MATH 221
FIRST SEMESTER
CALCULUS

fall 2007

Typeset:December 11, 2007

Math 221 – 1st Semester Calculus
Lecture notes version 1.0 (Fall 2007)

This is a self contained set of lecture notes for Math 221. The notes were written by Sigurd Angenent, starting from an extensive collection of notes and problems compiled by Joel Robbin.

The LaTeX and `Python` files which were used to produce these notes are available at the following web site

`www.math.wisc.edu/~angenent/Free-Lecture-Notes`

They are meant to be freely available in the sense that "free software" is free. More precisely:

Contents

I. Numbers, Points, Lines and Curves

1. What is a number?

The basic objects that we deal with in calculus are the so-called "real numbers" which you have already seen in pre-calculus. To refresh your memory let's look at the various kinds of "real" numbers that one runs into.

The simplest numbers are the ***positive integers***

$$1, 2, 3, 4, \cdots$$

the number ***zero***

$$0,$$

and the ***negative integers***

$$\cdots, -4, -3, -2, -1.$$

Together these form the integers or "whole numbers."

Next, there are the numbers you get by dividing one whole number by another (nonzero) whole number. These are the so called fractions or ***rational numbers*** such as

$$\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \frac{4}{3}, \cdots$$

or

$$-\frac{1}{2}, -\frac{1}{3}, -\frac{2}{3}, -\frac{1}{4}, -\frac{2}{4}, -\frac{3}{4}, -\frac{4}{3}, \cdots$$

By definition, any whole number is a rational number (in particular zero is a rational number.)

You can add, subtract, multiply and divide any pair of rational numbers and the result will again be a rational number (provided you don't try to divide by zero).

One day in middle school you were told that there are other numbers besides the rational numbers, and the first example of such a number is the square root of two. It has been known ever since the time of the greeks that no rational number exists whose square is exactly 2, i.e. you can't find a fraction $\frac{m}{n}$ such that

$$\left(\frac{m}{n}\right)^2 = 2, \text{ i.e. } m^2 = 2n^2.$$

Nevertheless, since

$$(1.4)^2 = 1.96 \text{ is less than 2, and}$$
$$(1.5)^2 = 2.25 \text{ is more than 2,}$$

it seems that there should be some number x between 1.4 and 1.5 whose square is exactly 2. So, we assume that there is such a number, and we call it the square root of 2, written as $\sqrt{2}$. This raises several questions. How do we know there really is a number between 1.4 and 1.5 for which $x^2 = 2$? How many other such numbers we are going to assume into existence? Do these new numbers obey the same algebra rules as the rational numbers? (e.g. when you add three numbers a, b and c the sum does not depend on the order in which you add them.) If we knew precisely what these numbers (like $\sqrt{2}$) were then we could perhaps answer such questions. It turns out to be rather difficult to give a precise description of what a number is, and in this course we won't try to get anywhere near the bottom of this issue. Instead we will think of numbers as "infinite decimal expansions" as follows.

One can represent certain fractions as decimal fractions, e.g.

$$\frac{279}{25} = \frac{1116}{100} = 11.16.$$

Not all fractions can be represented as decimal fractions. For instance, expanding $\frac{1}{3}$ into a decimal fraction leads to an unending decimal fraction

$$\frac{1}{3} = 0.333\,333\,333\,333\,333\,\cdots$$

It is impossible to write the complete decimal expansion of $\frac{1}{3}$ because it contains infinitely many digits. But we can describe the expansion: each digit is a three. An electronic calculator, which always represents numbers as ***finite*** decimal numbers, can never hold the number $\frac{1}{3}$ exactly.

Every fraction can be written as a decimal fraction which may or may not be finite. If the decimal expansion doesn't end, then it must repeat. For instance,

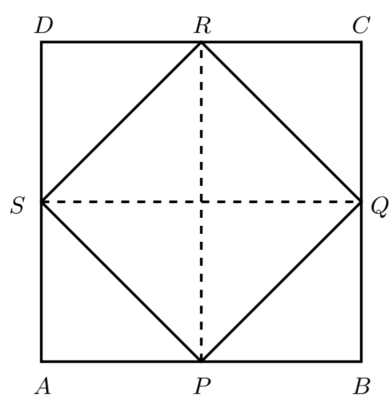
$$\frac{1}{7} = 0.142857\,142857\,142857\,142857\,\ldots$$

Conversely, any infinite repeating decimal expansion represents a rational number.

A ***real number*** is specified by a possibly unending decimal expansion. For instance,

$$\sqrt{2} = 1.414\,213\,562\,373\,095\,048\,801\,688\,724\,209\,698\,078\,569\,671\,875\,376\,9\ldots$$

Of course you can never write *all* the digits in the decimal expansion, so one only writes the first few digits and hides the others behind dots. To give a precise description of a real number (such as $\sqrt{2}$) you have to explain how one could in principle compute as many digits in the expansion as one would like. During the next three semesters of calculus we will not go into the details of how this should be done.



Another reason to believe in $\sqrt{2}$

The Pythagorean theorem says that the hypotenuse of a right triangle with sides 1 and 1 must be a line segment of length $\sqrt{2}$.

In middle or highschool you learned something similar to the following geometric construction of a line segment whose length is $\sqrt{2}$. Take a square $ABCD$ with sides of length 2. Let $PQRS$ be the square formed by connecting the midpoints of the square $ABCD$. Then the area of $PQRS$ is exactly half that of $ABCD$. Since $ABCD$ has area 4, the area of $PQRS$ must be 2, and therefore any side of $PQRS$ must have length $\sqrt{2}$.

Why are real numbers called real?

All the numbers we will use in this first semester of calculus are "real numbers." At some point (in 2nd semester calculus) it becomes useful to assume that there is a number whose square is -1. No real number has this property since the square of any real number is positive, so it was decided to call this new imagined number "imaginary" and to refer to the numbers we already have (rationals, $\sqrt{2}$-like things) as "real."

Exercises

1.1 – What is the 2007^{th} digit after the period in the expansion of $\frac{1}{7}$?

1.2 – Which of the following fractions have finite decimal expansions?

$$a = \frac{2}{3}, \quad b = \frac{3}{25}, \quad c = \frac{276937}{15625}.$$

1.3 – Write the numbers

$$x = 0.3131313131\ldots, y = 0.273273273273\ldots \text{ and } z = 0.21541541541541541\ldots$$

as fractions (i.e. write them as $\frac{m}{n}$, specifying m and n.)

(Hint: show that $100x = x + 31$. A similar trick works for y, but z is a little harder.)

1.4 – Is the number whose decimal expansion after the period consists only of nines, i.e.

$$x = 0.99999999999999999\ldots$$

an integer?

1.5 – There is a real number x which satisfies

$$\tfrac{1}{3}x^7 + x + 2 = 5.$$

Find the first three digits in the decimal expansion of x. [Use a calculator.]

2. The real number line and intervals

It is customary to visualize the real numbers as points on a straight line. We imagine a line, and choose one point on this line, which we call the ***origin***. We also decide which direction we call "left" and hence which we call "right." Some draw the number line vertically and use the words "up" and "down."

To plot any real number x one marks off a distance x from the origin, to the right (up) if $x > 0$, to the left (down) if $x < 0$.

The ***distance along the number line*** between two numbers x and y is $|x - y|$. In particular, the distance is never a negative number.

Almost every equation involving variables x, y, etc. we write down in this course will be true for some values of x but not for others. In modern abstract mathematics a collection of real numbers (or any other kind of mathematical objects) is called a ***set***. Below are some examples of sets of real numbers. We will use the notation from these examples throughout this course.

2.1. *Intervals*

The collection of all real numbers between two given real numbers form an interval. The following notation is used

- (a, b) is the set of all real numbers x which satisfy $a < x < b$.
- $[a, b)$ is the set of all real numbers x which satisfy $a \le x < b$.
- $(a, b]$ is the set of all real numbers x which satisfy $a < x \le b$.
- $[a, b)$ is the set of all real numbers x which satisfy $a \le x \le b$.

If the endpoint is not included then it may be ∞ or $-\infty$. E.g. $(-\infty, 2]$ is the interval of all real numbers (both positive and negative) which are ≤ 2.

2.2. *Set notation*

A common way of describing a set is to say it is the collection of all real numbers which satisfy a certain condition. One uses this notation

$$\mathcal{A} = \{x \mid x \text{ satisfies this or that condition}\}$$

Most of the time we will use upper case letters in a calligraphic font to denote sets. ($\mathcal{A}$,$\mathcal{B}$,$\mathcal{C}$,$\mathcal{D}$, ...)

For instance, the interval (a, b) can be described as

$$(a, b) = \{x \mid a < x < b\}$$

The set

$$\mathcal{B} = \{x \mid x^2 - 1 > 0\}$$

consists of all real numbers x for which $x^2 - 1 > 0$, i.e. it consists of all real numbers x for which either $x > 1$ or $x < -1$ holds. This set consists of two parts: the interval $(-\infty, -1)$ and the interval $(1, \infty)$.

You can try to draw a set of real numbers by drawing the number line and coloring the points belonging to that set red, or by marking them in some other way.

Some sets can be very difficult to draw. Consider

$$\mathcal{C} = \{x \mid x \text{ is a rational number}\}$$

or

$$\mathcal{D} = \{x \mid \text{the number 3 does not appear in the decimal expansion of } x\}.$$

Sets can also contain just a few numbers, like

$$\mathcal{E} = \{1, 2, 3\}$$

which is the set containing the numbers one, two and three. Or the set

$$\mathcal{F} = \{x \mid x^3 - 4x^2 + 1 = 0\}.$$

If $\mathcal{A}$ and $\mathcal{B}$ are two sets then ***the union of*** $\mathcal{A}$ ***and*** $\mathcal{B}$ is the set which contains all numbers that belong either to $\mathcal{A}$ or to $\mathcal{B}$. The following notation is used

$$\mathcal{A} \cup \mathcal{B} = \{x \mid x \text{ belongs to } \mathcal{A} \text{ or to } \mathcal{B}.\}$$

Similarly, the ***intersection of two sets*** $\mathcal{A}$ ***and*** $\mathcal{B}$ is the set of numbers which belong to both sets. This notation is used:

$$\mathcal{A} \cup \mathcal{B} = \{x \mid x \text{ belongs to both } \mathcal{A} \text{ and } \mathcal{B}.\}$$

Exercises

2.1 – Draw the following sets of real numbers

$$\mathcal{A} = \{x \mid x^2 - 3x + 2 \le 0\} \qquad \mathcal{B} = \{x \mid x^2 - 3x + 2 \ge 0\}$$
$$\mathcal{C} = \{x \mid x^2 - 3x > 3\} \qquad \mathcal{D} = \{x \mid x^2 - 5 > 2x\}$$
$$\mathcal{E} = \{t \mid t^2 - 3t + 2 \le 0\} \qquad \mathcal{F} = \{\alpha \mid \alpha^2 - 3\alpha + 2 \ge 0\}$$
$$\mathcal{G} = (0,1) \cup (5,7] \qquad \mathcal{H} = \big(\{1\} \cup \{2,3\}\big) \cap (0, 2\sqrt{2})$$
$$\mathcal{P} = \{x^2 - 2x \mid 0 \le x \le 2\} \qquad \mathcal{Q} = \{x^2 - 2x \mid 0 \le x \le 1\}$$
$$\mathcal{R} = \{\theta \mid \sin\theta = \tfrac{1}{2}\} \qquad \mathcal{S} = \{\varphi \mid \cos\varphi > 0\}$$

2.2 – Suppose $\mathcal{A}$ and $\mathcal{B}$ are intervals. Is it always true that $\mathcal{A} \cap \mathcal{B}$ is an interval? How about $\mathcal{A} \cup \mathcal{B}$?

2.3 – Consider the sets

$$\mathcal{M} = \{x \mid x > 0\} \text{ and } \mathcal{N} = \{y \mid y > 0\}.$$

Are these sets the same?

3. Sets of Points in the Plane

3.1. *Cartesian Coordinates*

The coordinate plane with its x and y axes are familiar from middle/high school mathematics.

Briefly, you can specify the location of any point in the plane by choosing two fixed orthogonal lines (called the x and y axes) and specifying the distances to each of the two axes. The distance to the x axes is y, and the distance to the y axis is x. By allowing the numbers x and y to be either positive or negative, one can keep track on which side of the axes the point with coordinates x and y lies.

The notation $P(x,y)$ is used as an abbreviation for the more cumbersome phrase "the point P whose coordinates are (x,y)."

3.2. *Sets*

Just as one can consider sets of real numbers, one can also consider sets of points in the plane. Examples of such sets are

$$\mathcal{A} = \text{ All points on the } x\text{-axis}$$
$$\mathcal{B} = \text{ All points on the } x \text{ or } y\text{-axes}$$
$$\mathcal{C} = \text{ All points whose distance to the origin is 25}$$
$$\mathcal{D} = \text{ All points whose coordinates } x \text{ and } y \text{ are both integers}$$
$$\mathcal{E} = \text{ All points one of whose coordinates } x \text{ and } y \text{ is an integer}$$
$$\mathcal{F} = \text{ All points whose coordinates } x \text{ and } y \text{ are both rational numbers}$$

One can also write some of these sets as follows

$$\mathcal{A} = \{(x,y) \mid y = 0\} = \{(x,0) \mid x \text{ arbitrary}\}$$
$$\mathcal{B} = \{(x,y) \mid x = 0 \text{ or } y = 0\} = \{(x,y) \mid xy = 0\}$$
$$\mathcal{C} = \left\{(x,y) \mid \sqrt{x^2 + y^2} = 25\right\} = \left\{(x,y) \mid x^2 + y^2 = 625\right\}.$$

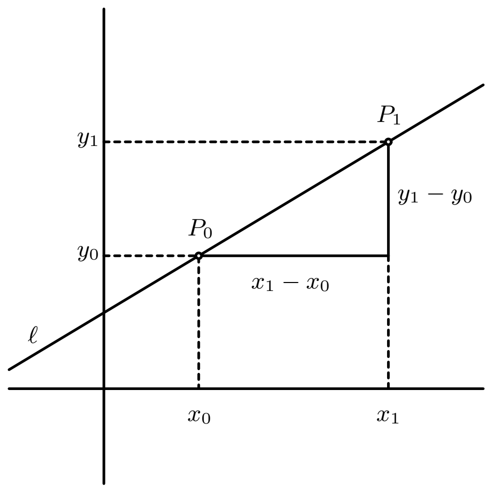


Figure 1. A straight line and its slope

3.3. *Lines*

A straight line through two points is also an example of a set of points in the plane. Unless the line is "vertical," (i.e. parallel to the y-axis) it is a set of the form

$$\ell = \{(x, y) \mid y = mx + n\} \tag{1}$$

where m is the ***slope*** of the line, and n is its y-intercept (the y coordinate of the point where ℓ intersects the ; y axis).

If $P_0(x_0, y_0)$ and $P_1(x_1, y_1)$ are two points on the line ℓ, then one can compute the slope m from the "rise-over-run" formula

$$m = \frac{y_1 - y_0}{x_1 - x_0}.$$

This formula actually contains a theorem, namely it says that the ratio $(y_1 - y_0) : (x_1 - x_0)$ is the same for every pair of points (x_0, y_0) and (x_1, y_1) that you could pick on the line.

Exercises

3.1 – Draw the sets from §3.2.

3.2 – What is the distance between the points $P(1, 3)$ and $Q(-3, 7)$?

Find the equation for the line ℓ through P and Q.

3.3 – Let ℓ be the line through $A(0, 1)$ and $B(2, 3)$, and let ℓ' be the line through $C(-2, -10)$ and $D(3, 3)$.

Find the equations for ℓ and ℓ'.

Where do the two lines ℓ and ℓ' intersect?

Draw the sets $\mathcal{A} = \ell \cup \ell'$ and $\mathcal{B} = \ell \cap \ell'$.

3.4 – Let ℓ be the line through the point $(-1, 0)$ with slope m, and let $\mathcal{C}$ be the circle centered at the origin with radius 1.

Make a drawing of ℓ and C, and find $\ell \cap \mathcal{C}$.

3.5 – Draw the following sets of points in the plane

$$\mathcal{A} = \{(x, y) \mid y > 0 \text{ and } x > 0\} \qquad \mathcal{B} = \{(x, y) \mid y < 0 \text{ or } y > x\}$$
$$\mathcal{C} = \{(x, y) \mid y + x = 0\} \qquad \mathcal{D} = \{(x, y) \mid x^2 - 5 > 2x\}$$
$$\mathcal{E} = \{(x, y) \mid x^2 + y^2 < 1\} \qquad \mathcal{F} = \{(x, y) \mid (x-1)^2 + y^2 = 1\}$$
$$\mathcal{G} = \mathcal{E} \cup \mathcal{F} \qquad \mathcal{H} = \mathcal{E} \cap \mathcal{F}$$

4. Functions

Definition. 4.1. *A **function** consists of a **rule** and a **domain**.*

*The **domain** is a set of real numbers, and the **rule** tells you how to compute a real number $f(x)$ for any given real number x in the domain.*

*The set of all possible numbers $f(x)$ as x runs over the domain is called the **range** of the function.*

The rule which specifies a function can come in many different forms. Most often it is a formula, as in

$$f(x) = x^2 - 2x + 1, \quad \text{domain of } f = \text{all real numbers.}$$

or a few formulas, as in

$$g(x) = \begin{cases} 2x & \text{for } x < 0 \\ x^2 & \text{for } x \geq 0 \end{cases} \qquad \text{domain of } g = \text{all real numbers.}$$

Functions which are defined by different formulas on different intervals are sometimes called "piecewise defined functions."

In this course we will usually not be careful about specifying the domain of the function. When this happens the domain is understood to be the set of all x for which the rule which tells you how to compute $f(x)$ is meaningful. For instance, if we say that h is the function

$$h(x) = \sqrt{x}$$

then the domain of h is understood to be the set of all nonnegative real numbers

$$\text{domain of } h = [0, \infty)$$

since $\sqrt{x}$ is well-defined for all $x \geq 0$ and undefined for $x < 0$.

A systematic way of finding the domain and range of a function for which you are only given a formula is as follows:

- The domain of f consists of all x for which $f(x)$ is well-defined ("makes sense")
- The range of f consists of all y for which you can solve the equation $f(x) = y$.

4.1. *Example: Find the domain and range of $f(x) = 1/x^2$*

The expression $1/x^2$ can be computed for all real numbers x except $x = 0$ since this leads to division by zero. Hence the domain of the function $f(x) = 1/x^2$ is

$$\{x \mid x \neq 0\} = (-\infty, 0) \cup (0, \infty).$$

To find the range we ask "for which y can we solve the equation $y = f(x)$ for x," i.e. we try to solve

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

for x.

Since x cannot be zero we have $x^2 > 0$ for all values of x we are allowed to choose. Therefore $1/x^2 > 0$, and the equation $y = 1/x^2$ will not have a solution if $y \leq 0$. On the other hand, if $y > 0$ then the equation has not just one but two solutions, namely $x = \pm 1/\sqrt{y}$. So any positive number y belongs to the range of f, and if y is not positive, then y does not belong to the range of f. We have found that

$$\text{range}(f) = (0, \infty).$$

4.2. *Functions in "real life"*

One can describe the motion of an object using a function. If some object is moving along a straight line, then you can define the following function: Let $x(t)$ be the distance from the object to a fixed marker on the line, at the time t. Here the domain of the function is the set of all times t for which we know the position of the object, and the rule is

Given t, measure the distance between the object and the marker.

There are many examples of this kind. For instance, a biologist could describe the growth of a cell by defining $m(t)$ to be the mass of the cell at time t (measured since the birth of the cell). Here the domain is the interval $[0, T]$, where T is the life time of the cell, and the rule that describes the function is *given t, weigh the cell at time t.*

5. **The graph of a function**

Definition. 5.1. *The **graph** of a function*

$$y = f(x)$$

is the set of all points $P(x, y)$ whose coordinates (x, y) satisfy the equation $y = f(x)$, i.e.

$$\text{Graph}(f) = \big\{(x, y) \mid y = f(x)\big\}.$$

5.1. *Vertical Line Property*

Generally speaking graphs of functions are curves in the plane but they distinguish themselves from arbitrary curves (or point sets) by the way they intersect vertical lines: *The graph of a function cannot intersect a vertical line* $x =$ constant *in more than one point.* To see why this is so, suppose that you have two points (x_0, y_0) and (x_1, y_1) on the graph of f which also lie on the same vertical line with equation $x = a$.

Since both points lie on the vertical line with $x = a$ we have

$$x_0 = a \text{ and } x_1 = a$$

Since both points lie on the graph of f we also have

$$y_0 = f(x_0) \text{ and } y_1 = f(x_1).$$

It follows that $x_0 = x_1 = a$ and $y_0 = f(x_0) = f(a)$ and $y_1 = f(x_1) = f(a)$ so that both points are given by $(x_0, y_0) = (x_1, y_1) = (a, f(a))$, i.e. they are the same point.

5.2. *Example*

The point set determined by the equation $x^2 + y^2 = 1$ is a circle; it is not the graph of a function since the vertical line $x = 0$ (the y-axis) intersects the graph in two points $P_1(0, 1)$ and $P_2(0, -1)$.

The graph of $f(x) = x^3 - x$ "goes up and down," and while it intersects the x-axis in three points ($(-1, 0)$, $(0, 0)$ and $(1, 0)$) it intersects every vertical line in exactly one point.

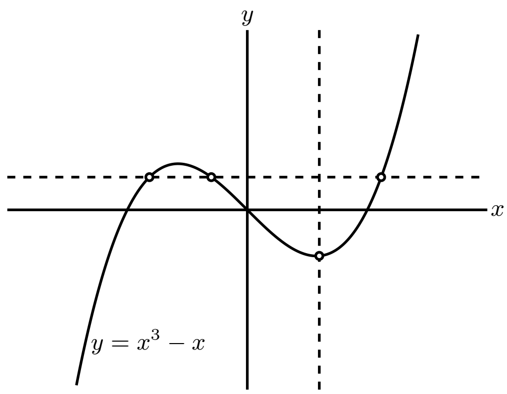


Figure 2. The graph of a function intersects vertical lines in at most one point, but may intersect horizontal lines more than once

6. **Inverse functions and Implicit functions**

For many functions the rule which tells you how to compute it is not an explicit formula, but instead an equation which you still must solve. A function which is defined in this way is called an "implicit function."

6.1. *Example*

One can define a function f by saying that for each x the value of $f(x)$ is the solution y of the equation

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

In this example you can solve the equation for y,

$$y = \frac{3 - x^2}{2}.$$

Thus we see that the function we have defined is $f(x) = (3 - x^2)/2$.

Here we have two definitions of the same function, namely (i) "f is defined by $y = f(x) \iff x^2 + 2y - 3 = 0$" and (ii) "$f$ is defined by $f(x) = (3 - x^2)/2$." The first definition is the implicit definition, the second is explicit. You see that with an "implicit function" it isn't the function itself, but rather the way it was defined that's implicit.

6.2. *Another example: domain of an implicitly defined function*

Define g by saying that for any x the value $y = g(x)$ is the solution of

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

Just as in the previous example one can then solve for y, and one finds that

$$g(x) = y = \frac{3 - x^2}{x}.$$

Unlike the previous example this formula does not make sense when $x = 0$, and indeed, for $x = 0$ our rule for g says that $g(0) = y$ is the solution of

$$0^2 + 0 \cdot y - 3 = 0, \text{ i.e. } y \text{ is the solution of } 3 = 0.$$

That equation has no solution and hence $x = 0$ does not belong to the domain of our function g.

Figure 3. The circle determined by $x^2+y^2=1$ is not the graph of a function, but it is the union of the graphs of the two functions $h_1(x)=\sqrt{1-x^2}$ and $h_2(x)=-\sqrt{1-x^2}$.

6.3. *Example: the equation alone does not determine the function*

Define $y=h(x)$ to be the solution of

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

If $x>1$ or $x<-1$ then $x^2>1$ and there is no solution, so $h(x)$ is at most defined when $-1\le x\le 1$. But when $-1<x<1$ there is another problem: not only does the equation have a solution, but it even has two solutions:

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

The rule which defines a function must be unambiguous, and since we have not specified which of these two solutions is $h(x)$ the function is not defined for $-1<x<1$.

One can fix this by making a choice, but there are many possible choices. Here are three possibilities:

$$\begin{aligned} h_1(x) &= \text{the nonnegative solution } y \text{ of } x^2+y^2=1 \\ h_2(x) &= \text{the nonpositive solution } y \text{ of } x^2+y^2=1 \\ h_3(x) &= \begin{cases} h_1(x) & \text{when } x<0 \\ h_2(x) & \text{when } x\ge 0 \end{cases} \end{aligned}$$

6.4. *Why use implicit functions?*

In all the examples we have done so far we could replace the implicit description of the function with an explicit formula. This is not always possible or if it is possible the implicit description is much simpler than the explicit formula. For instance, you can define a function f by saying that $y=f(x)$ if and only if

$$y^3+3y+2x=0. \tag{2}$$

This means that the recipe for computing $f(x)$ for any given x is "solve the equation $y^3+3y+2x=0$." E.g. to compute $f(0)$ you set $x=0$ and solve $y^3+3y=0$. The only solution is $y=0$, so $f(0)=0$. To compute $f(1)$ you have to solve $y^3+3y+2\cdot 1=0$, and if you're lucky you see that $y=-1$ is the solution, and $f(1)=-1$.

In general, no matter what x is, the equation (2) turns out to have exactly one solution y (which depends on x, this is how you get the function f). Solving (2) is not easy. In the early 1500s Cardano and Tartaglia discovered a formula[1] for the solution. Here it is:

$$y=f(x)=\sqrt[3]{-x+\sqrt{1+x^2}}-\sqrt[3]{x+\sqrt{1+x^2}}.$$

The implicit description looks a lot simpler, and when we try to differentiate this function later on, it will be much easier to use "implicit differentiation" than to use the Cardano-Tartaglia formula directly.

---

[1] To see the solution and its history visit
`www.gap-system.org/~history/HistTopics/Quadratic_etc_equations.html`

6.5. *Inverse functions*

If you have a function f, then you can try to define a new function g which will be called the ***inverse function of*** f by requiring

$$y = g(x) \iff x = f(y).$$

In other words, to find $y = g(x)$ you solve the equation $x = f(y)$. Depending on the function f it may happen that the equation $x = f(y)$ has no solution, or that it has more than one solution. In either case we won't get an unambiguous value for y. However, if it is the case for some x that the equation $x = f(y)$ has ***exactly one solution*** y, then we can define $g(x) = y$. The inverse function of f is usually written as f^{-1}.

More precisely:

Definition. 6.1. *If f is a given function then the domain of the inverse function of f consists of every number x for which the equation*

$$x = f(y) \tag{3}$$

has exactly one solution.

If (3) has exactly one solution y then $f^{-1}(x) = y$.

6.6. *Examples*

Consider the function f with $f(x) = 2x + 3$. Then the equation $f(y) = x$ works out to be

$$2y + 3 = x$$

and this has the solution

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

So $f^{-1}(x)$ is defined for all x, and it is given by $f^{-1}(x) = (x-3)/2$.

Next consider the function $g(x) = x^2$ with domain all real numbers. To see if this function has an inverse we try to solve the equation $g(y) = x$, i.e. we try to solve $y^2 = x$. If $x > 0$ then this equation has ***two*** solutions, $\pm\sqrt{x}$; if $x < 0$ then $y^2 = x$ has no solutions; if $x = 0$ then $y^2 = x$ has exactly one solution, namely $y = 0$. So we see that $g^{-1}(x)$ is only defined when $x = 0$. For all other x the equation defining $g^{-1}(x)$ either gives too few or too many solutions.

We now consider the function g again ***but we change its domain,*** i.e. we consider the function h defined by $h(x) = x^2$, and whose domain is $[0, \infty)$. So h is defined by the same rule as g ("square whatever number you are given"), but h is only allowed to apply the rule to nonnegative numbers, while g was allowed to apply its rule to all numbers.

What is the inverse of h? To find $h^{-1}(x)$ we solve the equation $h(y) = x$, i.e. $y^2 = x$. For negative x we again get no solution, just as in the case of g. But when x is positive we now get something else: the equation $h(y) = x$ ***has only one solution***. There are two numbers y which satisfy $y^2 = x$, but only one of them lies in the domain of h.

Conclusion: the inverse of h has domain $[0, \infty)$ and is given by $h(x) = \sqrt{x}$.

6.7. *Inverse trigonometric functions*

The familiar trigonometric functions Sine, Cosine and Tangent have inverses which are called arcsine, arccosine and arctangent.

$y = f(x)$		$x = f^{-1}(y)$	
$y = \sin x$	$(-\pi/2 \le x \le \pi/2)$	$x = \arcsin(y)$	$(-1 \le y \le 1)$
$y = \cos x$	$(0 \le x \le \pi)$	$x = \arccos(y)$	$(-1 \le y \le 1)$
$y = \tan x$	$(-\pi/2 < x < \pi/2)$	$x = \arctan(y)$	

The notations $\arcsin y = \sin^{-1} y$, $\arccos x = \cos^{-1} x$, and $\arctan u = \tan^{-1} u$ are also commonly used for the inverse trigonometric functions. We will avoid the $\sin^{-1} y$ notation because it is ambiguous. Namely, everybody writes the square of $\sin y$ as

$$\big(\sin y\big)^2 = \sin^2 y.$$

This suggests that $\sin^{-1} y$ should mean

$$\sin^{-1} y = \big(\sin y\big)^{-1} = \frac{1}{\sin y},$$

and not $\arcsin y$.

Exercises

6.1 – Draw the graphs of the functions h_1, h_2, h_3 from §6.3

6.2 – Find a formula for the function f which is defined by

$$y = f(x) \iff x^2 y + y = 7.$$

What is the domain of f?

6.3 – Find a formula for the function f which is defined by

$$y = f(x) \iff x^2 y - y = 6.$$

What is the domain of f?

6.4 – Let f be the function defined by $y = f(x) \iff y$ is the largest solution of

$$x^2 + xy + y^2 = 0.$$

Find a formula for f. What are the domain and range of f?

6.5 – Find a formula for the function f which is defined by

$$y = f(x) \iff 2x + 2xy + y^2 = 5 \text{ and } y > -x.$$

Find the domain of f.

6.6 – Use a calculator to compute $f(1.2)$ in three decimals where f is the implicitly defined function from §6.4. (There are (at least) two different ways of finding $f(1.2)$)

6.7 – On a graphing calculator plot the graphs of the following functions, and explain the results.

$$f(x) = \arcsin(\sin x) \qquad -2\pi \le x \le 2\pi$$
$$g(x) = \arcsin(x) + \arccos(x) \qquad 0 \le x \le 1$$
$$h(x) = \arctan \frac{\sin x}{\cos x} \qquad |x| < \pi/2$$
$$k(x) = \arctan \frac{\cos x}{\sin x} \qquad |x| < \pi/2$$
$$l(x) = \arcsin(\cos x) \qquad -\pi \le x \le \pi$$
$$m(x) = \cos(\arcsin x) \qquad -1 \le x \le 1$$

6.8 – A function f is given which satisfies

$$f(2x+3) = x^2$$

for all real numbers x.

Compute

(a) $f(0)$ (b) $f(3)$ (c) $f(x)$
(d) $f(y)$ (e) $f(f(2))$

where x and y are arbitrary real numbers.

What are the range and domain of f?

6.9 – A function f is given which satisfies

$$f\left(\frac{1}{x+1}\right) = 2x - 12.$$

for all real numbers x.

Compute

(a) $f(1)$ (b) $f(0)$ (c) $f(x)$
(d) $f(t)$ (e) $f(f(2))$

where x and t are arbitrary real numbers.

What are the range and domain of f?

6.10 – Does there exist a function f which satisfies

$$f(x^2) = x + 1$$

for ***all*** real numbers x?

II. Derivatives (1)

To work with derivatives you have to know what a limit is, but to motivate why we are going to study limits let's first look at the two classical problems that gave rise to the notion of a derivative: the tangent to a curve, and the instantaneous velocity of a moving object.

7. The tangent to a curve

Suppose you have a function $y = f(x)$ and you draw its graph. If you want to find the tangent to the graph of f at some given point on the graph of f, how would you do that?

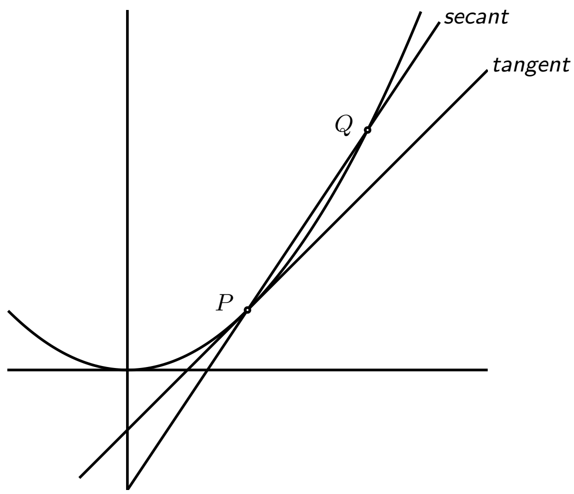


Figure 4. Constructing the tangent by letting $Q \to P$

Let P be the point on the graph at which want to draw the tangent. If you are making a real paper and ink drawing you would take a ruler, make sure it goes through P and then turn it until it doesn't cross the graph anywhere else.

If you are using equations to describe the curve and lines, then you could pick a point Q on the graph and construct the line through P and Q ("construct" means "find an equation for"). This line is called a "secant," and it is of course not the tangent that you're looking for. But if you choose Q to be very close to P then the secant will be close to the tangent.

So this is our recipe for constructing the tangent through P: pick another point Q on the graph, find the line through P and Q, and see what happens to this line as you take Q closer and closer to P. The resulting secants will then get closer and closer to some line, and that line is the tangent.

We'll write this in formulas in a moment, but first let's worry about how close Q should be to P. We can't set Q equal to P, because then P and Q don't determine a line (you need ***two*** points to determine a line). If you choose Q different from P then you don't get the tangent, but at best something that is "close" to it. Some people have suggested that one should take Q "infinitely close" to P, but it isn't clear what that would mean. The concept of a limit is meant to solve this confusing problem.

8. An example – tangent to a parabola

To make things more concrete, suppose that the function we had was $f(x) = x^2$, and that the point was $(1, 1)$. The graph of f is of course a parabola.

Any line through the point $P(1, 1)$ has equation

$$y - 1 = m(x - 1)$$

where m is the slope of the line. So instead of finding the equation of the secant and tangent lines we will find their slopes.

Let Q be the other point on the parabola, with coordinates (x, x^2). We can "move Q around on the graph" by changing x. Whatever x we choose, it must be different from 1, for otherwise P and Q would be the same point. What we want to find out is how the line through P and Q changes if x is changed (and in particular, if x is chosen very close to a). Now, as one changes x one thing stays the same, namely, the secant still goes through P. So to describe the secant we only need to know its slope. By the "rise over run" formula, the slope of the secant line joining P and Q is

$$m_{PQ} = \frac{\Delta y}{\Delta x} \tag{4}$$

where

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

is the difference between the vertical coordinate x^2 of Q and the vertical coordinate 1 of P and

$$\Delta x = x - 1$$

is the difference of the horizontal coordinates of P and Q. By factoring $x^2 - 1$ we can rewrite the formula for the slope as follows

$$m_{PQ} = \frac{\Delta y}{\Delta x} = \frac{x^2 - 1}{x - 1} = \frac{(x - 1)(x + 1)}{x - 1} = x + 1. \tag{5}$$

As x gets closer to 1, the slope m_{PQ}, being $x + 1$, gets closer to the value $1 + 1 = 2$. We say that

the limit of the slope m_{PQ} ***as*** Q ***approaches*** P ***is*** 2.

In symbols,

$$\lim_{Q \to P} m_{PQ} = 2,$$

or, since Q approaching P is the same as x approaching 1,

$$\lim_{x \to 1} m_{PQ} = 2. \tag{6}$$

So we find that the tangent line to the parabola $y = x^2$ at the point $(1, 1)$ has equation

$$y - 1 = 2(x - 1), \text{ i.e. } y = 2x - 1.$$

<u>*8.1*</u> – Repeat the above reasoning to find the slope at the point $(\frac{1}{2}, \frac{1}{4})$, or more generally at any point (a, a^2) on the parabola.

A warning: you cannot substitute $x = 1$ in equation (5) to get (6) even though it looks like that's what we did. The reason why you can't do that is that when $x = 1$ the point Q coincides with the point P so "the line through P and Q" is not defined; also, if $x = 1$ then $\Delta x = \Delta y = 0$ so that the rise-over-run formula for the slope gives

$$m_{PQ} = \frac{\Delta x}{\Delta y} = \frac{0}{0} = \text{ undefined.}$$

It is only after the algebra trick in (5) that setting $x = 1$ gives something that is well defined. But if the intermediate steps leading to $m_{PQ} = x + 1$ aren't valid for $x = 1$ why should the final result be worth anything for $x = 1$?

Something more complicated has happened. We did a calculation which is valid for all $x \neq 1$, and later looked at what happens if x gets "very close to 1." This is the concept of a limit and we'll study it in more detail later in this section, but first another example.

9. Instantaneous velocity

If you try to define "instantaneous velocity" you will again end up trying to divide zero by zero. Here is how it goes: When you are driving in your car the speedometer tells you how fast your are going, i.e. what your velocity is. What is this velocity? What does it mean if the speedometer says "50mph"?

We all know what ***average velocity*** is. Namely, if it takes you two hours to cover 100 miles, then your average velocity was

$$\frac{\text{distance traveled}}{\text{time it took}} = 50 \text{ miles per hour.}$$

This is not the number the speedometer provides you – it doesn't wait two hours, measure how far you went and compute distance/time. If the speedometer in your car tells you that you are driving 50mph, then that should be your velocity ***at the moment*** that you look at your speedometer, i.e. "distance traveled over time it took" at the moment you look at the speedometer. But during the moment you look at your speedometer no time goes by (because a moment has no length) and you didn't cover any distance, so your velocity at that moment is $\frac{0}{0}$, i.e. undefined. Your velocity at ***any*** moment is undefined. But then what is the speedometer telling you?

To put all this into formulas we need to introduce some notation. Let t be the time (in hours) that has passed since we got onto the road, and let $s(t)$ be the distance we have covered since then.

Instead of trying to find the velocity exactly at time t, we find a formula for the average velocity during some (short) time interval beginning at time t. We'll write Δt for the length of the time interval.

At time t we have traveled $s(t)$ miles. A little later, at time $t + \Delta t$ we have traveled $s(t+\Delta t)$. Therefore during the time interval from t to $t+\Delta t$ we have moved $s(t+\Delta t)-s(t)$ miles. Our average velocity in that time interval is therefore

$$\frac{s(t + \Delta t) - s(t)}{\Delta t} \text{ miles per hour.}$$

The shorter you make the time interval, i.e. the smaller you choose Δt, the closer this number should be to the instantaneous velocity at time t.

So we have the following formula (definition, really) for the velocity at time t

$$v(t) = \lim_{\Delta t \to 0} \frac{s(t + \Delta t) - s(t)}{\Delta t}. \tag{7}$$

10. Rates of change

The two previous examples have much in common. If we ignore all the details about geometry, graphs, highways and motion, the following happened in both examples:

We had a function $y = f(x)$, and we wanted to know how much $f(x)$ changes if x changes. If you change x to $x + \Delta x$, then y will change from $f(x)$ to $f(x + \Delta x)$. The change in y is therefore

$$\Delta y = f(x + \Delta x) - f(x),$$

and the average rate of change is

$$\frac{\Delta y}{\Delta x} = \frac{f(x + \Delta x) - f(x)}{\Delta x}. \tag{8}$$

This is the average rate of change of f over the interval from x to $x + \Delta x$. To define ***the rate of change of the function f at x*** we let the length Δx of the interval become smaller and smaller, in the hope that the average rate of change over the shorter and shorter time intervals will get closer and closer to some number. If that happens then that "limiting number" is called the rate of change of f at x, or, the ***derivative*** of f at x. It is written as

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}. \tag{9}$$

Derivatives and what you can do with them are what the first half of this semester is about. The description we just went through shows that to understand what a derivative is you need to know what a limit is. In this chapter we'll study limits so that we get a less vague understanding of formulas like (9).

Exercises

10.1 – Using a calculator (if you must) approximate the derivatives of the following functions at the indicated points (i.e. find $f'(a)$) by computing $\frac{\Delta y}{\Delta x}$ for various values of Δx.

$(a)\ f(x) = x^2 - 2x + 1,\ \ a = 1$ $\qquad$ $(b)\ f(x) = x^2,\ \ a = -1$

$(a)\ f(x) = x^{33},\ \ a = 1$ $\qquad$ $(d)\ f(x) = 2^x,\ \ a = 1$

Use the following values Δx

$$\Delta x = 0.1, 0.01, 0.001, 10^{-6}, 10^{-12}.$$

10.2 – Simplify the algebraic expressions you get when you compute Δy and $\Delta y/\Delta x$ for the following functions

$(a)\ y = x^2 - 2x + 1$ $\qquad$ $(b)\ y = \sin x$

$(c)\ y = \dfrac{1}{x}$ $\qquad$ $(d)\ y = 2^x$

10.3 – Suppose that some quantity y is a function of some other quantity x, and suppose that y is a mass, i.e. y is measured in pounds, and x is a length, measured in feet. What units do the increments Δy and Δx, and the derivative dy/dx have?

10.4 – Let $A(r)$ be the area enclosed by a circle of radius r, and let $L(r)$ be the length of the circle. Show that

$$A'(r) = L(r).$$

III. Limits and Continuous Functions

11. Informal definition of limits

While it is easy to define precisely in a few words what a square root is ($\sqrt{a}$ is the positive number whose square is a) the definition of the limit of a function runs over several terse lines, and most people don't find it very enlightening when they first see it. (See §12.) So we postpone this for a while and fine tune our intuition for another page.

Definition. 11.1 (Definition of limit (1st attempt)). *If g is some function then*

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

is read "the limit of $g(x)$ as x approaches a is L." It means that if you choose values of x which are close ***but not equal*** *to a, then $g(x)$ will be close to the value L; moreover, $g(x)$ gets closer and closer to L as x gets closer and closer to a.*

The following alternative notation is sometimes used

$$g(x) \to L \quad \text{as} \quad x \to a;$$

(read "$g(x)$ approaches L as x approaches a" or "$g(x)$ goes to L as x goes to a".)

11.1. *Example*

If $g(x) = x + 3$ then

$$\lim_{x \to 4} g(x) = 7,$$

is true, because if you substitute number x close to 4 in $g(x) = x + 3$ the result will be close to 7.

11.2. *Example: substituting numbers to guess a limit*

What (if anything) is

$$\lim_{x \to 2} \frac{x^2 - 2x}{x^2 - 4}?$$

Here $g(x) = (x^2 - 2x)/(x^2 - 4)$ and $a = 2$.

We first try to substitute $x = 2$, but this leads to

$$g(2) = \frac{2^2 - 2 \cdot 2}{2^2 - 4} = \frac{0}{0}$$

which does not exist. Next we try to substitute values of x close but not equal to 2. Table **??** suggests that $g(x)$ approaches 0.5.

x	$g(x)$
3.000000	0.600000
2.500000	0.555556
2.100000	0.512195
2.010000	0.501247
2.001000	0.500125

x	$h(x)$
1.000000	1.009990
0.500000	1.009980
0.100000	1.009899
0.010000	1.008991
0.001000	1.000000

Table 1. Finding limits by substituting values of x "close to a"

11.3. *Example: Substituting numbers can suggest the wrong answer*

The previous example shows that our first definition of "limit" is not very precise, because it says "x close to a," but how close is close enough? Suppose we had taken the function

$$h(x) = \frac{101\,000x}{100\,000x + 1}$$

and we had asked for the limit $\lim_{x\to 0} h(x)$.

Then substitution of some "small values of x" could lead us to believe that the limit is $1.000\ldots$. Only when you substitute even smaller values do you find that the limit is 0 (zero)!

Exercise

11.1 – Our definition of a derivative in (9) contains a limit. What is the function g there, and what is the variable?

Answer: Δx is the variable, and $g(\Delta x) = (f(x + \Delta x) - f(x))/\Delta x$; we want $\lim_{\Delta x\to 0} g(\Delta x)$.

