can use (14.9). To this end we set

$$au = x, \quad a\ du = dx.$$

Then

$$\begin{aligned} \int \frac{1}{\sqrt{a^2-x^2}}dx &= \int \frac{a\ du}{\sqrt{a^2-a^2u^2}} \\ &= \int \underbrace{\frac{a\ du}{a\sqrt{1-u^2}}}_{\text{since } a>0} = \int \frac{du}{\sqrt{1-u^2}} \\ &= \arcsin u + C = \arcsin\frac{x}{a} + C. \quad \square \end{aligned}$$

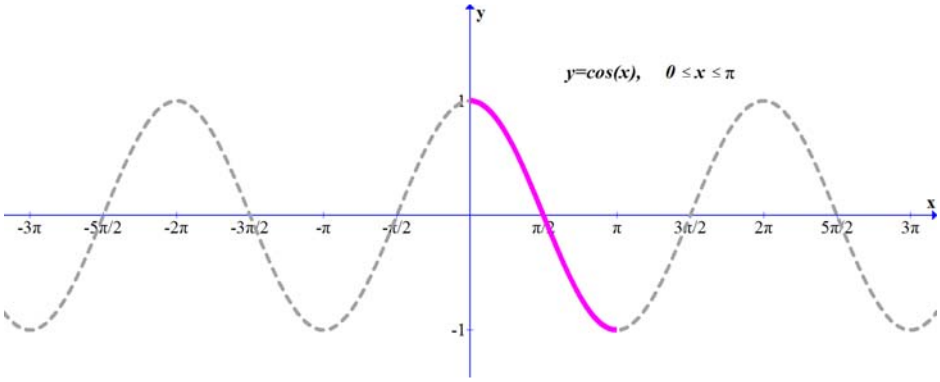


Fig. 14.15.

14.3.2 Inverse cosine

We're going to repeat the procedure from the previous section in order to understand the inverse cosine function. Start with the graph of $y = \cos(x)$:

Once again, no inverse. This time, restricting the domain to $[-\pi/2, \pi/2]$ won't work, since the horizontal line test would fail and also we'd be throwing away part of the range that would be useful. Already on the above graph, you can see that the section between $[0, \pi]$ is highlighted and obeys the horizontal line test, so that's what we'll use. We get an inverse function which we write as $\cos^{-1}$ or arccos. Like inverse sine, the domain of inverse cosine is $[-1, 1]$, since that's the range of cosine. On the other hand, the range of inverse cosine is $[0, \pi]$, since that's the restricted domain of cosine that we're using. The graph of $y = \cos^{-1} x$ is formed by reflecting the graph of $y = \cos(x)$ in the mirror $y = x$:

Example 14.23 *Evaluate* ***a)*** $\sin^{-1}\left(\frac{\sqrt{3}}{2}\right)$ *and* ***b)*** $\cos^{-1}\left(-\frac{1}{2}\right)$

Solution:
a) We see that

$$\sin^{-1}\left(\frac{\sqrt{3}}{2}\right) = \frac{\pi}{3}$$

because $\sin(\pi/3) = \sqrt{3}/2$ and $\pi/3$ belongs to the range $[-\pi/2, \pi/2]$ of the arcsin function (see Figure 14.14).
b) We have

$$\cos^{-1}\left(-\frac{1}{2}\right) = \frac{2\pi}{3}$$

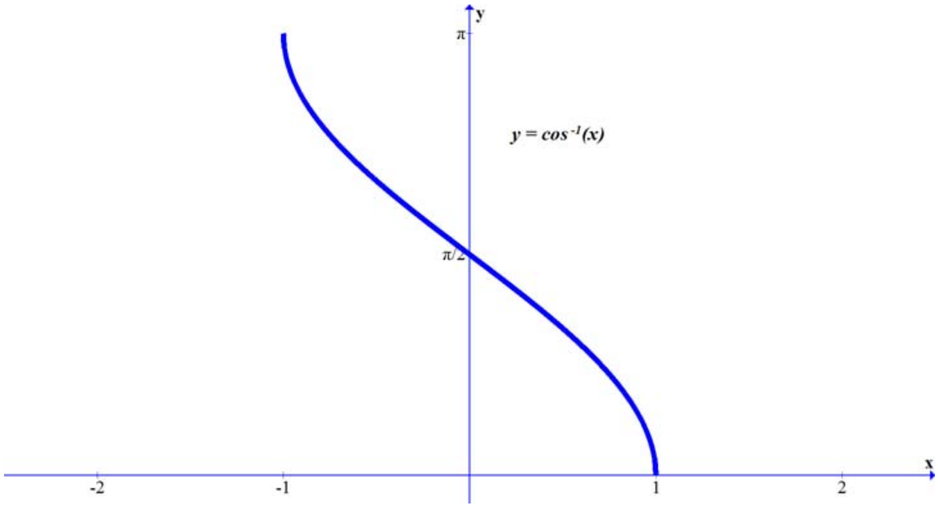


Fig. 14.16. The graph of $y = \cos^{-1}(x)$.

$\mathbf{x}$	$\sin^{-1}\mathbf{x}$	$\cos^{-1}\mathbf{x}$
$\sqrt{3}/2$	$\pi/3$	$\pi/6$
$\sqrt{2}/2$	$\pi/4$	$\pi/4$
$1/2$	$\pi/6$	$\pi/3$
$-1/2$	$-\pi/6$	$2\pi/3$
$-\sqrt{2}/2$	$-\pi/4$	$3\pi/4$
$-\sqrt{3}/2$	$-\pi/3$	$5\pi/6$

Table 14.1. Table of common values for the arcsine and arccosine functions

because $\cos\left(\frac{2\pi}{3}\right) = -\frac{1}{2}$ and $\frac{2\pi}{3}$ belongs to the range $[0, \pi]$ of the arccos function. □

Using the same procedure illustrated in Example 14.23 , we can create the following table of common values of the arcsine and arccosine function:

Remark 14.24 *Notice that the graph shows that* $\cos^{-1} x$ *is n*either even nor odd*. This is despite the fact that* $\cos(x)$ *is an even function of* x*!*

Now it's time to differentiate $y = \cos^{-1} x$ with respect to x. We do exactly the same thing we did in the previous section obtaining

$$\frac{dy}{dx} = \frac{-1}{\pm\sqrt{1 - x^2}}.$$

Unlike the case of inverse sine, the graph of inverse cosine is all downhill, which means that the slope is always negative, so we get

$$\boxed{\frac{d}{dx}\cos^{-1}(x) = -\frac{1}{\sqrt{1-x^2}} \qquad \text{for} \quad -1 < x < 1.}$$

Here are the other facts about inverse cosine that we collected above:

$$\boxed{\cos^{-1} \text{ is neither even nor odd;} \quad \text{it has domain } [-1,1] \quad \text{and range } [0,\pi].}$$

Before we move on to the inverse tangent function, let's just look at the derivatives of inverse sine and inverse cosine side by side:

$$\frac{d}{dx}\sin^{-1}(x) = \frac{1}{\sqrt{1-x^2}} \qquad \text{and} \qquad \frac{d}{dx}\cos^{-1}(x) = -\frac{1}{\sqrt{1-x^2}}$$

The derivatives are negatives of each other! So,

$$\frac{d}{dx}\left(\sin^{-1}(x) + \cos^{-1}(x)\right) = \frac{1}{\sqrt{1-x^2}} - \frac{1}{\sqrt{1-x^2}} = 0.$$

It means that the function $y = \sin^{-1}(x) + \cos^{-1}(x)$ has a constant slope 0, which means that it's flat. The value of this function at the point 0 can be easily calculated, It is equal to $\pi/2$. We've just used calculus to prove the following identity:

$$\sin^{-1}(x) + \cos^{-1}(x) = 1 \qquad \text{for any } x \text{ in the interval } [-1,1].$$

14.3.3 Inverse tangent

Let's remember the graph of $y = \tan(x)$ (Figure 14.17):We'll restrict the domain to $(-\pi/2,\pi/2)$ so that we can get an inverse function $\tan^{-1}$, also written as arctan. The domain of this function is the range of the tangent function, which is all of $\mathbb{R}$. The range of the inverse function is $(-\pi/2,\pi/2)$, which of course is the restricted domain of $\tan(x)$ that we're using. The graph of $y = \tan^{-1}(x)$ is presented on Figure 14.18.Now $\tan^{-1}(x)$ is an odd function of x, as you can see from the graph| it inherits its oddness from that of $\tan(x)$, in fact.

Now let's differentiate $y = \tan^{-1}(x)$ with respect to x. Write $x = \tan(y)$ and differentiate implicitly with respect to x. Check to make sure that you believe that

$$\frac{dy}{dx} = \frac{1}{\sec^2 y}.$$

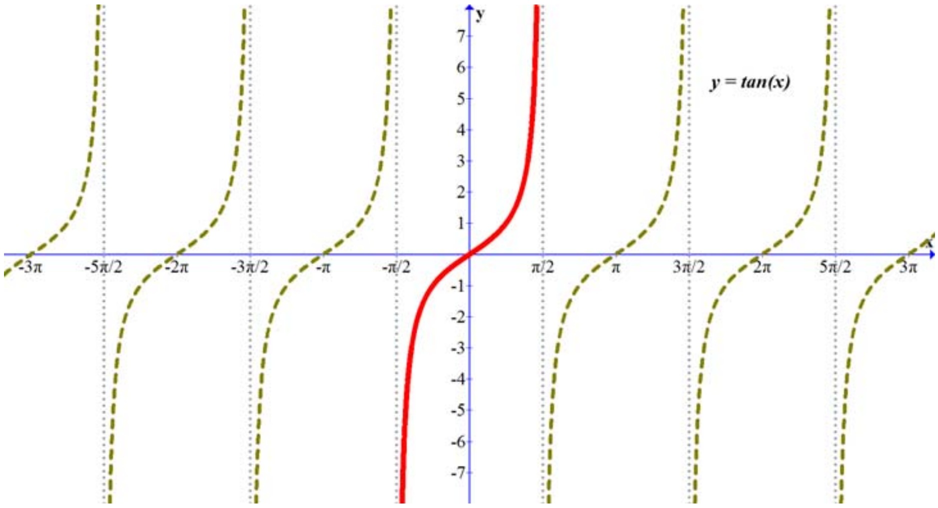


Fig. 14.17. The graph of $y = \tan(x)$.

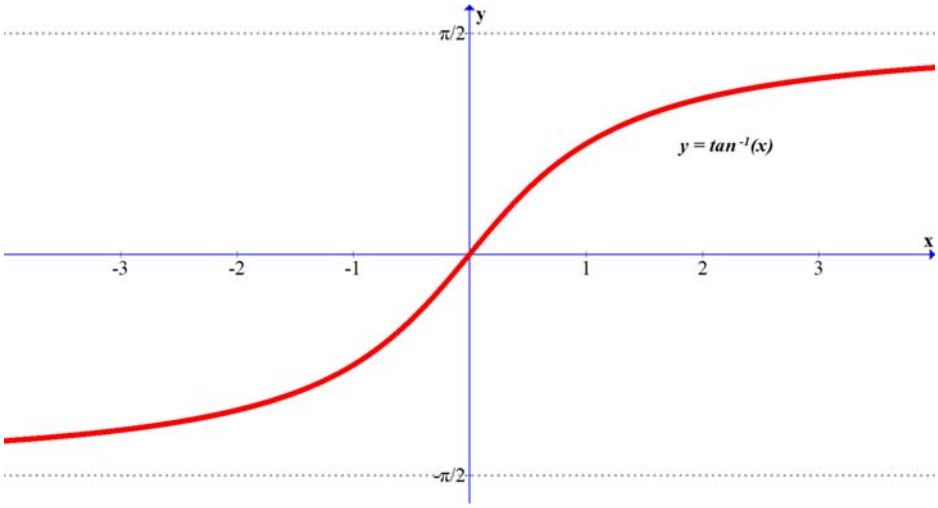


Fig. 14.18. The graph of $y = \tan^{-1}(x)$.

Since $\sec^2 y = 1 + \tan^2 y$, and $\tan(y) = x$,we see that $\sec^2(y) = 1 + x^2$. This means that

$$\boxed{\frac{d}{dx}\tan^{-1}(x) = \frac{1}{1+x^2} \qquad \text{for all real } x.} \tag{14.10}$$

We also have the following facts from above:

$$\boxed{\tan^{-1} \text{ is odd}; \quad \text{it has domain } \mathbb{R} \quad \text{and range } (-\pi/2, \pi/2).} \tag{14.11}$$

Unlike inverse sine and inverse cosine, the inverse tangent function has horizontal asymptotes. (The first two functions don't have a chance, since their domains are both [−1,1].) As you can see from the graph above, $\tan^{-1}(x)$ tends to $\pi/2$ as $x \to \infty$, and it tends to $-\pi/2$ as $x \to -\infty$. In fact, the vertical asymptotes $x = \pi/2$ and $x = -\pi/2$ of the tangent function have become horizontal asymptotes of the inverse tan function. This means that we have the following useful limits:

$$\boxed{\lim_{x\to\infty} \tan^{-1}(x) = \pi/2} \qquad \text{and} \qquad \boxed{\lim_{x\to-\infty} \tan^{-1}(x) = -\pi/2.} \tag{14.12}$$

The integral counterpart of (14.10) reads

$$\boxed{\int \frac{dx}{1+x^2} = \arctan x + C.} \tag{14.13}$$

Example 14.25 *Show that, for $a \neq 0$,*

$$\boxed{\int \frac{dx}{a^2+x^2} = \frac{1}{a}\arctan\frac{x}{a} + C.} \tag{14.14}$$

Solution: We change variables so that a^2 is replaced by 1 and we can use (14.13). We set

$$au = x, \quad a\,du = dx.$$

$$\begin{aligned}\int \frac{dx}{a^2+x^2} &= \int \frac{a\,du}{a^2\,(1+u^2)} = \frac{1}{a}\int \frac{du}{(1+u^2)} \\ &= \frac{1}{a}\arctan u + C = \frac{1}{a}\arctan\frac{x}{a} + C. \quad \square\end{aligned}$$

Example 14.26 *Evaluate*

$$\int \frac{dx}{4x^2+4x+2}.$$

Solution: We complete the square on the binomial $4x^2 + 4x$:

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

Then,

$$\int \frac{dx}{4x^2 + 4x + 2} = \int \frac{dx}{1 + (2x+1)^2}$$

We change variables[1] so that $(2x+1)^2$ is replaced by u^2 and we can use (14.13). We set

$$u = (2x+1), \quad \frac{du}{2} = dx.$$

$$\begin{aligned} \int \frac{dx}{1 + (2x+1)^2} &= \frac{1}{2} \int \frac{du}{1 + u^2} \\ &= \frac{1}{2} \arctan u + C = \frac{1}{2} \arctan(2x+1) + C \quad \square \end{aligned}$$

Example 14.27 *Evaluate* $\int_0^2 \dfrac{dx}{4 + x^2}$.

Solution: By 14.14

$$\int \frac{dx}{4 + x^2} = \int \frac{dx}{2^2 + x^2} = \frac{1}{2} \arctan \frac{x}{2} + C,$$

and therefore

$$\begin{aligned} \int_0^2 \frac{dx}{4 + x^2} &= \frac{1}{2} \arctan \frac{x}{2} \Big|_0^2 \\ &= \frac{1}{2} \arctan 1 - \frac{1}{2} \arctan 0 = \frac{1}{8} \pi \quad \square \end{aligned}$$

14.3.4 Arc cotangent, arc secant, arc cosecant

These functions are not as important to us as the arc sine, arc cosine and arc tangent, but they do deserve some attention.

Arc cotangent: The cotangent function is one-to-one on $(0, \pi)$ and maps that interval onto $(-\infty, \infty)$. The arc cotangent function

$$y = \operatorname{arccot} x, \quad x \in (-\infty, \infty)$$

[1] This technique of integration will be developed later

is the inverse of the function

$$y = \cot x, \quad x \in (0, \pi).$$

Arc secant, arc cosecant: These functions can be defined explicitly in terms of the arc cosine and the arc sine. For $|x| \geq 1$, we set

$$\operatorname{arcsec} x = \arccos(1/x), \quad \arccos x = \arcsin(1/x).$$

Easily you can check, that that for all $|x| \geq 1$

$$\sec(\operatorname{arcsec} x) = x \quad \text{and} \quad \csc(\arccos x) = x.$$

and

$$\begin{aligned} \arctan x + \operatorname{arccot} x &= \frac{\pi}{2}, \\ \operatorname{arcsec} x + \operatorname{arccsc} x &= \frac{\pi}{2}. \end{aligned} \tag{14.15}$$

Derivatives:

$$\boxed{\begin{aligned} \frac{d}{dx}(\operatorname{arccot} x) &= -\frac{1}{1+x^2} \\ \frac{d}{dx}(\operatorname{arcsec} x) &= \frac{1}{|x|\sqrt{x^2-1}} \\ \frac{d}{dx}(\operatorname{arccsc} x) &= -\frac{1}{|x|\sqrt{x^2-1}}. \end{aligned}} \tag{14.16}$$

The derivatives of the arc sine, arc cosine and the arc tangent were calculated earlier. That the derivatives of the arc cotangent is as stated follows immediately from (14.15). Once we show that

$$\frac{d}{dx}(\operatorname{arcsec} x) = \frac{1}{|x|\sqrt{x^2-1}}$$

the last formula will follow from (14.15). Hence we focus on the arc secant. Since

$$\operatorname{arcsec} x = \arccos(1/x)$$

the chain rule gives

$$\begin{aligned} \frac{d}{dx}(\operatorname{arcsec} x) &= -\frac{1}{\sqrt{1-(1/x)^2}} \cdot \frac{d}{dx}\left(\frac{1}{x}\right) \\ &= -\frac{\sqrt{x^2}}{\sqrt{x^2-1}}\left(-\frac{1}{x^2}\right) = \frac{\sqrt{x^2}}{x^2\sqrt{x^2-1}}. \end{aligned}$$

This tells us that

$$\frac{d}{dx}(\operatorname{arcsec} x) = \begin{cases} \dfrac{1}{x\sqrt{x^2-1}} & \text{for} \quad x > 1 \\[2ex] -\dfrac{1}{x\sqrt{x^2-1}} & \text{for} \quad x < -1. \end{cases}$$

The statement

$$\frac{d}{dx}(\operatorname{arcsec} x) = \frac{1}{|x|\sqrt{x^2-1}}$$

is just a summary of this result.

15
L'Hospital's rule and overview of limits

We've used limits to find derivatives. Now we'll turn things upside-down and use derivatives to find limits, by way of a nice technique called L'Hospital's Rule. After looking at various varieties of the rule, we'll give a summary, followed by an overview of all the methods we've used so far to evaluate limits. So, we'll look at:

- L'Hospital's Rule, and four types of limits which naturally lead to using the rule; and
- a summary of limit techniques from earlier chapters.

15.1 L'Hospital's Rule

Most of the limits we've looked at are naturally in one of the following forms:

$$\lim_{x\to c}\frac{f(x)}{g(x)} \quad \lim_{x\to c}(f(x)-g(x)) \quad \lim_{x\to c}f(x)g(x) \quad \text{and} \quad \lim_{x\to c}f(x)^{g(x)}$$

Sometimes you can just substitute $x = c$ and evaluate the limit directly, effectively using the continuity of f and g. This method doesn't always work, though for example, consider the limits

$$\lim_{x\to 3}\frac{x^{2-}9}{x-3} \quad \lim_{x\to 0}\left(\frac{1}{\sin x}-\frac{1}{x}\right) \quad \lim_{x\to 0^+}x\ln x \quad \text{and} \quad \lim_{x\to 0}(1+3\tan(x))^{\frac{1}{x}}$$

In the first case, replacing x by 3 gives the indeterminate form $0/0$. The second limit involves the difference between two terms which become infinite as $x \to 0$. Actually, they both go to ∞ as $x \to 0^+$ and -∞ as $x \to 0^-$, so you can think of the form in this case as $\pm(\infty - \infty)$. As for the third limit above (involving $x\ln(x)$), this leads to the form $0 \times (-\infty)$, remembering that $\ln(x) \to -\infty$ as $x \to 0^+$. Finally, the fourth limit looks like 1^∞, which is also problematic. When we encountered one of these indeterminate forms earlier in the text, we attempted to rewrite the expression by using various algebraic techniques.

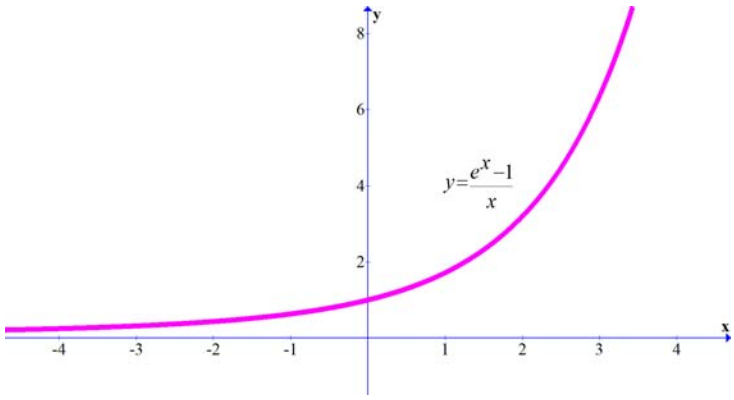


Fig. 15.1. The limit as x approaches 0 appears to be 1.

However, not all indeterminate forms can be evaluated by algebraic manipulation. This is often true when both algebraic and transcendental functions are involved. For instance, the limit

$$\lim_{x\to 0} \frac{e^x - 1}{x}$$

produces the indeterminate form 0/0. Rewriting the expression to obtain

$$\lim_{x\to 0} \left(\frac{e^x}{x} - \frac{1}{x} \right)$$

merely produces another indeterminate form, $\infty - \infty$. Of course, you could use technology to estimate the limit, as shown in Figure 15.1. From and the graph, the limit appears to be 1. (This limit will be verified in Example 15.4.) Luckily, all four types can often be solved using L'Hospital's Rule.

It turns out that the first type, involving the ratio $f(x)/g(x)$, is the most suitable for applying the rule, so we'll call it "Type A." The next two types, involving $f(x) - g(x)$ and $f(x)g(x)$, both reduce directly to Type A, so we'll call them Type $B1$ and Type $B2$, respectively. Finally, we'll say that limits involving exponentials like $f(x)^{g(x)}$ are Type C, since you can solve them by reducing them to Type $B2$ and then back to Type A. Let's look at all these types individually, then summarize the whole situation in Section 15.7 below.

15.2 Type A: 0/0 case

Consider limits of the form

$$\lim_{x\to c} \frac{f(x)}{g(x)}$$

where f and g are nice differentiable functions. If $g(c) \neq 0$, everything's great–you just substitute $x = c$ to see that the limit is $f(c)/g(c)$. If $g(c) = 0$ but $f(c) \neq 0$, then you're dealing with a vertical asymptote at $x = c$ and the above limit is either ∞, $-\infty$ or it doesn't exist.

The only other possibility is that $f(c) = 0$ and $g(c) = 0$. That is, the fraction $f(c)/g(c)$ is the **indeterminate form** $0/0$. The majority of the limits we've seen have been of this form. In fact, every derivative is of this form!

To prove the main theorem of this chapter, we can use a more general result called the Extended Mean Value Theorem.

Theorem 15.1 (The Extended Mean Value Theorem) *If f and g are differentiable on an open interval (a, b) and continuous on $[a, b]$such that $g'(x) \neq 0$ for any x in (a, b) then there exists a point c in (a, b) such*

$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

Proof. You can assume that $g(b) \neq g(a)$ because otherwise, by Rolle's Theorem, it would follow that $g'(x) = 0$ for some x in (a, b). Now, define $h(x)$ as

$$h(x) = f(x) - \left(\frac{f(b) - f(a)}{g(b) - g(a)}\right) g(x).$$

Then

$$h(a) = f(a) - \left(\frac{f(b) - f(a)}{g(b) - g(a)}\right) g(a) = \frac{f(a)g(b) - f(b)g(a)}{g(b) - g(a)}$$

and

$$h(b) = f(b) - \left(\frac{f(b) - f(a)}{g(b) - g(a)}\right) g(b) = \frac{f(a)g(b) - f(b)g(a)}{g(b) - g(a)}$$

and by Rolle's Theorem there exists a point c in (a, b) such that

$$h'(c) = f'(c) - \left(\frac{f(b) - f(a)}{g(b) - g(a)}\right) g'(c) = 0$$

which implies that

$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

■

To see why this is called the Extended Mean Value Theorem, consider the special case in which $g(x) = x$ For this case, you obtain the "standard" Mean Value Theorem as presented before

Theorem 15.2 *Let f and g be functions that are differentiable on an open interval (a, b) containing c, except possibly at c itself. Assume that $g'(x) \neq 0$ for all x in (a, b) except possibly at c itself. If the limit of $f(x)/g(x)$ as approaches c produces the indeterminate form $0/0$, then*

$$\lim_{x \to c} \frac{f(x)}{g(x)} = \lim_{x \to c} \frac{f'(x)}{g'(x)} \tag{15.1}$$

provided the limit on the right exists (or is infinite). This result also applies if the limit of $\frac{f(x)}{g(x)}$ as x approaches c produces any one of the indeterminate forms ∞/∞, $(-\infty)/\infty$, $\infty/(-\infty)$, or $(-\infty)/(-\infty)$

We can use the Extended Mean Value Theorem to prove L'Hopital's Rule. Of the several different cases of this rule, the proof of only one case is illustrated. The remaining cases where $x \to c^-$ and $x \to c$ are left for you to prove.

Proof. Consider the case for which $\lim_{x \to c^+} f(x) = 0$ and $\lim_{x \to c^+} g(x) = 0$. Define the following new functions:

$$F(x) = \begin{cases} f(x), & x \neq c \\ 0, & x = c \end{cases} \quad \text{and} \quad G(x) = \begin{cases} g(x), & x \neq c \\ 0, & x = c \end{cases}$$

For any x, $c < x < b$, F and G are differentiable on $(c, x]$ and continuous on $(c, x]$.We can apply the Extended Mean Value Theorem to conclude that there exists a number z in (c, x) such that

$$\frac{F'(z)}{G'(z)} = \frac{F(x) - F(c)}{G(x) - G(c)} = \frac{F(x)}{G(x)} = \frac{f'(x)}{g'(x)} = \frac{f(x)}{g(x)}$$

Finally, by letting approach x from the right $x \to c^+$, you have $z \to c^+$ because $c < z < x$, and

$$\lim_{x \to c^+} \frac{f(x)}{g(x)} = \lim_{z \to c^+} \frac{f'(z)}{g'(z)} = \lim_{x \to c^+} \frac{f'(x)}{g'(x)}.$$

■

Example 15.3 *Use L'Hospital's Rule to evaluate the from the beginning of the chapter:*

$$\lim_{x \to 3} \frac{x^{2-}9}{x - 3}.$$

Solution: $f(x) = x^{2-}9$ and $g(x) = x - 3$ are differentiable and $f(3) = g(3) = 0$, so the quotient is indeterminate at $x = 3$:

$$\frac{x^{2-}9}{x - 3}|_3 = \frac{9 - 9}{3 - 9} = \frac{0}{0} \qquad \text{(indeterminate)}.$$

Furthermore, $g'(3) = 1$, and thus $g'(3)$ is nonzero. Therefore, L'Hopital's Rule applies. We may replace the numerator and denominator by their derivatives to obtain

$$\lim_{x \to 3} \frac{x^{2-}9}{x-3} = \frac{2 \cdot 3}{1} = 6$$

By the way, you don't need to use L'Hopital's Rule here–you can just factor $x^{2-}9$ as $(x-3)(x+3)$, like this

$$\lim_{x \to 3} \frac{x^{2-}9}{x-3} = \lim_{x \to 3} \frac{(x-3)(x+3)}{x-3} = \lim_{x \to 3}(x+3) = 3+3 = 6.$$

□

Example 15.4 (Indeterminate form 0/0**)** *Evaluate*

$$\lim_{x \to 0} \frac{e^x - 1}{x}.$$

Solution: Because direct substitution results in the indeterminate form 0/0,we can apply L'Hopital's Rule, as shown below.

$$\lim_{x \to 0} \frac{e^x - 1}{x} = \lim_{x \to 0} \frac{\frac{d}{dx}\left[e^x - 1\right]}{\frac{d}{dx}\left[x\right]} = \lim_{x \to 0} \frac{e^x}{1} = 1$$

□

Here's a harder example where the factoring trick doesn't work:

Example 15.5 (Applying L'Hopital's Rule more than once) *Find*

$$\lim_{x \to 0} \frac{x - \sin x}{x^3}$$

Solution: If you put $x = 0$, then both top and bottom are 0. The principle that $\sin(x)$ behaves like x for small x is useless in this case, since we're taking the difference of the two quantities. So let's apply L'Hopital's Rule, differentiating $x - \sin x$ and x^3 separately:

$$\lim_{x \to 0} \frac{x - \sin x}{x^3} = \lim_{x \to 0} \frac{1 - \cos x}{3x^2}$$

We can actually solve right hand limit, using the trick of multiplying top and bottom by $1 + \cos(x)$. There's an easier way: just notice that the right-hand limit is also of the form 0/0 when you replace x by 0 (since $\cos(0) = 1$), so we can use L'Hopital's Rule again! We get

$$\lim_{x \to 0} \frac{x - \sin x}{x^3} = \lim_{x \to 0} \frac{1 - \cos x}{3x^2} = \lim_{x \to 0} \frac{\sin x}{6x}.$$

We could actually use L'Hopital's Rule once more to find the final limit, but a better way is to write

$$\lim_{x\to 0}\frac{\sin x}{6x} = \frac{1}{6}\lim_{x\to 0}\frac{\sin x}{x} = \frac{1}{6}\times 1 = \frac{1}{6}.$$

□

15.3 Type A: $\pm\infty/\pm\infty$ case

L'Hopital's Rule also works in the case where $\lim_{x\to c} f(x) = \infty$ and $\lim_{x\to c} g(x) = \infty$. That is, when you try to put $x = c$, the top and bottom both look infinite, so you are dealing with the indeterminate form ∞/∞.

Example 15.6 *Evaluate*

$$\lim_{x\to\infty}\frac{\ln x}{x}.$$

Solution: Because direct substitution results in the indeterminate form ∞/∞ you can apply L'Hopital's Rule to obtain

$$\lim_{x\to\infty}\frac{\ln x}{x} = \lim_{x\to\infty}\frac{\frac{d}{dx}[\ln x]}{\frac{d}{dx}[x]} = \lim_{x\to\infty}\frac{1}{x} = 0$$

□

Example 15.7 *Evaluate*

$$\lim_{x\to\infty}\frac{x}{e^x}.$$

Solution:

$$\lim_{x\to\infty}\frac{x}{e^x} = \lim_{x\to\infty}\frac{1}{e^x} = 0.$$

□

The last limit is 0 because $e^x \to \infty$ as $x \to \infty$. Also, the justification for using L'Hopital's Rule is that both x and e^x go to ∞ as $x \to \infty$. Notice that the denominator e^x was unscathed by the differentiation, but the numerator x was knocked down to 1. This is even clearer when you consider the following example.

