THE UNIVERSITY of WISCONSIN MADISON

MATH 221
FIRST SEMESTER
CALCULUS

Fall 2015

Typeset:June 19, 2015

MATH 221 – 1st Semester Calculus
Lecture notes version 1.3 (Fall 2015)

Contents

CHAPTER 1

Numbers and Functions

The subject of this course is "functions of one real variable" so we begin by talking about the real numbers, and then discuss functions.

1. Numbers

1.1. Different kinds of numbers. 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

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

or

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

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).

Approximating $\sqrt{2}$:

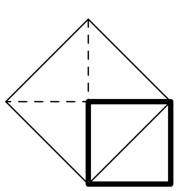
x	x^2	
1.2	1.44	
1.3	1.69	
1.4	***1.96***	¡2
1.5	***2.25***	¿2
1.6	2.56	

There are other numbers besides the rational numbers: it was discovered by the ancient Greeks that the square root of 2 is not a rational number. In other words, there does not exist a fraction $\frac{m}{n}$ such that

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

Nevertheless, if you compute x^2 for some values of x between 1 and 2, and check if you get more or less than 2, then it looks like there should be some number x between 1.4 and 1.5 whose square is exactly 2.

1.2. A 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 high school you learned something similar to the following geometric construction of a line segment whose length is $\sqrt{2}$. Take a square with side of length 1, and construct a new square one of whose sides is the diagonal of the first square. The figure you get consists of 5 triangles of equal area and by counting triangles you see that the larger square has exactly twice the area of the smaller square. Therefore the diagonal of the smaller square,



being the side of the larger square, is $\sqrt{2}$ as long as the side of the smaller square.

So, we will *assume* that there is such a number whose square is exactly 2, and we call it the square root of 2, written as $\sqrt{2}$. There are more than a few questions[1] raised by assuming the existence of numbers like the square root of 2, but we will not deal with these questions here. Instead, we will take a more informal approach of thinking of real numbers as "infinite decimal expansions".

1.3. Decimals.

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

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

Some rational numbers cannot be expressed with a finite decimal expansion. For instance, expanding $\frac{1}{3}$ as a decimal number leads to an unending "repeating decimal":

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

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 3.

Every fraction can be written as a decimal number which may or may not be finite. If the decimal expansion doesn't terminate, then it will repeat (although this fact is not so easy to show). 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 you only write the first few digits and hide the others behind dots. To give a precise description of a real number (such as $\sqrt{2}$) you have to explain how you could *in principle* compute as many digits in the expansion as you would like[2].

1.4. 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 history it became useful to assume that there is such a thing as $\sqrt{-1}$, a number whose square is -1. No real number has this property (since the square of any real number is nonnegative) so it was decided to call this newly-imagined number "imaginary" and to refer to the numbers we already had (rationals, $\sqrt{2}$-like things) as "real".

---

[1] Here are some questions: if we assume into existence a number a number x between 1.4 and 1.5 for which $x^2 = 2$, how many other such numbers must we also assume into existence? How do we know there is "only one" such square root of 2? And how can we be sure that these new numbers will obey the same algebra rules (like $a + b = b + a$) as the rational numbers?

[2] What exactly we mean by specifying how to compute digits in decimal expansions of real numbers is another issue that is beyond the scope of this course.

1.5. Reasons not to believe in ∞. In calculus we will often want to talk about very large and very small quantities. We will use the symbol ∞ (pronounced "infinity") all the time, and the way this symbol is traditionally used would suggest that we are thinking of ∞ as just another number. But ∞ is different. The ordinary rules of algebra don't apply to ∞. As an example of the many ways in which these rules can break down, just think about "$\infty + \infty$." What do you get if you add infinity to infinity? The elementary school argument for finding the sum is: "if you have a bag with infinitely many apples, and you add infinitely many more apples, you still have a bag with infinitely many apples." So, you would think that

$$\infty + \infty = \infty.$$

If ∞ were a number to which we could apply the rules of algebra, then we could cancel ∞ from both sides,

$$\infty + \cancel{\infty} = \cancel{\infty} \implies \infty = 0.$$

So infinity is the same as zero! If that doesn't bother you, then let's go on. Still assuming ∞ is a number we find that

$$\frac{\infty}{\infty} = 1,$$

but also, in view of our recent finding that $\infty = 0$,

$$\frac{\infty}{\infty} = \frac{0}{\infty} = 0.$$

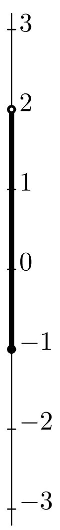
Therefore, combining these last two equations,

$$1 = \frac{\infty}{\infty} = 0.$$

In elementary school terms: "one apple is no apple."

This kind of arithmetic is not going to be very useful for scientists (or grocers), so we need to drop the assumption that led to this nonsense, i.e. we have to agree from here on that [3]

INFINITY IS NOT A NUMBER!



1.6. 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.

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.

---

[3]That is not the end of the story. Twentieth century mathematicians have produced a theory of "non standard real numbers" which includes infinitely large numbers. To keep the theory from running into the kind of nonsense we just produced, they had to assume that there are many different kinds of infinity: in particular $2 \times \infty$ is not the same as ∞, and $\infty \times \infty = \infty^2$ is yet another kind of infinity. Since there are many kinds of infinity in this theory you can't use the single symbol "∞" because it doesn't say ***which*** infinitely large number you would be talking about. In this course we will follow the traditional standard approach, and assume there are no "infinitely large numbers." But if you want to read the version of the theory where infinitely large and small numbers do exist, then you should see Keisler's calculus text at http://www.math.wisc.edu/~keisler/calc.html

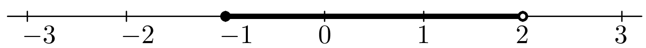


Figure 1. To draw the half open interval $[-1, 2)$ use a filled dot to mark the endpoint that is included and an open dot for an excluded endpoint. Some like to draw the number line vertically, like a thermometer.

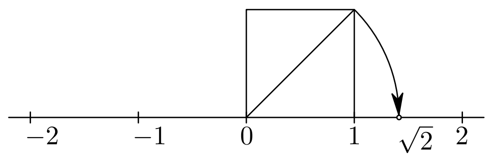


Figure 2. To find $\sqrt{2}$ on the real line you draw a square of side length 1 and drop the diagonal onto the real line.

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

- (a, b) is the set of all real numbers x that satisfy $a < x < b$.
- $[a, b)$ is the set of all real numbers x that satisfy $a \le x < b$.
- $(a, b]$ is the set of all real numbers x that satisfy $a < x \le b$.
- $[a, b]$ is the set of all real numbers x that 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) that are ≤ 2.

1.7. Set notation. A common way of describing a set is to say it is the collection of all real numbers that satisfy a certain condition. We use curly braces $\{\cdot\}$ to denote sets:

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

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

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

The set

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

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. For instance,

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

can't be accurately drawn. In this course we will try to avoid such sets.

Sets can also contain a finite collection of numbers which we can simply list, like

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

so that $\mathcal{D}$ is the set containing the numbers 1, 2, and 3. Or the set

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

which consists of the solutions of the equation $x^3 - 4x^2 + 1 = 0$. (There are three of them, but it is not easy to give a formula for the solutions.)

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

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

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

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

2. Problems

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

2. Which of the following fractions have finite decimal expansions?

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

3. Draw the following sets of real numbers. Each of these sets is the union of one or more intervals. Find those intervals. Which of these sets are finite?

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

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

5. Consider the sets

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

Are these sets the same?

6. **[Group Problem]** Write the numbers

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

as fractions (that is, 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.)

7. **[Group Problem]** **(a)** In §1.5 we agreed that infinitely large numbers don't exist. Do infinitely small numbers exist? In other words, does there exist a ***positive*** number x that is smaller than $\frac{1}{n}$ for all $n = 1, 2, 3, 4, \cdots$, i.e.

$$0 < x < \tfrac{1}{2}, \text{ and } 0 < x < \tfrac{1}{3}, \text{ and}$$
$$0 < x < \tfrac{1}{4}, \text{ and so on} \cdots ?$$

(b) Is the number whose decimal expansion after the period consists only of nines, i.e.

$$a = 0.99999999999999999\ldots$$

the same as the number 1? Or could it be that there are numbers between $0.9999\cdots$ and 1? That is, is it possible that there is some number x that satisfies

$$0.999999\cdots < x < 1\ ?$$

(c) Here is a very similar question: is

$$b = 0.3333333333333333333\ldots$$

the same as $\frac{1}{3}$? Or could there be a number x with

$$0.333333\cdots < x < \tfrac{1}{3}?$$

8. In §1.7 we said that the set

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

was difficult to draw. Explain why.

3. Functions

3.1. Dependence. Calculus deals with quantities that change. For instance, the water temperature T of Lake Mendota (as measured at the pier near the Memorial Union) is a well-defined quantity, but it changes with time. At each different time t we will find a different temperature T. Therefore, when we say "the temperature at the pier of Lake Mendota," we could mean two different things:

- On one hand we could mean the "temperature at some given time," e.g. the temperature at 3pm is 68F: here the temperature is just a number. The most common notation for this is $T(3) = 68$, or $T(3\text{pm}) = 68\text{F}$.
- On the other hand we could mean the "temperature in general," i.e. the temperatures at all times. In that second interpretation the temperature is not just a number, but a whole collection of numbers, listing ***all*** times t and the corresponding temperatures $T(t)$.

So $T(t)$ is a number while T by itself is not a number, but a more complicated thing. It is what in mathematics is called a ***function***. We say that *the water temperature is a function of time.*

Here is the definition of what a mathematical function is:

3.2. Definition. *To specify a* ***function*** f *you must*

(1) *give a* ***rule*** *that tells you how to compute the value* $f(x)$ *of the function for a given real number* x*, and:*
(2) *say for which real numbers* x *the rule may be applied.*

The set of numbers for which a function is defined is called its ***domain****. The set of all possible numbers* $f(x)$ *as* x *runs over the domain is called the* ***range*** *of the function. The rule must be* ***unambiguous:*** *the same* x *must always lead to the same* $f(x)$*.*

For instance, one can define a function f by putting $f(x) = 3x$ for all $x \geq 0$. Here the rule defining f is "multiply by 3 whatever number you're given", and the function f will accept all real numbers.

The rule that specifies a function can come in many different forms. Most often it is a formula, as in the square root example of the previous paragraph. Sometimes you need 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 whose definition involves different formulas on different intervals are sometimes called ***piecewise defined functions.***

3.3. Graphing a function. You get the ***graph of a function*** f by drawing all points whose coordinates are (x, y) where x is in the domain of f and $y = f(x)$.

3.4. Linear functions. A function f that is given by the formula

$$f(x) = mx + n$$

where m and n are constants is called a ***linear function***. Its graph is a straight line. The constants m and n are the ***slope*** and y***-intercept*** of the line, respectively. Conversely, any straight line which is not vertical (not parallel to the y-axis) is the graph of a linear

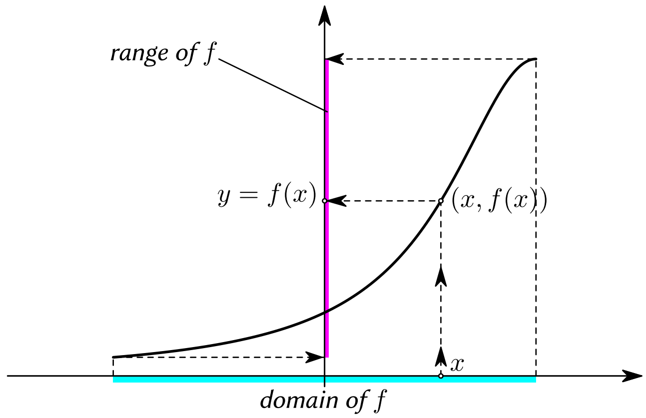


Figure 3. The graph of a function f. The domain of f consists of all x values at which the function is defined, and the range consists of all possible values f can have.

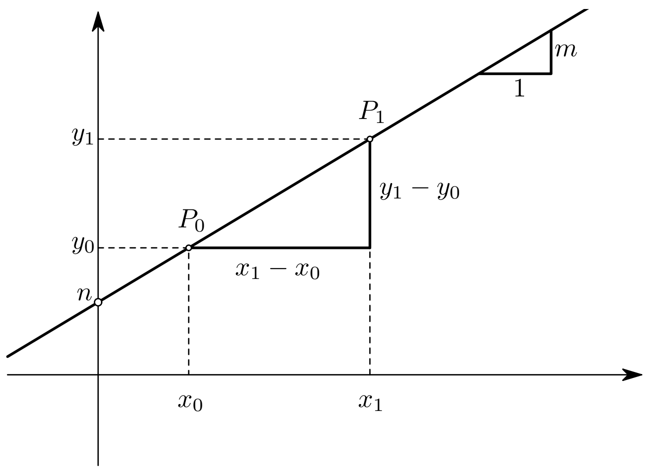


Figure 4. The graph of $f(x) = mx + n$ is a straight line. It intersects the y-axis at height n. The ratio between the amounts by which y and x increase as you move from one point to another on the line is $\frac{y_1 - y_0}{x_1 - x_0} = m$. This ratio is the same, no matter how you choose the points P_0 and P_1 as long as they are different and on the line.

function. If we know two points (x_0, y_0) and (x_1, y_1) on the line, then we 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 from Euclidean geometry, namely, it says that the ratio

$$(y_1 - y_0) : (x_1 - x_0)$$

is the same for every pair of distinct points (x_0, y_0) and (x_1, y_1) that you could pick on the line.

3.5. Domain and "biggest possible domain." In this course we will usually not be careful about specifying the domain of a function. When this happens the domain is understood to be the set of all x for which the rule that tells you how to compute $f(x)$ is a meaningful real number. 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 a well-defined real number for all $x \geq 0$ and not a real number 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 $f(x) = y$ to obtain at least one (real) value of x.

3.6. 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

$$\text{"all real numbers except 0"} = \{x \mid x \neq 0\} = (-\infty, 0) \cup (0, \infty).$$

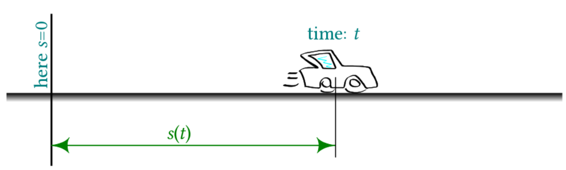
To find the range we ask "for which y can we solve the equation $y = f(x)$ for x," i.e. for which y can we solve $y = 1/x^2$ for x?

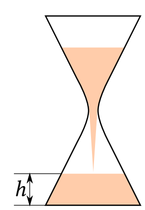
If $y = 1/x^2$ then we must have $x^2 = 1/y$, so first of all, since we have to divide by y, y can't be zero. Furthermore, $1/y = x^2$ says that y must be positive. On the other hand, if $y > 0$ then $y = 1/x^2$ has a solution (in fact two solutions), namely $x = \pm 1/\sqrt{y}$. This shows that the range of f is

$$\text{"all positive real numbers"} = \{x \mid x > 0\} = (0, \infty).$$

3.7. 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 $s(t)$ 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 at time t and the marker.





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

Given t, weigh the mouse at time t.

Here is another example: suppose you are given an hourglass. If you turn it over, then sand will pour from the top part to the bottom part. At any time t you could measure the height of the sand in the bottom and call it $h(t)$. Then, as in the previous examples, you can say that the height of the sand is a function of time. But in this example you can let the two variables height and time switch roles: given a value for h you wait until the pile

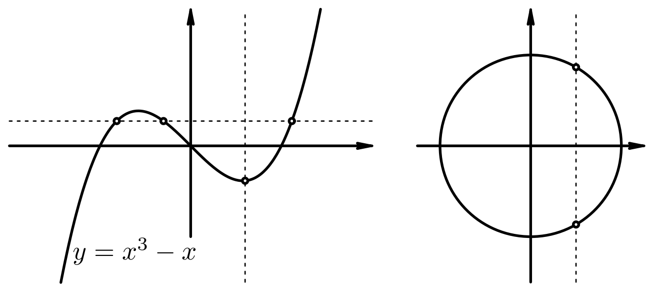


Figure 5. The graph of $y = x^3 - x$ fails the "horizontal line test," but it passes the "vertical line test." The circle fails both tests.

of sand in the bottom has reached height h and check what time it is when that happens: the resulting time $t(h)$ is determined by the specified height h. In this way you can regard time as a function of height.

3.8. The Vertical Line Property. Generally speaking, graphs of functions are curves in the plane but they distinguish themselves from arbitrary curves by the way they intersect vertical lines: ***The graph of a function cannot intersect a vertical line "$x =$ constant" in more than one point.*** The reason why this is true is very simple: if two points lie on a vertical line, then they have the same x coordinate, so if they also lie on the graph of a function f, then their y-coordinates must both be equal to $f(x)$, so in fact they are the same point.

3.9. Example – a cubic function. The graph of $f(x) = x^3 - x$ "goes up and down," and, even though it intersects several horizontal lines in more than one point, it intersects every vertical line in exactly one point. See Figure 5.

3.10. Example – a circle is not a graph of a function $y = f(x)$. The collection of points 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 $P_2(0, -1)$. See again Figure 5. This example continues in § 4.3 below.

4. Implicit functions

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

4.1. Example. We can define a function f by saying that if x is any given number, then $y = f(x)$ is the solution of the equation

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

In this example we 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) "$y = f(x)$ is defined by $x^2 + 2y - 3 = 0$," and
(ii) "f is defined by $f(x) = (3 - x^2)/2$."

The first definition is an implicit definition, the second is explicit. This example shows that with an "implicit function" it is not the function itself, but rather the way it was defined that is implicit.

4.2. 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 you can then solve for y, and you find 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.

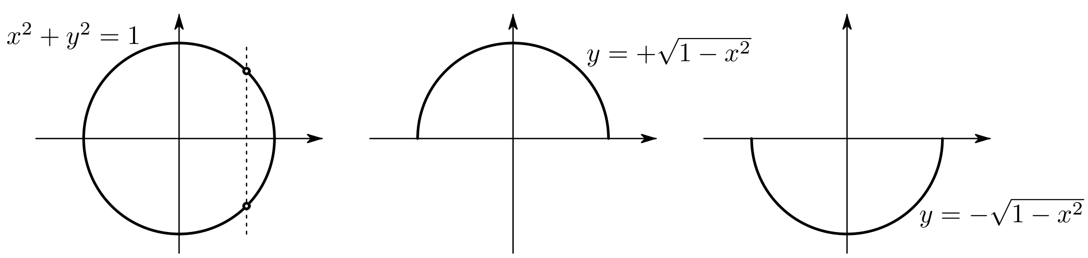


Figure 6. The circle determined by $x^2 + y^2 = 1$ is not the graph of a function, but it contains the graphs of the two functions $h_1(x) = \sqrt{1 - x^2}$ and $h_2(x) = -\sqrt{1 - x^2}$.

4.3. Example: the equation alone does not determine the function. We saw in § 3.10 that the unit circle is not the graph of a function (because it fails the vertical line test). What happens if you ignore this fact and try to use the equation $x^2 + y^2 = 1$ for the circle to define a function anyway? To find out, suppose we 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, it has ***two*** solutions:

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

The rule that 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 $-1 < x < 1$.

Strictly speaking, the domain of the function that is defined implicitly by the equation $x^2 + y^2 = 1$ consists of only two points, namely $x = \pm 1$. Why? Well, those are the only two values of x for which the equation has exactly one solution y (the solution is $y = 0$.)

To see this in the picture, look at Figure 6 and find all vertical lines that intersect the circle on the left exactly once.

To get different functions that are described by the equation $x^2 + y^2 = 1$, we have to specify for each x which of the two solutions $\pm\sqrt{1-x^2}$ we declare to be "$f(x)$". This leads to many possible choices. Here are three of them:

$$h_1(x) = \text{the non negative solution } y \text{ of } x^2 + y^2 = 1$$
$$h_2(x) = \text{the non positive solution } y \text{ of } x^2 + y^2 = 1$$
$$h_3(x) = \begin{cases} h_1(x) & \text{when } x < 0 \\ h_2(x) & \text{when } x \geq 0 \end{cases}$$

There are many more possibilities.

4.4. Why and when do we use implicit functions? Three examples. 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, even if it is possible, then the implicit description can still be much simpler than the explicit formula.

As a first example, define a function f by saying that $y = f(x)$ if and only if y is the largest of the solutions of

$$y^2 + 3y + 2x = 0. \tag{1}$$

This means that the recipe for computing $f(x)$ for any given x is "solve the equation $y^2+3y+2x = 0$ for y and, if the solutions are real numbers, then set y equal to the largest solution you find." For example, to compute $f(0)$, we set $x = 0$ and solve $y^2 + 3y = 0$. By factoring $y^2 + 3y = (y + 3)y$, we find that the solutions are $y = 0$ and $y = -3$. Since $f(0)$ is defined to be the *largest* of the solutions, we get $f(0) = 0$. Similarly, to compute $f(1)$ we solve $y^2 + 3y + 2 \cdot 1 = 0$: the solutions are $y = -1$ and $y = -2$, so $f(1) = -1$. For any other $x \leq \frac{9}{8}$ the quadratic formula tells us that the solutions are

$$y = \frac{-3 \pm \sqrt{3^2 - 4 \cdot 2x}}{2} = \frac{-3 \pm \sqrt{9-8x}}{2}.$$

By definition $f(x)$ is the largest solution, so

$$f(x) = -\frac{3}{2} + \frac{1}{2}\sqrt{9-8x}\,.$$

If you don't like square roots, then the equation (1) looks a lot simpler than this formula, and you would prefer to work with (1).

For a more extreme example, suppose you were asked to work with a function g defined implicitly by

$$y = g(x) \text{ if and only if } y^3 + 3y + 2x = 0. \tag{2}$$

This equation is a cubic equation and it is much harder to solve than the quadratic equation we had before. It turns out that for any value of x, there is exactly one real value of y for which $y^3 + 3y + 2x = 0$, and the solution was found in the early 1500s by Cardano and Tartaglia [4]. Here it is :

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

[4]The solution was actually found by Tartaglia and, according to some, stolen by Cardano. To see the solution and its history you can check the internet, and, in particular, the Wikipedia pages on Cardano and Tartaglia.

Don't worry about how this formula came about; let's just trust Cardano and Tartaglia. The implicit description (2) 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.

Finally, you could have been given the function h whose definition is

$$y = h(x) \text{ if and only if } \sin(y) + 3y + 2x = 0. \tag{3}$$

There is no formula involving only standard functions (exponents, trig and inverse trig functions, logarithms, etc. for the solution to this equation. Nonetheless it turns out that no matter how you choose x, the equation $\sin(y) + 3y + 2x = 0$ has exactly one solution y; in fact, you will prove this in Problem 12.54. So the function h is well defined, but for this function the implicit description is the only one available.

5. Inverse functions

If we have a function f, which takes input values and sends them to an output, we might want to try to define a function f^{-1} which "undoes" f, by the following prescription:

For any given x we say that $y = f^{-1}(x)$
if y is the solution of $f(y) = x$. (4)

Note that x and y have swapped their usual places in this last equation!

The prescription (4) defines the inverse function f^{-1}, but it does not say what the domain of f^{-1} is. By definition, *the domain of f^{-1} consists of all numbers x for which the equation $f(y) = x$ has* ***exactly one*** *solution.* ‘ So if for some x the equation $f(y) = x$ has no solution y, then that value of x does not belong to the domain of f^{-1}.

If, on the other hand, for some x the equation $f(y) = x$ has *more than one* solution y, then the prescription (4) for computing $f^{-1}(x)$ is ambiguous: which of the solutions y should be $f^{-1}(x)$? When this happens we throw away the whole idea of finding the inverse of the function f, and we say that the inverse function f^{-1} is undefined ("the function f has no inverse".)

5.1. Example – inverse of a linear function. Consider the function f with $f(x) = 2x + 3$. Then the equation $f(y) = x$ works out to be

$$2y + 3 = x$$

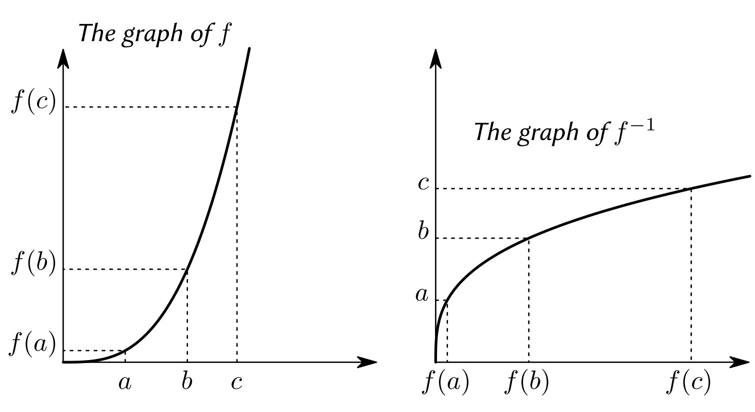


Figure 7. The graph of a function and its inverse are mirror images of each other. Can you draw the mirror?

and this has the solution

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

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

5.2. Example – inverse of $f(x) = x^2$. It is often said that "the inverse of x^2 is $\sqrt{x}$." This is not quite true, as you'll see in this and the next example.

Let f be the function $f(x) = x^2$ with domain all real numbers. What is f^{-1}?

The equation $f(y) = x$ is in this case $y^2 = x$. When $x > 0$ the equation has two solutions, namely $y = +\sqrt{x}$ and $y = -\sqrt{x}$. According to our definition, the function f does not have an inverse.

5.3. Example – inverse of x^2, again. Consider the function $g(x) = x^2$ with domain all ***positive*** real numbers. To see for which x the inverse $g^{-1}(x)$ is defined 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 no solutions since $y^2 \geq 0$ for all f. But if $x \geq 0$ then $y^2 = x$ does have a solution, namely $y = \sqrt{x}$.

So we see that $g^{-1}(x)$ is defined for all positive real numbers x, and that it is given by $g^{-1}(x) = \sqrt{x}$.

This example is shown in Figure 7. See also Problem 7.6.