12. The formal, authoritative, definition of limit

The informal description of the limit uses phrases like "closer and closer" and "really very small." In the end we don't really know what they mean, although they are suggestive. "Fortunately" there is a good definition, i.e. one which is unambiguous and can be used to settle any dispute about the question of whether $\lim_{x\to a} g(x)$ equals some number L or not. Here is the definition. It takes a while to digest, so read it once, look at the examples, do a few exercises, read the definition again. Go on to the next sections. Throughout the semester come back to this section and read it again.

Definition. 12.1. *Suppose we have an interval $p < x < q$ and a number a in that interval $p < a < q$. Let g be a function which is defined for all x in the interval $p < x < q$, except possibly at $x = a$.*

We say that L is the limit of $g(x)$ as $x \to a$, if for every $\varepsilon > 0$ one can find a $\delta > 0$ such that for all x in the interval $p < x < q$ one has

$$|x - a| < \delta \;\&\; x \neq a \implies |g(x) - L| < \varepsilon.$$

Why the absolute values? The quantity $|x-y|$ is the distance between the points x and y on the number line, and one can measure how close x is to y by calculating $|x - y|$.

What are ε and δ? The quantity ε is how close you would like $g(x)$ to be to its limit L; the quantity δ is how close you have to choose x to a to achieve this. If you make ε smaller then you are requiring $g(x)$ to be closer to L than before, and hence you may have to restrict the allowed values of x, and thus reduce δ. So, δ ***depends on*** ε***.*** See also figures 5 and 6.

A strategy for proving $\lim_{x\to a} g(x) = L$**.** using what you know about the function g write out the expression $|g(x) - L|$, and try to simplify it so that you get something of the form

$$|g(x) - L| \leq \text{ something that only depends on } |x - a|.$$

This kind of inequality is called an ***estimate of*** $|g(x) - L|$ ***in terms of*** $|x - a|$**.** Next, assume that $|x - a| < \delta$, and turn the estimate you have into one of the form

$$|g(x) - L| \leq \text{ something that only depends on } \delta.$$

Finally, figure out how small δ has to be for the Right Hand Side of this last inequality to be less than ε.

If you can find such a $\delta > 0$ no matter which $\varepsilon > 0$ you were given then you have proved that $\lim_{x\to a} g(x) = L$. If you cannot find such a δ then one of the following has occurred:

(1) there might be a better estimate for $|g(x) - L|$
(2) the limit $\lim_{x\to a} g(x)$ exists, but it isn't L,
(3) the limit $\lim_{x\to a} g(x)$ does not exist.

A common mistake: after using a lot of algebra to estimate $g(x) - L$ you get confused and in the end a find a δ that not only depends on ε, but also on x. The dependence on ε is OK (unavoidable), but ***the dependence on x is not allowed.***

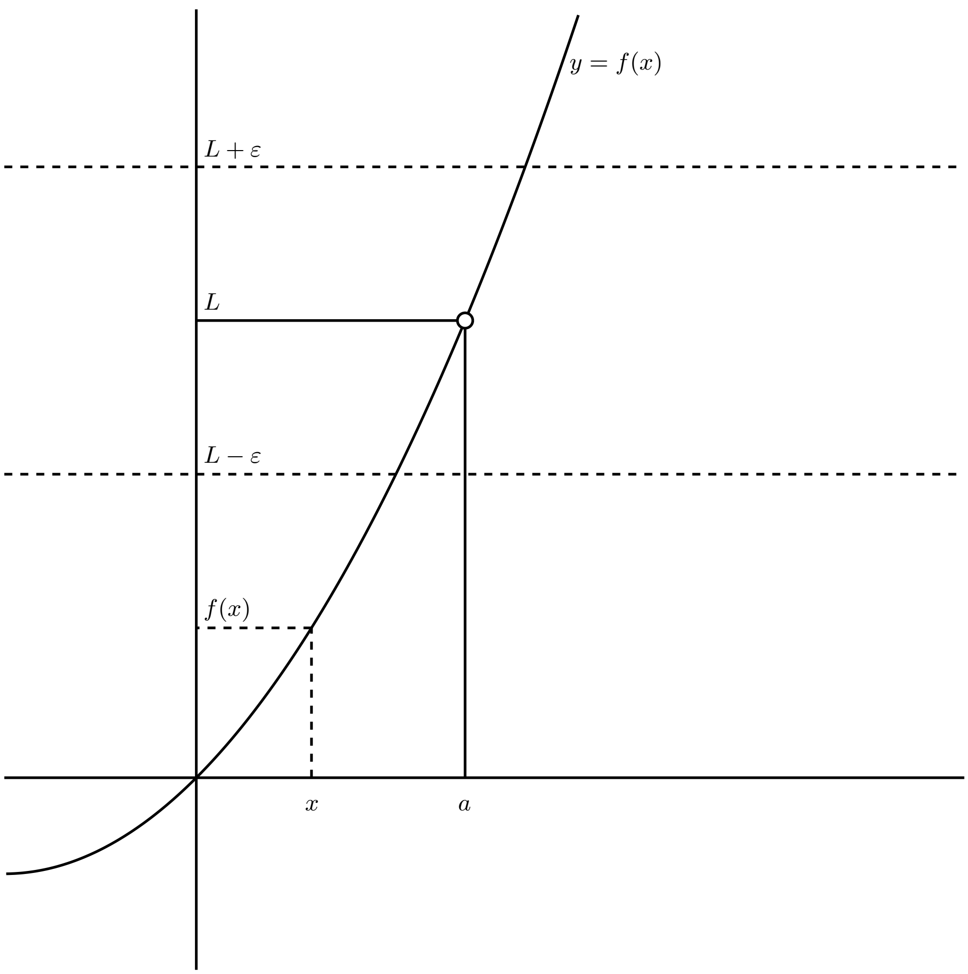


Figure 5.
It looks like $\lim_{x\to a} f(x) = L$. How close do you have to choose x to guarantee that $f(x)$ differs less than ε from the apparent limit L?
See the next figure for an answer.

12.1. *Show that* $\lim_{x\to 3} 3x+2 = 11$

We have $g(x) = 3x = 2$, $a = 3$ and $L = 11$. We want to show that $|g(x) - L| < \varepsilon$ follows from $|x - a| < \delta$ provided we choose $\delta > 0$ small enough (depending on ε). To see how small we have to choose δ we look at $|g(x) - L|$

$$|g(x) - L| = \big|(3x+2) - 11\big| = |3x - 9| = 3 \cdot |x - 3| = 3 \cdot |x - a|.$$

So $|x - a| < \delta$ implies $|g(x) - L| < 3\delta$. To guarantee $|g(x) - L| < \varepsilon$ we must therefore choose $3\delta \leq \varepsilon$. If we choose $\delta = \frac{1}{3}\varepsilon$, then the conclusion is that for all $x \neq 3$ it follows from $|x - 3| < \delta$ that $|g(x) - L| < \varepsilon$. Therefore $\lim_{x\to 3} g(x) = L$.

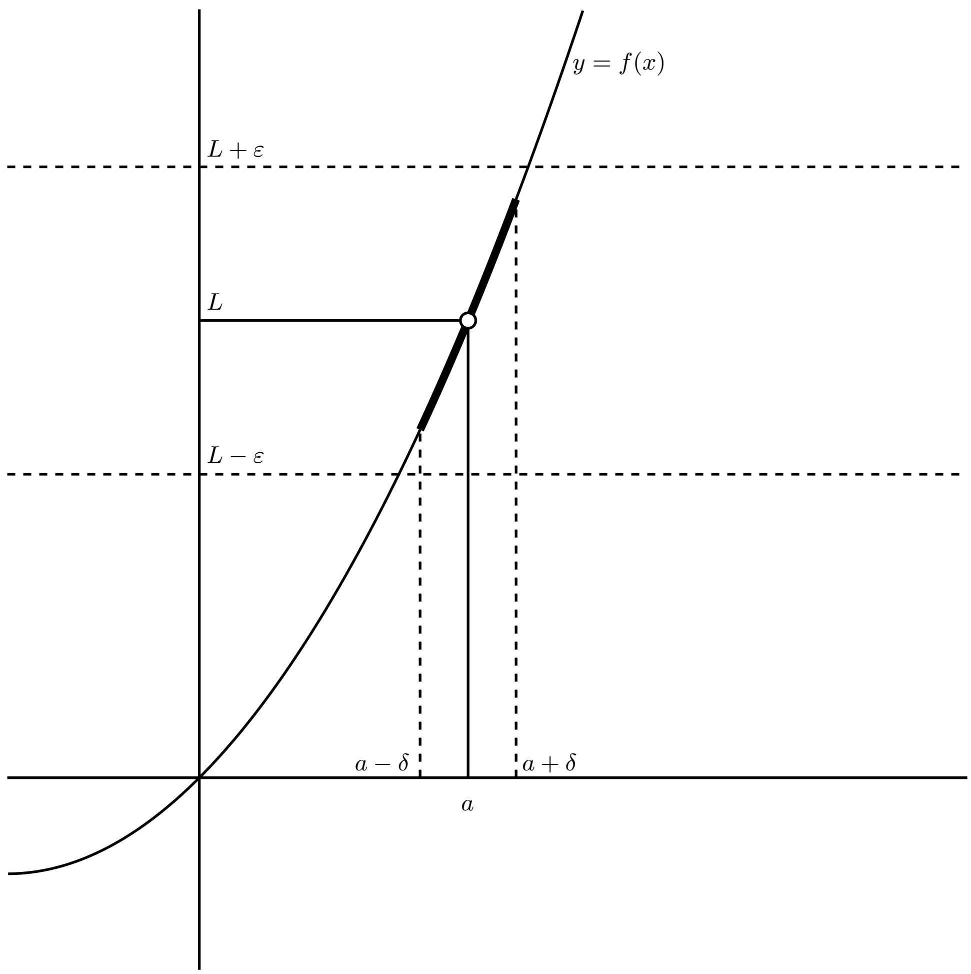


Figure 6.
If you choose δ as in this figure, then $|x - a| < \delta$ implies $|f(x) - L| < \varepsilon$. The same would have still been true if we had chosen a smaller δ, and a slightly larger δ would also work. The important thing is that one can find such a δ.

12.2. *Show that* $\lim_{x\to 1} x^2 = 1$

Here $g(x) = x^2$, $a = 1$ and $L = 1$. We want to show that choosing $|x - 1| < \delta$ implies $|x^2 - 1| < \varepsilon$ provided we choose the right δ (which is allowed to depend on ε.)

We begin by estimating the difference $|x^2 - 1|$

$$|x^2 - 1| = |(x - 1)(x + 1)| = |x - 1| \cdot |x + 1|.$$

If $|x - 1| < \delta$ then $1 - \delta < x < 1 + \delta$ and thus

$$|x + 1| < 2 + \delta$$

so that $|x - 1| < \delta$ implies

$$|x^2 - 1| < \delta(2 + \delta).$$

We want to make this less than ε by choosing δ small enough. One approach would be to solve the (quadratic) inequality

$$\delta(2 + \delta) < \varepsilon$$

for δ but there is a trick that steers us away from quadratic equations. Namely, if we are going to choose δ small anyway, why not agree to choose $\delta \le 1$? If we promise never to choose $\delta > 1$, then $|x - 1| < \delta$ implies

$$|x^2 - 1| < \delta(2 + \delta) \le 3\delta.$$

To guarantee that this does not exceed ε we choose $3\delta \le \varepsilon$, i.e. $\delta = \frac{1}{3}\varepsilon$. This requirement together with our earlier promise to choose $\delta \le 1$ leads us to this choice of δ

$$\delta = \text{the smaller of 1 and } \frac{\varepsilon}{3}.$$

We have shown that if you choose δ this way, then $|x - 1| < \delta$ implies $|x^2 - 1| < \varepsilon$, no matter what $\varepsilon > 0$ is.

12.3. *Show that* $\lim_{x\to 4} 1/x = 1/4$

Solution: We apply the definition with $a = 4$, $L = 1/4$ and $g(x) = 1/x$. Thus, for given $\varepsilon > 0$ we try to show that if $|x - 4|$ is small enough then one has $|g(x) - 1/4| < \varepsilon$.

We begin by estimating $|g(x) - \frac{1}{4}|$ in terms of $|x - 4|$:

$$|g(x) - 1/4| = \left|\frac{1}{x} - \frac{1}{4}\right| = \frac{|x - 4|}{|4x|}.$$

This quantity will be small if $|x - 4|$ is small, except if division by $4x$ makes the quotient larger. For that to happen x would have to be close to zero (division by a small number gives a large number). By requiring x to be close to 4 we can keep x away from zero. The way to do this is to agree now that we will always take $\delta \le 3$, no matter what ε is. Then $|x - 4| < \delta$ implies

$$|x - 4| < \delta \le 3 \implies 1 < x < 7 \implies |4x| > 4 \implies \frac{1}{|4x|} < \frac{1}{4}$$

and hence,

$$|g(x) - 1/4| = \frac{|x - 4|}{|4x|} < \frac{\delta}{4}.$$

Hence if we choose $\delta = 4\varepsilon$ or any smaller number, then $|x - 4| < \delta$ implies $|g(x) - 4| < \varepsilon$. Of course we have to honor our agreement never to choose $\delta > 3$, so our choice of δ is

$$\delta = \text{the smaller of 3 and } 4\varepsilon.$$

Exercises

<u>*12.1*</u> – Use the ε–δ definition to prove that

(**a**) $\lim_{x\to 1} 2x-4=6$ (**b**) $\lim_{x\to 2} x^2-7x+3=-7$

(**c**) $\lim_{x\to 3} x^3=27$ (**d**) $\lim_{x\to 0} \sqrt{|x|}=0$

13. **Variations on the limit theme**

Not all limits are "for $x\to a$." here we describe some possible variations on the concept of limit.

13.1. *Left and right limits*

When we let "x approach a" we allow x to be both larger or smaller than a, as long as x gets close to a. If we explicitly want to study the behaviour of $g(x)$ as x approaches a through values larger than a, then we write

$$\lim_{x\searrow a} g(x) \text{ or } \lim_{x\to a+} g(x) \text{ or } \lim_{x\to a+0} g(x) \text{ or } \lim_{x\to a, x>a} g(x).$$

All four notations are in use. Similarly, to designate the value which $g(x)$ approaches as x approaches a through values below a one writes

$$\lim_{x\nearrow a} g(x) \text{ or } \lim_{x\to a-} g(x) \text{ or } \lim_{x\to a-0} g(x) \text{ or } \lim_{x\to a, x<a} g(x).$$

The precise definition of right limits goes like this:

Definition. 13.1 (Definition of right-limits). *Let g be a function. Then*

$$\lim_{x\searrow a} g(x)=L. \tag{10}$$

means that for every $\varepsilon>0$ one can find a $\delta>0$ such that

$$a<x<a+\delta \implies |g(x)-L|<\varepsilon$$

holds for all x in the domain of g.

The left-limit, i.e. the one-sided limit in which x approaches a through values less than a is defined in a similar way. The following theorem tells you how to use one-sided limits to decide if a function $g(x)$ has a limit at $x=a$.

Theorem 13.2. *If both one-sided limits*

$$\lim_{x\searrow a} g(x)=L_+, \text{ and } \lim_{x\nearrow a} g(x)=L_-$$

exist, then

$$\lim_{x\to a} g(x) \text{ exists} \iff L_+=L_-.$$

In other words, if a function has both left- and right-limits at some $x=a$, then that function has a limit at $x=a$ if the left- and right-limits are equal.

13.2. *Limits at infinity.*

Instead of letting x approach some finite number, one can let x become "larger and larger" and ask what happens to $g(x)$. If there is a number L such that $g(x)$ gets arbitrarily close to L if one chooses x sufficiently large, then we write

$$\lim_{x\to\infty} g(x)=L, \quad \text{or} \quad \lim_{x\uparrow\infty} g(x)=L, \quad \text{or} \quad \lim_{x\nearrow\infty} g(x)=L.$$

("The limit for x going to infinity is L.")

13.3. *Example – Limit of* $1/x$

The larger you choose x, the smaller its reciprocal $1/x$ becomes. Therefore, it seems reasonable to say

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

Here is the precise definition:

Definition. 13.3 (Definition of limit at ∞). *Let g be some function which is defined on some interval $x_0 < x < \infty$. If there is a number L such that for every $\varepsilon > 0$ one can find an A such that*

$$x > A \implies |g(x) - L| < \varepsilon$$

for all x, then we say that the limit of $g(x)$ for $x \to \infty$ is L.

The definition is very similar to the original definition of the limit. Instead of δ which specifies how close x should be to a, we now have a number A which says how large x should be, which is a way of saying "how close x should be to infinity."

13.4. *Example – Limit of* $1/x$ *(again)*

To ***prove*** that $\lim_{x\to\infty} 1/x = 0$ we apply the definition to $g(x) = 1/x$, $L = 0$.

For given $\varepsilon > 0$ we need to show that

$$\left|\frac{1}{x} - L\right| < \varepsilon \text{ for all } x > A \tag{11}$$

provided we choose the right A.

How do we choose A? A is not allowed to depend on x, but it may depend on ε.

If we assume for now that we will only consider positive values of x, then (11) simplifies to

$$\frac{1}{x} < \varepsilon$$

which is equivalent to

$$x > \frac{1}{\varepsilon}.$$

This tells us how to choose A. Given any positive ε, we will simply choose

$$A = \frac{1}{\varepsilon}$$

Then one has $|\frac{1}{x} - 0| = \frac{1}{x} < \varepsilon$ for all $x > A$. Hence we have proved that $\lim_{x\to\infty} 1/x = 0$.

14. Properties of the Limit

The precise definition of the limit is not easy to use, and fortunately we won't use it very often in this class. Instead, there are a number of properties that limits have which allow you to compute them without having to resort to "epsiloncy."

The following properties also apply to the variations on the limit from 13. I.e. the following statements remain true if one replaces each limit by a one-sided limit, or a limit for $x \to \infty$.

Limits of constants and of x. If a and c are constants, then

$$\lim_{x\to a} c = c \tag{P_1}$$

and

$$\lim_{x\to a} x = a. \tag{P_2}$$

Limits of sums, products and quotients. Let F_1 and F_2 be two given functions whose limits for $x \to a$ we know,

$$\lim_{x\to a} F_1(x) = L_1, \qquad \lim_{x\to a} F_2(x) = L_2.$$

Then

$$(P_3) \qquad \lim_{x\to a} \big(F_1(x) + F_2(x)\big) = L_1 + L_2,$$

$$(P_4) \qquad \lim_{x\to a} \big(F_1(x) - F_2(x)\big) = L_1 - L_2,$$

$$(P_5) \qquad \lim_{x\to a} \big(F_1(x) \cdot F_2(x)\big) = L_1 \cdot L_2$$

Finally, if $\lim_{x\to a} F_2(x) \neq 0$,

$$(P_6) \qquad \lim_{x\to a} \frac{F_1(x)}{F_2(x)} = \frac{L_1}{L_2}.$$

In other words the limit of the sum is the sum of the limits, etc. One can prove these laws using the definition of limit in §12 but we will not do this here. However, I hope these laws seem like common sense: if, for x close to a, the quantity $F_1(x)$ is close to L_1 and $F_2(x)$ is close to L_2, then certainly $F_1(x) + F_2(x)$ should be close to $L_1 + L_2$.

There are two more properties of limits which we will add to this list later on. They are the "Sandwich Theorem" (§18) and the substitution theorem (§19).

15. Examples of limit computations

15.1. *Find* $\lim_{x\to 2} x^2$

One has

$$\begin{aligned} \lim_{x\to 2} x^2 &= \lim_{x\to 2} x \cdot x \\ &= \big(\lim_{x\to 2} x\big) \cdot \big(\lim_{x\to 2} x\big) && \text{by } (P_5) \\ &= 2 \cdot 2 = 4. \end{aligned}$$

Similarly,

$$\begin{aligned} \lim_{x\to 2} x^3 &= \lim_{x\to 2} x \cdot x^2 \\ &= \big(\lim_{x\to 2} x\big) \cdot \big(\lim_{x\to 2} x^2\big) && (P_5) \text{ again} \\ &= 2 \cdot 4 = 8, \end{aligned}$$

and, by (P_4)

$$\lim_{x\to 2} x^2 - 1 = \lim_{x\to 2} x^2 - \lim_{x\to 2} 1 = 4 - 1 = 3,$$

and, by (P_4) again,

$$\lim_{x\to 2} x^3 - 1 = \lim_{x\to 2} x^3 - \lim_{x\to 2} 1 = 8 - 1 = 7,$$

Putting all this together, one gets

$$\lim_{x\to 2} \frac{x^3 - 1}{x^2 - 1} = \frac{2^3 - 1}{2^2 - 1} = \frac{8 - 1}{4 - 1} = \frac{7}{3}$$

because of (P_6). To apply (P_6) we must check that the denominator ("L_2")is not zero. Since the denominator is 3 everything is OK, and we were allowed to use (P_6).

15.2. *Try the examples 11.2 and 11.3 using the limit properties*

To compute $\lim_{x\to 2}(x^2-2x)(x^2-4)$ we first use the limit properties to find

$$\lim_{x\to 2} x^2-2x = 0 \text{ and } \lim_{x\to 2} x^2-4 = 0.$$

to complete the computation we would like to apply the last property (P_6) about quotients, but this would give us

$$\lim_{x\to 2} g(x) = \frac{0}{0}.$$

The denominator is zero, so we were not allowed to use (P_6) (and the result doesn't mean anything anyway). We have to do something else.

The function we are dealing with is a ***rational function***, which means that its the quotient of two polynomials. For such functions there is an algebra trick which always allows you to compute the limit even if you first get $\frac{0}{0}$. The thing to do is to divide numerator and denominator by $x-2$. In our case we have

$$x^2-2x = (x-2)\cdot x, \qquad x^2-4 = (x-2)\cdot(x+2)$$

so that

$$\lim_{x\to 2} g(x) = \lim_{x\to 2} \frac{(x-2)\cdot x}{(x-2)\cdot(x+2)} = \lim_{x\to 2}\frac{x}{x+2}.$$

After this simplification we ***can*** use the properties ($P_{...}$) to compute

$$\lim_{x\to 2} g(x) = \frac{2}{2+2} = \frac{1}{2}.$$

The point in the above example is that

$$\frac{x^2-2x}{x^2-4} = \frac{x}{x+2}$$

for $x \neq 2$ and the right hand side (but not the left) is meaningful even when $x=2$.

15.3. *Example – Find* $\lim_{x\to 2}\sqrt{x}$

There is nothing in the limit properties which tells us how to deal with a square root, and using them we can't prove that there is a limit. However, if you assume that the limit exists, i.e. that there is a number L for which $\sqrt{x}$ gets "closer and closer" to L as x "approaches 2," then the limit properties allow us to find this the limit L.

So, suppose that there is a number L with

$$\lim_{x\to 2}\sqrt{x} = L.$$

Then property (P_5) implies that

$$L^2 = \big(\lim_{x\to 2}\sqrt{x}\big)\cdot\big(\lim_{x\to 2}\sqrt{x}\big) = \lim_{x\to 2}\sqrt{x}\cdot\sqrt{x} = \lim_{x\to 2} x = 2.$$

In other words, $L^2=2$, and hence L must be either $\sqrt{2}$ or $-\sqrt{2}$. We can reject the latter because whatever x does, its squareroot is always a positive number, and hence it can never "get close to" a negative number like $-\sqrt{2}$.

Our conclusion: if the limit exists, then

$$\lim_{x\to 2}\sqrt{x} = \sqrt{2}.$$

The result is not surprising: if x gets close to 2 then $\sqrt{x}$ gets close to $\sqrt{2}$.

15.4. *Example – Find* $\lim_{x\to 2}\sqrt{x}$

This is the same question as in the previous example. There we showed how the limit properties imply that the only possible value of the limit is $\sqrt{2}$, but we didn't show that $\sqrt{2}$ actually is the limit. For that one has to wield εs and δs. Here's how:

Assuming $|x-2|<\delta$ we estimate the difference $|\sqrt{x}-\sqrt{2}|$ as follows

$$|\sqrt{x}-\sqrt{2}| = \left|\frac{x-2}{\sqrt{x}+\sqrt{2}}\right| = \frac{|x-2|}{\sqrt{x}+\sqrt{2}}.$$

Since $\sqrt{x}\geq 0$ for all x we have $\sqrt{x}+\sqrt{2}\geq\sqrt{2}$, and thus if $|x-2|\leq\delta$

$$|\sqrt{x}-\sqrt{2}| \leq \frac{|x-2|}{\sqrt{2}} < \frac{\delta}{\sqrt{2}}.$$

If we want $|\sqrt{x}-\sqrt{2}|<\varepsilon$, then we should choose $\frac{\delta}{\sqrt{2}}\leq\varepsilon$. Thus with

$$\delta = \varepsilon\sqrt{2}$$

we find that $|x-2|<\delta$ always implies $|\sqrt{x}-\sqrt{2}|<\varepsilon$, no matter which $\varepsilon>0$ is given.

15.5. *Example – The derivative of* $\sqrt{x}$ *at* $x=2$.

Find

$$\lim_{x\to 2}\frac{\sqrt{x}-\sqrt{2}}{x-2}$$

assuming the result from the previous example.

Solution: The function is a rational function whose numerator and denominator vanish when $x=2$, i.e. the limit is of the form $\frac{0}{0}$. We use the same algebra trick as before, namely we factor numerator and denominator:

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

Now one can use the limit properties to compute

$$\lim_{x\to 2}\frac{\sqrt{x}-\sqrt{2}}{x-2} = \lim_{x\to 2}\frac{1}{\sqrt{x}+\sqrt{2}} = \frac{1}{2\sqrt{2}} = \frac{\sqrt{2}}{4}.$$

15.6. *Limit as* $x\to\infty$ *of rational functions*

A rational function is the quotient of two polynomials, so

$$R(x) = \frac{a_n x^n + \cdots + a_1 x + a_0}{b_m x^m + \cdots + b_1 x + b_0}. \tag{12}$$

We have seen that

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

We even proved this in example 13.4. Using this you can find the limit at ∞ for any rational function $R(x)$ as in (12). One could turn the outcome of the calculation of $\lim_{x\to\infty}R(x)$ into a recipe/formula involving the degrees n and m of the numerator and denominator, and also their coefficients a_i, b_j, which students would then memorize, but it is better to remember "the trick."

To find $\lim_{x\to\infty}R(x)$ divide numerator and denominator by x^m (the highest power of x occurring in the denominator).

For example, let's compute

$$\lim_{x\to\infty}\frac{3x^2+3}{5x^2+7x-39}.$$

Remember the trick and divide top and bottom by x^2, and you get

$$\begin{aligned}\lim_{x\to\infty}\frac{3x^2+3}{5x^2+7x-39} &= \lim_{x\to\infty}\frac{3+3/x^2}{5+7/x-39/x^2}\\ &= \frac{\lim_{x\to\infty}3+3/x^2}{\lim_{x\to\infty}5+7/x-39/x^2}\\ &= \frac{3}{5}\end{aligned}$$

Here we have used the limit properties (P_*) to break the limit down into little pieces like $\lim_{x\to\infty}39/x^2$ which we can compute as follows

$$\lim_{x\to\infty}39/x^2 = \lim_{x\to\infty}39\cdot\left(\frac{1}{x}\right)^2 = \left(\lim_{x\to\infty}39\right)\cdot\left(\lim_{x\to\infty}\frac{1}{x}\right)^2 = 39\cdot 0^2 = 0.$$

15.7. *Another example with a rational function*

Compute

$$\lim_{x\to\infty}\frac{x}{x^3+5}.$$

We apply "the trick" again and divide numerator and denominator by x^3. This leads to

$$\lim_{x\to\infty}\frac{x}{x^3+5} = \lim_{x\to\infty}\frac{1/x^2}{1+5/x^3} = \frac{\lim_{x\to\infty}1/x^2}{\lim_{x\to\infty}1+5/x^3} = \frac{0}{1} = 0.$$

To show all possible ways a limit of a rational function can turn out we should do yet another example, but that one belongs in the next section (see example 16.5.)

16. **When limits fail to exist**

In the last couple of examples we worried about the possibility that a limit $\lim_{x\to a}g(x)$ actually might not exist. This can actually happen, and in this section we'll see a few examples of what failed limits look like. First let's agree on what we will call a "failed limit."

Definition. 16.1. *If there is no number L such that $\lim_{x\to a}g(x) = L$, then we say that the limit $\lim_{x\to a}g(x)$ does not exist.*

16.1. *The sign function near $x = 0$*

The "sign function" is defined[2]

$$\text{sign}(x) = \begin{cases}-1 & \text{for } x<0\\ 0 & \text{for } x=0\\ 1 & \text{for } x>0\end{cases}$$

Note that the sign of zero is zero. But does the sign function has a limit at $x = 0$, i.e. does

$$\lim_{x\to 0}\text{sign}(x)$$

exist? And is it also zero? The answer is ***no*** and ***no***, and here is why: suppose that for some number L one had

$$\lim_{x\to 0}\text{sign}(x) = L,$$

[2] Some people don't like the notation $\text{sign}(x)$, and prefer to write

$$g(x) = \frac{x}{|x|}$$

instead of $g(x) = \text{sign}(x)$. If you think about this formula for a moment you'll see that $\text{sign}(x) = x/|x|$ for all $x \neq 0$. When $x = 0$ the quotient $x/|x|$ is of course not defined.

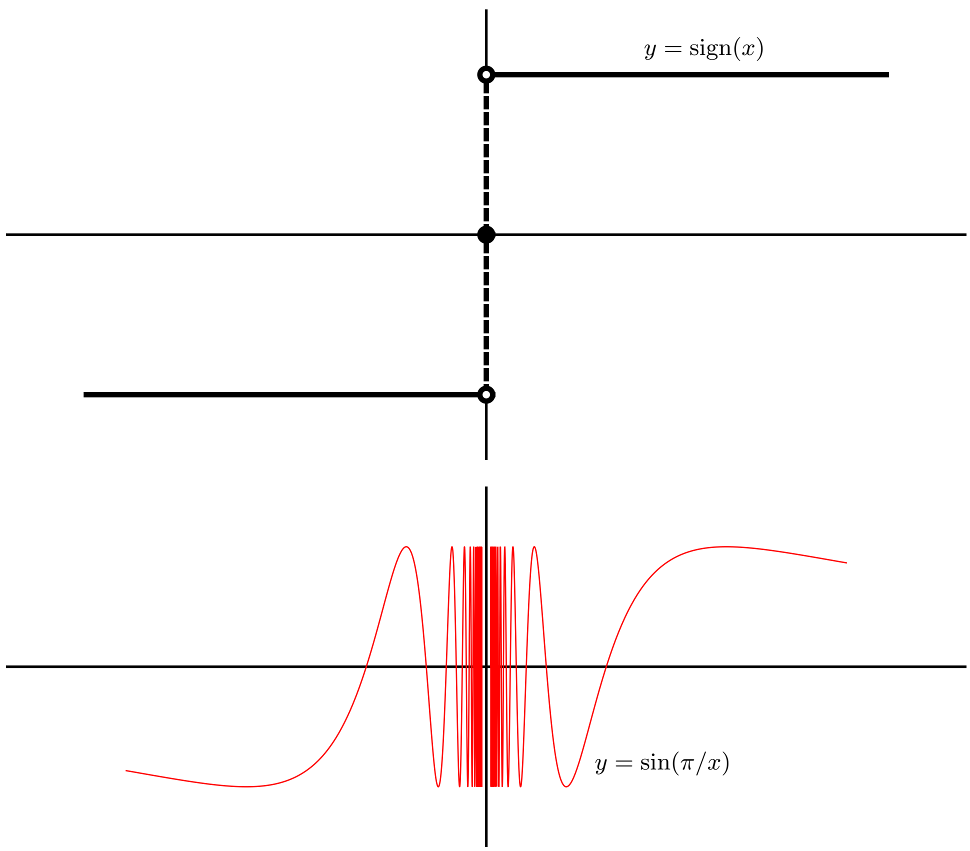
then since for arbitrary small positive values of x one has $\operatorname{sign}(x) = +1$ one would think that $L = +1$. But for arbitrarily small negative values of x one has $\operatorname{sign}(x) = -1$, so one would conclude that $L = -1$. But one number L can't be both $+1$ and -1 at the same time, so there is no such L, i.e. there is no limit.

$$\lim_{x\to 0} \operatorname{sign}(x) \text{ does not exist.}$$

In this example the one-sided limits do exist, namely,

$$\lim_{x\searrow 0} \operatorname{sign}(x) = 1 \text{ and } \lim_{x\nearrow 0} \operatorname{sign}(x) = -1.$$

All this says is that when x approaches 0 through positive values, its sign approaches $+1$, while if x goes to 0 through negative values, then its sign approaches -1.



Graphs of $g(x) = \operatorname{sign}(x)$ and $g(x) = \sin(1/x)$
for $-3 < x < 3$, $x \neq 0$.

16.2. *The example of the backward sine*

Contemplate the limit as $x \to 0$ of the "backward sine," i.e.

$$\lim_{x\to 0} \sin\left(\frac{1}{x}\right).$$

When $x = 0$ the function $g(x) = \sin(1/x)$ is not defined, because its definition involves division by x. What happens to $g(x)$ as $x \to 0$? First, $1/x$ becomes larger and larger ("goes to infinity") as $x \to 0$. Then, taking the sine, we see that $\sin(1/x)$ oscillates between $+1$ and -1 infinitely often as $x \to 0$. This means that $g(x)$ gets close to any number between -1 and $+1$ as $x \to 0$, but that the function $g(x)$ ***never stays close*** to any particular value because it keeps oscillating up and down.

Here again, the limit $\lim_{x\to 0} g(x)$ does not exist. We have arrived at this conclusion by only considering what $g(x)$ does for small positive values of x. So the limit fails to exist in a stronger way than in the example of the sign-function. There, even though the limit didn't exist, the one-sided limits existed. In the present example we see that even the one-sided limit

$$\lim_{x\searrow 0} \sin \frac{1}{x}$$

does not exist.

16.3. *Trying to divide by zero using a limit*

The expression $1/0$ is not defined, but what about

$$\lim_{x\to 0} \frac{1}{x}?$$

This limit also does not exist. Here are two reasons:

It is common wisdom that if you divide by a small number you get a large number, so if you divide 1 by "smaller and smaller" numbers, the result will get "larger and larger." In particular, it will not be able to stay close to any particular finite number. So the limit can't exist.

"Common wisdom" is not always a reliable tool in mathematical proofs, so here is a better argument. The limit can't exist, because that would contradict the limit properties $(P_1)\cdots(P_6)$. Namely, suppose that there were an number L such that

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

Then the limit property (P_5) would imply that

$$\lim_{x\to 0} \big(\frac{1}{x} \cdot x\big) = \big(\lim_{x\to 0} \frac{1}{x}\big) \cdot \big(\lim_{x\to 0} x\big) = L \cdot 0 = 0.$$

On the other hand $\frac{1}{x} \cdot x = 1$ so the above limit should be 1! A number can't be both 0 and 1 at the same time, so we have a contradiction. The assumption that $\lim_{x\to 0} 1/x$ exists is to blame, so it must go.

16.4. *Using limit properties to show a limit does* **not** *exist*

The limit properties tell us how to prove that certain limits exist (and how to compute them). Although it is perhaps not so obvious at first sight, they also allow you to prove that certain limits do not exist. The previous example shows one instance of such use. Here is another.

Property (P_3) says that if both $\lim_{x\to a} g(x)$ and $\lim_{x\to a} h(x)$ exist then $\lim_{x\to a} g(x) + h(x)$ also must exist. You can turn this around and say that if $\lim_{x\to a} g(x) + h(x)$ does not exist then either $\lim_{x\to a} g(x)$ or $\lim_{x\to a} h(x)$ does not exist (or both limits fail to exist).

For instance, the limit

$$\lim_{x\to 0} \frac{1}{x} - x$$

can't exist, for if it did, then the limit

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

would also have to exist, and we know $\lim_{x\to 0} \frac{1}{x}$ doesn't exist.

16.5. *Limits at ∞ which don't exist*

If you let x go to ∞, then x will not get "closer and closer" to any particular number L, so it seems reasonable to guess that

$$\lim_{x\to\infty} x \text{ does not exist.}$$

One can prove this from the limit definition (and see exercise 2).

Let's consider

$$L = \lim_{x\to\infty} \frac{x^2+2x-1}{x+2}.$$

Once again we divide numerator and denominator by the highest power in the denominator (i.e. x)

$$L = \lim_{x\to\infty} \frac{x+2-\frac{1}{x}}{1+2/x}$$

Here the denominator has a limit ('tis 1), but the numerator does not, for if $\lim_{x\to\infty} x + 2 - \frac{1}{x}$ existed then, since $\lim_{x\to\infty}(2-1/x) = 2$ exists,

$$\lim_{x\to\infty} x = \lim_{x\to\infty} \Big[\big(x+2-\frac{1}{x}\big) - \big(2-\frac{1}{x}\big)\Big]$$

would also have to exist, and $\lim_{x\to\infty} x$ doesn't exist.

So we see that L is the limit of a fraction in which the denominator has a limit, but the numerator does not. In this situation the limit L itself can never exist. If it did, then

$$\lim_{x\to\infty} \big(x+2-\frac{1}{x}\big) = \lim_{x\to\infty} \frac{x+2-\frac{1}{x}}{1+2/x} \cdot (1+2/x)$$

would also have to have a limit.

17. **What's in a name?**

There is a big difference between the variables x and a in the formula

$$\lim_{x\to a} 2x+1,$$

namely a is a ***free variable***, while x is a ***dummy variable*** (or "placeholder" or a "bound variable.")

The difference between these two kinds of variables is this:

- if you replace a dummy variable in some formula consistently by some other variable then the value of the formula does not change. On the other hand, it never makes sense to substitute a number for a dummy variable.
- the value of the formula may depend on the value of the free variable.

To understand what this means consider the example $\lim_{x\to a} 2x + 1$ again. The limit is easy to compute:

$$\lim_{x\to a} 2x+1 = 2a+1.$$

If we replace x by, say u (systematically) then we get

$$\lim_{u\to a} 2u+1$$

which is again equal to $2a+1$. This computation says that *if some number gets close to a then two times that number plus one gets close to* $2a+1$. This is a very wordy way of expressing the formula, and you can shorten things by giving a name (like x or u) to the number which approaches a. But the result of our computation shouldn't depend on the name we choose, i.e. it doesn't matter if we call it x or u.

Since the name of the variable x doesn't matter it is called a dummy variable. Some prefer to call x a bound variable, meaning that in

$$\lim_{x\to a} 2x + 1$$

the x in the expression $2x + 1$ is bound to the x written underneath the limit – you can't change one without changing the other.

Substituting a number for a dummy variable usually leads to complete nonsense. For instance, let's try setting $x = 3$ in our limit, i.e. what is

$$\lim_{3\to a} 2 \cdot 3 + 1\,?$$

Of course $2 \cdot 3 + 1 = 7$, but what does 7 do when 3 gets closer and closer to the number a? That's a silly question, because 3 is a constant and it doesn't "get closer" to some other number like a! If you ever see 3 get closer to another number then it's time to take a vacation.

On the other hand the variable a is free: you can assign it particular values, and its value will affect the value of the limit. For instance, if we set $a = 3$ (but leave x alone) then we get

$$\lim_{x\to 3} 2x + 1$$

and there's nothing strange about that (the limit is $2 \cdot 3 + 1 = 7$, no problem.) You could substitute other values of a and you would get a different answer. In general you get $2a + 1$.

18. Limits and Inequalities

This section has two theorems which let you compare limits of different functions. The properties in these theorems are not formulas that allow you to compute limits like the properties $(P_1) \ldots (P_6)$ from §14. Instead, they allow you to ***reason*** about limits, i.e. they let you say that this or that limit is positive, or that it must be the same as some other limit which you find easier to think about.

The first theorem should not surprise you – all it says is that bigger functions have bigger limits.

Theorem 18.1. *Let f and g be functions whose limits for $x \to a$ exist, and assume that $f(x) \le g(x)$ holds for all x. Then*

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

A useful special case arises when you set $f(x) = 0$. The theorem then says that if a function g never has negative values, then its limit will also never be negative.

The statement may seem obvious, but it still needs a proof, starting from the ε-δ definition of limit. This will be done in lecture.

Here is the second theorem about limits and inequalities.

Theorem 18.2 (The Sandwich Theorem). *Suppose that*

$$f(x) \le g(x) \le h(x)$$

(for all x) and that

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

Then

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

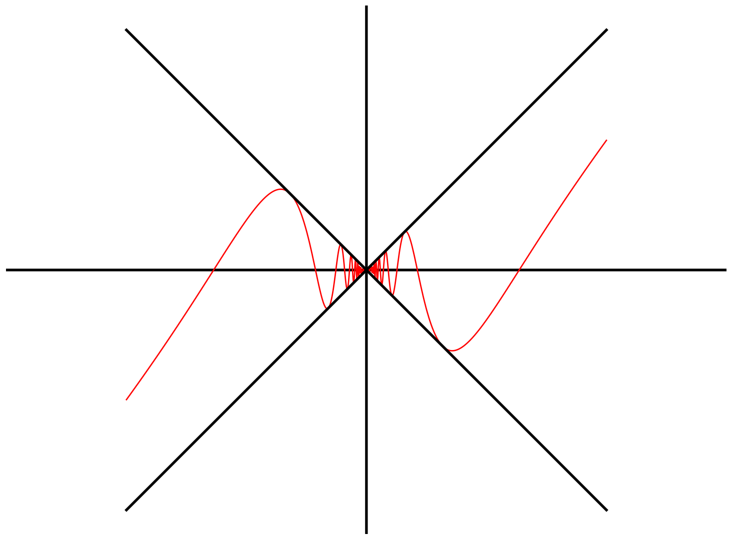


Figure 7. Graphs of $|x|$, $-|x|$ and $x\cos(1/x)$ for $-1.2 < x < 1.2$

The theorem is useful when you want to know the limit of g, and when you can ***sandwich*** it between two functions f and h whose limits are easier to compute. The Sandwich Theorem looks like the first theorem of this section, but there is an important difference: in the Sandwich Theorem you don't have to assume that the limit of g exists. The inequalities $f \le g \le h$ combined with the circumstance that f and h have the same limit are enough to guarantee that the limit of g exists.

18.1. *A backward cosine sandwich*

The Sandwich Theorem says that if the function $g(x)$ is sandwiched between two functions $f(x)$ and $h(x)$ and the limits of the outside functions f and h exist and are equal, then the limit of the inside function g exists and equals this common value. For example

$$-|x| \le x\cos\frac{1}{x} \le |x|$$

since the cosine is always between -1 and 1. Since

$$\lim_{x\to 0} -|x| = \lim_{x\to 0} |x| = 0$$

the sandwich theorem tells us that

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

Note that the limit $\lim_{x\to 0}\cos(1/x)$ does ***not*** exist, for the same reason that the "backward sine" did not have a limit for $x \to 0$ (see example 16.2). Multiplying with x changed that.

19. **Continuity**

Definition. 19.1. *A function g is **continuous** at a if*

$$\lim_{x\to a} g(x) = g(a) \tag{13}$$

A function is continuous if it is continuous at every a in its domain.

Note that when we say that a function is continuous on some interval it is understood that the domain of the function includes that interval. For example, the function $f(x) = 1/x^2$ is continuous on the interval $1 < x < 5$ but is ***not*** continuous on the interval $-1 < x < 1$.

19.1. *Polynomials are continuous*

For instance, let us show that $P(x) = x^2 + 3x$ is continuous at $x = 2$. To show that you have to prove that

$$\lim_{x\to 2} P(x) = P(2),$$

i.e.

$$\lim_{x\to 2} x^2 + 3x = 2^2 + 3 \cdot 2.$$

You can do this two ways: using the definition with ε and δ (i.e. the hard way), or using the limit properties $(P_1)\dots(P_6)$ from §14 (just as good, and easier, even though it still takes a few lines to write it out – do both!)

19.2. *Rational functions are continuous*

Let $R(x) = \frac{P(x)}{Q(x)}$ be a rational function, and let a be any number in the domain of R, i.e. any number for which $Q(a) \neq 0$. Then one has

$$\begin{aligned}
\lim_{x\to a} R(x) &= \lim_{x\to a} \frac{P(x)}{Q(x)} \\
&= \frac{\lim_{x\to a} P(x)}{\lim_{x\to a} Q(x)} && \text{property } (P_6) \\
&= \frac{P(a)}{Q(a)} && P \text{ and } Q \text{ are continuous} \\
&= R(a).
\end{aligned}$$

This shows that R is indeed continuous at a.