Example 15.8 *Find*

$$\lim_{x\to\infty}\frac{x^3}{e^x}.$$

Solution: Just use L'Hopital's Rule three times, noting that in each case we are dealing with the indeterminate form ∞/∞.

$$\lim_{x\to\infty}\frac{x^3}{e^x}=\lim_{x\to\infty}\frac{3x^2}{e^x}=\lim_{x\to\infty}\frac{6x}{e^x}=\lim_{x\to\infty}\frac{6}{e^x}=0.$$

Of course, the same technique applies to any power of x; you just have to apply the rule enough times, knocking the power down by 1 each time, while the e^x just sits there like some immovable lump. So we have proved the principle that exponentials grow quickly. □

Warning: Please, please, please check that you have an indeterminate form! The only acceptable forms for a quotient are $0/0$ or $\pm\infty/\pm\infty$. For example, if you try to use L'Hopital's Rule on the limit

$$\lim_{x\to 0}\frac{x^2}{\cos x}$$

you'll get into a real tangle. Let's see what happens:

$$\lim_{x\to 0}\frac{x^2}{\cos x}\stackrel{?}{=}\lim_{x\to 0}\frac{2x}{-\sin x}\stackrel{?}{=}-2\lim_{x\to 0}\frac{x}{\sin x}=-2.$$

This is clearly wrong, since x^2 and $\cos x$ are both positive when x is near 0.

In fact, the correct solution is

$$\lim_{x\to 0}\frac{x^2}{\cos x}=\frac{0^2}{\cos 0}=\frac{0}{1}=0.$$

Example 15.9 (Funy example) *Let us calculate*

$$\lim_{x\to\infty}\frac{x}{\sqrt{x^2+1}}$$

Solution: It is indeterminate form $\dfrac{\infty}{\infty}$. Let us apply L'Hospital's Rule

$$\lim_{x\to\infty}\frac{x}{\sqrt{x^2+1}}=\lim_{x\to\infty}\frac{\frac{d}{dx}[x]}{\frac{d}{dx}\left[\sqrt{x^2+1}\right]}=\lim_{x\to\infty}\frac{1}{\frac{x}{\sqrt{x^2+1}}}=\lim_{x\to\infty}\frac{\sqrt{x^2+1}}{x}$$

This is $\dfrac{\infty}{\infty}$ and we can try to use L'Hospital's Rule again

$$\lim_{x\to\infty}\frac{x}{\sqrt{x^2+1}}=\lim_{x\to\infty}\frac{\sqrt{x^2+1}}{x}=\lim_{x\to\infty}\frac{\frac{d}{dx}\left[\sqrt{x^2+1}\right]}{\frac{d}{dx}[x]}=\lim_{x\to\infty}\frac{x}{\sqrt{x^2+1}}$$

This is $\frac{\infty}{\infty}$. Should we use L'Hospital's Rule again? No! (Because we are looping ! We are now exactly at the initial position!) L'Hospital's Rule does not work here. But we can conclude, that If the limit is A, then the procedure tells that it is equal to $1/A$. The limit must therefore be 1. We can use elementary techniques to find that limit.

$$\lim_{x\to\infty} \frac{x}{\sqrt{x^2+1}} = \lim_{x\to\infty} \frac{\frac{x}{x}}{\frac{\sqrt{x^2+1}}{x}} = \lim_{x\to\infty} \frac{1}{\sqrt{1+\frac{1}{x^2}}} = 1$$

□

15.4 Type $B1$: $(\infty - \infty)$

Here's a limit from the beginning of this chapter:

$$\lim_{x\to 0} \left(\frac{1}{\sin x} - \frac{1}{x} \right)$$

As $x \to 0^+$, both $1/\sin x$ and $1/x$ go to ∞. As $x \to 0^-$, both quantities go to $-\infty$. Either way, you're looking at the difference of two huge (positive or negative) quantities, so we can express the indeterminate form as $(\infty - \infty)$.

Luckily, it's pretty easy to reduce this to Type A. Just take a common denominator:

$$\lim_{x\to 0} \left(\frac{1}{\sin x} - \frac{1}{x} \right) = \lim_{x\to 0} \frac{x - \sin x}{x \sin x}$$

Now you can put $x = 0$ and see that we are in the $0/0$ case. So we can apply L'Hopital's Rule:

$$\lim_{x\to 0} \left(\frac{1}{\sin x} - \frac{1}{x} \right) = \lim_{x\to 0} \frac{x - \sin x}{x \sin x} = \lim_{x\to 0} \frac{1 - \cos x}{\sin x + x \cos x}$$

Notice that we used the product rule to differentiate the denominator. In any case, we are again in $0/0$ territory|just put $x = 0$, and see that the top and bottom both become 0. So we use L'Hopital's Rule (and the product rule) once more:

$$\lim_{x\to 0} \frac{1 - \cos x}{\sin x + x \cos x} = \lim_{x\to 0} \frac{\sin x}{\cos x + \cos x - x \sin x}$$

Don't use L'Hopital's Rule again! At this stage, just put $x = 0$; the numerator is 0 and the denominator is 2, so the overall limit is 0. Putting everything together, we have shown that

$$\lim_{x\to 0} \left(\frac{1}{\sin x} - \frac{1}{x} \right) = 0$$

Taking a common denominator doesn't always work. Sometimes you might not even have a denominator at all, so you have to create one out of thin air (See Exercises book).

15.5 Type $B2$: $(0 \times \pm\infty)$

Here's a limit of this type:

$$\lim_{x \to 0^+} x \ln x$$

The limit has to be as $x \to 0^+$ since $\ln(x)$ isn't even defined when $x \leq 0$. Now, as $x \to 0^+$, we see that $x \to 0$ while $\ln x \to -\infty$, so we are dealing with the indeterminate form $0 \times -\infty$. Let's turn the limit into Type A by manufacturing a denominator. The idea is to move x into a new denominator by putting it there as $1/x$:

$$\lim_{x \to 0^+} x \ln x = \lim_{x \to 0^+} \frac{\ln x}{\frac{1}{x}}.$$

Now the form is $-\infty/\infty$, so we can use L'Hopital's Rule:

$$\lim_{x \to 0^+} x \ln x = \lim_{x \to 0^+} \frac{\ln x}{\frac{1}{x}} = \lim_{x \to 0^+} \frac{\frac{1}{x}}{-\frac{1}{x^2}}.$$

We can simplify the fraction on the right to $-x$, so that the overall limit is

$$\lim_{x \to 0^+} (-x) = 0.$$

We've solved the problem, but let's just check out something: why did we move x into the denominator and not $\ln x$? It's true that

$$\begin{aligned}
\lim_{x \to 0^+} x \ln x &= \lim_{x \to 0^+} \frac{x}{1/\ln x} \\
&= \lim_{x \to 0^+} \frac{1}{(1/x)\,(-1/(\ln x)^2)} \quad \text{(by L'Hopital's Rule)} \\
&= \lim_{x \to 0^+} -x(\ln x)^2.
\end{aligned}$$

This is actually worse than the original limit! So, take care when you choose which factor to move down the bottom. As you can see from the above example, moving a log term can be a bad idea–so avoid doing that.

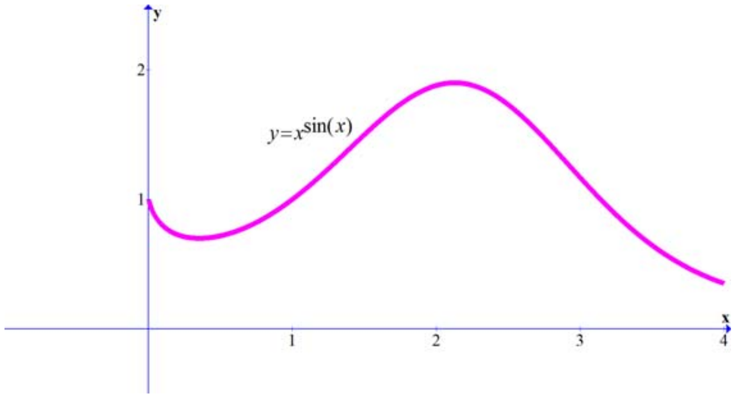


Fig. 15.2. The graph of $x^{\sin(x)}$.

15.6 Type C: $\left(1^{\pm\infty}, \quad 0^0, \quad \text{or} \quad \infty^0\right)$

Finally, the trickiest type involves limits like

$$\lim_{x \to 0^+} x^{\sin(x)}$$

where both the base and exponent involve the dummy variable (x in this case).

The corresponding graph is presented in Fig. 15.2. If you just put $x = 0$, you get 0^0, which is another indeterminate form. To find the limit, we'll use a technique very similar to logarithmic differentiation . The idea is to take the logarithm of the quantity $x^{\sin(x)}$ first, and work out its limit as $x \to 0^+$

$$\lim_{x \to 0^+} \ln\left(x^{\sin(x)}\right)$$

By our log rules, the exponent $\sin(x)$ comes down out front of the logarithm

$$\lim_{x \to 0^+} \ln\left(x^{\sin(x)}\right) = \lim_{x \to 0^+} \sin(x) \ln x.$$

As $x \to 0^+$, we have $\sin(x) \to 0$ and $\ln x \to -\infty$, so now we're dealing with a Type $B2$ problem. We can put the $\sin(x)$ into a new denominator as $1/\sin(x)$, which is just $\csc(x)$, then use L'Hopital's Rule on the resulting Type A problem:

$$\begin{aligned} \lim_{x \to 0^+} \sin(x) \ln x &= \lim_{x \to 0^+} \frac{\ln x}{\csc(x)} \\ &= \lim_{x \to 0^+} \frac{1/x}{-\csc(x)\cot(x).} \quad \text{(by L'Hopital's Rule)} \end{aligned}$$

This can be rearranged to

$$\lim_{x\to 0^+} -\frac{\sin(x)}{x} \times \tan(x) = -1 \times 0 = 0$$

Are we done? Not quite. We now know that

$$\lim_{x\to 0^+} \ln\left(x^{\sin(x)}\right) = 0;$$

so now we just have to exponentiate both sides to see that

$$\lim_{x\to 0^+} x^{\sin(x)} = 1.$$

(The exponentiation works because e^x is a continuous function of x.)

Let's review what we just did. Instead of finding the original limit, we took logarithms and then found that limit, using the Type $B2$ technique. Finally, we exponentiated at the end. In fact, sometimes you don't even have to go through the Type $B2$ step on your way to Type A. Here ia our old friend:

Example 15.10 *Evaluate*

$$\lim_{x\to\infty} \left(1 + \frac{1}{x}\right)^x$$

using L'Hopital's Rule.

Solution: Because direct substitution yields the indeterminate form 1^∞ you can proceed as follows. To begin, assume that the limit exists and is equal to y.

$$y = \lim_{x\to\infty} \left(1 + \frac{1}{x}\right)^x$$

Taking the natural logarithm of each side produces

$$\begin{aligned}
\ln y &= \ln\left[\lim_{x\to\infty} \left(1 + \frac{1}{x}\right)^x\right] \\
&= \lim_{x\to\infty} \left[\frac{\ln(1 + 1/x)}{1/x}\right] \quad \text{indeterminate form } 0/0 \\
&= \lim_{x\to\infty} \left[\frac{\left(-1/x^2\right)\left(1/\left(1 + 1/x\right)\right)}{-1/x^2}\right] \quad \text{L'Hopital's Rule} \\
&= \lim_{x\to\infty} \frac{1}{1/\left(1 + 1/x\right)} = 1
\end{aligned}$$

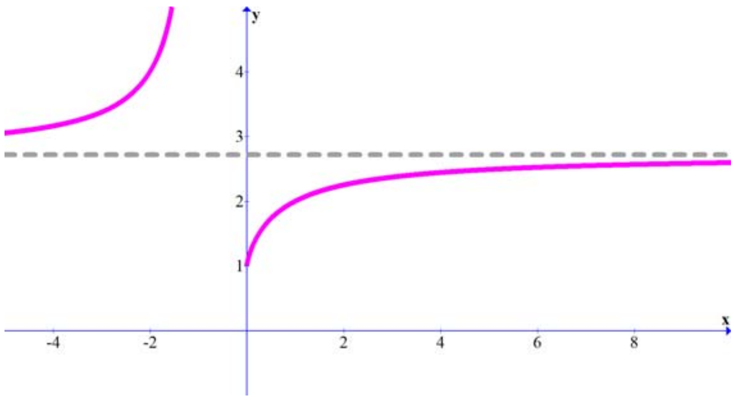


Fig. 15.3. The limit of $\left(1+\frac{1}{x}\right)^x$ as x approaches infinity is e. The constant function $y = e$ is a horizontal asymptote.

Now, because you have shown that $\ln y = 1$ you can conclude that and obtain

$$\lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x = e$$

(see Figure 15.3.) □

Check, that $\lim_{x\to 0^+}\left(1+\frac{1}{x}\right)^x = 1$.

15.7 Summary of L'Hopital's Rule types

Here are all the techniques we've looked at:

- Type A: if the limit involves a fraction, like

$$\lim_{x\to c}\frac{f(x)}{g(x)}$$

check that the form is indeterminate. It must be $0/0$ or $\pm\infty/\pm\infty$. Use the rule

$$\lim_{x\to c}\frac{f(x)}{g(x)} = \lim_{x\to c}\frac{f'(x)}{g'(x)}$$

Do not use the quotient rule here! Now, solve the new limit, perhaps even using L'Hopital's Rule again.

- Type $B1$: if the limit involves a difference, like

$$\lim_{x\to c}\left(f(x)-g(x)\right),$$

where the form is $\pm(\infty - \infty)$, try taking a common denominator or multiplying by a conjugate expression to reduce to a Type A form.

- Type $B2$: if the limit involves a product, like

$$\lim_{x \to c} f(x)g(x)$$

where the form is $0 \times \pm\infty$, pick the simplest of the two factors and put it on the bottom as its reciprocal. (Avoid picking a log term–keep that on the top.) You get something like

$$\lim_{x \to c} f(x)g(x) = \lim_{x \to c} \frac{g(x)}{\frac{1}{f(x)}}$$

This is now a Type A form.

- Type C: if the limit involves an exponential where both base and exponent involve the dummy variable, like

$$\lim_{x \to c} f(x)^{g(x)},$$

then first work out the limit of the logarithm:

$$\lim_{x \to c} \ln\left(f(x)^{g(x)}\right) = \lim_{x \to c} g(x) \ln f(x).$$

This should be either Type $B2$ or Type A (or else it's not indeterminate and you can just substitute). Once you've solved it, you can rewrite the equation as something like

$$\lim_{x \to c} \ln\left(f(x)^{g(x)}\right) = L$$

then exponentiate both sides to get

$$\lim_{x \to c} f(x)^{g(x)} = e^L.$$

Now all that's left is for you to practice doing as many l'H^opital's Rule problems as you can get your hands on!

16
Applications of integration

In Chapter 10, we calculated areas using definite integrals. We obtained the integral by slicing up the region, constructing a Riemann sum, and then taking a limit. In this section, we will discuss how integration allows us to calculate the volume of solid regions for which we're given the area of cross sections which are perpendicular to some axis. We will look at many examples in which the solids are solids of revolution, solids obtained by revolving a plane region around an axis and looking at the solid region which is "swept out". In the context of solids of revolution, integrating the cross-sectional area is referred to as the *disk method* or *washer method*, due to the shapes of the cross sections.

We will also look at a second method for finding the volumes of solids of revolution; this method has some aspects that are similar to the disk and washer methods, but it does not use planar cross sections of the solid. This second method for finding volumes of solids of revolution is the *cylindrical shell method*. Solids of revolution are used commonly in engineering and manufacturing. Some examples are axles, funnels, pills, bottles, pistons etc.

16.1 Volume by parallel cross sections

Figure 16.1 shows a plane region Ω and a solid formed by translating Ω along a line perpendicular to the plane of Ω. Such a solid is called a right cylinder with cross section Ω.

If Ω has area A and the solid has height h, then the volume of the solid is a simple product:

$$V = A \cdot h$$

To calculate the volume of a more general solid, we introduce a coordinate axis and then examine the cross sections of the solid that are perpendicular to that axis. In Figure 16.2 we depict a solid and a coordinate axis that we label the x-axis. As in the figure, we suppose that the solid lies entirely between $x = a$ and $x = b$. The figure shows an arbitrary cross section perpendicular to the x-axis. By $A(x)$ we mean the area of the cross section at coordinate x.

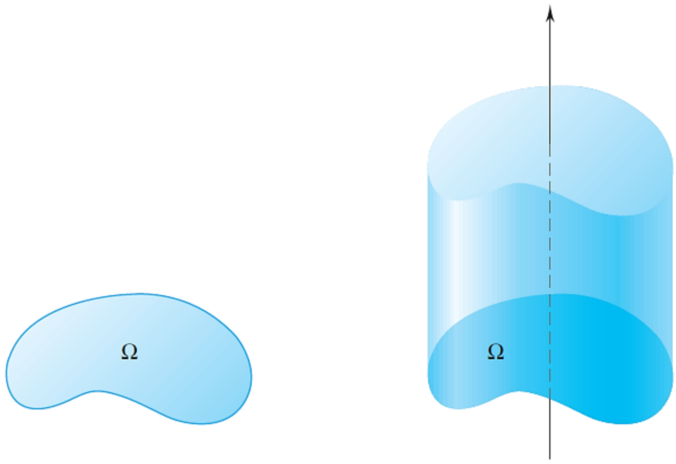


Fig. 16.1. Right cylinder with cross section Ω.

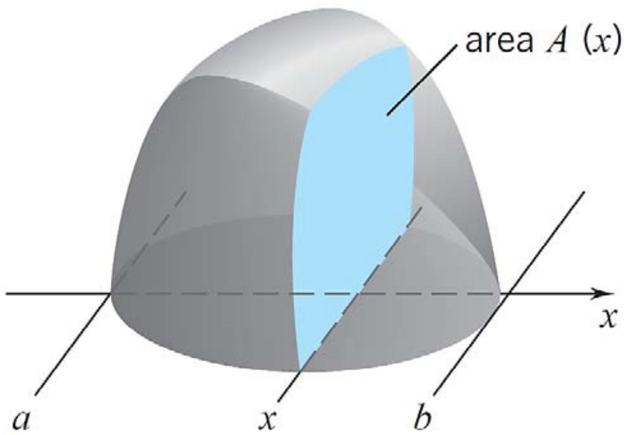


Fig. 16.2. An arbitrary cross section with area $A(x)$, perpendicular to the x-axis.

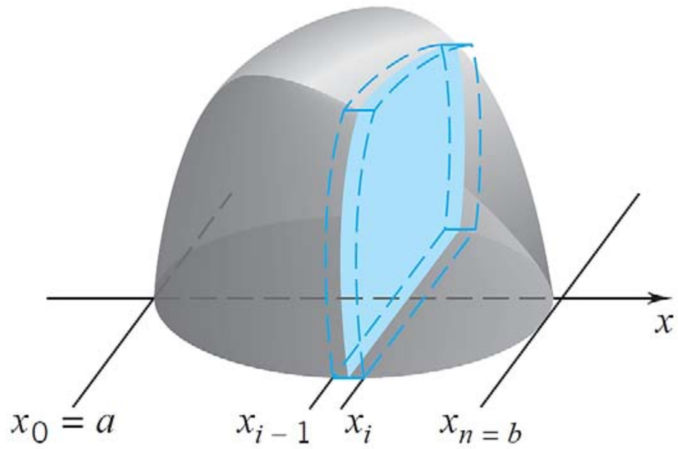


Fig. 16.3. A slab of cross-sectional area $A(x_i^*)$ and thickness Δx_i .

If the cross-sectional area $A(x)$ varies continuously with x, then we can find the volume V of the solid by integrating $A(x)$ from $x = a$ to $x = b$:

$$V = \int_a^b A(x)dx \tag{16.1}$$

Derivation of the formula. Let $P = \{x_0, x_1, x_2, ..., x_n\}$ be a partition of $[a, b]$. On each subinterval $[x_{i-1}, x_i]$ choose a point x_i^*. The solid from x_{i-1} to x_i can be approximated by a slab of cross-sectional area $A(x_i^*)$ and thickness Δx_i (Figure 16.3). The volume of this slab is the product

$$A(x_i^*)\Delta x_i \tag{16.2}$$

The sum of these products,

$$A(x_1^*)\Delta x_1 + A(x_2^*)\Delta x_2 + ... + A(x_n^*)\Delta x_n$$

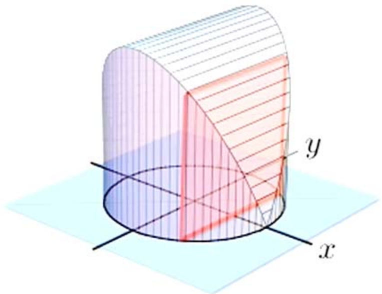
is a Riemann sum which approximates the volume of the entire solid. As $\Delta x = \max\{\Delta x_i : 1 \leq i \leq n\}$ tends to zero such Riemann sums converge to

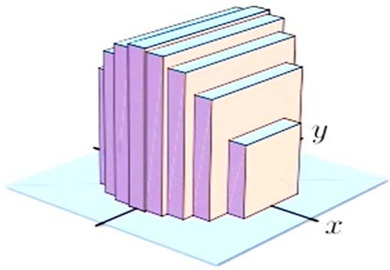
$$\int_a^b A(x)dx$$

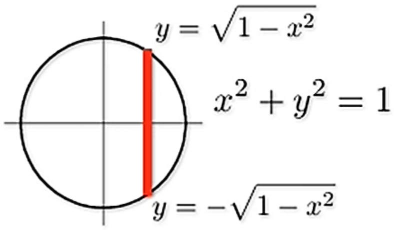
which is called the **volume of the solid**.

Example 16.1 *A solid's base is the unit disc in the xy-plane, and its vertical cross-sections parallel to the y-axis are squares. Find the volume of the solid.*

Solution:







Area of the cross-section: $A(x) = \left(2\sqrt{1-x^2}\right)^2 = 4\left(1-x^2\right)$

$$\begin{aligned} V &= \int_{-1}^{1} A(x)dx = \int_{-1}^{1} 4(1-x^2)dx \\ &= 2\int_{0}^{1} 4(1-x^2)dx = 8(x - \frac{1}{3}x^3)|_0^1 \\ &= 8\left(\frac{2}{3} - 0\right) = \frac{16}{3} \end{aligned}$$

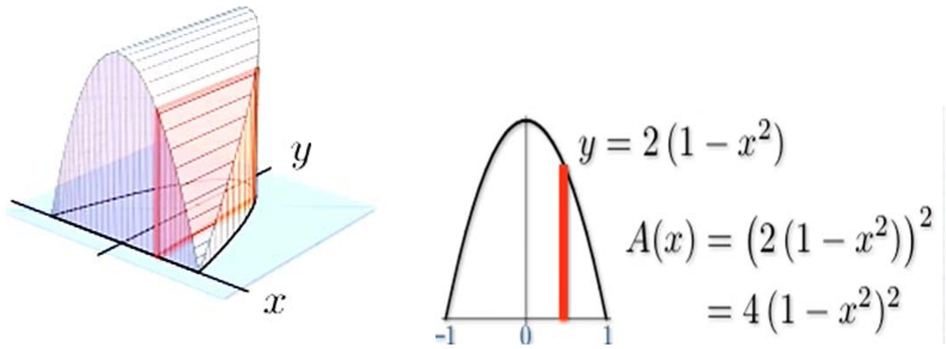
□

Example 16.2 *A solid whose base is the planar region in which*

$$0 \leq y \leq 2(1-x^2)$$

has square vertical cross-sections parallel to the y-axis. Find the volume of the solid.

Solution:



$$\begin{aligned} V &= \int_{-1}^{1} A(x)dx = \int_{-1}^{1} 4(1-x^2)^2 dx \\ &= 8\int_{0}^{1} (1-x^2)^2 dx \\ &= 8\int_{0}^{1} (x^4 - 2x^2 + 1)dx \\ &= 8(\frac{1}{5}x^5 - \frac{2}{3}x^3 + x)|_0^1 = \frac{64}{15} \end{aligned}$$

□

Remark 16.3 *$dV = A(x)dx$ is called the volume differential. Think of this as the volume of a typical "slice" with the thickness dx.*

Example 16.4 (Cavalieri's principle) *Cavalieri's principle says that solids with equal altitudes and identical cross-sectional areas at each height have the same volume (Figure 6.7). This follows immediately from the definition of volume, because the cross-sectional area function $A(x)$ and the interval $[a, b]$ are the same for both solids (see Figure 16.4)*

□

16.2 The disc method

If a region in the plane is revolved about a line, the resulting solid is a *solid of revolution*, and the line is called the *axis of revolution*. The simplest such solid is a right circular cylinder or *disk*, which is formed by revolving a rectangle

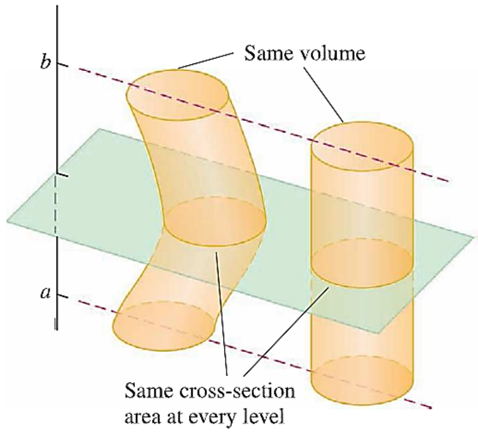


Fig. 16.4. *Cavalieri's principle:* These solids have the same volume, which can be illustrated with stacks of coins.

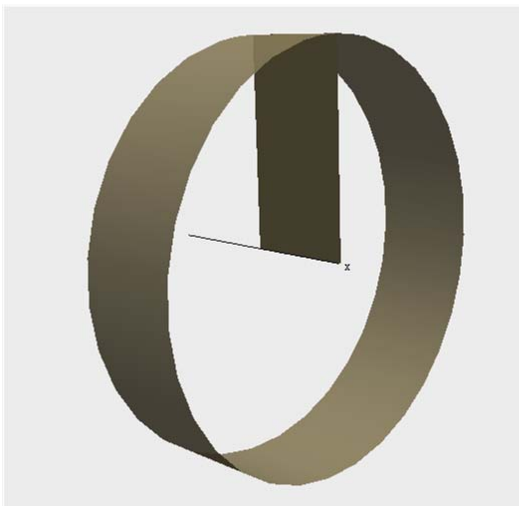


Fig. 16.5. Volume of a disc: $\pi R^2 w$

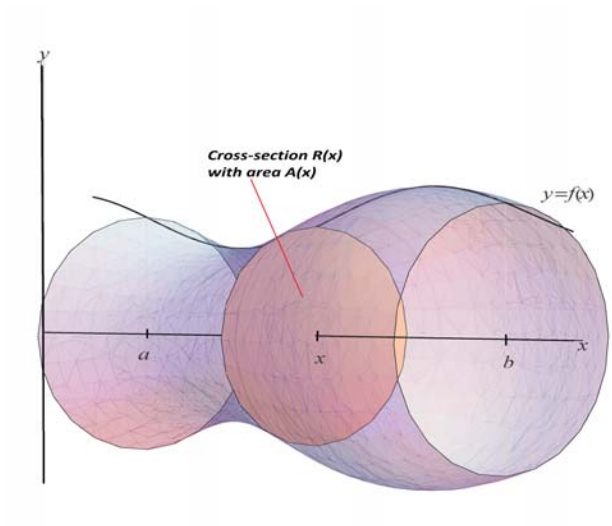


Fig. 16.6. A typical x cross section of a solid

about an axis adjacent to one side of the rectangle, as shown in Figure 16.5. The volume of such a discis

$$\text{Volume of disk} = (\text{area of disk}) \times (\text{width of disk}) = \pi R^2 w$$

where R is the radius of the disk and w is the width. We can suppose that $R(x) \geq 0$ for $a \leq x \leq b$.

To see how to use the volume of a disk to find the volume of a general solid of revolution, consider a solid of revolution formed by revolving the plane region in Figure 16.9 about the indicated axis. To determine the volume of this solid, consider a representative rectangle in the plane region with a base Δx_i. When this rectangle is revolved about the axis of revolution (see Fig. 16.10), it generates a representative disk whose volume is

$$\Delta V = \pi R^2 \Delta x_i.$$

Approximating the volume of the solid by such disks of width Δx_i and radius $R(x_i)$ produces

$$\text{Volume of solid} \approx \sum_{i=1}^{n} \pi R(x_i)^2 \Delta x_i = \pi \sum_{i=1}^{n} R(x_i)^2 \Delta x$$

Here each x_i belongs to the base of the corresponding revolved rectangle.

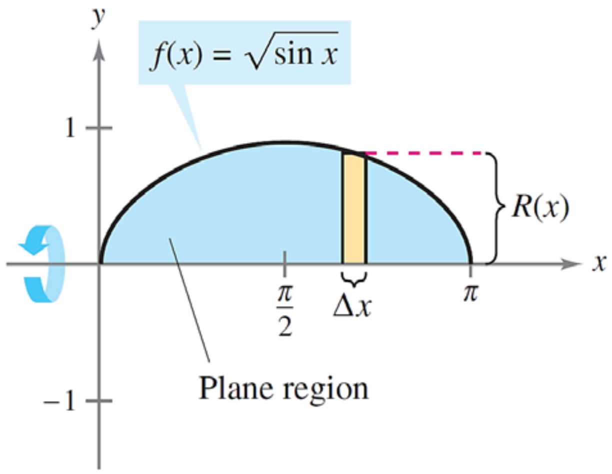


Fig. 16.7. Plane region (Example 16.5).

This approximation appears to become better and better as $\Delta x = \max\{\Delta x_i : 1 \le i \le n\} \to 0$ when $n \to \infty$. So, you can define the volume of the solid as

$$\textit{Volume of solid} = \lim_{\Delta x \to 0} \pi \sum_{i=1}^{n} R^2(x_i)\Delta x_i = \pi \int_a^b R^2(x)dx \tag{16.3}$$

Example 16.5 *Using the Disk Method find the volume of the solid formed by revolving the region bounded by the graph of*

$$f(x) = \sqrt{\sin x}$$

and the x-axis $(0 \le x \le \pi)$ *about the x-axis.*

Solution: From the representative rectangle in the upper graph in Figure 16.7, you can see that the radius of this solid is

$$R(x) = f(x) = \sqrt{\sin x}.$$

So, the volume of the solid of revolution is

$$\begin{aligned} V &= \pi \int_a^b [R(x)]^2\, dx = \pi \int_0^\pi \left[\sqrt{\sin x}\right]^2 dx \\ &= \pi \int_0^\pi \sin x dx = \pi\left[-\cos x\right]_0^\pi \\ &= \pi(1+1) = 2\pi. \end{aligned}$$

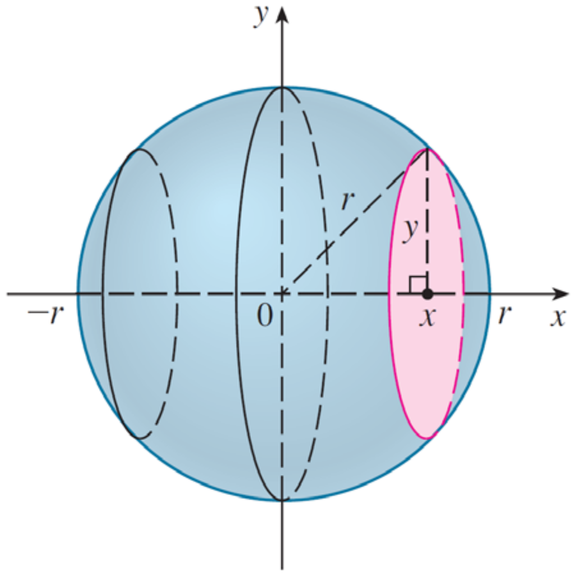
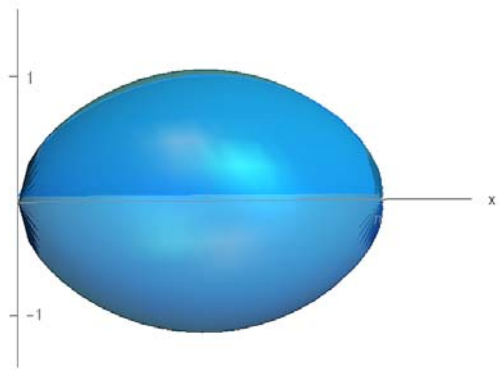
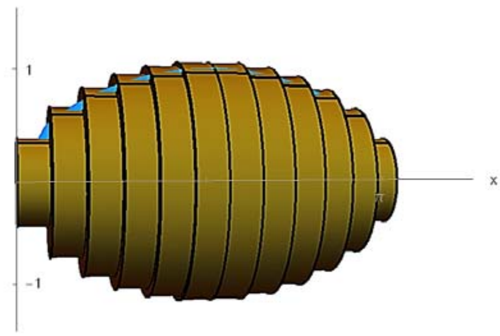


Fig. 16.8. Sphere of radius r.