6. Inverse trigonometric functions

The two most important inverse trigonometric functions are the ***arcsine*** and the ***arctangent***. The most direct definition of these functions is given in Figure 8. In words, $\theta = \arcsin x$ is the angle (in radians) whose sine is x. If $-1 \leq x \leq 1$ then there always is such an angle, and, in fact, there are many such angles. To make the definition of $\arcsin x$

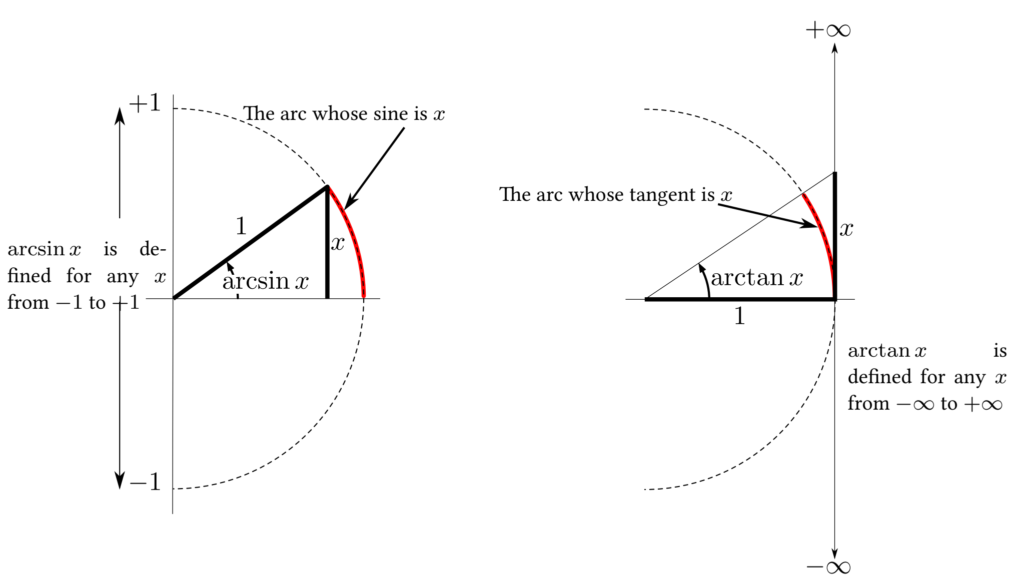


Figure 8. Definition of $\arcsin x$ and $\arctan x$. The dotted circles are unit circles. On the left a segment of length x and its arc sine are drawn. The length of the arc drawn on the unit circle is the subtended angle in radians, i.e. $\arcsin x$. So $\arcsin x$ is the length of "the arc whose sine is x."

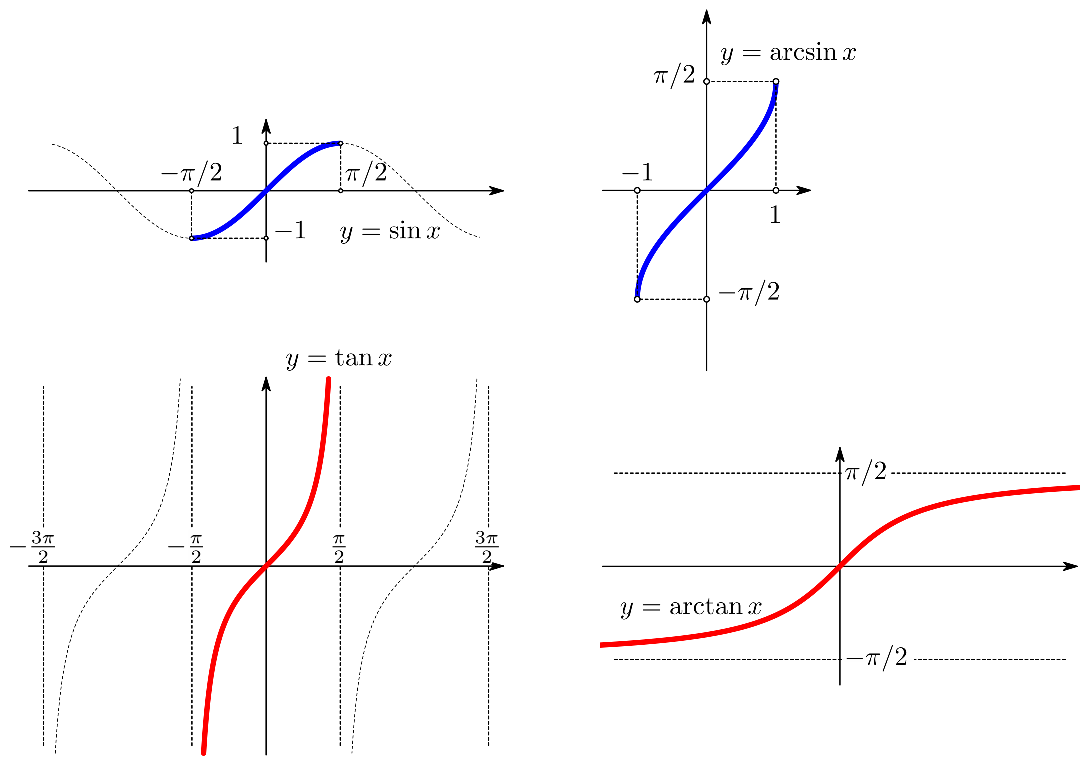


Figure 9. The graphs of the sine and tangent functions on the left, and their inverses, the arcsine and arctangent on the right. Note that the graph of arcsine is a mirror image of the graph of the sine, and that the graph of arctangent is a mirror image of the graph of the tangent.

unambiguous we always choose θ to be the angle that lies between $-\frac{\pi}{2}$ and $+\frac{\pi}{2}$. To see where the name "arcsine" comes from, look at Figure 8 on the left.

An equivalent way of defining the arcsine and arctangent is to say that they are the inverse functions of the sine and tangent functions on a restricted domain. E.g. if $y = f(x) = \sin x$, then the inverse of the function f is by definition (see (4)) the function f^{-1} with the property that

$$y = f^{-1}(x) \iff x = f(y) = \sin y.$$

If we restrict y to the interval $-\frac{\pi}{2} \le y \le \frac{\pi}{2}$ then this is just the definition of $\arcsin x$, so

$$y = \sin x \iff x = \arcsin y, \quad \text{provided } -\tfrac{\pi}{2} \le x \le \tfrac{\pi}{2}.$$

Likewise,

$$y = \tan x \iff x = \arctan y, \quad \text{provided } -\tfrac{\pi}{2} \le x \le \tfrac{\pi}{2}.$$

Forgetting about the requirement that $-\frac{\pi}{2} \le x \le \frac{\pi}{2}$ can lead to unexpected mistakes (see Problem 7.8).

Because of the interpretation of $y = \arcsin x$ as the inverse of the sine function, the notations

$$\arcsin x = \sin^{-1} x, \qquad \arctan x = \tan^{-1} x$$

are very commonly used.

In addition to the arcsine and arctangent, people have also defined the arccosine, the arcsecant and the arccosecant. However, because they can all be expressed in terms of the arcsine and arctangent, we will not bother with them.

7. Problems

1. The functions f and g are defined by

$$f(x) = x^2 \text{ and } g(s) = s^2.$$

Are f and g the same functions or are they different?

2. Find a formula for the function f that is defined by

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

What is the domain of f?

3. Find a formula for the function f that is defined by the requirement that for any x one has

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

What is the domain of f?

4. Let f be the function defined by the requirement that for any x one has

$$y = f(x) \iff \begin{array}{l} y \text{ is the largest of all} \\ \text{possible solutions of} \\ y^2 = 3x^2 - 2xy. \end{array}$$

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

5. Find a formula for the function f that is defined by

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

Find the domain of f.

6. (continuation of example 5.3.) Let k be the function with $k(x) = x^2$ whose domain is all *negative* real numbers. Find the domain of k^{-1}, and draw the graph of k^{-1}.

7. Use a calculator to compute $g(1.2)$ in three decimals where g is the implicitly defined function from §4.4. (There are (at least) two different ways of finding $g(1.2)$)

8. **[Group Problem] True or false?**

(a) For all real numbers x one has

$$\sin\bigl(\arcsin x\bigr) = x.$$

(b) For all real numbers x one has

$$\arcsin\bigl(\sin x\bigr) = x.$$

(c) For all real numbers x one has

$$\arctan\bigl(\tan x\bigr) = x.$$

(d) For all real numbers x one has

$$\tan\bigl(\arctan x\bigr) = x.$$

9. On a graphing calculator plot the graphs of the following functions, and explain the results. (Hint: first do the previous exercise.)

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

10. Find the inverse of the function f that is given by $f(x) = \sin x$ and ***whose domain is*** $\pi \le x \le 2\pi$. Sketch the graphs of both f and f^{-1}.

11. Find a number a such that the function $f(x) = \sin(x + \pi/4)$ with domain $a \le x \le a + \pi$ has an inverse. Give a formula for $f^{-1}(x)$ using the arcsine function.

12. Simplicio has found a new formula for the arcsine. His reasoning is as follows:

Since everybody writes "the square of $\sin y$*" as*

$$\bigl(\sin y\bigr)^2 = \sin^2 y.$$

we can replace the 2*'s by* -1*'s and we get*

$$\arcsin y = \sin^{-1} y = \bigl(\sin y\bigr)^{-1} = \frac{1}{\sin y}.$$

Is Simplicio right or wrong? Explain your opinion.

13. Draw the graph of the function h_3 from §4.3.

14. A function f is given that satisfies

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

for all real numbers x.

If x and y are arbitrary real numbers then compute

(a) $f(0)$

(b) $f(3)$

(c) $f(\pi)$

(d) $f(t)$

(e) $f(x)$

(f) $f(f(2))$

(g) $f(2f(x))$

15. A function f is given that satisfies

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

for all real numbers x.

If x and t are arbitrary real numbers, then compute the following quantities:

(a) $f(1)$

(b) $f(0)$

(c) $f(t)$

(d) $f(x)$

(e) $f(f(2))$

(f) $f(2f(x))$

16. Does there exist a function f that satisfies

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

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

* * *

The following exercises review precalculus material involving quadratic expressions $ax^2 + bx + c$ in one way or another.

17. Find the vertex (h, k) of the parabola $y = ax^2 + bx + c$. Use the result to find the range of this function. Note that the behavior depends on whether a is positive or negative. (Hint: "Complete the square" in the quadratic expression by writing it in the form $y = a(x-h)^2 + k$ for some h and k in terms of a, b, and c.)

18. Find the ranges of the following functions:

$$\begin{aligned} f(x) &= 2x^2 + 3 \\ g(x) &= -2x^2 + 4x \\ h(x) &= 4x + x^2 \\ k(x) &= 4\sin x + \sin^2 x \\ \ell(x) &= 1/(1+x^2) \\ m(x) &= 1/(3 + 2x + x^2). \end{aligned}$$

19. **[Group Problem]** For each real number a we define a line ℓ_a with equation $y = ax + a^2$.

(a) Draw the lines corresponding to $a = -2, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, 2$.

(b) Does the point with coordinates $(3, 2)$ lie on one or more of the lines ℓ_a (where a can be any number, not just the five values from part (a))? If so, for which values of a does $(3, 2)$ lie on ℓ_a?

(c) Which points in the plane lie on at least one of the lines ℓ_a?.

20. For which values of m and n does the graph of $f(x) = mx+n$ intersect the graph of $g(x) = 1/x$ in exactly one point and also contain the point $(-1, 1)$?

21. For which values of m and n does the graph of $f(x) = mx + n$ ***not*** intersect the graph of $g(x) = 1/x$?

CHAPTER 2

Derivatives (1)

To work with derivatives we first have to know what a limit is, but to motivate why we are going to study limits we essentially need to explain what a derivative is. In order to resolve this potentially circular logic, we will first motivate the idea of a derivative in this short chapter, then discuss how to compute limits in detail in the next chapter, and then we will return and discuss derivatives armed with a better understanding of limits.

Let's first look at the two classical problems that gave rise to the notion of a derivative: finding the equation of the line tangent to a curve at a point, and finding the instantaneous velocity of a moving object.

1. The tangent line to a curve

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

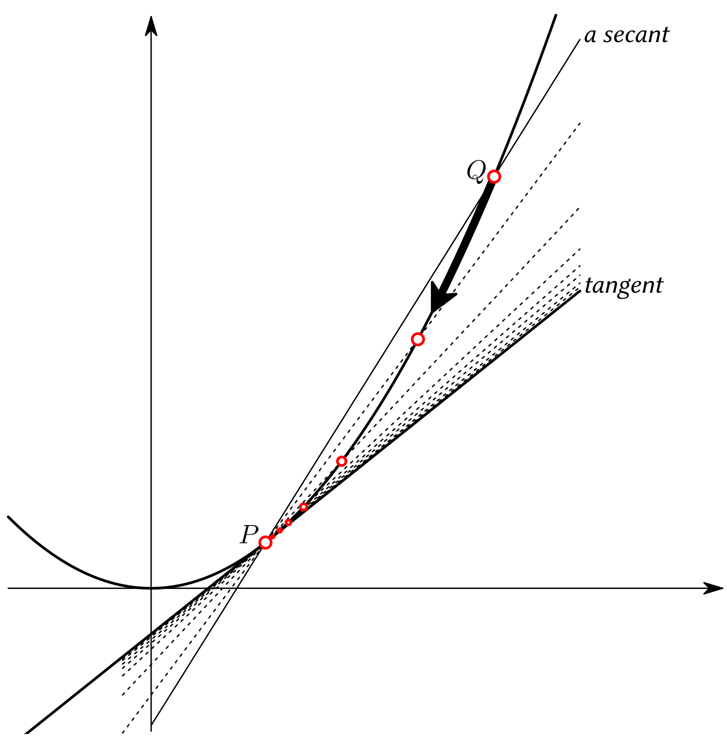


Figure 1. Constructing the tangent by letting $Q \to P$

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 (at least, nowhere else near P).

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 line", and it won't precisely be the tangent line, but if you choose Q to be very close to P then the secant line will be close to the tangent line.

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, since we need ***two*** points to determine a line. If we choose Q different from P then we won'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 needed to clarify this issue.

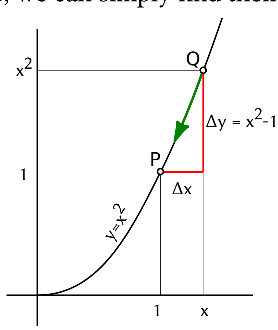
2. An example – tangent to a parabola

To make things more concrete, let us take the function $f(x) = x^2$, and attempt to find the equation of the tangent line to $y = f(x)$ at the point $P = (1, 1)$. 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 equations of the secant and tangent lines, we can simply find their slopes.



Finding the slope of the tangent at P

To find the slope of the tangent at P(1,1) pick another point Q (x, x^2) on the parabola and compute the slope m_{PQ} of the line connecting P and Q.

Then let Q approach P and see what happens to the slope m_{PQ}.

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, although we are not allowed to set $x = 1$ because P and Q have to be different points. By the "rise over run" formula, the slope of the secant line joining P and Q is

$$m_{PQ} = \frac{\Delta y}{\Delta x} \quad \text{where} \quad \Delta y = x^2 - 1 \quad \text{and} \quad \Delta x = x - 1.$$

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

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

A warning: we cannot substitute $x=1$ in equation (5) to get (6), even though it looks like that's what we did. The reason why we 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

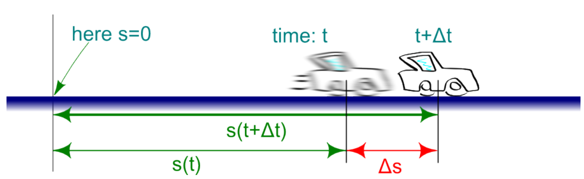
$$m_{PQ} = \frac{\Delta y}{\Delta x} = \frac{0}{0} = \text{ undefined.}$$

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 mean anything for $x=1$?

We did something more complicated than just setting $x=1$: 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 essence of a limit, and we'll study these ideas in detail soon.

3. Instantaneous velocity

When you are riding in a car, the speedometer tells you how fast you are going, i.e. what your velocity is. But what exactly does it mean when the speedometer says your car is traveling at a speed of (say) 50 miles per hour?



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 then 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 $\frac{0}{0}$, which is 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 our distance from our starting point (in miles).

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 are $s(t)$ miles from our start. A little later, at time $t+\Delta t$ we are $s(t+\Delta t)$ miles from our start. Therefore, during the time interval from t to $t+\Delta t$, we have moved

$$s(t+\Delta t) - s(t) \text{ miles,}$$

and therefore our average velocity in that time interval was

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

The shorter we make the time interval (the smaller we 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

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

4. 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 we 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 we can do with them are what the first portion of this course is about. The description we just went through shows that to understand what a derivative is, we need to understand more about this "limiting process" so that we can have a concrete understanding of statements like (9).

5. Examples of rates of change

5.1. Acceleration as the rate at which velocity changes. As you are driving in your car your velocity may change over time. Suppose $v(t)$ is your velocity at time t (measured in miles per hour). You could try to figure out how fast your velocity is changing by measuring it at one moment in time (you get $v(t)$), then measuring it a little later (you get $v(t+\Delta t)$)). You conclude that your velocity increased by $\Delta v = v(t+\Delta t) - v(t)$ during a time interval of length Δt, and hence

$$\left\{\begin{array}{c}\text{average rate at}\\ \text{which your}\\ \text{velocity changed}\end{array}\right\} = \frac{\text{change in velocity}}{\text{duration of time interval}} = \frac{\Delta v}{\Delta t} = \frac{v(t+\Delta t) - v(t)}{\Delta t}.$$

This rate of change is called your *average acceleration* (over the time interval from t to $t + \Delta t$). Your *instantaneous acceleration* at time t is the limit of your average acceleration as you make the time interval shorter and shorter:

$$\{\text{acceleration at time } t\} = a = \lim_{\Delta t \to 0} \frac{v(t+\Delta t) - v(t)}{\Delta t}.$$

The average and instantaneous accelerations are measured in "miles per hour per hour":

$$(\mathrm{mi/h})/\mathrm{h} = \mathrm{mi/h}^2.$$

Or, if you had measured distances in *meters* and time in *seconds* then velocities would be measured in *meters per second*, and acceleration in *meters per second per second*, which is the same as meters per second^2: "meters per squared second".

5.2. Reaction rates. Imagine a chemical reaction in which two substances A and B react in such a way that A converts B into A. The reaction could proceed by

$$\mathrm{A} + \mathrm{B} \longrightarrow 2\mathrm{A}.$$

If the reaction is taking place in a closed reactor, then the "amounts" of A and B will change with time. The amount of B will decrease, while the amount of A will increase. Chemists write [A] for the concentration of "A" in the chemical reactor (measured in moles per liter). We're mathematicians so we will write "$[\mathrm{A}](t)$" for the concentration of A present at time t.

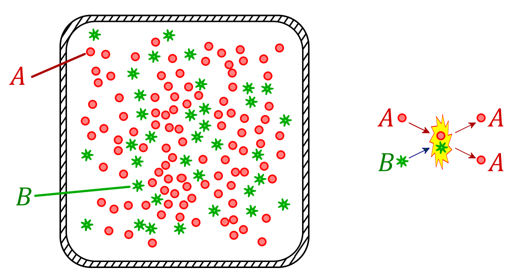


Figure 2. A chemical reaction in which A converts B into A.

To describe how fast the amount of A is changing we consider the derivative of [A] with respect to time:

$$[A]'(t) = \lim_{\Delta t \to 0} \frac{[\mathrm{A}](t+\Delta t) - [\mathrm{A}](t)}{\Delta t}.$$

This quantity is the rate of change of [A]. In chemistry and physics, it is more common to write the derivative in Leibniz notation:

$$\frac{d[\mathrm{A}]}{dt}.$$

How fast does the reaction take place? If you add more A or more B to the reactor then you would expect that the reaction would go faster (i.e., that more A would be produced per second). The law of *mass-action kinetics* from chemistry states this more precisely. For our particular reaction it would say that the rate at which A is consumed is given by

$$\frac{d[\mathrm{A}]}{dt} = k\,[\mathrm{A}]\,[\mathrm{B}],$$

where the constant k is called the *reaction constant*, which you could measure by timing how fast the reaction goes.

6. Problems

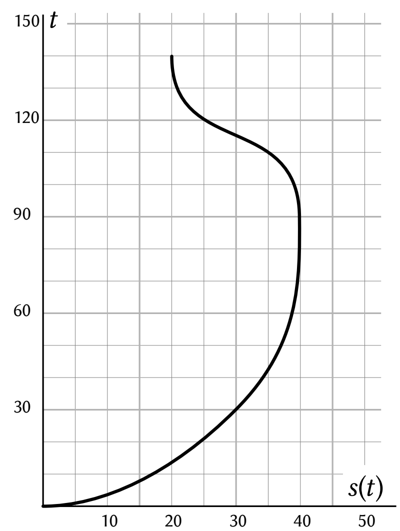
1. Repeat the reasoning in §2 to find the slope of the tangent line at the point $(\frac{1}{3}, \frac{1}{9})$, or more generally at any point (a, a^2) on the parabola with equation $y = x^2$.

2. Repeat the reasoning in §2 to find the slope of the tangent line at the point $(\frac{1}{2}, \frac{1}{8})$, or more generally at any point (a, a^3) on the curve with equation $y = x^3$.

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

$$(a)\ y = x^2 - 2x + 1$$
$$(b)\ y = \frac{1}{x}$$
$$(c)\ y = 2^x$$

4. This figure shows a plot of the distance traveled $s(t)$ (in miles) versus time t (in minutes):



(a) Something is wrong: the curve in the graph obviously doesn't pass the vertical line test, so it cannot be the graph of a function. How can it be the graph of $s(t)$ versus t?

(b) Use the plot to estimate the instantaneous velocity at the following times

t (min)	$v(t)$
30	
60	
90	
120	

Describe in one or two short sentences what you did to find your estimates.

(c) Make a graph of the instantaneous velocity $v(t)$.

5. Look ahead at Figure 3 in the next chapter. What is the derivative of $f(x) =$

$x \cos \frac{\pi}{x}$ at the points A and B on the graph?

6. Suppose that some quantity y is a function of some other quantity x, and suppose that y is a mass (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?

7. A tank is filling with water. The volume (in gallons) of water in the tank at time t (seconds) is $V(t)$. What units does the derivative $V'(t)$ have?

8. [Group Problem] Let $A(x)$ be the area of an equilateral triangle whose sides measure x inches.

(a) Show that $\frac{dA}{dx}$ has the units of a length.

(b) Which length does $\frac{dA}{dx}$ represent geometrically? [Hint: draw two equilateral triangles, one with side x and another with side $x + \Delta x$. Arrange the triangles so that they both have the origin as their lower left hand corner, and so their bases are on the x-axis.]

9. [Group Problem] Let $A(x)$ be the area of a square with side x, and let $L(x)$ be the perimeter of the square (sum of the lengths of all its sides). Using the familiar formulas for $A(x)$ and $L(x)$ show that $A'(x) = \frac{1}{2}L(x)$.

Give a geometric interpretation that explains why $\Delta A \approx \frac{1}{2}L(x)\Delta x$ for small Δx.

10. Let $A(r)$ be the area enclosed by a circle of radius r, and let $L(r)$ be the circumference of the circle. Show that $A'(r) = L(r)$. (Use the familiar formulas from geometry for the area and circumference of a circle.)

11. Let $V(r)$ be the volume enclosed by a sphere of radius r, and let $S(r)$ be the its surface area.

(a) Show that $V'(r) = S(r)$. (Use the formulas $V(r) = \frac{4}{3}\pi r^3$ and $S(r) = 4\pi r^2$.)

(b) Give a geometric explanation of the fact that $\frac{dV}{dr} = S$.

[Hint: to visualize what happens to the volume of a sphere when you increase the radius by a very small amount, imagine the sphere is the Earth, and you increase the radius by covering the Earth with a layer of water that is 1 inch deep. How much does the volume increase? What if the depth of the layer was "Δr"?]

12. [Group Problem] *Should you trust your calculator?*

Find the slope of the tangent to the parabola

$$y = x^2$$

at the point $(\frac{1}{3}, \frac{1}{9})$ (You have already done this: see exercise 6.1).

Instead of doing the algebra you could try to compute the slope by using a calculator. This exercise is about how you do that and what happens if you try (too hard).

Compute $\frac{\Delta y}{\Delta x}$ for various values of Δx:

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

As you choose Δx smaller your computed $\frac{\Delta y}{\Delta x}$ ought to get closer to the actual slope. Use at least 10 decimals and organize your results in a table like this: Look carefully at the ratios $\Delta y/\Delta x$. Do they look like they are converging to some number? Compare the values of $\frac{\Delta y}{\Delta x}$ with the true value you got in the beginning of this problem.

Δx	$a + \Delta x$	$f(a + \Delta x)$	Δy	$\Delta y / \Delta x$
0.1				
0.01				
0.001				
10-6				
10-12				

Table 1. The table for Problem 6.12. Approximate the derivative of $f(x) = x^2$ at $a = 1/3 \approx 0.333\,333\,333\,333$ by computing $\big(f(a + \Delta x) - f(a)\big)/\Delta x$ for smaller and smaller values of Δx. You know from Problem 6.1 that the derivative is $f'(a) = 2a = 2/3$. Which value of Δx above gives you the most accurate answer?

CHAPTER 3

Limits and Continuous Functions

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 is more complicated. In this chapter we will take three approaches to defining the limit. The first definition (§1) appeals to intuition and puts in words how most people think about limits. Unfortunately this definition contains language that is ambiguous and the more you think about it the more you realize that it actually doesn't mean anything. It is also not clear enough to answer all questions that come up about limits.

Most of the calculus you'll see in this semester was essentially invented in the 17th century, but the absence of a good definition of what derivatives and limits are led to centuries of confused arguments between experts. These more or less ended when a precise definition was developed in the late 19th century, 200 years after calculus was born! In §3 you'll see this precise definition. It runs over several terse lines, and unfortunately most people don't find it very enlightening when they first see it. The third approach we will take is the ***axiomatic approach***: instead of worrying about the details of what a limit is, we use our intuition and in §6 we write down a number of properties that we believe the limit should have. After that we try to base all our reasoning on those properties.