19.3. *Some discontinuous functions*

If $\lim_{x\to a} g(x)$ does not exist, then it certainly cannot be equal to $g(a)$, and therefore any failed limit provides an example of a discontinuous function.

For instance, the sign function $g(x) = \text{sign}(x)$ from example **??** is not continuous at $x = 0$.

Is the backward sine function $g(x) = \sin(1/x)$ from example 16.2 also discontinuous at $x = 0$? No, it is not, for two reasons: first, the limit $\lim_{x\to 0} \sin(1/x)$ does not exist, and second, we haven't even defined the function $g(x)$ at $x = 0$, so even if the limit existed, we would have no value $g(0)$ to compare it with.

19.4. *How to make functions discontinuous*

Here is a discontinuous function:

$$f(x) = \begin{cases} x^2 & \text{if } x \neq 3, \\ 47 & \text{if } x = 3. \end{cases}$$

In other words, we take a continuous function like $g(x) = x^2$, and change its value somewhere, e.g. at $x = 3$. Then

$$\lim_{x\to 3} f(x) = 9 \neq 47 = f(3).$$

The reason that the limit is 9 is that our new function $f(x)$ coincides with our old continuous function $g(x)$ for all x except $x = 3$. Therefore the limit of $f(x)$ as $x \to 3$ is the same as the limit of $g(x)$ as $x \to 3$, and since g is continuous this is $g(3) = 9$.

19.5. *Sandwich in a bow tie*

We return to the function from example 18.1. Consider

$$f(x) = \begin{cases} x\cos\left(\frac{1}{x}\right) & \text{for } x \neq 0, \\ 0 & \text{for } x = 0 \end{cases}$$

Then f is continuous at $x = 0$ by the Sandwich Theorem (see Example 18.1).

If we change the definition of f by picking a different value at $x = 0$ the new function will not be continuous, since changing f at $x = 0$ does not change the limit $\lim_{x\to 0} f(x)$. Since this limit is zero, $f(0) = 0$ is the only possible choice of $f(0)$ which makes f continuous at $x = 0$.

20. **Substitution in Limits**

Given two functions f and g one can consider their composition $h(x) = f(g(x))$. To compute the limit

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

we write $u = g(x)$, so that we want to know

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

Suppose that you can find the limits

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

Then it seems reasonable that as x approaches a, $u = g(x)$ will approach L, and $f(g(x))$ approaches M.

This is in fact a theorem:

Theorem 20.1. *If* $\lim_{x\to a} g(x) = a$*, and if the function* f *is continuous at* $u = L$*, then*

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

Another way to write this is

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

20.1. *Compute* $\lim_{x\to 3} \sqrt{x^3 - 3x^2 + 2}$

The given function is the composition of two functions, namely

$$\sqrt{x^3 - 3x^2 + 2} = \sqrt{u}, \text{ with } u = x^3 - 3x^2 + 2,$$

or, in function notation, we want to find $\lim_{x\to 3} h(x)$ where

$$h(x) = f(g(x)), \text{ with } g(x) = x^3 - 3x^2 + 2 \text{ and } g(x) = \sqrt{x}.$$

Either way, we have

$$\lim_{x\to 3} x^3 - 3x^2 + 2 = 2 \quad \text{and} \quad \lim_{u\to 2} \sqrt{u} = \sqrt{2}.$$

You get the first limit from the limit properties $(P_1)\ldots(P_5)$. The second limit says that taking the square root is a continuous function, which it is. We have not proved that (yet), but this particular limit is the one from example 15.3. Putting these two limits together we conclude that the limit is $\sqrt{2}$.

Normally, you write this whole argument as follows:

$$\lim_{x\to 3} \sqrt{x^3 - 3x^2 + 2} = \sqrt{\lim_{x\to 3} x^3 - 3x^2 + 2} = \sqrt{2},$$

where you must point out that $f(x) = \sqrt{x}$ is a continuous function to justify the first step.

Another possible way of writing this is

$$\lim_{x\to 3} \sqrt{x^3 - 3x^2 + 2} = \lim_{u\to 2} \sqrt{u} = \sqrt{2},$$

where you must say that you have substituted $u = x^3 - 3x^2 + 2$.

Exercises

20.1 – Find the following limits.

(**a**) $\lim_{x\to -7}(2x+5)$ (**b**) $\lim_{x\to 7-}(2x+5)$ (**c**) $\lim_{x\to -\infty}(2x+5)$

(**d**) $\lim_{x\to -4}(x+3)^{2006}$ (**e**) $\lim_{x\to -4}(x+3)^{2007}$ (**f**) $\lim_{x\to -\infty}(x+3)^{2007}$

(**g**) $\lim_{t\to 1}\dfrac{t^2+t-2}{t^2-1}$ (**h**) $\lim_{t\nearrow 1}\dfrac{t^2+t-2}{t^2-1}$ (**i**) $\lim_{t\to -1}\dfrac{t^2+t-2}{t^2-1}$

(**j**) $\lim_{x\to\infty}\dfrac{x^2+3}{x^2+4}$ (**k**) $\lim_{x\to\infty}\dfrac{x^5+3}{x^2+4}$ (**l**) $\lim_{x\to\infty}\dfrac{x^2+1}{x^5+2}$

(**m**) $\lim_{x\to\infty}\dfrac{(2x+1)^4}{(3x^2+1)^2}$ (**n**) $\lim_{u\to\infty}\dfrac{(2u+1)^4}{(3u^2+1)^2}$ (**o**) $\lim_{t\to 0}\dfrac{(2t+1)^4}{(3t^2+1)^2}$

20.2 – In the text we proved that $\lim_{x\to\infty}\frac{1}{x} = 0$. Show that this implies that $\lim_{x\to\infty} x$ does not exist. Hint: Suppose $\lim_{x\to\infty} x = L$ for some number L. Apply the limit properties to $\lim_{x\to\infty} x \cdot \dfrac{1}{x}$.

20.3 – Evaluate $\lim_{x\to 9}\dfrac{\sqrt{x}-3}{x-9}$. Hint: Multiply top and bottom by $\sqrt{x}+3$.

20.4 – Evaluate $\lim_{x\to 2}\dfrac{\frac{1}{x}-\frac{1}{2}}{x-2}$.

20.5 – Evaluate $\lim_{x\to 2}\dfrac{\frac{1}{\sqrt{x}}-\frac{1}{\sqrt{2}}}{x-2}$.

20.6 – A function f is defined by

$$f(x) = \begin{cases} x^3 & \text{for } x < -1 \\ ax+b & \text{for } -1 \le x < 1 \\ x^2+2 & \text{for } x \ge 1. \end{cases}$$

where a and b are constants. The function f is continuous. What are a and b?

21. Two Limits in Trigonometry

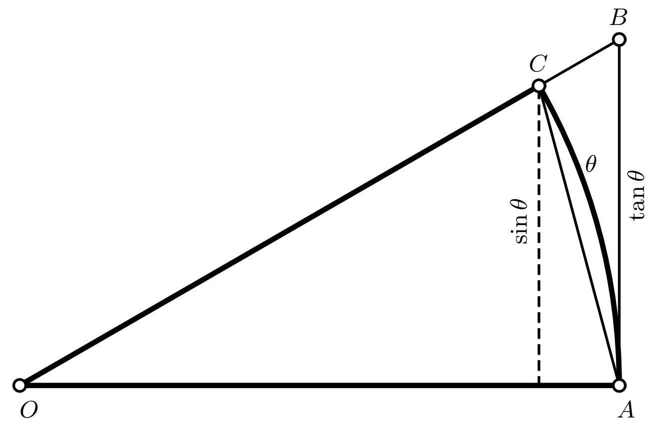
In this section we'll derive a few limits involving the trigonometric functions. You can think of them as saying that for small angles θ one has

$$\sin\theta \approx \theta \quad \text{and} \quad \cos\theta \approx 1 - \tfrac{1}{2}\theta^2.$$

We will use these limits when we compute the derivatives of Sine, Cosine and Tangent.

Theorem 21.1.

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



*The circular wedge OAC contains the triangle OAC and is contained in the right triangle OAB.
The area of triangle OAC is* $\frac{1}{2}\sin\theta$. *The area of circular wedge OAC is* $\frac{1}{2}\theta$.
The area of right triangle OAB is $\frac{1}{2}\tan\theta$.
Hence one has $\sin\theta < \theta < \tan\theta$ *for all angles* $0 < \theta < \pi/2$.

Figure 8. Proof that $\sin\theta \approx \theta$ for small values of θ

Proof. The proof requires a few sandwiches and some geometry.

We begin by only considering positive angles, and in fact we will only consider angles $0 < \theta < \pi/2$. By comparing areas in the drawing we see that for such angles one always has

$$\sin\theta < \theta < \tan\theta. \tag{14}$$

Since $\sin\theta > 0$ for $0 < \theta < \pi/2$ we get

$$0 < \sin\theta < \theta$$

for $0 < \theta < \pi/2$. As $\theta \searrow 0$ both 0 and θ go to zero, so the Sandwich Theorem implies that

$$\lim_{\theta\searrow 0} \sin\theta = 0.$$

Hence

$$\lim_{\theta\searrow 0} \cos\theta = \lim_{\theta\searrow 0} \sqrt{1-\sin^2\theta} = 1.$$

Finally we go back to the first sandwich (14) and divide it by θ

$$\frac{\sin\theta}{\theta} < 1 < \frac{\tan\theta}{\theta} = \frac{1}{\cos\theta}\,\frac{\sin\theta}{\theta}.$$

This implies

$$\cos\theta < \frac{\sin\theta}{\theta} < 1$$

The Sandwich Theorem can be used once again, and now it gives

$$\lim_{\theta\searrow 0} \frac{\sin\theta}{\theta} = 1.$$

This is a one-sided limit. To get the limit in which $\theta \nearrow 0$, you use that $\sin\theta$ is an odd function. □

Theorem 21.2.

$$\lim_{\theta\to 0}\frac{1-\cos\theta}{\theta^2}=\frac{1}{2}.$$

Proof. This follows from $\sin^2\theta+\cos^2\theta=1$. Namely,

$$\begin{aligned}\frac{1-\cos\theta}{\theta^2}&=\frac{1}{1+\cos\theta}\,\frac{1-\cos^2\theta}{\theta^2}\\&=\frac{1}{1+\cos\theta}\,\frac{\sin^2\theta}{\theta^2}\\&=\frac{1}{1+\cos\theta}\left\{\frac{\sin\theta}{\theta}\right\}^2.\end{aligned}$$

We have just shown that $\cos\theta\to 1$ and $\frac{\sin\theta}{\theta}\to 1$ as $\theta\to 0$, so the theorem follows. □

Exercises

21.1 – Find the limit or show that it does not exist. Distinguish between limits which are infinite and limits which do not exist.

(a) $\lim_{\theta\to 0}\frac{\tan\theta}{\theta}$. (b) $\lim_{x\to 0}\frac{1-\cos x}{x\sin x}$. (c) $\lim_{x\to\infty}\frac{2x^3+3x^2\cos x}{(x+2)^3}$.

(d) $\lim_{x\to 0}\frac{\sin(x^2)}{x^2}$. (e) $\lim_{x\to 0}\frac{x(1-\cos x)}{\tan^3 x}$. (f) $\lim_{x\to 0}\frac{\sin(x^2)}{1-\cos x}$.

(g) $\lim_{x\to 0}\frac{\cos x}{x^2+9}$. (h) $\lim_{x\to\pi}\frac{\sin x}{x-\pi}$. (i) $\lim_{x\to 0}\frac{\sin x}{x+\sin x}$.

21.2 – Compute

(**a**) $\lim_{x\to\infty}\frac{\sin x}{x}$ (**b**) $\lim_{x\to\infty}\frac{\cos x}{x}$.

21.3 – Find a constant k such that the function

$$f(x)=\begin{cases}3x+2 & \text{for } x<2\\ x^2+k & \text{for } x\ge 2.\end{cases}$$

is continuous. Hint: Compute the one-sided limits.

21.4 – Find constants a and b such that the function

$$f(x)=\begin{cases}x^3 & \text{for } x<-1\\ ax+b & \text{for } -1\le x<1\\ x^2+2 & \text{for } x\ge 1.\end{cases}$$

is continuous for all x.

21.5 – Is there a constant k such that the function

$$f(x)=\begin{cases}\sin(1/x) & \text{for } x\ne 0\\ k & \text{for } x=0.\end{cases}$$

is continuous? If so, find it; if not, say why.

21.6 – Compute $\lim_{x\to\infty}x\sin\frac{\pi}{x}$ and $\lim_{x\to\infty}x\tan\frac{\pi}{x}$.

21.7 – Let A_n be the area of the regular $2n$-gon inscribed in the unit circle, and let B_n be the area of the regular $2n$-gon whose inscribed circle has radius 1.

Show that

$$A_n = 2^n \sin\frac{\pi}{2^n} \text{ and } B_n = 2^{n+1} \tan\frac{\pi}{2^{n+1}}$$

Compute $\lim_{n\to\infty} A_n$ and $\lim_{n\to\infty} B_n$.

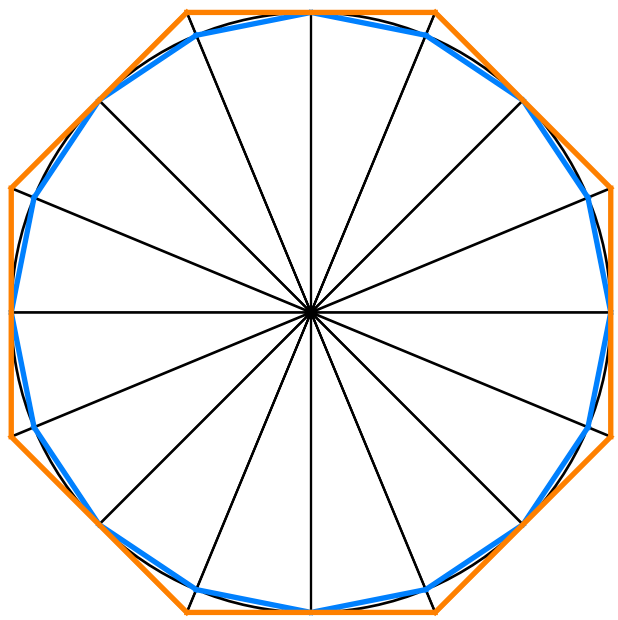


Figure 9. A_8, B_4 and π

IV. **Derivatives (2)**

"Leibniz never thought of the derivative as a limit. This does not appear until the work of d'Alembert."

http://www.gap-system.org/~history/Biographies/Leibniz.html

In chapter II we saw two mathematical problems which led to expressions of the form $\frac{0}{0}$. Now that we know how to handle limits, we can state the definition of the derivative of a function. After computing a few derivatives using the definition we will spend most of this section developing the *differential calculus,* which is a collection of rules that allow you to compute derivatives without always having to use basic definition.

22. **Derivatives Defined**

Definition. 22.1. *Let f be a function which is defined on some interval (c, d) and let a be some number in this interval.*

The ***derivative of the function*** *f* ***at*** *a is the value of the limit*

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}. \tag{15}$$

f is said to be ***differentiable at*** *a if this limit exists.*

f is called ***differentiable on the interval*** *(c, d) if it is differentiable at every point a in (c, d).*

22.1. *Other notations*

One can substitute $x = a + h$ in the limit (15) and let $h \to 0$ instead of $x \to a$. This gives the formula

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}, \tag{16}$$

Often you will find this equation written with x instead of a and Δx instead of h, which makes it look like this:

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

The interpretation is the same as in equation (8) from §10. The numerator $f(x + \Delta x) - f(x)$ represents the amount by which the function value of f changes if one increases its argument x by a (small) amount Δx. If you write $y = f(x)$ then we can call the increase in f

$$\Delta y = f(x + \Delta x) - f(x),$$

so that the derivative $f'(x)$ is

$$f'(x) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}.$$

Gottfried Wilhelm von Leibniz, one of the inventors of calculus, came up with the idea that one should write this limit as

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x},$$

the idea being that after letting Δx go to zero it didn't vanish, but instead became an infinitely small quantity which Leibniz called "dx." The result of increasing x by this infinitely small quantity dx is that $y = f(x)$ increased by another infinitely small quantity

dy. The ratio of these two infinitely small quantities is what we call the derivative of $y = f(x)$.

There are no "infinitely small real numbers," and this makes Leibniz' notation difficult to justify. In the 20th century mathematicians have managed to create a consistent theory of "infinitesimals" which allows you to compute with "dx and dy" as Leibniz and his contemporaries would have done. This theory is called "non standard analysis." We won't mention it any further[3]. Nonetheless, even though we won't use infinitely small numbers, Leibniz' notation is very useful and we will use it.

23. Direct computation of derivatives

23.1. *Example – The derivative of $f(x) = x^2$ is $f'(x) = 2x$*

We have done this computation before in §8. The result was

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

Leibniz would have written

$$\frac{dx^2}{dx} = 2x.$$

23.2. *The derivative of $g(x) = x$ is $g'(x) = 1$*

Indeed, one has

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

In Leibniz' notation:

$$\frac{dx}{dx} = 1.$$

This is an example where Leibniz' notation is most misleading, because if you divide dx by dx then you should of course get 1. Nonetheless, this is not what is going on. The expression $\frac{dx}{dx}$ is not really a fraction since there are no two "infinitely small" quantities dx which we are dividing.

23.3. *The derivative of any constant function is zero*

Let $k(x) = c$ be a constant function. Then we have

$$k'(x) = \lim_{h\to 0} \frac{k(x+h) - k(x)}{h} = \lim_{h\to 0} \frac{c - c}{h} = \lim_{h\to 0} 0 = 0.$$

Leibniz would have said that if c is a constant, then

$$\frac{dc}{dx} = 0.$$

---

[3] But if you want to read more on this you should see Keisler's calculus text at
`http://www.math.wisc.edu/~keisler/calc.html`
I would not recommend using Keisler's text and this text at the same time, but if you like math you should remember that it exists, and look at it (later, say, after you pass 221.)

23.4. *Derivative of* x^n *for* $n = 1, 2, 3, \ldots$

To differentiate $f(x) = x^n$ one proceeds as follows:

$$f'(a) = \lim_{x\to a} \frac{f(x) - f(a)}{x - a} = \lim_{x\to a} \frac{x^n - a^n}{x - a}.$$

We need to simplify the fraction $(x^n - a^n)/(x - a)$. For $n = 2$ we have

$$\frac{x^2 - a^2}{x - a} = x + a.$$

For $n = 1, 2, 3, \ldots$ the geometric sum formula tells us that

$$\frac{x^n - a^n}{x - a} = x^{n-1} + x^{n-2}a + x^{n-3}a^2 + \cdots + xa^{n-2} + a^{n-1}. \tag{17}$$

If you don't remember the geometric sum formula, then you could also just verify (17) by carefully multiplying both sides with $x - a$. For instance, when $n = 3$ you would get

$$\begin{array}{rclllll} x \times (x^2 + xa + a^2) & = & x^3 & +ax^2 & +a^2x & \\ -a \times (x^2 + xa + a^2) & = & & -ax^2 & -a^2x & -a^3 \\ \hline (x - a) \times (x^2 + ax + a^2) & = & x^3 & & & -a^3 \end{array}$$

With formula (17) in hand we can now easily find the derivative of x^n:

$$\begin{aligned} f'(a) &= \lim_{x\to a} \frac{x^n - a^n}{x - a} \\ &= \lim_{x\to a} \left\{x^{n-1} + x^{n-2}a + x^{n-3}a^2 + \cdots + xa^{n-2} + a^{n-1}\right\} \\ &= a^{n-1} + a^{n-2}a + a^{n-3}a^2 + \cdots + a\,a^{n-2} + a^{n-1}. \end{aligned}$$

Here there are n terms, and they all are equal to a^{n-1}, so the final result is

$$f'(a) = na^{n-1}.$$

One could also write this as $f'(x) = nx^{n-1}$, or, in Leibniz' notation

$$\frac{dx^n}{dx} = nx^{n-1}.$$

This formula turns out to be true in general, but here we have only proved it for the case in which n is a positive integer.

23.5. *Differentiable implies Continuous*

Theorem 23.1. *If a function f is differentiable at some a in its domain, then f is also continuous at a.*

Proof. We are given that

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

exists, and we must show that

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

This follows from the following computation

$$\begin{aligned}
\lim_{x\to a} f(x) &= \lim_{x\to a}\big(f(x) - f(a) + f(a)\big) && \text{(algebra)}\\
&= \lim_{x\to a} \frac{f(x)-f(a)}{x-a}\cdot (x-a) + f(a) && \text{(more algebra)}\\
&= \left\{\lim_{x\to a} \frac{f(x)-f(a)}{x-a}\right\}\cdot \lim_{x\to a}(x-a) + \lim_{x\to a} f(a) && \text{(Limit Properties)}\\
&= f'(a)\cdot 0 + f(a) && \text{($f'(a)$ exists)}\\
&= f(a).
\end{aligned}$$

□

23.6. *Some non-differentiable functions*

23.6.1. *A graph with a corner.* Consider the function

$$f(x) = |x| = \begin{cases} x & \text{for } x \geq 0,\\ -x & \text{for } x < 0.\end{cases}$$

This function is continuous at all x, but it is not differentiable at $x = 0$.

To see this try to compute the derivative at 0,

$$f'(0) = \lim_{x\to 0}\frac{|x| - |0|}{x - 0} = \lim_{x\to 0}\frac{|x|}{x} = \lim_{x\to 0}\operatorname{sign}(x).$$

We know this limit does not exist (see §16.1)

If you look at the graph of $f(x) = |x|$ then you see what is wrong: the graph has a corner at the origin and it is not clear which line, if any, deserves to be called the tangent to the graph at the origin.

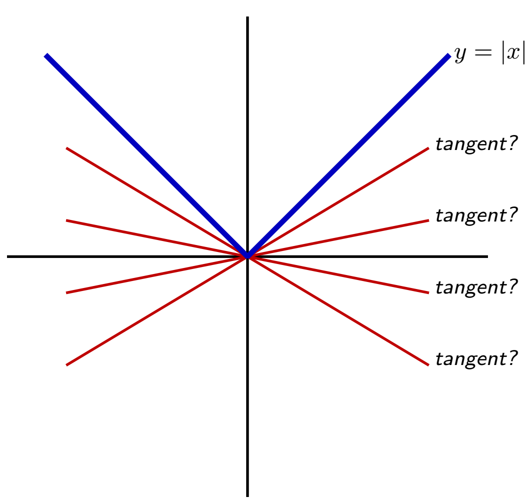


Figure 10. The graph of $y = |x|$ has no tangent at the origin.

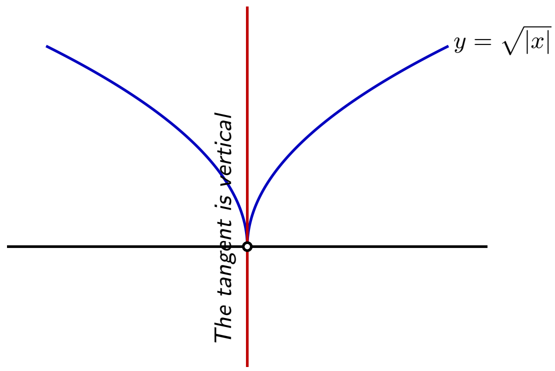


Figure 11. Tangent to the graph of $y = |x|^{1/2}$ at the origin

23.6.2. *A graph with a cusp.* Another example of a function without a derivative at $x = 0$ is

$$f(x) = \sqrt{|x|}.$$

When you try to compute the derivative you get this limit

$$f'(0) = \lim_{x\to 0} \frac{\sqrt{|x|}}{x} = ?$$

The limit from the right is

$$\lim_{x\searrow 0} \frac{\sqrt{|x|}}{x} = \lim_{x\searrow 0} \frac{1}{\sqrt{x}},$$

which does not exist (it is "$+\infty$"). Likewise, the limit from the left also does not exist ('tis "$-\infty$). Nonetheless, a drawing for the graph of f suggests an obvious tangent to the graph at $x = 0$, namely, the y-axis. That observation does not give us a derivative, because the y-axis is vertical and hence has no slope.

23.6.3. *A graph with absolutely no tangents anywhere.* The previous two examples were about functions which did not have a derivative at $x = 0$. In both examples the point $x = 0$ was the only point where the function failed to have a derivative. It is easy to give examples of functions which are not differentiable at more than one value of x, but here I would like to show you a function f which doesn't have a derivative ***anywhere in its domain.***

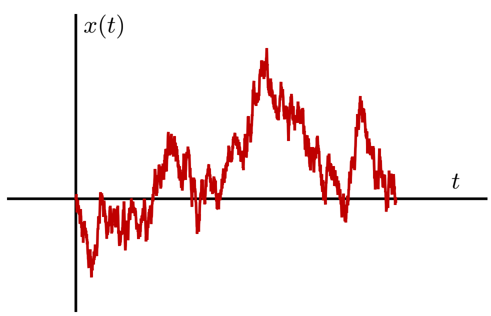


Figure 12. A one-dimensional Brownian motion

To keep things short I won't write a formula for the function, and merely show you a graph (see Figure 12.) In this graph you see a typical path of a Brownian motion, i.e. t is time, and $x(t)$ is the position of a particle which undergoes a Brownian motion – come to lecture for further explanation (see also the article on wikipedia).

To see a similar graph check the Dow Jones or Nasdaq in the upper left hand corner of the web page at `http://finance.yahoo.com` in the afternoon on any weekday.

Exercises

23.1 – Compute the derivative of the following functions

$$f(x) = x^2 - 2x \qquad g(x) = \frac{1}{x} \qquad k(x) = x^3 - 17x$$
$$u(x) = \frac{2}{1+x} \qquad v(x) = \sqrt{x} \qquad w(x) = \frac{1}{\sqrt{x}}$$

using either (15) or (16).

23.2 – Which of the following functions is differentiable at $x = 0$?

$$f(x) = x|x|, \qquad g(x) = x\sqrt{|x|}, \qquad h(x) = x + |x|.$$

23.3 – For which value(s) is the function defined by

$$f(x) = \begin{cases} ax + b & \text{for } x < 0 \\ x - x^2 & \text{for } x \geq 0 \end{cases}$$

differentiable at $x = 0$?

23.4 – For which value(s) is the function defined by

$$f(x) = \begin{cases} ax^2 & \text{for } x < 2 \\ x + b & \text{for } x \geq 2 \end{cases}$$

differentiable at $x = 0$?

23.5 – *True or false:* If a function f is continuous at some $x = a$ then it must also be differentiable at $x = a$?

23.6 – *True or false:* If a function f is differentiable at some $x = a$ then it must also be continuous at $x = a$?

24. The Differentiation Rules

You could go on and compute more derivatives from the definition. Each time you would have to compute a new limit, and hope that there is some trick that allows you to find that limit. This is fortunately not necessary. It turns out that if you know a few basic derivatives (such as $dx^n/dx = nx^{n-1}$) the you can find derivatives of arbitrarily complicated functions by breaking them into smaller pieces. In this section we'll look at rules which tell you how to differentiate a function which is either the sum, difference, product or quotient of two other functions.

The situation is analogous to that of the "limit-properties" $(P_1)\ldots(P_6)$ from the previous chapter which allowed us to compute limits without always having to go back to the epsilon-delta definition.

24.1. *Sum, product and quotient rules*

In the following c and n are constants, u and v are functions of x, and $'$ denotes differentiation. The Differentiation Rules in function notation, and Leibniz notation, are

Constant rule:	$c' = 0$	$\dfrac{dc}{dx} = 0$
Sum rule:	$(u \pm v)' = u' \pm v'$	$\dfrac{du \pm v}{dx} = \dfrac{du}{dx} \pm \dfrac{dv}{dx}$
Product rule:	$(u \cdot v)' = u' \cdot v + u \cdot v'$	$\dfrac{duv}{dx} = \dfrac{du}{dx} v + u \dfrac{dv}{dx}$
Quotient rule:	$\left(\dfrac{u}{v}\right)' = \dfrac{u' \cdot v - u \cdot v'}{v^2}$	$\dfrac{d\frac{u}{v}}{dx} = \dfrac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}$

Note that we already proved the Constant Rule in example 23.2. We will now prove the sum, product and quotient rules.

24.2. *Proof of the Sum Rule*

Suppose that $f(x) = u(x) + v(x)$ for all x where u and v are differentiable. Then

$$\begin{aligned}
f'(a) &= \lim_{x\to a} \frac{f(x) - f(a)}{x - a} && \text{(definition of } f'\text{)} \\
&= \lim_{x\to a} \frac{\big(u(x) + v(x)\big) - \big(u(a) + v(a)\big)}{x - a} && \text{(use } f = u + v\text{)} \\
&= \lim_{x\to a} \left(\frac{u(x) - u(a)}{x - a} + \frac{v(x) - v(a)}{x - a}\right) && \text{(algebra)} \\
&= \lim_{x\to a} \frac{u(x) - u(a)}{x - a} + \lim_{x\to a} \frac{v(x) - v(a)}{x - a} && \text{(limit property)} \\
&= u'(a) + v'(a) && \text{(definition of } u',\ v'\text{)}
\end{aligned}$$

24.3. *Proof of the Product Rule*

Let $f(x) = u(x)v(x)$. To find the derivative we must express the change of f in terms of the changes of u and v

$$\begin{aligned}
f(x) - f(a) &= u(x)v(x) - u(a)v(a) \\
&= u(x)v(x) - u(x)v(a) + u(x)v(a) - u(a)v(a) \\
&= u(x)\big(v(x) - v(a)\big) + \big(u(x) - u(a)\big)v(a)
\end{aligned}$$

Now divide by $x - a$ and let $x \to a$:

$$\begin{aligned}
\lim_{x\to a} \frac{f(x) - f(a)}{x - a} &= \lim_{x\to a} u(x)\frac{v(x) - v(a)}{x - a} + \frac{u(x) - u(a)}{x - a}v(a) \\
&\qquad \text{(use the limit properties)} \\
&= \Big(\lim_{x\to a} u(x)\Big)\Big(\lim_{x\to a} \frac{v(x) - v(a)}{x - a}\Big) + \Big(\lim_{x\to a} \frac{u(x) - u(a)}{x - a}\Big)v(a) \\
&= u(a)v'(a) + u'(a)v(a),
\end{aligned}$$

as claimed. In this last step we have used that

$$\lim_{x\to a} \frac{u(x) - u(a)}{x - a} = u'(a) \quad \text{and} \quad \lim_{x\to a} \frac{v(x) - v(a)}{x - a} = v'(a)$$

and also that

$$\lim_{x\to a} u(x) = u(a)$$

This last limit follows from the fact that u is continuous, which in turn follows from the fact that u is differentiable.

24.4. *Proof of the Quotient Rule*

We can break the proof into two parts. First we do the special case where $f(x) = 1/v(x)$, and then we use the product rule to differentiate

$$f(x) = \frac{u(x)}{v(x)} = u(x) \cdot \frac{1}{v(x)}.$$

So let $f(x) = 1/v(x)$. We can express the change in f in terms of the change in v

$$f(x) - f(a) = \frac{1}{v(x)} - \frac{1}{v(a)} = \frac{v(x) - v(a)}{v(x)v(a)}.$$

Dividing by $x - a$ we get

$$\frac{f(x) - f(a)}{x - a} = \frac{1}{v(x)v(a)} \frac{v(x) - v(a)}{x - a}.$$

Now we want to take the limit $x \to a$. We are given the v is differentiable, so it must also be continuous and hence

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

Therefore we find

$$\lim_{x\to a} \frac{f(x) - f(a)}{x - a} = \lim_{x\to a} \frac{1}{v(x)v(a)} \lim_{x\to a} \frac{v(x) - v(a)}{x - a} = \frac{v'(a)}{v(a)^2}.$$

That completes the first step of the proof. In the second step we use the product rule to differentiate $f = u/v$

$$f' = \left(\frac{u}{v}\right)' = \left(u \cdot \frac{1}{v}\right)' = u' \cdot \frac{1}{v} + u \cdot \left(\frac{1}{v}\right)' = \frac{u'}{v} - u\frac{v'}{v^2} = \frac{u'v - uv'}{v^2}.$$

24.5. *A shorter, but not quite perfect derivation of the Quotient Rule*

The Quotient Rule can be derived from the Product Rule as follows: if $w = u/v$ then

$$w \cdot v = u \tag{18}$$

By the product rule we have

$$w' \cdot v + w \cdot v' = u',$$

so that

$$w' = \frac{u' - w \cdot v'}{v} = \frac{u' - (u/v) \cdot v'}{v} = \frac{u' \cdot v - u \cdot v'}{v^2}.$$

Unlike the proof in §24.4 above, this argument does not prove that w is differentiable if u and v are. It only says that ***if the derivative exists*** then it must be what the Quotient Rule says it is.

The trick which is used here, is a special case of a method called "implicit differentiation." We have an equation (18) which the quotient w satisfies, and from by differentiating this equation we find w'.

24.6. *Differentiating a constant multiple of a function*

Note that the rule

$$(cu)' = cu'$$

follows from the Constant Rule and the Product Rule.

24.7. *Picture of the Product Rule*

If u and v are quantities which depend on x, and if increasing x by Δx causes u and v to change by Δu and Δv, then the product of u and v will change by

$$\Delta(uv) = (u + \Delta u)(v + \Delta v) - uv = u\Delta v + v\Delta u + \Delta u\Delta v. \tag{19}$$

If u and v are differentiable functions of x, then the changes Δu and Δv will be of the same order of magnitude as Δx, and thus one expects $\Delta u\Delta v$ to be much smaller. One therefore ignores the last term in (19), and thus arrives at

$$\Delta(uv) = u\Delta v + v\Delta u.$$

Leibniz would now divide by Δx and replace Δ's by d's to get the product rule:

$$\frac{\Delta(uv)}{\Delta x} = u\frac{\Delta v}{\Delta x} + v\frac{\Delta u}{\Delta x}.$$

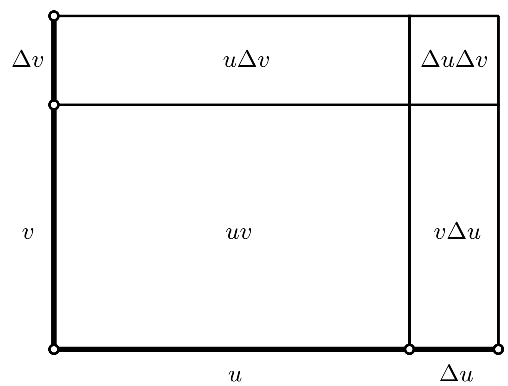


Figure 13. The Product Rule: how much does the area of a rectangle change if its sides u and v are increased by Δu and Δv?

25. **Differentiating powers of functions**

25.1. *Product rule with more than one factor*

If a function is given as the product of n functions, i.e.

$$f(x) = u_1(x) \times u_2(x) \times \cdots \times u_n(x),$$

then you can differentiate it by applying the product rule $n-1$ times (there are n factors, so there are $n-1$ multiplications.)

After the first step you would get

$$f' = u_1'\bigl(u_2 \cdots u_n\bigr) + u_1\bigl(u_2 \cdots u_n\bigr)'.$$

In the second step you apply the product rule to $(u_2u_3 \cdots u_n)'$. This yields

$$\begin{aligned} f' &= u_1'u_2 \cdots u_n + u_1\bigl[u_2'u_3 \cdots u_n + u_2(u_3 \cdots u_n)'\bigr] \\ &= u_1'u_2 \cdots u_n + u_1u_2'u_3 \cdots u_n + u_1u_2\bigl(u_3 \cdots u_n\bigr)'. \end{aligned}$$

Continuing this way one finds after $n-1$ applications of the product rule that

$$\bigl(u_1 \cdots u_n\bigr)' = u_1'u_2 \cdots u_n + u_1u_2'u_3 \cdots u_n + \cdots + u_1u_2u_3 \cdots u_n'. \tag{20}$$

25.2. *The Power rule*

If all n factors in the previous paragraph are the same, so that the function f is the n^{th} power of some other function,

$$f(x) = \bigl(u(x)\bigr)^n,$$

then all terms in the right hand side of (20) are the same, and, since there are n of them, one gets

$$f'(x) = nu^{n-1}(x)u'(x),$$

or, in Leibniz' notation,

$$\frac{du^n}{dx} = nu^{n-1}\frac{du}{dx}. \tag{21}$$

25.3. *The Power Rule for Negative Integer Exponents*

We have just proved the power rule (21) assuming n is a positive integer. The rule actually holds for all real exponents n, but the proof is harder.

Here we prove the Power Rule for negative exponents using the Quotient Rule. Suppose $n = -m$ where m is a positive integer. Then the Quotient Rule tells us that

$$(u^n)' = \bigl(u^{-m}\bigr)' = \left(\frac{1}{u^m}\right)' \stackrel{\text{Q.R.}}{=} -\frac{(u^m)'}{(u^m)^2}.$$

Since m is a positive integer, we can use (21), so $(u^m)' = mu^{m-1}$, and hence

$$(u^n)' = -\frac{mu^{m-1}\cdot u'}{u^{2m}} = -mu^{-m-1}\cdot u' = nu^{n-1}u'.$$

25.4. *The Power Rule for Rational Exponents*

So far we have proved that the power law holds if the exponent n is an integer.

We will now see how you can show that the power law holds even if the exponent n is any fraction, $n = p/q$. The following derivation contains the trick called ***implicit differentiation*** which we will study in more detail in chapter **??**.

So let $n = p/q$ where p and q are integers and consider the function

$$w(x) = u(x)^{p/q}.$$

Assuming that both u and w are differentiable functions, we will show that

$$w'(x) = \frac{p}{q}u(x)^{\frac{p}{q}-1}u'(x) \tag{22}$$

Raising both sides to the qth power gives

$$w(x)^q = u(x)^p.$$

Here the exponents p and q are integers, so we may apply the Power Rule to both sides. We get

$$qw^{q-1}\cdot w' = pu^{p-1}\cdot u'.$$

Dividing both sides by qw^{q-1} and substituting $u^{p/q}$ for w gives

$$w' = \frac{pu^{p-1}\cdot u'}{qw^{q-1}} = \frac{pu^{p-1}\cdot u'}{qu^{p(q-1)/q}} = \frac{pu^{p-1}\cdot u'}{qu^{p-(p/q)}} = \frac{p}{q}\cdot u^{(p/q)-1}\cdot u'$$

which is the Power Rule for $n = p/q$.

This proof is flawed because we did not show that $w(x) = u(x)^{p/q}$ is differentiable: we only showed what the derivative should be, *if it exists.*

25.5. *Derivative of x^n for integer n*

If you choose the function $u(x)$ in the Power Rule to be $u(x) = x$, then $u'(x) = 1$, and hence the derivative of $f(x) = u(x)^n = x^n$ is

$$f'(x) = nu(x)^{n-1}u'(x) = nx^{n-1} \cdot 1 = nx^{n-1}.$$

We already knew this of course.

25.6. *Example – differentiate a polynomial*

Using the Differentiation Rules you can easily differentiate any polynomial and hence any rational function. For example, using the Sum Rule, the Power Rule with $u(x) = x$, the rule $(cu)' = cu'$, the derivative of the polynomial

$$f(x) = 2x^4 - x^3 + 7$$

is

$$f'(x) = 8x^3 - 3x^2.$$

25.7. *Example – differentiate a rational function*

By the Quotient Rule the derivative of the function

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

is

$$\begin{aligned} g'(x) &= \frac{(8x^3 - 3x^2)(1 + x^2) - (2x^4 - x^3 + 7)2x}{(1 + x^2)^2} \\ &= \frac{6x^5 - x^4 + 8x^3 - 3x^2 - 14x}{(1 + x^2)^2}. \end{aligned}$$

If you compare this example with the previous then you see that polynomials simplify when you differentiate them while rational functions become more complicated.

25.8. *Derivative of the square root*

The derivative of $f(x) = \sqrt{x} = x^{1/2}$ is

$$f'(x) = \frac{1}{2}x^{1/2-1} = \frac{1}{2}x^{-1/2} = \frac{1}{2x^{1/2}} = \frac{1}{2\sqrt{x}}$$

where we used the power rule with $n = 1/2$ and $u(x) = x$.

Exercises

25.1 – Let $f(x) = (x^2 + 1)(x^3 + 3)$. Find $f'(x)$ in two ways:

(**a**) by multiplying and then differentiating,

(**b**) by using the product rule.

Are your answers the same?

25.2 – Let $f(x) = (1+x^2)^4$. Find $f'(x)$ in two ways, first by expanding to get an expression for $f(x)$ as a polynomial in x and then differentiating, and then by using the power rule. Are the answers the same?

25.3 – Prove the statement in §24.6, i.e. show that $(cu)' = c(u')$ follows from the product rule.

25.4 – Compute the derivatives of the following functions (try to simplify your answers)

(**a**) $f(x) = x + 1 + (x+1)^2$ (**b**) $f(x) = \dfrac{x-2}{x^4+1}$ (**c**) $f(x) = \left(\dfrac{1}{1+x}\right)^{-1}$

(**d**) $f(x) = \sqrt{1-x^2}$ (**e**) $f(x) = \dfrac{ax+b}{cx+d}$ (**f**) $f(x) = \dfrac{1}{(1+x^2)^2}$

(**g**) $f(x) = \dfrac{x}{1+\sqrt{x}}$ (**h**) $f(x) = \sqrt{\dfrac{1-x}{1+x}}$ (**i**) $f(x) = \sqrt[3]{x+\sqrt{x}}$

(**j**) $\varphi(t) = \dfrac{t}{1+\sqrt{t}}$ (**k**) $g(s) = \sqrt{\dfrac{1-s}{1+s}}$ (**l**) $h(\rho) = \sqrt[3]{\rho+\sqrt{\rho}}$

25.5 *Using derivatives to approximate numbers.*

(**a**) Find the derivative of $f(x) = x^{4/3}$.

(**b**) Use **(i)** to estimate the number

$$\frac{127^{4/3} - 125^{4/3}}{2}$$

approximately without a calculator. Your answer should have the form p/q where p and q are integers. [*Hint: Note that* $5^3 = 125$ *and take a good look at equation* (15).]

(**c**) Approximate in the same way the numbers $\sqrt{143}$ and $\sqrt{145}$ (Hint: $12 \times 12 = 144$).

25.6 *Making the product and quotient rules look nicer.* Instead of looking at the derivative of a function you can look at the ratio of its derivative to the function itself, i.e. you can compute f'/f. This quantity is called the ***logarithmic derivative of the function*** f for reasons that will become clear later this semester.

(**a**) Compute the logarithmic derivative of these functions (i.e. find $f'(x)/f(x)$)

$f(x) = x$ $g(x) = 3x$ $h(x) = x^2$

$k(x) = -x^2$ $\ell(x) = 2007x^2$ $m(x) = x^{2007}$

(**b**) Show that for any pair of functions u and v one has

$$\frac{(uv)'}{uv} = \frac{u'}{u} + \frac{v'}{v}$$

$$\frac{(u/v)'}{u/v} = \frac{u'}{u} - \frac{v'}{v}$$

$$\frac{\left(u^n\right)'}{u^n} = n\,\frac{u'}{u}$$

25.7 – (**a**) Find $f'(x)$ and $g'(x)$ if

$$f(x) = \frac{1+x^2}{2x^4+7}, \qquad g(x) = \frac{2x^4+7}{1+x^2}.$$

Note that $f(x) = 1/g(x)$.

(**b**) Is it true that $f'(x) = 1/g'(x)$?

(**c**) Is it true that $f(x) = g^{-1}(x)$?

(**d**) Is it true that $f(x) = g(x)^{-1}$?

25.8 – (**a**) Let $x(t) = (1-t^2)/(1+t^2)$, $y(t) = 2t/(1+t^2)$ and $u(t) = y(t)/x(t)$. Find dx/dt, dy/dt.

(**b**) Now that you've done **(i)** there are two different ways of finding du/dt. What are they, and use one of both to find du/dt.

26. Higher Derivatives

26.1. *The derivative is a function*

If the derivative $f'(a)$ of some function f exists for all a in the domain of f, then we have a new function: namely, for each number in the domain of f we compute the derivative of f at that number. This function is called the ***derivative function*** of f, and it is denoted by f'. Now that we have agreed that the derivative of a function is a function, we can repeat the process and try to differentiate the derivative. The result, if it exists, is called the ***second derivative of*** f. It is denoted f''. The derivative of the second derivative is called the third derivative, written f''', and so on.