Solid of revolution (Example 16.5)



Approximation by n disks

Example 16.6 *In this example we will use the disc method to show, that the volume of a sphere of radius* r *is*

$$V = \frac{4}{3}\pi r^3.$$

If we place the sphere so that its center is at the origin (see Figure 16.8), then the plane P_x *intersects the sphere in a circle whose radius (from the Pythagorean Theorem) is* $y = \sqrt{r^2 - x^2}$. *So the cross-sectional area is* $A(x) = \pi y^2 = \pi(r^2 - x^2)$. *Using the definition of volume with* $a = -r$ *and* $b = r$, *we*

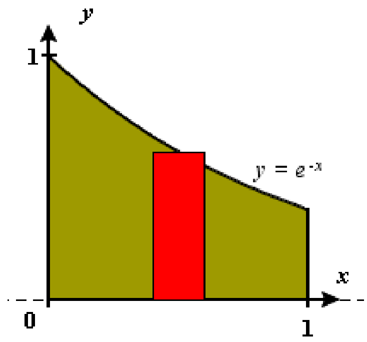


Fig. 16.9. Representative rectangle of width Δx

have

$$
\begin{aligned}
V &= \int_{-r}^{r} A(x)dx = \int_{-r}^{r} \pi \left(r^2 - x^2\right) dx \\
&= 2\pi \int_{0}^{r} \left(r^2 - x^2\right) dx \\
&= 2\pi \left[r^2 x - \frac{1}{3}x^3\right]_0^r \\
&= \frac{4}{3}\pi r^3
\end{aligned}
$$

□

Example 16.7 *This example illustrates a technique for calculating the volume of a solid of revolution. In particular, the solid we consider is formed by revolving the curve $y = e^{-x}$ from $x = 0$ to $x = 1$ about the x-axis. To find the volume of this solid, we first divide the region in the xy-plane into n thin vertical strips (rectangles) of thickness Δx ($\Delta x = 1/5$ in Fig. 16.9). As each rectangle is rotated about the x-axis, it forms a slice that looks like a circular disk (Fig. 16.10). It is easy to write the formula for the approximate radius of this disk in terms of x. (Here $R = e^{-x}$.) We can use this radius to state a formula for the approximate volume of this slice (where each x_i belongs to the base of the corresponding revolved rectangle)*

$$\text{Volume of the slice} \approx \pi y^2 \Delta x = \pi (e^{-x})^2 \Delta x.$$

We can approximate the total volume of the solid by adding up the volumes of a finite number of these slices (disks), so (Fig. 16.11)

$$\text{Total volume} \approx \sum \pi y^2 \Delta x = \sum \pi (e^{-x})^2 \Delta x.$$

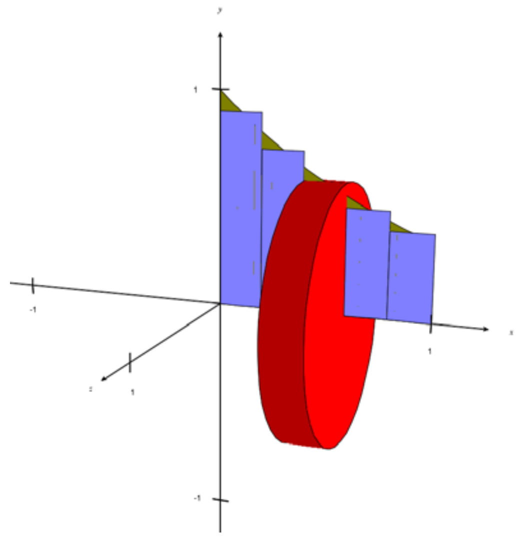


Fig. 16.10. A thin strip rotated around the x-axis to form a circular slice.

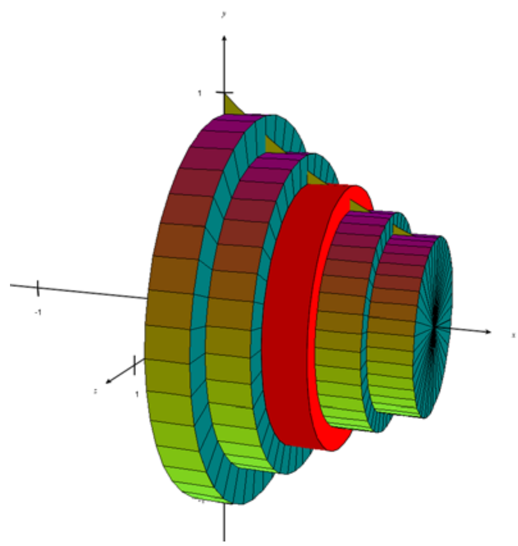


Fig. 16.11. Total volume of the solid is approximated by adding up the volumes of a finite number of slices .

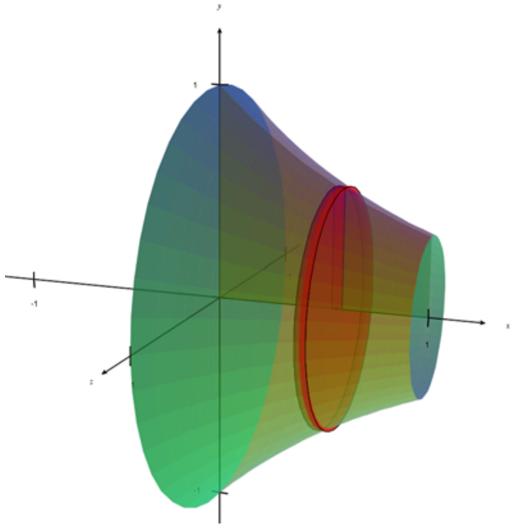


Fig. 16.12. As $\Delta x \to 0$ we get the exact volume of the solid using an integral.

Particularly, for different values of n we have (in this example)

n	Δx	*Volume of n slices (in units^3)*
5	1/5	1.349 20
10	1/10	1.355 95
30	1/30	1.357 90

As the thickness of the slices, Δx, tends to zero (Fig.16.12), we get the exact volume of the solid using an integral.

$$\begin{aligned}\text{Total volume} &\approx \int_0^1 \pi(e^{-x})^2 dx = \pi \int_0^1 (e^{-x})^2 dx = \pi \left(-\frac{1}{2}\right) e^{-2x}|_0^1 \\ &= \pi \left(-\frac{1}{2}\right)\left(e^{-2} - e^0\right) = \frac{\pi}{2}\left(1 - e^{-2}\right) = 1.358\,212 \quad (units^3).\end{aligned}$$

□

Example 16.8 (Revolving About a Line That Is't a Coordinate Axis)
Find the volume of the solid formed by revolving the region bounded by

$$f(x) = 2 - x^2$$

and $g(x) = 1$ about the line $y = 1$, as shown in Figure 16.13.

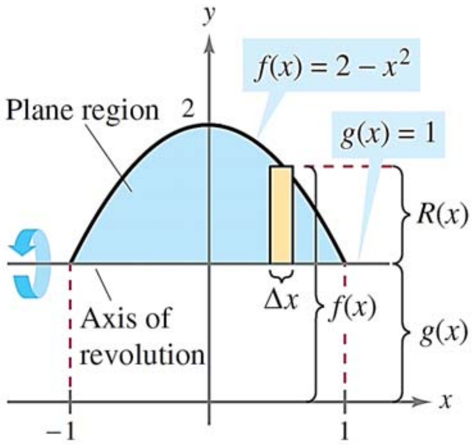


Fig. 16.13. Plane region (Example 16.8)

Solution: By equating $f(x)$ and $g(x)$ you can determine that the two graphs intersect when $x = \pm 1$. To find the radius, subtract $g(x)$ from $f(x)$

$$\begin{aligned} R(x) &= f(x) - g(x) \\ &= (2 - x^2) - 1 \\ &= 1 - x^2 \end{aligned}$$

Finally, integrate between -1 and 1 to find the volume

$$\begin{aligned} V &= \pi \int_a^b [R(x)]^2\, dx \\ &= \pi \int_{-1}^{1} (1 - x^2)^2 dx \\ &= \pi \int_{-1}^{1} (x^4 - 2x^2 + 1) dx \\ &= \pi \left[\frac{1}{5} x^5 - \frac{2}{3} x^3 + x \right]_{-1}^{1} \\ &= \frac{16\pi}{15} \end{aligned}$$

□

Note that you can determine the variable of integration by placing a representative rectangle in the plane region "perpendicular" to the axis of revolution. If the width of the rectangle is Δx integrate with respect to x and if the

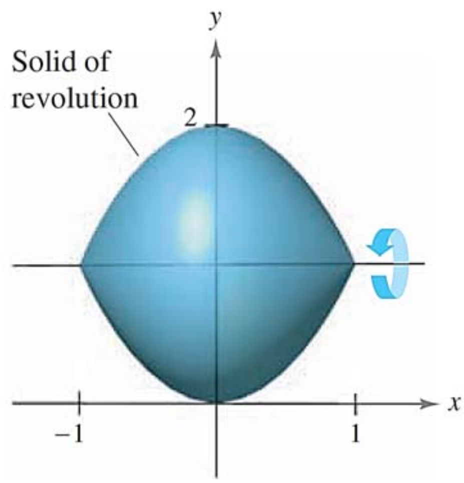


Fig. 16.14. Solid of revolution (Example 16.8)

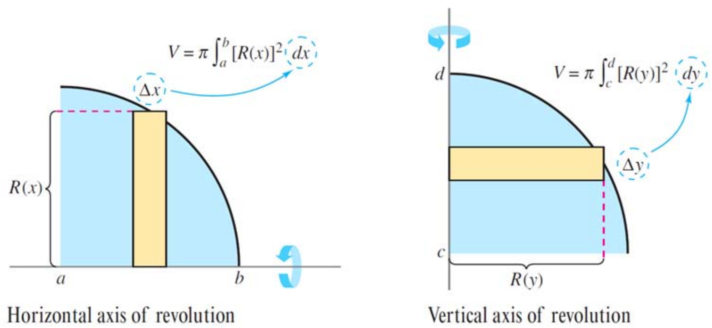


Fig. 16.15. Horizontal versus vertical axis of revolution.

width of the rectangle is Δy integrate with respect to y. To find the volume of a solid of revolution with the disk method, use one of the following, as shown in Figure 16.15.

Example 16.9 *In this example we will find the volume of the solid obtained by rotating the region bounded by* $y = x^3$*,* $y = 8$*, and* $x = 0$ *about the* y*-axis. The region is shown in Figure 16.16(a) and the resulting solid is shown in Figure 16.16(b). Because the region is rotated about the* y*-axis, it makes sense to slice the solid perpendicular to the* y*-axis and therefore to integrate*

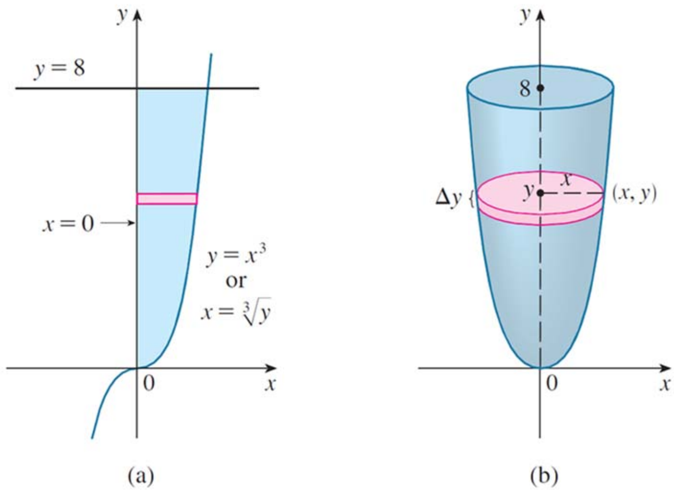


Fig. 16.16. Solid obtained by rotating the region bounded by $y = x^3$, $y = 8$, and $x = 0$ about the y-axis.

with respect to y. If we slice at height y, we get a circular disk with radius x, where $x = \sqrt[3]{y}$. So the area of a cross-section through y is

$$A(y) = \pi x^2 = \pi \left(\sqrt[3]{y}\right)^2 = \pi y^{2/3}$$

and the volume of the approximating cylinder pictured in Figure 16.16(b) is

$$A(y)\Delta y = \pi y^{2/3} \Delta y.$$

Since the solid lies between $y = 0$ and $y = 8$, its volume is

$$V = \int_0^8 \pi y^{2/3} dy = \left[\frac{3}{5}\pi y^{\frac{5}{3}}\right]_0^8 = \frac{96}{5}\pi.$$

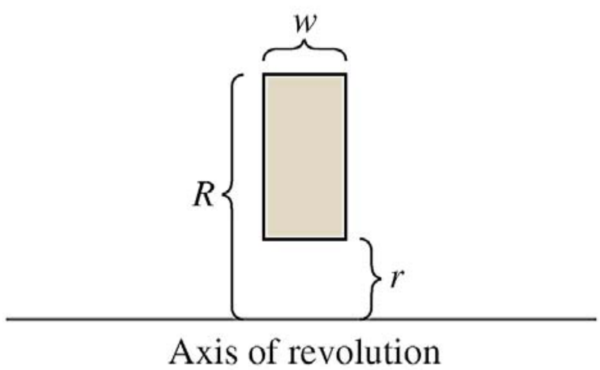
□

16.3 The washer method

The disk method can be extended to cover solids of revolution with holes by replacing the representative disk with a representative washer. The washer is

Fig. 16.17. Revolving rectangle

formed by revolving a rectangle



about an axis, as shown in Figure 16.17. If and are the inner r and outer radii R of the washer and w is the width of the washer, the volume is given by

$$\pi \left(R^2 - r^2\right) w. \tag{16.4}$$

To see how this concept can be used to find the volume of a solid of revolution, consider a region bounded by an *outer radius* $R(x)$ and an *inner radius* $r(x)$ as shown in Figure 16.18. If the region is revolved about its axis of revolution,

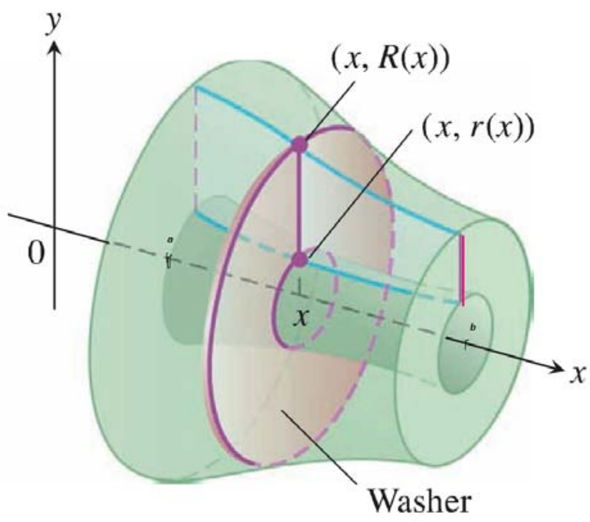


Fig. 16.18. Representative element in the washer method

the volume of the resulting solid is given by

$$V = \pi \int_a^b \left(R^2(x) - r^2(x)\right) dx$$

Note that the integral involving the inner radius represents the volume of the hole and is subtracted from the integral involving the outer radius.

Example 16.10 *In this another example where we are visualizing the process of finding the volume of a solid of revolution. It illustrates a technique for calculating the volume of a solid of revolution. In this example, we consider the solid that is formed by revolving the region bounded by the curves $y = x$ and $y = x^2$ about the line $y = 3$. To find the volume of this solid, we first divide the region in the xy-plane into n thin vertical strips (rectangles) of thickness Δx. As each rectangle (see Fig. 16.19)is rotated about the line $y = 3$, it forms a disk with a hole in it (Fig. 16.20). This disk with a hole in it has an approximate inner radius of $r_{inner}(x) = 3 - x$ and it has an approximate outer radius of $r_{outer}(x) = 3 - x^2$. Think of the slice as a circular disc of radius r_{outer} from which has been removed a smaller disc of radius r_{inner}. Using these radii we can generate a formula for the approximate volume of this disk with a hole in it (slice).*

$$V = \pi r_{outer}^2 \Delta x - \pi r_{inner}^2 \Delta x.$$

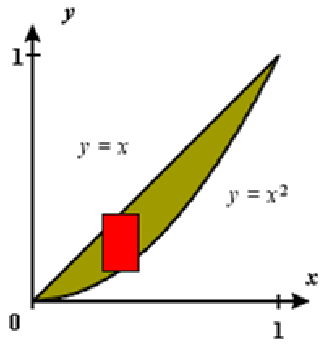


Fig. 16.19. Region in the xy-plane into $n = 7$ thin vertical strips

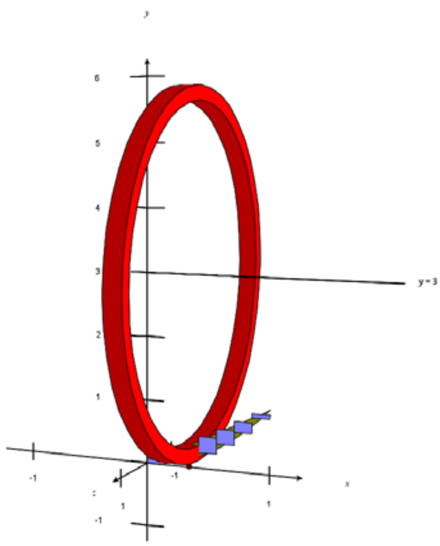


Fig. 16.20. Representative rectangle rotated about the line $y = 3$ forms a disk with a hole in it.

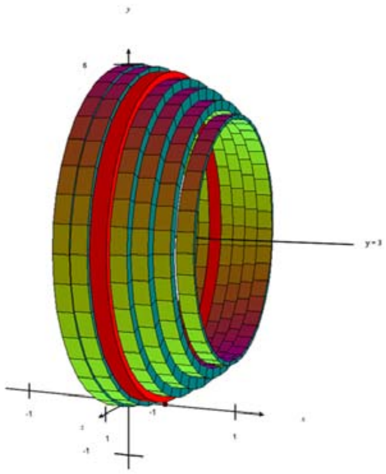


Fig. 16.21. We can approximate the total volume of the solid by adding up the volumes of a finite number of slices.

or

$$V = \pi(3-x^2)^2\Delta x - \pi(3-x)^2\Delta x = \pi[(3-x^2)^2 - (3-x)^2]\Delta x.$$

We can approximate the total volume of the solid by adding up the volumes of a finite number of these slices (Fig. 16.21)

$$\text{Total volume} \approx \sum \left(\pi r_{outer}^2 \Delta x - \pi r_{inner}^2 \Delta x\right) = \sum \pi[(3-x^2)^2 - (3-x)^2]\Delta x$$

Particularly, for different values of n we have (in this example)

n	Δx	*Volume of n slices (in units^3)*
5	1/5	2.775 22
10	1/10	2.733 58
30	1/30	2.724 68

As the thickness of the slices, Δx, tends to zero, we get the exact volume of the solid using an integral.Since the curves $y = x$ and $y = x^2$ intersect at $x = 0$ and $x = 1$, these are the limits of integration:

$$\begin{aligned} V &= \int_0^1 \pi[(3-x^2)^2 - (3-x)^2]dx = \pi \int_0^1 [(3-x^2)^2 - (3-x)^2]dx \\ &= \pi \int_0^1 \left(\left(x^4 - 6x^2 + 9\right) - \left(x^2 - 6x + 9\right)\right) dx \\ &= \pi \int_0^1 \left(x^4 - 7x^2 + 6x\right) dx = \frac{13}{15}\pi = 2.722\,714. \end{aligned}$$

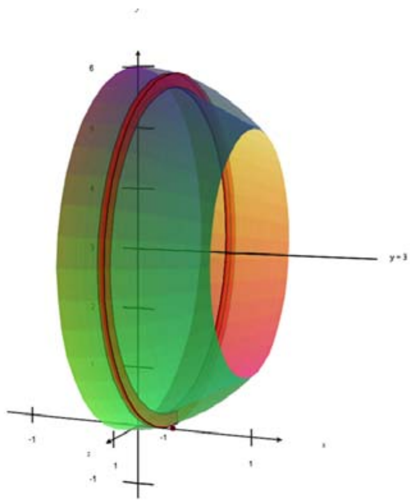


Fig. 16.22. As the thickness of the slices, Δx, tends to zero, we get the exact volume of the solid using an integral.

$\square$.

Example 16.11 *The region between the graphs of* $\sin x$ *and* x *on* $[0, \pi/2]$ *is revolved about the* x *axis. Sketch the resulting solid and find its volume.*

Solution: The solid is sketched in Fig. 16.23. It has the form of a hollowed-out cone.The volume is that of the cone minus that of the hole. The cone is obtained by revolving the region under the graph of x on $[0, \pi/2]$ about the axis, so its volume is

$$\pi \int_0^{\pi/2} x^2 dx = \frac{1}{24}\pi^4.$$

The hole is obtained by revolving the region under the graph of $\sin x$ on $[0, \pi/2]$ about the x axis, so its volume is

$$\begin{aligned} \pi \int_0^{\pi/2} \sin^2 x dx &= \pi \int_0^{\pi/2} \frac{1 - \cos 2x}{2} dx \qquad \text{since } \cos 2x = 1 - 2\sin^2 x \\ &= \pi \left(\frac{x}{2} - \frac{1}{4}\sin 2x \right) |_0^{\pi/2} \\ &= \frac{1}{4}\pi^2 \end{aligned}$$

Thus the volume of our solid is $V = \frac{1}{24}\pi^4 - \frac{1}{4}\pi^2 = 1.5913.$ $\square$

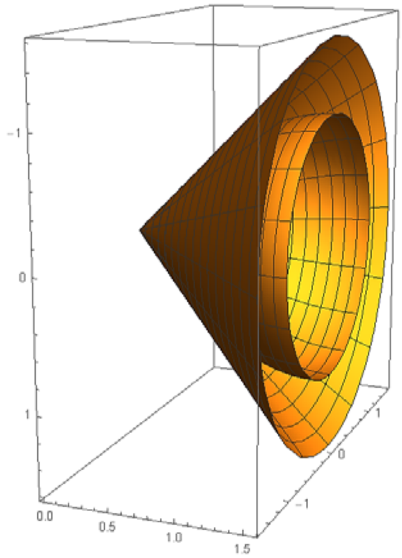
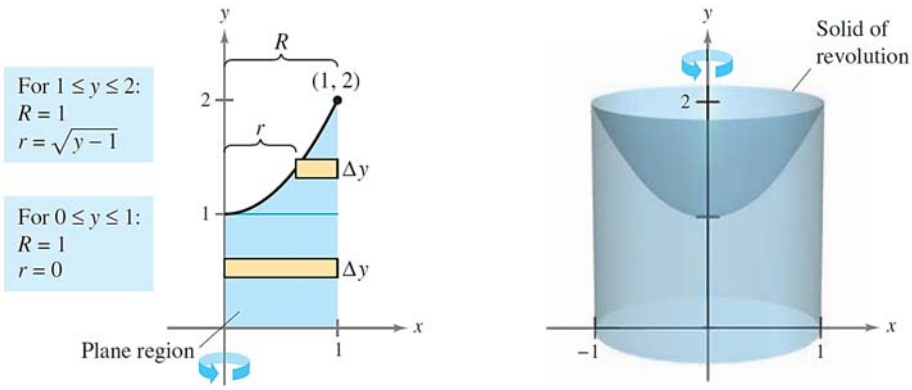


Fig. 16.23. The region between the graphs of $\sin x$ and x is revolved about the x axis.

Example 16.12 (Integrating with Respect to y, Two-Integral Case) *Find the volume of the solid formed by revolving the region bounded by the graphs of $y = x^2 + 1$, $y = 0$, $x = 0$, and $x - 1$ and about the y-axis, as shown in Figure below*



Solution: For the region shown in Figure above, the outer radius is simply $R = 1$, There is, however, no convenient formula that represents the inner radius. When $0 \leq y \leq 1$, $r = 0$ but when $1 \leq y \leq 2$, r is determined by the equation $y = x^2 + 1$ which implies that $r = \sqrt{y-1}$

$$r(y) = \begin{cases} 0, & 0 \le y \le 1 \\ \sqrt{y-1}, & 1 \le y \le 2 \end{cases}$$

Using this definition of the inner radius, you can use two integrals to find the volume

$$\begin{aligned} V &= \pi \int_0^1 (1^2 - 0^2)\, dy + \pi \int_1^2 \left(1^2 - \left(\sqrt{y-1}\right)^2\right) dy \\ &= \pi \int_0^1 1 dy + \pi \int_1^2 (2-y)\, dy \\ &= \pi y|_0^1 + \pi \left(2y - \frac{y^2}{2}\right)|_1^2 \\ &= \pi + \pi \left(4 - 2 - 2 + \frac{1}{2}\right) = \frac{3}{2}\pi \end{aligned}$$

□

16.4 The method of cylindrical shells

In this section, we will study an alternative method for finding the volume of a solid of revolution. This method is called the *shell method* because it uses cylindrical shells. A comparison of the advantages of the disk and shell methods is given later in this section.

To begin, consider a representative (brown) rectangle as shown in Figure 16.24, where w is the width of the rectangle, h is the height of the rectangle, and p is the distance between the axis of revolution and the center of the rectangle. When this rectangle is revolved about its axis of revolution, it forms a cylindrical shell (or tube) of thickness w.

To find the volume of this shell, consider two cylinders. The radius of the larger cylinder corresponds to the outer radius of the shell, and the radius of the smaller cylinder corresponds to the inner radius of the shell. Because p is the average radius of the shell, you know the outer radius is $p + (w/2)$ and the inner radius is $p + (w/2)$.

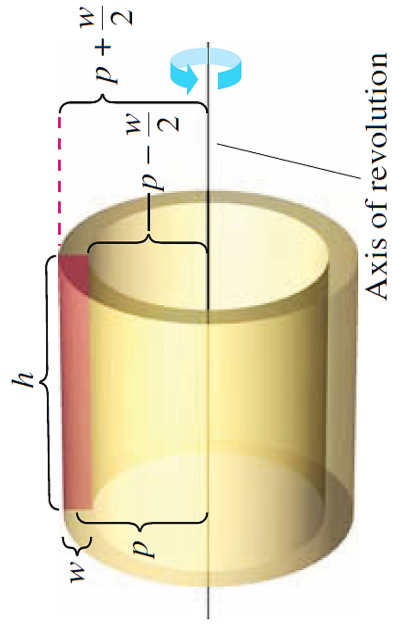


Fig. 16.24. Cylindrical shell

So, the volume of the shell is

$$\begin{aligned}\text{Volume of shell} &= (\text{volume of cylinder}) - (\text{volume of hole}) \\ &= \pi\left(p + \frac{w}{2}\right)^2 h - \pi\left(p - \frac{w}{2}\right)^2 h \\ &= 2\pi hpw \\ &= 2\pi\,(\text{average radius})\,(\text{height})\,(\text{thickness})\end{aligned}$$

Now, we can use this formula to find the volume of a solid of revolution. Assume that the plane region in Figure 16.25 is revolved about a line to form the indicated solid. If you consider a vertical rectangle of width Δx then, as the plane region is revolved about a line parallel to the y-axis, the rectangle generates a representative shell whose volume is

$$\Delta V = 2\pi\,[p(x)h(x)]\,\Delta x.$$

We can approximate the volume of the solid by such shells of thickness Δx height $h(x_i)$ and average radius $p(x_i)$

$$\text{Volume of solid} \approx \sum_{i=1}^{n} 2\pi\,[p(x_i)h(x_i)]\,\Delta x = 2\pi \sum_{i=1}^{n} [p(x_i)h(x_i)]\,\Delta x$$

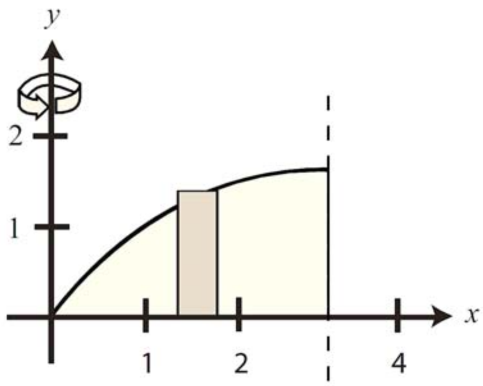


Fig. 16.25.

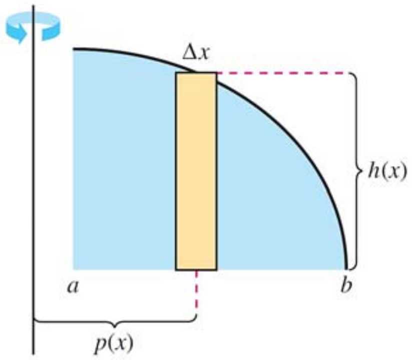
This approximation appears to become better and better as $\Delta x \to 0$ ($n \to \infty$). So, the volume of the solid is:

$$\text{Volume of solid} = \lim_{\Delta \to 0} 2\pi \sum_{i=1}^{n} [p(x_i)h(x_i)]\,\Delta x = 2\pi \int_a^b [p(x)h(x)]\,dx.$$

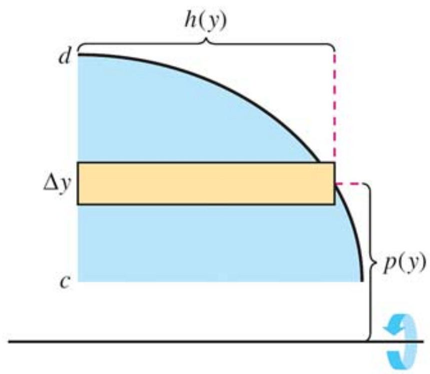
The shell method:
To find the volume of a solid of revolution with the shell method, use one of the following, as shown in Figure below.

Vertical axis of revolution
Volume=$V = 2\pi \int_a^b [p(x)h(x)]\,dx$

Horizontal axis of revolution
Volume=$V = 2\pi \int_c^d [p(y)h(y)]\,dy$



Vertical axis of revolution



Horizontal axis of revolution

Example 16.13 (Using the Shell Method) *Find the volume of the solid formed by revolving the region in the first quadrant bounded by* $y = \sqrt{x}$,

$0 \leq x \leq 3$, about the y-axis.