1. Informal definition of limits

1.1. Definition of limit (1st attempt). *If f is some function then*

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

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

The following alternative notation is sometimes used

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

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

Note that in the definition we require x to approach a without ever becoming equal to a. It's important that x never actually equals a because our main motivation for looking at limits was the definition of the derivative. In Chapter II, equation (9) we defined the derivative of a function as a limit in which some number Δx goes to zero:

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

The quantity whose limit we want to take here is not even defined when $\Delta x = 0$. Therefore any definition of limit we come up with had better not depend on what happens at $\Delta x = 0$; likewise, the limit $\lim_{x \to a} f(x)$ should not depend on what $f(x)$ does at $x = a$.

1.2. Example. If $f(x) = x + 3$ then

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

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

1.3. A complaint. Our first definition relies heavily on the phrase "gets close to" or "gets closer and closer to". What does this mean? When x gets closer and closer to a without ever being equal to a, how long does this take? ("are we there yet?") How close is close enough? Is it enough for x and a to be the same in five decimals? Fifty decimals? It is hard to answer these questions without generating new ones.

If we want to deal with limits with some measure of confidence that what we are doing isn't ultimately nonsense, then we will need a better definition of limit. Before going into that, let's look at a practical approach to finding limits. To compute $\lim_{x\to a} f(x)$ we need to let x get closer and closer to a; we don't really know how to do that, but we can certainly grab a calculator or a computer and compute $f(x)$ for several values of x that are somewhat close to a (a to two decimals, a to three decimals, etc.) If the values of $f(x)$ then begin to look like some fixed number we could guess that that number is the limit [1].

Here are two examples where we try to find a limit by calculating $f(x)$ for a few values of x:

1.4. Example: substituting numbers to guess a limit. What (if anything) is

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

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

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

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

which does not exist. Next we try to substitute values of x close but not equal to 2. Table 1 suggests that $f(x)$ approaches 0.5.

x	$f(x)$
3.000000	0.600000
2.500000	0.555556
2.100000	0.512195
2.010000	0.501247
2.001000	0.500125
↓	↓
2	limit?

x	$g(x)$
1.000000	1.009990
0.500000	1.009980
0.100000	1.009899
0.010000	1.008991
0.001000	1.000000
↓	↓
0	limit?

Table 1. Finding limits by substituting values of x "close to a." (Values of $f(x)$ and $g(x)$ rounded to six decimals.)

[1] This idea of making better and better approximations is actually a common approach in modern science: if you have a complicated problem (such as predicting tomorrow's weather, or the next three decades' climate) that you cannot solve directly, one thing often tried is changing the problem a little bit, and then making approximations with a computer. This activity is called ***Scientific Computation***, a very active branch of modern mathematics.

1.5. Example: substituting numbers can suggest the wrong answer. Suppose we had the function

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

and we want to find the limit $\lim_{x\to 0} g(x)$.

Then substitution of some "small values of x" could lead us to believe that the limit is $1.000\ldots$. But in fact, if we substitute sufficiently small values, we will see that the limit is 0 (zero)! As you see from this example, there's more to evaluating limits than just typing numbers into the computer and hitting return. See also problem 6.12.

2. Problems

1. Guess $\lim_{x\to 2} x^{10}$. Then try using the first definition of the limit to show that your guess is right.

2. Use a calculator to guess the value of

$$\lim_{x\to 0} (1+x)^{1/x}.$$

3. Is the limit $\lim_{x\to 0}(1+0.693x)^{1/x}$ an integer? Give reasons for your answer, and compare with your neighbor's answer.

4. Simplicio computed $2^{-10} \approx 0.001$ which is very close to zero. He therefore concludes that if x is very close to 2, then x^{-10} is very close to zero, so that, according to Simplicio, it is clearly true that

$$\lim_{x\to 2} x^{-10} = 0.$$

Comment on Simplicio's reasoning: do you agree with his answer? Do you agree with his reasoning?

5. Use a calculator to guess

(a) $\displaystyle\lim_{x\to 1} \frac{x^{100}}{1.01 + x^{100}}$.

(b) $\displaystyle\lim_{x\to 1} \frac{1.01 - x^{100}}{x^{100}}$.

(c) $\displaystyle\lim_{x\to 1} \frac{x^{100}}{1.01 - x^{100}}$.

3. The formal, authoritative, definition of limit

Our attempted definitions of the limit uses undefined and non-mathematical phrases like "closer and closer". In the end we don't really know what those statements really mean, although they are suggestive. Fortunately, there is a good definition, one that is unambiguous and can be used to settle any dispute about the question of whether $\lim_{x\to a} f(x)$ equals some number L or not. Here is the definition.

3.1. Definition of $\lim_{x\to a} f(x) = L$. *If $f(x)$ is a function defined for all x in some interval which contains a, except possibly at $x = a$, then we say that L is the limit of $f(x)$ as $x \to a$, if for every $\varepsilon > 0$, we can find a $\delta > 0$ (depending on ϵ) such that for all x in the domain of f it is true that*

$$0 < |x - a| < \delta \text{ implies } |f(x) - L| < \varepsilon. \tag{10}$$

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 $|x-y|$. The inequality $|x - y| < \delta$ says that "the distance between x and y is less than δ," or that "x and y are closer than δ."

What are ε and δ? The quantity ε is how close you would like $f(x)$ to be to its limit L; the quantity δ is how close you have to choose x to a to achieve this. To prove that $\lim_{x\to a} f(x) = L$ you must assume that someone has given you an unknown $\varepsilon > 0$, and then find a positive δ for which (10) holds. The δ you find will depend on ε.

3.2. Show that $\lim_{x\to 5}(2x+1) = 11$. We have $f(x) = 2x+1$, $a = 5$ and $L = 11$, and the question we must answer is "how close should x be to 5 if want to be sure that $f(x) = 2x+1$ differs less than ε from $L = 11$?"

To figure this out we try to get an idea of how big $|f(x) - L|$ is:

$$|f(x) - L| = \big|(2x+1) - 11\big| = |2x - 10| = 2 \cdot |x - 5| = 2 \cdot |x - a|.$$

So, if $2|x - a| < \varepsilon$ then we have $|f(x) - L| < \varepsilon$, i.e.,

$$\text{if } |x - a| < \tfrac{1}{2}\varepsilon \text{ then } |f(x) - L| < \varepsilon.$$

We can therefore choose $\delta = \frac{1}{2}\varepsilon$. No matter what $\varepsilon > 0$ we are given our δ will also be positive, and if $|x - 5| < \delta$ then we can guarantee $|(2x+1) - 11| < \varepsilon$. That shows that $\lim_{x\to 5} 2x + 1 = 11$.

3.3. The limit $\lim_{x\to 1} x^2 = 1$, the triangle inequality, and the "don't choose $\delta > 1$" trick. This example will show you two basic tricks that are useful in many ε-δ arguments.

The problem is to show that x^2 goes to 1 as x goes to 1: we have $f(x) = x^2$, $a = 1$, $L = 1$, and again the question is, "how small should $|x-1|$ be to guarantee $|x^2-1| < \varepsilon$?"

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

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

How big can the two factors $|x-1|$ and $|x+1|$ be when we assume $|x-1| < \delta$? Clearly the first factor $|x-1|$ satisfies $|x-1| < \delta$ because that is what we had assumed. For the second factor we have

$$|x+1| = \underbrace{|x-1+2| \leq |x-1| + |2|}_{\text{by the triangle inequality}} \leq \delta + 2.$$

The triangle inequality says that $|a+b| \leq |a| + |b|$ for any two real numbers.

It follows that if $|x-1| < \delta$ then

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

Our goal is to show that if δ is small enough then the estimate on the right will not be more than ε. Here is the second trick in this example: we agree *that we always choose our δ so that $\delta \leq 1$*. If we do that, then we will always have

$$(2+\delta)\delta < (2+1)\delta = 3\delta,$$

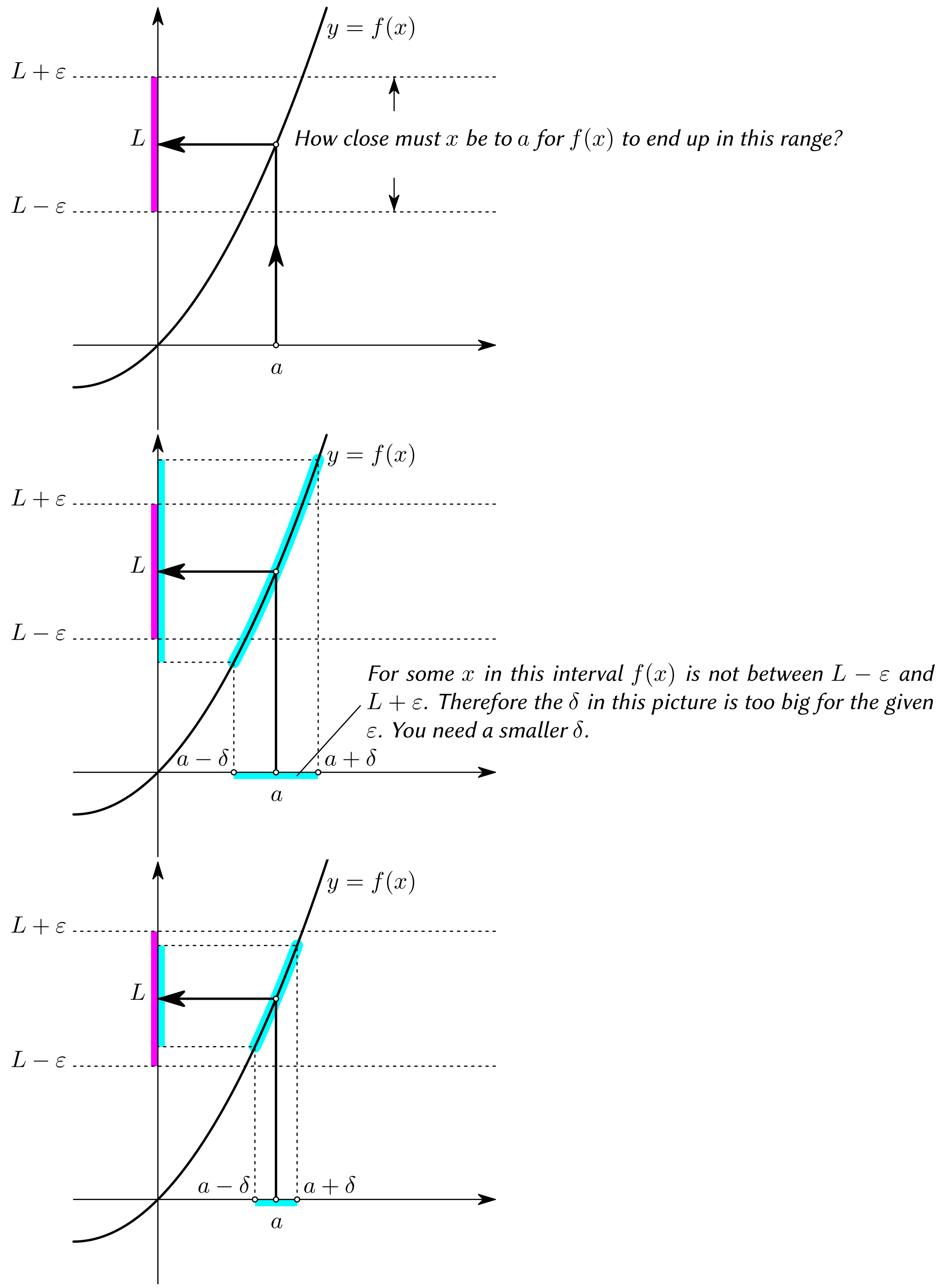
so that $|x-1| < \delta$ with $\delta < 1$ implies

$$|x^2 - 1| < 3\delta.$$

To ensure that $|x^2 - 1| < \varepsilon$, this calculation shows that we should require $3\delta \leq \varepsilon$, i.e. we should choose $\delta \leq \frac{1}{3}\varepsilon$. We must also live up to our promise never to choose $\delta > 1$, so if we are handed an ε for which $\frac{1}{3}\varepsilon > 1$, then we choose $\delta = 1$ instead of $\delta = \frac{1}{3}\varepsilon$. To summarize, we are going to choose

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

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


$y = f(x)$
$L+\varepsilon$
L
$L-\varepsilon$
How close must x be to a for $f(x)$ to end up in this range?
a
$y = f(x)$
$L+\varepsilon$
L
$L-\varepsilon$
For some x in this interval $f(x)$ is not between $L-\varepsilon$ and $L+\varepsilon$. Therefore the δ in this picture is too big for the given ε. You need a smaller δ.
$a-\delta$
$a+\delta$
a
$y = f(x)$
$L+\varepsilon$
L
$L-\varepsilon$
$a-\delta$
$a+\delta$
a

Propagation of errors – another interpretation of ε and δ

According to the limit definition "$\lim_{x\to R} \pi x^2 = A$" is true if *for every $\varepsilon > 0$ you can find a $\delta > 0$ such that $|x - R| < \delta$ implies $|\pi x^2 - A| < \varepsilon$.* Here's a more concrete situation in which ε and δ appear in exactly the same roles:

Suppose you are given a circle drawn on a piece of paper, and you want to know its area. You decide to measure its radius, R, and then compute the area of the circle by calculating

$$\text{Area} = \pi R^2.$$

The area is a function of the radius, and we'll call that function f:

$$f(x) = \pi x^2.$$

When you measure the radius R you will make an error, simply because you can never measure anything with infinite accuracy. Suppose that R is the real value of the radius, and that x is the number you measured. Then the size of the error you made is

$$\text{error in radius measurement} = |x - R|.$$

When you compute the area you also won't get the exact value: you would get $f(x) = \pi x^2$ instead of $A = f(R) = \pi R^2$. The error in your computed value of the area is

$$\text{error in area } = |f(x) - f(R)| = |f(x) - A|.$$

Now you can ask the following question:

Suppose you want to know the area
with an error of at most ε,
then what is the largest error
that you can afford to make
when you measure the radius?

The answer will be something like this: if you want the computed area to have an error of at most $|f(x) - A| < \varepsilon$, then the error in your radius measurement should satisfy $|x - R| < \delta$. You have to do the algebra with inequalities to compute δ when you know ε, as in the examples in this section.

You would expect that if your measured radius x is close enough to the real value R, then your computed area $f(x) = \pi x^2$ will be close to the real area A.

In terms of ε and δ this means that you would expect that no matter how accurately you want to know the area (i.e., how small you make ε) you can always achieve that precision by making the error in your radius measurement small enough (i.e. by making δ sufficiently small).

The expression "the smaller of a and b" shows up often, and is abbreviated to $\min(a, b)$. We could therefore say that in this problem we will choose δ to be

$$\delta = \min\bigl(1, \tfrac{1}{3}\varepsilon\bigr).$$

3.4. Show that $\lim_{x\to 4} 1/x = 1/4$. We apply the definition with $a = 4$, $L = 1/4$ and $f(x) = 1/x$. Thus, for any $\varepsilon > 0$ we try to show that if $|x - 4|$ is small enough then one has $|f(x) - 1/4| < \varepsilon$.

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

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

As before, things would be easier if $1/|4x|$ were a constant. To achieve that we again agree not to take $\delta > 1$. If we always have $\delta \le 1$, then we will always have $|x-4| < 1$, and hence $3 < x < 5$. How large can $1/|4x|$ be in this situation? Answer: the quantity $1/|4x|$ increases as you decrease x, so if $3 < x < 5$ then it will never be larger than $1/|4\cdot 3| = \frac{1}{12}$. We see that if we never choose $\delta > 1$, we will always have

$$|f(x) - \tfrac{1}{4}| \le \tfrac{1}{12}|x-4| \quad \text{for} \quad |x-4| < \delta.$$

To guarantee that $|f(x) - \frac{1}{4}| < \varepsilon$ we could therefore require

$$\tfrac{1}{12}|x-4| < \varepsilon, \quad \text{i.e.} \quad |x-4| < 12\varepsilon.$$

Hence if we choose $\delta = 12\varepsilon$ or any smaller number, then $|x-4| < \delta$ implies $|f(x) - 4| < \varepsilon$. Of course we have to honor our agreement never to choose $\delta > 1$, so our choice of δ is

$$\delta = \text{the smaller of } 1 \text{ and } 12\varepsilon = \min\big(1, 12\varepsilon\big).$$

4. Problems

1. **[Group Problem]** Joe offers to make square sheets of paper for Bruce. Given $x > 0$ Joe plans to mark off a length x and cut out a square of side x. Bruce asks Joe for a square with area 4 square foot. Joe tells Bruce that he can't measure ***exactly*** 2 feet and the area of the square he produces will only be approximately 4 square feet. Bruce doesn't mind as long as the area of the square doesn't differ more than 0.01 square feet from what he really asked for (namely, 4 square foot).

(a) What is the biggest error Joe can afford to make when he marks off the length x?

(b) Jen also wants square sheets, with area 4 square feet. However, she needs the error in the area to be less than 0.00001 square foot. (She's paying.)

How accurately must Joe measure the side of the squares he's going to cut for Jen?

2. **[Group Problem]** *(Joe goes cubic.)* Joe is offering to build cubes of side x. Airline regulations allow you take a cube on board provided its volume and surface area add up to less than 33 (everything measured in feet). For instance, a cube with 2-foot sides has volume+area equal to $2^3 + 6 \times 2^2 = 32$.

If you ask Joe to build a cube whose volume plus total surface area is 32 with an error of at most ε, then what error can he afford to make when he measures the side of the cube he's making?

3. Our definition of a derivative in (9) contains a limit. What is the function "f" there, and what is the variable?

Use the ε–δ definition to prove the following limits:

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

5. $\lim_{x\to 2} x^2 = 4.$

6. $\lim_{x\to 2} x^2 - 7x + 3 = -7$

7. $\lim_{x\to 3} x^3 = 27$

8. $\lim_{x\to 2} x^3 + 6x^2 = 32.$

9. $\lim_{x\to 4} \sqrt{x} = 2.$

10. $\lim_{x\to 3} \sqrt{x+6} = 3.$

11. $\lim_{x\to 2} \dfrac{1+x}{4+x} = \frac{1}{2}.$

12. $\lim_{x\to 1} \dfrac{2-x}{4-x} = \frac{1}{3}.$

13. $\lim_{x\to 3} \dfrac{x}{6-x} = 1.$

14. $\lim_{x\to 0} \sqrt{|x|} = 0$

5. Variations on the limit theme

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

5.1. Left and right limits. When we let "x approach a" we allow x to be larger or smaller than a, as long as x "gets close to a". If we explicitly want to study the behavior of $f(x)$ as x approaches a through values larger than a, then we write

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

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

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

The precise definition of these "one-sided" limits goes like this:

5.2. Definition of right-limits. *Let f be a function. Then*

$$\lim_{x \searrow a} f(x) = L. \tag{11}$$

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

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

holds for all x in the domain of f.

5.3. Definition of left-limits. *Let f be a function. Then*

$$\lim_{x \nearrow a} f(x) = L. \tag{12}$$

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

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

holds for all x in the domain of f. The following theorem tells you how to use one-sided limits to decide if a function $f(x)$ has a limit at $x = a$.

5.4. Theorem. *The two-sided limit $\lim_{x \to a} f(x)$ exists if and only if the two one-sided limits*

$$\lim_{x \searrow a} f(x), \quad \textit{and} \quad \lim_{x \nearrow a} f(x)$$

exist and have the same value.

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

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

("The limit for x going to infinity is L.") We have an analogous definition for what happens to $f(x)$ as x becomes very large and negative: we write

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

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

Here are the precise definitions: *Let $f(x)$ be a function which is defined on an interval $x_0 < x < \infty$. If there is a number L such that for every $\varepsilon > 0$ we can find an A such that*

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

for all x, then we say that the limit of $f(x)$ for $x \to \infty$ is L. Let $f(x)$ be a function which is defined on an interval $-\infty < x < x_0$. If there is a number L such that for every $\varepsilon > 0$ we can find an A such that

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

for all x, then we say that the limit of $f(x)$ for $x \to \infty$ is L. These definitions are 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 that says how large x should be, which is a way of saying "how close x should be to infinity" (or to negative infinity).

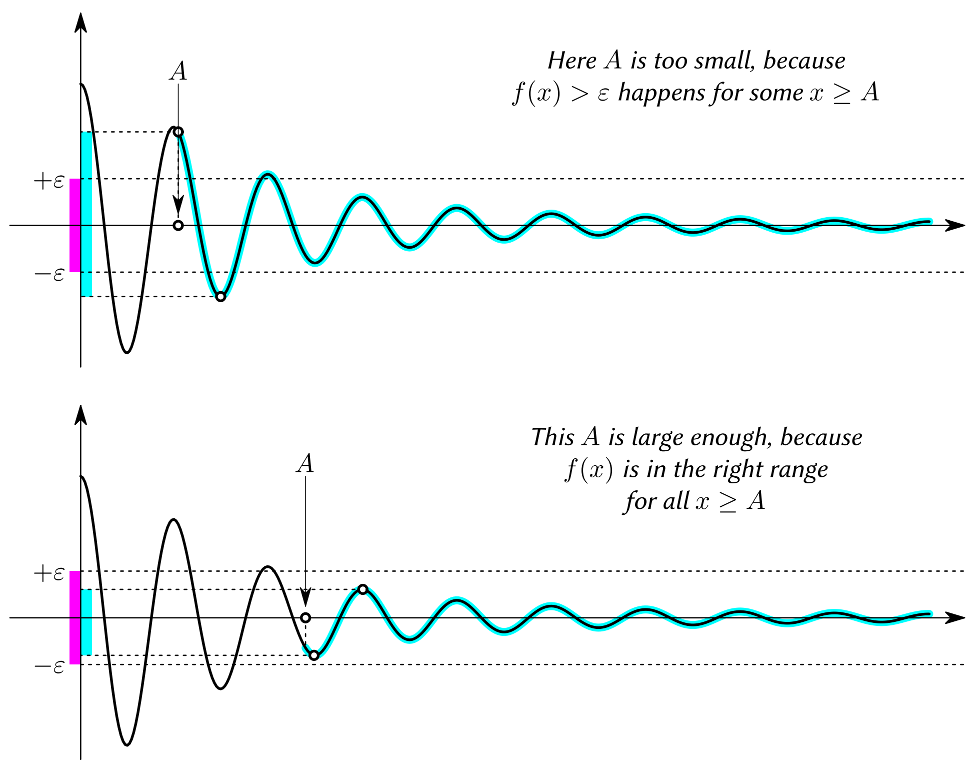


Figure 1. The limit of $f(x)$ as $x \to \infty$; how large must x be if you need $-\varepsilon < f(x) < \varepsilon$?

5.6. Example – Limit of $1/x$. The larger x is, the smaller its reciprocal is, so it seems natural that $1/x \to 0$ as $x \to \infty$. To ***prove*** that $\lim_{x\to\infty} 1/x = 0$ we apply the definition to $f(x) = 1/x$, $L = 0$.

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

$$\left|\frac{1}{x} - L\right| < \varepsilon \text{ for all } x > A \tag{13}$$

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.

Let's decide that we will always take $A > 0$, so that we only need consider positive values of x. Then (13) 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

$$A = \text{ the larger of } 0 \text{ and } \frac{1}{\varepsilon}$$

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

6. Properties of the Limit

The precise definition of the limit is not easy to use, and fortunately we don't need to use it very often. Instead, there are a number of properties that limits possess that will allow us to compute complicated limits without having to resort to "epsiloncy."

The following properties also apply to the variations on the limit from §5. 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, powers, 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

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

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

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

If $\lim_{x\to a} F_2(x) \neq 0$, then

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

Finally, if k is an integer, then

$$\lim_{x\to a} \big(F_1(x)^k\big) = {L_1}^k. \tag{P_7}$$

If $L_1 > 0$, then (P_7) holds for any real number k.

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 §3 but we will not do this here. However, these laws should 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 $L_1 + L_2$. (And so forth.)

Later in this chapter we will add two more properties of limits to this list. They are the "Sandwich Theorem" (§11) and the substitution theorem (§12).

7. Examples of limit computations

7.1. Find $\lim_{x\to 2} x^2$. We have

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

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, we get

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

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).

7.2. Try the examples 1.4 and 1.5 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

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

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 it is the quotient of two polynomials. For such functions there is an algebra trick that 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

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

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

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

7.3. Example – Find $\lim_{x\to 2}\sqrt{x}$. We can apply limit property (P_{5a}) with $k = 1/2$ to see that

$$\lim_{x\to 2}\sqrt{x} = \lim_{x\to 2} x^{1/2} = [\lim_{x\to 2} x]^{1/2} = 2^{1/2} = \sqrt{2}.$$

7.4. 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 (i.e., assuming that $\lim_{x\to 2}\sqrt{x} = \sqrt{2}$.)

Solution: The function is a fraction whose numerator and denominator vanish when $x = 2$, so the limit is of the form "$\frac{0}{0}$". We use an algebra trick; namely, multiplying the numerator and denominator by $\sqrt{x}+\sqrt{2}$:

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

Now we can use the limit properties to compute

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

7.5. Limit as $x\to\infty$ of Rational Functions. A rational function is the quotient of two polynomials:

$$R(x) = \frac{a_n x^n + \cdots + a_1 x + a_0}{b_m x^m + \cdots + b_1 x + b_0}. \tag{14}$$

We have seen that

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

We even proved this in example 5.6. Using this we can find the limit at ∞ for any rational function $R(x)$ as in (14). We could turn the outcome of the calculation of $\lim_{x\to\infty} R(x)$ into a boxed recipe involving the degrees n and m of the numerator and denominator, and also their coefficients a_i, b_j, which you, the student, would then memorize, but it is better to remember *the trick:*

> *To find the limit as $x\to\infty$,*
> *of some rational function you have been given,*
> *factor the highest occurring powers of x,*
> *both from numerator and from denominator.*

For example, let's compute

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

Remember the trick and factor x^2 from top and bottom. You get

$$\begin{aligned}
\lim_{x\to\infty} \frac{3x^2+3}{5x^2+7x-39} &= \lim_{x\to\infty} \frac{x^2}{x^2}\,\frac{3+3/x^2}{5+7/x-39/x^2} && \text{(algebra)}\\
&= \lim_{x\to\infty} \frac{3+3/x^2}{5+7/x-39/x^2} && \text{(more algebra)}\\
&= \frac{\lim_{x\to\infty}(3+3/x^2)}{\lim_{x\to\infty}(5+7/x-39/x^2)} && \text{(limit properties)}\\
&= \frac{3}{5}.
\end{aligned}$$

At the end of this computation, we used the limit properties (P_*) to break the limit down into simpler pieces like $\lim_{x\to\infty} 39/x^2$, which we can directly evaluate; for example, we have

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

The other terms are similar.