The nth derivative of f is denoted $f^{(n)}$. Thus

$$f^{(0)} = f, \qquad f^{(1)} = f', \qquad f^{(2)} = f'', \qquad f^{(3)} = f''', \ldots.$$

Leibniz' notation for the nth derivative of $y = f(x)$ is

$$\frac{d^n y}{dx^n} = f^{(n)}(x).$$

26.2. *Operator notation*

A common variation on Leibniz' notation for derivatives is the so-called ***operator notation***, as in

$$\frac{d(x^3 - x)}{dx} = \frac{d}{dx}(x^3 - x) = 3x^2 - 1.$$

For higher derivatives one can write

$$\frac{d^2 y}{dx^2} = \left(\frac{d}{dx}\right)^2 y$$

Be careful to distinguish the second derivative from the square of the first derivative. Usually

$$\frac{d^2 y}{dx^2} \neq \left(\frac{dy}{dx}\right)^2 \textbf{!!!!}$$

Exercises

<u>*26.1*</u> – The equation

$$\frac{2x}{x^2 - 1} = \frac{1}{x+1} + \frac{1}{x-1} \tag{†}$$

holds for all values of x (except $x = \pm 1$), so you should get the same answer if you differentiate both sides. Check this.

Compute the third derivative of $f(x) = 2x/(x^2 - 1)$ by using either the left or right hand side (your choice) of (†).

<u>*26.2*</u> – Compute the first, second and third derivatives of the following functions

$$f(x) = (x+1)^4 \qquad\qquad g(x) = \bigl(x^2 + 1\bigr)^4$$

$$h(x) = \sqrt{x - 2} \qquad\qquad k(x) = \sqrt[3]{x - \frac{1}{x}}$$

<u>*26.3*</u> – Find the derivatives of 10^{th} order of the functions

$$f(x) = x^{12} + x^8 \qquad g(x) = \frac{1}{x}$$

$$h(x) = \frac{12}{1-x} \qquad k(x) = \frac{x^2}{1-x}$$

<u>*26.4*</u> – Find $f'(x)$, $f''(x)$ and $f^{(3)}(x)$ if

$$f(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \frac{x^5}{120} + \frac{x^6}{720}.$$

<u>*26.5*</u> – (**a**) Find the 12^{th} derivative of the function $f(x) = \dfrac{1}{x+2}$.

(**b**) Find the n^{th} order derivative of $f(x) = \dfrac{1}{x+2}$(i.e. find a formula for $f^{(n)}(x)$ which is valid for all $n = 0, 1, 2, 3\ldots$).

(**c**) Find the n^{th} order derivative of $g(x) = \dfrac{x}{x+2}$.

<u>*26.6*</u> *About notation*

(**a**) Find dy/dx and d^2y/dx^2 if $y = x/(x+2)$. Hint: See previous problem.

(**b**) Find du/dt and d^2u/dt^2 if $u = t/(t+2)$. Hint: See previous problem.

(**c**) Find $\dfrac{d}{dx}\left(\dfrac{x}{x+2}\right)$ and $\dfrac{d^2}{dx^2}\left(\dfrac{x}{x+2}\right)$. Hint: See previous problem.

(**d**) Find $\dfrac{d}{dx}\left(\dfrac{x}{x+2}\right)\Big|_{x=1}$ and $\dfrac{d}{dx}\left(\dfrac{1}{1+2}\right)$.

<u>*26.7*</u> – Find d^2y/dx^2 and $(dy/dx)^2$ if $y = x^3$.

27. Differentiating Trigonometric functions

The trigonometric functions Sine, Cosine and Tangent are differentiable, and their derivatives are given by the following formulas

$$\frac{d\sin x}{dx} = \cos x, \quad \frac{d\cos x}{dx} = -\sin x, \quad \frac{d\tan x}{dx} = \frac{1}{\cos^2 x}. \tag{23}$$

Note the minus sign in the derivative of the cosine!

Proof. By definition one has

$$\sin'(x) = \lim_{h\to 0} \frac{\sin(x+h) - \sin(x)}{h}.$$

To simplify the numerator we use the trigonometric addition formula

$$\sin(\alpha + \beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta.$$

with $\alpha = x$ and $\beta = h$, which results in

$$\begin{aligned}\frac{\sin(x+h) - \sin(x)}{h} &= \frac{\sin(x)\cos(h) + \cos(x)\sin(h) - \sin(x)}{h} \\ &= \cos(x)\frac{\sin(h)}{h} + \sin(x)\frac{\cos(h) - 1}{h}\end{aligned}$$

Hence by the formulas

$$\lim_{h\to 0} \frac{\sin(h)}{h} = 1 \qquad \text{and} \qquad \lim_{h\to 0} \frac{\cos(h) - 1}{h} = 0$$

from Section 21 we have

$$\begin{aligned}\sin'(x) &= \lim_{h\to 0} \cos(x)\frac{\sin(h)}{h} + \sin(x)\frac{\cos(h)-1}{h} \\ &= \cos(x)\cdot 1 + \sin(x)\cdot 0 \\ &= \cos(x).\end{aligned}$$

A similar computation leads to the stated derivative of $\cos x$.

To find the derivative of $\tan x$ we apply the quotient rule to

$$\tan x = \frac{\sin x}{\cos x} = \frac{f(x)}{g(x)}.$$

We get

$$\tan'(x) = \frac{\cos(x)\sin'(x) - \sin(x)\cos'(x)}{\cos^2(x)} = \frac{\cos^2(x)+\sin^2(x)}{\cos^2(x)} = \frac{1}{\cos^2(x)}$$

as claimed. □

Exercises

27.1 – Find the derivatives of the following functions (try to simplify your answers)

(**a**) $f(x) = \sin(x) + \cos(x)$

(**b**) $f(x) = 2\sin(x) - 3\cos(x)$

(**c**) $f(x) = 3\sin(x) + 2\cos(x)$

(**d**) $f(x) = x\sin(x) + \cos(x)$

(**e**) $f(x) = x\cos(x) - \sin x$

(**f**) $f(x) = \dfrac{\sin x}{x}$

(**g**) $f(x) = \cos^2(x)$

(**h**) $f(x) = \sqrt{1-\sin^2 x}$

(**i**) $f(x) = \sqrt{\dfrac{1-\sin x}{1+\sin x}}$

(**j**) $\cot(x) = \dfrac{\cos x}{\sin x}$.

27.2 – Can you find a and b so that the function

$$f(x) = \begin{cases}\cos x & \text{for } x \le \frac{\pi}{4} \\ a + bx & \text{for } x > \frac{\pi}{4}\end{cases}$$

is differentiable at $x = \pi/4$?

27.3 – Can you find a and b so that the function

$$f(x) = \begin{cases}\tan x & \text{for } x < \frac{\pi}{6} \\ a + bx & \text{for } x \ge \frac{\pi}{6}\end{cases}$$

is differentiable at $x = \pi/4$?

27.4 – If f is a given function, and you have another function g which satisfies $g(x) = f(x) + 12$ for all x, then f and g have the same derivatives. Prove this. [*Hint: it's a short proof – use the differentiation rules.*]

27.5 – Show that the functions

$$f(x) = \sin^2 x \text{ and } g(x) = -\cos^2 x$$

have the same derivative by computing $f'(x)$ and $g'(x)$.

With hindsight this was to be expected – why?

28. The Chain Rule

28.1. *Composition of functions*

Given two functions f and g, one can define a new function called the ***composition of*** f ***and*** g. The notation for the composition is $f \circ g$, and it is defined by the formula

$$f \circ g(x) = f\big(g(x)\big).$$

The domain of the composition is the set of all numbers x for which this formula gives you something well-defined.

For instance, if $f(x) = x^2 + x$ and $g(x) = 2x + 1$ then

$$f \circ g(x) = f\big(2x+1\big) = (2x+1)^2 + (2x+1)$$
$$\text{and } g \circ f(x) = g\big(x^2+x\big) = 2(x^2+x) + 1$$

Note that $f \circ g$ and $g \circ f$ are not the same fucntion in this example (they hardly ever are the same).

If you think of functions as expressing dependence of one quantity on another, then the composition of functions arises as follows. If a quantity z is a function of another quantity y, and if y itself depends on x, then z depends on x via y.

To get $f \circ g$ from the previous example, we could say $z = f(y)$ and $y = g(x)$, so that

$$z = f(y) = y^2 + y \text{ and } y = 2x + 1.$$

Give x one can compute y, and from y one can then compute z. The result will be

$$z = y^2 + y = (2x+1)^2 + (2x+1),$$

in other notation,

$$z = f(y) = f\big(g(x)\big) = f \circ g(x).$$

One says that ***the composition of*** f ***and*** g ***is the result of subsituting*** g ***in*** f.

28.2. *A real world example*

A biologist is studying growth of yeast-cells. Assuming that a yeast-cell is spherical one can say how large it is by specifying its radius R. For a growing cell this radius will change with time t. The volume of the cell is a function of its radius, since the volume of a sphere of radius r is given by

$$V = \frac{4}{3}\pi r^3.$$

We now have two functions, the first f turns tells you the radius R of the cell at time t,

$$R = f(t)$$

and the second tells you the volume of the cell given its radius

$$V = g(r).$$

The volume of the cell at time t is then given by

$$V = g\big(f(t)\big) = g \circ f(t),$$

i.e. the function which tells you the volume of the cell at time t is the composition of first f and then g.

28.3. *Statement of the Chain Rule*

The chain rule tells you how to find the derivative of the composition $f \circ g$ of two functions f and g provided you now how to differentiate the two functions f and g.

Theorem 28.1 (*Chain Rule*). *If f and g are differentiable, so is the composition $f \circ g$*

The derivative of $f \circ g$ is given by

$$(f \circ g)'(x) = f'(g(x))\, g'(x).$$

When written in Leibniz' notation the chain rule looks particularly easy. Suppose that $y = g(x)$ and $z = f(y)$, then $z = f \circ g(x)$, and the derivative of z with respect to x is the derivative of the function $f \circ g$. The derivative of z with respect to y is the derivative of the function f, and the derivative of y with respect to x is the derivative of the function g. In short,

$$\frac{dz}{dx} = (f \circ g)'(x), \quad \frac{dz}{dy} = f'(y) \quad \text{and} \quad \frac{dy}{dx} = g'(x)$$

so that the chain rule says

$$\frac{dz}{dx} = \frac{dz}{dy}\,\frac{dy}{dx}. \tag{24}$$

First proof of the chain rule (using Leibniz' notation). We first consider difference quotients instead of derivatives, i.e. using the same notation as above, we consider the effect of an increase of x by an amount Δx on the quantity z.

If x increases by Δx, then $y = g(x)$ will increase by

$$\Delta y = g(x + \Delta x) - g(x),$$

and $z = f(y)$ will increase by

$$\Delta z = f(y + \Delta y) - f(y).$$

The ratio of the increase in $z = f(g(x))$ to the increase in x is

$$\frac{\Delta z}{\Delta x} = \frac{\Delta z}{\Delta y} \cdot \frac{\Delta y}{\Delta x}.$$

In contrast to dx, dy and dz in equation (24), the Δx, etc. here are finite quantities, so this equation is just algebra: you can cancel the two Δys. If you let the increase Δx go to zero, then the increase Δy will also go to zero, and the difference quotients converge to the derivatives,

$$\frac{\Delta z}{\Delta x} \longrightarrow \frac{dz}{dx}, \quad \frac{\Delta z}{\Delta y} \longrightarrow \frac{dz}{dy}, \quad \frac{\Delta y}{\Delta x} \longrightarrow \frac{dy}{dx}$$

which immediately leads to Leibniz' form of the quotient rule. □

Proof of the chain rule. We verify the formula in Theorem 28.1 at some arbitrary value $x = a$, i.e. we will show that

$$(f \circ g)'(a) = f'(g(a))\, g'(a).$$

By definition the left hand side is

$$(f \circ g)'(a) = \lim_{x \to a} \frac{(f \circ g)(x) - (f \circ g)(a)}{x - a} = \lim_{x \to a} \frac{f(g(x)) - f(g(a))}{x - a}.$$

The two derivatives on the right hand side are given by

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

and

$$f'(g(a)) = \lim_{y \to a} \frac{f(y) - f(g(a))}{y - g(a)}.$$

Since g is a differentiable function it must also be a continuous function, and hence $\lim_{x\to a} g(x) = g(a)$. So we can substitute $y = g(x)$ in the limit defining $f'(g(a))$

$$f'(g(a)) = \lim_{y\to a} \frac{f(y) - f(g(a))}{y - g(a)} = \lim_{x\to a} \frac{f(g(x)) - f(g(a))}{g(x) - g(a)}. \tag{25}$$

Put all this together and you get

$$\begin{aligned}(f\circ g)'(a) &= \lim_{x\to a} \frac{f(g(x)) - f(g(a))}{x - a}\\ &= \lim_{x\to a} \frac{f(g(x)) - f(g(a))}{g(x) - g(a)} \cdot \frac{g(x) - g(a)}{x - a}\\ &= \lim_{x\to a} \frac{f(g(x)) - f(g(a))}{g(x) - g(a)} \cdot \lim_{x\to a} \frac{g(x) - g(a)}{x - a}\\ &= f'(g(a)) \cdot g'(a)\end{aligned}$$

which is what we were supposed to prove – the proof seems complete.

There is one flaw in this proof, namely, we have divided by $g(x) - g(a)$, which is not allowed when $g(x) - g(a) = 0$. This flaw can be fixed but we will not go into the details here.[4] □

28.4. *First example*

We go back to the functions

$$z = f(y) = y^2 + y \text{ and } y = g(x) = 2x + 1$$

from the beginning of this section. The composition of these two functions is

$$z = f(g(x)) = (2x + 1)^2 + (2x + 1) = 4x^2 + 6x + 2.$$

We can compute the derivative of this composed function, i.e. the derivative of z with respect to x in two ways. First, you simply differentiate the last formula we have:

$$\frac{dz}{dx} = \frac{d(4x^2 + 6x + 2)}{dx} = 8x + 6. \tag{26}$$

The other approach is to use the chain rule:

$$\frac{dz}{dy} = \frac{d(y^2 + y)}{dy} = 2y + 1,$$

and

$$\frac{dy}{dx} = \frac{d(2x + 1)}{dx} = 2.$$

Hence, by the chain rule one has

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx} = (2y + 1) \cdot 2 = 4y + 2. \tag{27}$$

The two answers (26) and (27) should be the same. Once you remember that $y = 2x + 1$ you see that this is indeed true:

$$y = 2x + 1 \implies 4y + 2 = 4(2x + 1) + 2 = 8x + 6.$$

The two computations of dz/dx therefore lead to the same answer. In this example there was no clear advantage in using the chain rule. The chain rule becomes useful when the functions f and g become more complicated.

[4] Briefly, you have to show that the function

$$h(y) = \begin{cases} \{f(y) - f(g(a))\}/(y - g(a)) & y \neq a\\ f'(g(a)) & y = a\end{cases}$$

is continuous.

28.5. *Example where you really need the Chain Rule*

We know what the derivative of $\sin x$ with respect to x is, but none of the rules we have found so far tell us how to differentiate $f(x) = \sin(2x)$.

The function $f(x) = \sin 2x$ is the composition of two simpler functions, namely

$$f(x) = g(h(x)) \text{ where } g(u) = \sin u \text{ and } h(x) = 2x.$$

We know how to differentiate each of the two functions g and h:

$$g'(u) = \cos u, \quad h'(x) = 2.$$

Therefore the chain rule implies that

$$f'(x) = g'(h(x))h'(x) = \cos(2x) \cdot 2 = 2\cos 2x.$$

Leibniz would have decomposed the relation $y = \sin 2x$ between y and x as

$$y = \sin u, \quad u = 2x$$

and then computed the derivative of $\sin 2x$ with respect to x as follows

$$\frac{d\sin 2x}{dx} \overset{u=2x}{=} \frac{d\sin u}{dx} = \frac{d\sin u}{du} \cdot \frac{du}{dx} = \cos u \cdot 2 = 2\cos 2x.$$

28.6. *The Power Rule and the Chain Rule*

The Power Rule, which says that for any function f and any rational number n one has

$$\frac{d}{dx}\big(f(x)^n\big) = nf(x)^{n-1}f'(x),$$

is a special case of the Chain Rule, for one can regard $y = f(x)^n$ as the composition of two functions

$$y = g(u), \quad u = f(x)$$

where $g(u) = u^n$. Since $g'(u) = nu^{n-1}$ the Chain Rule implies that

$$\frac{du^n}{dx} = \frac{du^n}{du} \cdot \frac{du}{dx} = nu^{n-1}\frac{du}{dx}.$$

Setting $u = f(x)$ and $\frac{du}{dx} = f'(x)$ then gives you the Power Rule.

28.7. *The volume of a growing yeast cell*

Consider the "real world example" from §28.2 again. There we considered a growing spherical yeast cell of radius

$$r = f(t).$$

The volume of this cell is

$$V = \frac{4}{3}\pi r^3 = \frac{4}{3}\pi f(t)^3.$$

We can regard this as the composition of two functions, $V = g(r) = \frac{4}{3}\pi r^3$ and $r = f(t)$.

According to the chain rule the rate of change of the volume with time is now

$$\frac{dV}{dt} = \frac{dV}{dr}\frac{dr}{dt}$$

i.e. it is the product of the rate of change of the volume with the radius of the cell and the rate of change of the cell's radius with time. From

$$\frac{dV}{dr} = \frac{d\frac{4}{3}\pi r^3}{dr} = 4\pi r^2$$

we see that

$$\frac{dV}{dr} = 4\pi r^2\,\frac{dr}{dt}.$$

For instance, if the radius of the cell is growing at 0.5μm/sec, and if its radius is $r = 3.0\mu$m, then the volume is growing at a rate of

$$\frac{dV}{dt} = 4\pi(3.0\mu\text{m})^2 \times 0.5\mu\text{m/sec} \approx 57\mu\text{m}^3/\text{sec}.$$

28.8. *A more complicated example*

Suppose you needed to find the derivative of

$$y = h(x) = \frac{\sqrt{x+1}}{(\sqrt{x+1}+1)^2}$$

We can write this function as a composition of two simpler functions, namely,

$$y = f(u), \quad u = g(x),$$

with

$$f(u) = \frac{u}{(u+1)^2} \text{ and } g(x) = \sqrt{x+1}.$$

The derivatives of f and g are

$$f'(u) = \frac{1\cdot(u+1)^2 - u\cdot 2(u+1)}{(u+1)^4} = \frac{u+1-2}{(u+1)^3} = \frac{u-1}{(u+1)^3},$$

and

$$g'(x) = \frac{1}{2\sqrt{x+1}}.$$

Hence the derivative of the composition is

$$h'(x) = \frac{d}{dx}\left\{\frac{\sqrt{x+1}}{(\sqrt{x+1}+1)^2}\right\} = f'(u)g'(x) = \frac{u-1}{(u+1)^3} \cdot \frac{1}{2\sqrt{x+1}}.$$

The result should be a function of x, and we achieve this by replacing all u's with $u = \sqrt{x+1}$:

$$\frac{d}{dx}\left\{\frac{\sqrt{x+1}}{(\sqrt{x+1}+1)^2}\right\} = \frac{\sqrt{x+1}-1}{(\sqrt{x+1}+1)^3} \cdot \frac{1}{2\sqrt{x+1}}.$$

The last step (where you replace u by its definition in terms of x) is important because the problem was presented to you with only x and y as variables while u was a variable you introduced yourself to do the problem.

Sometimes it is possible to apply the Chain Rule without introducing new letters, and you will simply think "the derivative is the derivative of the outside with respect to the inside times the derivative of the inside." For instance, to compute

$$\frac{d\,4+\sqrt{7+x^3}}{dx}$$

you could set $u = 7 + x^3$, and compute

$$\frac{d\,4+\sqrt{7+x^3}}{dx} = \frac{d\,4+\sqrt{u}}{du} \cdot \frac{du}{dx}.$$

Instead of writing all this explicitly, you could think of $u = 7+x^3$ as the function "inside the square root," and think of $4+\sqrt{u}$ as "the outside function." You would then immediately write

$$\frac{d}{dx}(4+\sqrt{7+x^3}) = \frac{1}{2\sqrt{7+x^3}} \cdot 3x^2.$$

28.9. *The Chain Rule and composing more than two functions*

Often we have to apply the Chain Rule more than once to compute a derivative. Thus if $y = f(u)$, $u = g(v)$, and $v = h(x)$ we have

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dv} \cdot \frac{dv}{dx}.$$

In functional notation this is

$$(f \circ g \circ h)'(x) = f'(g(h(x)) \cdot g'(h(x)) \cdot h'(x).$$

Note that each of the three derivatives on the right is evaluated at a different point. Thus if $b = h(a)$ and $c = g(b)$ the Chain Rule is

$$\left.\frac{dy}{dx}\right|_{x=a} = \left.\frac{dy}{du}\right|_{u=c} \cdot \left.\frac{du}{dv}\right|_{v=b} \cdot \left.\frac{dv}{dx}\right|_{x=a}.$$

For example, if $y = \dfrac{1}{1+\sqrt{9+x^2}}$, then $y = 1/(1+u)$ where $u = 1+\sqrt{v}$ and $v = 9+x^2$ so

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dv} \cdot \frac{dv}{dx} = -\frac{1}{(1+u)^2} \cdot \frac{1}{2\sqrt{v}} \cdot 2x.$$

so

$$\left.\frac{dy}{dx}\right|_{x=4} = \left.\frac{dy}{du}\right|_{u=6} \cdot \left.\frac{du}{dv}\right|_{v=25} \cdot \left.\frac{dv}{dx}\right|_{x=4} = -\frac{1}{7} \cdot \frac{1}{10} \cdot 8.$$

Exercises

28.1 – Let $y = \sqrt{1+x^3}$ and find dy/dx using the Chain Rule. Say what plays the role of $y = f(u)$ and $u = g(x)$.

28.2 – Repeat the previous exercise with $y = (1+\sqrt{1+x})^3$.

28.3 – Alice and Bob differentiated $y = \sqrt{1+x^3}$ with respect to x differently. Alice wrote $y = \sqrt{u}$ and $u = 1+x^3$ while Bob wrote $y = \sqrt{1+v}$ and $v = x^3$. Assuming neither one made a mistake, did they get the same answer?

28.4 – Let $y = u^3 + 1$ and $u = 3x + 7$. Find $\dfrac{dy}{dx}$ and $\dfrac{dy}{du}$. Express the former in terms of x and the latter in terms of u.

28.5 – Suppose that $f(x) = \sqrt{x}$, $g(x) = 1 + x^2$, $v(x) = f \circ g(x)$, $w(x) = g \circ f(x)$. Find formulas for $v(x)$, $w(x)$, $v'(x)$, and $w'(x)$.

28.6 – Compute the following derivatives

(a) $f(x) = \sin 2x - \cos 3x$ **(b)** $f(x) = \sin \dfrac{\pi}{x}$

(c) $f(x) = \sin(\cos 3x)$ **(d)** $f(x) = \dfrac{\sin x^2}{x^2}$

(e) $f(x) = \tan \sqrt{1+x^2}$ **(f)** $f(x) = \cos^2 x - \cos x^2$.

28.7 – Suppose that $f(x) = x^2 + 1$, $g(x) = x + 5$, and

$$v = f \circ g, \qquad w = g \circ f, \qquad p = f \cdot g, \qquad q = g \cdot f.$$

Find $v(x)$, $w(x)$, $p(x)$, and $q(x)$.

28.8 – Suppose that the functions f and g and their derivatives with respect to x have the following values at $x = 0$ and $x = 1$.

x	$f(x)$	$g(x)$	$f'(x)$	$g'(x)$
0	1	1	5	1/3
1	3	-4	-1/3	-8/3

Define

$$v(x) = f(g(x)), \qquad w(x) = g(f(x)), \qquad p(x) = f(x)g(x), \qquad q(x) = g(x)f(x).$$

Evaluate $v(0)$, $w(0)$, $p(0)$, $q(0)$, $v'(0)$ and $w'(0)$, $p'(0)$, $q'(0)$. If there is insufficient information to answer the question, so indicate.

28.9 – A differentiable function f satisfies $f(3) = 5$, $f(9) = 7$, $f'(3) = 11$ and $f'(9) = 13$. Find an equation for the tangent line to the curve $y = f(x^2)$ at the point $(x, y) = (3, 7)$.

28.10 – There is a function f whose second derivative satisfies

$$(\dagger) \qquad f''(x) = -64f(x).$$

(**a**) One such function is $f(x) = \sin ax$, provided you choose the right constant a: Which value should a have?

(**b**) Can you find other functions f that satisfy $(\dagger)$

28.11 – A cubical sponge is absorbing water, which causes it to expand. Its side at time t is $S(t)$. Its volume is $V(t)$.

(**a**) What is the relation between $S(t)$ and $V(t)$ (i.e. can you find a function f so that $V(t) = f(S(t))$)?

(**b**) Describe the meaning of the derivatives $S'(t)$ and $V'(t)$ in one plain english sentence each. If we measure lengths in inches and time in minutes, then what units do $t, S(t), V(t), S'(t)$ and $V'(t)$ have?

(**c**) What is the relation between $S'(t)$ and $V'(t)$?

(**d**) At the moment that the sponge's volume is 8 cubic inches, it is absorbing water at a rate of 2 cubic inch per minute. How fast is its side $S(t)$ growing?

29. Implicit differentiation

29.1. *The recipe*

Recall that an implicitely defined function is a function $y = f(x)$ which is defined by an equation of the form

$$F(x, y) = 0.$$

We call this equation the ***defining equation*** for the function $y = f(x)$. To find $y = f(x)$ for a given value of x you must solve the defining equation $F(x, y) = 0$ for y.

Here is a recipe for computing the derivative of an implicitely defined function.

(1) Differentiate the equation $F(x, y) = 0$; you may need the chain rule to deal with the occurences of y in $F(x, y)$;

(2) You can rearrange the terms in the result of step 1 so as to get an equation of the form

$$G(x, y)\frac{dy}{dx} + H(x, y) = 0,$$

where G and H are expressions containing x and y but not the derivative.

(3) Solve the equation in step 2 for $\dfrac{dy}{dx}$:

$$\frac{dy}{dx} = -\frac{H(x, y)}{G(x, y)} \tag{28}$$

(4) If you also have an explicit description of the function (i.e. a formula expressing $y = f(x)$ in terms of x) then you can substitute $y = f(x)$ in the expression (28) to get a formula for dy/dx in terms of x only.

Often no explicit formula for y is available and you can't take this last step. In that case (28) is as far as you can go.

Observe that by following this procedure you will get a formula for the derivative $\frac{dy}{dx}$ which contains both x and y.

29.2. *Dealing with equations of the form* $F_1(x, y) = F_2(x, y)$

If the implicit definition of the function is not of the form $F(x, y) = 0$ but rather of the form $F_1(x, y) = F_2(x, y)$ then you move all terms to the left hand side, and proceed as above. E.g. to deal with a function $y = f(x)$ which satisfies

$$y^2 + x = xy$$

you rewrite this equation as

$$y^2 + x - xy = 0$$

and set $F(x, y) = y^2 + x - xy$.

29.3. *Example – Derivative of* $\sqrt[4]{1 - x^4}$

Consider the function

$$f(x) = \sqrt[4]{1 - x^4}, \quad -1 \le x \le 1.$$

We will compute its derivative in two ways: first the direct method, and then using the method f implicit differentiation (i.e. the recipe above).

The direct approach goes like this:

$$\begin{aligned} f'(x) &= \frac{d\big(1-x^4\big)^{1/4}}{dx} \\ &= \tfrac{1}{4}\big(1-x^4\big)^{-3/4}\frac{d(1-x^4)}{dx} \\ &= \tfrac{1}{4}\big(1-x^4\big)^{-3/4}\big(-4x^3\big) \\ &= -\frac{x^3}{\big(1-x^4\big)^{3/4}} \end{aligned}$$

To find the derivative using implicit differentiation we must first find a nice implicit description of the function. For instance, we could decide to get rid of all roots or fractional exponents in the function and point out that $y = \sqrt[4]{1-x^4}$ satisfies the equation $y^4 = 1-x^4$. So our implicit description of the function $y = f(x) = \sqrt[4]{1-x^4}$ is

$F(x,y) = x^4 + y^4 - 1$

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

Differentiate both sides with respect to x (and remember that $y = f(x)$, so y here is a function of x), and you get

$G(x,y) = 4y^3$
$H(x,y) = 4x^3$

$$\frac{dx^4}{dx} + \frac{dy^4}{dx} - \frac{d1}{dx} = 0 \implies 4x^3 + 4y^3\frac{dy}{dx} = 0.$$

This last equation can be solved for dy/dx:

$$\frac{dy}{dx} = -\frac{x^3}{y^3}.$$

This is a nice and short form of the derivative, but it contains y as well as x. To express dy/dx in terms of x only, and remove the y dependency we use $y = \sqrt[4]{1-x^4}$. The result is

$$f'(x) = \frac{dy}{dx} = -\frac{x^3}{y^3} = -\frac{x^3}{\big(1-x^4\big)^{3/4}}.$$

29.4. *Another example*

Let f be a function defined by

$$y = f(x) \iff 2y + \sin y = x, \text{ i.e. } 2y + \sin y - x = 0.$$

For instance, if $x = 2\pi$ then $y = \pi$, i.e. $f(2\pi) = \pi$.

To find the derivative dy/dx we differentiate the defining equation

$$\frac{d(2y + \sin y - x)}{dx} = \frac{d0}{dx} \implies 2\frac{dy}{dx} + \cos y\frac{dy}{dx} - \frac{dx}{dx} = 0 \implies (2+\cos y)\frac{dy}{dx} - 1 = 0.$$

Solve for $\frac{dy}{dx}$ and you get

$$f'(x) = \frac{1}{2+\cos y} = \frac{1}{2+\cos f(x)}.$$

If we were asked to find $f'(2\pi)$ then, since we know $f(2\pi) = \pi$, we could answer

$$f'(2\pi) = \frac{1}{2+\cos\pi} = \frac{1}{2-1} = 1.$$

If we were asked $f'(\pi/2)$, then all we would be able to say is

$$f'(\pi/2) = \frac{1}{2+\cos f(\pi/2)}.$$

To say more we would first have to find $y = f(\pi/2)$, which one does by solving

$$2y + \sin y = \frac{\pi}{2}.$$

29.5. *Derivatives of Arc Sine and Arc Tangent*

Recall that

$$y = \arcsin x \iff x = \sin y \text{ and } -\tfrac{\pi}{2} \le y \le \tfrac{\pi}{2},$$

and

$$y = \arctan x \iff x = \tan y \text{ and } -\tfrac{\pi}{2} < y < \tfrac{\pi}{2}.$$

Theorem 29.1.

$$\frac{d \arcsin x}{dx} = \frac{1}{\sqrt{1-x^2}}$$

$$\frac{d \arctan x}{dx} = \frac{1}{1+x^2}$$

Proof. If $y = \arcsin x$ then $x = \sin y$. Differentiate this relation

$$\frac{dx}{dx} = \frac{d \sin y}{dx}$$

and apply the chain rule. You get

$$1 = \left(\cos y\right) \frac{dy}{dx},$$

and hence

$$\frac{dy}{dx} = \frac{1}{\cos y}.$$

How do we get rid of the y on the right hand side? We know $x = \sin y$, and also $-\frac{\pi}{2} \le y \le \frac{\pi}{2}$. Therefore

$$\sin^2 y + \cos^2 y = 1 \implies \cos y = \pm\sqrt{1 - \sin^2 y} = \pm\sqrt{1-x^2}.$$

Since $-\frac{\pi}{2} \le y \le \frac{\pi}{2}$ we know that $\cos y \ge 0$, so we must choose the positive square root. This leaves us with $\cos y = \sqrt{1-x^2}$, and hence

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

The derivative of $\arctan x$ is found in the same way, and you should really do this yourself. □

Exercises on implicit differentiation

29.1 – Find the derivative $f'(x)$ if $y = f(x)$ satisfies

(a) $xy = \dfrac{\pi}{6}$ **(b)** $\sin(xy) = \frac{1}{2}$

(c) $\dfrac{xy}{x+y} = 1$ **(d)** $x + y = xy$

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

(g) $(y-x)^2 + x = 0$ **(h)** $(y+x)^2 + 2y - x = 0$

(i) $\left(y^2-1\right)^2 + x = 0$ **(j)** $\left(y^2+1\right)^2 - x = 0$

(k) $x^3 + xy + y^3 = 3$ **(l)** $\sin x + \sin y = 1$

(m) $\sin x + xy + y^5 = \pi$ **(n)** $\tan x + \tan y = 1$

For each of these problems state what the expressions $F(x,y)$, $G(x,y)$ and $H(x,y)$ from the recipe in the beginning of this section are.

If you can find an explicit description of the function $y = f(x)$, say what it is.

29.2 – For each of the following explicitly defined functions find an implicit definition which does not involve taking roots. Then use this description to find the derivative dy/dx.

(**a**) $y = f(x) = \sqrt{1-x}$ (**b**) $y = f(x) = \sqrt[4]{x+x^2}$

(**c**) $y = f(x) = \sqrt{1-\sqrt{x}}$ (**d**) $y = f(x) = \sqrt[4]{x-\sqrt{x}}$

(**e**) $y = f(x) = \sqrt[3]{\sqrt{2x+1}-x^2}$ (**f**) $y = f(x) = \sqrt[4]{x+x^2}$

(**g**) $y = f(x) = \sqrt[3]{x-\sqrt{2x+1}}$ (**h**) $y = f(x) = \sqrt[4]{\sqrt[3]{x}}$

29.3 *(Inverse trig review)* – Simplify the following expressions, and indicate for which values of x (or θ, or ...) your simplification is valid. In case of doubt, try plotting the function on a graphing calculator.

(**a**) $\sin \arcsin x$ (**b**) $\tan \arctan z$

(**c**) $\cos \arcsin x$ (**d**) $\tan \arcsin \theta$

(**e**) $\arctan(\tan \theta)$ (**f**) $\arcsin(\sin \theta)$

(**g**) $\cot \arctan x$ (**h**) $\cot \arcsin x$

29.4 – Now that you know the derivatives of arcsin and arctan, you can find the derivatives of the following functions. What are they?

(**a**) $f(x) = \arcsin(2x)$ (**b**) $f(x) = \arcsin\sqrt{x}$

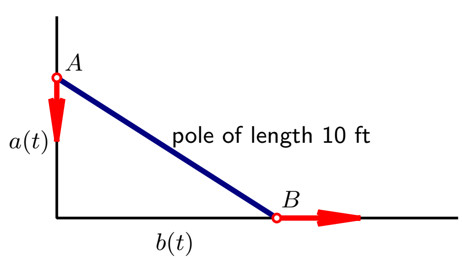
(**c**) $f(x) = \arctan(\sin x)$ (**d**) $f(x) = \sin \arctan x$

(**e**) $f(x) = \big(\arcsin x\big)^2$ (**f**) $f(x) = \dfrac{1}{1+(\arctan x)^2}$

(**g**) $f(x) = \sqrt{1-(\arcsin x)^2}$ (**h**) $f(x) = \dfrac{\arctan x}{\arcsin x}$

Exercises on rates of change

29.5 – A 10 foot long pole has one end (B) on the floor and another (A) against a wall.



If the bottom of the pole is 8 feet away from the wall, and if it is sliding away from the wall at 7 feet per second, then with what speed is the top (A) going down?

29.6 – A pole 10 feet long rests against a vertical wall. If the bottom of the pole slides away from the wall at a speed of 2 ft/s, how fast is the angle between the top of the pole and the wall changing when the angle is $\pi/4$ radians?

29.7 – A pole 13 meters long is leaning against a wall. The bottom of the pole is pulled along the ground away from the wall at the rate of 2 m/s. How fast is its height on the wall decreasing when the foot of the pole is 5 m away from the wall?

29.8 – A television camera is positioned 4000 ft from the base of a rocket launching pad. A rocket rises vertically and its speed is 600 ft/s when it has risen 3000 feet.

(**a**) How fast is the distance from the television camera to the rocket changing at that moment?

(**b**) How fast is the camera's angle of elevation changing at that same moment? (Assume that the television camera points toward the rocket.)

29.9 – A 2-foot tall dog is walking away from a streetlight which is on a 10-foot pole. At a certain moment, the tip of the dogs shadow is moving away from the streetlight at 5 feet per second. How fast is the dog walking at that moment?

29.10 – An isosceles triangle is changing its shape: the lengths of the two equal sides remain fixed at 2 inch, but the angle $\theta(t)$ between them changes.

Let $A(t)$ be the area of the triangle at time t. If the area increases at a constant rate of 0.5inch^2/sec, then how fast is the angle increasing or decreasing when $\theta = 60°$?

29.11 – A point P is moving in the first quadrant of the plane. Its motion is parallel to the x-axis; its distance to the x-axis is always 10 (feet). Its velocity is 3 feet per second to the left. We write θ for the angle between the positive x-axis and the line segment from the origin to P.

(**a**) Make a drawing of the point P.

(**b**) Where is the point when $\theta = \pi/3$?

(**c**) Compute the rate of change of the angle θ at the moment that $\theta = \frac{\pi}{3}$.

29.12 – The point Q is moving on the line $y = x$ with velocity 3 m/sec. Find the rate of change of the following quantities at the moment in which Q is at the point $(1, 1)$:

(**a**) the distance from Q to the origin,

(**b**) the distance from Q to the point $R(2, 0)$,

(**c**) the angle $\angle ORQ$ where R is again the point $R(2, 0)$.

29.13 – A point P is sliding on the parabola with equation $y = x^2$. Its x-coordinate is increasing at a constant rate of 2 feet/minute.

Find the rate of change of the following quantities at the moment that P is at $(3, 9)$:

(**a**) the distance from P to the origin,

(**b**) the area of the rectangle whose lower left corner is the origin and whose upper right corner is P,

(**c**) the slope of the tangent to the parabola at Q,

(**d**) the angle $\angle OPQ$ where Q is the point $(0, 3)$.

29.14 – A certain amount of gas is trapped in a cylinder with a piston. The ***ideal gas law*** from thermodynamics says that if the cylinder is not heated, and if the piston moves slowly, then one has

$$pV = CT$$

where p is the pressure in the gas, V is its volume, T its temperature (in degrees Kelvin) and C is a constant depending on the amount of gas trapped in the cylinder.

(**a**) If the pressure is 10psi (pounds per square inch), if the volume is 25inch^3, and if the piston is moving so that the gas volume is expanding at a rate of 2inch^3 per minute, then what is the rate of change of the pressure?

(**b**) The ideal gas law turns out to be only approximately true. A more accurate description of gases is given by ***van der Waals' equation of state***, which says that

$$\left(p + \frac{a}{V^2}\right)(V - b) = C$$

where a, b, C are constants depending on the temperature and the amount and type of gas in the cylinder.

Suppose that the cylinder contains fictitious gas for which one has $a = 12$ and $b = 3$. Suppose that at some moment the volume of gas is 12in^3, the pressure is 25psi and suppose the gas is expanding at 2 inch^3 per minute. Then how fast is the pressure changing?

V. Graph Sketching and Max-Min Problems

The signs of the first and second derivatives of a function tell us something about the shape of its graph. In this chapter we learn how to find that information.

30. Tangent and Normal lines to a graph

The slope of the tangent the tangent to the graph of f at the point $(a, f(a))$ is

$$m = f'(a) \tag{29}$$

and hence the equation for the tangent is

$$y = f(a) + f'(a)(x - a). \tag{30}$$

The slope of the normal line to the graph is $-1/m$ and thus one could write the equation for the normal as

$$y = f(a) - \frac{x-a}{f'(a)}. \tag{31}$$

When $f'(a) = 0$ the tangent is horizontal, and hence the normal is vertical. In this case the equation for the normal cannot be written as in (31), but instead one gets the simpler equation

$$y = f(a).$$

Both cases are covered by this form of the equation for the normal

$$x = a + f'(a)(f(a) - y). \tag{32}$$

Both (32) and (31) are formulas that you shouldn't try to remember. It is easier to remember that if the slope of the tangent is $m = f'(a)$, then the slope of the normal is $-1/m$.

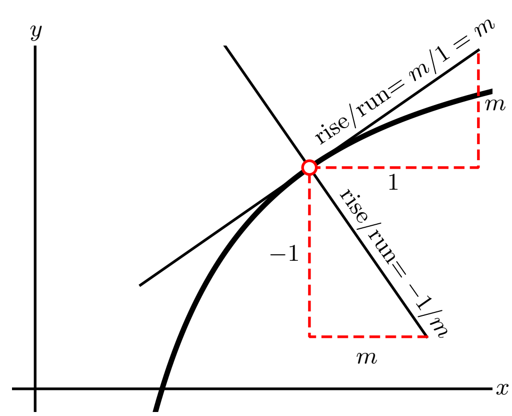


Figure 14. Why the slope of the normal is $-1/$(the slope of the tangent).

31. The intermediate value theorem

You could say that a function is continuous if you can draw its graph with out taking your pencil off the paper. A more precise version is the *intermediate value theorem*:

Theorem 31.1. *If f is a continuous function on an interval $a \le x \le b$, and if y is some number between $f(a)$ and $f(b)$, then there is a number c with $a \le c \le b$ such that $f(c) = y$.*

Here "y between $f(a)$ and $f(b)$" means that $f(a) \le y \le f(b)$ if $f(a) \le f(b)$, and $f(b) \le y \le f(a)$ if $f(b) \le f(a)$.

Example – Square root of 2

Consider the function $f(x) = x^2$. Since $f(1) < 2$ and $f(2) = 4 > 2$ the intermediate value theorem with $a = 1$, $b = 2$, $y = 2$ tells us that there is a number c between 1 and 2 such that $f(c) = 2$, i.e. for which $c^2 = 2$. So the theorem tells us that the square root of 2 exists.

Example – The equation $\theta + \sin\theta = \frac{\pi}{2}$

Consider the function $f(x) = x + \sin x$. It is a continuous function at all x, so from $f(0) = 0$ and $f(\pi) = \pi$ it follows that there is a number θ between 0 and π such that $f(\theta) = \pi/2$. In other words, the equation

$$\theta + \sin\theta = \frac{\pi}{2} \tag{33}$$

has a solution θ with $0 \le \theta \le \pi/2$. Unlike the previous example, where we knew the solution was $\sqrt{2}$, there is no simple formula for the solution to (33).

Example – Solving $1/x = 0$

If we apply the intermediate value theorem to the function $f(x) = 1/x$ on the interval $[a, b] = [-1, 1]$, then we see that for any y between $f(a) = f(-1) = -1$ and $f(b) = f(1) = 1$ there is a number c in the interval $[-1, 1]$ such that $1/c = y$. For instance, we could choose $y = 0$ (that's between -1 and $+1$), and conclude that there is some c with $-1 \le c \le 1$ and $1/c = 0$.

But there is no such c, because $1/c$ is never zero! So we have done something wrong, and the mistake we made is that we overlooked that our function $f(x) = 1/x$ is not defined on the ***whole*** interval $-1 \le x \le 1$ because it is not defined at $x = 0$. *The moral: always check the hypotheses of a theorem before you use it!*

32. Finding sign changes of a function

The intermediate value theorem implies the following very useful fact.