Solution: The volume of a representative shell is $2\pi p(x)h(x)\Delta x$, which equals $2\pi x\sqrt{x}\Delta x$ (Figure 16.25). The volume is the following definite integral:

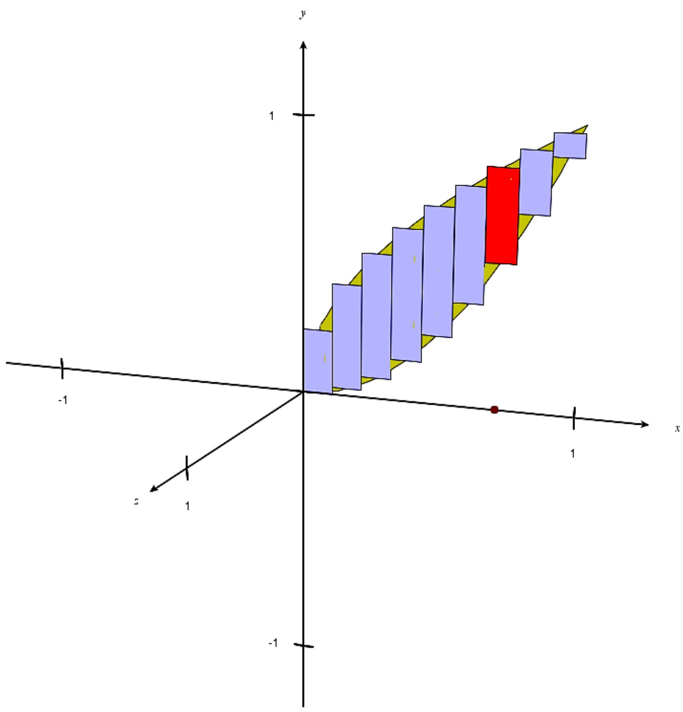
$$V = 2\pi \int_0^3 x\sqrt{x}dx = 2\pi \int_0^3 x^{3/2}dx = \frac{36}{5}\sqrt{3}\pi$$

□

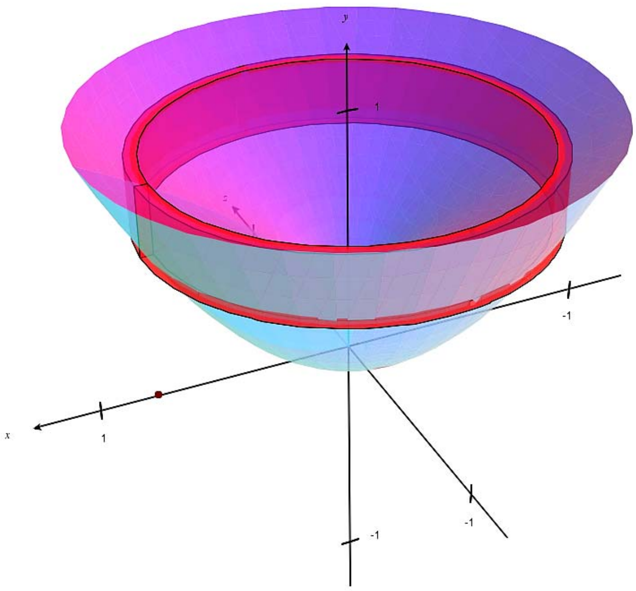
Example 16.14 *Let $\mathcal{R}$ be the region between the graph of $y = x^2$ (on the bottom) and $y = \sqrt{x}$ on top. Using the Shell Method find the volume of the solid obtained by revolving $\mathcal{R}$ about the y-axis.*

Solution:

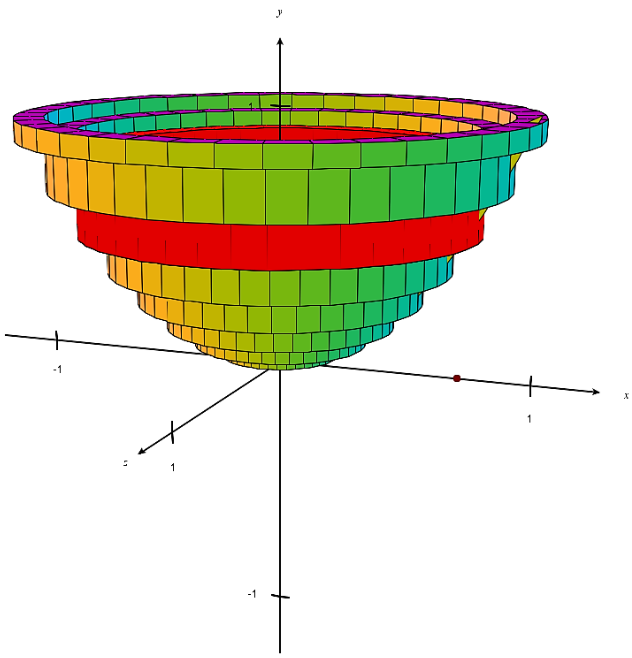
$$V = 2\pi \int_0^1 x(\sqrt{x} - x^2)dx = \frac{3}{10}\pi$$



The representative rectangle (in red).



A representative rectangle (in red) revolvs about x-axis.

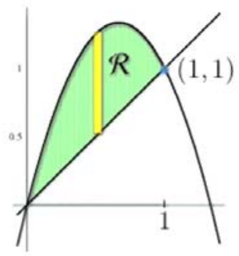


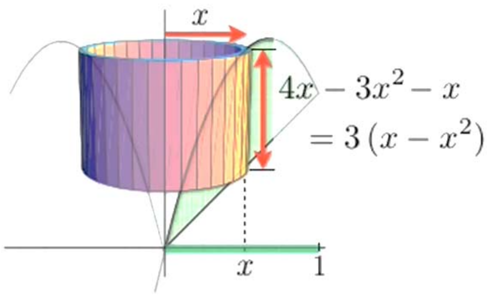
Volume of this solid approximates the volume of the original solid.

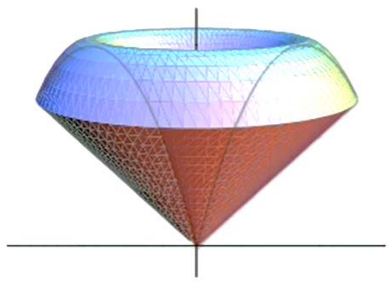
□

Example 16.15 *Let $\mathcal{R}$ be the region in which $x \le y \le 4x - 3x^2$. Find the volume of the solid obtained by revolving $\mathcal{R}$ about the y-axis.*

Solution:







$$dV = 2\pi x \cdot 3(x - x^2)dx = 6\pi(x^2 - x^3)dx$$

$$V = 6\pi \int_0^1 (x^2 - x^3)dx = -\frac{1}{2}\pi x^3 (3x - 4)\,|_0^1 = \frac{1}{2}\pi.$$

□

16.4.1 Comparison of Disk and Shell Methods

The disk and shell methods can be distinguished as follows. For the disk method, the representative rectangle is always perpendicular to the axis of revolution (Figure 16.26), whereas for the shell method, the representative rectangle is always parallel to the axis of revolution, as shown in Figure 16.27

Often, one method is more convenient to use than the other. The following example illustrates a case in which the shell method is preferable.

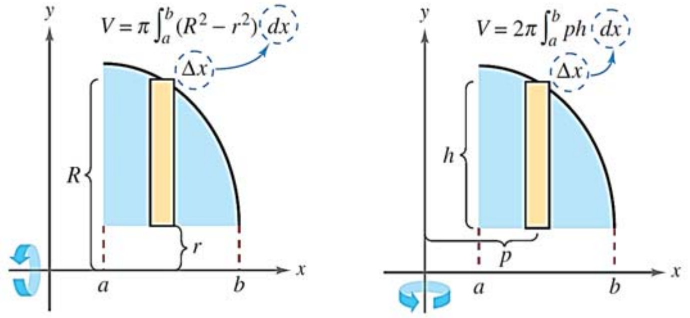


Fig. 16.26. Disk method: Representative rectangle is perpendicular to the axis of revolution.

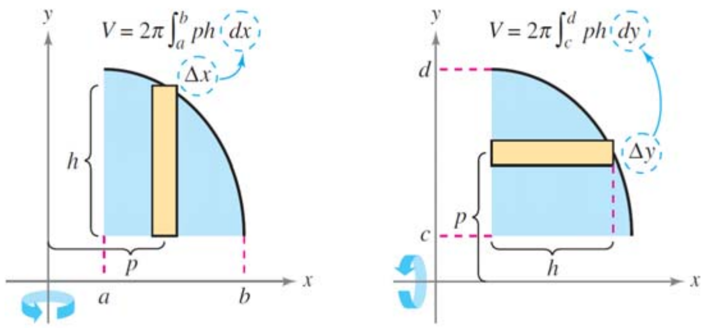


Fig. 16.27. Shell method: Representative rectangle is parallel to the axis of revolution.

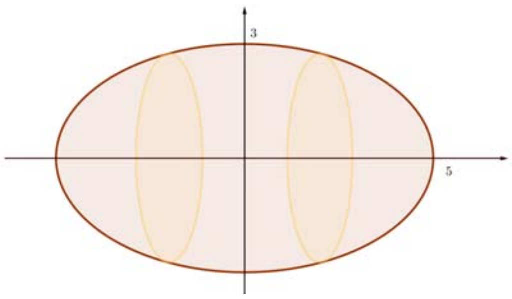
Example 16.16 *Find the volume of the solid formed by revolving the region bounded by the graphs of $y = x^2 + 1$, $y = 0$, $x = 0$, and $x = 1$ and about the y-axis, as shown in figure of Example 16.12.*

Solution: In Example 16.12 in the preceding section, you saw that the washer method requires two integrals to determine the volume of this solid. The shell method requires only one integral to find the volume.

$$\begin{aligned} V &= 2\pi \int_a^b p(x)h(x)dx \\ &= 2\pi \int_a^b x\left(x^2+1\right)\right) dx \\ &= 2\pi \left[\frac{x^4}{4} + \frac{x^2}{2}\right] |_0^1 \\ &= 2\pi \left(\frac{3}{4}\right) \\ &= \frac{3}{2}\pi. \end{aligned}$$

□

Example 16.17 *The ellipse $\frac{x^2}{25} + \frac{y^2}{9} = 1$ is revolved about the x-axis. Calculate the volume of the resulting football-shaped solid.*



Solution: We first solve for y^2

$$\frac{x^2}{25} + \frac{y^2}{9} = 1 \Rightarrow y^2 = \frac{9}{25}(25 - x^2).$$

Hence, the equation for the upper half of the ellipse is :

$$y = \frac{3}{5}\sqrt{25 - x^2}.$$

The volume is therefore

$$\begin{aligned} V &= \pi \int_{-5}^{5} \left[\frac{3}{5}\sqrt{25 - x^2} \right]^2 dx \\ &= \pi \int_{-5}^{5} \frac{9}{25}(25 - x^2) dx = 60\pi. \end{aligned}$$

Note that if the ellipse had been rotated about the y-axis, then the volume of the resulting solid would be 100π . Try using Figure above to set up the integral for the volume using the shell method. Does the integral seem more complicated?

□

17
Arc Length and Surface Area

In addition to being able to use definite integrals to find the areas of certain geometric regions, we can also use the definite integral to find the length of a portion of a curve.

17.1 Finding the length of a plane curve

Our first objective is to define what we mean by the length (also called the arc length) of a plane curve $y = f(x)$ over an interval $[a, b]$ (Figure 17.1). Once that is done we will be able to focus on the problem of computing arc lengths. To avoid some complications that would otherwise occur, we will impose the requirement that f' be continuous on $[a, b]$, in which case we will say that $y = f(x)$ is a *smooth curve* on $[a, b]$ or that f is a smooth function on $[a, b]$. Thus, we will be concerned with the following problem.

Suppose that $y = f(x)$ is a smooth curve on the in interval $[a, b]$. Define and find a formula for the arc length s of the curve $y = f(x)$ over the interval $[a, b]$.

To define the arc length of a curve we start by breaking the curve into small segments. Then we approximate the curve segments by line segments and add the lengths of the line segments to form a Riemann sum. Such line segments tend to become better and better approximations to a curve as the number of segments increases. As the number of segments increases, the corresponding Riemann sums approach a definite integral whose value we will take to be the arc *length s of the curve.*

To implement our idea for solving our problem , divide the interval $[a, b]$ into n subintervals by inserting points $a < x_1 < x_2 < ... < x_{n-1} < b$ between $a = x_0$ and $b = x_n$. As shown in Figure 17.2, let $P_0, P_1, ..., P_n$ be the points on the curve with x-coordinates $a = x_0$, x_1, x_2, . . . , x_{n-1}, $b = x_n$ and join these points with straight line segments. These line segments form a polygonal path that we can regard as an piecewise approximation to the curve $y = f(x)$.

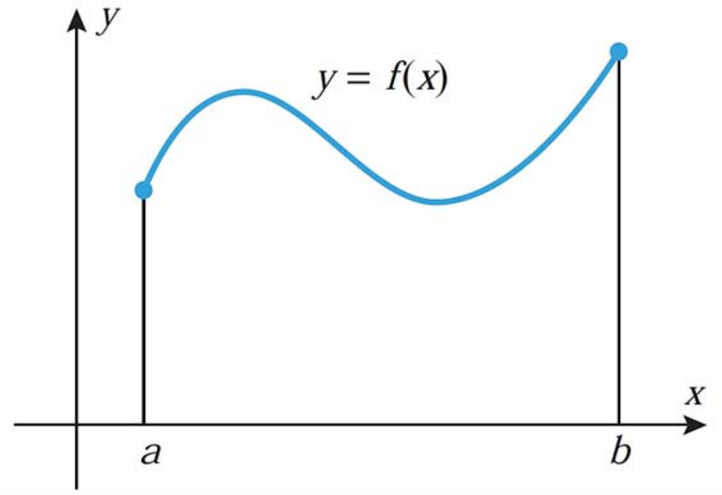


Fig. 17.1. Portion of a plane curve.

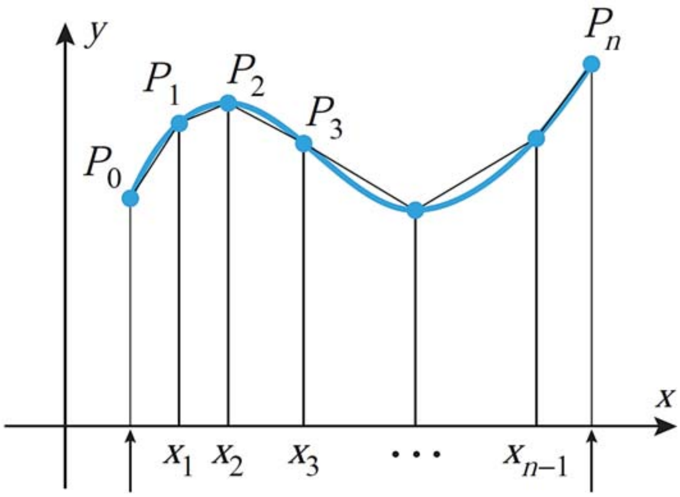


Fig. 17.2. Piecewise linear apptoximation to the curvr $f(x)$.

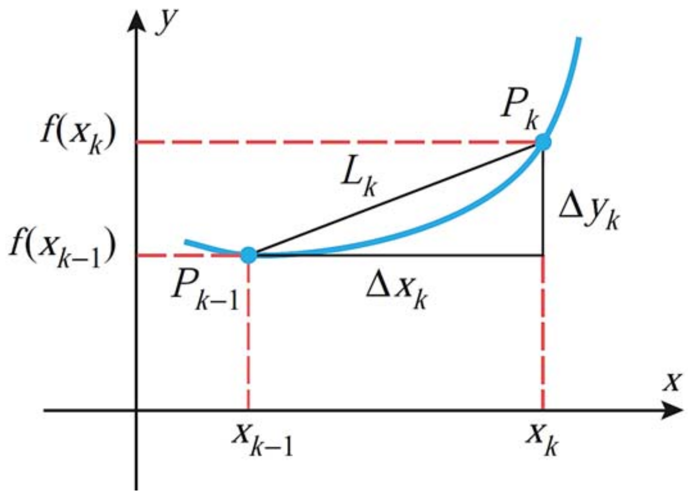


Fig. 17.3. kth line segment in the polygonal path.

As indicated in Figure 17.3, the length s_k of the kth line segment in the polygonal path is (by the Pythagorean theorem)

$$s_k = \sqrt{(\Delta x_k)^2 + (\Delta y_k)^2} = \sqrt{(\Delta x_k)^2 + (f(x_k) - f(x_{k-1}))^2} \tag{17.1}$$

If we now add the lengths of these line segments, we obtain the following approximation to the length s of the curve

$$s \approx \sum_{k=1}^{n} s_k = \sum_{k=1}^{n} \sqrt{(\Delta x_k)^2 + (f(x_k) - f(x_{k-1}))^2} \tag{17.2}$$

To put this in the form of a Riemann sum we will apply the Mean-Value Theorem. This theorem implies that there is a point x_k^* between x_{k-1} and x_k such that

$$\frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}} = f'(x_k^*) \quad \text{or} \quad f(x_k) - f(x_{k-1}) = f'(x_k^*)\,(x_k - x_{k-1})$$

and hence we can rewrite (17.2) as

$$s \approx \sum_{k=1}^{n} \sqrt{(\Delta x_k)^2 + \left(f'(x_k^*)\right)^2 (\Delta x_k)^2} = \sum_{k=1}^{n} \sqrt{1 + \left(f'(x_k^*)\right)^2}\,\Delta x_k.$$

Thus, taking the limit as n increases and the widths of all the subintervals approach zero yields the following integral that defines the arc length s:

$$s = \lim_{\max \Delta x_k \to 0} \sum_{k=1}^{n} \sqrt{1 + \left(f'(x_k^*)\right)^2}\,\Delta x_k = \int_a^b \sqrt{1 + (f'(x))^2}\,dx.$$

In summary, we have the following definition.

If $y = f(x)$ is a smooth curve on the interval $[a, b]$, then the arc length s of this curve over $[a, b]$ is defined as

$$s = \int_a^b \sqrt{1 + (f'(x))^2}dx. \tag{17.3}$$

This result provides both a definition and a formula for computing arc lengths. First we will check chow it works on a very elementary example.

Remark 17.1 *Even if the function $f(x)$ is relatively simple, the integral 17.3 can be very difficult to determine. Usually we use numerical procedures to approximate the exact value of s.*

Example 17.2 *Find the length of the line segment from* $(0, 0)$ *to* $(3, 4)$.

Solution: We already know that the length of the segment can be calculated by the distance formula:

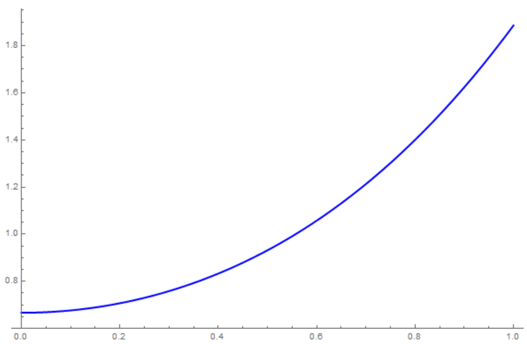
$$s = \sqrt{3^2 + 4^2} = \sqrt{25} = 5.$$

Let's verify this with the arc length formula. The equation of the line segment is $f(x) = \frac{4}{3}x$, $0 \le x \le 3$. Because $f'(x) = \frac{4}{3}$ the length of the line segment is

$$s = \int_a^b \sqrt{1 + (f'(x))^2}dx = \int_0^3 \sqrt{1 + \left(\frac{4}{3}\right)^2}dx = \int_0^3 \frac{5}{3}dx = 5,$$

as expected. □

Example 17.3 *Find the arc length of the graph of the function* $f(x) = \frac{2}{3}(x^2 + 1)^{3/2}$ *over the interval* $0 \le x \le 1$.



Solution:

$$f(x) = \frac{2}{3}(x^2+1)^{3/2}$$
$$f'(x) = 2x\sqrt{x^2+1}$$
$$1 + \big(f'(x)\big)^2 = \big(2x^2+1\big)^2$$

Fortunately, $1 + (f'(x))^2$ is a square, so we can easily compute the arc length

$$s = \int_0^1 \sqrt{1 + (f'(x))^2}\,dx = \int_0^1 \big(2x^2+1\big)\,dx = \frac{5}{3}$$

□

Example 17.4 *Find the arc length of the graph of the function* $f(x) = \dfrac{x^4}{8} + \dfrac{1}{4x^2}$ *over the interval* $1 \le x \le 3$.

Solution:

$$f(x) = \frac{x^4}{8} + \frac{1}{4x^2}$$
$$f'(x) = \frac{1}{2}x^3 - \frac{1}{2x^3}$$
$$1 + \big(f'(x)\big)^2 = \left(\frac{1}{2x^3} - \frac{1}{2}x^3\right)^2 + 1 = \left(\frac{1}{2x^3} + \frac{1}{2}x^3\right)^2$$
$$s = \int_1^3 \sqrt{1 + (f'(x))^2}\,dx = \int_1^3 \left(\frac{1}{2x^3} + \frac{1}{2}x^3\right)dx = \frac{92}{9}$$

□

17.2 Surface Area

In this section we will consider the problem of finding the area of a surface that is generated by revolving a plane curve about a line.

In this section we will be concerned with the following problem

Suppose that $y = f(x)$ is a smooth nonnegative function on $[a, b]$, and that a surface of revolution is generated by revolving the portion of the curve $y = f(x)$ between $x = a$ and $x = b$ about the x-axis (Fig. 17.4) Define what is meant by the *area* S of the surface, and find a formula for computing it.

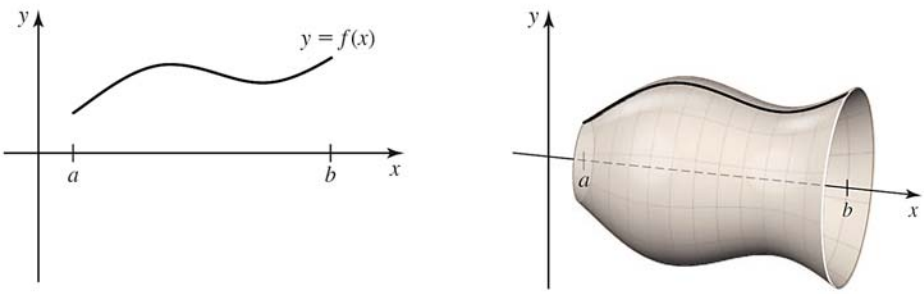


Fig. 17.4. Surface of revolution generated by a revolving curve $y = f(x)$.

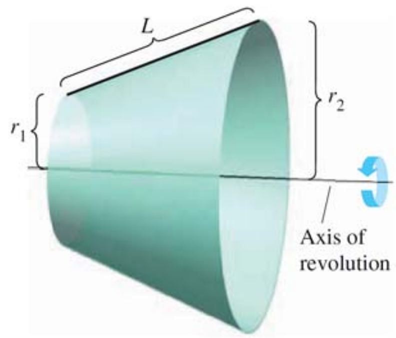


Fig. 17.5. Lateral surface area of the frustum of a right circular cone.

A *surface of revolution* is a surface that is generated by revolving a plane curve about an axis that lies in the same plane as the curve. For example, the surface of a sphere can be generated by revolving a semicircle about its diameter, and the lateral surface of a right circular cylinder can be generated by revolving a line segment about an axis that is parallel to it (Figure 17.4).

The area of a surface of revolution is derived from the formula for the lateral surface area of the frustum of a right circular cone with the vertex at the origin. Consider the line segment in Figure 17.5, where L is the length of the line segment, r_1 is the radius at the left end of the line segment, and r_2 is the radius at the right end of the line segment. When the line segment is revolved about its axis of revolution, it forms a frustum of a right circular cone, with lateral surface area of frustum

$$S = 2\pi r L$$

where

$$r = \frac{1}{2}(r_1 + r_2)$$

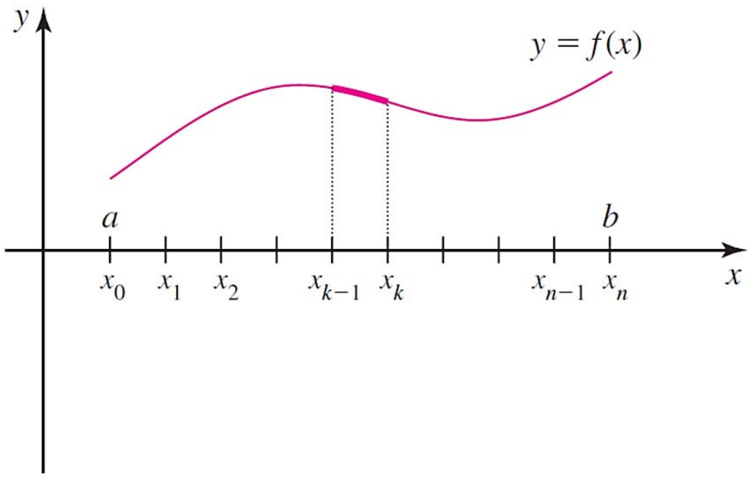


Fig. 17.6. Line segment between points $(x_{k-1}, f(x_{k-1}))$ and $(x_k, f(x_k))$.

is an average radius of frustum.

This result can be generalized to any linear function $g(x) = cx + d$ that is positive on the interval (a, b). That is, the surface area of the frustum generated by revolving the line segment between $(a, g(a))$ and $(b, g(b))$ about the x-axis is given by

$$\pi(g(b) + g(a))l \tag{17.4}$$

where

$$l = \sqrt{(b-a)^2 + (bc - ac)^2} = (b-a)\sqrt{c^2 + 1} \tag{17.5}$$

With the surface area formula for a frustum of a cone, we now derive a general area formula for a surface of revolution. We assume the surface is generated by revolving the graph of a positive, continuously differentiable function f on the interval $[a, b]$ about the x-axis. We begin by subdividing the interval $[a, b]$ into n subintervals of equal length

$$\Delta x = \frac{b-a}{n}$$

The grid points in this partition are

$$x_0 = a < x_1 < x_2, ..., x_{n-1}, x_n = b.$$

Now consider the kth subinterval $[x_{k-1}, x_k]$ and the line segment between the points $(x_{k-1}, f(x_{k-1}))$ and $(x_k, f(x_k))$ (Figure 17.6). We let the change in the y-coordinates between these points be $\Delta y_k = f(x_k) - f(x_{k-1})$.When this line

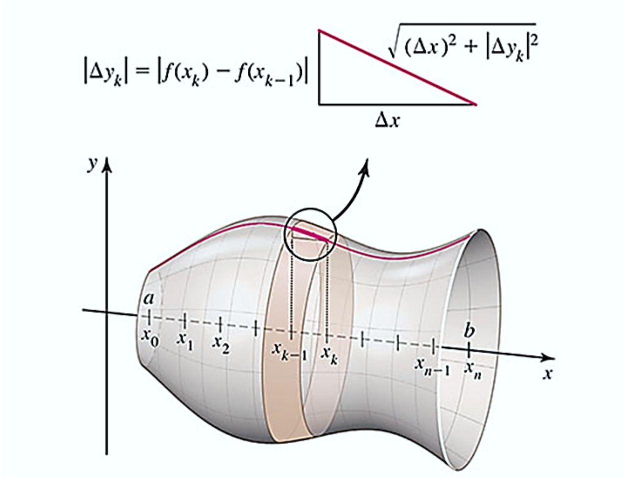


Fig. 17.7. A frustum of a cone generated by a line segment.

segment is revolved about the x-axis, it generates a frustum of a cone (Figure 17.7). The slant height of this frustum is the length of the hypotenuse of a right triangle whose sides have lengths Δx and $|\Delta y_k|$. Therefore, the slant height of the kth frustum is

$$\sqrt{(\Delta x)^2 + |\Delta y_k|^2} = \sqrt{(\Delta x)^2 + (\Delta y_k)^2}$$

and its surface area is

$$S_k = \pi(f(x_k) + f(x_{k-1}))\sqrt{(\Delta x)^2 + (\Delta y_k)^2}$$

It follows that the area S of the entire surface of revolution is approximately the sum of the surface areas of the individual frustums S_k, for $k = 1, ..., n$; that is,

$$S \approx \sum_{k=1}^{n} \pi(f(x_k) + f(x_{k-1}))\sqrt{(\Delta x)^2 + (\Delta y_k)^2}. \qquad (17.6)$$

We would like to identify this sum as a Riemann sum. However, one more step is required to put it in the correct form. We apply the Mean Value Theorem[1] on the kth subinterval $[x_{k-1}, x_k]$ and observe that

$$\frac{f(x_k) - f(x_{k-1})}{\Delta x} = f'(x_k^*)$$

[1] Notice that f satisfies the conditions of the Mean Value Theorem.

for some number x_k^* the interval (x_{k-1}, x_k), for $k = 1, ..., n$. It follows that $\Delta y_k = f(x_k) - f(x_{k-1}) = f'(x_k^*)\Delta x$.

We now replace Δy_k by $f'(x_k^*)\Delta x$ in the expression for the approximate surface area.

The result is

$$\begin{aligned} S &\approx \sum_{k=1}^{n} \pi(f(x_k) + f(x_{k-1}))\sqrt{(\Delta x)^2 + (\Delta y_k)^2} \\ &= \sum_{k=1}^{n} \pi(f(x_k) + f(x_{k-1}))\sqrt{(\Delta x)^2 \left(1 + \left(f'(x_k^*)\right)^2\right)} \\ &= \sum_{k=1}^{n} \pi(f(x_k) + f(x_{k-1}))\sqrt{\left(1 + \left(f'(x_k^*)\right)^2\right)}\,(\Delta x)\,. \end{aligned}$$

When Δx is small, we have $x_{k-1} \approx x_k \approx x_k^*$, and by the continuity of f, it follows that $f(x_k) \approx f(x_{k-1}) \approx f(x_k^*)$, for $k = 1, ..., n$. These observations allow us to write

$$\begin{aligned} S &\approx \sum_{k=1}^{n} \pi(f(x_k^*) + f(x_k^*))\sqrt{\left(1 + \left(f'(x_k^*)\right)^2\right)}\,(\Delta x) \\ &= \sum_{k=1}^{n} 2\pi f(x_k^*)\sqrt{\left(1 + \left(f'(x_k^*)\right)^2\right)}\,(\Delta x) \end{aligned}$$

This approximation to S, which has the form of a Riemann sum, improves as the number of subintervals increases and as the length of the subintervals approaches 0. Specifically, as $n \to \infty$ and as $\Delta x \to 0$, we obtain an integral for the surface area:

$$\begin{aligned} S &= \lim_{\Delta x \to 0} \sum_{k=1}^{n} 2\pi f(x_k^*)\sqrt{\left(1 + \left(f'(x_k^*)\right)^2\right)}\,(\Delta x) \\ &= \int_a^b 2\pi f(x)\sqrt{\left(1 + (f'(x))^2\right)}dx \qquad (17.7) \end{aligned}$$

In a similar manner, if the graph of is revolved about the y axis, then is

$$S = \int_a^b 2\pi x\sqrt{\left(1 + (f'(x))^2\right)}dx \qquad (17.8)$$

In these two formulas for you can regard the products $2\pi f(x)$ and $2\pi x$ as the circumferences of the circles traced by a point on the graph of as it is revolved about the x-axis and the y-axis. In one case the radius is $r = f(x)$ and in the

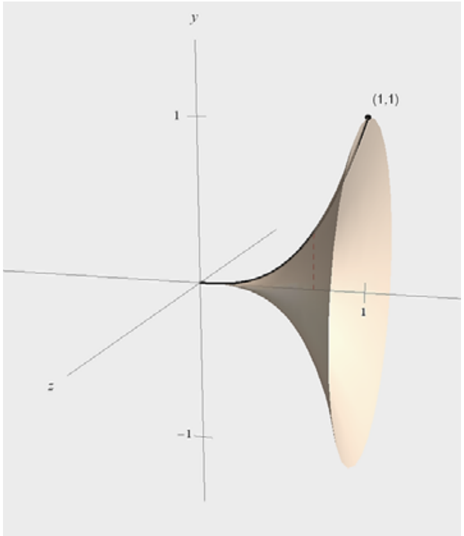


Fig. 17.8. The surface formed by revolving the graph of $f(x) = x^3$ on the interval about the axis.

other case the radius is x. Moreover, by appropriately adjusting r, you can generalize the formula for surface area to cover *any* horizontal or vertical axis of revolution, as indicated in the following definition.