7.6. Another example with a rational function. Compute

$$\lim_{x\to\infty}\frac{2x}{4x^3+5}.$$

We apply "the trick" again and factor x out of the numerator and x^3 out of the denominator. This leads to

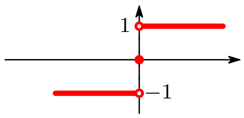
$$\begin{aligned}
\lim_{x\to\infty}\frac{2x}{4x^3+5} &= \lim_{x\to\infty}\left(\frac{x}{x^3}\,\frac{2}{4+5/x^3}\right)\\
&= \lim_{x\to\infty}\left(\frac{1}{x^2}\,\frac{2}{4+5/x^3}\right)\\
&= \lim_{x\to\infty}\left(\frac{1}{x^2}\right)\cdot\left(\lim_{x\to\infty}\frac{2}{4+5/x^3}\right)\\
&= 0\cdot\tfrac{2}{4}\\
&= 0.
\end{aligned}$$

This example and the previous cover two of the three possible combinations of degrees n and m, namely $n=m$ and $n<m$. To show the remaining case, in which the numerator has higher degree than the denominator, we should do yet another example, but that one will have to wait a few pages (see § 9.3)

8. 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."

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



$y = \operatorname{sign}(x)$

There are those who do not like the notation "$\operatorname{sign}(x)$," and prefer to write $g(x) = \frac{x}{|x|}$ instead of $g(x) = \operatorname{sign}(x)$. If you think about this formula for a moment you'll see that $\operatorname{sign}(x)$ and $x/|x|$ are the same for all $x \neq 0$. When $x = 0$ the quotient $x/|x|$ is of course not defined.

8.2. The sign function near $x = 0$. The "sign function" is defined by

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

Note that "the sign of zero" is defined to be zero. Our question is: does the sign function have a limit at $x = 0$? The answer is ***no***, and here is why: since $\operatorname{sign}(x) = +1$ for all positive values of x, we see that

$$lim_{x\searrow 0} \operatorname{sign}(x) = +1.$$

Similarly, since $\operatorname{sign}(x) = -1$ for all negative values of x, we see that

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

Then by the result of 5.4, we see that because the two one-sided limits have different values, the two-sided limit does not exist. (Essentially, if the two-sided limit existed, then it would have to be equal to both $+1$ and -1 at the same time, which is impossible.)

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

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

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

When $x = 0$ the function $f(x) = \sin(\pi/x)$ is not defined, because its definition involves division by x. What happens to $f(x)$ as $x \to 0$? First, π/x becomes larger and larger ("goes to infinity") as $x \to 0$. Then, taking the sine, we see that $\sin(\pi/x)$ oscillates between $+1$ and -1 infinitely often as $x \to 0$, since, for example, we can calculate that $f(\frac{2}{4k+1}) = +1$ and $f(\frac{2}{4k+3}) = -1$ for each integer k. This means that $f(x)$ gets close to any number between -1 and $+1$ as $x \to 0$, but that the function $f(x)$ ***never stays close*** to any particular value because it keeps oscillating up and down between $+1$ and -1.

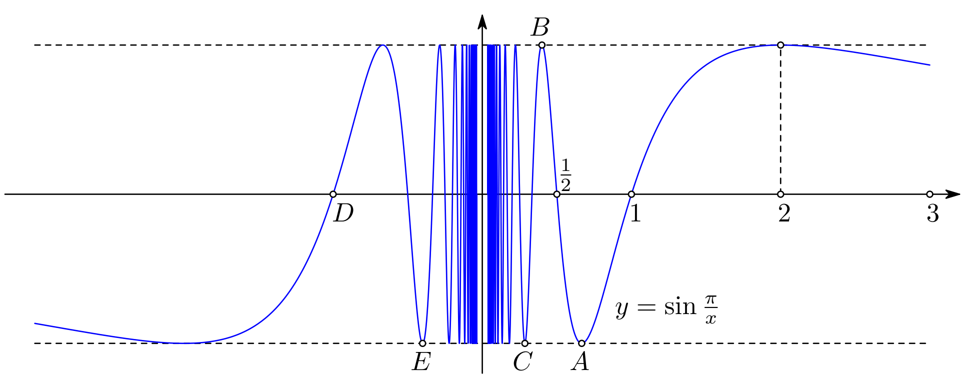


Figure 2. Graph of $y = \sin\frac{\pi}{x}$ for $-3 < x < 3,\ x \neq 0$.

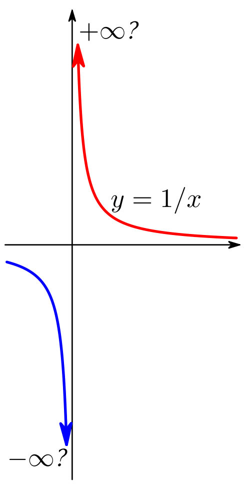
Here again, the limit $\lim_{x\to 0} f(x)$ does not exist. We have arrived at this conclusion by only considering what $f(x)$ does for small positive values of x. So the limit fails to exist in a stronger way than in the example of the sign-function. There, even though the limit didn't exist, the one-sided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

$$\lim_{x\searrow 0}\sin\left(\frac{\pi}{x}\right) \text{ and } \lim_{x\nearrow 0}\sin\left(\frac{\pi}{x}\right)$$

do not exist.

8.4. Limit properties for limits that don't exist. The limit properties all say that some limit exists (and what it's equal to), but by applying some logic you can sometimes also use the limit properties to conclude that a limit does ***not*** exist. For instance:

Theorem. *If* $\lim_{x\to a} f(x) + g(x)$ *does not exist, then at least one of the two limits* $\lim_{x\to a} f(x)$ *or* $\lim_{x\to a} g(x)$ *does not exist.* In English: "if the sum of two functions has no limit, then one of the terms also has no limit." This follows directly from Property (P_3) which says that if $\lim f(x)$ and $\lim g(x)$ both exist then $\lim f(x) + g(x)$ also must exist.



9. Limits that equal ∞

For some limits that don't exist, we can be more descriptive about the reason that the limit doesn't exist. Here are some examples.

9.1. The limit of $1/x$ at $x = 0$. Consider the limit

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

As x decreases to $x = 0$ through smaller and smaller positive values, its reciprocal $1/x$ becomes larger and larger. We say that instead of going to some finite number, the quantity $1/x$ "goes to infinity" as $x \searrow 0$. In symbols:

$$\lim_{x\searrow 0}\frac{1}{x} = \infty. \tag{15}$$

Likewise, as x approaches 0 through negative numbers, its reciprocal $1/x$ drops lower and lower, and we say that $1/x$ "goes to $-\infty$" as $x \nearrow 0$. Symbolically,

$$\lim_{x\nearrow 0}\frac{1}{x} = -\infty. \tag{16}$$

The limits (15) and (16) are not like the normal limits we have been dealing with so far. Namely, when we write something like

$$\lim_{x\to 2} x^2 = 4$$

we mean that the limit actually exists and that it is equal to 4. On the other hand, since we have agreed that ∞ is not a number, the meaning of (15) cannot be to say that "the limit exists and its value is ∞."

Instead, when we write

$$\lim_{x\to a} f(x) = \infty \tag{17}$$

for some function $y = f(x)$, we mean, ***by definition,*** that the limit of $f(x)$ does not exist, and that it fails to exist in a specific way: as $x \to a$, the value of $f(x)$ becomes "larger and larger," and in fact eventually becomes larger than any finite number.

The language in that last paragraph shows you that this is an intuitive definition, at the same level as the first definition of limit we gave in §1. It contains the usual suspect phrases such as "larger and larger," or "finite number" (as if there were any other kind.) A more precise definition involving epsilons can be given, but in this course we will not go into this much detail. As a final comment on infinite limits, it is important to realize that, since (17) is not a normal limit ***you cannot apply the limit rules to infinite limits.*** Here is an example of what goes wrong if you try anyway.

9.2. Trouble with infinite limits. If you apply the limit properties to $\lim_{x\searrow 0} 1/x = \infty$, then you could conclude

$$1 = \lim_{x\searrow 0} x \cdot \frac{1}{x} = \lim_{x\searrow 0} x \times \lim_{x\searrow 0} \frac{1}{x} = 0 \times \infty = 0,$$

because "anything multiplied with zero is zero."

After using the limit properties in combination with this infinite limit we reach the absurd conclusion that $1 = 0$. The moral of this story is that you can't use the limit properties when some of the limits are infinite.

9.3. Limit as $x \to \infty$ of rational functions, again. In § 7.5 we computed the limits of rational functions $R(x) = \frac{P(x)}{Q(x)}$ as $x \to \infty$ in all cases where the degree of the numerator P was not more than the degree of the denominator Q. Here we complete the story by looking at the remaining case where $\deg P > \deg Q$. We'll do the following examples:

$$A = \lim_{x\to\infty} x, \quad B = \lim_{x\to\infty} -7x^2, \quad C = \lim_{x\to\infty} \frac{x^3 - 2x}{1 + 25x^2}.$$

The first limit is the easiest: "as x goes to infinity, x goes to infinity." So we have

$$A = \lim_{x\to\infty} x = \infty.$$

Remember, we don't have a precise definition that says when a limit equals "∞," so we can't prove this here. On the other hand, if someone were to come up with a definition of the above limit that would make it equal to something else, we would ask them to change their definition instead of changing our mind about $\lim_{x\to\infty} x$. The second limit is only slightly harder: as $x \to \infty$, the square x^2 also becomes arbitrarily large, and thus $-7x^2$ will go to $-\infty$. So we'll say $B = -\infty$. Finally, let's look at C. We apply the trick from § 7.5:

$$C = \lim_{x\to\infty} \frac{x^3 - 2x}{1 + 25x^2} = \lim_{x\to\infty} \frac{x^3}{x^2} \frac{1 - \frac{2}{x}}{\frac{1}{x} + 25}.$$

At this point we would like to use the limit properties to say

$$C = \lim_{x\to\infty} \frac{x^3}{x^2} \frac{1 - \frac{2}{x}}{\frac{1}{x} + 25} = \lim_{x\to\infty} \frac{x^3}{x^2} \lim_{x\to\infty} \frac{1 - \frac{2}{x}}{\frac{1}{x} + 25} = \bigl(\lim_{x\to\infty} x\bigr) \cdot \frac{1}{25} = \infty \cdot \frac{1}{25} = \infty.$$

Unfortunately, as we have just seen, the limit properties are unreliable when infinite limits are involved, so we have no justification for using them here. Nevertheless, the answer we got this time seems reasonable; in an attempt to make it look more reasonable you could write it this way:

$$\frac{x^3}{x^2} \frac{1 - \frac{2}{x}}{\frac{1}{x} + 25} = x \cdot \frac{1 - \frac{2}{x}}{\frac{1}{x} + 25} \approx x \cdot \frac{1}{25}$$

for large x. Therefore, as $x \to \infty$ the fraction

$$\frac{x^3}{x^2} \frac{1 - \frac{2}{x}}{\frac{1}{x} + 25} \approx x \cdot \frac{1}{25} \to \infty.$$

10. What's in a name? – Free Variables and Dummy variables

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 that approaches a. But the result of our computation shouldn't depend on the name we choose: it doesn't matter if we call it x or u, or anything else. 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: 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, since the limit is $2\cdot 3+1=7$ (no problem). We could substitute other values of a and we would get different answers. In general you get $2a+1$.

11. Limits and Inequalities

This section has two theorems that let you compare limits of different functions. The properties in these theorems are not formulas that allow you to compute limits like the properties $(P_1)\dots(P_6)$ from §6. Instead, they allow us to ***reason*** about limits.

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

11.1. Theorem. *Let f and g be functions whose limits as $x \to a$ 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 we 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. Here is the second theorem about limits and inequalities.

11.2. The Sandwich Theorem. *Suppose that*

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

(for all x) and that both the limits

$$\lim_{x\to a} f(x) \text{ and } \lim_{x\to a} h(x)$$

exist and are equal to the same value L. Then

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

also exists and is equal to L.

The theorem is useful when we want to know the limit of g, and when we 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 we 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.

11.3. Example: 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

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

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

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

Note that the limit $\lim_{x\to 0}\cos(\pi/x)$ does ***not*** exist, for the same reason that the "backward sine" did not have a limit for $x \to 0$ (see § 8.3). Multiplying with x changed that.

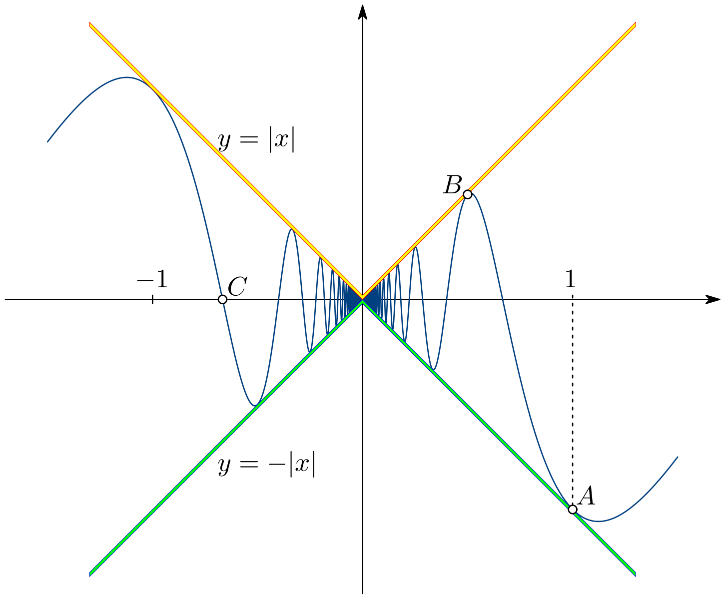


Figure 3. The graph of $x \cos \frac{\pi}{x}$ is "sandwiched" in between the graphs of $y = +|x|$ and of $y = -|x|$.

12. Continuity

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

$$\lim_{x \to a} g(x) = g(a) \tag{18}$$

*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 has to be defined on 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 $-1 < x < 1$ because it isn't defined at $x = 0$. Continuous functions are "nice", and most functions we will want to talk about will be continuous. For example, all polynomials are continuous functions. The ideas in the proof are the same as in the following example

12.2. Example. A polynomial that is continuous. Let us show that $P(x) = x^2 + 3x$ is continuous at $x = 2$. To show that we have to prove that

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

which is to say,

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

We can do this two ways: using the definition with ε and δ (the hard way). It is much easier to the limit properties $(P_1) \dots (P_6)$ from §6: for example,

$$\begin{aligned}
\lim_{x \to 2} x^2 + 3x &= \lim_{x \to 2} x^2 + \lim_{x \to 2} 3x \\
&= \big(\lim_{x \to 2} x\big) \cdot \big(\lim_{x \to 2} x\big) + \big(\lim_{x \to 2} 3\big) \cdot \big(\lim_{x \to 2} x\big) \\
&= 2 \cdot 2 + 3 \cdot 2
\end{aligned}$$

and this is what we wanted.

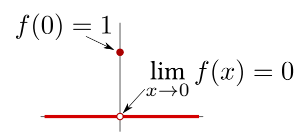
12.3. Example. Rational functions are continuous wherever they are defined. Let $R(x) = \frac{P(x)}{Q(x)}$ be a rational function, and let a be any number in the domain of R; i.e., any number for which $Q(a) \neq 0$. Then one has

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

This shows that R is indeed continuous at a.

12.4. Example of a discontinuous function. Here is an artificial example of a discontinuous function. Consider

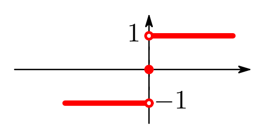
$$f(x) = \begin{cases} 0 & \text{when } x \neq 0 \\ 1 & \text{when } x = 0 \end{cases}.$$



This function is not continuous at $x = 0$ because $\lim_{x\to 0} f(x) = 0$, while $f(0) = 1$: the limit for $x \to 0$ exists, but it is different from the function value at $x = 0$.

This example is artificial, because we took a continuous function ($f(x) = 0$) and changed its definition at one point ($f(0) = 1$), which then made the function discontinuous. We get less artificial examples by looking at functions that do not have limits. 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 $g(x) = \operatorname{sign}(x)$ from § 8.2 is not continuous at $x = 0$. This kind of discontinuity is fairly common and has a name:



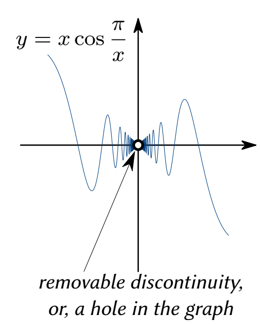
$y = \operatorname{sign}(x)$: a function with a jump discontinuity. Left and right limits at $x = 0$ exist, but they are not the same

12.5. Definition. *A function $y = f(x)$ has a **jump discontinuity** at $x = a$ if the left- and right-hand limits of $f(x)$ at $x = a$ both exist but are different:*

$$\lim_{x\nearrow a} f(x) \neq \lim_{x\searrow a} f(x).$$

12.6. Removable discontinuities. Consider the function from § 11.3

$$f(x) = x\cos\frac{\pi}{x}.$$



It is defined for all x except $x = 0$, and therefore it is not continuous at $x = 0$, but for the silly reason that f is not defined there: to decide if f is continuous at $x = 0$ we must compare $\lim_{x\to 0} f(x)$ with $f(0)$, but $f(0)$ isn't defined. However, the limit

$$\lim_{x\to 0} f(x) = \lim_{x\to 0} x\cos\frac{\pi}{x} = 0$$

does exist (by the Sandwich theorem, see § 11.3 again). Since $f(0)$ is not defined yet we can extend the definition of f to include $f(0) = 0$, i.e., we define

$$g(x) = \begin{cases} x\cos\pi/x & \text{for } x \neq 0 \\ 0 & \text{when } x = 0. \end{cases}$$

This new function $g(x)$ then is continuous everywhere, including at $x = 0$.

This is an example of a more general situation in which the domain of a function "has a hole," meaning that the function is defined everywhere in some interval $a-\delta < x < a+\delta$, except at $x = a$. If the limit $\lim_{x\to a} f(x) = L$ exists, then you can extend the definition of the function by declaring that $f(a) = L$. The extended function is then continuous at $x = a$. The point $x = a$ is called a ***removable discontinuity*** of the function f. In § 15.3 we'll see another example of a function with a removable discontinuity whose graph isn't as messy as the graph of $x \cos \pi/x$.

13. Substitution in Limits

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

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

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

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

If we 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:

13.1. Theorem. *If $\lim_{x\to a} g(x) = L$, 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

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

What this statement means is that we can freely move continuous functions outside of limits.

13.2. Example: 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

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

Either way, we have

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

We can 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 this is true, but this particular limit is the one from example 7.3. Putting these two limits together, we conclude that the limit is $\sqrt{2}$. We could more compactly write this whole argument as follows:

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

with the remark that we need to know 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 we would need to say that we have substituted $u = x^3 - 3x^2 + 2$.

14. Problems

Find the following limits:

1. $\lim_{x\to -7}(2x+5)$

2. $\lim_{x\to 7-}(2x+5)$

3. $\lim_{x\to -\infty}(2x+5)$

4. $\lim_{x\to -4}(x+3)^{2006}$

5. $\lim_{x\to -4}(x+3)^{2007}$

6. $\lim_{x\to -\infty}(x+3)^{2007}$

7. $\lim_{t\to 1} \dfrac{t^2+t-2}{t^2-1}$

8. $\lim_{t\nearrow 1} \dfrac{t^2+t-2}{t^2-1}$

9. $\lim_{t\to -1} \dfrac{t^2+t-2}{t^2-1}$

10. $\lim_{x\to\infty} \dfrac{x^2+3}{x^2+4}$

11. $\lim_{x\to\infty} \dfrac{x^5+3}{x^2+4}$

12. $\lim_{x\to\infty} \dfrac{x^2+1}{x^5+2}$

13. $\lim_{x\to\infty} \dfrac{(2x+1)^4}{(3x^2+1)^2}$

14. $\lim_{u\to\infty} \dfrac{(2u+1)^4}{(3u^2+1)^2}$

15. $\lim_{t\to 0} \dfrac{(2t+1)^4}{(3t^2+1)^2}$

16. What are the coordinates of the points labeled $A, \ldots, E$ in Figure 2 (the graph of $y = \sin \pi/x$).

17. If $\lim_{x\to a} f(x)$ exists then f is continuous at $x = a$. *True or false?*

18. Give two examples of functions for which $\lim_{x\searrow 0} f(x)$ does not exist.

19. **[Group Problem]** If $\lim_{x\to 0} f(x)$ and $\lim_{x\to 0} g(x)$ both do not exist, then $\lim_{x\to 0}\big(f(x) + g(x)\big)$ also does not exist. *True or false?*

20. **[Group Problem]** If $\lim_{x\to 0} f(x)$ and $\lim_{x\to 0} g(x)$ both do not exist, then $\lim_{x\to 0}(f(x)/g(x))$ also does not exist. *True or false?*

21. True or false:

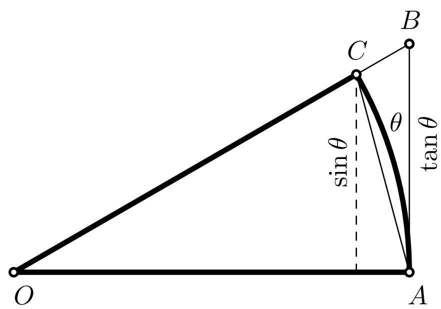
(a) If $\lim_{x\to a} f(x)$ exists and $\lim_{x\to a} g(x)$ does not exist, then $\lim_{x\to a} f(x) + g(x)$ could still exist.

(b) If $\lim_{x\to a} f(x)$ exists and $\lim_{x\to a} g(x)$ does not exist, then $\lim_{x\to a} f(x)g(x)$ could still exist.

(c) If $\lim_{x\to a} f(x)$ exists and $\lim_{x\to a} g(x)$ does not exist, then $\lim_{x\to a} f(x)/g(x)$ could still exist.

(d) If $\lim_{x\to a} f(x)$ does not exist but $\lim_{x\to a} g(x)$ does exist, then $\lim_{x\to a} f(x)/g(x)$ could still exist.

22. **[Group Problem]** 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 $\lim_{x\to\infty} x \cdot (\frac{1}{x})$.



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 4. Proving $\lim_{\theta\to 0}\dfrac{\sin\theta}{\theta} = 1$ by comparing the areas of two triangles and a circular wedge.

23. Evaluate $\displaystyle\lim_{x\to 9}\frac{\sqrt{x}-3}{x-9}$. Hint: Multiply top and bottom by $\sqrt{x}+3$.

24. Evaluate $\displaystyle\lim_{x\to 2}\frac{\frac{1}{x}-\frac{1}{2}}{x-2}$.

25. Evaluate $\displaystyle\lim_{x\to 2}\frac{\frac{1}{\sqrt{x}}-\frac{1}{\sqrt{2}}}{x-2}$.

26. 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?

27. 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.

28. Find all possible pairs of constants a and c such that the function

$$f(x) = \begin{cases} x^3+c & \text{for } x < 0 \\ ax+c^2 & \text{for } 0 \le x < 1 \\ \arctan x & \text{for } x \ge 1. \end{cases}$$

is continuous for all x. (There is more than one possibility.)

15. Two Limits in Trigonometry

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

$$\sin\theta \approx \theta \quad \text{and} \quad \cos\theta \approx 1 - \tfrac{1}{2}\theta^2. \tag{19}$$

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

Theorem.

$$\lim_{\theta\to 0}\frac{\sin\theta}{\theta} = 1. \tag{20}$$

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 $0 < \theta < \pi/2$.

Look at Figure 4. Since the wedge OAC contains the triangle OAC its area must be larger. The area of the wedge is $\frac{1}{2}\theta$ and the area of the triangle is $\frac{1}{2}\sin\theta$, so we find that

$$0 < \sin\theta < \theta \text{ for } 0 < \theta < \frac{\pi}{2}. \tag{21}$$

The Sandwich Theorem implies that

$$\lim_{\theta \searrow 0} \sin\theta = 0. \tag{22}$$

Moreover, we also have

$$\lim_{\theta \searrow 0} \cos\theta = \lim_{\theta \searrow 0} \sqrt{1 - \sin^2\theta} = 1. \tag{23}$$

Next we compare the areas of the wedge OAC and the larger triangle OAB. Since OAB has area $\frac{1}{2}\tan\theta$ we find that

$$\theta < \tan\theta$$

for $0 < \theta < \frac{\pi}{2}$. Since $\tan\theta = \frac{\sin\theta}{\cos\theta}$ we can multiply with $\cos\theta$ and divide by θ to get

$$\cos\theta < \frac{\sin\theta}{\theta} \text{ for } 0 < \theta < \frac{\pi}{2}$$

If we go back to (21) and divide by θ, then we get

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

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. □

15.1. The Cosine counterpart of (20). We will show that

$$\lim_{\theta \to 0} \frac{1 - \cos\theta}{\theta^2} = \frac{1}{2}. \tag{24}$$

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 (24) follows.

This limit claims that for small values of θ one has

$$\frac{1 - \cos\theta}{\theta^2} \approx \frac{1}{2},$$

and hence

$$\cos\theta \approx 1 - \frac{1}{2}\theta^2.$$

15.2. The Sine and Cosine of a small angle. As promised in (19), you can use the limits in this section to approximate $\sin\theta$ when θ is small. Namely, since $\lim_{\theta\to 0}\frac{\sin\theta}{\theta} = 1$ it is reasonable to assume that for small angles θ one has $\frac{\sin\theta}{\theta}\approx 1$, or, $\sin\theta\approx\theta$. Similarly, (24) implies that for small angles θ one has $\cos\theta\approx 1-\frac{1}{2}\theta^2$.