Theorem 32.1. *If f is continuous function on some interval $a < x < b$, and if $f(x) \ne 0$ for all x in this interval, then $f(x)$ is either positive for all $a < x < b$ or else it is negative for all $a < x < b$.*

Proof. The theorem says that there can't be two numbers $a < x_1 < x_2 < b$ such that $f(x_1)$ and $f(x_2)$ have opposite signs. If there were two such numbers then the intermediate value theorem would imply that somewhere between x_1 and x_2 there was a c with $f(c) = 0$. But we are assuming that $f(c) \ne 0$ whenever $a < c < b$. □

32.1. *Example*

Consider

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

The zeros of f (i.e. the solutions of $f(x) = 0$) are

$$-\frac{1}{2}, 1, 3.$$

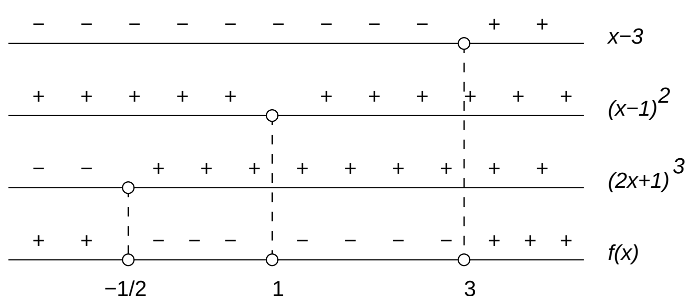
Theorem 32.1 tells us that $f(x)$ must have the same sign for all x between these zeros. We can find those signs by computing $f(x)$ for one x from each interval between two consecutive zeros. We find

$$\begin{aligned} f(-1) = (-4)(-2)^2(-3)^3 > 0 &\implies f(x) > 0 \text{ for } x < -\tfrac{1}{2} \\ f(0) = (-3)(-1)^2(1)^3 < 0 &\implies f(x) < 0 \text{ for } -\tfrac{1}{2} < x < 1 \\ f(2) = (-1)(1)^2(5)^3 < 0 &\implies f(x) < 0 \text{ for } 1 < x < 3 \\ f(4) = (1)(3)^2(9)^3 > 0 &\implies f(x) > 0 \text{ for } x > 3. \end{aligned}$$

When the given function is factored into simple functions, as in this example, there is a different way of finding out where f is positive, and where f is negative. For each of the factors $x-3$, $(x-1)^2$ and $(2x+1)^3$ it is easy to determine the sign, for any given x. These signs can only change at a zero of the factor. Thus we have

- $x - 3$ is positive for $x > 3$ and negative for $x < 3$;
- $(x-1)^2$ is always positive (except at $x = 1$);
- $(2x+1)^3$ is positive for $x > -\frac{1}{2}$ and negative for $x < -\frac{1}{2}$.

Multiplying these signs we get the same conclusions as above. We can summarize this computation in the following diagram:



33. **Increasing and decreasing functions**

Here are four very similar definitions – look closely to see how they differ.

- A function is called ***increasing*** if $a < b$ implies $f(a) < f(b)$ for all numbers a and b in the domain of f.
- A function is called ***decreasing*** if $a < b$ implies $f(a) > f(b)$ for all numbers a and b in the domain of f.
- The function f is called ***non-increasing*** if $a < b$ implies $f(a) \leq f(b)$ for all numbers a and b in the domain of f.
- The function f is called ***non-decreasing*** if $a < b$ implies $f(b) \geq f(b)$ for all numbers a and b in the domain of f.

You can summarize these definitions as follows:

f **is . . .**	**if for all** a **and** b **one has. . .**
Increasing:	$a < b \implies f(a) < f(b)$
Decreasing:	$a < b \implies f(a) > f(b)$
Non-increasing:	$a < b \implies f(a) \geq f(b)$
Non-decreasing:	$a < b \implies f(a) \leq f(b)$

The sign of the derivaitve of f tells you if f is increasing or not. More precisely:

Theorem 33.1. *If a function is non-decreasing on an interval* $a < x < b$ *then* $f'(x) \geq 0$ *for all* x *in that interval.*

If a function is non-increasing on an interval $a < x < b$ *then* $f'(x) \leq 0$ *for all* x *in that interval.*

For instance, if f is non-decreasing, then for any given x and any positive Δx one has $f(x + \Delta x) \geq f(x)$ and hence

$$\frac{f(x + \Delta x) - f(x)}{\Delta x} \geq 0.$$

now let $\Delta x \searrow 0$ and you find that

$$f'(x) = \lim_{\Delta x \searrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \geq 0.$$

What about the converse, i.e. if you know the sign of f' then what can you say about f? For this we have the following

Theorem 33.2. *Suppose* f *is a differentiable function on an interval* (a, b).

If $f'(x) > 0$ *for all* $a < x < b$*, then* f *is increasing.*

If $f'(x) < 0$ *for all* $a < x < b$*, then* f *is decreasing.*

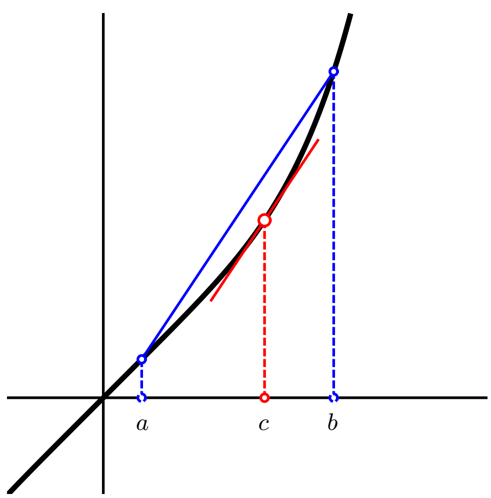


Figure 15. According to the Mean Value Theorem there always is some number c between a and b such that the tangent to the graph of f is parallel to the line segment connecting the two points $(a, f(a))$ and $(b, f(b))$. This is true for any choice of a and b; c depends on a and b of course.

The proof is based on the Mean Value theorem which also finds use in many other situations:

Theorem 33.3 (The Mean Value Theorem). *If f is a differentiable function on the interval $a \leq x \leq b$, then there is some number c, with $a < c < b$ such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Proof of theorem 33.2. We show that $f'(x) > 0$ for all x implies that f is increasing. Let $x_1 < x_2$ be two numbers between a and b. Then the Mean Value Theorem implies that there is some c between x_1 and x_2 such that

$$f'(c) = \frac{f(x_2) - f(x_1)}{x_2 - x_1},$$

or

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1).$$

Since we know that $f'(c) > 0$ and $x_2 - x_1 > 0$ it follows that $f(x_2) - f(x_1) > 0$, i.e. $f(x_2) > f(x_1)$. □

34. **Examples**

Armed with these theorems we can now split the graph of any function into increasing and decreasing parts simply by computing the derivative $f'(x)$ and finding out where $f'(x) > 0$ and where $f'(x) < 0$ – i.e. we apply the method form the previous section to f' rather than f.

34.1. *Example: the parabola* $y = x^2$

The familiar graph of $f(x) = x^2$ consists of two parts, one decreasing and one increasing. You can see this from the derivative which is

$$f'(x) = 2x \begin{cases} > 0 & \text{for } x > 0 \\ < 0 & \text{for } x < 0. \end{cases}$$

Therefore the function $f(x) = x^2$ is

decreasing for $x < 0$
increasing for $x > 0$.

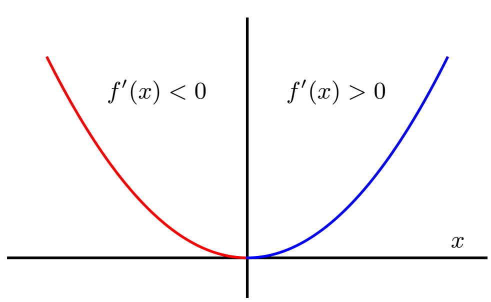


Figure 16. Increasing and decreasing branches of the parabola

34.2. *Example: the hyperbola $y = 1/x$*

The derivative of the function $f(x) = 1/x = x^{-1}$ is

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

which is always negative. You would therefore think that this function is decreasing, or at least non-increasing: if $a < b$ then $1/a \geq 1/b$. But this isn't true if you take $a = -1$ and $b = 1$:

$$a = -1 < 1 = b, \text{ but } \frac{1}{a} = -1 < 1 = \frac{1}{b} \text{ !!}$$

The problem is that we used theorem 33.2, but it you carefully read that theorem then you see that it applies to functions ***that are defined on an interval***. The function in this example, $f(x) = 1/x$, is not defined on the interval $-1 < x < 1$ because it isn't defined at $x = 0$. That's why you can't conclude that the $f(x) = 1/x$ is increasing from $x = -1$ to $x = +1$.

On the other hand, the function is defined and differentiable on the interval $0 < x < \infty$, so theorem 33.2 tells us that $f(x) = 1/x$ is decreasing for $x > 0$. This means, that as long as x is positive, increasing x will decrease $1/x$.

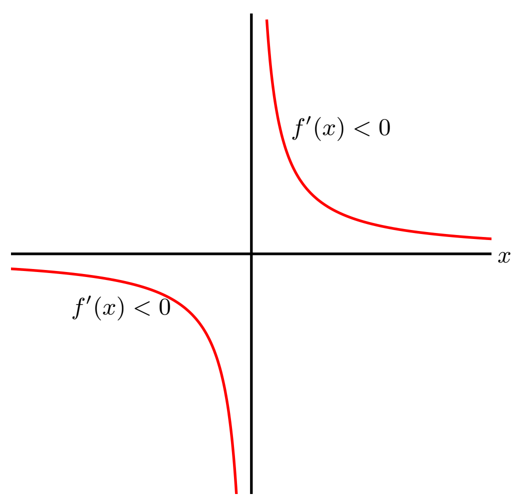


Figure 17. The graph of $y = 1/x$ consists of two decreasing pieces.

34.3. *Graph of a cubic function*

Consider the function

$$y = f(x) = x^3 - x.$$

Its derivative is

$$f'(x) = 3x^2 - 1.$$

We try to find out where f' is positive, and where it is negative by factoring $f'(x)$

$$f'(x) = 3\big(x^2 - \tfrac{1}{3}\big) = 3\big(x + \tfrac{1}{3}\sqrt{3}\big)\big(x - \tfrac{1}{3}\sqrt{3}\big)$$

from which you see that

$$f'(x) > 0 \text{ for } x < -\tfrac{1}{3}\sqrt{3}$$
$$f'(x) < 0 \text{ for } -\tfrac{1}{3}\sqrt{3} < x < \tfrac{1}{3}\sqrt{3}$$
$$f'(x) > 0 \text{ for } x > \tfrac{1}{3}\sqrt{3}$$

Therefore the function f is

$$\text{increasing on } \left(-\infty, -\tfrac{1}{3}\sqrt{3}\right)$$
$$\text{decreasing on } \left(-\tfrac{1}{3}\sqrt{3}, \tfrac{1}{3}\sqrt{3}\right)$$
$$\text{increasing on } \left(\tfrac{1}{3}\sqrt{3}, \infty\right)$$

At the two points $x = \pm\frac{1}{3}\sqrt{3}$ one has $f'(x) = 0$ so there the tangent will be horizontal. This leads us to the following picture of the graph of f:

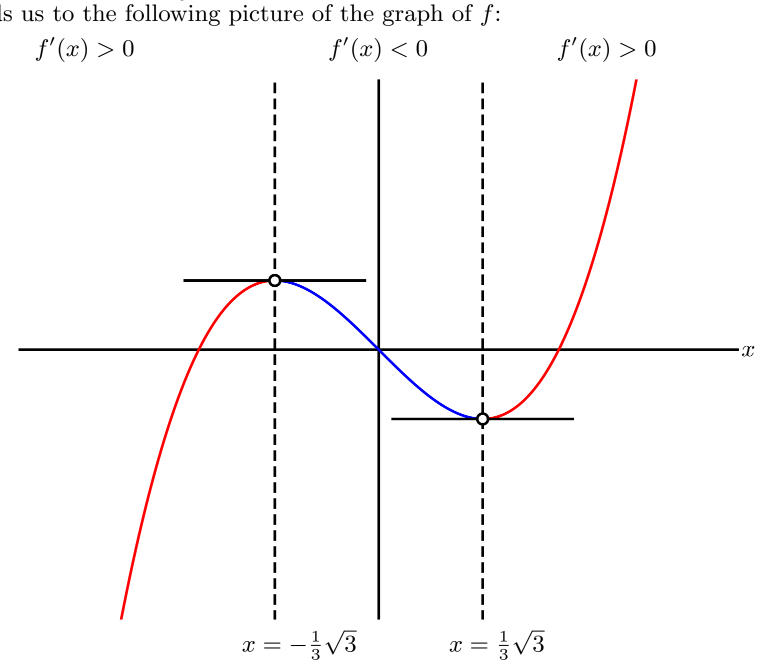


Figure 18. The graph of $f(x) = x^3 - x$.

34.4. *A function whose tangent turns up and down infinitely often near the origin*

We end with a weird example. Somewhere in the mathematician's zoo of curious functions the following will be on exhibit. Consider the function

$$f(x) = \frac{x}{2} + x^2 \sin\frac{\pi}{x}.$$

For $x = 0$ this formula is undefined, and we are free to define $f(0) = 0$. This makes the function continuous at $x = 0$. In fact, this function is differentiable at $x = 0$, with derivative given by

$$f'(0) = \lim_{x\to 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x\to 0} \frac{1}{2} + x\sin\frac{\pi}{x} = \frac{1}{2}.$$

(To find the limit apply the sandwich theorem to $-|x| \le x\sin\frac{\pi}{x} \le |x|$.)

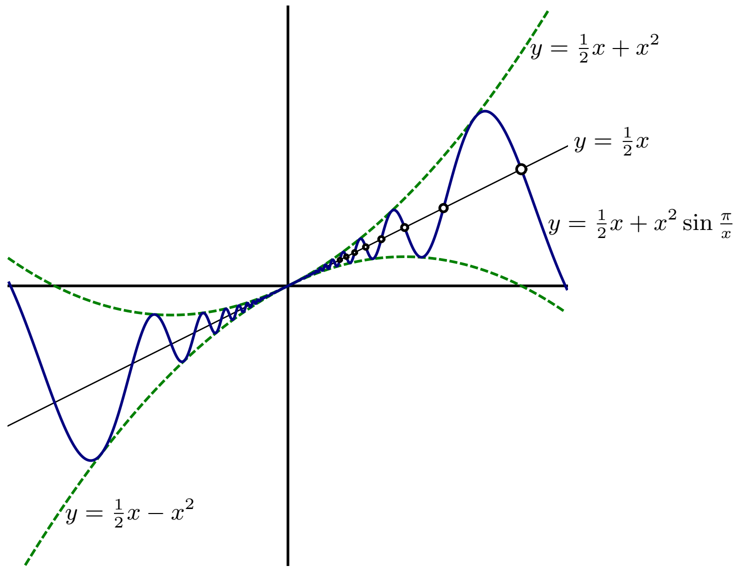


Figure 19. Positive derivative at a point ($x = 0$) does not mean that the function is "increasing near that point." The slopes at the intersection points alternate between $\frac{1}{2} - \pi$ and $\frac{1}{2} + \pi$.

So the slope of the tangent to the graph at the origin is positive ($\frac{1}{2}$), and one would *think* that the function should be increasing near $x = 0$ (i.e. bigger x gives bigger $f(x)$.) The point of this example is that this turns out not to be true.

To explain why not, we must compute the derivative of this function for $x \neq 0$. It is given by

$$f'(x) = \frac{1}{2} - \pi\cos\frac{\pi}{x} + 2x\sin\frac{\pi}{x}.$$

Now consider the sequence of intersection points $P_1, P_2, \ldots$ of the graph with the line $y = x/2$. They are

$$P_k(x_k, y_k), \qquad x_k = \frac{1}{k}, \quad y_k = f(x_k).$$

For larger and larger k the points P_k tend to the origin (the x coordinate is $\frac{1}{k}$ which goes to 0 as $k \to \infty$). The slope of the tangent at P_k is given by

$$\begin{aligned} f'(x_k) &= \frac{1}{2} - \pi\cos\frac{\pi}{1/k} + 2\frac{1}{k}\sin\frac{\pi}{1/k} \\ &= \frac{1}{2} - \pi\cos k\pi + \frac{2}{k}\sin k\pi \\ &= \begin{cases} -\frac{1}{2} - \pi \approx -2.64159265358979\ldots & \text{for } k \text{ even} \\ \frac{1}{2} + \pi \approx +3.64159265358979\ldots & \text{for } k \text{ odd} \end{cases} \end{aligned}$$

In other words, along the sequence of points P_k the slope of the tangent flip-flops between $\frac{1}{2} - \pi$ and $\frac{1}{2} + \pi$, i.e. between a positive and a negative number.

In particular, the slope of the tangent at the odd intersection points is negative, and so you would expect the function to be decreasing there. In other words we see that ***even though the derivative at $x = 0$ of this function is positive, there are points on the graph arbitrarily close to the origin where the tangent has negative slope.***

35. Maxima and Minima

A function has a ***global maximum*** at some a in its domain if $f(x) \le f(a)$ for all other x in the domain of f. Global maxima are sometimes also called "absolute maxima."

A function has a ***local maximum*** at some a in its domain if there is a small $\delta > 0$ such that $f(x) \le f(a)$ for all x with $a - \delta < x < a + \delta$ which lie in the domain of f.

Every global maximum is a local maximum, but a local maximum doesn't have to be a global maximum.

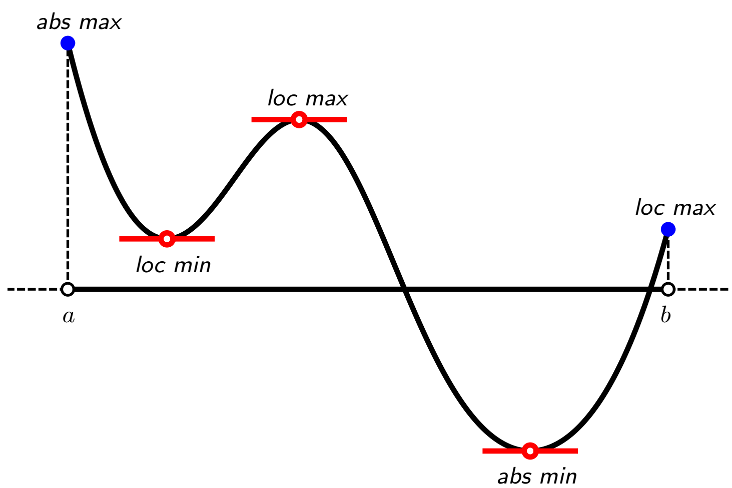


Figure 20. A function defined on an interval $[a, b]$ with one interior absolute minimum, another interior local minimum, an interior local maximum, and two local maxima on the boundary, one of which is in fact an absolute maximum.

35.1. *Where to find local maxima and minima*

Any x value for which $f'(x) = 0$ is called a ***stationary point*** for the function f.

Theorem 35.1. *Suppose f is a differentiable function on some interval $[a, b]$.*

Every local maximum or minimum of f is either one of the end points of the interval $[a, b]$, or else it is a stationary point for the function f.

Proof. Suppose that f has a local maximum at x and suppose that x is not a or b. By assumption the left and right hand limits

$$f'(x) = \lim_{\Delta x \nearrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \text{ and } f'(x) = \lim_{\Delta x \searrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

both exist and they are equal.

Since f has a local maximum at x we have $f(x + \Delta x) - f(x) \le 0$ if $-\delta < \Delta x < \delta$. In the first limit we also have $\Delta x < 0$, so that

$$\lim_{\Delta x \nearrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \le 0$$

Hence $f'(x) \le 0$.

In the second limit we have $\Delta x > 0$, so

$$\lim_{\Delta x \searrow 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} \ge 0$$

which implies $f'(x) \ge 0$.

Thus we have shown that $f'(x) \le 0$ and $f'(x) \ge 0$ at the same time. This can only be true if $f'(x) = 0$. □

35.2. *How to tell if a stationary point is a maximum, a minimum, or neither*

If $f'(c) = 0$ then c is a stationary point (by definition), and it might be local maximum or a local minimum. You can tell what kind of stationary point c is by looking at the signs of $f'(x)$ for x near c.

Theorem 35.2. *If in some small interval $(c-\delta, c+\delta)$ you have $f'(x) < 0$ for $x < c$ and $f'(x) > 0$ for $x > c$ then f has a local maximum at $x = c$.*

If in some small interval $(c-\delta, c+\delta)$ you have $f'(x) < 0$ for $x < c$ and $f'(x) > 0$ for $x > c$ then f has a local maximum at $x = c$.

The reason is simple: if f increases to the left of c and decreases to the right of c then it has a maximum at c. More precisely:

> if $f'(x) > 0$ for x between $c-\delta$ and c, then f is increasing for $c-\delta < x < c$ and therefore $f(x) < f(c)$ for x between $c-\delta$ and c.
>
> If in addition $f'(x) < 0$ for $x > c$ then f is decreasing for x between c and $c+\delta$, so that $f(x) < f(c)$ for those x.
>
> Combine these two facts and you get $f(x) \le f(c)$ for $c-\delta < x < c+\delta$.

35.3. *Example – local maxima and minima of $f(x) = x^3 - x$*

In §34.3 we had found that the function $f(x) = x^3 - x$ is decreasing when $-\infty < x < -\frac{1}{3}\sqrt{3}$, and also when $\frac{1}{3}\sqrt{3} < x < \infty$, while it is increasing when $-\frac{1}{3}\sqrt{3} < x < \frac{1}{3}\sqrt{3}$. It follows that the function has a local minimum at $x = -\frac{1}{3}\sqrt{3}$, and a local maximum at $x = \frac{1}{3}\sqrt{3}$.

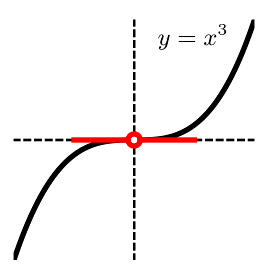
Neither the local maximum nor the local minimum are global max or min since

$$\lim_{x\to-\infty} f(x) = +\infty \text{ and } \lim_{x\to\infty} f(x) = -\infty.$$

35.4. *A stationary point that is neither a maximum nor a minimum*

If you look for stationary points of the function $f(x) = x^3$ you find that there's only one, namely $x = 0$. The derivative $f'(x) = 3x^2$ does not change sign at $x = 0$, so the test in Theorem 35.2 does not tell us anything.

And in fact, $x = 0$ is neither a local maximum nor a local minimum since $f(x) < f(0)$ for $x < 0$ and $f(x) > 0$ for $x > 0$.



36. Must there always be a maximum?

Theorem 35.1 is very useful since it tells you how to find (local) maxima and minima. The following theorem is also useful, but in a different way. It doesn't say how to find maxima or minima, but it tells you that they do exist, and hence that you are not wasting your time trying to find a maximum or minimum.

Theorem 36.1. *Let f be continuous function defined on the* ***closed*** *interval $a \le x \le b$. Then f attains its maximum and also its minimum somewhere in this interval. In other words there exist real numbers c and d such that*

$$f(c) \le f(x) \le f(d)$$

whenever $a \le x \le b$.

The proof of this theorem requires a more careful definition of the real numbers than we have given in Chapter 1, and we will take the theorem for granted.

37. Examples – functions with and without maxima or minima

In the following three example we explore what can happen if some of the hypotheses in Theorem 36.1 are not met.

Question: Does the function

$$f(x) = \begin{cases} x & \text{for } 0 \le x < 1 \\ 0 & \text{for } x = 1. \end{cases}$$

have a maximum on the interval $0 \le x \le 1$?

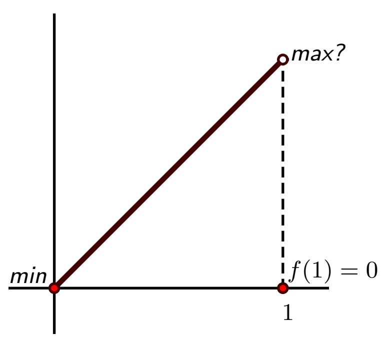


Figure 21. A function without a maximum

Answer: No. What would the maximal value be? Since

$$\lim_{x \nearrow 1} f(x) = \lim_{x \nearrow 1} x = 1$$

The maximal value cannot be less than 1. On the other hand the function is never larger than 1. So if there were a number a in the interval $[0, 1]$ such that $f(a)$ was the maximal value of f, then we would have $f(a) = 1$. If you now search the interval for numbers a with $f(a) = 1$, then you notice that such an a does not exist. Conclusion: this function does ***not*** attain its maximum on the interval $[0, 1]$.

What about Theorem 36.1? That theorem only applies to continuous functions, and the function f in this example is not continuous at $x = 1$. For at $x = 1$ one has

$$f(1) = 0 \neq 1 = \lim_{x \nearrow 1} f(x).$$

So all it takes for the Theorem to fail is that the function f be discontinuous at just one point.

Question: Does the function

$$f(x) = \frac{1}{x}, \quad 1 \leq x < \infty$$

have a maximum or minimum?

Answer: The function has a maximum at $x = 1$, but it has no minimum.

Concerning the maximum: if $x \geq 1$ then $f(x) = 1/x \leq 1$, while $f(1) = 1$. Hence $f(x) \leq f(1)$ for all x in the interval $[1, \infty)$ and that is why f attains its maximum at $x = 1$.

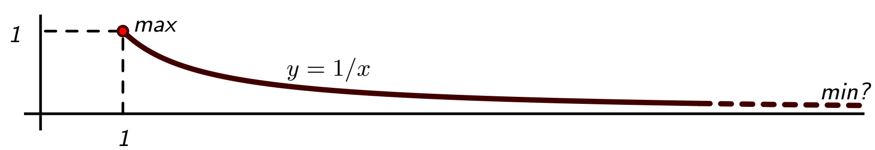


Figure 22. $f(x) = 1/x$ has a maximum but no minimum on the interval $1 \leq x < \infty$.

If we look for a minimal value of f then we note that $f(x) \geq 0$ for all x in the interval $[1, \infty)$, and also that

$$\lim_{x \to \infty} f(x) = 0,$$

so that ***if*** f attains a minimum at some a with $1 \leq a < \infty$, then the minimal value $f(a)$ must be zero. However, the equation $f(a) = 0$ has no solution – f does not attain its minimum.

Why does Theorem 36.1 not apply? In this example the function f is continuous on the whole interval $[1, \infty)$, but this interval is not a closed interval, i.e. it is not of the form $[a, b]$ (it does not include its endpoints).

38. General method for sketching the graph of a function

Given a differentiable function f defined on some interval $a \leq x \leq b$, you can find the increasing and decreasing parts of the graph, as well as all the local maxima and minima by following this procedure:

(1) find all solutions of $f'(x) = 0$ in the interval $[a, b]$: these are called the *critical* or *stationary* points for f.
(2) find the sign of $f'(x)$ at all other points
(3) each stationary point at which $f'(x)$ actually changes sign is a local maximum or local minimum. Compute the function value $f(x)$ at each stationary point.
(4) compute the function values at the endpoints of the interval, i.e. compute $f(a)$ and $f(b)$.
(5) the absolute maximum is attained at the stationary point or the boundary point with the highest function value; the absolute minimum occurs at the boundary or stationary point with the smallest function value.

If the interval is unbounded, i.e. if the function is defined for $-\infty < x < \infty$ then you can't compute the values $f(a)$ and $f(b)$, but instead you should compute $\lim_{x\to\pm\infty} f(x)$.

38.1. *Example – the graph of a rational function*

Let's "sketch the graph" of the function

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

By looking at the signs of numerator and denominator we see that

$$f(x) > 0 \text{ for } 0 < x < 1$$
$$f(x) < 0 \text{ for } x < 0 \text{ and also for } x > 1.$$

We compute the derivative of f

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

Hence $f'(x) = 0$ holds if and only if

$$1 - 2x - x^2 = 0$$

and the solutions to this quadratic equation are $-1 \pm \sqrt{2}$. These two roots will appear several times and it will shorten our formulas if we abbreviate

$$A = -1 - \sqrt{2} \text{ and } B = -1 + \sqrt{2}.$$

To see if the derivative changes sign we factor the numerator and denominator. The denominator is always positive, and the numerator is

$$-x^2 - 2x + 1 = -(x^2 + 2x - 1) = -(x - A)(x - B).$$

Therefore

$$f'(x) \begin{cases} < 0 & \text{for } x < A \\ > 0 & \text{for } A < x < B \\ < 0 & \text{for } x > B \end{cases}$$

It follows that f is decreasing on the interval $(-\infty, A)$, increasing on the interval (A, B) and decreasing again on the interval (B, ∞). Therefore

$$A \text{ is a local minimum, and } B \text{ is a local maximum.}$$

Are these global maxima and minima?

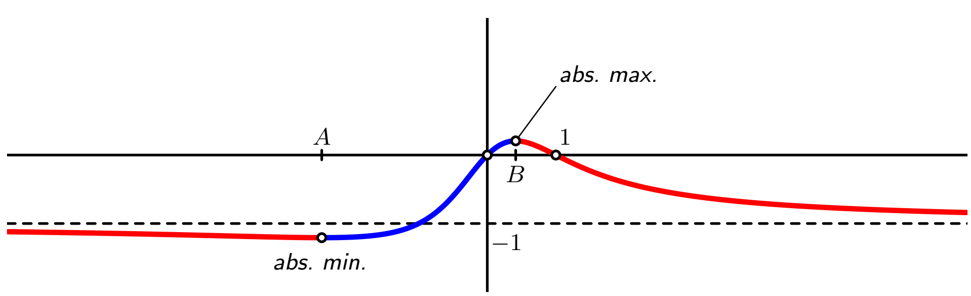


Figure 23. The graph of $f(x) = (x - x^2)/(1 + x^2)$

Since we are dealing with an unbounded interval we must compute the limits of $f(x)$ as $x \to \pm\infty$. You find

$$\lim_{x\to\infty} f(x) = \lim_{x\to-\infty} f(x) = -1.$$

Since f is decreasing between $-\infty$ and A, it follows that

$$f(A) \le f(x) < -1 \text{ for } -\infty < x \le A.$$

Similarly, f is decreasing from B to $+\infty$, so

$$-1 < f(x) \le f(-1+\sqrt{2}) \text{ for } B < x < \infty.$$

Between the two stationary points the function is increasing, so

$$f(-1-\sqrt{2}) \le f(x) \le f(B) \text{ for } A \le x \le B.$$

From this it follows that $f(x)$ is the smallest it can be when $x = A = -1-\sqrt{2}$ and at its largest when $x = B = -1+\sqrt{2}$: the local maximum and minimum which we found are in fact a global maximum and minimum.

39. Convexity, Concavity and the Second Derivative

By definition, a function f is ***convex*** on some interval $a < x < b$ if the line segment connecting any pair of points on the graph lies *above* the piece of the graph between those two points.

The function is called ***concave*** if the line segment connecting any pair of points on the graph lies *below* the piece of the graph between those two points.

A point on the graph of f where $f''(x)$ changes sign is called an ***inflection point***.

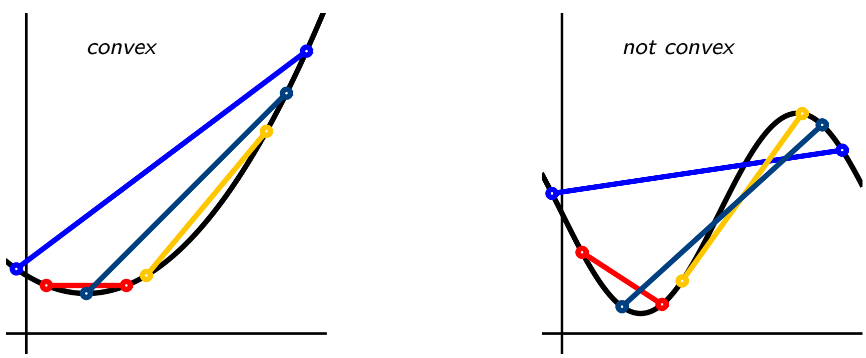


Figure 24. If a graph is convex then all chords lie above the graph. If it is not convex then some chords will cross the graph or lie below it.

Instead of “convex” and “concave” one often says ”curved upwards” or ”curved downwards.”

You can use the second derivative to tell if a function is concave or convex.

Theorem 39.1. *A function f is convex on some interval $a < x < b$ if and only if $f''(x) \ge 0$ for all x on that interval.*

Theorem 39.2. *A function f is convex on some interval $a < x < b$ if and only if the derivative $f'(x)$ is a nondecreasing function on that interval.*

A proof using the Mean Value Theorem will be given in class.

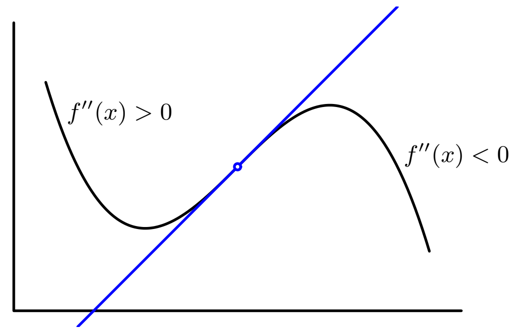


Figure 25. At an inflection point the tangent crosses the graph.

39.1. *Example – the cubic function* $f(x) = x^3 - x$

The second derivative of the function $f(x) = x^3 - x$ is

$$f''(x) = 6x$$

which is positive for $x > 0$ and negative for $x < 0$. Hence, in the graph in §34.3, the origin is an inflection point, and the piece of the graph where $x > 0$ is convex, while the piece where $x < 0$ is concave.

39.2. *The second derivative test*

In §35.2 we saw how you can tell if a stationary point is a local maximum or minimum by looking at the sign changes of $f'(x)$. There is another way of distinguishing between local maxima and minima which involves computing the second derivative.

Theorem 39.3. *If c is a stationary point for a function f, and if $f''(c) < 0$ then f has a local maximum at $x = c$.*

If $f''(c) > 0$ then f has a local minimum at c.

The theorem doesn't say what happens when $f''(c) = 0$. In that case you must go back to checking the signs of the first derivative near the stationary point.

The basic reason why this theorem is true is that if c is a stationary point with $f''(c) > 0$ then "$f'(x)$ is increasing near $x = c$" and hence $f'(x) < 0$ for $x < c$ and $f'(x) > 0$ for $x > c$. So the function f is decreasing for $x < c$ and increasing for $x > c$, and therefore it reaches a local minimum at $x = c$.

39.3. *Example – that cubic function again*

Consider the function $f(x) = x^3 - x$ from §34.3 and §39.1. We had found that this function has two stationary points, namely at $x = \pm\frac{1}{3}\sqrt{3}$. By looking at the sign of $f'(x) = 3x^2 - 1$ we concluded that $-\frac{1}{3}\sqrt{3}$ is a local maximum while $+\frac{1}{3}\sqrt{3}$ is a local minimum. Instead of looking at $f'(x)$ we could also have computed $f''(x)$ at $x = \pm\frac{1}{3}\sqrt{3}$ and applied the second derivative test. Here is how it goes:

Since $f''(x) = 6x$ we have

$$f''\big(-\tfrac{1}{3}\sqrt{3}\big) = -2\sqrt{3} < 0 \text{ and } f''\big(\tfrac{1}{3}\sqrt{3}\big) = 2\sqrt{3} > 0.$$

Therefore f has a local maximum at $-\frac{1}{3}\sqrt{3}$ and a local minimum at $\frac{1}{3}\sqrt{3}$.

39.4. *When the second derivative test doesn't work*

Usually the second derivative test will work, but sometimes a stationary point c has $f''(c) = 0$. In this case the second derivative test gives no information at all. The figure below shows you the graphs of three functions, all three of which have a stationary point at $x = 0$. In all three cases the second derivative vanishes at $x = 0$ so the second derivative test says nothing. As you can see, the stationary point can be a local maximum, a local minimum, or neither.

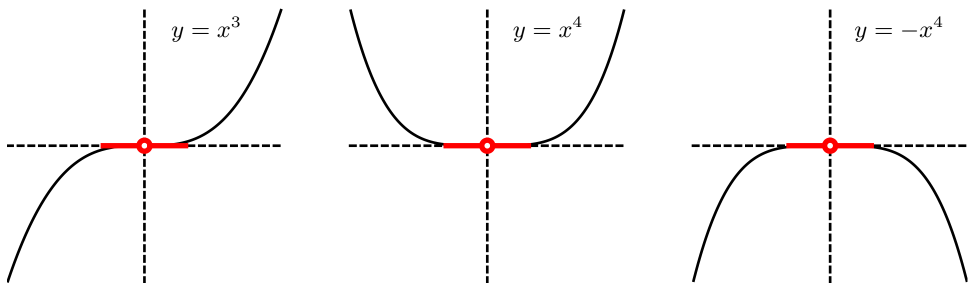


Figure 26. Three functions for which the second derivative test doesn't work.

40. **Proofs of some of the theorems**

40.1. *Proof of the Mean Value Theorem*

Let m be the slope of the chord connecting the points $(a, f(a))$ and $(b, f(b))$, i.e.

$$m = \frac{f(b) - f(a)}{b - a},$$

and consider the function

$$g(x) = f(x) - f(a) - m(x - a).$$

This function is continuous (since f is continuous), and g attains its maximum and minimum at two numbers $c_{\min}$ and $c_{\max}$.

There are now two possibilities: either at least one of $c_{\min}$ or $c_{\max}$ is an interior point, or else both $c_{\min}$ and $c_{\max}$ are endpoints of the interval $a \le x \le b$.

Consider the first case: one of these two numbers is an interior point, i.e. if $a < c_{\min} < b$ or $a < c_{\max} < b$, then the derivative of g must vanish at $c_{\min}$ or $c_{\max}$. If one has $g'(c_{\min}) = 0$, then one has

$$0 = g'(c_{\min}) = f'(c_{\min}) - m, \text{i.e. } m = f'(c_{\min}).$$

The definition of m implies that one gets

$$f'(c_{\min}) = \frac{f(b) - f(a)}{b - a}.$$

If $g'(c_{\max}) = 0$ then one gets $m = f'(c_{\max})$ and hence

$$f'(c_{\max}) = \frac{f(b) - f(a)}{b - a}.$$

We are left with the remaining case, in which both $c_{\min}$ and $c_{\max}$ are end points. To deal with this case note that at the endpoints one has

$$g(a) = 0 \text{ and } g(b) = 0.$$

Thus the maximal and minimal values of g are both zero! This means that $g(x) = 0$ for all x, and thus that $g'(x) = 0$ for all x. Therefore we get $f'(x) = m$ for ***all*** x, and not just for some c.

40.2. *Proof of Theorem 33.1*

If f is a non-increasing function and if it is differentiable at some interior point a, then we must show that $f'(a) \geq 0$.

Since f is non-decreasing, one has $f(x) \geq f(a)$ for all $x > a$. Hence one also has

$$\frac{f(x) - f(a)}{x - a} \geq 0$$

for all $x > a$. Let $x \searrow a$, and you get

$$f'(a) = \lim_{x \searrow a} \frac{f(x) - f(a)}{x - a} \geq 0.$$

40.3. *Proof of Theorem 33.2*

Suppose f is a differentiable function on an interval $a < x < b$, and suppose that $f'(x) \geq 0$ on that interval. We must show that f is non-decreasing on that interval, i.e. we have to show that if $x_1 < x_2$ are two numbers in the interval (a, b), then $f(x_1) \geq f(x_2)$. To prove this we use the Mean Value Theorem: given x_1 and x_2 the Mean Value Theorem hands us a number c with $x_1 < c < x_2$, and

$$f'(c) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

We don't know where c is exactly, but it doesn't matter because we do know that wherever c is we have $f'(c) \geq 0$. Hence

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \geq 0.$$

Multiply with $x_2 - x_1$ (which we are allowed to do since $x_2 > x_1$ so $x_2 - x_1 > 0$) and you get

$$f(x_2) - f(x_1) \geq 0,$$

as claimed.

Exercises

40.1 – Find equations for the tangent and normal lines

to the curve ...	**at the point...**
$y = 4x/(1 + x^2)$	$(1, 2)$
$y = 8/(4 + x^2)$	$(2, 1)$
$y^2 = 2x + x^2$	$(2, 2)$
$xy = 3$	$(1, 3)$

40.2 – Does the curve $y = x^4 - 2x^2 + 2$ have any horizontal tangents? If so, where?

40.3 – At some point $(a, f(a))$ on the graph of $f(x) = -1 + 2x - x^2$ the tangent to this graph goes through the origin. Which point is it?

40.4 – (**a**) What does the Intermediate Value Theorem say?

(**b**) What does the Mean Value Theorem say?

(**c**) ***If*** $f(a) = 0$ ***and*** $f(b) = 0$ ***then there is a*** c ***between*** a ***and*** b ***such that*** $f'(c) = 0$***.*** Show that this follows from the Mean Value Theorem. (Help! A proof! Relax: this one is not difficult. Make a drawing of the situation, then read the Mean Value Theorem again.)

(**d**) What is a stationary point?

(**e**) How can you tell if a local maximum is a global maximum?

(**f**) What is an inflection point?

(**g**) Give an example of a function for which $f''(0) = 0$ even though the graph of does not have an inflection point at $x = 0$.

(**h**) Draw four graphs of functions, one for each of the following four combinations

$f' > 0$ and $f'' > 0$ $\qquad$ $f' > 0$ and $f'' < 0$

$f' < 0$ and $f'' > 0$ $\qquad$ $f' < 0$ and $f'' < 0$

40.5 – Sketch the graph of the following functions. You should

(1) find where f, f' and f'' are positive or negative
(2) find all stationary points
(3) decide which stationary points are local maxima or minima
(4) decide which local max/minima are in fact global max/minima
(5) find all inflection points
(6) find "horizontal asymptotes," i.e. compute the limits $\lim_{x\to\pm\infty} f(x)$ when appropriate.

(**a**) $y = x^3 + 2x^2$ $\qquad$ (**b**) $y = x^3 - 4x^2$

(**c**) $y = x^4 + 27x$ $\qquad$ (**d**) $y = x^4 - 27x$

(**e**) $y = x^4 + 2x^2 - 3$ $\qquad$ (**f**) $y = x^4 - 5x^2 + 4$

(**g**) $y = x^5 + 16x$ $\qquad$ (**h**) $y = x^5 - 16x$

(**i**) $y = \dfrac{x}{x+1}$ $\qquad$ (**j**) $y = \dfrac{x}{1+x^2}$

(**k**) $y = \dfrac{x^2}{1+x^2}$ $\qquad$ (**l**) $y = \dfrac{1+x^2}{1+x}$

(**m**) $y = x + \dfrac{1}{x}$ $\qquad$ (**n**) $y = x - \dfrac{1}{x}$

(**o**) $y = x^3 + 2x^2 + x$ $\qquad$ (**p**) $y = x^3 + 2x^2 - x$

(**q**) $y = x^4 - x^3 - x$ $\qquad$ (**r**) $y = x^4 - 2x^3 + 2x$

(**s**) $y = \sqrt{1+x^2}$ $\qquad$ (**t**) $y = \sqrt{1-x^2}$

(**u**) $y = \sqrt[4]{1+x^2}$ $\qquad$ (**v**) $y = \dfrac{1}{1+x^4}$

40.6 – The following functions are periodic, i.e. they satisfy $f(x+L) = f(x)$ for all x, where the constant L is called the period of the function. The graph of a periodic function repeats itself indefinitely to the left and to the right. It therefore has infinitely many (local) minima and maxima, and infinitely many inflections points. Sketch the graphs of the following functions as in the previous problem, but only list those "interesting points" that lie in the interval $0 \le x < 2\pi$.

(**a**) $y = \sin x$ (**b**) $y = \sin x + \cos x$

(**c**) $y = \sin x + \sin^2 x$ (**d**) $y = 2\sin x + \sin^2 x$

(**e**) $y = 4\sin x + \sin^2 x$ (**f**) $y = 2\cos x + \cos^2 x$

(**g**) $y = \dfrac{4}{2+\sin x}$ (**h**) $y = \bigl(2+\sin x\bigr)^2$

40.7 – Find the domain and sketch the graphs of each of the following functions

(**a**) $y = \arcsin x$ (**b**) $y = \arctan x$ (**c**) $y = 2\arctan x - x$

(**d**) $y = \arctan(x^2)$ (**e**) $y = 3\arcsin(x) - 5x$ (**f**) $y = 6\arcsin(x) - 10x^2$

40.8 – In the following two problems it is not possible to solve the equation $f'(x) = 0$, but you can tell something from the second derivative.

(**a**) Show that the function $f(x) = x\arctan x$ is convex. Then sketch the graph of f.

(**b**) Show that the function $g(x) = x\arcsin x$ is convex. Then sketch the graph of g.

40.9 – For each of the following functions use the derivative to decide if they are increasing, decreasing or neither on the indicated intervals

(**a**) $f(x) = \dfrac{x}{1+x^2}$ $10 < x < \infty$

(**b**) $f(x) = \dfrac{2+x^2}{x^3-x}$ $1 < x < \infty$

(**c**) $f(x) = \dfrac{2+x^2}{x^3-x}$ $0 < x < 1$

(**d**) $f(x) = \dfrac{2+x^2}{x^3-x}$ $0 < x < \infty$

41. Optimization Problems

Often a problem can be phrased as

For which value of x in the interval $a \le x \le b$ is $f(x)$ the largest?