Definition of the area of a surface of revolution

Let $y = f(x)$ have a continuous derivative on the interval $[a, b]$. The area of the surface of revolution formed by revolving the graph of about a horizontal or vertical axis is

$$S = 2\pi \int_a^b r(x)\sqrt{\left(1 + (f'(x))^2\right)}dx \quad y \text{ is a function of } x$$

where $r(x)$ is the distance between the graph of f and the axis of revolution. If $x = g(y)$ on the interval $[c, d]$ then the surface area is

$$S = 2\pi \int_a^b r(y)\sqrt{\left(1 + (g'(y))^2\right)}dy \quad x \text{ is a function of } y$$

where $r(y)$ is the distance between the graph of g and the axis of revolution.

Example 17.5 *Find the area of the surface formed by revolving the graph of $f(x) = x^3$ on the interval about the axis, as shown in Figure 17.8*

Solution: The distance between the x-axis and the graph of is f is $r(x) = f(x)$, and because $f'(x) = 3x^2$the surface area is

$$\begin{aligned} S &= 2\pi \int_a^b r(x)\sqrt{\left(1 + (f'(x))^2\right)}dx \\ &= 2\pi \int_0^1 x^3\sqrt{\left(1 + (3x^2)^2\right)}dx \\ &= \frac{2\pi}{36}\int_0^1 \left(36x^3\right)\sqrt{(1+9x^4)}dx \\ &= \frac{\pi}{18}\left[\frac{\left(1+9x^4\right)^{3/2}}{3/2}\right]_0^1 \\ &= \frac{\pi}{27}(10^{3/2} - 1) \approx 3.5631 \end{aligned}$$

□

Example 17.6 (Surface area of a sphere) *Find the area of the surface of a sphere of radius a.*

Solution: Such a sphere can be generated by rotating the semicircle with equation $f(x) = \sqrt{a^2 - x^2}, \quad -a \le x \le a,$ about the x-axis. Since

$$f'(x) = -\frac{x}{\sqrt{a^2 - x^2}} = \frac{-x}{f(x)}$$

the area of the sphere is given by

$$\begin{aligned} & 2\pi \int_{-a}^a f(x)\sqrt{\left(1 + \left(\frac{x}{f(x)}\right)^2\right)}dx \\ = \ & 4\pi \int_0^a \sqrt{\left(f^2(x) + (x)^2\right)}dx \\ = \ & 4\pi \int_0^a \sqrt{a^2}dx = 4\pi a x|_0^a =: 4\pi a^2 \end{aligned}$$

□

Example 17.7 (Surface area of a parabolic dish) *Find the surface area of a parabolic reflector whose shape is obtained by rotating the parabolic arc $y = x^2, \quad 0 \le x \le 1,$ about the y-axis.*

Solution: In this case, the distance between the graph f of and the y-axis is $r(x) = x$. Using $f'(x) = 2x$ we can determine that the surface area is

$$\begin{aligned} & \int_a^b 2\pi x \sqrt{\left(1 + (f'(x))^2\right)} dx \\ = & \ 2\pi \int_0^1 x \sqrt{\left(1 + (2x)^2\right)} dx \\ = & \ \frac{2\pi}{8} \int_0^1 (8x) \left(1 + 4x^2\right)^{1/2} dx \\ = & \ \frac{\pi}{4} \frac{2}{3} \left(4x^2 + 1\right)^{\frac{3}{2}} |_0^1 \\ = & \ \left(\frac{5}{6}\sqrt{5} - \frac{1}{6}\right) \pi \end{aligned}$$

□

17.3 Gabriel's Horn and Improper Integrals

The definition of a definite integral

$$\int_a^b f(x) dx$$

requires that the interval $[a, b]$ be finite. In this section we will study a procedure for evaluating integrals that do not satisfy this requirement because either one or both of the limits of integration are infinite. Integrals that possess either property are *improper integrals.*

To get an idea of how to evaluate an improper integral, consider the integral

$$\int_1^b \frac{1}{x^2} dx = -\frac{1}{x}|_1^b = 1 - \frac{1}{b}$$

Taking the limit as $b \to \infty$ produces

$$\int_1^\infty \frac{1}{x^2} dx = \lim_{b \to \infty} \int_1^b \frac{1}{x^2} dx = \lim_{b \to \infty} \left(1 - \frac{1}{b}\right) = 1.$$

This improper integral can be interpreted as the area of the unbounded region between the graph of $f(x) = 1/x^2$ and the x-axis (to the right of $x = 1$).

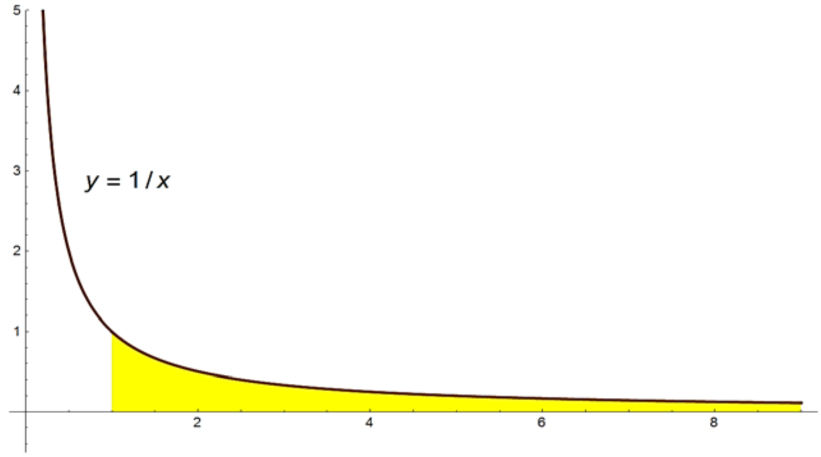


Fig. 17.9. Function $f(x) = 1/x$.

Definition of improper integrals with infinite integration limits
1. If f is continuous on the interval $[a, \infty)$ then $\int_a^\infty f(x)dx = \lim_{b\to\infty} \int_a^b f(x)dx$
2. If f is continuous on the interval $(-\infty, b]$ then $\int_{-\infty}^b f(x)dx = \lim_{a\to-\infty} \int_a^b f(x)dx$
3. If f is continuous on the interval $(-\infty, \infty)$ then $\int_{-\infty}^\infty f(x)dx = \int_{-\infty}^c f(x)dx + \int_c^\infty f(x)dx$ where c is any real number.

Example 17.8 *Here we will calculate the volume of a solid of revolution, even though the solid "goes out to infinity". Consider the region B under the graph of* $y = 1/x$ *and above the interval* $[1; \infty)$ *on the x-axis (Figure 17.9). When you revolve B around the x-axis, the solid of revolution S that is generated looks like some sort of long, narrow, trumpet, a "horn", where the mouthpiece has been stretched out to infinity. Of course, the solid region should be filled in, but we've drawn just the surface in order to make it look more horn-like. We'll use disks to calculate the volume of S. Our disks are cross sections perpendicular to the x-axis; the radii are given by* $r(x) = 1/x$*, and so* $dV = \pi(1/x)^2 dx$*. The*

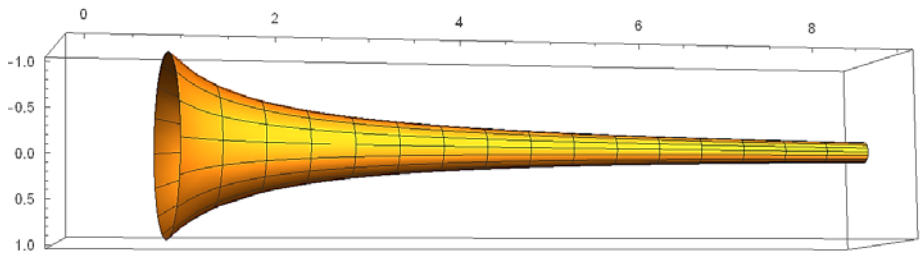


Fig. 17.10. Gabriel's Horn (initial part)

total volume is

$$\begin{aligned} V &= \int_1^\infty \pi(1/x)^2 dx = \lim_{b\to\infty}\int_1^b \pi(1/x)^2 dx \\ &= \pi\left[\lim_{b\to\infty}\frac{x^{-1}}{-1}|_1^b\right] = \pi\lim_{b\to\infty}\left(1-\frac{1}{b}\right) = \pi. \end{aligned}$$

Now, we wish to show that the surface area of Gabriel's horn is ***infinite****. (When we write "surface area" here, we mean the area of the "sides", i.e., we are excluding the disk that could fill the flared end of the horn at* $x = 1$*. However, as the surface area is infinite without the disk at* $x = 1$*, the surface area would certainly still be infinite if we included the disk.)*
The curve that we are revolving around the x*-axis is the graph of* $f(x) = 1/x = x^{-1}$ *for* $x \geq 1$*. We find*

$$\sqrt{1+(f'(x))^2} = \sqrt{1+(-x^{-2})^2} = \sqrt{\frac{1}{x^4}+1}$$

and r*, the distance from a point on the graph to the* x*-axis, in terms of* x*, is given by the* y*-coordinate of the graph of* $y = 1/x$*,* $r = 1/x$*.*
Therefore

$$\textit{surface area} = \int_1^\infty 2\pi\cdot\frac{1}{x}\sqrt{\frac{1}{x^4}+1}dx = 2\pi\lim_{b\to\infty}\int_1^b \frac{1}{x}\sqrt{\frac{1}{x^4}+1}dx$$

An easy substitution will not let us find an anti-derivative of $\frac{1}{x}\sqrt{\frac{1}{x^4}+1}$*. However, note that, for* $x \geq 1$*,*

$$\frac{1}{x}\sqrt{\frac{1}{x^4}+1} \geq \frac{1}{x}.$$

Thus (from the properties of definite integrals), if $b \geq 1$

$$\int_1^b \frac{1}{x}\sqrt{\frac{1}{x^4}+1}dx \geq \int_1^b \frac{1}{x}dx = \ln x|_1^b = \ln b.$$

As b goes to infinity, so does $\ln b$*, and this forces the larger quantity*

$$\int_1^b \frac{1}{x}\sqrt{\frac{1}{x^4}+1}dx$$

to go to infinity also. We conclude:

$$\textit{surface area of Gabriel's horn} = 2\pi\infty = \infty.$$

□

Remark 17.9 *The results of Example 17.8 tell us that Gabriel's horn has finite volume, but infinite surface area. These results are sometimes described as "you can fill Gabriel's horn, but you can't paint it". The clever student then asks "What if you filled the horn with paint? Wouldn't that paint the surface?". This seeming contradiction is caused by a lack of precision in claiming that having infinite surface area means that a surface can't be painted. What is true is that having infinite surface area implies that the surface cannot be painted with a finite amount of paint, if we are required to have a uniformly thick layer of paint everywhere. However, if it were possible to have arbitrarily thin layers of paint, then the surface of Gabriel's horn could be painted.*

We'll try to describe a similar problem, in which it's hopefully easier to see what's going on. Suppose you took a cube that's 1 foot long on each side, and you fill it with paint. Then, the volume of paint is finite; it's 1 ft^3*. The surface area of the cube is the combined area of the* 6 *sides, namely,* 6 *ft*^{2}*.*

Now, imagine chopping the cube in half by a cut which is parallel to two of the faces, while simultaneously sealing the two new exposed sides (or, you could think of inserting dividers into the cube first, then chopping the cube in half). The total volume of paint in the two half-cubes is still 1 ft^3*, but now the surface area has gone up, because we created two new faces; the surface area is now* $6 + 2 = 8$ *ft*^{2}*.*

Now, by making a cut parallel to the original cut, divide (and seal) one of the two half-cubes from above; the volume of paint remains 1 *ft*^{3}*, but we added two more faces, for a new surface area of* 10 *ft*^{2}*. Imagine continuing this process indefinitely, each time, taking one of your smallest two pieces, and dividing it into two pieces by making a cut parallel to all of the other cuts. The volume of paint is always* 1 *ft*^{3}*, but the surface area gets arbitrarily large or, in the* *limit, is infinite. Is there a contradiction here? No, but note that the layer of paint on the sides of the smaller and smaller pieces can't be any thicker than the width of each piece, which is getting arbitrary small (close to zero).*

18

Basic techniques of integration, part one

"*Learning the art of integration requires practice.*"

18.1 Fitting Integrands to Basic Integration Rules

In this chapter, we first collect in a more systematic way some of the integration formulas derived before. We then present the two most important general techniques: *integration by substitution* and *integration by parts*. As the techniques for evaluating integrals are developed, you will see that integration is a more subtle process than differentiation and that it takes practice to learn which method should be used in a given problem. A major step in solving any integration problem is recognizing which basic integration rule to use. As shown in Example 18.1, slight differences in the integrand can lead to very different solution techniques.

Example 18.1 (A Comparison of Three Similar Integrals) *Find each integral.*

$$\textbf{\textit{a.}}\ \int \frac{4}{x^2+9}dx \qquad \textbf{\textit{b.}}\ \int \frac{4x}{x^2+9}dx \qquad \textbf{\textit{c.}}\ \int \frac{4x^2}{x^2+9}dx$$

Solution:

a) Use the Arctangent Rule:

$$\begin{aligned}\int \frac{4}{x^2+9}dx &= 4\int \frac{1}{x^2+9}dx = 4\int \frac{1}{x^2+(3)^3}dx \\ &= 4\left(\frac{1}{3}\arctan\frac{x}{3}\right)+C \\ &= \frac{4}{3}\arctan\frac{x}{3}+C\end{aligned}$$

b) Here the Arctangent Rule does not apply because the numerator contains a factor of x.Consider the Log Rule

$$\begin{aligned}\int \frac{4x}{x^2+9}dx &= 2\int \frac{2x}{x^2+9}dx \\ &= 2\ln\left(x^2+9\right)+C\end{aligned}$$

c) Because the degree of the numerator is equal to the degree of the denominator, we should first use division to rewrite the improper rational function as the sum of a polynomial and a proper rational function.

$$\begin{aligned}\int \frac{4x^2}{x^2+9}dx &= \int \left(4-\frac{36}{x^2+9}\right)dx \\ &= 4\int 1dx - 36\int \frac{1}{x^2+9}dx \\ &= 4x - 12\arctan\frac{x}{3}+C\end{aligned}$$

□

Sometimes it is difficult to recognize how to calculate the integral. Here is an example.

Example 18.2 (Tricky, clever example) *Evaluate*

$$\int \frac{1}{1-\sin x}dx.$$

Solution:

$$\begin{aligned}\int \frac{1}{1-\sin x}dx &= \int \frac{1}{1-\sin x}\frac{1+\sin x}{1+\sin x}dx \\ &= \int \frac{1+\sin x}{1-\sin^2 x}dx \\ &= \int \frac{1+\sin x}{\cos^2 x}dx \\ &= \int \frac{1}{\cos^2 x}dx + \int \frac{\sin x}{\cos^2 x}dx \\ &= \tan x + \frac{1}{\cos x} + C. \quad \square\end{aligned}$$

18.2 Calculating Integrals

In this section, we review the basic integration formulas learned in before.

Given a function $f(x)$, $\int f(x)dx$ denotes the *general antiderivative* of f, also called the *indefinite integral*. Thus

$$\int f(x)dx = F(x) + C$$

where $F'(x) = f(x)$ and C is a constant. Therefore

$$\frac{d}{dx}\int f(x)dx = f(x).$$

The definite integral is obtained via the fundamental theorem of calculus by evaluating the indefinite integral at the two limits and subtracting. Thus:

$$\int_a^b f(x)dx = F(x)|_a^b = F(b) - F(a).$$

We recall the following general rules for antiderivatives, which may be deduced from the corresponding differentiation rules. To check the sum rule, for instance, we must see if

$$\frac{d}{dx}\left[\int f(x)dx + \int g(x)dx\right] = f(x) + g(x).$$

But this is true by the sum rule for derivatives

Sum and Constant Multiple Rules for Antiderivatives

$\int [f(x) + g(x)]\, dx = \int f(x)dx + \int g(x)dx.$

$\int cf(x)dx = c\int f(x)dx.$

The antiderivative rule for powers is given as follows

Power Rule For Antiderivatives

$$\int x^n dx = \begin{cases} \dfrac{x^{n+1}}{n+1} + C & n \neq -1 \\ \ln|x| + C & n = -1 \end{cases}$$

Example 18.3 *Calculate*

a) $\int \left(3x^{2/3} + \frac{8}{x}\right) dx;$

b) $\int \left(\frac{x^3 + 8x + 3}{x}\right) dx;$

c) $\int \left(x^\pi + x^3\right) dx.$

Solution:

a) By the sum and constant multiple rules,

$$\int \left(3x^{2/3} + \frac{8}{x}\right) dx = 3\int x^{2/3} dx + 8\int \frac{1}{x} dx.$$

By the power rule, this becomes

$$\frac{9}{5}x^{\frac{5}{3}} + 8\ln|x| + C.$$

b)

$$\int \left(\frac{x^3 + 8x + 3}{x}\right) dx = \int \left(\frac{3}{x} + x^2 + 8\right) dx = 8x + 3\ln|x| + \frac{1}{3}x^3 + C.$$

c)

$$\int \left(x^\pi + x^3\right) dx = \frac{x^{\pi+1}}{\pi+1} + \frac{x^4}{4} + C.$$

□

Applying the fundamental theorem to the power rule, we obtain the rule for definite integrals of powers:

integer, a and b must be positive (or zero if $n > 0$).

Definite Integral of a Power

$$\int_a^b x^n dx = \frac{x^{n+1}}{n+1}\Big|_a^b = \frac{b^{n+1} - a^{n+1}}{n+1} \qquad \text{for } n \text{ real } n \neq 1$$

If $n = -2, -3, -4, ...$, a and b must have the same sign. If n is not an integer, a and b must be positive (or zero if $n > 0$).

$$\int_a^b \frac{1}{x} dx = \ln|x|\,\Big|_a^b = \ln|b| - \ln|a| = \ln\tfrac{b}{a}$$

Again a and b must have the same sign.

The extra conditions on a and b are imposed because the integrand must be defined and continuous on the domain of integration; otherwise the fundamental theorem does not apply.

Example 18.4 *Evaluate*

a) $\int_0^1 \left(x^4 - 3\sqrt{x}\right) dx$;

b) $\int_1^2 \left(\sqrt{x} + \dfrac{2}{x}\right) dx$;

c) $\dfrac{x^4 + x^2 + 1}{x^2} dx$

Solution:

a) $\int_0^1 \left(x^4 - 3\sqrt{x}\right) dx = \frac{1}{5}x^{\frac{3}{2}}\left(x^{\frac{7}{2}} - 10\right)|_0^1 = -\frac{9}{5}$;

b) $\int_1^2 \left(\sqrt{x} + \dfrac{2}{x}\right) dx = 2\ln|x| + \frac{2}{3}x^{\frac{3}{2}}|_1^2 = 2\ln 2 + \frac{4}{3}\sqrt{2} - \frac{2}{3}$;

c) $\int_{1/2}^1 \left(\dfrac{x^4 + x^6 + 1}{x^2}\right) dx = \int_{1/2}^1 \left(\frac{1}{x^2} + x^2 + x^4\right) dx = \frac{1}{15x}\left(3x^6 + 5x^4 - 15\right)|_{1/2}^1 = \frac{713}{480}$.

□

In the following box, we recall some general properties satisfied by the definite integral.

Properties of the Definite Integral

1. *Inequality rule*: If $f(x) \leq g(x)$ for all x in $[a, b]$, then

$$\int_a^b f(x)dx \leq \int_a^b g(x)dx$$

2. *Sum rule*:

$$\int_a^b [f(x) + g(x)]\, dx = \int_a^b f(x)dx + \int_a^b g(x)dx$$

3. *Constant multiple rule*:

$$\int_a^b cf(x)dx = c\int_a^b f(x)dx$$

4. *Endpoint additivity rule*:

$$\int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx$$

5. *Wrong-way integrals:*

$$\int_b^a f(x)dx = -\int_a^b f(x)dx$$

Example 18.5 *Let*

$$f(t) = \begin{cases} \frac{1}{2} & 0 \leq t < \dfrac{1}{2} \\ t & \dfrac{1}{2} \leq t \leq 1 \end{cases}$$

Draw a graph of f and evaluate $\int_0^1 f(t)dt$.

Solution: The graph of f is drawn in Figure 18.1. To evaluate the integral,

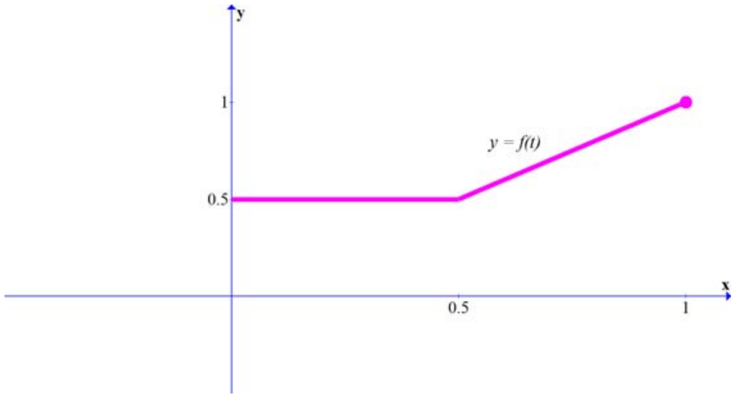


Fig. 18.1. The integral of f on $[0, 1]$ is the sum of its integrals on $[0, 1/2]$ and $[1/2, 1]$.

we apply the endpoint additivity rule with $a = 0$, $b = \frac{1}{2}$, and $c = 1$.

$$\begin{aligned}\int_0^1 f(t)dt &= \int_0^{1/2} f(t)dt + \int_{1/2}^1 f(t)dt \\ &= \int_0^{1/2} \left(\frac{1}{2}\right) dt + \int_{1/2}^1 tdt \\ &= \frac{1}{2}t|_0^{1/2} + \frac{1}{2}t^2|_{1/2}^1 \\ &= \frac{5}{8}.\end{aligned}$$

□

Let us recall that the alternative form of the fundamental theorem of calculus states that if f is continuous, then

$$\frac{d}{dx}\int_a^x f(t)dt = f(x)$$

Example 18.6 *Find*

$$\frac{d}{dt}\int_a^{t^2} \sqrt{1 + 2s^3}ds.$$

Solution: We write $g(t) = \int_a^{t^2} \sqrt{1 + 2s^3}ds$ as $f(t^2)$, where $f(u) = \int_a^u \sqrt{1 + 2s^3}ds$. By the fundamental theorem of calculus

$$f'(u) = 1 + 2u^3,$$

by the chain rule

$$g'(t) = f'(t^2)\frac{d\left(t^2\right)}{dt} = 2t\sqrt{2t^6+1}\ .$$

□

As we developed the calculus of the trigonometric and exponential functions, we obtained formulas for the antiderivatives of certain of these functions. For convenience, we summarize those formulas.

Trigonometric Formulas

1. $\int \cos\theta d\theta = \sin\theta + C$	2. $\int \sin\theta d\theta = -\cos\theta + C$
3. $\int \sec^2\theta d\theta = \tan\theta + C$	4. $\int \csc^2\theta d\theta = -\cot\theta + C$

Inverse Trigonometric Formulas

1.	$\int \frac{dx}{\sqrt{a^2-x^2}} = \arcsin\frac{x}{a} + C$	$-1 < x < 1, \quad a \neq 0,$
2.	$\int \frac{-dx}{\sqrt{a^2-x^2}} = \arccos\frac{x}{a} + C$	$-1 < x < 1 \quad a \neq 0,$
3.	$\int \frac{dx}{a^2+x^2} = \frac{1}{a}\ \arctan\frac{x}{a} + C$	$-\infty < x < \infty \quad a \neq 0.$

Example 18.7 *Evaluate :*

a) $\int_0^{\pi}(x^4 + 2x + \sin x)dx$

b) $\int_0^{\pi/6}\cos 3x dx$

c) $\int_{-1/2}^{1/2}\frac{dy}{\sqrt{1-y^2}}$

Solution:

a) $\int (x^4 + 2x + \sin x)dx = \frac{1}{5}x^5 + x^2 - \cos x + C$. The fundamental theorem of calculus gives

$$\begin{aligned}\int_0^{\pi} (x^4 + 2x + \sin x)dx &= \left(\frac{1}{5}x^5 + x^2 - \cos x\right)|_0^{\pi} \\ &= \pi^2 + \frac{1}{5}\pi^5 + 2 \\ &\approx 73.074\end{aligned}$$

b) An antiderivative of $\cos 3x$ is, by guesswork, $\dfrac{\sin 3x}{3}$. Thus

$$\int_0^{\pi/6} \cos 3x dx = \frac{\sin 3x}{3}|_0^{\pi/6} = \frac{1}{3}.$$

c) From the preceding box, we have $\int \dfrac{dy}{\sqrt{1-y^2}} = \arcsin y + C$, and so by the fundamental theorem,

$$\int_{-1/2}^{1/2} \frac{dy}{\sqrt{1-y^2}} = \arcsin y|_{-1/2}^{1/2} = \frac{\pi}{6} - \left(-\frac{\pi}{6}\right) = \frac{1}{3}\pi$$

□

The following box summarizes the antidifferentiation formulas obtained for the Exponential and Logarithmic functions

Exponential and Logarithm	
1.	$\int e^x dx = e^x + C$
2.	$\int b^x dx = \dfrac{b^x}{\ln b} + C$
3.	$\int \dfrac{1}{x} dx = \ln\lvert x\rvert + C$

Example 18.8 *Find*

a) $\int_{-1}^{1} 2^x dx$

b) $\int_0^1 (3e^x + 2\sqrt{x})\, dx$

c) $\int_0^1 2^{2y} dy$

Solution:

a) $\int 2^x dx = \dfrac{2^x}{\ln 2} + C$, so

$$\int_{-1}^{1} 2^x dx = \frac{2^x}{\ln 2}|_{-1}^{1} = \frac{3}{2\ln 2}.$$

b) $\int (3e^x + 2\sqrt{x})\, dx = 3e^x + \frac{4}{3}x^{\frac{3}{2}} + C$, and so by the fundamental theorem,

$$\begin{aligned}\int_0^1 \left(3e^x + 2\sqrt{x}\right) dx &= \left(3e^x + \frac{4}{3}x^{\frac{3}{2}}\right)|_0^1 \\ &= 3e - \frac{5}{3} \\ &\approx 6.4882.\end{aligned}$$

c) $\int 2^{2y} dy = \frac{1}{2}\dfrac{2^{2y}}{\ln 2}$, therefore

$$\begin{aligned}\int_0^1 2^{2y} dy &= \frac{1}{2}\frac{2^{2y}}{\ln 2}|_0^1 \\ &= \frac{3}{2\ln 2}\end{aligned}$$

□

Example 18.9

a) *Differentiate* $x \ln x$

b) *Find* $\int \ln x dx$

c) *Find* $\int_2^5 \ln x dx$

Solution:

a) $\dfrac{d}{dx}(x \ln x) = \ln x + 1$

b) From *a)* $\int (\ln x + 1)\, dx = x \ln x + C$, Hence,

$$\int \ln x dx = x \ln x - x + C$$

c) $\int_2^5 \ln x dx = (x \ln x - x)\,|_2^5 = 5\ln 5 - 2\ln 2 - 3.$

□

18.3 Integration by Substitution

The method of integration by substitution is based on the chain rule for differentiation. If F and g are differentiable functions, the chain rule tells us that $(F(g(x))' = F'(g(x))g'(x)$; that is, $F(g(x))$ is an antiderivative of $F'(g(x))g'(x)$. In indefinite integral notation, we have

$$\int F'(g(x))g'(x)dx = (F(g(x)) + C \tag{18.1}$$

As in differentiation, it is convenient to introduce an intermediate variable $u = g(x)$; then the preceding formula becomes

$$\int F'(u)\frac{du}{dx}dx = F(u) + C \tag{18.2}$$

If we write $f(u)$ for $F'(u)$, so that $\int f(u)du = F(u)+C$, we obtain, the formula

$$\int f(u)\frac{du}{dx}dx = \int f(u)du \tag{18.3}$$

This formula is easy to remember, since one may "cancel the dx's."

To apply the method of substitution one must find in a given integrand an expression $u = g(x)$ whose derivative $du/dx = g'(x)$ also occurs in the integrand.

Example 18.10 *Find* $\int 2x\sqrt{x^2+1}dx$ *and check the answer by differentiation,*

Solution: None of the rules in previous section apply to this integral, so we try integration by substitution. Noticing that $2x$, the derivative of $x^2 + 1$, occurs in the integrand, we are led to write $u = x^2 + 1$; then we have

$$\int 2x\sqrt{x^2+1}dx = \int \sqrt{x^2+1}2xdx = \int \sqrt{u}\frac{du}{dx}dx.$$

By formula (18.3), the last integral equals

$$\int \sqrt{u}du = \frac{2}{3}u^{\frac{3}{2}} + C.$$

At this point we substitute $x^2 + 1$ for u, which gives

$$\int 2x\sqrt{x^2+1}dx = \frac{2}{3}\left(x^2+1\right)^{\frac{3}{2}} + C$$

Checking our answer by differentiating has educational as well as insurance value, since it will show how the chain rule produces the integrand we started with:

$$\frac{d}{dx}\left(\frac{2}{3}\left(x^2+1\right)^{\frac{3}{2}}+C\right)=\frac{2}{3}\frac{3}{2}\left(x^2+1\right)^{\frac{1}{2}}\frac{d}{dx}\left(x^2+1\right)=2x\sqrt{x^2+1}$$

as it should be. □

Sometimes the derivative of the intermediate variable is "hidden" in the integrand. If we are clever, however, we can still use the method of substitution, as the next example shows.

Example 18.11 *Find* $\int \cos^2 x \sin x dx$.

Solution: We are tempted to make the substitution $u=\cos x$, but du/dx is then $-\sin x$ rather than $\sin x$. No matter-we can rewrite the integral as

$$\int \left(-\cos^2 x\right)\left(-\sin x\right)dx$$

Setting $u=\cos x$, we have

$$\int -u^2\frac{du}{dx}dx=\int -u^2 du=-\frac{1}{3}u^3+C,$$

so

$$\int \cos^2 x\sin x dx=-\frac{1}{3}\cos^3 x+C$$

You may check this by differentiating. □

Example 18.12 *Find* $\int \frac{e^x}{1+e^{2x}}dx$.

Solution: We cannot just let $u=1+e^{2x}$, because $du/dx=2e^{2x}\neq e^x$; but we may recognize that $e^{2x}=\left(e^x\right)^2$ and remember that the derivative of e^x is e^x.

Making the substitution $u=e^x$ and $du/dx=e^x$, we have

$$\begin{aligned}\int \frac{e^x}{1+e^{2x}}dx &= \int \frac{e^x}{1+\left(e^x\right)^2}dx\\ &= \int \frac{1}{1+u^2}\frac{du}{dx}\cdot dx=\int \frac{1}{1+u^2}du\\ &= \arctan u+C=\arctan\left(e^x\right)+C.\end{aligned}$$

Again you should check this by differentiation □

We may summarize the method of substitution as developed so far:

Integration by Substitution

To integrate a function which involves an intermediate variable u and its derivative du/dx, write the integrand in the form $f(u)(du/dx)$, incorporating constant factors as required in $f(u)$. Then apply the formula

$$\int f(u)\frac{du}{dx}dx = \int f(u)du$$

Finally, evaluate $\int f(u)du$ if you can; then substitute for u its expression in terms of x.