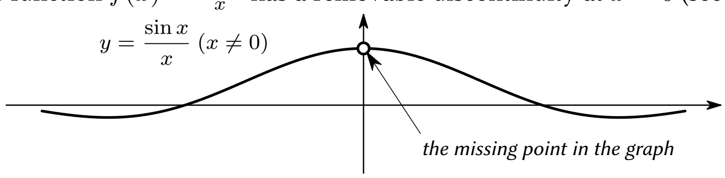
For instance, to get a quick estimate of $\sin(0.1)$ and $\cos(0.1)$ (where, as always, 0.1 is measured in radians) we use (19) and get

$$\sin 0.1 \approx 0.1, \qquad \cos 0.1 \approx 1-\tfrac{1}{2}(0.1)^2 = 0.995.$$

The formula (19) does not say how accurate these approximations are. To see how good the approximations are you could compare with the actual values of $\sin 0.1$ and $\cos 0.1$. These are, rounded to six decimals,

$$\sin 0.1 \approx 0.099\,833, \qquad \cos 0.1 \approx 0.995\,004.$$

15.3. Another example of a removable discontinuity. Since $(\sin x)/x \to 1$ as $x\to 0$, the function $f(x) = \frac{\sin x}{x}$ has a removable discontinuity at $x = 0$ (see § 12.6.)



The function

$$f(x) \stackrel{\text{def}}{=} \begin{cases} (\sin x)/x & \text{for } x\neq 0,\\ 1 & \text{when } x = 0\end{cases}$$

is continuous at $x = 0$.

16. Problems

For each of the following limits, either evaluate them ***or*** *show that they do not exist. Distinguish between limits which are infinite and limits which do not exist.*

1. $\displaystyle\lim_{\theta\to 0}\frac{\tan\theta}{\theta}$

(Hint: $\tan\theta = \frac{\sin\theta}{\cos\theta}$).

2. $\displaystyle\lim_{\theta\to 0}\frac{\theta}{\sin\theta}$

3. $\displaystyle\lim_{\alpha\to 0}\frac{\sin 2\alpha}{\alpha}$

(Hint: a substitution)

4. $\displaystyle\lim_{\alpha\to 0}\frac{\sin 2\alpha}{\sin\alpha}$

(two ways: with and without the double angle formula!)

5. $\displaystyle\lim_{x\to 0}\frac{\sin 3x}{2x}$.

6. $\displaystyle\lim_{\alpha\to 0}\frac{\tan 4\alpha}{\sin 2\alpha}$.

7. $\displaystyle\lim_{x\to 0}\frac{1-\cos x}{x\sin x}$.

8. $\displaystyle\lim_{\theta\to\pi/2}\frac{1-\sin\theta}{\theta-\pi/2}$

9. $\displaystyle\lim_{x\to 0}\frac{\sin^2 x}{1-\cos x}$.

10. $\displaystyle\lim_{x\to 0}\frac{\sin(x^2)}{x^2}$.

11. $\displaystyle\lim_{x\to 0}\frac{x(1-\cos x)}{\tan^3 x}$.

12. $\displaystyle\lim_{x\to 0}\frac{\sin(x^2)}{1-\cos x}$.

13. $\displaystyle\lim_{x\to\pi/2}\frac{x-\frac{\pi}{2}}{\cos x}$.

14. $\displaystyle\lim_{x\to\pi/2}(x-\tfrac{\pi}{2})\tan x$.

15. $\displaystyle\lim_{x\to 0}\frac{\cos x}{x^2+9}$.

16. $\displaystyle\lim_{x\to\pi}\frac{\sin x}{x-\pi}$.

17. $\displaystyle\lim_{x\to 0}\frac{\sin x}{x+\sin x}$.

18. $\displaystyle A = \lim_{x\to\infty}\frac{\sin x}{x}$.

19. $\displaystyle B = \lim_{x\to\infty}\frac{\cos x}{x}$.

20. $\displaystyle\lim_{x\to\infty}\frac{\tan x}{x}$

21. $\displaystyle\lim_{x\to\infty}\frac{\sin 2x}{x}$

22. $\displaystyle\lim_{x\to\infty}\frac{\sin 2x}{1+x^2}$

23. $\displaystyle\lim_{x\to\infty} \frac{x}{\cos x + x^2}$

24. $\displaystyle\lim_{x\to\infty} \frac{2x^3 + 3x^2\cos x}{(x+2)^3}$

25. Since both $\lim_{\theta\to 0} \frac{\sin\theta}{\theta} = 1$ and $\lim_{\theta\to 0} \frac{\tan\theta}{\theta} = 1$, you would think that for small angles θ

$$\sin\theta \approx \theta \approx \tan\theta.$$

In other words the sine and tangent of small angles are almost the same. This problem goes into the question of how small the difference really is.

(a) Compute $\displaystyle\lim_{\theta\to 0} \frac{\tan\theta - \sin\theta}{\theta^3}$.

(Hint: use $\tan\theta = \sin\theta / \cos\theta$.)

(b) Use your answer from part (a) to estimate the difference between $\tan 0.1$ and $\sin 0.1$.

26. Suppose you are lost on an island and you have no calculator. Approximate the following quantities as best as you can:

(a) $\sin 0.2$ **(b)** $\cos 0.2$ **(c)** $\tan 0.2$
(d) $\sin\left(\frac{\pi}{2} - 0.2\right)$ **(e)** $\cos\left(\frac{\pi}{2} + 0.2\right)$
(f) $\tan\left(\frac{\pi}{2} - 0.2\right)$

27. You are still on the island, and you still have no calculator. Approximate the following quantities as best as you can:

(a) $\sin 10^\circ$ **(b)** $\cos 10^\circ$ **(c)** $\tan 10^\circ$
(d) $\sin 100^\circ$ **(e)** $\cos 10^\circ$ **(f)** $\tan 80^\circ$

Note that here the angles are specified in degrees rather than radians.

28. **(a)** Simplicio says that while a *million* is clearly a lot, *two* is not a large number. After reading (19) he then concludes that

$$\sin(2) \approx 2.$$

Comment on his conclusion, and explain the meaning of "small" in the sentence "*if θ is small then* $\sin\theta \approx \theta$."

(b) The next day Simplicio makes an accurate drawing of a three degree angle by dividing a right angle (90°) into 30 equal angles. The drawing convinces him that three degrees is a very small angle indeed, and so he concludes that

$$\sin 3^\circ \approx 3.$$

Again, comment on Simplicio's reasoning.

29. Is there a constant k such that the function

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

is continuous? If so, find it; if not, say why no such k exists.

30. Find a constant A so that the function

$$f(x) = \begin{cases} \dfrac{\sin x}{2x} & \text{for } x \neq 0 \\ A & \text{when } x = 0 \end{cases}$$

is continuous everywhere.

31. Compute $\lim_{x\to\infty} x\sin\frac{\pi}{x}$ and $\lim_{x\to\infty} x\tan\frac{\pi}{x}$. (Hint: substitute something).

17. Asymptotes

Asymptotes can be vaguely defined by saying that a line is an asymptote of the graph of a function $y = f(x)$ if the graph "approaches the line as x approaches infinity." For a more precise definition we distinguish between three cases, depending on the line.

17.1. Vertical Asymptotes. If $y = f(x)$ is a function for which either

$$\lim_{x\searrow a} f(x) = \infty, \text{ or } \lim_{x\searrow a} f(x) = -\infty, \text{ or } \lim_{x\nearrow a} f(x) = \infty, \text{ or } \lim_{x\nearrow a} f(x) = -\infty$$

holds, then the line $x = a$ is a ***vertical asymptote*** of the function $y = f(x)$.

17.2. Horizontal Asymptotes. If $y = f(x)$ is a function for which

$$\lim_{x\to\infty} f(x) = a \text{ or } \lim_{x\to-\infty} f(x) = a,$$

then the line $y = a$ is a ***horizontal asymptote*** of the function $y = f(x)$.

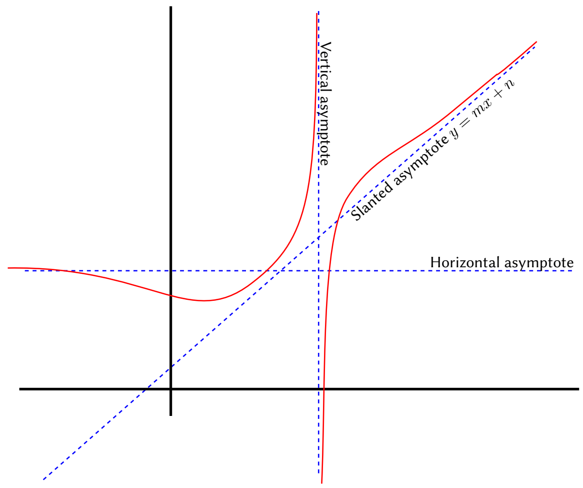


Figure 5. A function with all three kinds of asymptote.

17.3. Slanted Asymptotes. If $y = f(x)$ is a function for which

$$\lim_{x\to\infty}\big[f(x) - \big(mx+n\big)\big] = 0, \text{ or } \lim_{x\to-\infty}\big[f(x) - \big(mx+n\big)\big] = 0,$$

then the line $y = mx + n$ is a ***slanted asymptote*** for the function $y = f(x)$. In this case we say that ***the function*** $f(x)$ ***is asymptotic to*** $mx + n$ ***as*** $x \to \infty$ (or $x \to -\infty$, depending on the case).

Slanted asymptotes are the hardest to find, but the following observations can be helpful.

17.4. Asymptotic slope and intercept. If $y = mx + n$ is a slanted asymptote of $y = f(x)$ then

$$m = \lim_{x\to\infty}\frac{f(x)}{x}, \qquad n = \lim_{x\to\infty} f(x) - mx. \tag{25}$$

You should be able to prove this yourself. See problem 18.5.

17.5. Example. The function

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

has both a horizontal and a vertical asymptote. To find the vertical asymptote we look for values of x where the function "may become infinite." The function is well defined everywhere except when the denominator $1 - x$ vanishes. This happens when $x = 1$. To verify that $x = 1$ is a vertical asymptote we compute the limit

$$\lim_{x\searrow 1} f(x) = -\infty \text{ and } \lim_{x\nearrow 1} f(x) = +\infty.$$

Either of these two limits shows that the line $x = 1$ is a vertical asymptote for $y = x/(1-x)$.

To find the horizontal asymptotes we compute

$$\lim_{x\to\infty} f(x) = -1 \text{ and } \lim_{x\to-\infty} f(x) = -1.$$

Again, either of these limits implies that the line $y = -1$ is a horizontal asymptote for $y = x/(1-x)$.

17.6. Example. Find the asymptotes of the function

$$f(x) = \sqrt{x^2 + 1}.$$

The function is well defined and continuous at all x. This means that $\lim_{x \to a} f(x)$ is always finite (it's $f(a)$) and therefore this function has no vertical asymptotes.

Both limits $\lim_{x \to \pm\infty} f(x)$ are infinite so the function also has no horizontal asymptotes.

To find possible slanted asymptotes we first see what the slope of such an asymptote would be:

$$m = \lim_{x\to\infty} \frac{f(x)}{x} = \lim_{x\to\infty} \frac{\sqrt{x^2+1}}{x} = \lim_{x\to\infty} \sqrt{1 + \frac{1}{x^2}} = 1.$$

This does not yet prove that there is a slanted asymptote; to prove that one exists we must also find the intercept of the asymptote. This intercept is given by

$$n = \lim_{x\to\infty} f(x) - mx = \lim_{x\to\infty} \sqrt{x^2+1} - x = \lim_{x\to\infty} \frac{1}{\sqrt{x^2+1} + x} = \lim_{x\to\infty} \frac{1/x}{\sqrt{1 + 1/x^2} + 1} = \frac{0}{2}.$$

This last limit implies that

$$\lim_{x\to\infty} f(x) - x = 0$$

and therefore $y = x$ is the slanted asymptote of the function.

18. Problems

1. Find the asymptotes (horizontal, vertical and slanted) of the following functions:

(a) $f(x) = \dfrac{x}{x^2+1}$

(b) $f(x) = \dfrac{x}{x^2-4}$

(c) $f(x) = \dfrac{5x^2}{x^2-2}$

(d) $f(x) = \dfrac{x}{x^2-4}$

(e) $f(x) = \dfrac{x}{x-4}$

(f) $f(x) = \dfrac{x^3}{x^2+4}$

2. Which of the following functions have asymptotes?

(a) $f(x) = \sqrt{x}$

(b) $f(x) = \sqrt{x} - \sqrt{x-1}$

(c) $f(x) = x + \cos x$

(d) $f(x) = x \sin x$

(e) $f(x) = \dfrac{\sin x}{x}$

(f) $f(x) = \sqrt{x^2 + x}$

(g) $f(x) = \dfrac{x}{\sqrt{x^2+1}}$

3. Find all asymptotes for the graphs of the following functions:

(a) $f(x) = \dfrac{\sin x}{x^2+4}$

(b) $f(x) = \dfrac{\sin x}{x-\pi}$

(c) $f(x) = \dfrac{\sin x}{x-1}$

(d) $f(x) = \tan x$

(e) $f(x) = \sec \dfrac{x}{2}$

(f) $f(x) = \arctan x$

(g) $f(x) = \arctan(x/\pi)$

(h) $f(x) = \dfrac{\arctan x}{x}$

(i) $f(x) = \tan \pi x$

(j) $f(x) = \sec(x^2)$

4. Give an example of a function $y = f(x)$ that is defined for all $x > 0$, whose graph has no slanted asymptote, but which still satisfies

$$\lim_{x\to\infty} \frac{f(x)}{x} = 1.$$

5. Do a proof! Derive the formulas (25) from the definition of slanted asymptote in §17.3.

About finding proofs: you can use all the material in this chapter of the text. Proofs usually don't "just come to you." Instead they tend to require some puzzling and most often our first few written attempts at the proof belong in the trash. The proof we keep for posterity should be a short and polished write-up of the argument we find.

6. Suppose the function $y = f(x)$ has exactly one vertical asymptote, which happens to be located at $x = 1$. Which of the following functions have vertical asymptotes, and where are they located?

(a) $g(x) = f(2x)$

(b) $h(x) = xf(x)$

(c) $h(x) = (x-1)f(x)$

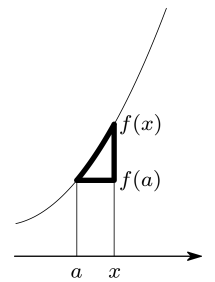
(d) $k(x) = f(x) + \sin x$

CHAPTER 4

Derivatives (2)

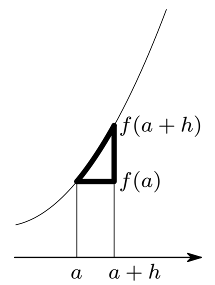
"Leibniz never thought of the derivative as a limit"

www-history.mcs.st-and.ac.uk/Biographies/Leibniz.html



In Chapter 2 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 us to compute derivatives without always having to use the basic definition.

1. Derivatives Defined



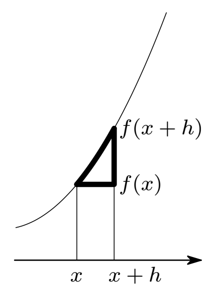
1.1. Definition. *Let $f(x)$ be a function that is defined on an interval (c, d) and let a be a number in this interval.*

*We say that $f(x)$ is **differentiable at** a if the limit*

$$\lim_{x\to a}\frac{f(x)-f(a)}{x-a} \tag{26}$$

*exists, and (if it exists) we call the value of the limit the **derivative of the function** f **at** a, and denote it as $f'(a)$.*

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

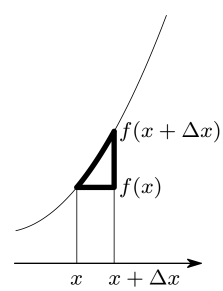


1.2. Other notations. We can substitute $x = a + h$ in the limit (26) and let $h \to 0$ instead of $x \to a$. This gives the formula

$$f'(a) = \lim_{h\to 0}\frac{f(a+h)-f(a)}{h}, \tag{27}$$

Often this equation is written with x instead of a,

$$f'(x) = \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}, \tag{28}$$



and Δx instead of h, which makes it look like this:

$$f'(x) = \lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}. \tag{29}$$

The interpretation is the same as in equation (8) from § 4 in Chapter II. The numerator $f(x+\Delta x) - f(x)$ represents the amount by which the function value of f changes if we increase its argument x by a (small) amount Δx. If we 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

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

Leibniz (1646–1716)

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

We began the semester by agreeing that there are no "infinitely small numbers", and this makes Leibniz's notation difficult to justify. In spite of this we will often use his notation because it shortens many formulas for derivatives, and it often makes intuitive sense.[1]

If $f(x)$ is a complicated function, we will often use "operator notation" for the derivative $\dfrac{df}{dx}$, by writing it as

$$\frac{d}{dx}\left[f\right].$$

The symbol $\frac{d}{dx}$ is shorthand for "take the derivative of what follows with respect to the variable x".

2. Direct computation of derivatives

2.1. Example – The derivative of $f(x) = x^2$ is $f'(x) = 2x$. We have done this computation before in Chapter II, §2. Using one of our equivalent definitions for the derivative, say (29), we get the same result as before:

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

Leibniz would have written

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

and he would have said (in German) that *"when you increase x by an infinitely small amount dx, then x^2 will increase by the infinitely small amount $2x\,dx$; hence the ratio between the infinitesimal changes in x^2 and x is $2x$."*

[1] So many people use Leibniz' notation that mathematicians have tried hard to create a consistent theory of "infinitesimals" that would allow you to compute with "dx and dy" as Leibniz and his contemporaries would have done. In the mid 20th century such a theory was finally created, and dubbed "non-standard analysis." We won't mention it any further, but, as pointed out in a footnote in Chapter I, Keisler's calculus text using infinitesimals at

http://www.math.wisc.edu/~keisler/calc.html

is aimed at undergraduates, so you could have a look if you're interested.

2.2. The derivative of $g(x) = x$ is $g'(x) = 1$. We'll use the form (28) of the definition of the derivative (since writing h is easier than Δx's). We get:

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

In Leibniz's notation, this says

$$\frac{d}{dx}\,[x] = 1.$$

This is an example where Leibniz' notation is most persuasive. In Leibniz' words, *if you increase x by an infinitely small amount dx, then the increase of x is of course dx, and the ratio of these two infinitely small increases is $dx/dx = 1$.* Sadly, we do not believe in infinitely small numbers, so we cannot take this explanation seriously. The expression $\frac{dx}{dx}$ is not really a fraction since there are no two "infinitely small" quantities dx which we are dividing.

2.3. 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 *an infinitely small increase dx of the quantity x will not change c, so the corresponding infinitely small change in c is $dc = 0$; therefore*

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

2.4. Derivative of x^n for $n = 1, 2, 3, \ldots$. To differentiate $f(x) = x^n$, it turns out to be easier to use the first definition (26). This definition gives

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

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

$$x^{n-1} + x^{n-2}a + x^{n-3}a^2 + \cdots + xa^{n-2} + a^{n-1} = \frac{x^n - a^n}{x-a}. \tag{30}$$

If you don't remember the geometric sum formula, then you could also just verify (30) by carefully multiplying both sides with $x - a$. For instance, when $n = 3$ you would get

$$\begin{array}{rllll} x\cdot(x^2+ax+a^2) = & x^3 & +ax^2+a^2x & & \\ -a\cdot(x^2+ax+a^2) = & & -ax^2-a^2x & -a^3 & \textit{(add)} \\ \hline (x-a)\cdot(x^2+ax+a^2) = & x^3 & & -a^3 & \end{array}$$

With formula (30) in hand we can now easily find the derivative of x^n:

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

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

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

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

$$\frac{d}{dx}\left[x^n\right] = nx^{n-1}.$$

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.

3. Differentiable implies Continuous

3.1. Theorem. *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

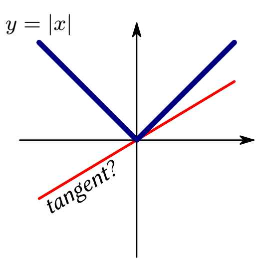
$$\begin{aligned}
\lim_{x\to a} f(x) &= \lim_{x\to a} \big(f(x) - f(a) + f(a)\big) && \text{(algebra)}\\
&= \lim_{x\to a} \left(\frac{f(x) - f(a)}{x - a} \cdot (x - a) + f(a)\right) && \text{(more algebra)}\\
&= \left(\lim_{x\to a} \frac{f(x) - f(a)}{x - a}\right) \cdot \lim_{x\to a} (x - a) + \lim_{x\to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

□

4. Some non-differentiable functions

4.1. A graph with a corner. Consider the function

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



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

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

To see this we can try to compute the derivative at 0: we get

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

But we know this limit does not exist (see Chapter III, §8.2)

If we look at the graph of $f(x) = |x|$ then we can 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 line to the graph at the origin.

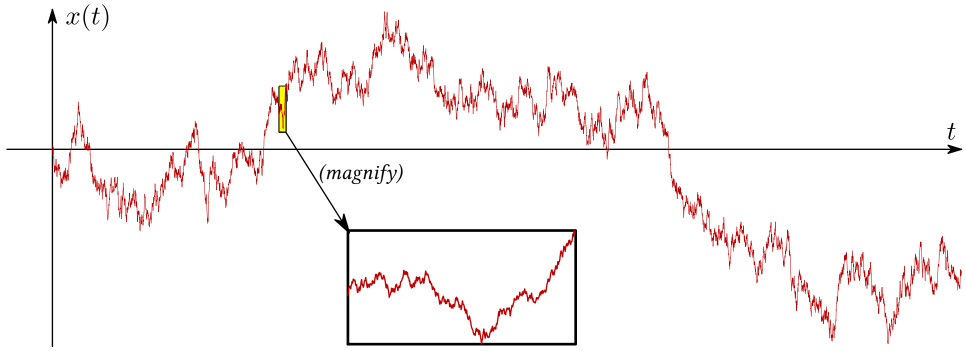


Figure 1. A "Brownian motion" provides an example of a function $y = f(x)$ that is not differentiable at any x. The graph looks like it zig-zags up and down everywhere. If you magnify any small piece of the graph it still has the same jagged appearance.

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

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

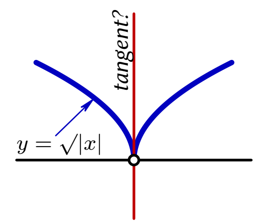
When you try to compute the derivative you get this limit

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

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 (it's "$-\infty$"). Nonetheless, a drawing for the graph of f suggests an obvious tangent line 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 its slope is not defined.



The tangent to the graph of $y = |x|^{1/2}$ at the origin is vertical, so its slope is not defined. The origin is the only point on the graph of $y = |x|^{1/2}$ where the tangent is vertical.

4.3. A graph with absolutely no tangents, *anywhere*. The previous examples were about functions that 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 that are not differentiable at more than one value of x, but here I would like to show you a function f that doesn't have a derivative ***anywhere in its domain.***

To keep things short I won't write a formula for the function, and merely show you a graph. 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, Nasdaq or S&P 500 at the top of the page at `http://finance.yahoo.com` in the afternoon on any weekday.

5. Problems

Compute the derivatives of the following functions, using either (26) or (27).

1. $f(x) = x^2 - 2x$.

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

3. $k(x) = x^3 - 17x$.

4. $u(r) = \frac{2}{1+r}$.

5. $v(\theta) = \sqrt{\theta}$.

6. $\varphi(m) = 1/\sqrt{m}$.

7. $f(t) = \sqrt[3]{t}$

8. $f(t) = \sqrt{1+2t}$

9. $f(x) = 1/x^2$

10. $f(x) = \sqrt{1+2x} + 1/x^2$

11. Which of the following functions is differentiable at $x = 0$?

$$f(x) = x|x|, \qquad g(x) = x\sqrt{|x|},$$
$$h(x) = x + |x|, \qquad k(x) = x^2 \sin\frac{\pi}{x},$$
$$\ell(x) = x \sin\frac{\pi}{x}.$$

These formulas do not define k and ℓ at $x = 0$. We define $k(0) = \ell(0) = 0$.

12. For which value(s) of a and b is the function defined by

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

differentiable at $x = 0$? Sketch the graph of the function f for the values of a and b you found.

13. For which value(s) of a and b is the function defined by

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

differentiable at $x = 1$? Sketch the graph of the function f for the values of a and b you found.

14. For which value(s) of a and b is the function defined by

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

differentiable at $x = 2$? Sketch the graph of the function f for the values of a and b you found.

15. **[Group Problem]** *True or false?* If a function f is continuous at some $x = a$ then it must also be differentiable at $x = a$.

16. **[Group Problem]** *True or false?* If a function f is differentiable at some $x = a$ then it must also be continuous at $x = a$.

6. The Differentiation Rules

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

The situation is analogous to that of the “limit properties” (P_1)–(P_6) from the previous chapter which allowed us to compute limits without always having to go back to the formal definition.

6.1. Sum, Product, and Quotient Rules.
In the following, c and n are constants, f and g are functions of x, and $'$ denotes differentiation. The differentiation rules in function notation, and Leibniz notation, are listed in Table 1.

Note that we already proved the Constant Rule in § 2.2. We will now prove the Sum, Product and Quotient Rules.

6.2. Proof of the Sum Rule. Suppose that $h(x) = f(x) + g(x)$ for all x where f and g are differentiable. Then

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

6.3. Proof of the Product Rule. The Product Rule is strange, at first sight. In fact, Leibniz first thought that it should look more like the Sum Rule, and that $(fg)'$ should equal $f'\,g'$. After a while he discovered his mistake and figured out that $(fg)' = f'\,g + f\,g'$. Below is a proof that this rule is correct. There is also a picture proof, which we get around to in § 6.7.