In other words you are given a function f on an interval $[a, b]$ and you must find all global maxima of f on this interval.

If the function is continuous then according to theorem 36.1 there always is at least one x in the interval $[a, b]$ which maximizes $f(x)$.

If f is differentiable then we know what to do: any local maximum is either a stationary point or one of the end points a and b. Therefore you can find the global maxima by following this recipe:

(1) Find all stationary points of f;
(2) Compute $f(x)$ at each stationary point you found in step (1);
(3) Compute $f(a)$ and $f(b)$;
(4) The global maxima are those stationary- or endpoints from steps (2) and (3) which have the largest function value.

Usually there is only one global maximum, but sometimes there can be more.

If you have to *minimize* rather than *maximize* a function, then you must look for global minima. The same recipe works (of course you should look for the smallest function value instead of the largest in step 4.)

The difficulty in optimization problems frequently lies not with the calculus part, but rather with setting up the problem. Choosing which quantity to call x and finding the function f is half the job.

41.1. *Example – The rectangle with largest area and given perimeter*

Which rectangle has the largest area, among all those rectangles for which the total length of the sides is 1?

Solution. If the sides of the rectangle have lengths x and y, then the total length of the sides is

$$L = x + x + y + y = 2(x + y)$$

and the area of the rectangle is

$$A = xy.$$

So are asked to find the largest possible value of $A = xy$ provided $2(x + y) = 1$. The lengths of the sides can also not be negative, so x and y must satisfy $x \ge 0$, $y \ge 0$.

We now want to turn this problem into a question of the form "maximize a function over some interval." The quantity which we are asked to maximize is A, but it depends on two variables x and y instead of just one variable. However, the variables x and y are not independent since we are only allowed to consider rectangles with $L = 1$. From this equation we get

$$L = 1 \implies y = \tfrac{1}{2} - x.$$

Hence we must find the maximum of the quantity

$$A = xy = x\left(\tfrac{1}{2} - x\right)$$

The values of x which we are allowed to consider are only limited by the requirements $x \ge 0$ and $y \ge 0$, i.e. $x \le \frac{1}{2}$. So we end up with this problem:

Find the maximum of the function $f(x) = x\left(\frac{1}{2} - x\right)$ on the interval $0 \le x \le \frac{1}{2}$.

Before we start computing anything we note that the function f is a polynomial so that it is differentiable, and hence continuous, and also that the interval $0 \le x \le \frac{1}{2}$ is closed. Therefore the theory guarantees that there is a maximum and our recipe will show us where it is.

The derivative is given by

$$f'(x) = \tfrac{1}{2} - 2x,$$

and hence the only stationary point is $x = \frac{1}{4}$. The function value at this point is

$$f\tfrac{1}{4}) = \tfrac{1}{4}(\tfrac{1}{2} - \tfrac{1}{4}) = \tfrac{1}{16}.$$

At the endpoints one has $x = 0$ or $x = \frac{1}{2}$, which corresponds to a rectangle one of whose sides has length zero. The area of such rectangles is zero, and so this is not the maximal value we are looking for.

We conclude that the largest area is attained by the rectangle whose sides have lengths

$$x = \tfrac{1}{4}, \text{ and } y = \tfrac{1}{2} - \tfrac{1}{4} = \tfrac{1}{4},$$

i.e. by a square with sides $\frac{1}{4}$.

41.2. *Exercises*

41.1 – By definition, the perimeter of a rectangle is the sum of the lengths of its four sides. Which rectangle, of all those whose perimeter is 1, has the smallest area? Which one has the largest area?

41.2 – Which rectangle of area 100in^2 minimizes its height plus two times its length?

41.3 – You have 1 yard of string from which you make a circular wedge with radius R and opening angle θ. Which choice of θ and R will give you the wedge with the largest area? Which choice leads to the smallest area?

[A circular wedge is the figure consisting of two radii of a circle and the arc connecting them. So the yard of string is used to form the two radii and the arc.]

41.4 – A rounded rectangle is a rectangle of width W and height H, with two half circles on the left and right sides glued on (so the circles have diameter H.) Which rounded rectangle with area 10in^2 has the smallest perimeter?

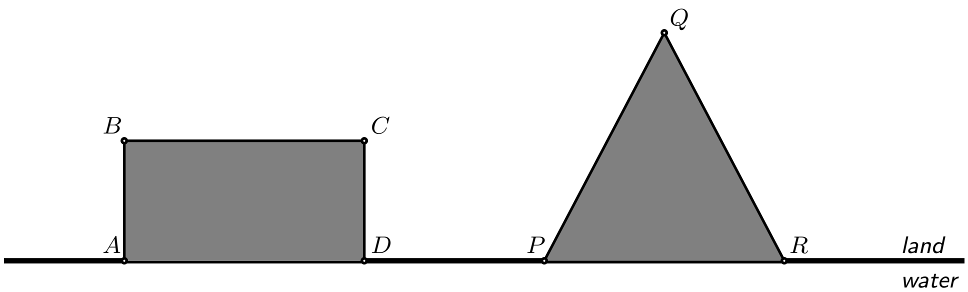
41.5 – (**a**) You have a sheet of metal with area $100\ \text{in}^2$ from which you are to make a cylindrical soup can. If r is the radius of the can and h its height, then which h and r will give you the can with the largest volume?

(**b**) If instead of making a plain cylinder you replaced the flat top and bottom of the cylinder with two spherical caps, then (using the same 100in^2 os sheet metal), then which choice of radius and height of the cylinder give you the container with the largest volume?

(**c**) Suppose you only replace the top of the cylinder with a spherical cap, and leave the bottom flat, then which choice of height and radius of the cylinder result in the largest volume?

41.6 – A triangle has one vertex at the origin $O(0,0)$, another at the point $A(2a,0)$ and the third at $(a, a/(1+a^3))$. What are the largest and smallest areas this triangle can have if $0 \le a < \infty$?

41.7 – According to tradition Dido was the founder and first Queen of Carthage. When she arrived on the north coast of Africa (∼800BC) the locals allowed her to take as much land as could be enclosed with the hide of one ox. She cut the hide into thin strips and put these together to form a length of 100 yards[5].



(**a**) If Dido wanted a rectangular region, then how wide should she choose it to enclose as much area as possible (the coastal edge of the boundary doesn't count, so in this problem the length $AB + BC + CD$ is 100 yards.)

(**b**) If Dido chose a region in the shape of an isosceles triangle PQR, then how wide should she make it to maximize its area (again, don't include the coast in the perimiter: $PQ+QR$ is 100 yards long, and $PQ = QR$.)

41.8 – The product of two numbers x, y is 16. We know $x \geq 1$ and $y \geq 1$. What is the greatest possible sum of the two numbers?

41.9 – What are the smallest and largest values that $(\sin x)(\sin y)$ can have if $x + y = \pi$ and if x and y are both nonnegative?

41.10 – What are the smallest and largest values that $(\cos x)(\cos y)$ can have if $x + y = \frac{\pi}{2}$ and if x and y are both nonnegative?

41.11 – (**a**) What are the smallest and largest values that $\tan x + \tan y$ can have if $x+y = \frac{\pi}{2}$ and if x and y are both nonnegative?

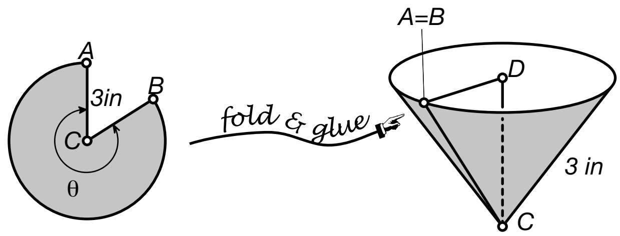
(**b**) What are the smallest and largest values that $\tan x + 2\tan y$ can have if $x + y = \frac{\pi}{2}$ and if x and y are both nonnegative?

41.12 – The cost per hour of fuel to run a locomotive is $v^2/25$ dollars, where v is speed (in miles per hour), and other costs are \$100 per hour regardless of speed. What is the speed that minimizes cost per mile ?

41.13 – Josh is in need of coffee. He has a circular filter with 3 inch radius. He cuts out a wedge and glues the two edges AC and BC together to make a conical filter to hold the ground coffee. The volume V of the coffee cone depends the angle θ of the piece of filter paper Josh made.

---

[5] I made that number up. For the rest start at `http://en.wikipedia.org/wiki/Dido`



(**a**) Find the volume in terms of the angle θ. (Hint: how long is the circular arc AB on the left? How long is the circular top of the cone on the right? If you know that you can find the radius $AD = BD$ of the top of the cone, and also the height CD of the cone.)

(**b**) Which angle θ maximizes the volume V?

VI. Exponentials and Logarithms

In this chapter we first recall some facts about exponentials (x^y with $x > 0$ and y arbitrary): they should be familiar from algebra, or "precalculus." What is new is perhaps the definition of x^y when y is not a fraction: e.g., $2^{3/4}$ is the 4th root of the third power of 2 ($\sqrt[4]{2^3}$), but what is $2^{\sqrt{2}}$?

Then we ask "what is the derivative of $f(x) = a^x$?" The answer leads us to the famous number $e \approx 2.718\,281\,828\,459\,045\,235\,360\,287\,471\,352\,662\,497\,757\,247\,093\,699\,95\cdots$.

Finally, we compute the derivative of $f(x) = \log_a x$, and we look at things that "grow exponentially."

42. Exponents

Here we go over the definition of x^y when x and y are arbitrary real numbers, with $x > 0$.

For any real number x and any positive integer $n = 1, 2, 3, \ldots$ one defines

$$x^n = \overbrace{x \cdot x \cdot \cdots \cdot x}^{n \text{ times}}$$

and, if $x \neq 0$,

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

One defines $x^0 = 1$ for any $x \neq 0$.

To define $x^{p/q}$ for a general fraction $\frac{p}{q}$ one must assume that the number x is positive. One then defines

$$x^{p/q} = \sqrt[q]{x^p}. \tag{34}$$

This does not tell us how to define x^a is the exponent a is not a fraction. One can define x^a for irrational numbers a by taking limits. For example, to define $2^{\sqrt{2}}$, we look at the sequence of numbers you get by truncating the decimal expansion of $\sqrt{2}$, i.e.

$$a_1 = 1, \quad a_2 = 1.4 = \tfrac{14}{10}, \quad a_3 = 1.41 = \tfrac{141}{100}, \quad a_4 = 1.414 = \tfrac{1414}{1000}, \quad \cdots.$$

Each a_n is a fraction, so that we know what 2^{a_n} is, e.g. $2^{a_4} = \sqrt[1000]{2^{1414}}$. Our definition of $2^{\sqrt{2}}$ then is

$$2^{\sqrt{2}} = \lim_{n\to\infty} 2^{a_n},$$

i.e. we define $2^{\sqrt{2}}$ as the limit of the sequence of numbers

$$2,\ \sqrt[10]{2^{14}},\ \sqrt[100]{2^{141}},\ \sqrt[1000]{2^{1414}}, \ldots$$

(See table 2.)

Here one ought to prove that this limit exists, and that its value does not depend on the particular choice of numbers a_n tending to a. We will not go into these details in this course.

It is shown in precalculus texts that the exponential functions satisfy the following properties:

$$\boxed{x^a x^b = x^{a+b}, \qquad \frac{x^a}{x^b} = x^{a-b}, \qquad \left(x^a\right)^b = x^{ab}} \tag{35}$$

x	2^x
1.0000000000	$\mathbf{2}.000000000000$
1.4000000000	$\mathbf{2.6}39015821546$
1.4100000000	$\mathbf{2.6}57371628193$
1.4140000000	$\mathbf{2.66}4749650184$
1.4142000000	$\mathbf{2.6651}19088532$
1.4142100000	$\mathbf{2.6651}37561794$
1.4142130000	$\mathbf{2.66514}3103798$
1.4142135000	$\mathbf{2.665144}027466$
$\vdots$	$\vdots$

Table 2. Approximating $2^{\sqrt{2}}$. Note that as x gets closer to $\sqrt{2}$ the quantity 2^x appears to converge to some number. This limit is our definition of $2^{\sqrt{2}}$.

provided a and b are fractions. One can show that these properties still hold if a and b are real numbers (not necessarily fractions.) Again, we won't go through the proofs here.

Now instead of considering x^a as a function of x we can pick a positive number a and consider the function $f(x) = a^x$. This function is defined for all real numbers x (as long as the base a is positive.).

42.1. *The trouble with powers of negative numbers*

The cube root of a negative number is well defined. For instance $\sqrt[3]{-8} = -2$ because $(-2)^3 = -8$. In view of the definition (34) of $x^{p/q}$ we can write this as

$$(-8)^{1/3} = \sqrt[3]{(-8)^1} = \sqrt[3]{-8} = -2.$$

But there is a problem: since $\frac{2}{6} = \frac{1}{3}$ you would think that $(-8)^{2/6} = (-8)^{1/3}$. However our definition (34) tells us that

$$(-8)^{2/6} = \sqrt[6]{(-8)^2} = \sqrt[6]{+64} = +2.$$

Another example:

$$(-4)^{1/2} = \sqrt{-4} \text{ is not defined}$$

but, even though $\frac{1}{2} = \frac{2}{4}$,

$$(-4)^{2/4} = \sqrt[4]{(-4)^2} = \sqrt[4]{+16} = 2 \text{ is defined.}$$

There are two ways out of this mess:

(1) avoid taking fractional powers of negative numbers
(2) when you compute $x^{p/q}$ first simplify the fraction by removing common divisors of p and q.

The safest is just not to take fractional powers of negative numbers.

Given that fractional powers of negative numbers cause all these headaches it is not surprising that we didn't try to define x^a for negative x if a is irrational. For example, $(-8)^\pi$ is not defined[6].

[6]There is a definition of $(-8)^\pi$ which uses complex numbers. You will see this next semester if you take math 222.

43. Logarithms

Briefly, $y = \log_a x$ is the inverse function to $y = a^x$. This means that, by definition,

$$y = \log_a x \iff x = a^y.$$

In other words, $\log_a x$ is the answer to the question "for which number y does one have $x = a^y$?" The number $\log_a x$ is called ***the logarithm with base a of x.*** In this definition both a and x must be positive.

For instance,

$$2^3 = 8, \quad 2^{1/2} = \sqrt{2}, \quad 2^{-1} = \frac{1}{2}$$

so

$$\log_2 8 = 3, \quad \log_2\big(\sqrt{2}\big) = \frac{1}{2}, \quad \log_2 \frac{1}{2} = -1.$$

Also:

$$\log_2(-3) \text{ doesn't exist}$$

because there is no number y for which $2^y = -3$ (2^y is always positive) and

$$\log_{-3} 2 \text{ doesn't exist either}$$

because $y = \log_{-3} 2$ would have to be some real number which satisfies $(-3)^y = 2$, and we don't take non-integer powers of negative numbers.

44. Properties of logarithms

In general one has

$$\log_a a^x = x, \text{ and } a^{\log_a x} = x.$$

There is a subtle difference between these formulas: the first one holds for all real numbers x, but the second only holds for $x > 0$, since $\log_a x$ doesn't make sense for $x \le 0$.

Again, one finds the following formulas in precalculus texts:

$$\boxed{\begin{aligned} \log_a xy &= \log_a x + \log_a y \\ \log_a \frac{x}{y} &= \log_a x - \log_a y \\ \log_a x^y &= y \log_a x \\ \log_a x &= \frac{\log_b x}{\log_b a} \end{aligned}} \tag{36}$$

They follow from (35).

45. Graphs of exponential functions and logarithms

Figure 27 shows the graphs of some exponential functions $y = a^x$ with different values of a, and figure 28 shows the graphs of $y = \log_2 x$, $y = \log_3 x$, $\log_{1/2} x$, $\log_{1/3}(x)$ and $y = \log_{10} x$. Can you tell which is which? (Yes, you can.)

From algebra/precalc recall:

$$\boxed{\text{If } a > 1 \text{ then } f(x) = a^x \text{ is an increasing function.}}$$

and

$$\boxed{\text{If } 0 < a < 1 \text{ then } f(x) = a^x \text{ is a decreasing function.}}$$

In other words, for $a > 1$ it follows from $x_1 < x_2$ that $a^{x_1} < a^{x_2}$; if $0 < a < 1$, then $x_1 < x_2$ implies $a^{x_1} > a^{x_2}$.

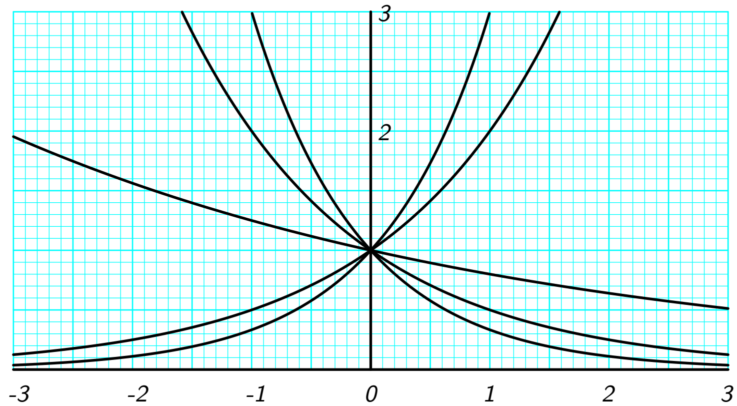


Figure 27. The graphs of $y = 2^x, 3^x, (1/2)^x, (0.1)^x$ and $y = (4/5)^x$.

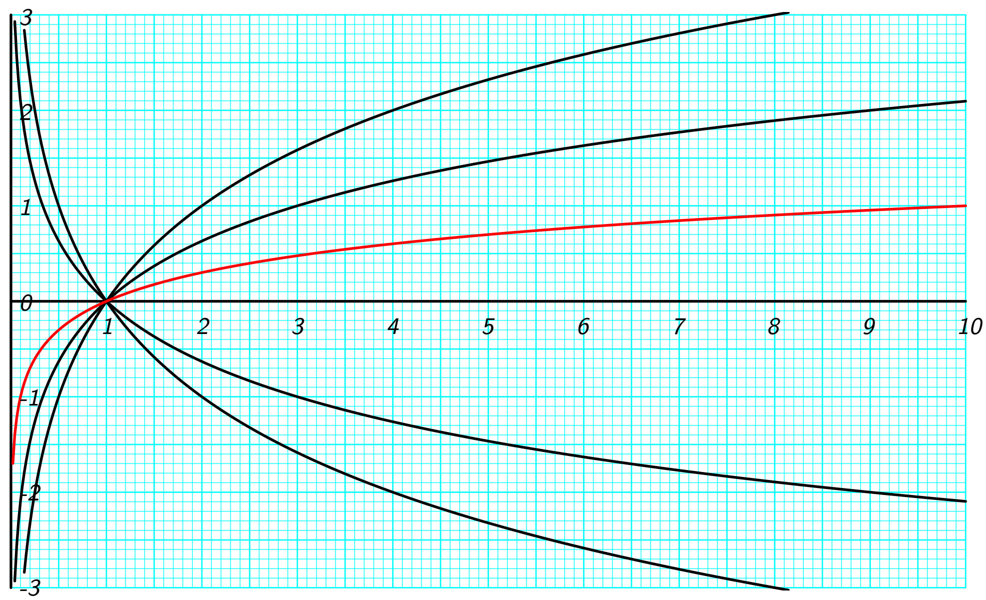


Figure 28. Graphs of some logarithms

46. The derivative of a^x and the definition of e

To begin, we try to differentiate the function $y = 2^x$:

$$\begin{aligned}\frac{d2^x}{dx} &= \lim_{\Delta x\to 0}\frac{2^{x+\Delta x}-2^x}{\Delta x}\\ &= \lim_{\Delta x\to 0}\frac{2^x 2^{\Delta x}-2^x}{\Delta x}\\ &= \lim_{\Delta x\to 0} 2^x\frac{2^{\Delta x}-1}{\Delta x}\\ &= 2^x\lim_{\Delta x\to 0}\frac{2^{\Delta x}-1}{\Delta x}.\end{aligned}$$

So if we assume that the limit

$$\lim_{\Delta x \to 0} \frac{2^{\Delta x} - 1}{\Delta x} = C$$

exists then we have

$$\frac{d2^x}{dx} = C2^x. \tag{37}$$

On your calculator you can compute $\frac{2^{\Delta x}-1}{\Delta x}$ for smaller and smaller values of Δx, which leads you to suspect that the limit actually exists, and that $C \approx 0.693\,147\,\ldots$. One can in fact prove that the limit exists, but we will not do this here.

Once we know (37) we can compute the derivative of a^x for any other positive number a. To do this we write $a = 2^{\log_2 a}$, and hence

$$a^x = \left(2^{\log_2 a}\right)^x = 2^{x \cdot \log_2 a}.$$

By the chain rule we therefore get

$$\begin{aligned} \frac{da^x}{dx} &= \frac{d2^{x \cdot \log_2 a}}{dx} \\ &= C\, 2^{x \cdot \log_2 a}\, \frac{dx \cdot \log_2 a}{dx} \\ &= (C \log_2 a)\, 2^{x \cdot \log_2 a} \\ &= (C \log_2 a)\, a^x. \end{aligned}$$

So the derivative of a^x is just some constant times a^x, the constant being $C \log_2 a$. This is essentially our formula for the derivative of a^x, but one can make the formula look nicer by introducing a special number, namely, we define

$$e = 2^{1/C} \text{ where } C = \lim_{\Delta x \to 0} \frac{2^{\Delta x} - 1}{\Delta x}.$$

One has

$$e \approx 2.718\,281\,818\,459\,\cdots$$

This number is special because if you set $a = e$, then

$$C \log_2 a = C \log_2 e = C \log_2 2^{1/C} = C \cdot \frac{1}{C} = 1,$$

and therefore the derivative of the function $y = e^x$ is

$$\boxed{\frac{de^x}{dx} = e^x.} \tag{38}$$

Read that again: the function e^x is its own derivative!

The logarithm with base e is called the ***Natural Logarithm***, and is written

$$\ln x = \log_e x.$$

Thus we have

$$\boxed{e^{\ln x} = x \qquad \ln e^x = x} \tag{39}$$

where the second formula holds for all real numbers x but the first one only makes sense for $x > 0$.

For any positive number a we have $a = e^{\ln a}$, and also

$$a^x = e^{x \ln a}.$$

By the chain rule you then get

$$\boxed{\frac{da^x}{dx} = a^x \ln a.} \tag{40}$$

47. Derivatives of Logarithms

Since the natural logarithm is the inverse function of $f(x) = e^x$ we can find its derivative by implicit differentiation. Here is the computation (which you should do yourself)

The function $f(x) = \log_a x$ satisfies

$$a^{f(x)} = x$$

Differentiate both sides, and use the chain rule on the left,

$$(\ln a)a^{f(x)} f'(x) = 1.$$

Then solve for $f'(x)$ to get

$$f'(x) = \frac{1}{(\ln a)a^{f(x)}}.$$

Finally we remember that $a^{f(x)} = x$ which gives us the derivative of a^x

$$\frac{da^x}{dx} = \frac{1}{x \ln a}.$$

In particular, the natural logarithm has a very simple derivative, namely, since $\ln e = 1$ we have

$$\boxed{\frac{d \ln x}{dx} = \frac{1}{x}.} \tag{41}$$

48. Limits involving exponentials and logarithms

Theorem 48.1. *Let r be any real number. Then, if $a > 1$,*

$$\lim_{x\to\infty} x^r a^{-x} = 0,$$

i.e.

$$\lim_{x\to\infty} \frac{x^r}{a^x} = 0.$$

This theorem says that *any exponential will beat any power of x as $x \to \infty$.* For instance, as $x \to \infty$ both x^{1000} and $(1.001)^x$ go to infinity, but

$$\lim_{x\to\infty} \frac{x^{1000}}{(1.001)^x} = 0,$$

so, in the long run, for very large x, 1.001^x will be much larger than 1000^x.

Proof when $a = e$. We want to show $\lim_{x\to\infty} x^r e^{-x} = 0$. To do this consider the function $f(x) = x^{r+1}e^{-x}$. Its derivative is

$$f'(x) = \frac{dx^{r+1}e^{-x}}{dx} = \big((r+1)x^r - x^{r+1}\big)e^{-x} = (r+1-x)x^r e^{-x}.$$

Therefore $f'(x) < 0$ for $x > r+1$, i.e. $f(x)$ is decreasing for $x > r+1$. It follows that $f(x) < f(r+1)$ for all $x > r+1$, i.e.

$$x^{r+1}e^{-x} < (r+1)^{r+1}e^{-(r+1)} \text{ for } x > r+1.$$

Divide by x, abbreviate $A = (r+1)^{r+1}e^{-(r+1)}$, and we get

$$0 < x^r e^{-x} < \frac{A}{x} \text{ for all } x > r+1.$$

The Sandwich Theorem implies that $\lim_{x\to\infty} x^r e^{-x} = 0$, which is what we had promised to show.

□

Here are some related limits:

$$\begin{aligned} a>1 &\implies \lim_{x\to\infty}\frac{a^x}{x^r}=\infty \quad (D.N.E.) \\ m>0 &\implies \lim_{x\to\infty}\frac{\ln x}{x^m}=0 \\ m>0 &\implies \lim_{x\to 0} x^m\ln x=0 \end{aligned}$$

The second limit says that even though $\ln x$ becomes infinitely large as $x\to\infty$, it is always much less than any power x^m with $m>0$ real. To prove it you set $x=e^t$ and then $t=s/m$, which leads to

$$\lim_{x\to\infty}\frac{\ln x}{x^m} \overset{x=e^t}{=} \lim_{t\to\infty}\frac{t}{e^{mt}} \overset{t=s/m}{=} \frac{1}{m}\lim_{t\to\infty}\frac{s}{e^s}=0.$$

The third limit follows from the second by substituting $x=1/y$ and using $\ln\frac{1}{x}=-\ln x$.

49. **Exponential growth and decay**

A quantity X which depends on time t is said to grow or decay exponentially if it is given by

$$X(t)=X_0e^{kt}. \tag{42}$$

The constant X_0 is the value of $X(t)$ at time $t=0$ (sometimes called "the initial value of X").

The derivative of an exponentially growing quantity, i.e. its rate of change with time, is given by $X'(t)=X_0\,ke^{kt}$ so that

$$\frac{dX(t)}{dt}=kX(t). \tag{43}$$

In words, *for an exponentially growing quantity the rate of change is always proportional to the quantity itself.* The proportionality constant is k and is sometimes called "the relative growth rate."

This property of exponential functions completely describes them, by which I mean that any function which satisfies (43) automatically satisfies (42). To see that this is true, suppose you have a function $X(t)$ for which $X'(t)=kX(t)$ holds at all times t. Then

$$\begin{aligned} \frac{dX(t)e^{-kt}}{dt} &= X(t)\frac{de^{-kt}}{dt}+\frac{dX(t)}{dt}e^{-kt} \\ &= -kX(t)e^{-kt}+X'(t)e^{-kt} \\ &= (X'(t)-kX(t))e^{-kt} \\ &= 0. \end{aligned}$$

It follows that $X(t)e^{-kt}$ does not depend on t. At $t=0$ one has

$$X(t)e^{-kt}=X(0)e^0=X_0$$

and therefore we have

$$X(t)e^{-kt}=X_0 \text{ for all } t.$$

Multiply with e^{kt} and we end up with

$$X(t)=X_0e^{kt}.$$

49.1. *Half time and doubling time*

If $X(t) = X_0 e^{kt}$ then one has

$$X(t+T) = X_0 e^{kt+kT} = X_0 e^{kt} e^{kT} = e^{kT} X(t).$$

In words, after time T goes by an exponentially growing (decaying) quantity changes by a factor e^{kT}. If $k > 0$, so that the quantity is actually growing, then one calls

$$T = \frac{\ln 2}{k}$$

the ***doubling time*** for X because $X(t)$ changes by a factor $e^{kT} = e^{\ln 2} = 2$ every T time units: $X(t)$ doubles every T time units.

If $k < 0$ then $X(t)$ is decaying and one calls

$$T = \frac{\ln 2}{-k}$$

the ***half life*** because $X(t)$ is reduced by a factor $e^{kT} = e^{-\ln 2} = \frac{1}{2}$ every T time units.

49.2. *Determining X_0 and k*

The general exponential growth/decay function (42) contains only two constants, X_0 and k, and if you know the values of $X(t)$ at two different times then you can compute these constants.

Suppose that you know

$$X_1 = X(t_1) \text{ and } X_2 = X(t_2).$$

Then we have

$$X_0 e^{kt_1} = X_1 \text{ and } X_2 = X_0 e^{kt_2}$$

in which t_1, t_2, X_1, X_2 are given and k and X_0 are unknown. One first finds k from

$$\frac{X_1}{X_2} = \frac{X_0 e^{kt_1}}{X_0 e^{kt_2}} = e^{k(t_1 - t_2)} \implies \ln \frac{X_1}{X_2} = k(t-1-t_2)$$

which implies

$$k = \frac{\ln X_1 - \ln X_2}{t_1 - t_2}.$$

Once you have computed k you can find X_0 from

$$X_0 = \frac{X_1}{e^{kt_1}} = \frac{X_2}{e^{kt_2}}.$$

(both expressions should give the same result.)

Exercises

49.1 – Sketch the graphs of the following functions.

(**a**) $y = e^x$ (**b**) $y = e^{-x}$

(**c**) $y = e^x + e^{-2x}$ (**d**) $y = e^{3x} - 4e^x$

(**e**) $y = \dfrac{e^x}{1+e^x}$ (**f**) $y = \dfrac{2e^x}{1+e^{2x}}$

(**g**) $y = xe^{-x}$ (**h**) $y = \sqrt{x}e^{-x/4}$

(**i**) $y = x^2 e^{x+2}$ (**j**) $y = e^{x/2} - x$

Hint for some of these: if you have to solve something like $e^{4x} - 3e^{3x} + e^x = 0$, then call $w = e^x$, and you get a polynomial equation for w, namely $w^4 - 3w^3 + w = 0$.

49.2 – Sketch the graphs of the following functions.

(**a**) $y = \ln\sqrt{x}$

(**b**) $y = \ln\dfrac{1}{x}$

(**c**) $y = x\ln x$

(**d**) $y = \dfrac{-1}{\ln x}$ $\quad(0 < x < \infty, x \neq 1)$

(**e**) $y = (\ln x)^2$ $\quad(x > 0)$

(**f**) $y = \dfrac{\ln x}{x}$ $\quad(x > 0)$

(**g**) $y = \ln\sqrt{\dfrac{1+x}{1-x}}$ $\quad(|x| < 1)$

(**h**) $y = \ln\bigl(1 + x^2\bigr)$

(**i**) $y = \ln\bigl(x^2 - 3x + 2\bigr)$ $\quad(x > 2)$

(**j**) $y = \ln\cos x$ $\quad(|x| < \frac{\pi}{2})$

49.3 – The function $f(x) = e^{-x^2}$ plays a central in statistics and its graph is called ***the bell curve*** (because of its shape). Sketch the graph of f.

49.4 – Sketch the part of the graph of the function

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

with $x > 0$.

Find the limits

$$\lim_{x\searrow 0} \frac{f(x)}{x^n} \text{ and } \lim_{x\to\infty} f(x)$$

where n can be any positive integer (hint: substitute $x = \ldots$?)

49.5 – A ***damped oscillation*** is a function of the form

$$f(x) = e^{-ax}\cos bx \text{ or } f(x) = e^{-ax}\sin bx$$

where a and b are constants.

Sketch the graph of $f(x) = e^{-x}\sin 10x$ (i.e. find zeroes, local max and mins, inflection points) and draw (with pencil on paper) the piece of the graph with $0 \leq x \leq 2\pi$.

49.6 – Find the inflection points on the graph of $f(x) = (1 + x)\ln x$ $(x > 0)$.

49.7 – (**a**) If x is large, which is bigger: 2^x or x^2?

(**b**) The graphs of $f(x) = x^2$ and $g(x) = 2^x$ intersect at $x = 2$ (since $2^2 = 2^2$). How many more intersections do these graphs have (with $-\infty < x < \infty$)?

49.8 – Find the following limits

(**a**) $\displaystyle\lim_{x\to\infty} \frac{e^x - 1}{e^x + 1}$

(**b**) $\displaystyle\lim_{x\to\infty} \frac{e^x - x^2}{e^x + x}$

(**c**) $\displaystyle\lim_{x\to\infty} \frac{2^x}{3^x - 2^x}$

(**d**) $\displaystyle\lim_{x\to\infty} \frac{e^x - x^2}{e^{2x} + e^{-x}}$

(**e**) $\displaystyle\lim_{x\to\infty} \frac{e^{-x} - e^{-x/2}}{\sqrt{e^x + 1}}$

(**f**) $\displaystyle\lim_{x\to\infty} \frac{\sqrt{x + e^{4x}}}{e^{2x} + x}$

(**g**) $\displaystyle\lim_{x\to\infty} \frac{e^{\sqrt{x}}}{\sqrt{e^x + 1}}$

(**h**) $\displaystyle\lim_{x\to\infty} \ln(1 + x) - \ln x$

(**i**) $\displaystyle\lim_{x\to\infty} \frac{\ln x}{\ln x^2}$

(**j**) $\displaystyle\lim_{x\to 0} x\ln x$

(**k**) $\displaystyle\lim_{x\to\infty} \frac{\ln x}{\sqrt{x} + \ln x}$

(**l**) $\displaystyle\lim_{x\to 0} \frac{\ln x}{\sqrt{x} + \ln x}$

49.9 – Find the tenth derivative of xe^x.

49.10 – For which real number x is $2^x - 3^x$ the largest?

49.11 – Find $\dfrac{dx^x}{dx}$, $\dfrac{dx^{x^x}}{dx}$, and $\dfrac{d(x^x)^x}{dx}$. Hint: $x^x = e^?$.

49.12 About logarithmic differentiation:

(**a**) Let $y = (x+1)^2(x+3)^4(x+5)^6$ and $u = \ln y$. Find du/dx. Hint: Use the fact that ln converts multiplication to addition before you differentiate. It will simplify the calculation.

(**b**) Check that the derivative of $\ln u(x)$ is the logarithmic derivative of the function u (as defined in the exercises following §25, chapter 4.)

49.13 – After 3 days a sample of radon-222 decayed to 58% of its original amount.
(a) What is the half life of radon-222?
(b) How long would it take the sample to decay to 10% of its original amount?

49.14 – Polonium-210 has a half life of 140 days.
(a) If a sample has a mass of 200 mg find a formula for the mass that remains after t days.
(b) Find the mass after 100 days.
(c) When will the mass be reduced to 10 mg?
(d) Sketch the graph of the mass as a function of time.

49.15 – Current agricultural experts believe that the world's farms can feed about 10 billion people. The 1950 world population was 2.517 billion and the 1992 world population was 5.4 billion. When can we expect to run out of food?

49.16 – The Archer Daniel Midlands company runs two ads on Sunday mornings. One says that "when this baby is old enough to vote, the world will have one billion new mouths to feed" and the other says "in thirty six years, the world will have to set eight billion places at the table." What does ADM think the population of the world is at present? How fast does ADM think the population is increasing? Use units of billions of people so you can write 8 instead of $8,000,000,000$. (Hint: $36 = 2 \times 18$.)

49.17 – The population of California grows exponentially at an instantaneous rate of 2% per year. The population of California on January 1, 2000 was 20,000,000.

(**a**) Write a formula for the population $N(t)$ of California t years after January 1, 2000.

(**b**) Each Californian consumes pizzas at the rate of 70 pizzas per year. At what rate is California consuming pizzas t years after 1990?

(**c**) How many pizzas were consumed in California from January 1, 2005 to January 1, 2009?

49.18 – The population of the country of Slobia grows exponentially.

(**a**) If its population in the year 1980 was 1,980,000 and its population in the year 1990 was 1,990,000, what is its population in the year 2000?

(**b**) How long will it take the population to double? (Your answer may be expressed in terms of exponentials and natural logarithms.)

49.19 – The *hyperbolic functions* are defined by

$$\sinh x = \frac{e^x - e^{-x}}{2}, \quad \cosh x = \frac{e^x + e^{-x}}{2}, \quad \tanh x = \frac{\sinh x}{\cosh x}.$$

(**a**) Prove the following identities

$$\cosh^2 x - \sinh^2 x = 1$$
$$\cosh 2x = \cosh^2 x + \sinh^2 x$$
$$\sinh 2x = 2 \sinh x \cosh x.$$

(**b**) Show that

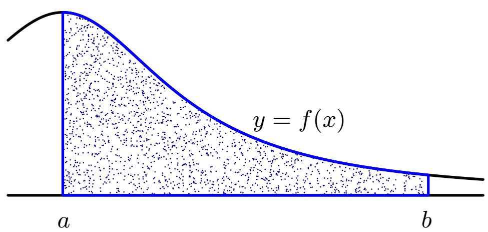
$$\frac{d\sinh x}{dx} = \cosh x, \quad \frac{d\cosh x}{dx} = \sinh x, \quad \frac{d\tanh x}{dx} = \frac{1}{\cosh^2 x}.$$

(**c**) Sketch the graphs of the three hyperbolic functions.

VII. The Integral

In this chapter we define the integral of a function on some interval $[a, b]$. The most common interpretation of the integral is in terms of the area under the graph of the given function, so that is where we begin.

50. Area under a Graph



Let f be a function which is defined on some interval $a \le x \le b$ and assume it is positive, i.e. assume that its graph lies above the x axis. *How large is the area of the region caught between the x axis, the graph of $y = f(x)$ and the vertical lines $y = a$ and $y = b$?*

One can try to compute this area by approximating the region with many thin rectangles. To do this you choose a *partition* of the interval $[a, b]$, i.e. you pick numbers $x_1 < \cdots < x_n$ with

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b.$$

These numbers split the interval $[a, b]$ into n sub-intervals

$$[x_0, x_1], \quad [x_1, x_2], \quad \ldots, \quad [x_{n-1}, x_n]$$

whose lengths are

$$\Delta x_1 = x_1 - x_0, \quad \Delta x_2 = x_2 - x_1, \quad \ldots, \quad \Delta x_n = x_n - x_{n-1}.$$

In each interval we choose a point c_k, i.e. in the first interval we choose $x_0 \le c_1 \le x_1$, in the second interval we choose $x_1 \le c_2 \le x_2$, ..., and in the last interval we choose some number $x_{n-1} \le c_n \le x_n$. See figure 29.

We then define n rectangles: the base of the k^{th} rectangle is the interval $[x_{k-1}, x_k]$ on the x-axis, while its height is $f(c_k)$ (here k can be any integer from 1 to n.)

The area of the k^{th} rectangle is of course the product of its height and width, i.e. its area is $f(c_k)\Delta x_k$. Adding these we see that the total area of the rectangles is

$$R = f(c_1)\Delta x_1 + f(c_2)\Delta x_2 + \cdots + f(c_n)\Delta x_n. \tag{44}$$

This kind of sum is called a ***Riemann sum.***

If the partition is sufficiently fine then one would expect this sum, i.e. the total area of all rectangles to be a good approximation of the area of the region under the graph. Replacing the partition by a finer partition, with more division points, should improve the approximation. So you would expect the area to be the limit of Riemann-sums like R "as the partition becomes finer and finer." A precise formulation of the definition goes like this:

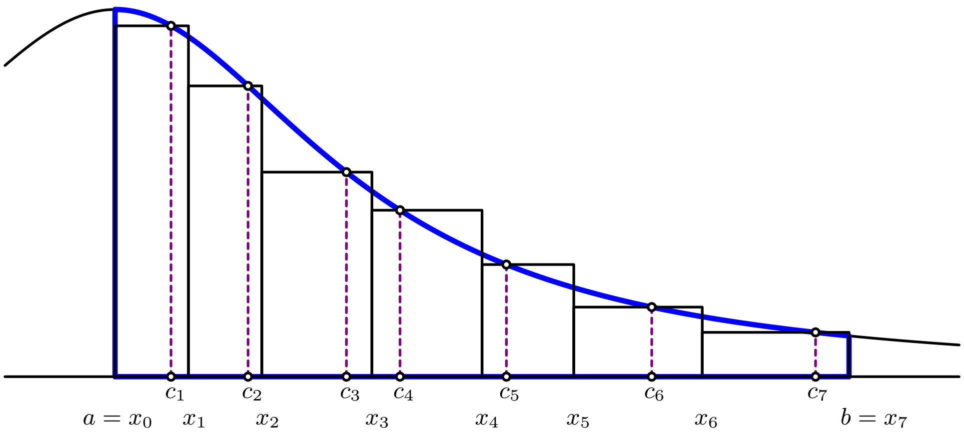


Figure 29. Forming a Riemann sum

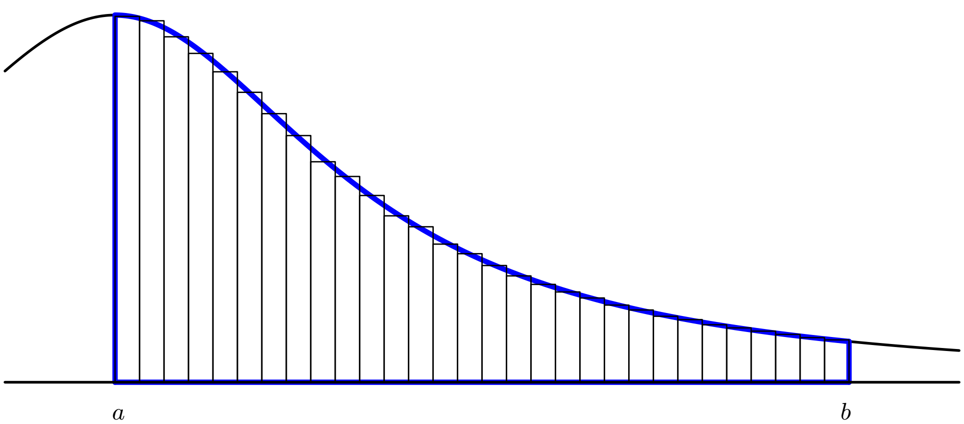


Figure 30. Refining the partition.

Definition. 50.1. *If f is a function defined on an interval $[a,b]$, then we say that*

$$\int_a^b f(x)dx = I,$$

i.e. the integral of "$f(x)$ from $x = a$ to b" equals I, if for every $\varepsilon > 0$ one can find a $\delta > 0$ such that

$$\Big| f(c_1)\Delta x_1 + f(c_2)\Delta x_2 + \cdots + f(c_n)\Delta x_n - I \Big| < \varepsilon$$

holds for every partition all of whose intervals have length $\Delta x_k < \delta$.

51. When f changes its sign

If the function f is not necessarily positive everywhere in the interval $a \leq x \leq b$, then we still define the integral in exactly the same way: as a limit of Riemann sums whose mesh size becomes smaller and smaller. However the interpretation of the integral as "the area of the region between the graph and the x-axis" has a twist to it.

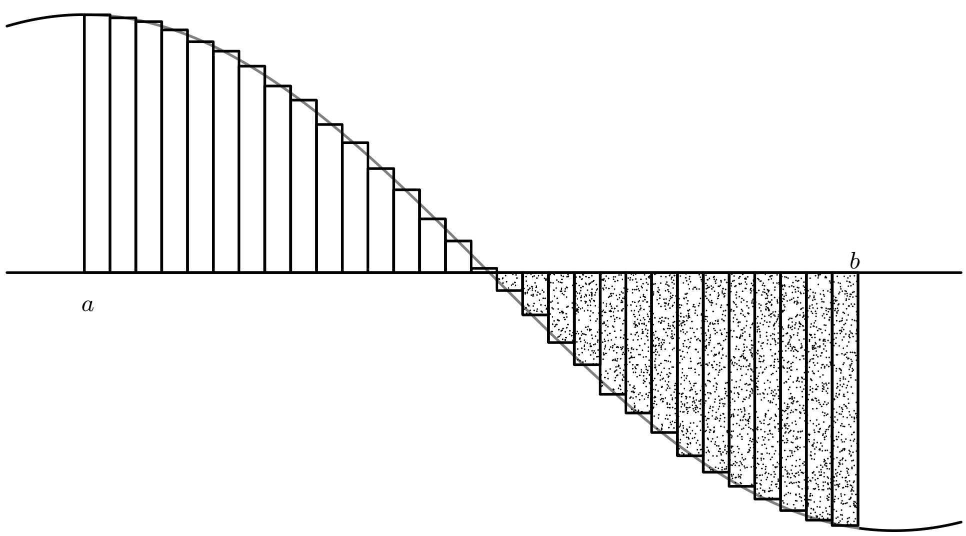


Figure 31. Illustrating a Riemann sum for a function whose sign changes

Let f be some function on an interval $a \leq x \leq b$, and form the Riemann sum

$$R = f(c_1)\Delta x_1 + f(c_2)\Delta x_2 + \cdots + f(c_n)\Delta x_n$$

that goes with some partition, and some choice of c_k.