Example 18.13 *Find*

a) $\int x^2 \sin(x^3)dx$

b) $\int \sin 2x dx$

Solution:

a) We observe that the factor x^2 is, apart from a factor of 3, the derivative of x^3. Substitute $u = x^3$, so $\frac{du}{dx} = 3x^2$ and $x^2 = \frac{1}{3}\frac{du}{dx}$. Thus

$$\begin{aligned}\int x^2 \sin(x^3)dx &= \int \frac{1}{3}\frac{du}{dx}\sin(u)dx = \frac{1}{3}\int \sin(u)\frac{du}{dx}dx \\ &= \frac{1}{3}\int \sin(u)du = -\frac{1}{3}\cos u + C.\end{aligned}$$

Hence

$$\int x^2 \sin(x^3)dx = -\frac{1}{3}\cos\left(x^3\right) + C$$

b) Substitute $u = 2x$, so $du/dx = 2$. Then

$$\begin{aligned}\int \sin 2x dx &= \frac{1}{2}\int (\sin 2x)\, 2dx = \frac{1}{2}\int (\sin u)\frac{du}{dx}dx \\ &= \frac{1}{2}\int (\sin u)\, du = -\frac{1}{2}\cos u + C.\end{aligned}$$

Thus

$$\int \sin 2x dx = -\frac{1}{2}\cos 2x + C.$$

□

Example 18.14 *Evaluate:*

a) $\int \frac{x^2}{x^3+5}dx$

b) $\int \frac{1}{t^2-6t+10}dt$, (HINT: Complete the square in the denominator)

c) $\int \sin^2 2x \cos 2x dx$,

d) $\int \frac{1}{\sqrt{-7-8x-x^2}}dx$, (HINT: Complete the square)

Solution:

a) Set $u = x^3+5$; $du/dx = 3x^2$. Then

$$\begin{aligned} \int \frac{x^2}{3(x^3+5)} 3x^2 dx &= \frac{1}{3}\int \frac{1}{u}\frac{du}{dx}dx \\ &= \frac{1}{3}\int \frac{1}{u}du = \frac{1}{3}\ln|u| + C \\ &= \frac{1}{3}\ln\left|x^3+5\right| + C \end{aligned}$$

b) Completing the square, we find

$$\begin{aligned} t^2-6t+10 &= (t^2-6t+9)-9+10 \\ &= (t-3)^2+1 \end{aligned}$$

We set $u = t-3$; $du/dt = 1$. Then

$$\begin{aligned} \int \frac{1}{t^2-6t+10}dt &= \int \frac{1}{1+(t-3)^2}dt = \int \frac{1}{1+(u)^2}\frac{du}{dt}dt \\ &= \int \frac{1}{1+(u)^2}du = \arctan u + C, \end{aligned}$$

so

$$\int \frac{1}{t^2-6t+10}dt = \arctan(t-3) + C.$$

c) We first substitute $u = 2x$. Since $du/dx = 2$,

$$\begin{aligned} \int \sin^2 2x \cos 2x dx &= \int \sin^2 u \cos u \frac{1}{2}\frac{du}{dx}dx \\ &= \frac{1}{2}\int \sin^2 u \cos u du \end{aligned}$$

At this point, we notice that another substitution is appropriate: we set $s = \sin u$ and $ds/du = \cos u$. Then

$$\begin{aligned}\frac{1}{2}\int \sin^2 u \cos u du &= \frac{1}{2}\int s^2 \frac{ds}{du} du = \frac{1}{2}\int s^2 ds \\ &= \frac{1}{2}\cdot\frac{1}{3}s^3 + C = \frac{1}{6}s^3 + C\end{aligned}$$

Now we must put our answer in terms of x. Since $s = \sin u$ and $u = 2x$, we have

$$\int \sin^2 2x \cos 2x dx = \frac{\sin^3 2x}{6} + C.$$

You should check this formula by differentiating.

d)

$$\int \frac{1}{\sqrt{-7-8x-x^2}} dx = \int \frac{1}{\sqrt{9-(x+4)^2}} dx$$

We set $u = x + 4$; $du/dx = 1$. Then

$$\int \frac{1}{\sqrt{(3)^2 - u^2}} du = \arcsin \frac{1}{3} u + C$$

Now we must put our answer in terms of x, replacing u by $(x + 4)$

$$\int \frac{1}{\sqrt{-7-8x-x^2}} dx = \arcsin \frac{1}{3}(x+4) + C$$

□

You may have noticed that we could have done this problem in one step by substituting $u = \sin 2x$ in the beginning. We did the problem the long way to show that you can solve an integration problem even if you do not see everything at once. □

Two simple substitutions are so useful that they are worth noting explicitly. We have already used them in the preceding examples. The first is the shifting rule, obtained by the substitution $u = x + a$, where a is a constant. Here $du/dx = 1$.

Shifting Rule

To evaluate $\int f(x+a)dx$, first evaluate $\int f(u)du$, then substitute $x + a$ for u

$$\int f(x+a)dx = F(x+a) + C, \qquad \text{where} \quad F(u) = \int f(u)du.$$

The second rule is the scaling rule, obtained by substituting $u = bx$, where b is a constant. Here $du/dx = b$. The substitution corresponds to a change of scale on the x axis.

Scaling Rule

To evaluate $\int f(bx)dx$, evaluate $\int f(u)du$, divide by b, then substitute bx for u

$$\int f(bx)dx = \frac{1}{b}F(bx) + C, \qquad \text{where} \quad F(u) = \int f(u)du.$$

Example 18.15 *Find*

a) $\int \sec^2(x+7)dx$

b) $\int \cos(10x)dx$

Solution:

a) Since $\int \sec^2 u du = \tan u + C$, the shifting rule gives

$$\int \sec^2(x+7)dx = \tan(x+7) + C.$$

b) Since $\int \cos u du = \sin u + C$, the scaling rule gives

$$\int \cos(10x)dx = \frac{1}{10}\sin(10x) + C.$$

□

You do not need to memorize the shifting and scaling rules as such; however, the underlying substitutions are so common that you should learn to use them rapidly and accurately.

To conclude this section, we shall introduce a useful device called *differential notation*, which makes the substitution process more mechanical. In particular, this notation helps keep track of the constant factors which must be distributed between the $f(u)$ and du/dx parts of the integrand. We illustrate the device with an example before explaining why it works.

Example 18.16 *Find* $\int \frac{x^4+2}{(x^5+10x)^5}dx$.

Solution: We wish to substitute $u = x^5 + 10x$; note that $du/dx = 5x^4 + 10$. Pretending that du/dx is a fraction, we may "solve for dx," writing $dx = du/(5x^4 + 10)$. Now we substitute u for $x^5 + 10x$ and $du/(5x^4 + 10)$ for dx in our integral to obtain

$$\int \frac{x^4+2}{(x^5+10x)^5}dx = \int \frac{x^4+2}{u^5} \cdot \frac{du}{5x^4+10} = \int \frac{x^4+2}{5(x^4+2)}\frac{du}{u^5} = \int \frac{1}{5}\frac{du}{u^5}.$$

Notice that the $(x^4 + 2)$ is cancelled, leaving us an integral in u which we can evaluate:

$$\frac{1}{5}\int \frac{du}{u^5} = \frac{1}{5}\left(-\frac{1}{4u^4}\right) + C = -\frac{1}{20u^4} + C.$$

Substituting $x^5 + 10x$ for u gives

$$\int \frac{x^4+2}{(x^5+10x)^5}dx = -\frac{1}{20(x^5+10x)^4} + C$$

□

Although du/dx is not really a fraction, we can still justify "solving for dx" when we integrate by substitution. Suppose that we are trying to integrate $\int h(x)dx$ by substituting $u = g(x)$. Solving $du/dx = g'(x)$ for dx amounts to replacing dx by $du/g'(x)$ and hence writing

$$\int h(x)dx = \int \frac{h(x)}{g'(x)}du \tag{18.4}$$

Now suppose that we can express $\frac{h(x)}{g'(x)}$ in terms of u, i.e., $\frac{h(x)}{g'(x)} = f(u)$ for some function f. Then we are saying that $h(x) = f(u)g'(x) = f(g(x))g'(x)$, and equation (18.4) just says

$$\int f(g(x))g'(x)dx = \int f(u)du,$$

which is the form of integration by substitution we have been using all along.

Example 18.17 *Find* $\int \left(\frac{e^{1/x}}{x^2}\right) dx.$

Solution 18.18 *Let* $u = 1/x$; $du/dx = -1/x^2$ *and* $dx = -x^2du$, *so*

$$\int \left(\frac{1}{x^2}\right) e^{1/x}dx = \int \left(\frac{1}{x^2}\right) e^u \left(-x^2du\right) = -\int e^u du = -e^u + C$$

and therefore

$$\int \left(\frac{1}{x^2}\right) e^{1/x} dx = -e^{1/x} + C.$$

□

Integration by Substitution
(Differential Notation)

To integrate $\int h(x)dx$ by substitution:

1. Choose a new variable $u = g(x)$.
2. Differentiate to get $du/dx = g'(x)$ and then solve for dx.
3. Replace dx in the integral by the expression found in step 2.
4. Try to express the new integrand completely in terms of u, eliminating x. (If you cannot, try another substitution or another method.)
5. Evaluate the new integral $\int f(u)du$ (if you can).
6. Express the result in terms of x.
7. Check by differentiating

Example 18.19 *Calculate the following integrals*

a) $\int \frac{x^2 + 2x}{\sqrt[3]{x^3 + 3x^2 + 1}} dx$

b) $\int \cos x[\cos(\sin x)]dx$,

c) $\int \left(\frac{\sqrt{1 + \ln x}}{x}\right) dx$

Solution:

a) Let $u = x^3 + 3x^2 + 1$; $du/dx = 3x^2 + 6x$, so $dx = du/(3x^2 + 6x)$ and

$$\begin{aligned} \int \frac{x^2 + 2x}{\sqrt[3]{x^3 + 3x^2 + 1}} dx &= \int \frac{1}{\sqrt[3]{u}} \frac{x^2 + 2x}{3x^2 + 6x} du \\ &= \frac{1}{3} \int \frac{1}{\sqrt[3]{u}} du = \frac{1}{3} \cdot \frac{3}{2} u^{2/3} + C. \end{aligned}$$

Thus

$$\int \frac{x^2 + 2x}{\sqrt[3]{x^3 + 3x^2 + 1}} dx = \frac{1}{2} \left(x^3 + 3x^2 + 1\right)^{2/3} + C.$$

b) Let $u = \sin x$; $du/dx = \cos x$, $dx = du/\cos x$, so

$$\begin{aligned}\int \cos x[\cos(\sin x)]dx &= \int \cos x[\cos(\sin x)]\frac{du}{\cos x} \\ &= \int \cos u du = \sin u + C.\end{aligned}$$

and therefore

$$\int \cos x[\cos(\sin x)]dx = \sin(\sin x) + C.$$

c) Let $u = 1 + \ln x$; $du/dx = 1/x$, $dx = xdu$, so

$$\int \left(\frac{\sqrt{1+\ln x}}{x}\right) dx = \int \left(\frac{\sqrt{1+\ln x}}{x}\right)(xdu) = \int u^{1/2}du = \frac{2}{3}u^{3/2} + C,$$

and therefore

$$\int \left(\frac{\sqrt{1+\ln x}}{x}\right) dx = \frac{2}{3}(1+\ln x)^{3/2} + C.$$

□

18.4 Changing Variables in the Definite Integral

We have just learned how to evaluate many indefinite integrals by the method of substitution. Using the fundamental theorem of calculus, we can use this knowledge to evaluate definite integrals as well.

Example 18.20 *Find* $\int_0^2 \sqrt{x+3}dx$

Solution: Substitute $u = x + 3$, $du = dx$. Then

$$\int \sqrt{x+3}dx = \int \sqrt{u}du = \frac{2}{3}u^{3/2} + C = \frac{2}{3}(x+3)^{3/2} + C.$$

By the fundamental theorem of calculus,

$$\int_0^2 \sqrt{x+3}dx = \frac{2}{3}(x+3)^{3/2}\Big|_0^2 = \frac{10}{3}\sqrt{5} - 2\sqrt{3} \approx: 3.9895.$$

□

Notice that we must express the indefinite integral in terms of x before plugging in the endpoints 0 and 2, since they refer to values of x. It is possible, however, to evaluate the definite integral directly in the u variable-*provided that we change the endpoints*. We offer an example before stating the general procedure.

Example 18.21 *Find* $\int_1^4 \frac{x}{1+x^4}dx$.

Solution: Substitute $u = x^2$, $du = 2xdx$, that is, $xdx = du/2$. As x runs from 1 to 4, $u = x^2$ runs from 1 to 16, so we have

$$\int_1^4 \frac{x}{1+x^4}dx = \frac{1}{2}\int_1^{16} \frac{1}{1+u^2}du = \frac{1}{2}\arctan 16 - \frac{1}{8}\pi \approx 0.36149$$

□

In general, suppose that we have an integral of the form $\int f(g(x))g'(x)dx$. If $F'(u) = f(u)$, then $F(g(x))$ is an antiderivative of $f(g(x))g'(x)$; by the fundamental theorem of calculus, we have

$$\int_a^b f(g(x))g'(x)dx = F(g(b)) - F(g(a)).$$

However, the right-hand side is equal to

$$\int_{g(a)}^{g(b)} f(u)du,$$

so we have the formula

$$\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(u)du. \tag{18.5}$$

Example 18.22 *Evaluate* $\int_0^{\pi/4} \cos(2\theta)d\theta$.

Solution: Let $u = 2\theta$, $d\theta = \frac{1}{2}du$; $u = 0$ when $\theta = 0$, $u = \pi/2$ when $\theta = \pi/4$. Thus

$$\int_0^{\pi/4} \cos(2\theta)d\theta = \frac{1}{2}\int_0^{\pi/2} \cos u du = \frac{1}{2}\sin u \,|_0^{\pi/2} = \frac{1}{2}$$

□

Definite Integral by Substitution

Given an integral $\int_a^b h(x)dx$ and a new variable $u = g(x)$.

1. Substitute $du/g'(x)$ for dx and then try to express the integrand $h(x)/g'(x)$ in terms of u
2. Change the endpoints a and b to $g(a)$ and $g(b)$, the corresponding values of u.

Then

$$\int_a^b h(x)dx = \int_{g(a)}^{g(b)} f(u)du$$

where $f(u) = h(x)/(du/dx)$. Since $h(x) = f(g(x))g'(x)$, this can be written as

$$\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(u)du.$$

Example 18.23 *Evaluate* $\int_1^5 \dfrac{x}{x^4 + 10x^2 + 25}dx$.

Solution: Seeing that the denominator can be written in terms of x^2, we try $u = x^2$, $dx = du/(2x)$; $u = 1$ when $x = 1$ and $u = 25$ when $x = 5$. Thus

$$\int_1^5 \frac{x}{x^4 + 10x^2 + 25}dx = \frac{1}{2}\int_1^{25} \frac{du}{u^2 + 10u + 25}.$$

Now we notice that the denominator is $(u+5)^2$, so we set $v = u+5$, $du = dv$; $u = 6$ when $u = 1$, $v = 30$ when $u = 25$. Therefore

$$\begin{aligned}\frac{1}{2}\int_1^{25} \frac{du}{u^2 + 10u + 25} &= \frac{1}{2}\int_6^{30} \frac{dv}{v^2} = \frac{1}{2}\left(-\frac{1}{v}\right)\Big|_6^{30} \\ &= -\frac{1}{60} + \frac{1}{12} = \frac{1}{15}.\end{aligned}$$

If you see the substitution $v = x^2 + 5$ right away, you can do the problem in one step instead of two. □

Example 18.24 *Evaluate* $\int_0^{\pi/4} \left(\cos^2\theta - \sin^2\theta\right) d\theta$.

Solution: It is not obvious what substitution is appropriate here, so a little trial and error is called for. If we remember the trigonometric identity

$\cos 2\theta = \cos^2\theta - \sin^2\theta$ we can proceed easily:

$$\begin{aligned}\int_0^{\pi/4} \left(\cos^2\theta - \sin^2\theta\right) d\theta &= \int_0^{\pi/4} \cos 2\theta d\theta \\ &= \int_0^{\pi/2} \cos u \frac{du}{2} \\ &= \frac{\sin u}{2}\Big|_0^{\pi/2} \\ &= \frac{1}{2}.\end{aligned}$$

□

Example 18.25 *Evaluate*

$$\int_0^1 \frac{e^x}{1+e^x} dx.$$

Solution: Let $u = 1 + e^x$; $du = e^x dx$, $dx = du/e^x$; $u = 1 + e^0 = 2$ when $x = 0$ and $u = 1 + e$ when $x = 1$. Thus

$$\int_0^1 \frac{e^x}{1+e^x} dx = \int_2^{1+e} \frac{1}{u} dx = \ln u|_2^{1+e} = \ln\left(\frac{1+e}{2}\right)$$

□

Remark 18.26 *Substitution does not always work. We can always make a substitution, but sometimes it leads nowhere.*

18.5 Integration by Parts

"*Integrating the product rule leads to the method of integration by parts.*"

The second of the two important new methods of integration is developed in this section. The method parallels that of substitution, with the chain rule replaced by the product rule.

The product rule for derivatives asserts that

$$(FG)'(x) = F'(x)G(x) + F(x)G'(x)$$

Since $F(x)G(x)$ is an antiderivative for $F'(x)G(x) + F(x)G'(x)$, we can write

$$\int \left[F'(x)G(x) + F(x)G'(x)\right] dx = F(x)G(x) + C.$$

Applying the sum rule and transposing one term leads to the formula

$$\int F(x)G'(x)dx = F(x)G(x) - \int F'(x)G(x)dx + C$$

If the integral on the right-hand side can be evaluated, it will have its own constant C, so it need not be repeated. We thus write

$$\boxed{\int F(x)G'(x)dx = F(x)G(x) - \int F'(x)G(x)dx} \qquad (18.6)$$

which is the formula for integration by parts. To apply formula (18.6) we need to break up a given integrand as a product $F(x)G'(x)$, write down the right-hand side of formula (18.6), and hope that we can integrate $F'(x)G(x)$. Integrands involving trigonometric, logarithmic, and exponential functions are often good candidates for integration by parts, but practice is necessary to learn the best way to break up an integrand as a product.

Example 18.27 *Evaluate* $\int x \sin x dx =$

Solution: If we remember that $\sin x$ is the derivative of $-\cos x$, we can write $x \sin x$ as$F(x)G'(x)$, where $F(x) = x$ and $G(x) = -\cos x$. Applying formula (18.6), we have

$$\begin{aligned} \int x \sin x dx &= x \cdot (-\cos x) - \int 1 \cdot (-\cos x)\, dx \\ &= -x \cos x + \sin x + C. \end{aligned}$$

Checking by differentiation, we have

$$\frac{d}{dx}(-x \cos x + \sin x + C) = x \sin x$$

as required. □

It is often convenient to write formula (18.6) using differential notation. Here we write$u = F(x)$ and $v = G(x)$. Then $du/dx = F'(x)$ and $dv/dx = G'(x)$. Treating the derivatives as if they were quotients of "differentials" du, dv, and dx, we have $du = F'(x)dx$ and $dv = G'(x)dx$. Substituting these into formula (18.6) gives

$$\int u dv = uv - \int v du \qquad (18.7)$$

(see Figure 18.2.)

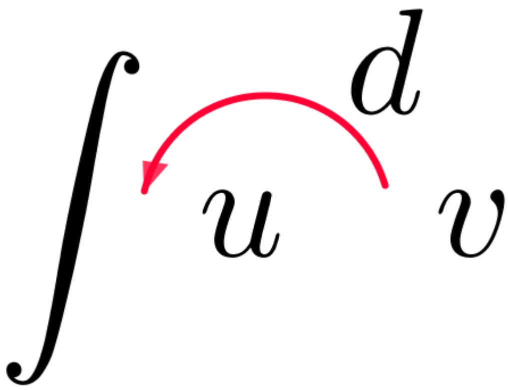


Fig. 18.2. You can move d from v to u if you switch the sign and add uv.

When you use integration by parts, to integrate a function h write $h(x)$ as a product $F(x)G'(x) = udv/dx$; the factor $G'(x)$ is a function whose antiderivative $v = G(x)$ can be found. With a good choice of $u = F(x)$ and $v = G(x)$, the integral $\int F'(x)G(x)dx = \int vdu$ becomes simpler than the original problem $\int udv$. The ability to make good choices of u and v *comes with practice.* A last reminder- *don't forget the minus sign.*

Integration by Parts

To evaluate $\int h(x)dx$ by parts:

1. Write $h(x)$ as a product $F(x)G'(x)$, where the antiderivative $G(x)$ of $G'(x)$ is known.

2. Take the derivative $F'(x)$ of $F(x)$.

3. Use the formula

$$\int F(x)G'(x)dx = F(x)G(x) - \int F'(x)G(x)dx$$

i.e., with $u = F(x)$ and $v = G(x)$,

$$\int udv = uv - \int vdu.$$

Example 18.28 *Find*

a) $\int x \sin x dx$

b) $\int x^2 \sin x dx$

Solution:

a) Let $F(x) = x$ and $G'(x) = \sin x$. Integrating $G'(x)$ gives $G(x) = -\cos x$; also, $F'(x) = 1$, so

$$\begin{aligned} \int x \sin x dx &= -x\cos x - \int -\cos x dx \\ &= -x\cos x - (-\sin x) + C \\ &= \sin x - x\cos x + C. \end{aligned}$$

b) (Using formula 18.7) Let $u = x^2$, $dv = \sin x dx$. To apply formula (18.7) for integration by parts, we need to know v. But $v = \int dv = \int \sin x dx = -\cos x$. (We leave out the arbitrary constant here and will put it in at the end of the problem.)

Now

$$\int x^2 \sin x dx = uv - \int v du$$
$$= -x^2 \cos x - \int (-\cos x) \cdot 2x dx$$
$$= -x^2 \cos x + 2 \int x \cos x dx.$$

Using the same method again, we can find that

$$\int x \cos x dx = \cos x + x \sin x + C.$$

Hence

$$\int x^2 \sin x dx = -x^2 \cos x + 2 (\cos x + x \sin x + C)$$
$$= 2 \cos x - x^2 \cos x + 2x \sin x + C$$

Check this result by differentiating-it is nice to see all the cancellation. □

Integration by parts is also commonly used in integrals involving e^x and $\ln x$.

Example 18.29 *Find*

a) $\int \ln x dx$ *(using integration by parts)*

b) $\int x e^x dx$

Solution:

a) Here, let $u = \ln x$, $dv = 1dx$. Then $du = dx/x$ and $v = \int 1dx = x$. Applying the formula for integration by parts, we have

$$\int \ln x dx = uv - \int v du = (\ln x) x - \int x \frac{dx}{x}$$
$$= x \ln x - \int 1dx = x \ln x - -x + C.$$

b) Let $u = x$ and $v = e^x$, so $dv = e^x dx$. Thus, using integration by parts,

$$\int x e^x dx = \int u dv = uv - \int v du$$
$$= x e^x - \int e^x dx$$
$$= x e^x - e^x + C.$$

□

Next we consider an example involving both e^x and $\sin x$.

Example 18.30 *Apply integration by parts twice to find* $\int e^x \sin x dx$.

Solution: Let $u = \sin x$ and $v = e^x$, so $dv = e^x dx$ and

$$\int e^x \sin x dx = e^x \sin x - \int e^x \cos x dx. \qquad (18.8)$$

Repeating the integration by parts,

$$\int e^x \cos x dx = e^x \cos x + \int e^x \sin x dx, \qquad (18.9)$$

where, this time, $u = \cos x$ and $u = e^x$. Substituting formula (18.9) into (18.8), we get

$$\int e^x \sin x dx = e^x \sin x - \left(e^x \cos x + \int e^x \sin x dx \right).$$

The unknown integral $\int e^x \sin x dx$ appears twice in this equation. Writing "I" for this integral, we have

$$I = e^x \sin x - e^x \cos x - I,$$

and solving for I gives

$$\begin{aligned} I &= \frac{1}{2}\left(e^x \sin x - e^x \cos x\right) + C \\ &= \frac{e^x}{2}\left(\sin x - \cos x\right) + C \end{aligned}$$

□

Using integration by parts and then the fundamental theorem of calculus, we can calculate definite integrals

Example 18.31 *Find* $\int_{-\pi/2}^{\pi/2} x \sin x dx$.

Solution: From Example 18.27 we have $\int x \sin x dx = -x \cos x + \sin x + C$, so

$$\begin{aligned} \int_{-\pi/2}^{\pi/2} x \sin x dx &= \left(-x \cos x + \sin x\right) \Big|_{-\pi/2}^{\pi/2} \\ &= (0 + 1) - [0 + (-1)] \\ &= 2. \end{aligned}$$

□

Example 18.32 *Find*

a) $\int_0^{\ln 2} e^x \ln(e^x + 1)dx$

b) $\int_1^e \sin(\ln x)dx$

Solution:

a) Notice that e^x is the derivative of $(e^x + 1)$, so we first make the substitution $t = e^x + 1$. Then

$$\int_0^{\ln 2} e^x \ln(e^x + 1)dx = \int_2^3 \ln t dt$$

and, from Example 18.29 $\int \ln t dt = t \ln t - t + C$. Therefore

$$\begin{aligned} \int_0^{\ln 2} e^x \ln(e^x + 1)dx &= (t \ln t - t) \vert_2^3 \\ &= 3 \ln 3 - 2 \ln 2 - 1 \approx: 0.90954 \end{aligned}$$

b)

Again we begin with a substitution. Let $u = \ln x$, so that $x = e^u$ and $du = (1/x)dx$. Then $\int \sin(\ln x)dx = \int (\sin u)e^u du$, which was evaluated in Example 18.30 . Hence

$$\begin{aligned} \int_1^e \sin(\ln x)dx &= \int_0^1 \sin u e^u du \\ &= e^u (\sin u - \cos u) \vert_0^1 \\ &= 1 - e (\cos 1 - \sin 1) \end{aligned}$$

□

19

Basic techniques of integration, part two

19.1 Trigonometric Integrals

Besides the basic methods of integration associated with reversing the differentiation rules, there are special methods for integrands of particular forms. Using these methods, we can solve some interesting length and area problems.

fundamental identities:	$\sin^2 x + \cos^2 x = 1; \quad \tan^2 x + 1 = \sec^2 x.$
half angle formulas:	$\sin^2 x = \dfrac{1-\cos 2x}{2}; \quad \cos^2 x = \dfrac{1+\cos 2x}{2}$

The integrals treated in this section fall into two groups. First, there are some purely trigonometric integrals that can be evaluated using trigonometric identities. Second, there are integrals involving quadratic functions and their square roots which can be evaluated using trigonometric substitutions.

We begin by considering integrals of the form

$$\int \sin^m x \cos^n x dx,$$

where m and n are integers. The case $n = 1$ is easy, for if we let $u = \sin x$, we find

$$\int \sin^m x \cos x dx = \int u^m du = \frac{u^{m+1}}{m+1} + C = \frac{\sin^{m+1}(x)}{m+1} + C$$

(or $\ln|\sin x| + C$, if $n = -1$). The case $m = 1$ is similar

$$\int \sin x \cos^n x dx = -\frac{\cos^{n+1}(x)}{n+1} + C$$

(or -$\ln|\cos x| + C$, if $n = -1$). If either m or n is odd, we can use the identity $\sin^2 x + \cos^2 x = 1$ to reduce the integral to one of the types just treated.

Example 19.1 *Evaluate*

$$\int \sin^2 x \cos^3 x dx$$

Solution: $\int \sin^2 x \cos^2 x \cos x dx = \int \sin^2 x \left(1 - \sin^2 x\right) \cos x dx$, which can be integrated by the substitution $u = \sin x$. We get

$$\int u^2(1-u^2)du = \frac{1}{3}u^3 - \frac{1}{5}u^5 + C = \frac{1}{3}\sin^3 x - \frac{1}{5}\sin^5 x + C \quad \square$$

If $m = 2k$ and $n = 21$ are both even, we can use the half-angle formulas $\sin^2 x = (1 - \cos 2x)/2$ and $\cos^2 x = (1 + \cos 2x)/2$ to write

$$\begin{aligned} \int \sin^{2k} x \cos^{2l} x dx &= \int \left(\frac{1-\cos 2x}{2}\right)^k \left(\frac{1+\cos 2x}{2}\right)^l dx \\ &= \frac{1}{2}\int \left(\frac{1-\cos y}{2}\right)^k \left(\frac{1+\cos y}{2}\right)^l dy \end{aligned}$$

where $y = 2x$. Multiplying this out, we are faced with a sum of integrals of the form $\int \cos^m y dy$, with m ranging from zero to $k + l$. The integrals for odd m can be handled by the previous method; to those with even m we apply the half-angle formula once again. The whole process is repeated as often as necessary until everything is integrated.

Example 19.2 *Evaluate* $\int \sin^2 x \cos^2 x dx$.

Solution:

$$\begin{aligned} \int \sin^2 x \cos^2 x dx &= \int \left(\frac{1-\cos 2x}{2}\right) \left(\frac{1+\cos 2x}{2}\right) dx \\ &= \frac{1}{4}\int (1 - \cos^2 2x) dx \\ &= \frac{x}{4} - \frac{1}{4}\int \cos^2 2x dx \\ &= \frac{x}{4} - \frac{1}{4}\int \frac{1+\cos 4x}{2} dx \\ &= \frac{x}{4} - \frac{x}{8} - \frac{1}{8}\int \cos 4x dx \\ &= \frac{x}{8} - \frac{\sin 4x}{32} + C. \quad \square \end{aligned}$$

Trigonometric Integrals

To evaluate $\int \sin^m x \cos^n x dx$:

1. If m is odd, write $m = 2k + 1$, and

$$\int \sin^m x \cos^n x dx = \int \sin^{2k} x \cos^n x \sin x dx$$

$$= \int \left(1 - \cos^2 x\right)^k \cos^n x \sin x dx$$

Now integrate by substituting $u = \cos x$.

2. If n is odd, write $n = 2l + 1$, and

$$\int \sin^m x \cos^n x dx = \int \sin^m x \cos^{2l} x \cos x dx$$

$$= \int \sin^m x \left(1 - \sin^2 x\right)^l x \cos x dx$$

Now integrate by substituting $u = \sin x$.

3. *a*) If m and n are even, write $m = 2k$ and $n = 2l$ and

$$\int \sin^{2k} x \cos^{2l} x dx = \int \left(\frac{1 - \cos 2x}{2}\right)^k \left(\frac{1 + \cos 2x}{2}\right)^l dx$$

Substitute $y = 2x$. Expand and apply step 2 to the odd powers of $\cos y$.

b) Apply step 3(a) to the even powers of $\cos y$ and continue until the integration is completed.

Example 19.3 *Evaluate the integral*

$$\int \sin^3 x \cos^2 x dx$$

Solution:

$$\begin{aligned}
\int \sin^3 x \cos^2 x dx &= \int \sin^2 x \cos^2 x \sin x dx \\
&= \int \left(1-\cos^2 x\right) \cos^2 x \sin x dx \\
&= \int \left(\cos^2 x - \cos^4 x\right) \sin x dx \\
&= \int \left(\cos^4 x - \cos^2 x\right) (-\sin x)\, dx \\
&= \frac{\cos^5 x}{5} - \frac{\cos^3 x}{3} + C. \qquad \square
\end{aligned}$$

Example 19.4 *Evaluate the integral* $\int \frac{\cos^3 x}{\sin x} dx$.