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

$$\begin{aligned}
h(x) - h(a) &= f(x)g(x) - f(a)f(a)\\
&= f(x)g(x) \underbrace{-f(x)g(a) + f(x)g(a)}_{\text{add and subtract the same term}} - f(a)g(a)\\
&= f(x)\big(f(x) - g(a)\big) + \big(f(x) - g(a)\big)v(a)
\end{aligned}$$

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

$$\begin{aligned}
\lim_{x\to a} \frac{h(x)-h(a)}{x-a} &= \lim_{x\to a} f(x)\frac{g(x)-g(a)}{x-a} + \frac{f(x)-f(a)}{x-a}g(a)\\
&= f(a)g'(a) + f'(a)g(a),
\end{aligned}$$

as claimed. In this last step we have used that

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

Constant Rule:	$c' = 0$	$\dfrac{dc}{dx} = 0$
Sum Rule:	$(f \pm g)' = f' \pm g'$	$\dfrac{d}{dx}(f \pm g) = \dfrac{df}{dx} \pm \dfrac{dg}{dx}$
Product Rule:	$(f \cdot g)' = f' \cdot g + f \cdot g'$	$\dfrac{d}{dx}(f \cdot g) = \dfrac{df}{dx} \cdot g + f \cdot \dfrac{dg}{dx}$
Quotient Rule:	$\left(\dfrac{f}{g}\right)' = \dfrac{f' \cdot g - f \cdot g'}{g^2}$	$\dfrac{d}{dx}\left(\dfrac{f}{g}\right) = \dfrac{\frac{df}{dx} \cdot g - f \cdot \frac{dg}{dx}}{g^2}$

Table 1. The differentiation rules: if you know the derivatives of two functions $u(x)$ and $v(x)$, then these rules tell you what the derivatives of their sum, product and quotient are.

and also that

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

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

6.4. Proof of the Quotient Rule. We can break the proof into two parts. First we do the special case where $h(x) = 1/g(x)$, and then we use the Product Rule to differentiate

$$q(x) = \frac{f(x)}{g(x)} = f(x)\cdot\frac{1}{g(x)}\ .$$

So let $h(x) = 1/g(x)$. We can express the change in h in terms of the change in g:

$$h(x) - h(a) = \frac{1}{g(x)} - \frac{1}{g(a)} = \frac{g(x)-g(a)}{g(x)g(a)}.$$

Dividing by $x - a$, we get

$$\frac{h(x)-h(a)}{x-a} = \frac{1}{g(x)g(a)}\frac{g(x)-g(a)}{x-a}.$$

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

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

Therefore we find

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

That completes the first step of the proof. In the second step, we use the Product Rule to differentiate $q = f/g$:

$$q' = \left(\frac{f}{g}\right)' = \left(f\cdot\frac{1}{g}\right)' = f'\cdot\frac{1}{g} + f\cdot\left(\frac{1}{g}\right)' = \frac{f'}{g} - f\frac{g'}{g^2} = \frac{f'g - fg'}{g^2}.$$

6.5. A shorter, but not quite perfect derivation of the Quotient Rule. The Quotient Rule can be derived from the Product Rule as follows: for $q = f/g$ then

$$q\cdot g = f \tag{31}$$

By the Product Rule, we have

$$q'\cdot g + q\cdot g' = f',$$

so that

$$q' = \frac{f' - q\cdot g'}{g} = \frac{f' - (f/g)\cdot g'}{g} = \frac{f'\cdot g - f\cdot g'}{g^2}.$$

Unlike the proof in §6.4 above, this argument does not prove that q is differentiable if f and g are. It only says that ***if the derivative exists*** then it must be what the Quotient Rule says it is.

The trick we have used here is a special case of a method called “implicit differentiation”, which we will discuss much more in § 15.

6.6. Differentiating a constant multiple of a function. Note that the rule

$$(cf)' = c\,f' \quad\text{or}\quad \frac{d\,cu}{dx} = c\frac{du}{dx} \quad\text{or}\quad \frac{d}{dx}(cf) = c\frac{df}{dx}$$

follows from the Constant Rule and the Product Rule.

6.7. Picture of the Product Rule. If u and v are quantities that 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

$$\Delta(uv) = (u + \Delta u)(v + \Delta v) - uv = u\Delta v + v\Delta u + \Delta u\Delta v. \tag{32}$$

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 (32), 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:

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

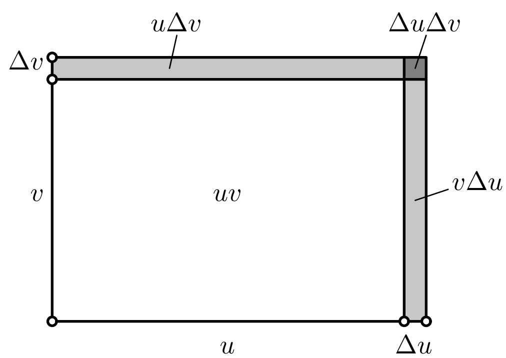


Figure 2. The Product Rule. *How much does the area of a rectangle change if its sides u and v are increased by Δu and Δv?* Most of the increase is accounted for by the two thin rectangles whose areas are $u\Delta v$ and $v\Delta u$. So the increase in area is approximately $u\Delta v + v\Delta u$, which explains why the product rule says $(uv)' = uv' + vu'$.

7. Differentiating powers of functions

7.1. 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.)

In the first step you write $f(x)$ as the product of two functions

$$f(x) = u_1(x) \times \big(u_2(x)u_3(x)\cdots u_n(x)\big),$$

would get

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

In the second step you apply the product rule to $(u_2 u_3 \cdots u_n)'$. This yields

$$\begin{aligned} f' &= u_1' u_2 \cdots u_n + u_1 \left[u_2' u_3 \cdots u_n + u_2 (u_3 \cdots u_n)' \right] \\ &= u_1' u_2 \cdots u_n + u_1 u_2' u_3 \cdots u_n + u_1 u_2 \big(u_3 \cdots u_n \big)'. \end{aligned}$$

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

$$\big(u_1 \cdots u_n\big)' = u_1' u_2 \cdots u_n + u_1 u_2' u_3 \cdots u_n + \cdots + u_1 u_2 u_3 \cdots u_n'. \tag{33}$$

7.2. The Power rule. If all n factors in the previous paragraph are the same, i.e. $u_1(x) = u_2(x) = \cdots = u_n(x) = u(x)$, then the function f is the n^{th} power of some other function,

$$f(x) = \big(u(x)\big)^n.$$

If we use the product rule to find the derivative of $f(x)$, we find that all terms in the right hand side of (33) are the same. Since there are n of them, we get

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

or, in Leibniz' notation,

$$\frac{du^n}{dx} = nu^{n-1} \frac{du}{dx}. \tag{34}$$

7.3. The Power Rule for Negative Integer Exponents. We have just proved the power rule (34) 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

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

Since m is a positive integer, we can use (34), so $(u^m)' = mu^{m-1}$, and hence

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

7.4. The Power Rule for Rational Exponents. So far we have proved that the power law holds if the exponent n is an integer.

We now 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 Section 15.

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

$$w'(x) = \frac{p}{q} u(x)^{\frac{p}{q}-1} u'(x) \tag{35}$$

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.*

7.5. 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.

7.6. Example – differentiate a polynomial. Using the Differentiation Rules we can easily differentiate any polynomial and hence any rational function. For example, using the Sum Rule, the Power Rule, and the rule $(cu)' = cu'$, we see that the derivative of the polynomial

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

is

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

7.7. 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

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

By comparing this example with the previous one, it seems to be the case that polynomials simplify when we differentiate them, while rational functions become more complicated.

7.8. 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 $u(x) = x$.

8. Problems

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?

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?

3. Prove the statement in §6.6, that $(cu)' = c(u')$ follows from the Product Rule.

Compute the derivatives of the following functions (Try to simplify your answers.)

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

5. $f(x) = \dfrac{x-2}{x^4+1}$

6. $f(x) = \left(\dfrac{1}{1+x}\right)^{-1}$

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

8. $f(x) = \dfrac{ax+b}{cx+d}$

9. $f(x) = \dfrac{1}{(1+x^2)^2}$

10. $f(x) = \dfrac{x}{1+\sqrt{x}}$

11. $f(x) = \sqrt{\dfrac{1-x}{1+x}}$

12. $f(x) = \sqrt[3]{x+\sqrt{x}}$

13. $\varphi(t) = \dfrac{t}{1+\sqrt{t}}$

14. $g(s) = \sqrt{\dfrac{1-s}{1+s}}$

15. $h(\rho) = \sqrt[3]{\rho+\sqrt{\rho}}$

16. **[Group Problem]** *Using derivatives to approximate numbers.*

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

(b) Use **(a)** 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* (26).]

(c) Approximate in the same way the numbers $\sqrt{99}$ and $\sqrt{101}$. (Hint: $10 \times 10 = 100$.)

17. *(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 $f'(x)/f(x)$):

$$f(x) = x, \quad g(x) = 3x,$$
$$h(x) = x^2 \quad k(x) = -x^2,$$
$$\ell(x) = 2007x^2, \quad m(x) = x^{2007}$$

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

$$\frac{(uv)'}{uv} = \frac{u'}{u} + \frac{v'}{v}$$
$$\frac{(u/v)'}{u/v} = \frac{u'}{u} - \frac{v'}{v}$$
$$\frac{(u^n)'}{u^n} = n\,\frac{u'}{u}$$

18. **(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 $f(x) = g(x)^{-1}$?

19. **(a)** Let

$$x(t) = (1-t^2)/(1+t^2),$$
$$y(t) = 2t/(1+t^2),$$
$$u(t) = y(t)/x(t).$$

Find dx/dt, dy/dt.

(b) Now that you've done **(a)** there are two different ways of finding du/dt. Do both of them, and compare the results you get.

9. Higher Derivatives

9.1. The derivative of a function 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

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

9.2. Two examples. If $f(x) = x^2 - 2x + 3$ and $g(x) = x/(1-x)$ then

$$\begin{aligned}
f(x) &= x^2 - 2x + 3 & g(x) &= \frac{x}{1-x} \\
f'(x) &= 2x - 2 & g'(x) &= \frac{1}{(1-x)^2} \\
f''(x) &= 2 & g''(x) &= \frac{2}{(1-x)^3} \\
f^{(3)}(x) &= 0 & g^{(3)}(x) &= \frac{2 \cdot 3}{(1-x)^4} \\
f^{(4)}(x) &= 0 & g^{(4)}(x) &= \frac{2 \cdot 3 \cdot 4}{(1-x)^5} \\
&\vdots & &\vdots
\end{aligned}$$

All further derivatives of f are zero, but no matter how often we differentiate $g(x)$ we will never get zero. Instead of multiplying the numbers in the numerator of the derivatives of g we left them as "$2 \cdot 3 \cdot 4$." A good reason for doing this is that we can see a pattern in the derivatives, which would allow us to guess what (say) the 10th derivative is, without actually computing ten derivatives:

$$g^{(10)}(x) = \frac{2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10}{(1-x)^{11}}.$$

9.3. 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

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

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

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

10. Problems

1. The equation

$$\frac{2x}{x^2 - 1} = \frac{1}{x+1} + \frac{1}{x-1} \qquad (\dagger)$$

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 (†).

2. Compute the first, second and third derivatives of the following functions:

(a) $f(x) = (x+1)^4$

(b) $g(x) = \left(x^2+1\right)^4$

(c) $h(x) = \sqrt{x-2}$

(d) $k(x) = \sqrt[3]{x - \frac{1}{x}}$

3. Find the derivatives of 10^{th} order of the following functions. (The problems have been chosen so that after doing the first few derivatives in each case, you should start seeing a pattern that will let you guess the 10^{th} derivative without actually computing 10 derivatives.)

(a) $f(x) = x^{12} + x^8$.

(b) $g(x) = 1/x$.

(c) $h(x) = 12/(1-x)$.

(d) $k(x) = x^2/(1-x)$.

(e) $\ell(x) = 1/(1-2x)$

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

4. Find $f'(x)$, $f''(x)$ and $f^{(3)}(x)$ if

$$f(x) = 1+x+\frac{x^2}{2}+\frac{x^3}{6}+\frac{x^4}{24}+\frac{x^5}{120}+\frac{x^6}{720}.$$

5. **[Group Problem]** Suppose we have these two functions

$$f(x) = \frac{1}{x+2}, \quad g(x) = \frac{x}{x+2}.$$

(a) Find the 12^{th} derivative of $f(x)$.

(b) Find the n^{th} order derivative of $f(x)$ (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 $g(x)$.

(d) Find $h^{(n)}(x)$ where $h(x) = \dfrac{x^2}{x+2}$.

6. **[Group Problem]** *(Mostly about notation)*

(a) Find dy/dx and d^2y/dx^2 if $y = x/(x+2)$.

(b) Find du/dt and d^2u/dt^2 if $u = t/(t+2)$.

(c) Find $\dfrac{d}{dx}\left(\dfrac{x}{x+2}\right)$ and $\dfrac{d^2}{dx^2}\left(\dfrac{x}{x+2}\right)$.

(d) Compute

$$A = \frac{d\dfrac{x}{x+2}}{dx} \text{ at } x = 1,$$

and

$$B = \frac{d\dfrac{1}{1+2}}{dx}.$$

(e) Simplicio just thought of this argument: *if you set $x = 1$ in $x/(x+2)$ you get $1/(1+2)$; therefore these two quantities have the same derivative when $x = 1$, and hence in the previous problem you should have found $A = B$.*

Is he right? What do you conclude from his reasoning?

7. Find d^2y/dx^2 and $(dy/dx)^2$ if $y = x^3$.

11. Differentiating trigonometric functions

The trigonometric functions sine, cosine and tangent are differentiable, and their derivatives are given by the following formulas

$$\frac{d\sin x}{dx} = \cos x, \qquad \frac{d\cos x}{dx} = -\sin x, \qquad \frac{d\tan x}{dx} = \frac{1}{\cos^2 x}. \tag{36}$$

Note the minus sign in the derivative of the cosine!

11.1. The derivative of $\sin x$.

By definition one has

$$\sin'(x) = \lim_{h\to 0} \frac{\sin(x+h) - \sin(x)}{h}. \tag{37}$$

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 Chapter III, § 15 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.

12. Problems

Find the derivatives of the following functions (try to simplify your answers)

1. $f(x) = \sin(x) + \cos(x)$

2. $f(x) = 2\sin(x) - 3\cos(x)$

3. $f(x) = 3\sin(x) + 2\cos(x)$

4. $f(x) = x\sin(x) + \cos(x)$

5. $f(x) = x\cos(x) - \sin x$

6. $f(x) = \dfrac{\sin x}{x}$

7. $f(x) = \cos^2(x)$

8. $f(x) = \sqrt{1-\sin^2 x}$

9. $f(x) = \sqrt{\dfrac{1-\sin x}{1+\sin x}}$

10. $\cot(x) = \dfrac{\cos x}{\sin x}$.

11. 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 $x = \pi/4$?

12. Can you find a and b so that the function

$$f(x) = \begin{cases}\tan x & \text{for } x < \frac{\pi}{4}\\ a+bx & \text{for } x \ge \frac{\pi}{4}\end{cases}$$

is differentiable at $x = \pi/4$?

13. 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?

14. Simplicio claims that our formula for the derivative of $\tan x$ must be wrong because he has a book that says

$$\frac{d\tan x}{dx} = 1 + \tan^2 x.$$

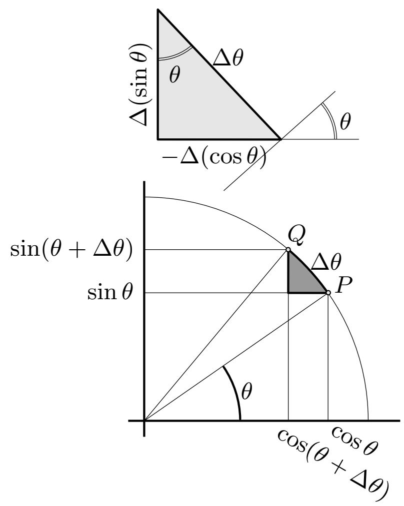
Show that both formulas are right.

15. Find the first, second and third derivatives of the functions

$$f(x) = \tan^2 x \text{ and } g(x) = \frac{1}{\cos^2 x}.$$

Hint: remember your trig to reduce work!

16. *A geometric derivation of the derivatives of sine and cosine.*



In the lower picture you see a unit circle, a point P with angle θ, whose coordinates are $(\cos\theta, \sin\theta)$. The figure also shows the point Q you get if you increase the angle θ by a small amount $\Delta\theta$. Drawing horizontal and vertical lines through these two points then produces the shaded triangular area. The picture above the circle is a magnification of the shaded triangle. The hypotenuse of the triangle is not straight since it is part of the unit circle, but the smaller you make $\Delta\theta$ the closer it will be to a straight line segment. The lengths of the sides of the triangle are $\Delta\theta$, the change in $\sin\theta$ and the change of $\cos\theta$. Since it is a right triangle you can express

$$\frac{\Delta\sin\theta}{\Delta\theta} \text{ and } \frac{\Delta\cos\theta}{\Delta\theta}$$

as the sine and/or cosine of some angle. Find the relevant angle, and show how this leads to our formulas for the derivatives of $\sin\theta$ and $\cos\theta$.

Make sure you explain where the "$-$" in the derivative of $\cos\theta$ comes from.

17. **[Group Problem]** *(Yet another derivation of the formula for $d\sin x/dx$)*

In § 11.1 above we used the addition formula for $\sin(\alpha+\beta)$ to compute the limit (37). Instead of using the addition formula, use this formula from trigonometry

$$\sin\alpha - \sin\beta = 2\sin\Big(\frac{\alpha-\beta}{2}\Big)\cos\Big(\frac{\alpha+\beta}{2}\Big)$$

to compute the limit (37).

13. The Chain Rule

13.1. Composition of functions. Given two functions f and g, we 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 composite function is the set of all numbers x for which this formula gives we something well-defined.

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. See Figure 3 for a concrete example of such a chain of dependencies, and thus of composing two functions.

13.2. Example. Suppose we have three quantities, x, y, and z which are related by

$$z = y^2 + y, \text{ and } y = 2x + 1.$$

Here we have a chain of dependencies, with z depending on y and y depending on x. By combining (composing) these dependencies you see that z depends on x via

$$z = y^2 + y = (2x+1)^2 + 2x + 1.$$

A depends on B depends on C depends on…

Someone is pumping water into a balloon. Assuming that the balloon is a sphere, we can say how large it is by specifying its radius R. For an expanding balloon this radius will change with time t.

The volume of the balloon 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) tells us the radius R of the balloon at time t,

$$R = f(t),$$

and the second function (g) tells us the volume of the balloon given its radius

$$V = g(R).$$

The volume of the balloon at time t is then given by

$$V = g\big(f(t)\big) = g \circ f(t),$$

and it is the function that tells us the volume of the balloon at time t, and is the composition of first f and then g.

Schematically we can summarize this chain of cause-and-effect relations as follows: we could either say that V depends on R, and R depends on t,

$$\boxed{\text{time } t} \xrightarrow{f} \boxed{\begin{array}{c}\text{radius } R\\ \text{depends}\\ \text{on time } t\end{array}} \xrightarrow{g} \boxed{\begin{array}{c}\text{volume } V\\ \text{depends}\\ \text{on radius}\\ R\end{array}}$$

or we could say that V depends directly on t:

$$\boxed{\text{time } t} \xrightarrow{g \circ f} \boxed{\begin{array}{c}\text{volume } V\\ \text{depends on}\\ \text{time } t\end{array}}$$

Figure 3. A "real world example" of a composition of functions.

We can express the same ideas in function notation. The first dependency $z = y^2 + y$ defines a function, which we'll call f. It is given by

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

The second dependency, $y = 2x + 1$, also defines a function

$$g(x) = 2x + 1.$$

The way the quantity z depends on x also defines a function, namely, the composition of f and g. The composition is

$$f \circ g(x) = f\big(2x + 1\big) = (2x + 1)^2 + (2x + 1)$$

In other notation,

$$\left.\begin{array}{r} z = f(y) \\ y = g(x) \end{array}\right\} \implies z = f\big(g(x)\big) = f \circ g(x).$$

One says that ***the composition of*** f ***and*** g ***is the result of substituting*** g ***in*** f.

How can we find the derivative of the composition of two functions? This is what the Chain Rule tells us:

13.3. Theorem (Chain Rule). *If f and g are differentiable, so is the composition $f \circ g$. The derivative of $f \circ g$ is given by*

$$(f \circ g)'(x) = f'(g(x))\, g'(x).$$

The Chain Rule looks a lot simpler, perhaps even "obvious", if we write it in Leibniz notation. Let's translate from function notation to "$d$this$/d$that":

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,

$$\frac{dz}{dx} = (f \circ g)'(x), \quad \frac{dz}{dy} = f'(y) \quad \text{and } \frac{dy}{dx} = g'(x)$$

so now the Chain Rule says

$$\frac{dz}{dx} = \frac{dz}{dy}\,\frac{dy}{dx}. \tag{38}$$

13.4. 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 (38), 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's form of the Chain Rule.

13.5. Second proof of the Chain Rule. We verify the formula in Theorem 13.3 at some arbitrary value $x = a$. Specifically, 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

$$f'(g(a)) = \lim_{y \to g(a)} \frac{f(y) - f(g(a))}{y - g(a)}.$$

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 $f'(g(a))$

$$f'(g(a)) = \lim_{y\to g(a)} \frac{f(y) - f(g(a))}{y - g(a)} = \lim_{x\to a} \frac{f(g(x)) - f(g(a))}{g(x) - g(a)}. \tag{39}$$

(Keep in mind that as $x \to a$ one has $y \to g(a)$.) By putting all this together, we 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.[2]

13.6. Using the Chain Rule – 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, we could simply differentiate the last formula:

$$\frac{dz}{dx} = \frac{d(4x^2 + 6x + 2)}{dx} = 8x + 6. \tag{40}$$

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

$$\frac{dz}{dx} = \frac{dz}{dy}\,\frac{dy}{dx} = (2y + 1)\cdot 2 = 4y + 2. \tag{41}$$

The two answers (40) and (41) should be the same. Once you remember that $y = 2x + 1$ you see that this is indeed true:

$$y = 2x + 1 \implies 4y + 2 = 4(2x + 1) + 2 = 8x + 6.$$

---

[2] Briefly, one must 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 at $y = a$, and then use

$$f(g(x)) - f(g(a)) = h(g(x))\big(g(x) - g(a)\big).$$

The two computations of dz/dx therefore lead to the same answer (which they should!). In this example there was no clear advantage in using the Chain Rule.

13.7. Example where we 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 $f(x) = \sin(\pi x)$.

The function $f(x) = \sin \pi x$ is the composition of two simpler functions, namely

$$f(x) = g(h(x)) \text{ where } g(u) = \sin u \text{ and } h(x) = \pi x.$$

We know how to differentiate each of the two functions g and h:

$$g'(u) = \cos u, \quad h'(x) = \pi.$$

Therefore the Chain Rule implies that

$$f'(x) = g'(h(x))h'(x) = \cos(\pi x) \cdot \pi = \pi \cos \pi x.$$

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

$$\frac{d \sin 2x}{dx} \overset{u=2x}{=} \frac{d \sin u}{dx} = \frac{d \sin u}{du} \cdot \frac{du}{dx} = \cos u \cdot 2 = 2\cos 2x.$$

In words, this computation tells us this:

- when x changes, $\sin(2x)$ changes $\cos(2x)$ times as fast as $2x$,
- and $2x$ changes twice as fast as x,
- therefore $\sin 2x$ changes $\cos(2x) \times 2$ times as fast as x.

13.8. 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 the Power Rule.

13.9. A more complicated example. Let us 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 we replaced u by its definition in terms of x) is important because the problem was presented to us with only x as a variable, while u was a variable we introduced ourselves 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

$$\frac{d}{dx}\left[4 + \sqrt{7+x^3}\right]$$

we could set $u = 7 + x^3$, and compute

$$\frac{d}{dx}\left[4 + \sqrt{7+x^3}\right] = \frac{d}{du}\left[4 + \sqrt{u}\right] \cdot \frac{du}{dx}.$$

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

$$\frac{d}{dx}(4 + \sqrt{7+x^3}) = \frac{1}{2\sqrt{7+x^3}} \cdot 3x^2 = \frac{3x^2}{2\sqrt{7+x^3}},$$

doing the computation of $\dfrac{d\, 4+\sqrt{u}}{du}$ in your head.

13.10. Related Rates – the volume of an inflating balloon. You can use the Chain Rule to compute the derivative of the composition of two functions, but that's not all it's good for. A very common use of the Chain Rule arises when you have two related quantities that are changing in time. If you know the relation between the quantities, then the Chain Rule allows you to find a relation between the rates at which these quantities change in time. To see a concrete example, let's go back to the inflating water balloon (Figure 3) of radius

$$R = f(t).$$

The volume of this balloon 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 balloon and the rate of change of the balloon's radius with time. From

$$\frac{dV}{dR} = \frac{d}{dR}\left[\frac{4}{3}\pi R^3\right] = 4\pi R^2$$

we see that

$$\frac{dV}{dt} = 4\pi R^2\,\frac{dR}{dt}.$$

For instance, if the radius of the balloon is growing at $0.5\mathrm{in/sec}$, and if its radius is $R = 3.0\mathrm{in}$, then the volume is growing at a rate of

$$\frac{dV}{dt} = 4\pi(3.0\mathrm{in})^2\cdot 0.5\mathrm{in/sec} \approx 57\mathrm{in}^3/\mathrm{sec}.$$

13.11. 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

$$\left.\frac{dy}{dx}\right|_{x=4} = \left.\frac{dy}{du}\right|_{u=6}\cdot\left.\frac{du}{dv}\right|_{v=25}\cdot\left.\frac{dv}{dx}\right|_{x=4} = -\frac{1}{7}\cdot\frac{1}{10}\cdot 8.$$

14. Problems

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 $u = g(x)$.

2. Repeat the previous exercise with

$$y = (1+\sqrt{1+x})^3.$$

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?

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.

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 $w'(x)$.

6. Compute the derivatives of the following functions:

(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) $f(x) = \cos^2 x - \cos(x^2)$

7. **[Group Problem]** Moe is pouring water into a glass. At time t (seconds) the height of the water in the glass is $h(t)$ (inches). The ACME glass company, which made the glass, says that the volume in the glass to height h is $V = 1.2\, h^2$ (fluid ounces).

(a) The water height in the glass is rising at 2 inches per second at the moment that the height is 2 inches. How fast is Moe pouring water into the glass?

(b) If Moe pours water at a rate of 1 ounce per second, then how fast is the water level in the glass going up when it is 3 inches?

(c) Moe pours water at 1 ounce per second, and at some moment the water level is going up at 0.5 inches per second. What is the water level at that moment?