When f can be positive or negative, then the terms in the Riemann sum can also be positive or negative. If $f(c_k) > 0$ then the quantity $f(c_k)\Delta x_k$ is the area of the corresponding rectangle, but if $f(c_k) < 0$ then $f(c_k)\Delta x_k$ is a negative number, namely *minus* the area of the corresponding rectangle. The Riemann sum is therefore the area of the rectangles above the x-axis *minus* the area below the axis and above the graph.

Taking the limit over finer and finer partitions, we conclude that

$$\int_a^b f(x)dx = \begin{array}{c}\text{area above the } x\text{-axis, below the graph} \\ \textbf{\textit{minus}} \text{ the area below the } x\text{-axis, above the graph.}\end{array}$$

52. The Fundamental Theorem of Calculus

Definition. 52.1. *A function F is called an **antiderivative** of f on the interval $[a, b]$ if one has $F'(x) = f(x)$ for all x with $a < x < b$.*

For instance, $F(x) = \frac{1}{2}x^2$ is an antiderivative of $f(x) = x$, but so is $G(x) = \frac{1}{2}x^2 + 2008$.

Theorem 52.2. *If f is a function whose integral $\int_a^b f(x)dx$ exists, and if F is an antiderivative of f on the interval $[a, b]$, then one has*

$$\int_a^b f(x)dx = F(b) - F(a). \tag{45}$$

(a proof was given in lecture.)

Because of this theorem the expression on the right appears so often that various abbreviations have been invented. We will abbreviate

$$F(b) - F(a) \stackrel{\text{def}}{=} \big[F(x)\big]_{x=a}^{b} = \big[F(x)\big]_{a}^{b}.$$

52.1. *Terminology*

In the integral

$$\int_a^b f(x)\, dx$$

the numbers a and b are called the *bounds of the integral*, the function $f(x)$ which is being integrated is called *the integrand*, and the variable x is *integration variable.*

The integration variable is a *dummy variable.* If you systematically replace it with another variable, the resulting integral will still be the same. For instance,

$$\int_0^1 x^2\, dx = \big[\tfrac{1}{3}x^3\big]_{x=0}^{1} = \tfrac{1}{3},$$

and if you replace x by φ you still get

$$\int_0^1 \varphi^2\, d\varphi = \big[\tfrac{1}{3}\varphi^3\big]_{\varphi=0}^{1} = \tfrac{1}{3}.$$

Another way to appreciate that the integration variable is a dummy variable is to look at the Fundamental Theorem again:

$$\int_a^b f(x)\, dx = F(b) - F(a).$$

The right hand side tells you that the value of the integral depends on a and b, and does not anything to do with the variable x.

Exercises

52.1 – Find an antiderivative $F(x)$ for each of the following functions $f(x)$. Finding antiderivatives involves a fair amount of guess work, but with experience it gets easier to guess antiderivatives.

(**a**) $f(x) = 2x + 1$ (**b**) $f(x) = 1 - 3x$

(**c**) $f(x) = x^2 - x + 11$ (**d**) $f(x) = x^4 - x^2$

(**e**) $f(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4}$ (**f**) $f(x) = \frac{1}{x}$

(**g**) $f(x) = e^x$ (**h**) $f(x) = \frac{2}{x}$

(**i**) $f(x) = e^{2x}$ (**j**) $f(x) = \frac{1}{2+x}$

(**k**) $f(x) = \frac{e^x - e^{-x}}{2}$ (**l**) $f(x) = \frac{1}{1+x^2}$

(**m**) $f(x) = \frac{e^x + e^{-x}}{2}$ (**n**) $f(x) = \frac{1}{\sqrt{1-x^2}}$

(**o**) $f(x) = \sin x$ (**p**) $f(x) = \frac{2}{1-x}$

(**q**) $f(x) = \cos x$ (**r**) $f(x) = \cos 2x$

(**s**) $f(x) = \sin(x - \pi/3)$ (**t**) $f(x) = \sin x + \sin 2x$

(**u**) $f(x) = \sin(x) + \frac{1}{x}$ (**v**) $f(x) = 2x(1 + x^2)^5$

In each of the following exercises you should compute the area of the indicated region, and also of the smallest enclosing rectangle with horizontal and vertical sides.

Before computing anything draw the region.

52.2 – The region between the vertical lines $x = 0$ and $x = 1$, and between the x-axis and the graph of $y = x^3$.

52.3 – The region between the vertical lines $x = 0$ and $x = 1$, and between the x-axis and the graph of $y = x^n$ (here $n > 0$, draw for $n = \frac{1}{2}, 1, 2, 3, 4$).

52.4 – The region above the graph of $y = \sqrt{x}$, below the line $y = 2$, and between the vertical lines $x = 0$, $x = 4$.

52.5 – The region above the x-axis and below the graph of $f(x) = x^2 - x^3$.

52.6 – The region above the x-axis and below the graph of $f(x) = 4x^2 - x^4$.

52.7 – The region above the x-axis and below the graph of $f(x) = 1 - x^4$.

52.8 – The region above the x-axis, below the graph of $f(x) = \sin x$, and between $x = 0$ and $x = \pi$.

52.9 – The region above the x-axis, below the graph of $f(x) = 1/(1 + x^2)$ (a curve known as *Maria Agnesi's witch*), and between $x = 0$ and $x = 1$.

52.10 – The region between the graph of $y = 1/x$ and the x-axis, and between $x = a$ and $x = b$ (here $0 < a < b$ are constants, e.g. choose $a = 1$ and $b = \sqrt{2}$ if you have something against either letter a or b.)

52.11 – The region above the x-axis and below the graph of

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

52.12 – Compute

$$\int_0^1 \sqrt{1-x^2}dx$$

without find an antiderivative for $\sqrt{1-x^2}$ (you can find such an antiderivative, but it's not easy. This integral is the area of some region: which region is it, and what is that area?)

52.13 – Compute

$$(\mathbf{a}) \int_0^{1/2} \sqrt{1-x^2}dx \qquad (\mathbf{b}) \int_{-1}^{1} |1-x|dx \qquad (\mathbf{c}) \int_{-1}^{1} |2-x|dx$$

without finding antiderivatives.

53. The indefinite integral

The fundamental theorem tells us that in order to compute the integral of some function f over an interval $[a,b]$ you should first find an antiderivative F of f. In practice, much of the effort required to find an integral goes into finding the antiderivative. In order to simplify the computation of the integral

$$\int_a^b f(x)dx = F(b) - F(a) \tag{46}$$

the following notation is commonly used for the antiderivative:

$$F(x) = \int f(x)dx. \tag{47}$$

For instance,

$$\int x^2\, dx = \frac{1}{3}x^3, \quad \int \sin 5x\, dx = -\tfrac{1}{5}\cos 5x, \quad \text{etc}\ldots$$

The integral which appears here does not have the integration bounds a and b. It is called an ***indefinite integral***, as opposed to the integral in (46) which is called a ***definite integral***. You use the indefinite integral if you expect the computation of the antiderivative to be a lengthy affair, and you do not want to write the integration bounds a and b all the time.

It is important to distinguish between the two kinds of integrals. Here is a list of differences:

INDEFINITE INTEGRAL	DEFINITE INTEGRAL
$\int f(x)dx$ is a function of x.	$\int_a^b f(x)dx$ is a number.
By definition $\int f(x)dx$ is *any function of x whose derivative is $f(x)$*.	$\int_a^b f(x)dx$ was defined in terms of Riemann sums and can be interpreted as "area under the graph of $y = f(x)$", at least when $f(x) > 0$.
x is not a dummy variable, for example, $\int 2xdx = x^2+C$ and $\int 2tdt = t^2+C$ are functions of diffferent variables, so they are not equal.	x is a dummy variable, for example, $\int_0^1 2xdx = 1$, and $\int_0^1 2tdt = 1$, so $\int_0^1 2xdx = \int_0^1 2tdt$.

53.1. *You can always check the answer*

Suppose you want to find an antiderivative of a given function $f(x)$ and after a long and messy computation which you don't really trust you get an "answer", $F(x)$. You can then throw away the dubious computation and differentiate the $F(x)$ you had found. If $F'(x)$ turns out to be equal to $f(x)$, then your $F(x)$ is indeed an antiderivative and your computation isn't important anymore.

For example, suppose that we want to find $\int \ln x\,dx$. My cousin Louie says it might be $F(x) = x\ln x - x$. Let's see if he's right:

$$\frac{d}{dx}\big(x\ln x - x\big) = x\cdot\frac{1}{x} + 1\cdot\ln x - 1 = \ln x.$$

Who knows how Louie thought of this[7], but it doesn't matter: he's right! We now know that $\int \ln x dx = x\ln x - x + C$.

53.2. *About "+C"*

Let $f(x)$ be a function defined on some interval $a \le x \le b$. If $F(x)$ is an antiderivative of $f(x)$ on this interval, then for any constant C the function $\tilde{F}(x) = F(x)+C$ will also be an antiderivative of $f(x)$. So one given function $f(x)$ has many different antiderivatives, obtained by adding different constants to one given antiderivative.

Theorem 53.1. *If $F_1(x)$ and $F_2(x)$ are antiderivatives of the same function $f(x)$ on some interval $a \le x \le b$, then there is a constant C such that $F_1(x) = F_2(x) + C$.*

Proof. Consider the difference $G(x) = F_1(x) - F_2(x)$. Then $G'(x) = F_1'(x) - F_2'(x) = f(x) - f(x) = 0$, so that $G(x)$ must be constant. Hence $F_1(x) - F_2(x) = C$ for some constant. □

It follows that there is some ambiguity in the notation $\int f(x)\,dx$. Two functions $F_1(x)$ and $F_2(x)$ can both equal $\int f(x)\,dx$ without equaling each other. When this happens, they (F_1 and F_2) differ by a constant. This can sometimes lead to confusing situations, e.g. you can check that

$$\int 2\sin x\cos x\,dx = \sin^2 x$$
$$\int 2\sin x\cos x\,dx = -\cos^2 x$$

are both correct. (Just differentiate the two functions $\sin^2 x$ and $-\cos^2 x$!) These two answers look different until you realize that because of the trig identity $\sin^2 x + \cos^2 x = 1$ they really only differ by a constant: $\sin^2 x = -\cos^2 x + 1$.

To avoid this kind of confusion we will from now on never forget to include the "arbitrary constant $+C$" in our answer when we compute an antiderivative.

---

[7]He took math 222 and learned to integrate by parts.

53.3. *Standard Integrals*

Here is a list of the standard integrals everyone should know.

$$\int f(x)\,dx = F(x) + C$$

$$\int x^n\,dx = \frac{x^{n+1}}{n+1} + C \qquad \text{for all } n \neq -1$$

$$\int \frac{1}{x}\,dx = \ln|x| + C \qquad \text{(Note the absolute values)}$$

$$\int e^x\,dx = e^x + C$$

$$\int a^x\,dx = \frac{a^x}{\ln a} + C \qquad \text{(don't memorize: use } a^x = e^{x\ln a}\text{)}$$

$$\int \sin x\,dx = -\cos x + C$$

$$\int \cos x\,dx = \sin x + C$$

$$\int \tan x\,dx = -\ln|\cos x| + C \qquad \text{(Note the absolute values)}$$

$$\int \frac{1}{1+x^2}\,dx = \arctan x + C$$

$$\int \frac{1}{\sqrt{1-x^2}}\,dx = \arcsin x + C$$

The following integral is also useful, but not as important as the ones above:

$$\int \frac{dx}{\cos x} = \frac{1}{2}\ln\frac{1+\sin x}{1-\sin x} + C \text{ for } -\frac{\pi}{2} < x < \frac{\pi}{2}.$$

All of these integrals should be familiar from the differentiation rules we have learned so far, except for for the integrals of $\tan x$ and of $\frac{1}{\cos x}$. You can check those by differentiation (using $\ln\frac{a}{b} = \ln a - \ln b$ simplifies things a bit).

54. **Properties of the Integral**

Just as we had a list of properties for the limits and derivatives of sums and products of functions, the integral has similar properties.

Suppose we have two functions $f(x)$ and $g(x)$ with antiderivatives $F(x)$ and $G(x)$, respectively. Then we know that

$$\frac{d}{dx}\{F(x) + G(x)\} = F'(x) + G'(x) = f(x) + g(x),$$

in words, $F + G$ is an antiderivative of $f + g$, which we can write as

$$\int \{f(x) + g(x)\}\,dx = \int f(x)\,dx + \int g(x)\,dx. \tag{48}$$

Similarly, $\frac{d}{dx}\big(cF(x)\big) = cF'(x) = cf(x)$ implies that

$$\int cf(x)\,dx = c\int f(x)\,dx \tag{49}$$

if c is a constant.

These properties imply analogous properties for the definite integral. For any pair of functions on an interval $[a,b]$ one has

$$\int_a^b \big[f(x)+g(x)\big]dx = \int_a^b f(x)dx + \int_a^b g(x)dx, \tag{50}$$

and for any function f and constant c one has

$$\int_a^b cf(x)dx = c\int_a^b f(x)dx. \tag{51}$$

Definite integrals have one other property for which there is no analog in indefinite integrals: if you split the interval of integration into two parts, then the integral over the whole is the sum of the integrals over the parts. The following theorem says it more precisely.

Theorem 54.1. *Given $a < c < b$, and a function on the interval $[a,b]$ then*

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx. \tag{52}$$

Proof. Let F be an antiderivative of f. Then

$$\int_a^c f(x)dx = F(c) - F(a) \text{ and } \int_c^b f(x)dx = F(b) - F(a),$$

so that

$$\begin{aligned}\int_a^b f(x)dx &= F(b) - F(a)\\ &= F(b) - F(c) + F(c) - F(a)\\ &= \int_a^c f(x)dx + \int_c^b f(x)dx.\end{aligned}$$

□

So far we have always assumed theat $a < b$ in all indefinitie integrals $\int_a^b \ldots$. The fundamental theorem suggests that when $b < a$, we should define the integral as

$$\int_a^b f(x)dx = F(b) - F(a) = -\big(F(a) - F(b)\big) = -\int_b^a f(x)dx. \tag{53}$$

For instance,

$$\int_1^0 xdx = -\int_0^1 xdx = -\tfrac{1}{2}.$$

55. **The definite integral as a function of its integration bounds**

Consider the expression

$$I = \int_0^x t^2\, dt.$$

What does I depend on? To see this, you calculate the integral and you find

$$I = \left[\tfrac{1}{3}t^3\right]_0^x = \tfrac{1}{3}x^3 - \tfrac{1}{3}0^3 = \tfrac{1}{3}x^3.$$

So the integral depends on x. It does not depend on t, since t is a "dummy variable" (see §52.1 where we already discussed this point.)

In this way you can use integrals to define new functions. For instance, we could define

$$I(x) = \int_0^x t^2 dt,$$

which would be a roundabout way of defining the function $I(x) = x^3/3$. Again, since t is a dummy variable we can replace it by any other variable we like. Thus

$$I(x) = \int_0^x \alpha^2\, d\alpha$$

defines the same function (namely, $I(x) = \frac{1}{3}x^3$).

The previous example does not define a new function ($I(x) = x^3/3$). An example of a *new* function defined by an integral is the "error-function" from statistics. It is given by

$$\operatorname{erf}(x) \stackrel{\text{def}}{=} \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt, \tag{54}$$

so $\operatorname{erf}(x)$ is the area of the shaded region in figure 32. The integral in (54) cannot be com-

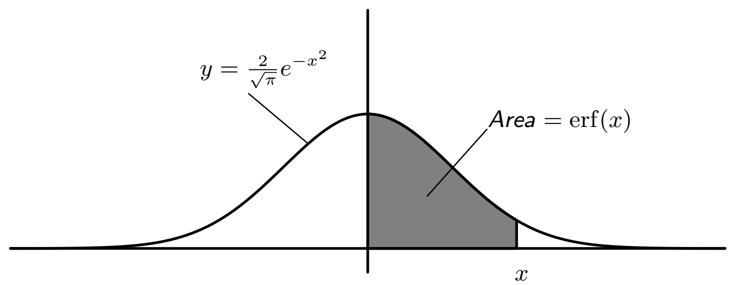


Figure 32. Definition of the Error function.

puted in terms of the standard functions (square and higher roots, sine, cosine, exponential and logarithms). Since the integral in (54) occurs very often in statistics (in relation with the so-called normal distribution) it has been given a name, namely, "erf(x)".

How do you differentiate a function that is defined by an integral? The answer is simple, for if $f(x) = F'(x)$ then the fundamental theorem says that

$$\int_a^x f(t)\, dt = F(x) - F(a),$$

and therefore

$$\frac{d}{dx}\int_a^x f(t)\, dt = \frac{d}{dx}\bigl\{F(x) - F(a)\bigr\} = F'(x) = f(x),$$

i.e.

$$\frac{d}{dx}\int_a^x f(t)\, dt = f(x).$$

A similar calculation gives you

$$\frac{d}{dx}\int_x^b f(t)\, dt = -f(x).$$

So what is the derivative of the error function? We have

$$\begin{aligned}
\operatorname{erf}'(x) &= \frac{d}{dx}\Bigl\{\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\, dt\Bigr\} \\
&= \frac{2}{\sqrt{\pi}}\frac{d}{dx}\int_0^x e^{-t^2}\, dt \\
&= \frac{2}{\sqrt{\pi}}e^{-x^2}.
\end{aligned}$$

56. **Method of substitution**

The chain rule says that

$$\frac{dF(G(x))}{dx} = F'(G(x)) \cdot G'(x),$$

so that

$$\int F'(G(x)) \cdot G'(x)\, dx = F(G(x)) + C.$$

56.1. *Example*

Consider the function $f(x) = 2x \sin(x^2 + 3)$. It does not appear in the list of standard antiderivatives we know by heart. But we do notice[8] that $2x = \frac{d}{dx}(x^2 + 3)$. So let's call $G(x) = x^2 + 3$, and $F(u) = -\cos u$, then

$$F(G(x)) = -\cos(x^2 + 3)$$

and

$$\frac{dF(G(x))}{dx} = \underbrace{\sin(x^2 + 3)}_{F'(G(x))} \cdot \underbrace{2x}_{G'(x)} = f(x),$$

so that

$$\int 2x \sin(x^2 + 3)\, dx = -\cos(x^2 + 3) + C. \tag{55}$$

56.2. *Leibniz' notation for substitution*

The most transparent way of computing an integral by substitution is by following Leibniz and introduce new variables. Thus to do the integral

$$\int f(G(x))G'(x)\, dx$$

where $f(u) = F'(u)$, we introduce the substitution $u = G(x)$, and agree to write

$$du = dG(x) = G'(x)\, dx.$$

Then we get

$$\int f(G(x))G'(x)\, dx = \int f(u)\, du = F(u) + C.$$

At the end of the integration we must remember that u really stands for $G(x)$, so that

$$\int f(G(x))G'(x)\, dx = F(u) + C = F(G(x)) + C.$$

As an example, let's do the integral (55) using Leibniz' notation. We want to find

$$\int 2x \sin(x^2 + 3)\, dx$$

and decide to substitute $z = x^2 + 3$ (the substitution variable doesn't always have to be called u). Then we compute

$$dz = d\big(x^2 + 3\big) = 2x\, dx \text{ and } \sin(x^2 + 3) = \sin z,$$

so that

$$\int 2x \sin(x^2 + 3)\, dx = \int \sin z\, dz = -\cos z + C.$$

Finally we get rid of the substitution variable z, and we find

$$\int 2x \sin(x^2 + 3)\, dx = -\cos\big(x^2 + 3\big) + C.$$

---

[8] You *will* start noticing things like this after doing several examples.

When we do integrals in this calculus class, we always get rid of the substitution variable because it is a variable we invented, and which does not appear in the original problem. But if you are doing an integral which appears in some longer discussion of a real-life (or real-lab) situation, then it may be that the substitution variable actually has a meaning (e.g. "the effective stoichiometric modality of CQF self-inhibition") in which case you may want to skip the last step and leave the integral in terms of the (meaningful) substitution variable.

56.3. *Substitution for definite integrals*

For definite integrals the chain rule

$$\frac{d}{dx}\bigl(F(G(x))\bigr) = F'(G(x))G'(x) = f(G(x))G'(x)$$

implies

$$\int_a^b f(G(x))G'(x)\,dx = F(G(b)) - F(G(a)).$$

which you can also write as

$$\int_{x=a}^{b} f(G(x))G'(x)\,dx = \int_{u=G(a)}^{G(b)} f(u)\,du. \tag{56}$$

56.4. *Example of substitution in a definite integral*

Let's compute

$$\int_0^1 \frac{x}{1+x^2}\,dx,$$

using the substitution $u = G(x) = 1 + x^2$. Since $du = 2x\,dx$, the associated ***indefinite*** integral is

$$\int \underbrace{\frac{1}{1+x^2}}_{\frac{1}{u}} \underbrace{x\,dx}_{\frac{1}{2}du} = \tfrac{1}{2}\int \frac{1}{u}\,du.$$

To find the definite integral you must compute the new integration bounds $G(0)$ and $G(1)$ (see equation (56).) If x runs between $x = 0$ and $x = 1$, then $u = G(x) = 1 + x^2$ runs between $u = 1 + 0^2 = 1$ and $u = 1 + 1^2 = 2$, so the definite integral we must compute is

$$\int_0^1 \frac{x}{1+x^2}\,dx = \tfrac{1}{2}\int_1^2 \frac{1}{u}\,du, \tag{57}$$

which is in our list of memorable integrals. So we find

$$\int_0^1 \frac{x}{1+x^2}\,dx = \tfrac{1}{2}\int_1^2 \frac{1}{u}\,du = \tfrac{1}{2}\bigl[\ln u\bigr]_1^2 = \tfrac{1}{2}\ln 2.$$

Sometimes the integrals in (57) are written as

$$\int_{x=0}^{1} \frac{x}{1+x^2}\,dx = \tfrac{1}{2}\int_{u=1}^{2} \frac{1}{u}\,du,$$

to emphasize (and remind yourself) to which variable the bounds in the integral refer.

Exercises

56.1 – Compute these derivatives

(**a**) $\dfrac{d}{dx}\displaystyle\int_0^x \left(1+t^2\right)^4\,dt$ (**b**) $\dfrac{d}{dx}\displaystyle\int_x^1 \ln z\,dz$

(**c**) $\dfrac{d}{ds}\displaystyle\int_s^0 \frac{dx}{1+x^2}$ (**d**) $\dfrac{d}{dx}\displaystyle\int_x^{2x} s^2 ds$

(**e**) $\dfrac{d}{d\theta}\displaystyle\int_0^{\sin\theta} \frac{dx}{1-x^2}$ (**f**) $\dfrac{d}{dt}\displaystyle\int_0^{t^2} e^{2x} dx$

56.2 – Compute the second derivative of the error function. How many inflection points does the graph of the error function have?

Compute (some of) the following integrals

56.3 – $\displaystyle\int \left\{6x^5 - 2x^{-4} - 7x + 3/x - 5 + 4e^x + 7^x\right\} dx$

56.4 – $\displaystyle\int (x/a + a/x + x^a + a^x + ax)\,dx$

56.5 – $\displaystyle\int \left\{\sqrt{x} - \sqrt[3]{x^4} + \frac{7}{\sqrt[3]{x^2}} - 6e^x + 1\right\} dx$

56.6 – $\displaystyle\int \left\{2^x + \left(\tfrac{1}{2}\right)^x\right\} dx$

56.7 – $\displaystyle\int_{-2}^{4} (3x-5)\,dx$

56.8 – $\displaystyle\int_1^4 x^{-2}\,dx$ (hm...)

56.9 – $\displaystyle\int_1^4 t^{-2}\,dt$ (**!**)

56.10 – $\displaystyle\int_1^4 x^{-2}\,dt$ (**!!!**)

56.11 – $\displaystyle\int_0^1 (1-2x-3x^2)\,dx$

56.12 – $\displaystyle\int_1^2 (5x^2-4x+3)\,dx$

56.13 – $\displaystyle\int_{-3}^{0} (5y^4-6y^2+14)\,dy$

56.14 – $\displaystyle\int_0^1 (y^9-2y^5+3y)\,dy$

56.15 – $\displaystyle\int_0^4 \sqrt{x}\,dx$

56.16 – $\displaystyle\int_0^1 x^{3/7}\,dx$

56.17 – $\displaystyle\int_1^3 \left(\frac{1}{t^2} - \frac{1}{t^4}\right) dt$

56.18 – $\displaystyle\int_1^2 \frac{t^6-t^2}{t^4}\,dt$

56.19 – $\displaystyle\int_1^2 \frac{x^2+1}{\sqrt{x}}\,dx$

56.20 – $\displaystyle\int_0^2 (x^3-1)^2\,dx$

56.21 – $\displaystyle\int_0^1 u(\sqrt{u}+\sqrt[3]{u})\,du$

56.22 – $\displaystyle\int_1^2 (x+1/x)^2\,dx$

56.23 – $\displaystyle\int_3^3 \sqrt{x^5+2}\,dx$

56.24 – $\displaystyle\int_1^{-1} (x-1)(3x+2)\,dx$

56.25 – $\displaystyle\int_1^4 (\sqrt{t} - 2/\sqrt{t})\,dt$

56.26 – $\displaystyle\int_1^8 \left(\sqrt[3]{r} + \frac{1}{\sqrt[3]{r}}\right) dr$

56.27 – $\displaystyle\int_{-1}^{0} (x+1)^3\,dx$

56.28 – $\displaystyle\int_{-5}^{-2} \frac{x^4-1}{x^2+1}\,dx$

56.29 – $\displaystyle\int_1^e \frac{x^2+x+1}{x}\,dx$

56.30 – $\displaystyle\int_4^9 \left(\sqrt{x} + \frac{1}{\sqrt{x}}\right)^2 dx$

56.31 – $\displaystyle\int_0^1 \left(\sqrt[4]{x^5} + \sqrt[5]{x^4}\right) dx$

56.32 – $\displaystyle\int_1^8 \frac{x-1}{\sqrt[3]{x^2}}\,dx$

56.33 – $\displaystyle\int_{\pi/4}^{\pi/3} \sin t\,dt$

56.34 – $\displaystyle\int_0^{\pi/2} (\cos\theta + 2\sin\theta)\,d\theta$

56.35 – $\displaystyle\int_0^{\pi/2} (\cos\theta + \sin 2\theta)\,d\theta$

56.36 – $\displaystyle\int_{2\pi/3}^{\pi} \frac{\tan x}{\cos x}\,dx$

56.37 – $\displaystyle\int_{\pi/3}^{\pi/2} \frac{\cot x}{\sin x}\,dx$

56.38 – $\displaystyle\int_1^{\sqrt{3}} \frac{6}{1+x^2}\,dx$

56.39 – $\displaystyle\int_0^{0.5} \frac{dx}{\sqrt{1-x^2}}$

56.40 – $\displaystyle\int_4^8 (1/x)\,dx$

56.41 – $\displaystyle\int_{\ln 3}^{\ln 6} 8e^x\,dx$

56.42 – $\displaystyle\int_8^9 2^t\,dt$

56.43 – $\displaystyle\int_{-e^2}^{-e} \frac{3}{x}\,dx$

56.44 – $\displaystyle\int_{-2}^{3} |x^2-1|\,dx$

56.45 – $\displaystyle\int_{-1}^{2} |x-x^2|\,dx$

56.46 – $\displaystyle\int_{-1}^{2} (x-2|x|)\,dx$

56.47 – $\displaystyle\int_0^2 (x^2-|x-1|)\,dx$

56.48 – $\displaystyle\int_0^2 f(x)\,dx$ where

$$f(x) = \begin{cases} x^4 & \text{if } 0 \le x < 1, \\ x^5, & \text{if } 1 \le x \le 2. \end{cases}$$

56.49 – $\displaystyle\int_{-\pi}^{\pi} f(x)\,dx$ where

$$f(x) = \begin{cases} x, & \text{if } -\pi \le x \le 0, \\ \sin x, & \text{if } 0 < x \le \pi. \end{cases}$$

56.50 – Compute

$$I = \int_0^2 2x\big(1+x^2\big)^3\,dx$$

in two different ways:
(**a**) Expand $(1+x^2)^3$, multiply with $2x$, and integrate each term.
(**b**) Use the substitution $u = 1 + x^2$.

56.51 – Compute

$$I_n = \int 2x\big(1+x^2\big)^n\,dx.$$

56.52 – If $f'(x) = x - 1/x^2$ and $f(1) = 1/2$ find $f(x)$.

56.53 – Sketch the graph of the curve $y = \sqrt{x+1}$ and determine the area of the region enclosed by the curve, the x-axis and the lines $x = 0$, $x = 4$.

56.54 – Find the area under the curve $y = \sqrt{6x+4}$ and above the x-axis between $x = 0$ and $x = 2$. Draw a sketch of the curve.

56.55 – Graph the curve $y = 2\sqrt{1-x^2}$, $x \in [0,1]$, and find the area enclosed between the curve and the x-axis. (Don't evaluate the integral, but compare with the area under the graph of $y = \sqrt{1-x^2}$.)

56.56 – Determine the area under the curve $y = \sqrt{a^2-x^2}$ and between the lines $x = 0$ and $x = a$.

56.57 – Graph the curve $y = 2\sqrt{9-x^2}$ and determine the area enclosed between the curve and the x-axis.

56.58 – Graph the area between the curve $y^2 = 4x$ and the line $x = 3$. Find the area of this region.

56.59 – Find the area bounded by the curve $y = 4 - x^2$ and the lines $y = 0$ and $y = 3$.

56.60 – Find the area enclosed between the curve $y = \sin 2x$, $0 \le x \le \pi/4$ and the axes.

56.61 – Find the area enclosed between the curve $y = \cos 2x$, $0 \le x \le \pi/4$ and the axes.

56.62 – Graph $y^2 + 1 = x$, and find the area enclosed by the curve and the line $x = 2$.

56.63 – Find the area of the region bounded by the parabola $y^2 = 4x$ and the line $y = 2x$.

56.64 – Find the area bounded by the curve $y = x(2 - x)$ and the line $x = 2y$.

56.65 – Find the area bounded by the curve $x^2 = 4y$ and the line $x = 4y - 2$.

56.66 – Calculate the area of the region bounded by the parabolas $y = x^2$ and $x = y^2$.

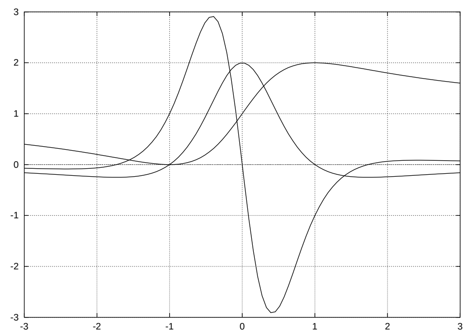
56.67 – Find the area of the region included between the parabola $y^2 = x$ and the line $x + y = 2$.

56.68 – Find the area of the region bounded by the curves $y = \sqrt{x}$ and $y = x$.

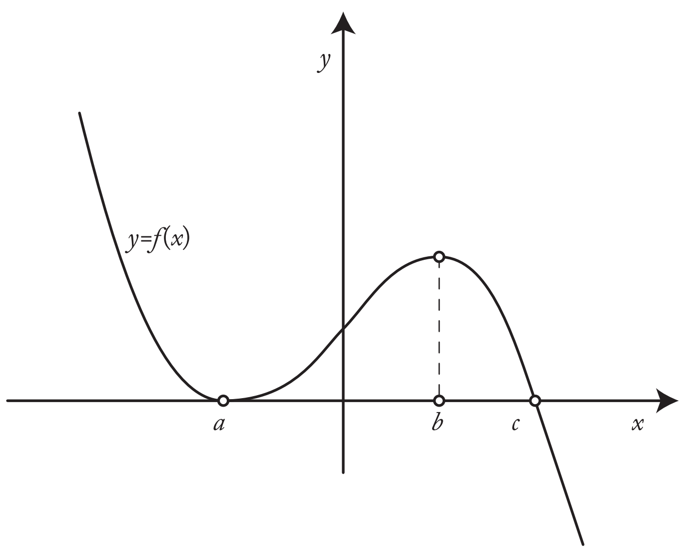
56.69 – Here are the graphs of a function $f(x)$, its derivative $f'(x)$ and an antiderivative $F(x)$ of $f(x)$.

Unfortunately the graphs are not labeled. Identify which graph is which.

Explain your answer.



56.70 – Below is the graph of a function $y = f(x)$



Which among the following statements true?

(**b**) $F(a) = F(c)$?	*True/False*	Reason:
(**c**) $F(b) = 0$?	*True/False*	Reason:
(**d**) $F(b) > F(c)$?	*True/False*	Reason:
(**e**) The graph of $y = F(x)$ has ***two*** inflection points?	*True/False*	Reason:

Use a substitution to evaluate the following integrals.

56.71 – $\displaystyle\int_1^2 \frac{u\,du}{1+u^2}$

56.72 – $\displaystyle\int_0^5 \frac{x\,dx}{\sqrt{x+1}}$

56.73 – $\displaystyle\int_1^2 \frac{x^2\,dx}{\sqrt{2x+1}}$

56.74 – $\displaystyle\int_0^5 \frac{s\,ds}{\sqrt[3]{s+2}}$

56.75 – $\displaystyle\int_1^2 \frac{x\,dx}{1+x^2}$

56.76 – $\displaystyle\int_0^\pi \cos(\theta+\tfrac{\pi}{3})d\theta$

56.77 – $\displaystyle\int \sin\frac{\pi+x}{5}dx$

56.78 – $\displaystyle\int \frac{\sin 2x}{\sqrt{1+\cos 2x}}\,dx$

56.79 – $\displaystyle\int_{\pi/4}^{\pi/3} \sin^2\theta\cos\theta\,d\theta$

56.80 – $\displaystyle\int_2^3 \frac{1}{r\ln r}, dr$

56.81 – $\displaystyle\int \frac{\sin 2x}{1+\cos^2 x}\,dx$

56.82 – $\displaystyle\int \frac{\sin 2x}{1+\sin x}\,dx$

56.83 – $\displaystyle\int_0^1 z\sqrt{1-z^2}\,dz$

56.84 – $\displaystyle\int_1^2 \frac{\ln 2x}{x}\,dx$

56.85 – $\displaystyle\int_{\xi=0}^{\sqrt{2}} \xi(1+2\xi^2)^{10}\,d\xi$

56.86 – $\displaystyle\int_2^3 \sin\rho\big(\cos 2\rho\big)^4\,d\rho$

56.87 – $\displaystyle\int \alpha e^{-\alpha^2}\,d\alpha$

56.88 – $\displaystyle\int \frac{e^{\frac{1}{t}}}{t^2}\,dt$

VIII. **Applications of the integral**

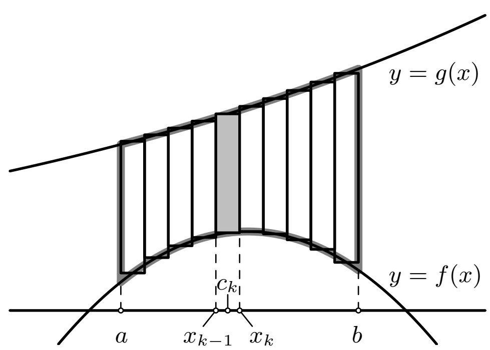
The integral appears as the answer to many different questions. In this chapter we will describe a number of "things which are an integral." In each example there is a quantity we want to compute, and which we can approximate through Riemann-sums. After letting the partition become arbitrarily fine we then find that the quantity we are looking for is given by an integral. The derivations are an important part of the subject.

57. **Areas between graphs**

Suppose you have two functions f and g on an interval $[a, b]$, one of which is always larger than the other, i.e. for which you know that $f(x) \le g(x)$ for all x in the interval $[a, b]$. Then the area of the region between the graphs of the two functions is

$$\text{Area} = \int_a^b \big(g(x) - f(x)\big)dx. \tag{58}$$

To get this formula you approximate the region by a large number of thin rectangles.



Choose a partition $a = x_0 < x_1 < \cdots < x_n = b$ of the interval $[a, b]$; choose a number c_k in each interval $[x_{k-1}, x_k]$; form the rectangles

$$x_{k-1} \le x \le x_k, \qquad f(c_k) \le y \le g(c_k).$$

The area of this rectangle is

$$\text{width} \times \text{height} = \Delta x_k \times \big(g(c_k) - f(c_k)\big).$$

Hence the combined area of the rectangles is

$$R = \big(g(c_1) - f(c_1)\big)\Delta x_1 + \cdots + \big(g(c_n) - f(c_n)\big)\Delta x_n$$

which is just the Riemann-sum for the integral

$$I = \int_a^b (g(x) - f(x))dx.$$

So, since the area of the region between the graphs of f and g is the limit of the combined areas of the rectangles, and since this combined area is equal to the Riemann sum R, which converges to the integral I The area between the graphs of f and g is exactly the integral I.

Exercises

57.1 – Find the area of the region bounded by the parabola $y^2 = 4x$ and the line $y = 2x$.

57.2 – Find the area bounded by the curve $y = x(2 - x)$ and the line $x = 2y$.

57.3 – Find the area bounded by the curve $x^2 = 4y$ and the line $x = 4y - 2$.

57.4 – Calculate the area of the region bounded by the parabolas $y = x^2$ and $x = y^2$.

57.5 – Find the area of the region included between the parabola $y^2 = x$ and the line $x + y = 2$.

57.6 – Find the area of the region bounded by the curves $y = \sqrt{x}$ and $y = x$.

57.7 – Use integration to find the area of the triangular region bounded by the lines $y = 2x + 1$, $y = 3x + 1$ and $x = 4$.

57.8 – Find the area bounded by the parabola $x^2 - 2 = y$ and the line $x + y = 0$.

57.9 – Where do the graphs of $f(x) = x^2$ and $g(x) = 3/(2 + x^2)$ intersect? Find the area of the region which lies above the graph of g and below the graph of g.

(Hint: if you need to integrate $1/(2+x^2)$ you could substitute $x = u\sqrt{2}$.)

57.10 – Graph the curve $y = (1/2)x^2 + 1$ and the straight line $y = x + 1$ and find the area between the curve and the line.

57.11 – Find the area of the region between the parabolas $y^2 = x$ and $x^2 = 16y$.

57.12 – Find the area of the region enclosed by the parabola $y^2 = 4ax$ and the line $y = mx$.

57.13 – Find a so that the curves $y = x^2$ and $y = a\cos x$ intersect at the points $(x, y) = (\frac{\pi}{4}, \frac{\pi^2}{16})$. Then find the area between these curves.

57.14 – Write a definite integral whose value is the area of the region between the two circles $x^2+y^2 = 1$ and $(x-1)^2+y^2 = 1$. Find this area. If you cannot evaluate the integral by calculus you may use geometry to find the area. Hint: The part of a circle cut off by a line is a circular sector with a triangle removed.

58. Cavalieri's principle and volumes of solids

You can use integration to derive the formulas for volumes of spheres, cylinder, cones, and many many more solid objects in a systematic way. In this section we'll see the "method of slicing."

58.1. *Example – Volume of a pyramid*

As an example let's compute the volume of a pyramid whose base is a square of side 1, and whose height is 1. Our strategy will be to divide the pyramid into thin horizontal slices whose volumes we can compute, and to add the volumes of the slices to get the volume of the pyramid.

To construct the slices we choose a partition of the (height) interval $[0, 1]$ into N subintervals, i.e. we pick numbers

$$0 = x_0 < x_1 < x_2 < \cdots < x_{N-1} < x_N = 1,$$

and as usual we set $\Delta x_k = x_k - x_{k-1}$, we define the mesh size of the partition to be the largest of the Δx_k.

The k^{th} slice consists of those points on the pyramid whose height is between x_{k-1} and x_k. The intersection of the pyramid with the plane at height x is a square, and by similarity the length of the side of this square is $1 - x$. Therefore the bottom of the k^{th} slice is a square with side $1 - x_{k-1}$, and its top is a square with side $1 - x_k$. The height of the slice is $x_k - x_{k-1} = \Delta x_k$.

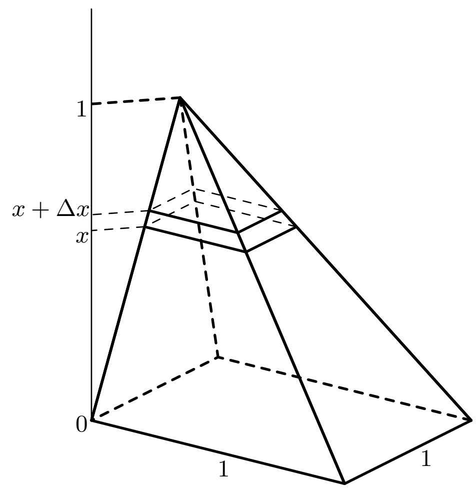


Figure 33. The slice at height x is a square with side $1 - x$.

Thus the k^{th} slice *contains* a block of height Δx_k whose base is a square with sides $1 - x_k$, and its volume must therefore be larger than $(1 - x_k)^2 \Delta x_k$. On the other hand the k^{th} slice *is contained in* a block of the same height whose base is a square with sides $1 - x_{k-1}$. The volume of the slice is therefore not more than $(1 - x_{k-1})^2 \Delta x_k$. So we have

$$(1 - x_k)^2 \Delta x_k < \text{volume of } k^{\text{th}} \text{ slice} < (1 - x_{k-1})^2 \Delta x_k.$$

Therefore there is some c_k in the interval $[x_{k-1}, x_k]$ such that

$$\text{volume of } k^{\text{th}} \text{ slice} = (1 - c_k)^2 \Delta x_k.$$

Adding the volumes of the slices we find that the volume V of the pyramid is given by

$$V = (1 - c_1)^2 \Delta x_1 + \cdots + (1 - c_N)^2 \Delta x_N.$$

The right hand side in this equation is a Riemann sum for the integral

$$I = \int_0^1 (1 - x)^2 dx$$

and therefore we have

$$I = \lim_{\cdots} \big\{ (1 - c_1)^2 \Delta x_1 + \cdots + (1 - c_N)^2 \Delta x_N \big\} = V.$$

Compute the integral and you find that the volume of the pyramid is

$$V = \frac{1}{3}.$$

58.2. *General case*

The "method of slicing" which we just used to compute the volume of a pyramid works for solids of any shape. The strategy always consists of dividing the solid into many thin (horizontal) slices, compute their volumes, and recognize that the total volume of the slices is a Riemann sum for some integral. That integral then is the volume of the solid.

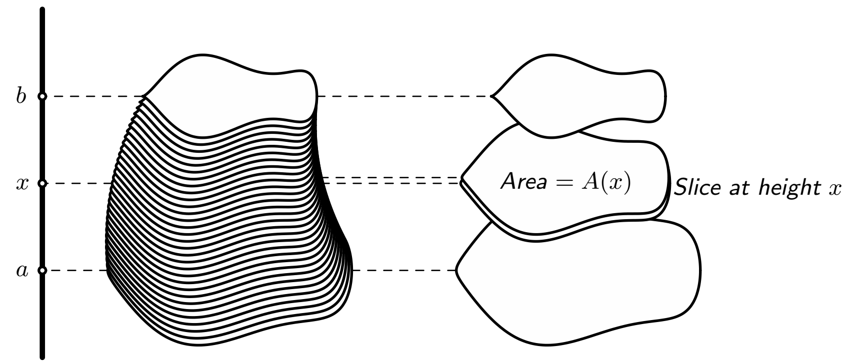


Figure 34. **Slicing a solid to compute its volume.** The volume of one slice is approximately the product of its thickness (Δx) and the area $A(x)$ of its top. Summing the volume $A(x)\Delta x$ over all slices leads approximately to the integral $\int_a^b f(x)dx$.