Solution:

$$\begin{aligned}
\int \frac{\cos^3 x}{\sin x} dx &= \int \frac{\cos^2 x}{\sin x} \cos x dx \\
&= \int \frac{1-\sin^2 x}{\sin x} \cos x dx \\
&= \int \frac{1-u^2}{u} du \\
&= \int \left(\frac{1}{u} - u\right) du \\
&= \ln|u| - \frac{1}{2}u^2 + C \\
&= \ln|\sin x| - \frac{1}{2}\sin^2 x + C \qquad \square
\end{aligned}$$

Certain other integration problems yield to the use of the addition formulas:

$$\boxed{\begin{aligned}
\sin(x+y) &= \sin x \cos y + \cos x \sin y, \\
\cos(x+y) &= \cos x \cos y - \sin x \sin y
\end{aligned}}$$

and the product formulas:

$$\boxed{\begin{aligned}
\sin x \cos y &= \tfrac{1}{2}[\sin(x-y) + \sin(x+y)], \\
\sin x \sin y &= \tfrac{1}{2}[\cos(x-y) - \cos(x+y)], \\
\cos x \cos y &= \tfrac{1}{2}[\cos(x-y) + \cos(x+y)].
\end{aligned}}$$

Example 19.5 *Evaluate*

a) $\int \sin x \cos 2x dx$,

b) $\int \cos 3x \cos 5x dx$.

Solution:

a) $\int \sin x \cos 2x dx = \frac{1}{2} \int (\sin 3x - \sin x)\, dx = \frac{1}{2} \cos x - \frac{1}{6} \cos 3x + C$

b) $\int \cos 3x \cos 5x dx = \frac{1}{2} \int (\cos 8x - \cos 2x)\, dx = \frac{1}{4} \sin 2x + \frac{1}{16} \sin 8x + C$

□

Example 19.6 *Evaluate* $\int \sin(5\pi x) \sin(2\pi x)\, dx$.

Solution:

$$\begin{aligned} \int \sin(5\pi x)\sin(2\pi x)\, dx &= \frac{1}{2}\int (\cos(3\pi x) - \cos(7\pi x))\, dx \\ &= \frac{1}{2}\left(\frac{1}{3\pi}\sin(3\pi x) - \frac{1}{7\pi}\sin(7\pi x)\right) + C \\ &= \frac{1}{42\pi}(7\sin(3\pi x) - 3\sin(7\pi x)) + C. \end{aligned}$$ □

Example 19.7 *Evaluate* $\int \sin ax \cos bx dx$ *where* a *and* b *are constants.*

Solution:

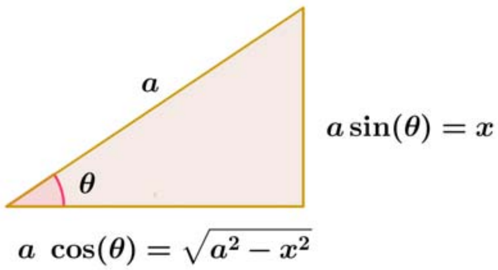
$$\begin{aligned} \int \sin ax \cos bx dx &= \frac{1}{2}\int [\cos(a-b)x - \cos(a+b)x]\, dx \\ &= \begin{cases} \frac{1}{2}\frac{\sin(a-b)x}{a-b} & \text{if} \quad a \neq \pm b \\ \frac{x}{2} - \frac{1}{4a}\sin 2ax + C & \text{if} \quad a = b \\ \frac{1}{4a}\sin 2ax - \frac{x}{2} + C & \text{if} \quad a = -b \end{cases} \end{aligned}$$

□

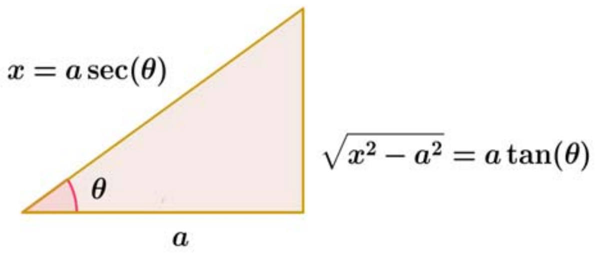
Many integrals containing factors of the form $\sqrt{a^2 \pm x^2}$, $\sqrt{x^2 - a^2}$, or a^2+x^2 can be evaluated or simplified by means of trigonometric substitutions. In order to remember what to substitute, it is useful to draw the appropriate right-angle triangle, as in the following box. From this right triangle we can easily read off all of the six trigonometric functions of θ in terms of x.

Trigonometric Substitutions

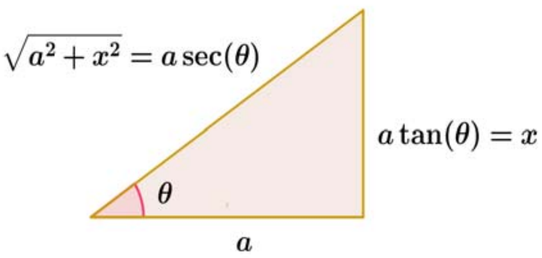
1. If $\sqrt{a^2 - x^2}$ occurs, try $x = a\sin\theta$; then $dx = a\cos\theta d\theta$ and $\sqrt{a^2 - x^2} = a\cos\theta$; ($a > 0$ and θ is an acute angle).



2. If $\sqrt{x^2 - a^2}$ occurs, try $x = a\sec\theta$; then $dx = a\tan\theta\sec\theta d\theta$ and $\sqrt{x^2 - a^2} = a\tan\theta$.



3. If $\sqrt{a^2 + x^2}$ or $a^2 + x^2$ occurs, try $x = a\tan\theta$; then $dx = a\sec^2\theta d\theta$ and $\sqrt{a^2 + x^2} = a\sec\theta$



Example 19.8 (A Sine Substitution) *Evaluate the inyegral* $\int \dfrac{dx}{x^2\sqrt{9 - x^2}}$

Solution: We begin by making the substitution $x = 3\sin\theta$. Then, we have $x^2 = 9sin^2\theta$ and $dx = \cos\theta d\theta$

$$\sqrt{9-x^2} = \sqrt{9-9\sin^2\theta} = 3\sqrt{1-\sin^2\theta} = 3\sqrt{\cos^2\theta} = 3\cos\theta.$$

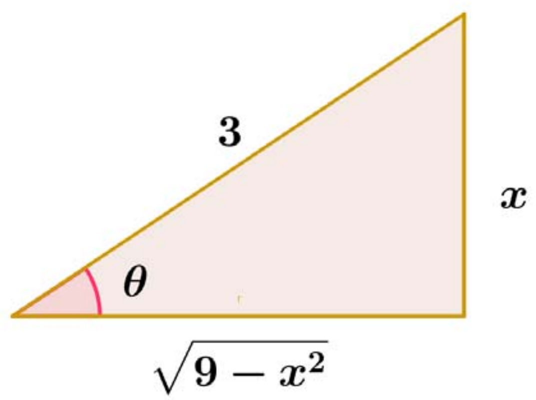
Next, go from the variable x to the variable θ , as follows.

$$\begin{aligned}\int \frac{dx}{x^2\sqrt{9-x^2}} &= \int \frac{3\cos\theta d\theta}{(9sin^2\theta)\,(3\cos\theta)} \\ &= \frac{1}{9}\int \frac{1}{sin^2\theta} d\theta = -\frac{1}{9}\cot\theta + C.\end{aligned}$$

Finally, return to the original variable $x = -\frac{1}{9}\cot\theta = -\frac{1}{9}\frac{\sqrt{9-x^2}}{x} + C$, that is

$$\int \frac{dx}{x^2\sqrt{9-x^2}} = -\frac{1}{9}\frac{\sqrt{9-x^2}}{x} + C$$

You can summarize this substitution by the following useful right triangle.



□

Example 19.9 (A Tangent Substitution) *Evaluate the integral*

$$\int \frac{dx}{(x^2+1)^{3/2}}.$$

Solution: We begin with the substitution $x = a\tan\theta = \tan\theta$. Then, we have

$$dx = \sec^2\theta d\theta, \quad x^2+1 = \tan^2\theta + 1 = \sec^2\theta, \quad (x^2+1)^{3/2} = \sec^3\theta.$$

Moving to the new variable,

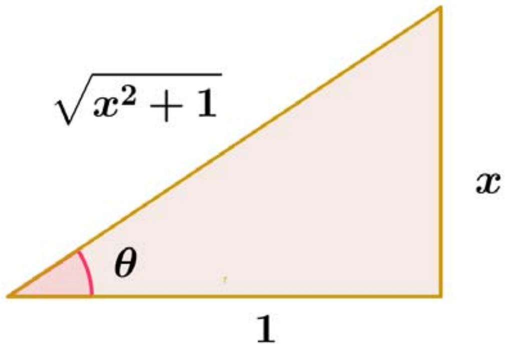
$$\int \frac{dx}{(x^2+1)^{3/2}} = \int \frac{\sec^2\theta d\theta}{\sec^3\theta} = \int \frac{d\theta}{\sec\theta} = \int \cos\theta d\theta = \sin\theta + C.$$

Converting back to the original variable,

$$\sin\theta + C = \frac{x}{\sqrt{x^2+1}} + C.$$

In other words,

$$\int \frac{dx}{(x^2+1)^{3/2}} = \frac{x}{\sqrt{x^2+1}} + C.$$



□

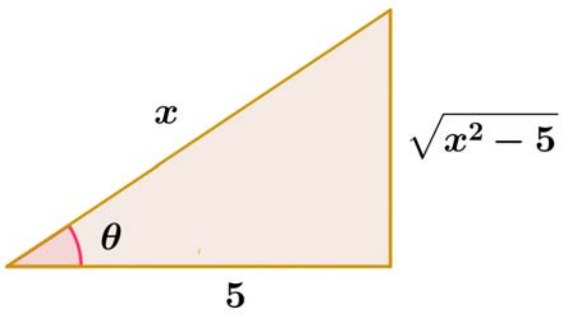
Example 19.10 (A secant Substitution) *Evaluate the integral* $\int \frac{dx}{\sqrt{x^2-25}}$.

Solution: We begin with the substitution $x = 5\sec\theta$. Then, we have

$$dx = 5\sec\theta\tan\theta d\theta, \quad \sqrt{x^2-25} = \sqrt{\sec^2\theta - 25} = 5\sqrt{\sec^2\theta - 1} = 5\tan\theta.$$

With this substitution, we have

$$\begin{aligned}\int \frac{dx}{\sqrt{x^2-25}} &= \int \frac{5\sec\theta\tan\theta dx}{5\tan\theta} = \int \sec\theta d\theta \\ &= \ln|\sec\theta + \tan\theta| + C = \ln\left|\frac{x}{5} + \frac{\sqrt{x^2-25}}{5}\right| + C.\end{aligned}$$



□

19.2 Even Powers by Parts

Example 19.11 *Integrating by parts we have*

$$\int \sin^2 x dx = -\sin x \cos x + \int \cos^2 x dx$$

where

$$\begin{array}{cc} u = \sin x & dv = \sin x dx \\ du = \cos x dx & v = -\cos x \end{array}$$

So

$$\begin{aligned} \int \sin^2 x dx &= -\sin x \cos x + \int \cos^2 x dx \\ &= -\sin x \cos x + \int \left(1 - \sin^2\right) x dx \\ &= -\sin x \cos x + x - \int \sin^2 x dx \end{aligned}$$

Now, the same integral appears on the both equality sizes, hence

$$\int \sin^2 x dx = \frac{1}{2}(-\sin x \cos x + x) + C. \qquad \square$$

In a similar way we can obtain the following Reduction Formulas:

Reduction Formulas

$\int \cos^n x dx = \frac{1}{n}\cos^{n-1} x \sin x + \frac{n-1}{n}\int \cos^{n-2} x dx$

$\int \sin^n x dx = -\frac{1}{n}\sin^{n-1} x \cos x + \frac{n-1}{n}\int \sin^{n-2} x dx$

Example 19.12 *Evaluate* $\int \cos^4 x \sin^2 x dx$.

Solution: Using identity

$$\sin^2 x = 1 - \cos^2 x,$$

we have

$$\int \cos^4 x \sin^2 x dx == \int \cos^4 x dx - \int \cos^6 x dx.$$

Next, from the reduction formula

$$\int \cos^n x dx = \frac{1}{n}\cos^{n-1} x \sin x + \frac{n-1}{n}\int \cos^{n-2} x dx$$

(applied for $n = 6$, 4, and 2) we have:

$$\begin{aligned}
&\int \cos^4 x \sin^2 x dx \\
&= \int \cos^4 x dx - \int \cos^6 x dx \\
&= \int \cos^4 x dx - \frac{1}{6}\cos^5 x \sin x - \frac{5}{6}\int \cos^4 x dx \\
&= \frac{1}{6}\left(-\cos^5 x \sin x + \int \cos^4 x dx\right) \\
&= \frac{1}{6}\left(-\cos^5 x \sin x + \cos^3 x \sin x + \frac{3}{4}\int \cos^2 x dx\right) \\
&= \frac{1}{6}\left(-\cos^5 x \sin x + \cos^3 x \sin x + \frac{3}{4}\left(\frac{1}{2}\cos x \sin x + \frac{1}{2}x\right)\right) + C \\
&= \frac{1}{16}x + \frac{3}{64}\sin 2x - \frac{1}{192}\sin 6x + C \qquad \square
\end{aligned}$$

Example 19.13 *Let R be the region bounded by the graph of $f(x) = \sin^2(x)$ and the x-axis for $0 \le x \le \pi$ (see Figure 19.1 and 19.2). Find the volume of the solid that is generated by revolving R about the x-axis.*

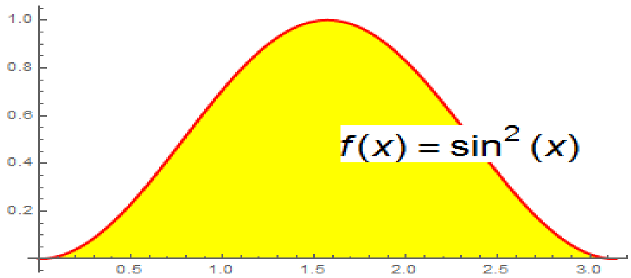


Fig. 19.1. The graph of the function $f(x) = \sin^2(x)$.

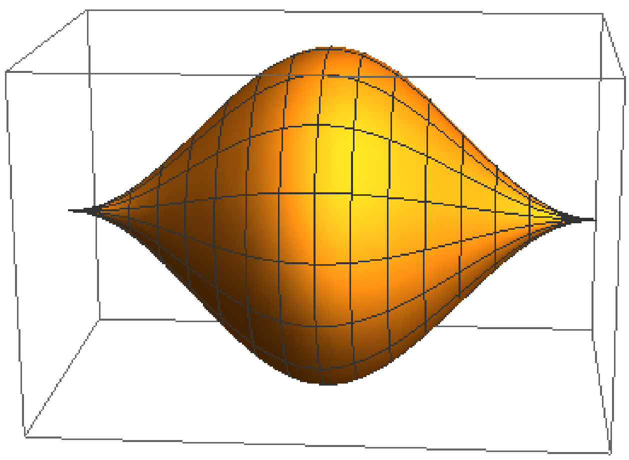


Fig. 19.2. A solid of revolution from Example 19.13

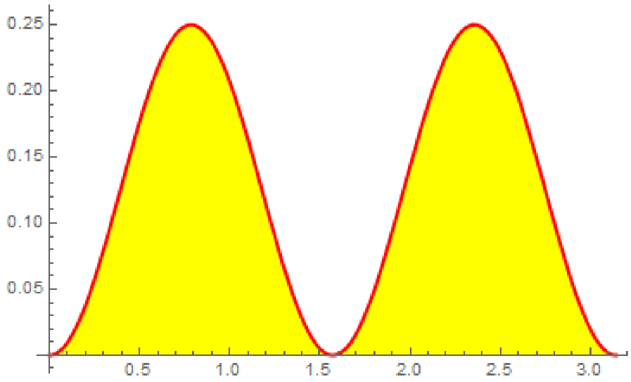


Fig. 19.3. The graph of the function $f(x) = \sin^2 x \cos^2 x$.

Solution: The volume V of the solid of revolution is:

$$\begin{aligned} V &= \int_0^{\pi} \pi \left(\sin^2 x\right)^2 dx \\ &= \int_0^{\pi} \pi \left(\frac{1-\cos 2x}{2}\right)^2 dx \\ &= \int_0^{\pi} \frac{\pi}{4} \left(1 - 2\cos 2x + \cos^2 2x\right) dx \\ &= \frac{\pi}{4} \int_0^{\pi} \left(1 - 2\cos 2x + \frac{1+\cos 4x}{2}\right) dx \\ &= \frac{\pi}{4} \int_0^{\pi} \left(\frac{1}{2}\cos 4x - 2\cos 2x + \frac{3}{2}\right) dx \\ &= \frac{1}{4}\pi \left(\frac{3}{2}x - \sin 2x + \frac{1}{8}\sin 4x\right) |_0^{\pi} \\ &= \frac{3}{8}\pi^2 \qquad \square \end{aligned}$$

Example 19.14 ***Example 19.15*** *Find the average value of* $f(x) = \sin^2 x \cos^2 x$ *on the interval* $[0, 2\pi]$ *(see Figure 19.3).*

Solution: By definition, the average value is the integral divided by the length of the interval:

$$\frac{1}{2\pi} \int_0^{2\pi} \sin^2 x \cos^2 x.$$

By Example 19.1

$$\int \sin^2 x \cos^2 x = \frac{1}{3}\sin^3 x - \frac{1}{5}\sin^5 x + C.$$

Thus

$$\int_0^{2\pi} \sin^2 x \cos^2 x = \left(\frac{1}{3}\sin^3 x - \frac{1}{5}\sin^5 x\right)\Big|_0^{2\pi} = \frac{\pi}{4}$$

so the average value is

$$\frac{1}{2\pi}\cdot\frac{\pi}{4} = \frac{1}{8} \qquad \square$$

19.3 Completing the square

Completing the square can be useful in simplifying integrals involving the expression $ax^2 + bx + c$. The following two examples illustrate the method.

Example 19.16 *Evaluate* $\int \frac{dx}{\sqrt{10+4x-x^2}}$.

Solution: To complete the square, write $10 + 4x - x^2 = -(x+a)^2 + b$; solving for a and b, we find $a = -2$ and $b = 14$, so $10+4x-x^2 = -(x-2)^2+14$. Hence

$$\int \frac{dx}{\sqrt{10+4x-x^2}} = \int \frac{dx}{\sqrt{14-(x-2)^2}} = \int \frac{du}{\sqrt{14-u^2}}$$

where $u = x - 2$. This integral is

$$\arcsin\left(\frac{u}{\sqrt{14}}\right) + C,$$

so our final answer is

$$\arcsin\left(\frac{x-2}{\sqrt{14}}\right) + C. \qquad \square$$

Completing the square

If an integral involves $ax^2 + bx + c$, complete the square and then use a trigonometric substitution or some other method to evaluate the integral.

Example 19.17 *Evaluate*

a) $\int \frac{dx}{x^2+x+1}$,

b) $\int \frac{dx}{\sqrt{x^2+x+1}}$.

Solution:

a)

$$\begin{aligned}\int \frac{dx}{x^2+x+1} &= \int \frac{dx}{\left(x+\frac{1}{2}\right)^2+\frac{3}{4}} \\ &= \int \frac{du}{u^2+\frac{3}{4}} \qquad (u = x+\frac{1}{2}) \\ &= \frac{1}{\sqrt{3/4}} \arctan\left(\frac{u}{\sqrt{3/4}}\right)+C \\ &= \frac{2}{\sqrt{3}} \arctan\left(\frac{2x+1}{\sqrt{3}}\right)+C.\end{aligned}$$

b)

$$\begin{aligned}\int \frac{dx}{\sqrt{x^2+x+1}} &= \int \frac{du}{\sqrt{u^2+3/4}} \qquad (u = x+\frac{1}{2}) \\ &= \ln\left|u+\sqrt{u^2+3/4}\right|+C \\ &= \ln\left|x+\frac{1}{2}+\sqrt{\left(x+\frac{1}{2}\right)^2+3/4}\right|+C \\ &= \ln\left|x+\frac{1}{2}+\sqrt{\left(x+\frac{1}{2}\right)^2+3/4}\right|+C \\ &= \ln\left|x+\frac{1}{2}+\sqrt{x^2+x+1}\right|+C \qquad \square\end{aligned}$$

19.4 Partial Fractions

By the method of partial fractions, one can evaluate any integral of the form $\int \frac{P(x)}{Q(x)}dx$*, where* P *and* Q *are polynomials.*

The integral of a polynomial can be expressed simply by the formula

$$\begin{aligned}&\int \left(a_n x^n + a_{n-1}x^{n-1} + .. + a_1 x + a_0\right) dx \\ = \;& \frac{a_n x^{n+1}}{n+1} + \frac{a_{n-1}x^n}{n} + ... + \frac{a_1 x^2}{2} + a_0 x + C\end{aligned}$$

but there is no simple general formula for integrals of quotients of polynomials, i.e., for rational functions. There is, however, a general method for integrating rational functions, which we shall learn in this section. This method demonstrates clearly the need for evaluating integrals by hand or by a computer program such as Mathematica ot Maple, which automatically carries out the procedures to be described in this section, since tables cannot include the infinitely many possible integrals of this type.

One class of rational functions which we can integrate simply are the reciprocal powers. Using the substitution $u = ax + b$, we find that

$$\int \frac{dx}{(ax+b)^n} = \int \frac{du}{dx},$$

which is evaluated by the power rule. Thus, we get

$$\int \frac{dx}{(ax+b)^n} = \begin{cases} \dfrac{-1}{a(n-1)(ax+b)^{n-1}} + C, & \text{if} \quad n \neq 1, \\[2ex] \dfrac{1}{a} \ln |ax+b| + C & \text{if} \quad n = 1. \end{cases}$$

More generally, we can integrate any rational function whose denominator can be factored into linear factors. We shall give several examples before presenting the general method.

Example 19.18 *Evaluate*

$$\int \frac{x+1}{(x-1)(x-3)} dx$$

Solution: We shall try to write

$$\frac{x+1}{(x-1)(x-3)} = \frac{A}{(x-1)} + \frac{B}{(x-3)},$$

for constants A and B. To determine them, note that

$$\frac{A}{(x-1)} + \frac{B}{(x-3)} = \frac{(A+B)x - 3A - B}{(x-1)(x-3)}$$

Thus, we should choose

$$A + B = 1 \qquad \text{and} \qquad -3A - B = 1.$$

Solving system of two linear equations with two unknowns we get, $A = -1$ and $B = 2$. Thus,

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

so

$$\begin{aligned}\int \frac{x+1}{(x-1)(x-3)}dx &= 2\ln|x-3| - \ln|x-1| + C \\ &= \ln\left(\frac{|x-3|^2}{|x-1|}\right) + C\end{aligned}$$

□

Example 19.19 *Evaluate*

a) $\int \frac{4x^2+2x+3}{(x-2)^2(x+3)}dx,$

b) $\int_{-1}^{1} \frac{4x^2+2x+3}{(x-2)^2(x+3)}dx.$

a) As in Example 19.18, we might expect to decompose the quotient in terms of $\frac{1}{(x-2)}$ and $\frac{1}{(x+3)}$. In fact, we shall see that we can write

$$\frac{4x^2+2x+3}{(x-2)^2(x+3)} = \frac{A}{(x-2)} + \frac{B}{(x-2)^2} + \frac{C}{(x+3)} \tag{19.1}$$

if we choose the constants A, B, and C suitably. Adding the terms on the right-hand side of equation (1) over the common denominator, we get

$$\frac{A(x-2)(x-3) + B(x+3) + C(x-2)^2}{(x-2)^2(x+3)}$$

The numerator, when multiplied out, would be a polynomial $a_2x^2+a_1x+a_0$, where the coefficients a_2, a_1, and a$_0$ depend on A, B, and C. The idea is to choose A, B, and C so that we get the numerator $4x^2+2x+3$ of our integration problem. (Notice that we have exactly three unknowns A, B, and C at our disposal to match the three coefficients in the numerator.)

To choose A, B, and C, it is easiest not to multiply out but simply to write

$$4x^2+2x+3 = A(x-2)(x+3) + B(x+3) + C(x-2)^2 \tag{19.2}$$

and make judicious substitutions for x. For instance, $x = -3$ gives

$$C = \frac{33}{25}.$$

Next, $x = 2$ gives

$$B = \frac{23}{5}.$$

To solve for A, we may use either of two methods.

Method 1. Let $x = 0$ in equation (19.2):

$$\begin{aligned} 3 &= -6A + 3B + 4C \\ 3 &= -6A + 3\frac{23}{5} + 4\frac{33}{25}, \\ A &= \frac{67}{25}. \end{aligned}$$

Method 2. Differentiate equation (19.2) to give

$$8x + 2 = A[(x-2) + (x+3)] + B + 2C(x-2)$$

and then substitute $x = 2$ again:

$$\begin{aligned} 8 \cdot 2 + 2 &= A(2+3) + B \\ 18 &= 5A + \frac{23}{5} \\ A &= \frac{67}{25}. \end{aligned}$$

This gives

$$\frac{4x^2 + 2x + 3}{(x-2)^2(x+3)} = \frac{67}{25}\frac{1}{(x-2)} + \frac{23}{5}\frac{1}{(x-2)^2} + \frac{33}{25}\frac{1}{(x+3)}$$

(At this point, it is a good idea to check your answer, either by adding up the right-hand side or by substituting a few values of x, using a calculator.)

We can now integrate:

$$\begin{aligned} &\int \frac{4x^2 + 2x + 3}{(x-2)^2(x+3)} dx \\ &= \frac{67}{25}\int \frac{dx}{(x-2)} + \frac{23}{5}\int \frac{dx}{(x-2)^2} + \frac{33}{25}\int \frac{dx}{(x+3)} \\ &= \frac{67}{25}\ln|x-2| - \frac{23}{5}\frac{1}{x-2} + \frac{33}{25}\ln|x+3| + C. \end{aligned}$$

b)

Since the integrand "blows up" at $x = -3$ and $x = 2$, it only makes sense to evaluate definite integrals over intervals which do not contain these points; $[-1, 1]$ is such an interval. Thus, by (a), the definite integral is

$$\left(\frac{67}{25}\ln|x-2| - \frac{23}{5}\frac{1}{x-2} + \frac{33}{25}\ln|x+3|\right)|_{-1}^{1}$$
$$= \frac{33}{25}\ln 4 - \frac{67}{25}\ln 3 - \frac{33}{25}\ln 2 + \frac{46}{15} \approx 1.0373$$

□

Not every polynomial can be written as a product of linear factors. For instance, $x^2 + 1$ cannot be factored further (unless we use complex numbers) nor can any other quadratic function $ax^2 + bx + c$ for which $b^2 - 4ac < 0$; but any polynomial can, in principle, be factored into linear and quadratic factors. (This is proved in more advanced algebra texts.) This factorization is not always so easy to carry out in practice, but whenever we manage to factor the denominator of a rational function, we can integrate that function by the method of partial fractions.

Example 19.20 *Integrate* $\int \frac{1}{x^3-1}dx$.

Solution: The denominator factors as $(x-1)(x^2+x+1)$, and x^2+x+1 cannot be further factored (since $b^2 - 4ac = 1 - 4 = -3 < 0$). Now write

$$\frac{1}{x^3-1} = \frac{a}{x-1} + \frac{Ax+B}{x^2+x+1}$$

Thus

$$1 = a\left(x^2+x+1\right) + (x-1)(Ax+B).$$

We substitute values for x:

$$x = 1; \quad 1 = 3a \quad \text{so} \quad a = \frac{1}{3};$$
$$x = 0; \quad 1 = \frac{1}{3} - B \quad \text{so} \quad B = -\frac{2}{3}.$$

Comparing the x^2 terms, we get $0 = a + A$, so $A = -\frac{1}{3}$. Hence

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

(This is a good point to check your work.)

Now

$$\int \frac{dx}{x-1} = \ln|x-1| + C$$

and, writing $x+2 = \frac{1}{2}(2x+1) + \frac{3}{2}$,

$$\begin{aligned}\int \frac{x+2}{x^2+x+1}dx &= \frac{1}{2}\int \frac{2x+1}{x^2+x+1}dx + \frac{3}{2}\int \frac{dx}{(x+1/2)^2+3/4} \\ &= \frac{1}{2}\ln\left|x^2+x+1\right| + \frac{3}{2}\sqrt{\frac{4}{3}}\arctan\left(\frac{x+1/2}{\sqrt{3/4}}\right) + C \\ &= \frac{1}{2}\ln\left|x^2+x+1\right| + \sqrt{3}\arctan\left(\frac{2x+1}{\sqrt{3}}\right) + C\end{aligned}$$

Thus

$$\begin{aligned}\int \frac{1}{x^3-1}dx &= \frac{1}{3}\ln|x-1| - \frac{1}{6}\ln\left|x^2+x+1\right| - \frac{1}{\sqrt{3}}\arctan\left(\frac{2x+1}{\sqrt{3}}\right) + C \\ &= \frac{1}{3}\left[\frac{1}{2}\ln\left|\frac{(x-1)^2}{x^2+x+1}\right|\right] - \sqrt{3}\arctan\left(\frac{2x+1}{\sqrt{3}}\right) + C.\end{aligned}$$

Observe that the innocuous-looking integrand $\frac{1}{x^3-1}$ has brought forth both logarithmic and trigonometric functions.

Now we are ready to set out a systematic method for the integration of $P(x)/Q(x)$ by partial fractions. (See the box on p. 456.) A few remarks may clarify the procedures given in the box. In case the denominator Q factors into n distinct linear factors, which we denote $Q = (x-r_1)(x-r_2)...(x-r_n)$, we write

$$\frac{P}{Q} = \frac{\alpha_1}{x-r_1} + \frac{\alpha_2}{x-r_2} + ... + \frac{\alpha_n}{x-r_n}$$

and determine the n coefficients $\alpha_1, ..., \alpha_n$ by multiplying by Q and matching P to the resulting polynomial. The division in step 1 has guaranteed that P has degree at most $n-1$, containing n coefficients. This is consistent with the number of constants $\alpha_1, ..., \alpha_n$, we have at our disposal. Similarly, if the denominator has repeated roots, or if there are quadratic factors in the denominator, it can be checked that the number of constants at our disposal is equal to the number of coefficients in the numerator to be matched. A system of n equations in n unknowns is likely to have a unique solution, and in this case, one can prove that it does.

Partial Fractions

To integrate $P(x)/Q(x)$, where P and Q are polynomials containing no common factor:

1. If the degree of P is larger than or equal to the degree of Q, divide Q into P by long division, obtaining a polynomial plus $R(x)/Q(x)$, where the degree of R is less than that of Q. Thus we need only investigate the case where the degree of P is less than that of Q.

2. Factor the denominator Q into linear and quadratic factors-that is, factors of the form $(x-r)$ and ax^2+bx+c. (Factor the quadratic expressions if $b^2-4ac>0$.)

3. If $(x-r)^m$ occurs in the factorization of Q, write down a sum of the

$$\frac{a_1}{(x-r)} + \frac{a_2}{(x-r)^2} + ... + \frac{a_m}{(x-r)^m},$$

where $a_1, a_2, ..., a_n$ are constants. Do so for each factor of this form (using constants $b_1, b_2, ..., c_1, c_2, ...$, and so on) and add the expressions you get. The constants $a_1, a_2, ..., b_1, b_2, ...$ will be determined in step 5.