8. Find the derivative of $f(x) = x \cos \frac{\pi}{x}$ at the point C in Figure 3.

9. Suppose that $f(x) = x^2 + 1$, $g(x) = x + 5$, and

$$v = f \circ g, \qquad w = g \circ f,$$
$$p = f \cdot g, \qquad q = g \cdot f.$$

Find $v(x)$, $w(x)$, $p(x)$, and $q(x)$.

10. **[Group Problem]** 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

$$v(x) = f(g(x)), \qquad w(x) = g(f(x)),$$
$$p(x) = f(x)g(x), \qquad q(x) = g(x)f(x).$$

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.

11. 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 $(x, y) = (3, 7)$.

12. There is a function f whose second derivative satisfies

$$f''(x) = -64 f(x). \tag{†}$$

for all x.

(a) One such function is $f(x) = \sin ax$, provided you choose the right constant a. Which value should a have?

(b) For which choices of the constants A, a and b does the function $f(x) = A \sin(ax + b)$ satisfy (†)?

13. **[Group Problem]** A cubical sponge, hereafter referred to as 'Bob', is absorbing water, which causes him to expand. The length of his side at time t is $S(t)$ units. His volume is $V(t)$ cubic units.

(a) Find a function f so that $V(t) = f(S(t))$.

(b) Describe the meaning of the derivatives $S'(t)$ and $V'(t)$ (in words). 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 Bob's volume is 8 cubic inches, he is absorbing water at a rate of 2 cubic inches per minute. How fast is his side $S(t)$ growing?

15. Implicit differentiation

15.1. The recipe. Recall that an implicitly 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 implicitly defined function.

(1) Differentiate the equation $F(x, y) = 0$; you may need the chain rule to deal with the occurrences 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

$$G(x, y)\frac{dy}{dx} + H(x, y) = 0, \tag{42}$$

where G and H are expressions containing x and y but not the derivative.

(3) Solve the equation in step 2 for dy/dx:

$$\frac{dy}{dx} = -\frac{H(x, y)}{G(x, y)} \tag{43}$$

(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 (43) 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 (43) is as far as you can go.

Observe that by following this procedure you will get a formula for the derivative dy/dx that contains both x and y.

15.2. 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 $y = f(x)$ that satisfies

$$y^2 + x = xy$$

you rewrite this equation as

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

and set $F(x, y) = y^2 + x - xy$.

15.3. 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 of implicit differentiation (i.e. the recipe above).

The direct approach goes like this:

$$\begin{aligned} f'(x) &= \frac{d\bigl(1-x^4\bigr)^{1/4}}{dx} \\ &= \tfrac14\bigl(1-x^4\bigr)^{-3/4}\frac{d(1-x^4)}{dx} \\ &= \tfrac14\bigl(1-x^4\bigr)^{-3/4}\bigl(-4x^3\bigr) \\ &= -\frac{x^3}{\bigl(1-x^4\bigr)^{3/4}} \end{aligned}$$

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

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

The defining function is therefore $F(x,y) = x^4 + y^4 - 1$.

Differentiate both sides with respect to x and don't forget that $y = f(x)$, so y here is a function of x. You get

$$\frac{dx^4}{dx} + \frac{dy^4}{dx} - \frac{d1}{dx} = 0 \implies 4x^3 + 4y^3\frac{dy}{dx} = 0,$$

$$\text{i.e. } H(x,y) + G(x,y)\frac{dy}{dx} = 0.$$

The expressions G and H from equation (42) in the recipe are $G(x,y) = 4y^3$ and $H(x,y) = 4x^3$.

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

$$f'(x) = \frac{dy}{dx} = -\frac{x^3}{y^3} = -\frac{x^3}{\bigl(1-x^4\bigr)^{3/4}}.$$

15.4. Another example. Let $y = f(x)$ be a function defined by

$$2y + \sin y - x = 0.$$

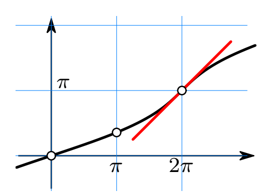
It turns out to be impossible to find a formula that tells you what $f(x)$ is for any given x (i.e. there's no formula for the solution y of the equation $2y + \sin y = x$.) But you can find many points on the graph by picking some y value and computing the corresponding x.

For instance, if $y = \pi$ then $x = 2\pi$, so that $f(2\pi) = \pi$: the point $(2\pi, \pi)$ lies on the graph of f.

Implicit differentiation lets us compute the derivative $f'(2\pi)$, or "dy/dx at $x = 2\pi$." To find that derivative we differentiate the defining equation

$$\frac{d(2y + \sin y - x)}{dx} = \frac{d\,0}{dx} \implies 2\frac{dy}{dx} + \cos y\frac{dy}{dx} - \frac{dx}{dx} = 0 \implies (2+\cos y)\frac{dy}{dx} - 1 = 0.$$

Solve for $\frac{dy}{dx}$ and you get



The graph of $x = 2y + \sin y$ contains the point $(2\pi, \pi)$. What is the slope of the tangent at that point?

$$f'(x) = \frac{1}{2+\cos y} = \frac{1}{2+\cos f(x)}.$$

If we were asked to find $f'(2\pi)$ then, since we know $f(2\pi) = \pi$, we could answer

$$f'(2\pi) = \frac{1}{2+\cos\pi} = \frac{1}{2-1} = 1.$$

The graph of f also contains a point with $x = \pi$, but we don't know the y coordinate of that point, i.e. we don't know $f(\pi)$. If we wanted to know the slope of the tangent at this point, i.e. $f'(\pi)$, then all we would be able to say is

$$f'(\pi) = \frac{1}{2+\cos f(\pi)}.$$

To say more we would first have to find $y = f(\pi)$, which would require us to solve

$$2y + \sin y = \pi.$$

There is no simple formula for the solution, but using a calculator you can find approximations of the solution y.

15.5. Theorem – The Derivatives of Inverse Trigonometric Functions. *The derivatives of arc sine and arc tangent are*

$$\frac{d \arcsin x}{dx} = \frac{1}{\sqrt{1-x^2}}, \qquad \frac{d \arctan x}{dx} = \frac{1}{1+x^2}.$$

Proof: Recall that by definition the function $y = \arcsin x$ satisfies $x = \sin y$, and $-\frac{\pi}{2} \le y \le \frac{\pi}{2}$. Differentiate this relation,

$$\frac{dx}{dx} = \frac{d \sin y}{dx}$$

and apply the chain rule. You get

$$1 = (\cos y)\,\frac{dy}{dx},$$

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.

16. Problems

For each of the following problems use Implicit Differentiation to find the derivative $f'(x)$ if $y = f(x)$ satisfies the given equation. 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.

1. $xy = \dfrac{\pi}{6}$
2. $\sin(xy) = \frac{1}{2}$
3. $\dfrac{xy}{x+y} = 1$
4. $x + y = xy$
5. $(y-1)^2 + x = 0$
6. $(y+1)^2 + y - x = 0$
7. $(y-x)^2 + x = 0$
8. $(y+x)^2 + 2y - x = 0$
9. $\left(y^2-1\right)^2 + x = 0$
10. $\left(y^2+1\right)^2 - x = 0$
11. $x^3 + xy + y^3 = 3$
12. $\sin x + \sin y = 1$
13. $\sin x + xy + y^5 = \pi$
14. $\tan x + \tan y = 1$

For each of the following explicitly defined functions find an implicit definition that does not involve taking roots. Then use this description to find the derivative dy/dx.

15. $y = f(x) = \sqrt{1-x}$
16. $y = f(x) = \sqrt[4]{x+x^2}$
17. $y = f(x) = \sqrt{1-\sqrt{x}}$
18. $y = f(x) = \sqrt[4]{x-\sqrt{x}}$
19. $y = f(x) = \sqrt[3]{\sqrt{2x+1}-x^2}$
20. $y = f(x) = \sqrt[4]{x+x^2}$
21. $y = f(x) = \sqrt[3]{x-\sqrt{2x+1}}$
22. $y = f(x) = \sqrt[4]{\sqrt[3]{x}}$
23. **[Group Problem]** *(Inverse trig review)* Simplify the following expressions, and indicate for which values of x (or θ, or ...) your simplification is valid. Figure 9 from Chapter I may be helpful. In case of doubt, try plotting the function on a graphing calculator (or just Google the graph, i.e. to find a picture of the graph of $y = \sin(x)$ just type `y=sin(x)` into the Google search box).

 (a) $\sin \arcsin x$ **(b)** $\tan \arctan z$
 (c) $\cos \arcsin x$ **(d)** $\tan \arcsin \theta$
 (e) $\arctan(\tan\theta)$ **(f)** $\arcsin(\sin x)$
 (g) $\cot \arctan x$ **(h)** $\cot \arcsin x$

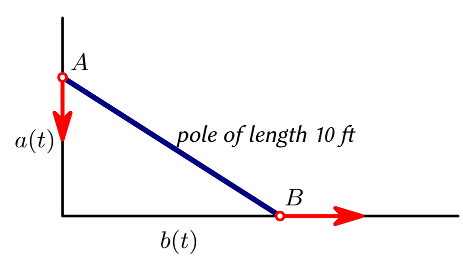
Now that you know the derivatives of arcsin and arctan, you can find the derivatives of the following functions. What are they?

24. $f(x) = \arcsin(2x)$
25. $f(x) = \arcsin\sqrt{x}$
26. $f(x) = \arctan(\sin x)$
27. $f(x) = \sin \arctan x$
28. $f(x) = \left(\arcsin x\right)^2$
29. $f(x) = \dfrac{1}{1+(\arctan x)^2}$
30. $f(x) = \sqrt{1-(\arcsin x)^2}$
31. $f(x) = \dfrac{\arctan x}{\arcsin x}$

17. Problems on Related Rates

1. A particle moves on the hyperbola $x^2 - 18y^2 = 9$ in such a way that its y coordinate increases at a constant rate of 9 units per second. How fast is the x-coordinate changing when $x = 9$?

2. 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?



3. A pole of length 10 feet 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?

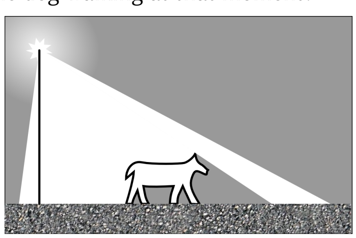
4. A 13 meter long pole is leaning against a wall. The bottom of the pole is pulled along the ground away from the wall at the rate of 2 m/s. How fast is its height on the wall decreasing when the foot of the pole is 5 m away from the wall?

5. **[Group Problem]** A video 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 video 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 video camera points toward the rocket.)

6. **[Group Problem]** 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 dog's shadow is moving away from the streetlight at 5 feet per second. How fast is the dog walking at that moment?



7. The sides of an isosceles triangle are rotating: 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 $\theta = 60°$?

8. 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 $\theta = \frac{\pi}{3}$.

9. 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 $(2, 0)$,

(c) the angle $\angle ORQ$ where R is the point $(2, 0)$.

10. 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 P,

(d) the angle $\angle OPQ$ where Q is the point $(0, 3)$ and O is the origin $(0, 0)$.

11. **(a)** The kinetic energy of a moving object with mass m and velocity v is $K = \frac{1}{2}mv^2$. If an object of mass $m = 5$ grams is accelerating at a rate of $9.8\ \mathrm{m/sec}^2$, how fast is the kinetic energy increasing when the speed is $30\ \mathrm{m/sec}$?

(b) The sum of all the forces acting on the moving object is, according to Newton, $F = ma$, where $a = \frac{dv}{dt}$. Show that the rate at which the kinetic energy K of the object is changing is

$$\frac{dK}{dt} = Fv.$$

12. Here are three versions of the same problem:

ffft During a chemotherapy treatment, a tumor of spherical shape decreases in size at a rate proportional to its surface area. Show that the tumor's radius decreases at a constant rate.

fftt Helium is released from a spherical balloon so that its volume decreases at a rate proportional to the balloon's surface area. Show that the radius of the balloon decreases at a constant rate.

ffQu A spherical snowball melts at a rate proportional to its surface area. Show that its radius decreases at a constant rate.

13. Suppose that we have two resistors connected in parallel with resistances R_1 and R_2 measured in ohms (Ω). The total resistance, R, is then given by:

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}.$$

Suppose that R_1 is increasing at a rate of $0.3\Omega/$min and R_2 is decreasing at a rate of $0.5\Omega/$min. At what rate is R changing when $R_1 = 80\Omega$ and $R_2 = 100\Omega$?

14. A runner sprints around a circular track of radius 100 meters at a constant speed of 7 m/s. The runner's friend is standing at a distance 130 meters from the center of the track. How fast is the distance between two friends changing when the distance between them is 130 meters?

15. Ship A is 50 miles north of ship B and is sailing due south at a constant speed of 25 mph. Ship B is sailing due east at a constant speed of 20 mph. At what rate is the distance between the ships changing after one hour? Is the distance increasing or decreasing?

16. A conical water tank with vertex down has a radius of 12 ft at the top and is 30 ft high. If water flows out of the tank at a rate of 14 ft^3/min, how fast is the depth of the water decreasing when the water is 20 ft deep?

17. A train, starting at noon, travels east at 50 mph, while another train leaves an hour later from the same point, traveling north at 90 mph. How fast are the trains moving apart at 3 pm?

18. The radius of a right circular cylinder is increasing at a rate of 3 in/min and the height is decreasing at a rate of 5 in/min. At what rate is the volume changing when the radius is 10 in and the height is 15 in? Is the volume increasing of decreasing?

19. A hemispherical bowl of radius 10 cm contains water to a depth of h cm. Find the radius r of the surface of the water as a function of h. The water level drops at a rate of 0.1 cm/hour. At what rate is the radius of water decreasing when the depth is 5 cm?

20. A cylindrical swimming pool (whose center axis is vertical) is being filled from a fire hose at rate of 5 cubic feet per second. If the pool is 40 feet across, how fast is the water level increasing when the pool is one third full?

21. Consider a clock whose minute hand is 5 cm long and whose hour hand is 4 cm long.

(a) Find the rate of change of the angle between the minute hand and hour hand when it is $3:00$ o'clock.

(b) Find the rate of change of the distance between the hands when it is $3:00$ o'clock.

22. A baseball diamond has the shape of a square, and each side is 80 feet long. A player is running from second to third base, and he is 60 feet from reaching third. He is running at a speed of 28 feet per second. At what rate is the player's distance from home plate decreasing?

23. Doppler radar measures the rate of change of the distance from an object to the observer. A police officer 10 meters from a straight road points a radar gun at a car traveling along the road, 20 meters away, and measures a speed of 30m/sec. What is the car's actual speed?

24. A particle A moves on a circular trajectory centered at the origin O, and of radius 1, so that the angle α subtended by the line segment OA and the x-axis changes at a constant rate k. Determine the rates of change of the x and y coordinates of the particle.

25. A crystal in the shape of a cube dissolves in acid so that the edge of the cube decreases by 0.3 mm/min. How fast is the volume of the cube changing when the edge is 5 mm?

CHAPTER 5

Graph Sketching and Max-Min Problems

When we first discuss functions, one thing we want to know how to do is determine what the graph of a function looks like. Naively, we can produce graphs by plugging in a number of values into a function and connecting the points with a smooth curve. However, this is an unsatisfying method for graphing for a variety of reasons: first, how do we know that if we find a few points on a graph, that joining them up will actually produce the "real" shape of the curve? Second, how can we know where the "interesting" features of the graph are? And third, how can we be sure that we have not missed anything?

We can provide good answers to all of these questions using calculus: specifically, by using the first and second derivatives of a function, we can find essentially all of the interesting features of the function's graph.

1. Tangent and Normal lines to a graph

The slope of the tangent to the graph of f at the point $(a, f(a))$ is

$$m = f'(a) \tag{44}$$

and hence the equation for the tangent line is

$$y = f(a) + f'(a)(x - a). \tag{45}$$

The normal line at a point $(a, f(a))$ on the graph of a function is defined to be the line perpendicular to the tangent line passing through that point. The slope of the normal line to the graph is $-1/m$ and thus one could write the equation for the normal as

$$y = f(a) - \frac{x - a}{f'(a)}. \tag{46}$$

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 (46), but instead one has the simpler equation

$$x = a.$$

Both cases are covered by this form of the equation for the normal

$$x = a + f'(a)(f(a) - y). \tag{47}$$

Both (47) and (46) are formulas that you shouldn't try to memorize. It is better to remember that if the slope of the tangent is $m = f'(a)$, then the slope of the normal is $-1/m$.

2. The Intermediate Value Theorem

It is said that a function is continuous if you can draw its graph without taking your pencil off the paper. A more precise version of this statement is the *Intermediate Value Theorem*:

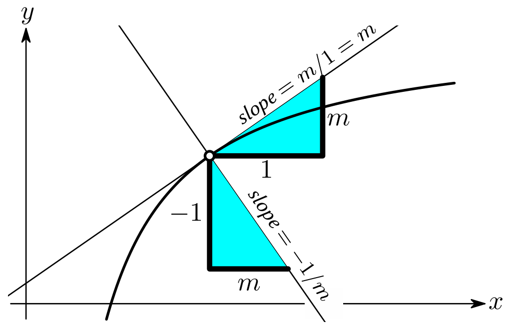


Figure 1. Why "slope of normal $= \dfrac{-1}{\text{slope of tangent}}$."

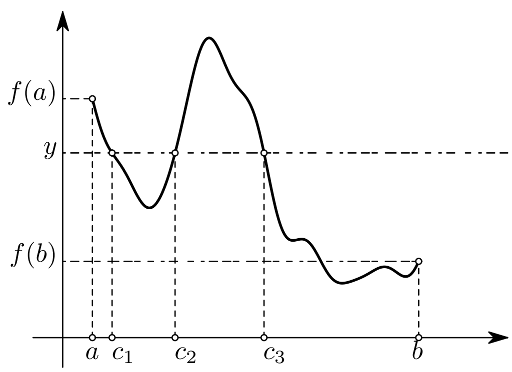


Figure 2. The Intermediate Value Theorem says that a continuous function must attain any given value y between $f(a)$ and $f(b)$ at least once. In this example there are three values of c for which $f(c) = y$ holds.

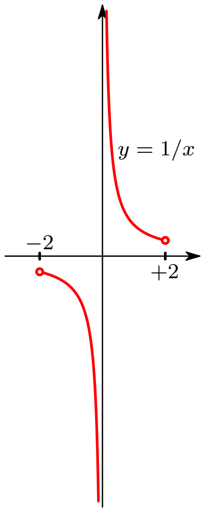
2.1. Intermediate Value Theorem. *If f is a continuous function defined on an interval $a \le x \le b$, and if y is any number between $f(a)$ and $f(b)$, then there is a number c with $a \le c \le b$ such that $f(c) = y$.*

Here "y between $f(a)$ and $f(b)$" means either $f(a) \le y \le f(b)$ or $f(b) \le y \le f(a)$ depending on which of $f(a)$ or $f(b)$ is larger.

The theorem should be intuitively obvious: to draw the graph you have to start at $(a, f(a))$ and somehow move your pencil to $(b, f(b))$. If you do this without taking your pencil off the paper, then at some point the pencil will have to cross the horizontal line at height y if y is between $f(a)$ and $f(b)$. That's what the theorem says.

However, this argument is about pencils and paper and it is not about real numbers and solving equations. It turns out to be difficult to give a complete proof of the theorem. In particular, it requires a much better description of what a real number is than we are using in this course.

2.2. Example – Square root of 2. Consider the function $f(x) = x^2$, which we know is continuous. 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".



2.3. Example – The equation $2\theta + \sin\theta = \pi$. Consider the function $f(x) = 2x + \sin x$. It is a continuous function at all x, so from $f(0) = 0$ and $f(\pi) = 2\pi$ it follows that there is a number θ between 0 and π such that $f(\theta) = \pi$. In other words, the equation

$$2\theta + \sin\theta = \pi \tag{48}$$

has a solution θ with $0 \le \theta \le \pi$. Unlike the previous example, where we knew the solution was $\sqrt{2}$, there is no simple formula for the solution to (48).

2.4. Example – "Solving $1/x = 0$". If we attempt to apply the Intermediate Value Theorem to the function $f(x) = 1/x$ on the interval $[a, b] = [-2, 2]$, then we see that for any y between $f(a) = f(-2) = -\frac{1}{2}$ and $f(b) = f(2) = \frac{1}{2}$ there is a number c in the interval $[-2, 2]$ such that $1/c = y$. For instance, we could choose $y = 0$ (which lies between -2 and $+2$), and conclude that there is some c with $-2 \le c \le 2$ and $1/c = 0$.

But there is no such c, because $1/c$ is never zero! So we have done something wrong: the mistake we made is that we overlooked that our function $f(x) = 1/x$ is not defined on the ***entire*** interval $-2 \le x \le 2$ because it is not defined at $x = 0$. *The moral: always check the hypotheses of a theorem before you use it!*

3. Problems

1. Where does the normal line to the graph of $y = x^2$ at the point (1,1) intersect the x-axis?

2. Where does the tangent line to the graph of $y = x^2$ at the point (a, a^2) intersect the x-axis?

3. Where does the normal line to the graph of $y = x^2$ at the point (a, a^2) intersect the x-axis?

4. Find all points on the parabola with the equation $y = x^2 - 1$ such that the normal line at the point goes through the origin.

5. Where does the normal to the graph of $y = \sqrt{x}$ at the point $(a, \sqrt{a})$ intersect the x-axis?

6. Does the graph of $y = x^4 - 2x^2 + 2$ have any horizontal tangents? If so, where?

Does the graph of the same function have any vertical tangents?

Does it have vertical normals?

Does it have horizontal normals?

7. 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?

8. The line from the origin to the point $(a, f(a))$ on the graph of $f(x) = 1/x^2$ is perpendicular to the tangent line to that graph. What is a?

9. The line from the origin to the point $(a, f(a))$ on the graph of $f(x) = 4/x$ is perpendicular to the tangent line to the graph. What is a?

10. Find equations for the tangent and normal lines

to the curve ...	**at the point...**
(a) $y = 4x/(1 + x^2)$	$(1, 2)$
(b) $y = 8/(4 + x^2)$	$(2, 1)$
(c) $y^2 = 2x + x^2$	$(2, 2)$
(d) $xy = 3$	$(1, 3)$

11. [Group Problem] The function

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

satisfies $f(-1) = -2$ and $f(+1) = +2$, so, by the Intermediate Value Theorem, there should be some value c between -1 and $+1$ such that $f(c) = 0$. ***True or False?***

12. Find the equation for the tangents to the graph of the Backward Sine

$$f(x) = \sin\left(\frac{\pi}{x}\right)$$

at the points $x = 1$, $x = \frac{1}{2}$ and at D (see Figure 2 in §8.3.)

13. Find the equation for the tangent to the graph of the Backward Cosine in a Bow Tie $g(x) = x\cos\left(\frac{\pi}{x}\right)$ at the point C (see Figure 3 in §11.3)

4. Finding sign changes of a function

The Intermediate Value Theorem implies the following very useful fact.

4.1. Theorem. *If f is a continuous function on an interval $a \le x \le b$, and if $f(x) \ne 0$ for all x in this interval, then $f(x)$ does not change its sign in the interval $a \le x \le b$: either $f(x)$ is positive for all $a \le x \le b$, or it is negative for all $a \le x \le b$.*

PROOF. The theorem says that there cannot be two numbers $a \le x_1 < x_2 \le b$ 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 would be a c with $f(c) = 0$. But we are assuming that $f(c) \ne 0$ whenever $a \le c \le b$. This is a contradiction, so no such x_1 and x_2 can exist. □

4.2. Example. Consider

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

The zeros of f (the solutions of $f(x) = 0$) are $-\frac{1}{2}, 1, 3$. These numbers split the real line into four intervals

$$(-\infty, -\tfrac{1}{2}), \quad (-\tfrac{1}{2}, 1), \quad (1, 3), \quad (3, \infty).$$

Theorem 4.1 tells us that $f(x)$ cannot change its sign in any of these intervals. For instance, $f(x)$ has the same sign for all x in the first interval $(\infty, -\frac{1}{2})$. Now we choose a number we like from this interval, say -1, and determine that the sign of $f(-1)$: $f(-1) = (-4)(-2)^2(-1)^3$ is positive. Therefore $f(x) > 0$ for all x in the interval $(-\infty, -\frac{1}{2})$. In the same way we find

$$\begin{aligned}
f(-1) = (-4)(-2)^2(-1)^3 > 0 \quad &\Longrightarrow \quad f(x) > 0 \text{ for } x < -\tfrac{1}{2}\\
f(0) = (-3)(-1)^2(1)^3 < 0 \quad &\Longrightarrow \quad f(x) < 0 \text{ for } -\tfrac{1}{2} < x < 1\\
f(2) = (-1)(1)^2(5)^3 < 0 \quad &\Longrightarrow \quad f(x) < 0 \text{ for } 1 < x < 3\\
f(4) = (1)(3)^2(9)^3 > 0 \quad &\Longrightarrow \quad f(x) > 0 \text{ for } x > 3.
\end{aligned}$$

If we know all the zeroes of a continuous function, then this method allows us to decide where the function is positive or negative. However, when the given function is factored into easy functions, as in this example, there is a different way of finding the signs of f. 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

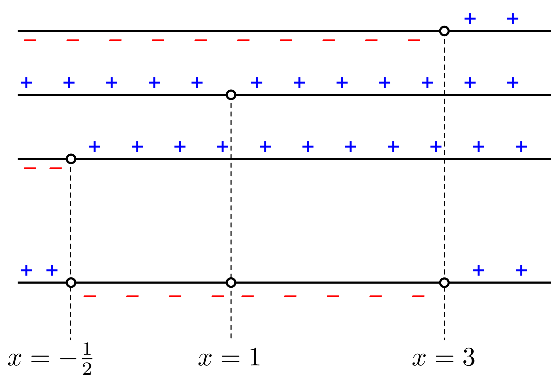
- $x-3$ is positive for $x > 3$ and negative for $x < 3$;
- $(x-1)^2$ is always positive (except at $x = 1$);
- $(2x+1)^3$ is positive for $x > -\frac{1}{2}$ and negative for $x < -\frac{1}{2}$.

Multiplying these signs we get the same conclusions as above. We can summarize this computation in the following diagram:

5. Increasing and decreasing functions

Here are four very similar definitions – look closely to see how they differ.