To be more precise, let a and b be the heights of the lowest and highest points on the solid, and let $a = x_0 < x_1 < x_2 < \ldots < x_{N-1} < x_N = b$ be a partition of the interval $[a, b]$. Such a partition divides the solid into N distinct slices, where slice number k consists of all points in the solid whose height is between x_{k-1} and x_k. The thickness of the k^{th} slice is $\Delta x_k = x_k - x_{k-1}$. If

$$A(x) = \text{area of the intersection of the solid with the plane at height } x.$$

then we can approximate the volume of the k^{th} slice by

$$A(c_k)\Delta x_k$$

where c_k is any number (height) between x_{k-1} and x_k.

The total volume of all slices is therefore approximately

$$V \approx A(c_1)\Delta x_1 + \cdots + A(c_N)\Delta x_N.$$

While this formula only holds approximately, we expect the approximation to get better as we make the partition finer, and thus

$$V = \lim_{\cdots} \big\{A(c_1)\Delta x_1 + \cdots + A(c_N)\Delta x_N\big\}. \tag{59}$$

On the other hand the sum on the right is a Riemann sum for the integral $I = \int_a^b A(x)dx$, so the limit is exactly this integral. Therefore we have

$$V = \int_a^b A(x)dx. \tag{60}$$

58.3. *Cavalieri's principle*

The formula (60) for the volume of a solid which we have just derived shows that the volume only depends on the areas $A(x)$ of the cross sections of the solid, and not on the particular shape these cross sections may have. This observation is older than calculus itself and goes back at least to Bonaventura Cavalieri (1598 – 1647) who said: *If the intersections of two solids with a horizontal plane always have the same area, no matter what the height of the horizontal plane may be, then the two solids have the same volume.*

This principle is often illustrated by considering a stack of coins: If you put a number of coins on top of each other then the total volume of the coins is just the sum of the volumes of the coins. If you change the shape of the pile by sliding the coins horizontally

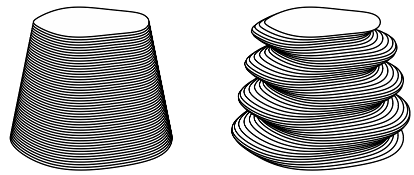


Figure 35. Cavalieri's principle. Both solids consist of a pile of horizontal slices. The solid on the right was obtained from the solid on the left by sliding some of the slices to the left and others to the right. This operation does not affect the volumes of the slices, and hence both solids have the same volume.

then the volume of the pile will still be the sum of the volumes of the coins, i.e. it doesn't change.

58.4. *Solids of revolution*

In principle, formula (60) allows you to compute the volume of any solid, provided you can compute the areas $A(x)$ of all cross sections. One class of solids for which the areas of the cross sections are easy are the so-called "solids of revolution."

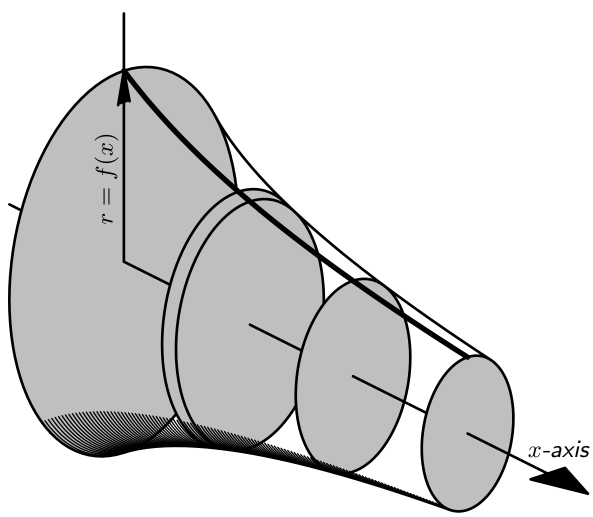


Figure 36. A solid of revolution consists of all points in three-dimensional space whose distance r to the x-axis satisfies $r \leq f(x)$.

A solid of revolution is created by rotating (revolving) the graph of a positive function around the x-axis. More precisely, let f be a function which is defined on an interval $[a, b]$ and which is always positive ($f(x) > 0$ for all x). If you now imagine the x-axis floating in three dimensional space, then the solid of revolution obtained by rotating the graph of f around the x-axis consists of all points in three-dimensional space with $a \leq x \leq b$, and whose distance to the x-axis is no more than $f(x)$.

Yet another way of describing the solid of revolution is to say that the solid is the union of all discs which meet the x-axis perpendicularly and whose radius is given by $r = f(x)$.

If we slice the solid with planes perpendicular to the x-axis, then (60) tells us the volume of the solid. Each slice is a disc of radius $r = f(x)$ so that its area is $A(x) = \pi r^2 = \pi f(x)^2$. We therefore find that

$$V = \pi \int_a^b f(x)^2 dx. \tag{61}$$

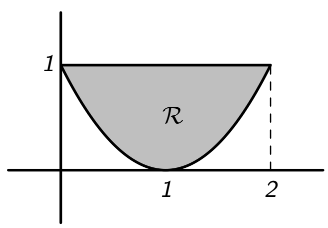
59. Examples of volumes of solids of revolution

59.1. *Problem 1: Revolve $\mathcal{R}$ around the y-axis*

Consider the solid obtained by revolving the region

$$\mathcal{R} = \{(x,y) \mid 0 \le x \le 2, \quad (x-1)^2 \le y \le 1\}$$

around the y-axis.



Solution: The region we have to revolve around the y-axis consists of all points above the parabola $y = (x-1)^2$ but below the line $y = 1$.

If we intersect the solid with a plane at height y then we get a ring shaped region, or "annulus", i.e. a large disc with a smaller disc removed. You can see it in the figure below: if you cut the region $\mathcal{R}$ horizontally at height y you get the line segment AB, and if you rotate this segment around the y-axis you get the grey ring region pictured below the graph. Call the radius of the outer circle r_{out} and the radius of the inner circle r_{in}. These radii are the two solutions of

$$y = (1-r)^2$$

so they are

$$r_{\text{in}} = 1 - \sqrt{y}, \qquad r_{\text{out}} = 1 + \sqrt{y}.$$

The area of the cross section is therefore given by

$$A(y) = \pi r_{\text{out}}^2 - \pi r_{\text{in}}^2 = \pi\big(1+\sqrt{y}\big)^2 - \pi\big(1-\sqrt{y}\big)^2 = 4\pi\sqrt{y}.$$

The y-values which occur in the solid are $0 \le y \le 1$ and hence the volume of the solid is given by

$$V = \int_0^1 A(y)dy = 4\pi \int_0^1 \sqrt{y}\, dy = 4\pi \times \tfrac{2}{3} = \frac{8\pi}{3}.$$

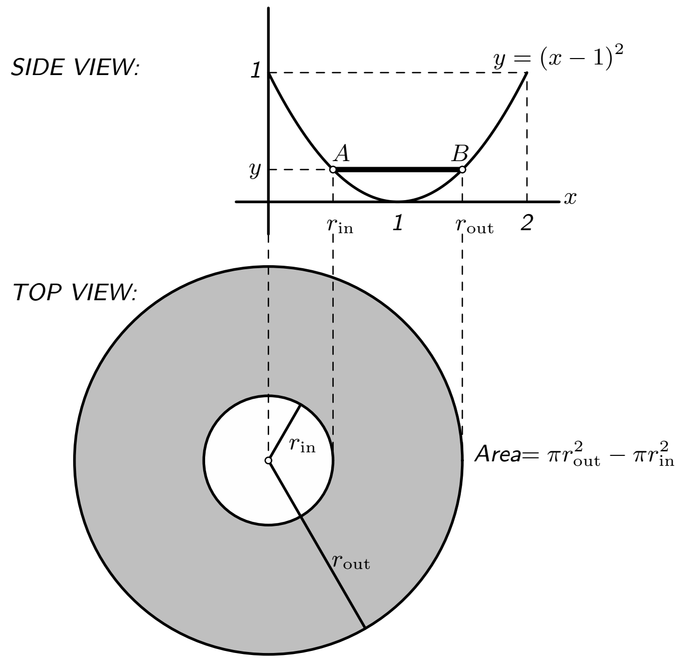
59.2. *Problem 2: Revolve $\mathcal{R}$ around the line $x = -1$*

Find the volume of the solid of revolution obtained by revolving the same region $\mathcal{R}$ around the line $x = -1$.

Solution: The line $x = -1$ is vertical, so we slice the solid with horizontal planes. The height of each plane will be called y.

As before the slices are ring shaped regions but the inner and outer radii are now given by

$$r_{\text{in}} = 1 + x_{\text{in}} = 2 - \sqrt{y}, \qquad r_{\text{out}} = 1 + x_{\text{out}} = 2 + \sqrt{y}.$$



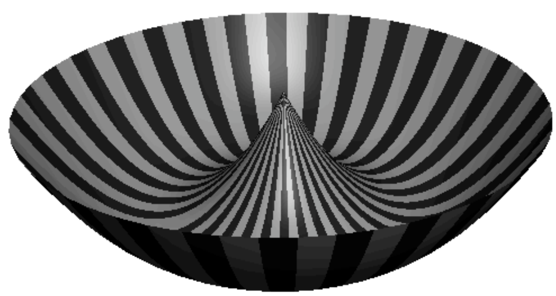


Figure 37. What is the volume of this bowl? Answer: $8\pi/3$.

The volume is therefore given by

$$V = \int_0^1 \left(\pi r_{\text{out}}^2 - \pi r_{\text{in}}^2\right) dy = \pi \int_0^1 8\sqrt{y}\, dy = \frac{16\pi}{3}.$$

59.3. *Problem 3: Revolve $\mathcal{R}$ around the line $y = 2$*

Compute the volume of the solid you get when you revolve the same region $\mathcal{R}$ around the line $y = 2$.

Solution: This time the line around which we rotate $\mathcal{R}$ is horizontal, so we slice the solid with planes perpendicular to the x-axis.

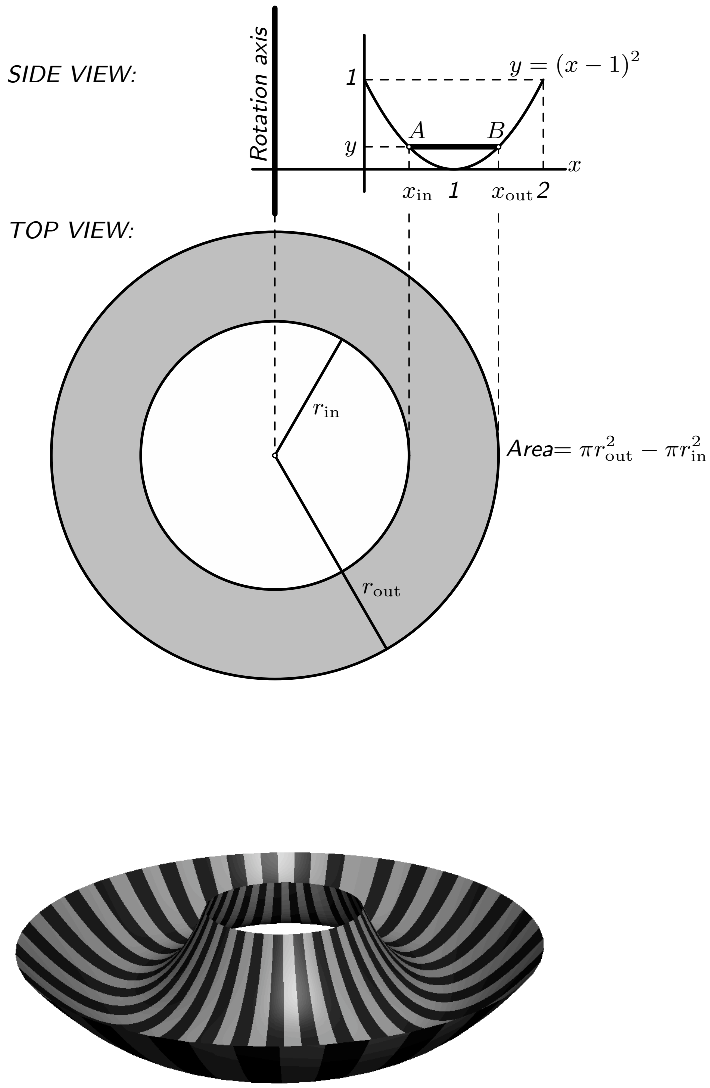


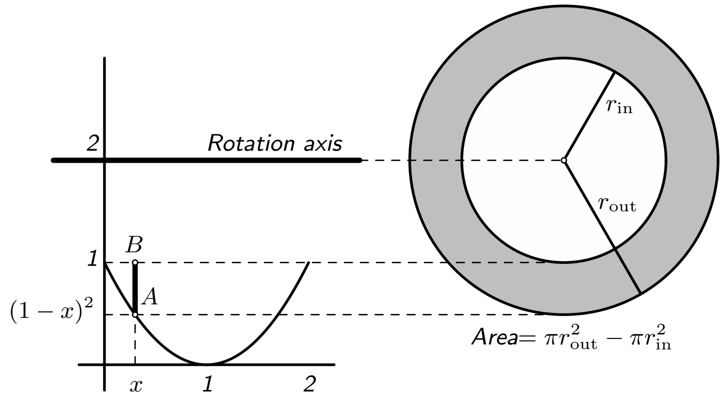
Figure 38. What is the volume of the cake you could bake in this form? Answer: $16\pi/3$. (That's twice as much as in the bowl from Problem 1 – double your recipe!)

A typical slice is obtained by revolving the line segment AB about the line $y = 2$. The result is again an annulus, and from the figure we see that the inner and outer radii of the annulus are

$$r_{\text{in}} = 1, \qquad r_{\text{out}} = 2 - (1-x)^2.$$

The area of the slice is therefore

$$A(x) = \pi\big\{2 - (1-x)^2\big\}^2 - \pi 1^2 = \pi\left\{3 - 4(1-x)^2 + (1-x)^4\right\}.$$



The x values which occur in the solid are $0 \le x \le 2$, and so its volume is

$$\begin{aligned} V &= \pi \int_0^2 \left\{3 - 4(1-x)^2 + (1-x)^4\right\} dx \\ &= \pi \left[3x + \tfrac{4}{3}(1-x)^3 - \tfrac{1}{5}(1-x)^5\right]_0^2 \\ &= \pi \left[6 - \tfrac{8}{3} + \tfrac{2}{5}\right] \\ &= \tfrac{56}{15}\pi \end{aligned}$$

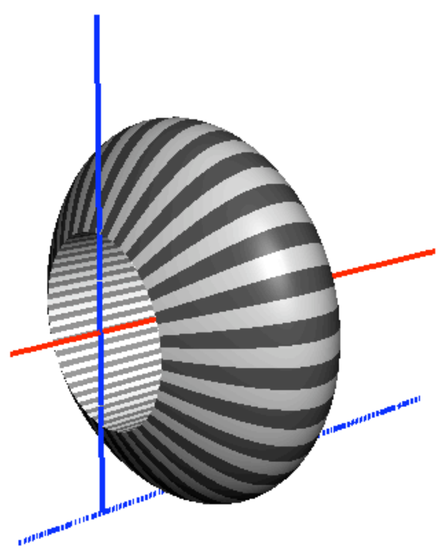
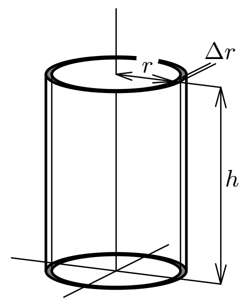


Figure 39. The region $\mathcal{R}$ revolved about the line $y = 2$ produces this solid. Shown are the x-axis, y-axis (vertical) and the axis of rotation.

60. **Volumes by cylindrical shells**

Instead of slicing a solid with planes you can also try to decompose it into cylindrical shells. The volume of a cylinder of height h and radius r is $\pi r^2 h$ (height times area base). Therefore the volume of a cylindrical shell of height h, (inner) radius r and thickness Δr is

$$\begin{aligned}\pi h(r+\Delta r)^2 - \pi h r^2 \\ &= \pi h(2r+\Delta r)\Delta r \\ &\approx 2\pi h r \Delta r.\end{aligned}$$

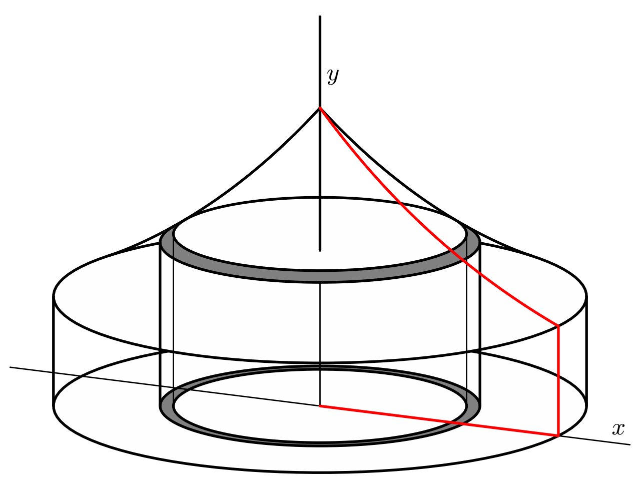


Now consider the solid you get by revolving the region

$$\mathcal{R} = \{(x,y) \mid a \le x \le b, 0 \le y \le f(x)\}$$

around the y-axis. By partitioning the interval $a \le x \le b$ into many small intervals we can decompose the solid into many thin shells. The volume of each shell will approximately be given by $2\pi x f(x)\Delta x$. Adding the volumes of the shells, and taking the limit over finer and finer partitions we arrive at the following formula for the volume of the solid of revolution:

$$V = 2\pi \int_a^b x f(x)\, dx. \tag{62}$$



If the region $\mathcal{R}$ is not the region under the graph, but rather the region between the graphs of two functions $f(x) \le g(x)$, then we get

$$V = 2\pi \int_a^b x\big\{g(x) - f(x)\big\}\, dx.$$

60.1. *Example – The solid obtained by rotating $\mathcal{R}$ about the y-axis, again*

The region $\mathcal{R}$ from §59.1 can also be described as

$$\mathcal{R} = \{(x, y) \mid 0 \le x \le 2, f(x) \le y \le g(x)\},$$

where

$$f(x) = (x-1)^2 \text{ and } g(x) = 1.$$

The volume of the solid which we already computed in §59.1 is thus given by

$$\begin{aligned} V &= 2\pi \int_0^1 x\{1-(x-1)^2\}\, dx \\ &= 2\pi \int_0^2 \{-x^3 + 2x^2\}\, dx \\ &= 2\pi \left[-\tfrac{1}{4}x^4 + \tfrac{2}{3}x^3\right]_0^2 \\ &= 8\pi/3, \end{aligned}$$

which coincides with the answer we found in §59.1.

Exercises

60.1 – What do the dots in "lim..." in equation (59) stand for? (i.e. what approaches what in this limit?)

Draw and describe the solids whose volume you are asked to compute in the following problems:

60.2 – Find the volume enclosed by the paraboloid obtained by rotating the graph of $f(x) = R\sqrt{x/H}$ $(0 \le x \le H)$ around the x-axis. Here R and H are positive constants. Draw the solid whose volume you are asked to compute, and indicate what R and H are in your drawing.

60.3 – Find the volume of the solids you get by rotating each of the following graphs around the x-axis:

(**a**) $f(x) = x$, $0 \le x \le 2$
(**b**) $f(x) = \sqrt{2-x}$, $0 \le x \le 2$
(**c**) $f(x) = (1+x^2)^{-1/2}$, $|x| \le 1$
(**d**) $f(x) = \sin x$, $0 \le x \le \pi$
(**e**) $f(x) = 1 - x^2$, $|x| \le 1$
(**f**) $f(x) = \cos x$, $0 \le x \le \pi$ (**!!**)
(**g**) $f(x) = 1/\cos x$, $0 \le x \le \pi/4$

60.4 – Find the volume that results by rotating the semicircle $y = \sqrt{R^2 - x^2}$ about the x-axis.

60.5 – Let $\mathcal{T}$ be triangle $1 \le x \le 2$, $0 \le y \le 3x - 3$.

(**a**) Find the volume of the solid obtained by rotating the triangle $\mathcal{T}$ around the x-axis.
(**b**) Find the volume that results by rotating the triangle $\mathcal{T}$ around the y axis.
(**c**) Find the volume that results by rotating the triangle $\mathcal{T}$ around the line $x = -1$.
(**d**) Find the volume that results by rotating the triangle $\mathcal{T}$ around the line $y = -1$.

60.6 – (**a**) A spherical bowl of radius a contains water to a depth $h < 2a$. Find the volume of the water in the bowl. (Which solid of revolution is implied in this problem?)

(**b**) Water runs into a spherical bowl of radius 5 ft at the rate of 0.2 ft^3/sec. How fast is the water level rising when the water is 4 ft deep?

61. Distance from velocity, velocity from acceleration

61.1. *Motion along a line*

If an object is moving on a straight line, and if its position at time t is $x(t)$, then we had defined the velocity to be $v(t) = x'(t)$. Therefore the position is an antiderivative of the velocity, and the fundamental theorem of calculus says that

$$\int_{t_a}^{t_b} v(t)\,dt = x(t_b) - x(t_a), \tag{63}$$

or

$$x(t_b) = x(t_a) + \int_{t_a}^{t_b} v(t)\,dt.$$

In words, the integral of the velocity gives you the distance travelled of the object (during the interval of integration).

Equation (63) can also be obtained using Riemann sums. Namely, to see how far the object moved between times t_a and t_b we choose a partition $t_a = t_0 < t_1 < \cdots < t_N = t_b$. Let Δs_k be the distance travelled during the time interval (t_{k-1}, t_k). The length of this time interval is $\Delta t_k = t_k - t_{k-1}$. During this time interval the velocity $v(t)$ need not be constant, but if the time interval is short enough then we can estimate the velocity by $v(c_k)$ where c_k is some number between t_{k-1} and t_k. We then have

$$\Delta s_k = v(c_k)\Delta t_k$$

and hence the total distance travelled is the sum of the travel distances for all time intervals $t_{k-1} < t < t_k$, i.e.

$$\text{Distance travelled} \approx \Delta s_1 + \cdots + \Delta s_N = v(c_1)\Delta t_1 + \cdots + v(c_N)\Delta t_N.$$

The right hand side is again a Riemann sum for the integral in (63). As one makes the partition finer and finer you therefore get

$$\text{Distance travelled} = \int_{t_a}^{t_b} v(t)\,dt.$$

The return of the dummy. Often you want to write a formula for $x(t) = \cdots$ rather than $x(t_b) = \cdots$ as we did in (63), i.e. you want to say what the position is at time t, instead of at time t_a. For instance, you might want to express the fact that the position $x(t)$ is equal to the initial position $x(0)$ plus the integral of the velocity from 0 to t. To do this you cannot write

$$x(t) = x(0) + \int_0^t v(t)\,dt \quad \Leftarrow \quad \textbf{BAD FORMULA}$$

because the variable t gets used in two incompatible ways: the t in $x(t)$ on the left, and in the upper bound on the integral ($\int^t$) are the same, but they are not the same as the two t's in $v(t)dt$. The latter is a dummy variable (see §17 and §52.1). To fix this formula we should choose a different letter or symbol for the integration variable. A common choice in this situation is to decorate the integration variable with a prime (t'), a tilde ($\tilde{t}$) or a bar ($\bar{t}$). So you can write

$$x(t) = x(0) + \int_0^t v(\bar{t})\,d\bar{t}$$

61.2. *Velocity from acceleration*

The acceleration of the object is by definition the rate of change of its velocity,

$$a(t) = v'(t),$$

so you have

$$v(t) = v(0) + \int_0^t a(\bar{t})d\bar{t}.$$

Conclusion: *If you know the acceleration $a(t)$ at all times t, and also the velocity $v(0)$ at time $t = 0$, then you can compute the velocity $v(t)$ at all times by integrating.*

61.3. *Free fall in a constant gravitational field*

If you drop an object then it will fall, and as it falls its velocity increases. The object's motion is described by the fact that *its acceleration is constant.* This constant is called g and is about $9.8\text{m}/\text{sec}^2 \approx 32\text{ft}/\text{sec}^2$. If we dsignate the upward direction as positive then $v(t)$ is the upward velocity of the object, and this velocity is actually decreasing. Therefore the constant acceleration is negative: it is $-g$.

If you write $h(t)$ for the height of the object at time t then its velocity is $v(t) = h'(t)$, and its acceleration is $h''(t)$. Since the acceleration is constant you have the following formula for the velocity at time t:

$$v(t) = v(0) + \int_0^t (-g)\, d\bar t = v(0) - gt.$$

Here $v(0)$ is the velocity at time $t = 0$ (the "initial velocity").

To get the height of the object at any time t you must integrate the velocity:

$$\begin{aligned} h(t) &= h(0) + \int_0^t v(\bar t)\, d\bar t && \text{(Note the use of the dummy } \bar t\text{)} \\ &= h(0) + \int_0^t \{v(0) - g\bar t\}\, d\bar t && \text{(use } v(\bar t) = v(0) - g\bar t\text{)} \\ &= h(0) + \left[v(0)\bar t - \tfrac12 g\bar t^2\right]_{\bar t=0}^{\bar t=t} \\ &= h(0) + v(0)t - \tfrac12 gt^2. \end{aligned}$$

For instance, if you launch the object upwards with velocity 5ft/sec from a height of 10ft, then you have

$$h(0) = 10\text{ft}, \quad v(0) = +5\text{ft/sec},$$

and thus

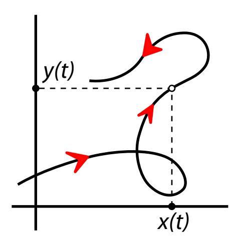
$$h(t) = 10 + 5t - 32t^2/2 = 10 + 5t - 16t^2.$$

The object reaches its maximum height when $h(t)$ has a maximum, which is when $h'(t) = 0$. To find that height you compute $h'(t) = 5 - 32t$ and conclude that $h(t)$ is maximal at $t = \frac{5}{32}$sec. The maximal height is then

$$h_{\max} = h(\tfrac{5}{32}) = 10 + \tfrac{25}{32} - \tfrac{25}{64} = 10\tfrac{25}{64}\text{ft}.$$

61.4. *Motion in the plane – parametric curves*

To describe the motion of an object in the plane you could keep track of its x and y coordinates at all times t. This would give you *two* functions of t, namely, $x(t)$ and $y(t)$, both of which are defined on the same interval $t_0 \le t \le t_1$ which describes the duration of the motion you are describing. In this context a pair of functions $(x(t), y(t))$ is called ***a parametric curve.***



A moving point in the plane

As an example, consider the motion described by

$$x(t) = \cos t, \quad y(t) = \sin t (0 \le t \le 2\pi).$$

In this motion the point $(x(t), y(t))$ lies on the unit circle since

$$x(t)^2 + y(t)^2 = \cos^2 t + \sin^2 t = 1.$$

As t increases from 0 to 2π the point $(x(t), y(t))$ goes around the unit circle exactly once, in the counter-clockwise direction.

In another example one could consider

$$x(t) = t, \quad y(t) = \sqrt{1 - t^2}, \quad (-1 \le t \le 1).$$

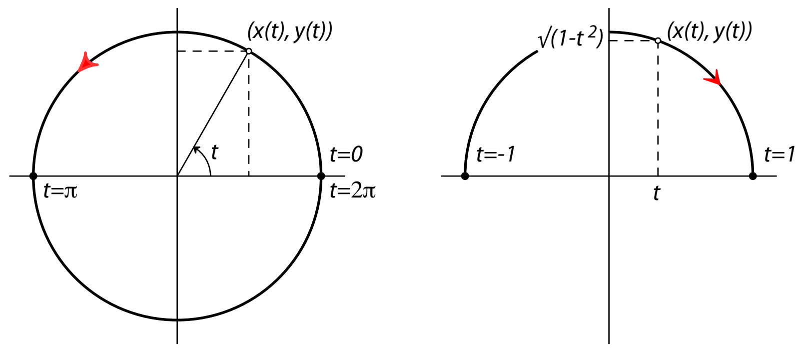


Figure 40. Two motions in the plane. On the left $x(t) = \cos t$, $y(t) = \sin t$ with $0 \le t \le 2\pi$, and on the right $x(t) = t$, $y(t) = \sqrt{(1-t^2)}$ with $-1 \le t \le 1$.

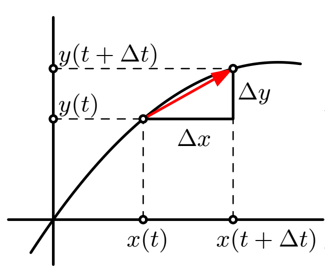
Here at all times the x and y coordinates satisfy

$$x(t)^2 + y(t)^2 = 1$$

again so that the point $(x(t), y(t)) = \big(t, \sqrt{1-t^2}\big)$ again lies on the unit circle. Unlike the previous example we now always have $y(t) \ge 0$ (since $y(t)$ is the square root of something), and unlike the previous example the motion is only defined for $-1 \le t \le 1$. As t increases from -1 to $+1$, $x(t) = t$ does the same, and hence the point $(x(t), y(t))$ moves along the *upper half of the unit circle* from the leftmost point to the rightmost point.

61.5. *The velocity of an object moving in the plane*

We have seen that the velocity of an object which is moving along a line is the derivative of its position. If the object is allowed to move in the plane, so that its motion is described by a parametric curve $(x(t), y(t))$, then we can differentiate both $x(t)$ and $y(t)$, which gives us $x'(t)$ and $y'(t)$, and which leaves us with the following



Question: *At what speed is a particle moving if it is undergoing the motion $(x(t), y(t))$ $(t_a \le t \le t_b)$?*

To answer this question we consider a short time interval $(t, t+\Delta t)$. During this time interval the particle moves from $(x(t), y(t))$ to $(x(t+\Delta t), t(t+\Delta t))$. Hence it has traveled a distance

$$\Delta s = \sqrt{(\Delta x)^2 + (\Delta y)^2}$$

where

$$\Delta x = x(t+\Delta t) - x(t), \text{ and } \Delta y = y(t+\Delta t) - y(t).$$

Dividing by Δt you get

$$\frac{\Delta s}{\Delta t} = \sqrt{\left(\frac{\Delta x}{\Delta t}\right)^2 + \left(\frac{\Delta y}{\Delta t}\right)^2},$$

for the average velocity over the time interval $[t, t+\Delta t]$. Letting $\Delta t \to 0$ you find the velocity at time t to be

(64) $$v = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}.$$

61.6. *Example – the two motions on the circle from §61.4*

If a point moves along a circle according to $x(t) = \cos t$, $y(t) = \sin t$ (figure 40 on the left) then

$$\frac{dx}{dt} = -\sin t, \qquad , \frac{dy}{dt} = +\cos t$$

so

$$v(t) = \sqrt{(-\cos t)^2 + (\sin t)^2} = 1.$$

The velocity of this motion is therefore always the same; the point $(\cos t, \sin t)$ moves along the unit circle with constant velocity.

In the second example in §61.4 we had $x(t) = t$, $y(t) = \sqrt{1-t^2}$, so

$$\frac{dx}{dt} = 1, \quad \frac{dy}{dt} = \frac{-t}{\sqrt{1-t^2}}$$

whence

$$v(t) = \sqrt{1^2 + \frac{t^2}{1-t^2}} = \frac{1}{\sqrt{1-t^2}}.$$

Therefore the point $(t, \sqrt{1-t^2})$ moves along the upper half of the unit circle from the left to the right, and its velocity changes according to $v = 1/\sqrt{1-t^2}$.

62. **The length of a curve**

62.1. *Length of a parametric curve*

Let $(x(t), y(t))$ be some parametric curve defined for $t_a \le t \le t_b$. To find the length of this curve you can reason as follows: The length of the curve should be the distance travelled by the point $(x(t), y(t))$ as t increases from t_a to t_b. At each moment in time the velocity $v(t)$ of the point is given by (64), and therefore the distance traveled should be

$$s = \int_{t_a}^{t_b} v(t)\, dt = \int_{t_a}^{t_b} \sqrt{x'(t)^2 + y'(t)^2}\, dt. \tag{65}$$

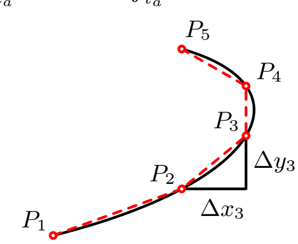


Figure 41. Approximate a curve by a polygon to compute its length.

Alternatively, you can try to compute the distance travelled by means of Riemann sums. Choose a partition $t_a = t_0 < t_1 < \cdots < t_N = t_b$ of the interval $[t_a, t_b]$. You then get a sequence of points $P_0(x(t_0), y(t_0))$, $P_1(x(t_1), y(t_1))$, …, $P_N(x(t_N), y(t_N))$, and we can approximate the l;ength of the curve by computing the length of the polygon you get by drawing straight line segments from P_0 to P_1, from P_1 to P_2, etc. The distance between two consecutive points P_{k-1} and P_k is

$$\begin{aligned}\Delta s_k &= \sqrt{(\Delta x_k)^2 + (\Delta y_k)^2} \\ &= \sqrt{\left(\frac{\Delta x_k}{\Delta t_k}\right)^2 + \left(\frac{\Delta y_k}{\Delta t_k}\right)^2}\ \Delta t_k \\ &\approx \sqrt{x'(c_k)^2 + y'(c_k)^2}\ \Delta t_k\end{aligned}$$

where we have approximated the difference quotients

$$\frac{\Delta x_k}{\Delta t_k} \text{ and } \frac{\Delta y_k}{\Delta t_k}$$

by the derivatives $x'(c_k)$ and $y'(c_k)$ for some c_k in the interval $[t_{k-1}, t_k]$.

The total length of the polygon is then

$$\sqrt{x'(c_1)^2 + y'(c_1)^2}\ \Delta t_1 + \cdots + \sqrt{x'(c_1)^2 + y'(c_1)^2}\ \Delta t_1$$

This is a Riemann sum for the integral $\int_{t_a}^{t_b} \sqrt{x'(t)^2 + y'(t)^2}\, dt$, and hence we find (once more) that the length of the curve is

$$s = \int_{t_a}^{t_b} \sqrt{x'(t)^2 + y'(t)^2}\, dt.$$

62.2. *The length of the graph of a function*

The graph of a function ($y = f(x)$ with $a \le x \le b$) is also a curve in the plane, and you can ask what its length is. We will now find this length by *representing the graph as a parametric curve* and applying the formula (65) from the previous section.

The standard method of representing the graph of a function $y = f(x)$ by a parametric curve is to choose

$$x(t) = t, \text{ and } y(t) = f(t), \text{ for } a \le t \le b.$$

This parametric curve traces the graph of $y = f(x)$ from left to right as t increases from a to b.

Since $x'(t) = 1$ and $y'(t) = f(t)$ we find that the length of the graph is

$$L = \int_a^b \sqrt{1 + f'(t)^2}\, dt.$$

The variable t in this integral is a dummy variable and we can replace it with any other variable we like, for instance, x:

$$L = \int_a^b \sqrt{1 + f'(x)^2}\, dx. \tag{66}$$

In Leibniz' notation we have $y = f(x)$ and $f'(x) = dy/dx$ so that Leibniz would have written

$$L = \int_a^b \sqrt{1 + (\frac{dy}{dx})^2}\, dx.$$

62.3. *Examples of length computations*

62.3.1. *Length of a circle.* In §62 we parametrized the unit circle by

$$x(t) = \cos t, \quad y(t) = \sin t, \quad (0 \le t \le 2\pi)$$

and computed $\sqrt{x'(t)^2 + y'(t)^2} = 1$. Therefore our formula tells us that the length of the unit circle is

$$L = \int_0^{2\pi} \sqrt{x'(t)^2 + y'(t)^2}\, dt = \int_0^{2\pi} 1\, dt = 2\pi.$$

This cannot be a PROOF that the unit circle has length 2π since we have already used that fact to define angles in radians, to define the trig functions Sine and Cosine, and to find their derivatives. But our computation shows that the length formula (66) is at least consistent with what we already knew.

62.3.2. *Length of a parabola.* Consider our old friend, the parabola $y = x^2$, $0 \le x \le 1$. While the area under its graph was easy to compute $(1/3)$, its length turns out to be much more complicated.

Our length formula (66) says that the length of the parabola is given by

$$L = \int_0^1 \sqrt{1 + (\frac{dx^2}{dx})^2}\, dx = \int_0^1 \sqrt{1+4x^2}\, dx.$$

To find this integral you would have to use one of the following (not at all obvious) substitutions[9]

$$x = \tfrac{1}{4}\big(z - \frac{1}{z}\big) \qquad \text{(then } 1+4x^2 = \frac{1}{4}(z+1/z)^2 \text{ so you can simplify the } \sqrt{\cdot}\text{)}$$

or (if you like hyperbolic functions)

$$x = \tfrac{1}{2}\sinh w \qquad \text{(in which case } \sqrt{1+4x^2} = \cosh w.\text{)}$$

62.3.3. *Length of the graph of the Sine function.* To compute the length of the curve given by $y = \sin x$, $0 \le x \le \pi$ you would have to compute this integral:

$$L = \int_0^\pi \sqrt{1 + (\frac{d\sin x}{dx})^2}\, dx = \int_0^\pi \sqrt{1+\cos^2 x}\, dx. \tag{67}$$

Unfortunately this is not an integral which can be computed in terms of the functions we know in this course (it's an "elliptic integral of the second kind.") This happens very often with the integrals that you run into when you try to compute the length of a curve. In spite of the fact that we get stuck when we try to compute the integral in (67), the formula is not useless. For example, since $-1 \le< \cos x \le 1$ we know that

$$1 \le \sqrt{1+\cos^2 x} \le \sqrt{1+1} = \sqrt{2},$$

and therefore the length of the Sine graph is bounded by

$$\int_0^\pi 1dx \;\le\; \int_0^\pi \sqrt{1+\cos^2 x}\, dx \;\le\; \int_0^\pi \sqrt{2}\, dx,$$

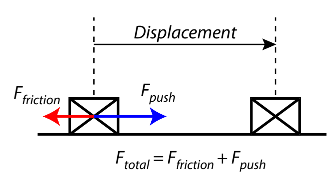
i.e.

$$\pi \le L \le \pi\sqrt{2}.$$

63. **Work done by a force**

63.1. *Work as an integral*

In Newtonian mechanics a force which acts on an object in motion performs a certain amount of work, i.e. it spends a certain amount of energy. If the force which acts is constant, then the work done by this force is



$$\text{Work} \;=\; \text{Force} \;\times\; \text{Displacement}.$$

For example if you are pushing a box forward then there will be two forces acting on the box: the force you apply, and the friction force of the floor on the box.

The amount of work you do is the product of the force you exert and the length of the displacement. Both displacement and the force you apply are pointed towards the right, so both are positive, and the work you do (energy you provide to the box) is positive.

The amount of work done by the friction is similarly the product of the friction force and the displacement. Here the displacement is still to the right, but the friction force points to the left, so it is negative. The work done by the friction force is therefore negative. Friction extracts energy from the box.

[9] Many calculus textbooks will tell you to substitute $x = \tan\theta$, but the resulting integral is still not easy.

Suppose now that the force $F(t)$ on the box is not constant, and that its motion is described by saying that its position at time t is $x(t)$. The basic formula work = force × displacement does not apply directly since it assumes that the force is constant. To compute the work done by the varying force $F(t)$ we choose a partition of the time interval $t_a \le t \le t_b$ into

$$t_a = t_0 < t_1 < \cdots < t_{N-1} < t_N = t_b$$

In each short time interval $t_{k-1} \le t \le t_k$ we assume the force is (almost) constant and we approximate it by $F(c_k)$ for some $t_{k-1} \le c_k \le t_k$. If we also assume that the velocity $v(t) = x'(t)$ is approximately constant between times t_{k-1} and t_k then the displacement during this time interval will be

$$x(t_k) - x(t_{k-1}) \approx v(c_k)\Delta t_k,$$

where $\Delta t_k = t_k - t_{k-1}$. Therefore the work done by the force F during the time interval $t_{k-1} \le t \le t_k$ is

$$\Delta W_k = F(c_k)v(c_k)\Delta t_k.$$

Adding the work done during each time interval we get the total work done by the force between time t_a and t_b:

$$W = F(c_1)v(c_1)\Delta t_1 + \cdots + F(c_N)v(c_N)\Delta t_N.$$

Again we have a Riemann sum for an integral. If we take the limit over finer and finer partitions we therefore find that the work done by the force $F(t)$ on an object whose motion is described by $x(t)$ is

$$W = \int_{t_a}^{t_b} F(t)v(t)dt, \tag{68}$$

in which $v(t) = x'(t)$ is the velocity of the object.

63.2. *Kinetic energy*

Newton's famous law relating the force exerted on an object and its motion says $F = ma$, where a is the acceleration of the object, m is its mass, and F is the combination of all forces acting on the object. If the position of the object at time t is $x(t)$, then its velocity and acceleration are $v(t) = x'(t)$ and $a(t) = v'(t) = x''(t)$, and thus the total force acting on the object is

$$F(t) = ma(t) = m\frac{dv}{dt}.$$

The work done by the total force is therefore

$$W = \int_{t_a}^{t_b} F(t)v(t)dt = \int_{t_a}^{t_b} m\frac{dv(t)}{dt}\ v(t)\ dt. \tag{69}$$

Even though we have not assumed anything about the motion, so we don't know anything about the velocity $v(t)$, we can still do this integral. The key is to notice that, by the chain rule,

$$m\frac{dv(t)}{dt}\ v(t) = \frac{d\frac{1}{2}mv(t)^2}{dt}.$$

(Remember that m is a constant.) This says that the quantity

$$K(t) = \tfrac{1}{2}mv(t)^2$$

is the antiderivative we need to do the integral (69). We get

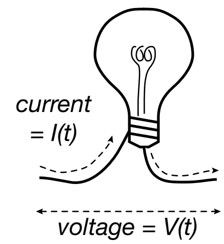
$$W = \int_{t_a}^{t_b} m\frac{dv(t)}{dt}\ v(t)\ dt = \int_{t_a}^{t_b} K'(t)\ dt = K(t_b) - K(t_a).$$

In Newtonian mechanics the quantity $K(t)$ is called ***the kinetic energy*** of the object, and our computation shows that *the amount by which the kinetic energy of an object increases is equal to the amount of work done on the object.*

64. Work done by an electric current

If at time t an electric current $I(t)$ (measured in Ampère) flows through an electric circuit, and if the voltage across this circuit is $V(t)$ (measured in Volts) then the energy supplied tot the circuit per second is $I(t)V(t)$. Therefore the total energy supplied during a time interval $t_0 \le t \le t_1$ is the integral

$$\text{Energy supplied} = \int_{t_0}^{t_1} I(t)V(t)dt.$$



(measured in Joule; the energy consumption of a circuit is defined to be how much energy it consumes per time unit, and the power consumption of a circuit which consumes 1 Joule per second is said to be one Watt.)

If a certain voltage is applied to a simple circuit (like a light bulb) then the current flowing through that circuit is determined by the resistance R of that circuit by *Ohm's law*[10] which says

$$I = \frac{V}{R}.$$

Example. If the resistance of a light bulb is $R = 200\Omega$, and if the voltage applied to it is

$$V(t) = 150 \sin 2\pi f t$$

where $f = 50\text{sec}^{-1}$ is the frequency, then how much energy does the current supply to the light bulb in one second?

To compute this we first find the current using Ohm's law,

$$I(t) = \frac{V(t)}{R} = \frac{150}{200} \sin 2\pi f t = 0.75 \sin 2\pi f t. \quad \text{(Amp)}$$

The energy supplied in one second is then

$$\begin{aligned} E &= \int_0^{1 \text{ sec}} I(t)V(t)dt \\ &= \int_0^1 (150 \sin 2\pi f t) \times (0.75 \sin 2\pi f t)\, dt \\ &= 112.5 \int_0^1 \sin^2(2\pi f t)\, dt \end{aligned}$$

You can do this last integral by using the double angle formula for the cosine, to rewrite

$$\sin^2(2\pi f t) = \tfrac{1}{2}\{1 - \cos 4\pi f t\} = \tfrac{1}{2} - \tfrac{1}{2}\cos 4\pi f t.$$

Keep in mind that $f = 50$, and you find that the integral is

$$\int_0^1 \sin^2(2\pi f t)\, dt = \Big[\frac{t}{2} - \frac{1}{4\pi f}\sin 4\pi f t\Big]_0^1 = \tfrac{1}{2},$$

and hence the energy supplied to the light bulb during one second is

$$E = 112.5 \times \tfrac{1}{2} = 56.25(\text{Joule}).$$

[10] http://en.wikipedia.org/wiki/Ohm's_law