4. If $(ax^2+bx+c)^p$ occurs in the factorization of Q with $b^2-4ac<0$, write down a sum of the form

$$\frac{A_1x+B_1}{ax^2+bx+c} + \frac{A_2x+B_2}{(ax^2+bx+c)^2} + ... + \frac{A_px+B_p}{(ax^2+bx+c)^p}$$

Do so for each factor of this form and add the expressions you get The constants $A_1, A_2, ..., B_1, B_2, ...$ are determined in step 5. Add

.

this expression to the one obtained in step 3.

5. Equate the expression obtained in steps 3 and 4 to $P(x)/Q(x)$. Multiply through by $Q(x)$ to obtain an equation between two polynomials. Comparing coefficients of these polynomials, determine equations for the constants $a_1, a_2, \ldots$, $A_1, A_2, \ldots, B_1, B_2, \ldots$ and solve these equations. Sometimes the constants can be determined by substituting convenient values of x in the equality or by differentiation of the equality.

6. Check your work by adding up the partial fractions or substituting a few values of x.

7. Integrate the expression obtained in step 5 by using

$$\int \frac{dx}{(x-r)^j} = -\left[\frac{1}{(j-1)(x-r)^{j-1}}\right] + C, \qquad j > 1$$

and

$$\int \frac{dx}{(x-r)} = \ln|x-r| + C.$$

The terms with a quadratic denominator may be integrated by a manipulation which makes the derivative of the denominator appear in the numerator, together with completing the square.

Example 19.21 (Repeated, nonreducible quadratic factor.) *Calculate*

$$\int \frac{x^3}{(x^2+2x+2)^2} dx.$$

Solution: It will be easier to compute if we complete the square and write the integral as

$$\int \frac{x^3}{(x^2+2x+2)^2} dx = \int \frac{x^3}{\left((x+1)^2+1\right)^2} dx.$$

Now we can take a substitution

$$u = x+1, \qquad du = dx,$$

and write

$$\begin{aligned} & \int \frac{x^3}{\left((x+1)^2+1\right)^2} dx \\ = & \int \frac{(u-1)^3}{(u^2+1)^2} du \end{aligned}$$

Now, let us perform the partial fractions expansion

$$\begin{aligned} \frac{(u-1)^3}{(u^2+1)^2} &= \frac{u^3-3u^2+3u-1}{(u^2+1)^2} \\ &= \frac{Au+B}{(u^2+1)^2} + \frac{Cu+D}{(u^2+1)} \end{aligned}$$

So, we must have

$$u^3-3u^2+3u-1 = (Au+B) + \left(u^2+1\right)(Cu+D)$$

and

$$u^3-3u^2+3u-1 = Cu^3 + u^2D + (A+C)\,u + B + D.$$

From that it follows, that

$$C=1, \qquad D=-3, \qquad , A=2, \qquad B=2,$$

hence

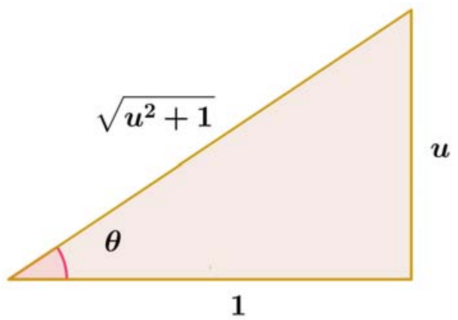
$$\frac{(u-1)^3}{(u^2+1)^2} = \frac{2u+2}{(u^2+1)^2} + \frac{u-3}{(u^2+1)},$$

and our integral becomes

$$\begin{aligned} \int \frac{(u-1)^3}{(u^2+1)^2} du &= \int \left(\frac{2u}{(u^2+1)^2} + \frac{2}{(u^2+1)^2} + \frac{u}{u^2+1} - \frac{3}{u^2+1} \right) du \\ &= -\frac{1}{u^2+1} + \int \frac{2}{(u^2+1)^2} du + \frac{1}{2} \ln\left(u^2+1\right) - 3\arctan u. \end{aligned}$$

Now we have everything except the integral of the second term. Let us compute that using the trigonometric substitution.

$$\begin{aligned}
&\int \frac{2}{(u^2+1)^2} du \\
=\ &\int \frac{2\sec^2\theta}{\sec^4\theta} d\theta \qquad u=\tan\theta, \quad du=\sec^2\theta d\theta \\
=\ &\int 2\cos^2\theta d\theta = \theta + \sin\theta\cos\theta + C \\
=\ &\arctan u + \frac{u}{u^2+1} + C.
\end{aligned}$$



Therefore, our original u integral can be presented as

$$\begin{aligned}
&\int \frac{(u-1)^3}{(u^2+1)^2} du \\
=\ &-\frac{1}{u^2+1} + \frac{u}{u^2+1} + \frac{1}{2}\ln\left(u^2+1\right) + \arctan u - 3\arctan u + C \\
=\ &\frac{u-1}{u^2+1} + \frac{1}{2}\ln\left(u^2+1\right) - 2\arctan u + C.
\end{aligned}$$

Now, the only thing to do is replace u with $x+1$. So, our original integral becomes

$$\begin{aligned}
&\int \frac{x^3}{(x^2+2x+2)^2} dx \\
=\ &\frac{x}{(x+1)^2+1} + \frac{1}{2}\ln\left((x+1)^2+1\right) - 2\arctan(x+1) + C \\
=\ &\frac{x}{x^2+2x+2} + \frac{1}{2}\ln\left(x^2+2x+2\right) - 2\arctan\left(x^2+2x+2\right) + C.
\end{aligned}$$

□

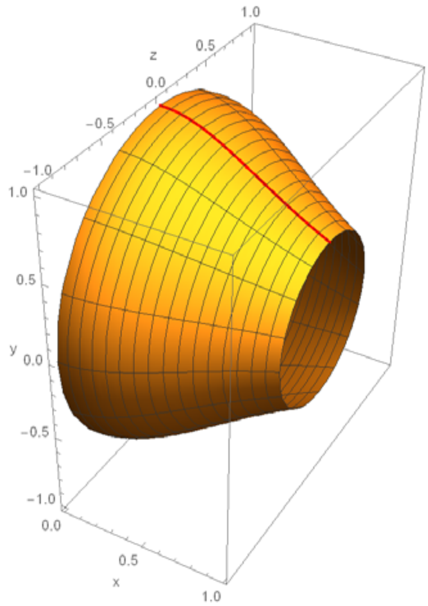


Fig. 19.4. Solid generated by revolving R about the x-axis (in Example 19.22).

Example 19.22 *Ler R be the region bounded by the graph of $f(x) = \frac{1}{1+x^2}$ and the x-axis for $0 \leq x \leq 1$.*

a) *Find the volume of the solid generated by revolving R about the x-axis (see Figure 19.4),*

b) *Find the volume of the solid generated by revolving R about the y-axis (see Figure 19.5),*

Solution:

a) The volume of the solid generated by revolving R about the x-axis is

$$V = \int_0^1 \pi \left(\frac{1}{1+x^2} \right)^2 dx.$$

We will calculate the indefinite integral $\int \pi \left(\frac{1}{1+x^2} \right)^2 dx$ first and then use it to evaluate the definite integral.

Let

$$\begin{aligned} x &= \tan\theta \\ dx &= \sec^2\theta d\theta \\ \sec\theta &= \sqrt{1+x^2}. \end{aligned}$$

Then

$$\begin{aligned} \int \pi\left(\frac{1}{1+x^2}\right)^2 dx &= \int \pi\frac{\sec^2\theta}{\sec^4\theta}d\theta \\ &= \pi\int\frac{1}{\sec^2\theta}d\theta \\ &= \frac{\pi}{2}\int\cos^2\theta d\theta \\ &= \frac{\pi}{2}\int(1+\cos 2\theta)\,d\theta \\ &= \frac{\pi}{2}\left(\theta+\frac{1}{2}\sin 2\theta\right)+C \\ &= \frac{\pi}{2}(\theta+\sin\theta\cos\theta)+C \\ &= \frac{\pi}{2}\left(\arctan x+\frac{x}{\sqrt{1+x^2}}\frac{1}{\sqrt{1+x^2}}\right)+C \\ &= \frac{\pi}{2}\left(\arctan x+\frac{x}{1+x^2}\right)+C. \end{aligned}$$

Now

$$\begin{aligned} V &= \int_0^1 \pi\left(\frac{1}{1+x^2}\right)^2 dx \\ &= \left[\frac{\pi}{2}\left(\arctan x+\frac{x}{1+x^2}\right)\right]_0^1 \\ &= \frac{1}{2}\pi\left(\frac{1}{4}\pi+\frac{1}{2}\right). \end{aligned}$$

b) The volume of the solid generated by revolving R about the y-axis is

$$\begin{aligned} V &= \int_0^1 2\pi x\left(\frac{1}{1+x^2}\right)dx = 2\pi\int_0^1\frac{x}{1+x^2}dx \\ &= \pi\int_1^2\frac{1}{u}du \qquad (\text{where } u=1+x^2, \quad du=2xdx) \\ &= \pi\ln 2 \end{aligned}$$

□

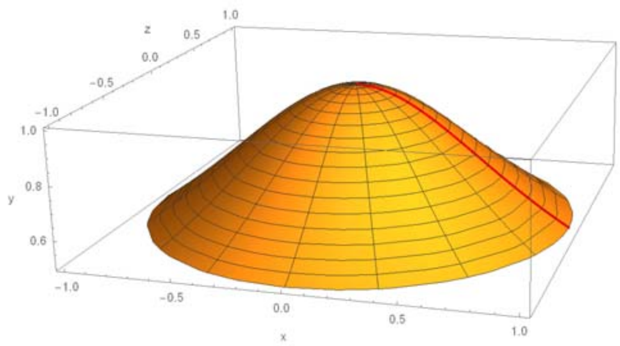


Fig. 19.5. Solid generated by revolving R about the y-axis (in Example 19.22).

19.5 Final Note

The preceding discussion of partial fraction decomposition assumes that $f(x) = P(x)/Q(x)$ is a proper rational function. If this is not the case and we are faced with an improper rational function f, we divide the denominator into the numerator and express f in two parts. One part will be a polynomial, and the other will be a proper rational function.

For example, given the function

$$f(x) = \frac{3x^4 + 5x^3 - 4x^2 + 7x - 1}{x^2 + 2x - 3}$$

we perform long division.

It follows that

$$f(x) = 3x^2 - x + 7 + \frac{-10x + 20}{x^2 + 2x - 3}$$

The first piece (a polynomial) is easily integrated, and the second piece (rational function) now qualifies for the methods described in this section.

20
Taylor polynomials

20.1 Taylor and Maclaurin polynomials

We have seen that approximation is a basic theme in calculus. For example, we may approximate a differentiable function $f(x)$ at $x = a$ by the linear function (called the *linearization*)

$$L(x) = f(a) + f'(a)(x - a)$$

However, a drawback of the linearization is that it is accurate only in a small interval around $x = a$. Taylor polynomials are higher-degree approximations that generalize the linearization using the higher derivatives $f^{(k)}(a)$. They are useful because, by taking sufficiently high degree, we can approximate transcendental functions such as $\sin x$ and e^x to arbitrary accuracy on any given interval.

Assume that $f(x)$ is defined on an open interval I and that all higher derivatives $f^{(k)}(a)$ exist on I . Fix a number $a \in I$. The nth **Taylor polynomial** for f centered at $x = a$ is the polynomial

$$\boxed{T_n(x) = f(a) + \frac{f'(a)}{1!}(x - a) + \frac{f''(a)}{2!}(x - a)^2 + ... + \frac{f^{(n)}(a)}{n!}(x - a)^n}$$

It is convenient to regard $f(x)$ itself as the zeroth derivative $f^{(0)}(x)$. Then we may write the Taylor polynomial in summation notation

$$T_n(x) = \sum_{j=0}^{n} \frac{f^{(j)}(a)}{j!}(x - a)^j.$$

When $a = 0$, $T_n(x)$ is also called the nth **Maclaurin polynomial**. The first few Taylor polynomials are

$$T_0(x) = f(a) \quad \text{(a constant term)}$$

$$T_1(x) = f(a) + \frac{f'(a)}{1!}(x-a)$$

$$T_2(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2$$

$$T_3(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3$$

Note that $T_1(x)$ is the linearization of $f(x)$ at a. In most cases, the higher-degree Taylor polynomials provide increasingly better approximations to $f(x)$. Before computing some Taylor polynomials, we record two important properties that follow from the definition:

- $T_n(a) = f(a)$ [since all terms in $T_n(x)$ after the first are zero at $x = a$].
- $T_n(x)$ is obtained from $T_{n-1}(x)$ by adding on a term of degree n:

$$T_n(x) = T_{n-1}(x) + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

Example 20.1 *Let* $f(x) = \sqrt{x+1}$. *Compute* $T_n(x)$ *at* $a = 3$ *for* $n = 0, 1, 2, 3,$ *and* 4.

$$f(x) = \sqrt{x+1}$$

Solution: First evaluate the derivatives $f^{(j)}(3)$:

$$f(x) = \sqrt{x+1} \qquad f(3) = 2$$

$$f'(x) = \tfrac{1}{2\sqrt{x+1}} \qquad f'(3) = \frac{1}{4}$$

$$f''(x) = -\tfrac{1}{4(x+1)^{\frac{3}{2}}} \qquad f''(3) = -\tfrac{1}{32}$$

$$f'''(x) = \tfrac{3}{8(x+1)^{\frac{5}{2}}} \qquad f'''(3) = \tfrac{3}{256}$$

$$f^{(4)}(x) = -\tfrac{15}{16(x+1)^{\frac{7}{2}}} \qquad f^{(4)}(3) = -\tfrac{15}{2048}$$

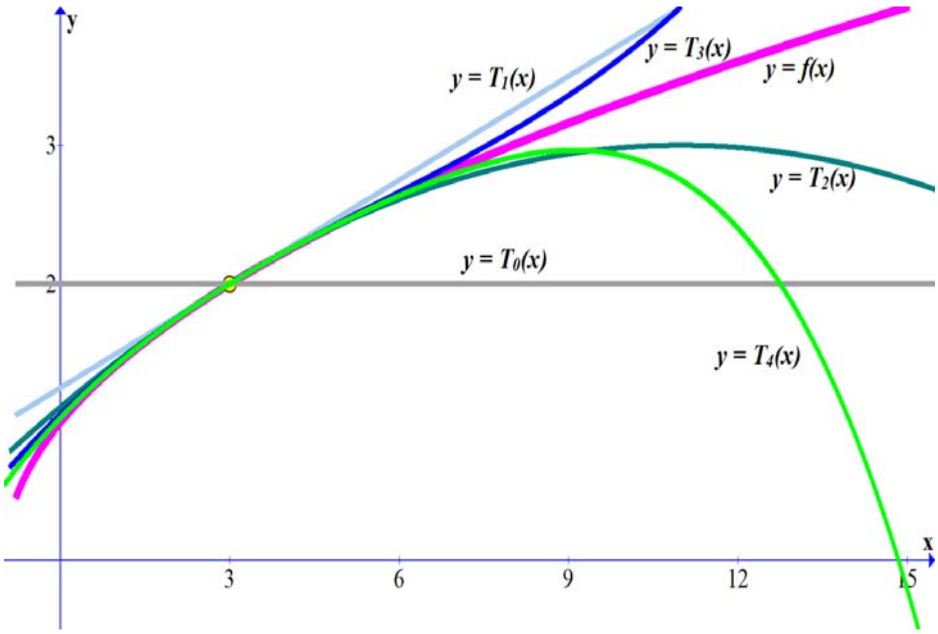


Fig. 20.1. Graph of $f(x) = \sqrt{1+x}$ and its first five Taylor polynomials centered at $x = 3$.

Then compute the coefficients $\dfrac{f^{(j)}(3)}{j!}$

$$\begin{aligned}
\text{Constant term} &= f(3) = 2 \\
\text{Coefficient of } (x-3) &= f'(3) = \frac{1}{4} \\
\text{Coefficient of } (x-3)^2 &= \frac{f''(3)}{2!} = -\frac{1}{64} \\
\text{Coefficient of } (x-3)^3 &= \frac{f'''(3)}{3!} = \frac{1}{512} \\
\text{Coefficient of } (x-3)^4 &= \frac{f^{(4)}(3)}{4!} = -\frac{5}{16\,384}
\end{aligned}$$

The first five Taylor polynomials centered at $x = 3$ are:

$$
\begin{aligned}
T_0(x) &= 2 \\
T_1(x) &= 2 + \frac{1}{4}(x-3) \\
T_2(x) &= 2 + \frac{1}{4}(x-3) - \frac{1}{64}(x-3)^2 \\
T_3(x) &= 2 + \frac{1}{4}(x-3) - \frac{1}{64}(x-3)^2 + \frac{1}{512}(x-3)^3 \\
T_4(x) &= 2 + \frac{1}{4}(x-3) - \frac{1}{64}(x-3)^2 + \frac{1}{512}(x-3)^3 - \frac{5}{16\,384}(x-3)^4
\end{aligned}
$$

□

Example 20.2 (Maclaurin Polynomials for $f(x) = \cos x$**)** *Find the Maclaurin polynomials of* $f(x) = \cos x$*.*

Solution: Recall that the Maclaurin polynomials are the Taylor polynomials centered at $a = 0$. The key observation is that the derivatives of $f(x) = \cos x$ form a pattern that repeats with period 4:

$$f'(x) = -\sin x \qquad f''(x) = -\cos x \qquad f'''(x) = \sin x \qquad f^{(4)}(x) = \sin x$$

and, in general, $f^{(j+4)}(x) = f^{(j)}(x)$. At $x = 0$, the derivatives form the repeating pattern 1, 0, -1, and 0

$f(0)$	$f'(0)$	$f''(0)$	$f'''(0)$	$f^{(4)}(0)$	$f^{(5)}(0)$	$f^{(6)}(0)$	$f^{(7)}(0)$	...
1	0	-1	0	1	0	-1	0	...

In other words, the even derivatives are $f^{(2k)}(0) = (-1)^k$ and the odd derivatives are zero: $f^{(2k+1)}(0) = 0$. Therefore, the coefficient of x^{2k} is $(-1)^k/(2k)!$ and the coefficient of x^{2k+1} is zero. We have

$$
\begin{aligned}
T_0(x) = T_1(x) &= 1 \\
T_2(x) = T_3(x) &= 1 - \frac{1}{2!}x^2 \\
T_4(x) = T_5(x) &= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4
\end{aligned}
$$

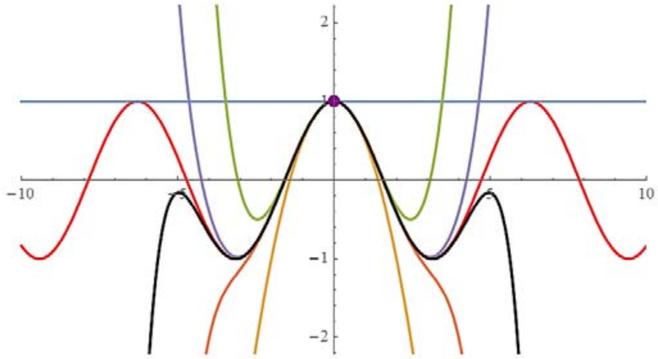


Fig. 20.2. Graphs of Maclaurin polynomials for $f(x) = \cos(x)$. The graph of $f(x)$ is shown in red.

and, in general,

$$T_{2n}(x) = T_{2n+1}(x) = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \frac{1}{6!}x^6 + ... + (-1)^n \frac{1}{(2n)!}x^{2n}$$

□

Taylor polynomials $T_n(x)$ are designed to approximate $f(x)$ in an interval around $x = a$. Figure 20.2 shows the first few Maclaurin polynomials for $f(x) = \cos x$. Observe that as n gets larger, $T_n(x)$ approximates $f(x) = \cos x$ well over larger and larger intervals. Outside this interval, the approximation fails. Notice additionaly, that all of the polynomials $T_n(x)$ are the even functions (as $\cos x$ is).

Figure 20.3 shows the first few Maclaurin polynomials for $f(x) = \sin x$. Observe that again, as n gets larger, $T_n(x)$ approximates $f(x) = \sin x$ well over larger and larger intervals. Outside this interval, the approximation fails. Notice, that this time all of the polynomials $T_n(x)$ are the odd functions (as $\sin x$ is). The reader is asked to calculate the explicit form of several first Maclaurin polynomials for the function $\sin x$.

20.2 The Remainder Term

Our next goal is to study the error $|T_n(x) - f(x)|$ in the approximation provided by the nth Taylor polynomial centered at $x = a$. Define the nth remain-

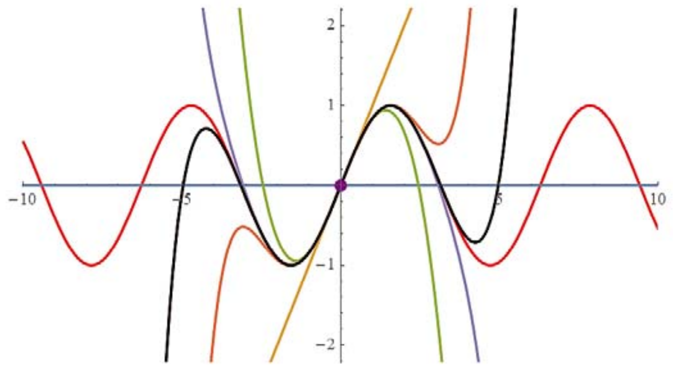


Fig. 20.3. Graphs of Maclaurin polynomials for $f(x) = \sin(x)$. The graph of $f(x)$ is shown in red.

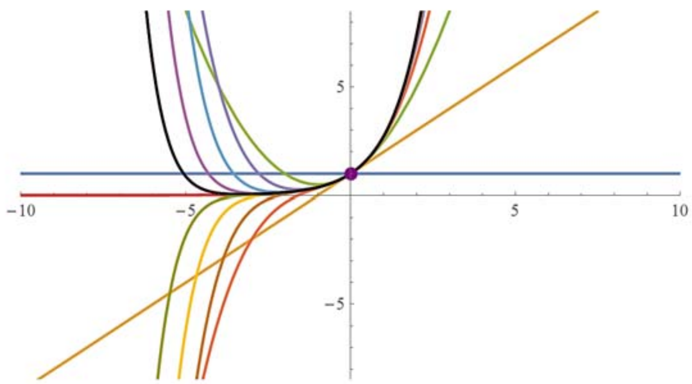


Fig. 20.4. Graphs of Maclaurin polynomials for $f(x) = e^x$. The graph of $f(x)$ is shown in red.

der for $f(x)$ at $x = a$ by

$$\boxed{R_n(x) = f(x) - T_n(x)}$$

The error is the absolute value of the remainder. Also, $f(x) = T_n(x) + R_n(x)$, so

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + ... + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n(x).$$

Theorem 20.3 (Taylor's Theorem) *Assume that $f^{(n+1)}(x)$ exists and is continuous. Then*

$$\boxed{R_n(x) = \frac{1}{n!}\int_a^x (x-u)^n f^{(n+1)}(u)du} \tag{20.1}$$

Proof. Set

$$I_n(x) = \frac{1}{n!}\int_a^x (x-u)^n f^{(n+1)}(u)du$$

Our goal is to show that $R_n(x) = I_n(x)$. For $n = 0$, $R_0(x) = f(x) - f(a)$ and the desired result is just a restatement of the Fundamental Theorem of Calculus:

$$I_0(x) = \int_a^x f^{'}(u)du = f(x) - f(u) = R_0(x).$$

To prove the formula for $n > 0$, we apply Integration by Parts to $I_n(x)$ with

$$h(u) = \frac{(x-u)^n}{n!}, \qquad g(u) = f^{(n)}(u).$$

$$\begin{aligned} I_n(x) &= \int_a^x h(u)g'(u)du \\ &= h(u)g(u)|_a^x - \int_a^x h'(u)g(u)du \\ &= \frac{1}{n!}(x-u)^n f^{(n)}(u)|_a^x - \frac{1}{n!}\int_a^x (-n)\,(x-u)^{n-1} f^{(n)}(u)du \\ &= -\frac{(x-a)^n}{n!}f^{(n)}(a) + I_{n-1}(x). \end{aligned}$$

This result can be rewritten as

$$I_{n-1}(x) = \frac{f^{(n)}(a)}{n!}(x-a)^n + I_n(x).$$

Now apply this relation n times:

$$\begin{aligned}
f(x) &= f(a) + I_0(x) \\
&= f(a) + \frac{f'(a)}{1!}(x-a) + I_1(x) \\
&= f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + I_2(x) \\
&\vdots \\
&= f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + ... + \frac{f^{(n)}(a)}{n!}(x-a)^n + I_n(x)
\end{aligned}$$

This shows that $f(x) = T_n(x) + I_n(x)$ and hence $I_n(x) = R_n(x)$ as desired. ■

Although Taylor's Theorem gives us an explicit formula for the remainder, we will not use this formula directly. Instead, we will use it to *estimate* the size of the error.

Theorem 20.4 (Error Bound) *Assume that $f^{(n+1)}(x)$ exists and is continuous. Let K be a number such that $\left|f^{(n+1)}(u)\right| \leq K$ for all u between a and x. Then*

$$\boxed{|T_n(x) - f(x)| \leq K\frac{|x-a|^{n+1}}{(n+1)!}} \tag{20.2}$$

Proof. Assume that $x \geq a$ (the case $x \leq a$ is similar). Then, since $\left|f^{(n+1)}(u)\right| \leq K$ for $a \leq u \leq x$,

$$|T_n(x) - f(x)| = |R_n(x)| = \left|\frac{1}{n!}\int_a^x (x-u)^n f^{(n+1)}(u)du\right|$$

and on the basis of the Triangle Inequality for Integrals (see p. 258)

$$\begin{aligned}
&\leq \frac{1}{n!}\left|\int_a^x (x-u)^n f^{(n+1)}(u)du\right| \\
&\leq \frac{K}{n!}\int_a^x (x-u)^n(u)du \\
&= \frac{K}{n!}\frac{-(x-u)^{n+1}}{n+1}\Big|_{u=a}^{x} = K\frac{|x-a|^{n+1}}{(n+1)!}
\end{aligned}$$

■

Example 20.5 *Let $T_n(x)$ be the nth Maclaurin polynomial for $f(x) = \cos x$. Find a value of n such that*

$$|T_n(0.2) - \cos(0.2)| < 10^{-5}.$$

Solution: Since $\left|f^{(n)}(x)\right|$ is $|\cos x|$ or $|\sin x|$ depending on whether n is even or odd, we have $\left|f^{(n)}(x)\right| \leq 1$ for all u. Thus, we may apply Error Bound with $K = 1$;

$$|T_n(0.2) - \cos(0.2)| \leq K\frac{|0.2 - 0|^{n+1}}{(n+1)!} = \frac{|0.2|^{n+1}}{(n+1)!}$$

To make the error less than 10^{-5}, we must choose n so that $\frac{|0.2|^{n+1}}{(n+1)!} < 10^{-5}$

n	2	3	4
$\frac{\lvert 0.2\rvert^{n+1}}{(n+1)!}$	$\frac{0.2^3}{3!} \approx 0.00134$	$\frac{0.2^4}{4!} \approx 6.667 \times 10^{-5}$	$\frac{0.2^5}{5!} = 2.67 \times 10^{-6} < 10^{-5}$

We see that the error is less than 10^{-5} for n =4. To verify this, recall that

$$T_4(x) = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4$$

by Example 20.2. The following values from a calculator confirm that the error is significantly less than 10^{-5} as required:

$$\textit{Actual error} = |T_4(0.2) - \cos(0.2)| = |0.98006667 - 0.98006657| \approx 10^{-7}.$$

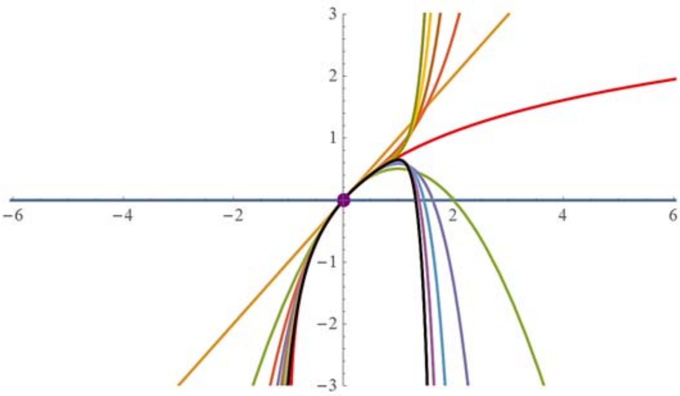


Fig. 20.5. Graphs of Maclaurin polynomials for $f(x) = \ln(x + 1)$. The graph of $f(x)$ is shown in red.

Appendixes

A1. Greek letters used in mathematics, science, and engineering

The Greek letter forms used in mathematics are often different from those used in Greek-language text: they are designed to be used in isolation, not connected to other letters, and some use variant forms which are not normally used in current Greek typography. The table below shows Greek letters rendered in $\TeX$

Table 20.1. Greek letters used in mathematics

α		alpha	ν		nu
β		beta	ξ	Ξ	xi
γ	Γ	gamma	π	Π	pi
δ	Δ	delta	ρ		rho
ϵ		epsilon	σ	Σ	sigma
ζ		zeta	τ		tau
η		eta	υ		upsilon
θ	Θ	theta	ϕ	Φ	phi
ι		iota	χ		chi
κ		kappa	ψ	Ψ	psi
λ	Λ	lambda	ω	Ω	omega
μ		mu	$\dagger$		dagger

$\TeX$is a typesetting system designed and mostly written by Donald Knuth at Stanford and released in 1978.

Together with the Metafont language for font description and the Computer Modern family of typefaces, TeX was designed with two main goals in mind: to allow anybody to produce high-quality books using a reasonably minimal amount of effort, and to provide a system that would give exactly the same results on all computers, now and in the future.

Bibliography

[1] A. Banner, *The Calculus Lifesaver.* Princeton 2007.

[2] J.W. Brown and R.V. Churchill, *Complex Variables and Applications*, McGraw-Hill 2013.

[3] S.C. Campbell and R. Haberman, *Introduction to Differntial Equations,* Princeton 2008.

[4] W. H. Fleming, *Functions of Several Variables.* Addison-Wesley, 2nd edition, 1977.

[5] B.R. Gelbaum and J.M.H. Olmsted, *Countrexamples in Analysis,* Dover Publications, 2nd edition, 1992.

[6] J.M. Howie, *Real Analysis,* Springer, London 2001.

[7] R. Larson and B. Edwards, *Calculus.* Brooks\Cole 2012.

[8] R. Larson and B. Edwards, *Multivariable Calculus.* Brooks\Cole 2010.

[9] J. Marsden and A. Weinstein, *Calculus I,* Springer 1985.

[10] J. Marsden and A. Weinstein, *Calculus II,* Springer 1985.

[11] J. Marsden and A. Weinstein, *Calculus III,* Springer 1985.

[12] A. Mattuck, *Introduction to Analysis*, Prentice Hall 2013.

[13] D. McMahon, *Complex Variables Demystified,* McGraw-Hill*, 2008.*

[14] J. Rogawski, *Calculus, Early Transcendentals.* W. H. Freeman and Company 2008.

[15] W. Rudin, *Principles of Mathematical Analysis.* International Series in Pure and Applied Mathematics. McGraw-Hill, 1953.

[16] T. Sundstrom, *Mathematical Reasoning. Writing and Proof ,* Ted Sundstrom, Grand Valley State University, 2014.

[17] W F. Trench, *Introduction to Real Analysis.* Prentice Hall, 2003.

Index

Index