- A function is called ***increasing*** on an interval if $a < b$ implies $f(a) < f(b)$ for all numbers a and b in the interval.
- A function is called ***decreasing*** on an interval if $a < b$ implies $f(a) > f(b)$ for all numbers a and b in the interval.
- The function f is called ***non-decreasing*** on an interval if $a < b$ implies $f(a) \le f(b)$ for all numbers a and b in the interval.



- The function f is called ***non-increasing*** on an interval if $a < b$ implies $f(a) \geq f(b)$ for all numbers a and b in the interval.

We can summarize these definitions as follows:

$on an interval, f$ **is ...**	**if for all a and b in the interval...**
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 definitions of increasing, decreasing, non-increasing, and non-decreasing do not seem like they are easy to check for a particular function. But, in fact, there is an easy test for whether a function is increasing or decreasing: look at the sign of the derivative of f. More precisely:

5.1. Theorem. *If a differentiable function is non-decreasing on an interval then $f'(x) \geq 0$ for all x in that interval.*

If a differentiable function is non-increasing on an interval then $f'(x) \leq 0$ for all x in that interval.

The idea of the proof is the following: if f is non-decreasing, then for any given x and any positive Δx, we have $f(x + \Delta x) \geq f(x)$, so

$$\frac{f(x + \Delta x) - f(x)}{\Delta x} \geq 0.$$

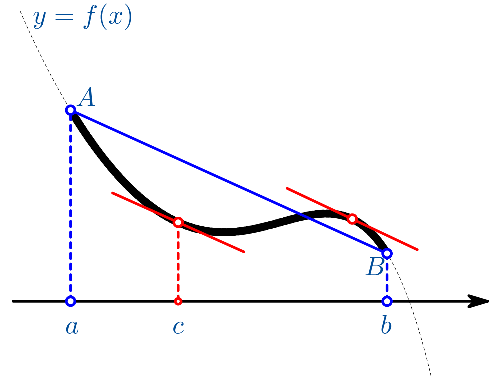
Now if we let $\Delta x \searrow 0$, we see

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

What about the converse? In other words, if we know the sign of f' then what can we say about f? For this we have the following

5.2. Theorem. *Suppose f is a differentiable function on an interval (a, b).*
If $f'(x) > 0$ for all $a < x < b$, then f is increasing on the interval.
If $f'(x) < 0$ for all $a < x < b$, then f is decreasing on the interval.

The Mean Value Theorem says that if you pick any two points A and B on the graph of a differentiable function, then there always is a point on the graph between A and B where the tangent line to the graph is parallel to the line segment AB. Depending on the shape of the graph and the location of the points A, B, there may even be more than one such point (there are two in this figure). The MVT is used in the proofs of many facts about functions; for instance, Theorem 5.2. Here is a picture proof of the Mean Value Theorem:



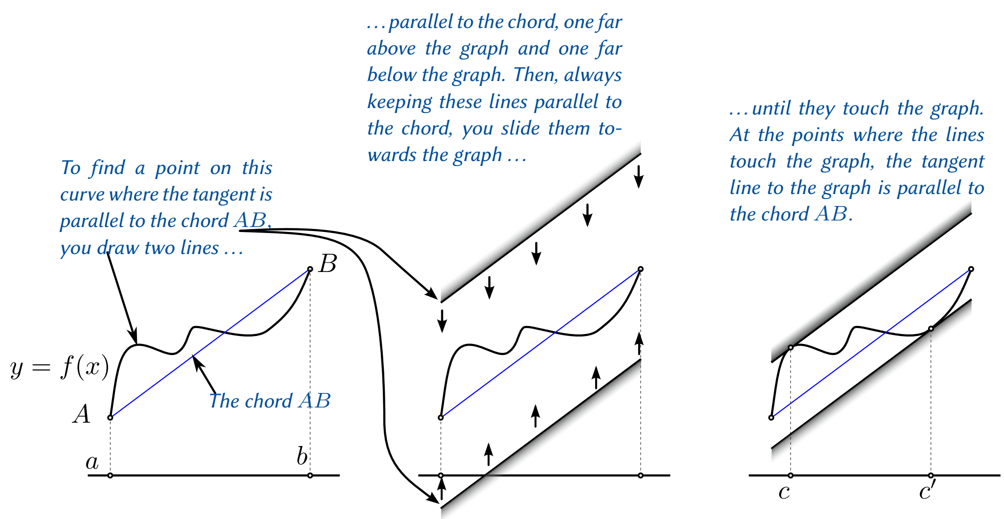


Figure 3. The Mean Value Theorem

The proof is based on the Mean Value Theorem, which also finds use in many other situations:

5.3. 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 5.2. We show that $f'(x) > 0$ for all x implies that f is increasing. (The proof of the other statement is very similar.) 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 $f(x_2) - f(x_1) > 0$, or $f(x_2) > f(x_1)$. □

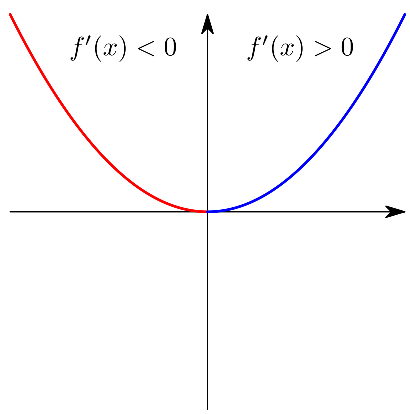
6. Examples

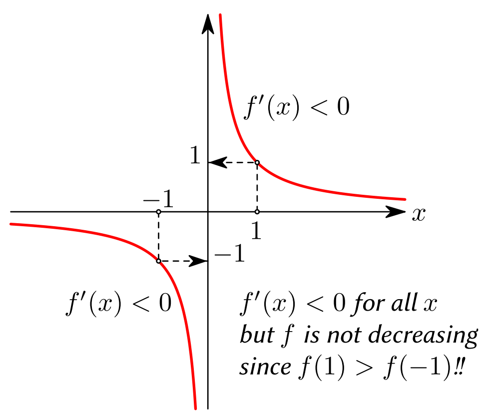
Armed with these theorems we can now split the graph of any function into increasing and decreasing parts simply by computing the derivative $f'(x)$ and finding out where $f'(x) > 0$ and where $f'(x) < 0$, and to do this we need only apply the method from 4 to f' rather than f.

6.1. Example: the parabola $y = x^2$. The familiar graph of $f(x) = x^2$ consists of two parts, one decreasing and one increasing, separated by $x = 0$. 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*** $x < 0$ and ***increasing for*** $x > 0$**.**





6.2. 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. One might 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 we 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 attempted to use theorem 5.2, but this function only 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 we can't conclude that $f(x) = 1/x$ is increasing from $x = -1$ to $x = +1$.

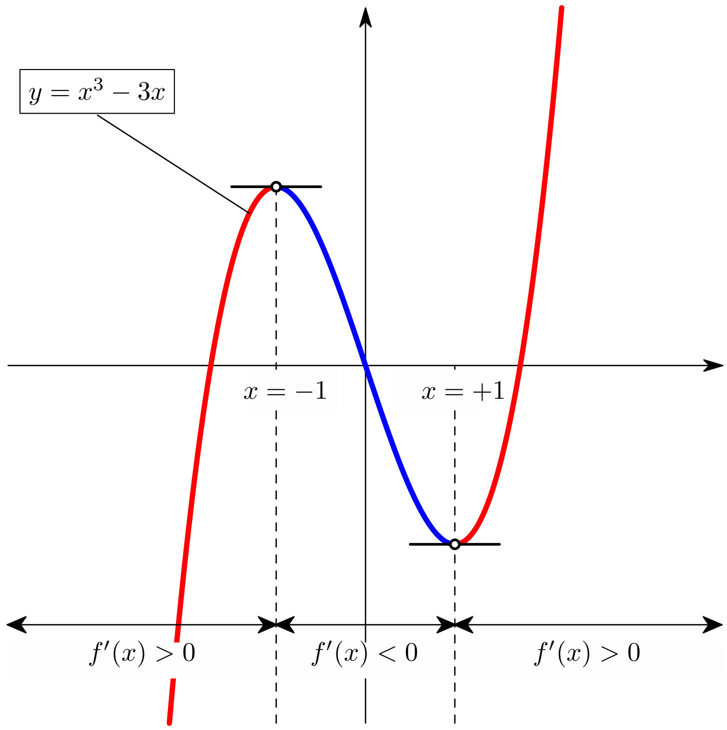


Figure 4. The graph of $f(x) = x^3 - 3x$.

On the other hand, the function is defined and differentiable on the interval $0 < x < \infty$, so theorem 5.2 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 $1/x$.

6.3. Graph of a cubic function. Consider the function

$$y = f(x) = x^3 - 3x.$$

Its derivative is $f'(x) = 3x^2 - 3$. We try to find out where f' is positive, and where it is negative by factoring $f'(x)$:

$$f'(x) = 3(x^2 - 1) = 3(x - 1)(x + 1)$$

from which we see that

$$\begin{aligned} &f'(x) > 0 \text{ for } x < -1 \\ &f'(x) < 0 \text{ for } -1 < x < 1 \\ &f'(x) > 0 \text{ for } x > 1 \end{aligned}$$

Therefore the function f is

$$\text{increasing on } (-\infty, -1), \quad \text{decreasing on } (-1, 1), \quad \text{increasing on } (1, \infty).$$

At the two points $x = \pm 1$ one has $f'(x) = 0$ so there the tangent is horizontal. This leads to Figure 4.

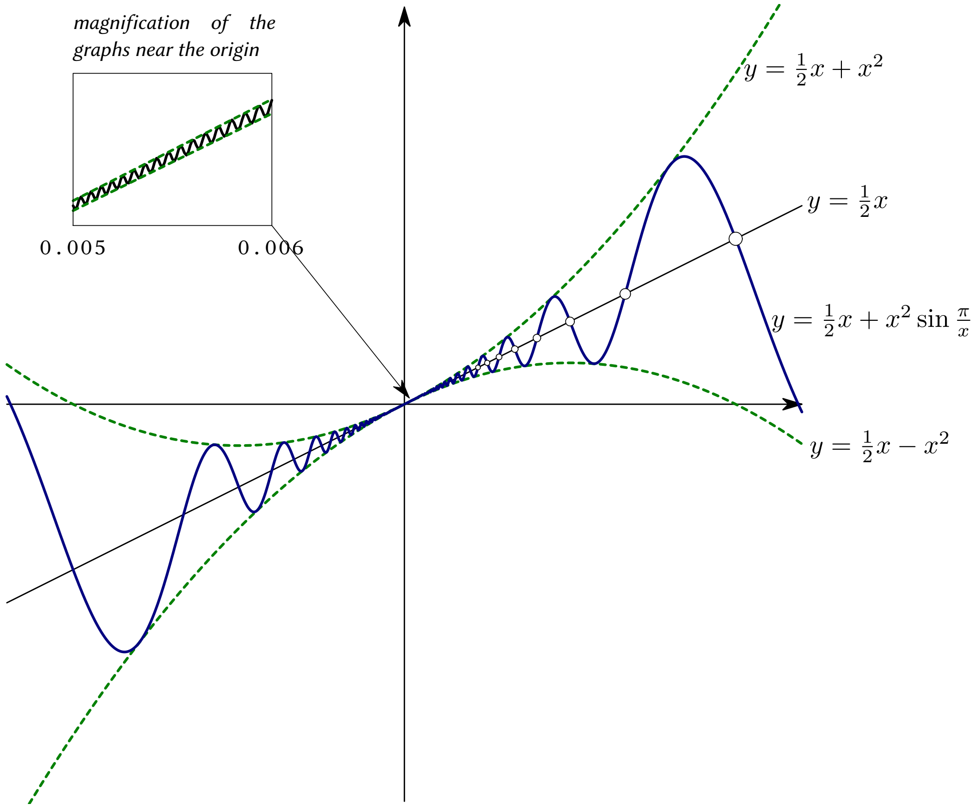


Figure 5. The function $y = \frac{1}{2}x + x^2 \sin \frac{\pi}{x}$ is differentiable, and at $x = 0$ its derivative is positive (namely, $1/2$). You would think that this means that the function therefore has to be increasing near $x = 0$, but this example shows that the function doesn't have to be increasing near $x = 0$. You can see that by zooming in on the graph near the origin. As you follow the graph of $y = \frac{1}{2}x + x^2 \sin \frac{\pi}{x}$ to the origin it alternates between increasing and decreasing infinitely often.

6.4. A function whose tangent turns up and down infinitely often near the origin. We end this section with a weird example. Somewhere in the mathematician's zoo of curious functions the following will be on exhibit. Consider the function

$$f(x) = \frac{x}{2} + x^2 \sin \frac{\pi}{x}.$$

(See Figure 5.) For $x = 0$ this formula is undefined so that we still have to define $f(0)$. We choose $f(0) = 0$. 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 $-|x| \leq x \sin \frac{\pi}{x} \leq |x|$, as in § 11.3.) The slope of the tangent to the graph at the origin is positive ($\frac{1}{2}$), and you would *think* that the function should be "increasing near $x = 0$" But this turns out not to be true!

To see why not, compute the derivative of this function for $x \neq 0$:

$$f'(x) = \frac{1}{2} - \pi \cos \frac{\pi}{x} + 2x \sin \frac{\pi}{x}.$$

We could try to find where $f'(x)$ is positive, and where it's negative, just as we did for $y = x^3 - 3x$ in § 6.3, but for this function solving $f'(x) = 0$ turns out to be pretty much impossible. Fortunately, even though we can't solve $f'(x) = 0$, we can still find out something about the derivative by looking at the intersection points of the graph with the line $y = x/2$ and checking the sign of $f'(x)$ at those points. We will find that $f'(x)$ flip-flops infinitely often between positive and negative as $x \searrow 0$.

There are infinitely many intersection points of $y = f(x)$ with $y = x/2$ and the coordinates are

$$x_k = \frac{1}{k}, \quad y_k = f(x_k).$$

For larger and larger k the points (x_k, y_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 $x = x_k$ is given by

$$\begin{aligned} f'(x_k) &= \frac{1}{2} - \pi \cos \frac{\pi}{1/k} + 2\frac{1}{k} \sin \frac{\pi}{1/k} \\ &= \frac{1}{2} - \pi \underbrace{\cos k\pi}_{=(-1)^k} + \frac{2}{k} \underbrace{\sin k\pi}_{=0} \\ &= \begin{cases} \frac{1}{2} - \pi < 0 & \text{for } k \text{ even} \\ \frac{1}{2} + \pi > 0 & \text{for } k \text{ odd} \end{cases} \end{aligned}$$

Therefore, along the sequence of points (x_k, y_k) the slope of the tangent jumps back and forth 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. We see that *even though the derivative at $x = 0$ of this particular function is positive, there are points on the graph arbitrarily close to the origin where the tangent has negative slope.*

7. Maxima and Minima

A function has a ***global maximum*** at some a in its domain if $f(x) \leq f(a)$ for all other x in the domain of f. Global maxima are sometimes also called "absolute maxima".

A function has a ***global minimum*** at some a in its domain if $f(x) \geq f(a)$ for all other x in the domain of f. Global minima are sometimes also called "absolute minima".

A function has a ***local maximum*** at some a in its domain if there is a small $\delta > 0$ such that $f(x) \leq f(a)$ for all x with $a - \delta < x < a + \delta$ that lie in the domain of f.

A function has a ***local minimum*** at some a in its domain if there is a small $\delta > 0$ such that $f(x) \leq f(a)$ for all x with $a - \delta < x < a + \delta$ that 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.

7.1. How to find local maxima and minima. Any x value for which $f'(x) = 0$ is called a ***stationary point*** for the function f.

7.2. Theorem. *Suppose f is a differentiable function on an 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.

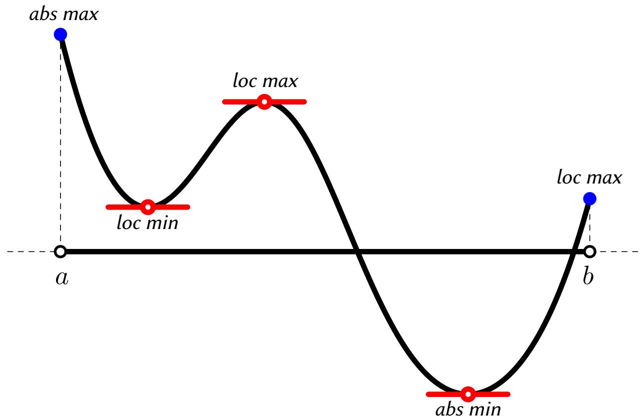


Figure 6. A function defined on an interval $[a, b]$ with one interior absolute minimum, another interior local minimum, an interior local maximum, and two local maxima on the boundary, one of which is in fact an absolute maximum.

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 $f(x+\Delta x) - f(x) \leq 0$ if $-\delta < \Delta x < \delta$. In the first limit we also have $\Delta x < 0$, so that

$$\lim_{\Delta x \nearrow 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} \geq 0.$$

Hence $f'(x) \geq 0$.

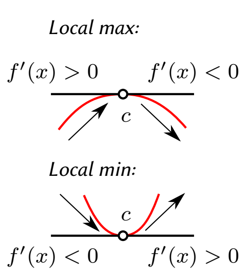
In the second limit we have $\Delta x > 0$, so

$$\lim_{\Delta x \searrow 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} \leq 0,$$

which implies $f'(x) \leq 0$.

Thus we have shown that $f'(x) \leq 0$ and $f'(x) \geq 0$ at the same time. This can only be true if $f'(x) = 0$. □

7.3. 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 sign of $f'(x)$ for x near c.



7.4. Theorem. *Suppose $y = f(x)$ is function that is defined in an interval $(c-\delta, c+\delta)$, where $\delta > 0$ is some (small) number.*

- *If $f'(x) > 0$ for $x < c$ and $f'(x) < 0$ for $x > c$ then f has a local maximum at $x = c$.*
- *If $f'(x) < 0$ for $x < c$ and $f'(x) > 0$ for $x > c$ then f has a local minimum 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:

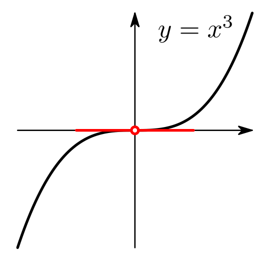
(1) if $f'(x) > 0$ for x between $c-\delta$ and c, then f is increasing for $c-\delta < x < c$ and therefore $f(x) < f(c)$ for x between $c-\delta$ and c.
(2) If in addition $f'(x) < 0$ for $x > c$ then f is decreasing for x between c and $c+\delta$, so that $f(x) < f(c)$ for those x.
(3) Combining (1) & (2) gives $f(x) \le f(c)$ for $c-\delta < x < c+\delta$.

The statement about when f has a local minimum follows from an analogous argument.

7.5. Example – local maxima and minima of $f(x) = x^3 - 3x$. In §6.3 we found that the function $f(x) = x^3 - 3x$ is increasing when $-\infty < x < -1$, and also when $1 < x < \infty$, while it is decreasing when $-1 < x < 1$. It follows that the function has a local maximum at $x = -1$, and a local minimum at $x = 1$.

Neither the local maximum nor the local minimum are global since

$$\lim_{x\to-\infty} f(x) = +\infty \text{ and } \lim_{x\to\infty} f(x) = -\infty.$$



7.6. 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 will 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 7.4 does not say anything.

And in fact, $x = 0$ is neither a local maximum nor a local minimum because $f(x) < f(0)$ for $x < 0$ and $f(x) > 0$ for $x > 0$.

8. Must a function always have a maximum?

Theorem 7.2 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 them.

8.1. Theorem. *Let f be continuous function defined on the **closed** interval $a \le x \le b$. Then f attains its absolute maximum somewhere in this interval, and it also attains its absolute minimum somewhere in the interval. In other words, there exist real numbers c and d between a and b such that*

$$f(c) \le f(x) \le f(d)$$

whenever $a \le x \le b$.

This is another one of those theorems whose proof requires a more careful definition of the real numbers than we have given in Chapter 1. In this course we will take the theorem for granted.

9. 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 8.1 are not met.

9.1. 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 $0 \le x \le 1$?

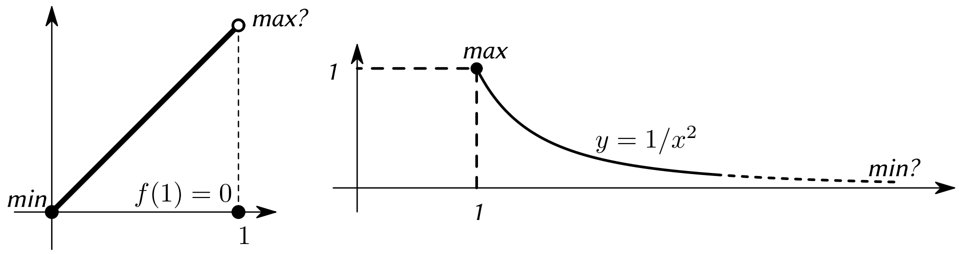


Figure 7. The function on the left has no maximum, and the one on the right has no minimum.

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 to have $f(a) = 1$. But such an a does not exist! Conclusion: this function does ***not*** attain its maximum on the interval $[0, 1]$.

What about Theorem 8.1? That theorem only applies to continuous functions, and the function f in this example is not continuous at $x = 1$. It's not continuous because at $x = 1$ we have

$$f(1) = 0, \text{ which is not the same as } \lim_{x \nearrow 1} f(x) = 1.$$

So all it takes for Theorem 8.1 to fail is that the function f be discontinuous at *just one point*.

9.2. Question: *Does the function*

$$f(x) = \frac{1}{x^2}, \quad 1 \le x < \infty$$

have a maximum or minimum?

Answer: The function has a maximum at $x = 1$, but it has no minimum.

Concerning the maximum: if $x \ge 1$ then $f(x) = 1/x^2 \le 1$, while $f(1) = 1$. Hence $f(x) \le f(1)$ for all x in the interval $[1, \infty)$. By definition, this means that f attains its maximum at $x = 1$.

If we look for a minimal value of f then we note that $f(x) \ge 0$ 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 $1 \le a < \infty$, 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 did Theorem 8.1 fail this time? In this example the function f is continuous on the whole interval $[1, \infty)$, but this interval is not a closed interval of the form $[a, b]$.

10. General method for sketching the graph of a function

Given a differentiable function f defined on an interval $a \le x \le b$, we can find the increasing and decreasing parts of the graph along with 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, then instead of computing the values $f(a)$ or $f(b)$, but instead you should compute $\lim_{x\to\pm\infty} f(x)$.

10.1. Example – the graph of a rational function. Let's sketch the graph of the function

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

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:

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

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

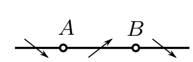
$$A = -1 - \sqrt{2} \text{ and } B = -1 + \sqrt{2}.$$

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?

Since we are dealing with an unbounded interval we must compute the limits of $f(x)$ as $x \to \pm\infty$. We find

$$\lim_{x\to\infty} f(x) = \lim_{x\to-\infty} f(x) = -3.$$

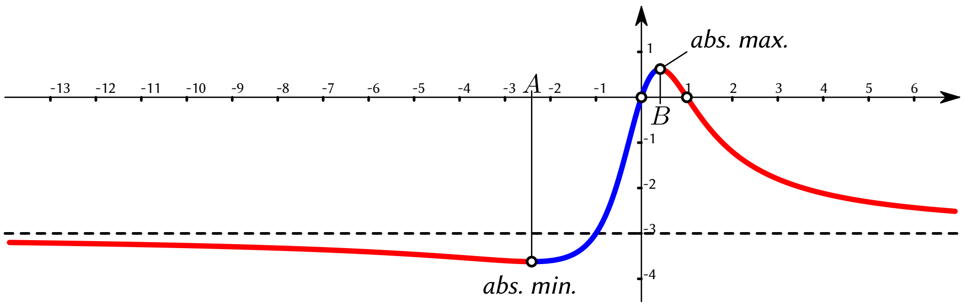


Figure 8. The graph of $f(x) = 3(x - x^2)/(1 + x^2)$

Since f is decreasing between $-\infty$ and A, it follows that

$$f(A) \leq f(x) < -3 \text{ for } -\infty < x \leq A.$$

Similarly, f is decreasing from B to $+\infty$, so

$$-3 < f(x) \leq f(B) \text{ for } B < x < \infty.$$

Between the two stationary points the function is increasing, so

$$f(A) \leq f(x) \leq f(B) \text{ for } A \leq x \leq B.$$

From this it follows that $f(x)$ has a global minimum when $x = A = -1 - \sqrt{2}$ and has a global maximum when $x = B = -1 + \sqrt{2}$.

11. Convexity, Concavity and the Second Derivative

11.1. Definition. *A function f is **convex** on some interval $a < x < b$ if the derivative $f'(x)$ is **increasing** on that interval.*

*Likewise, the function is called **concave** for $a < x < b$ if the derivative $f'(x)$ is **decreasing** on that interval.*

In the following drawings the graph on the left is convex, while the one in the middle

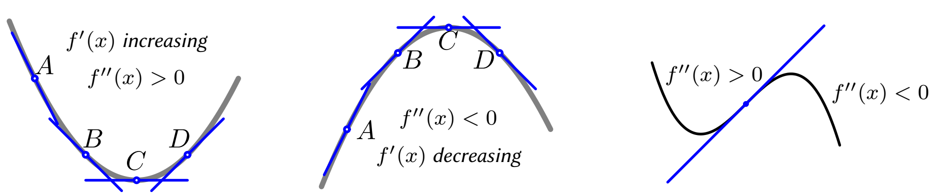


Figure 9. Convex, concave and mixed graphs.

is concave, because, going from left to right along either of these graphs (from A to B to C to D), you see that the slope of the tangent increases on the left graph, and decreases on the middle graph. The third graph is neither convex nor concave, although it does have one concave piece and one convex piece.

By our results on increasing and decreasing functions, we know that f is convex if the derivative of $f'(x)$ is positive. In other words, if $f''(x) > 0$ for $a < x < b$, then $f'(x)$ is increasing and therefore, by definition, $f(x)$ is convex.

11.2. Theorem (convexity and concavity from the second derivative). *If $f''(x)$ is positive for $a < x < b$, then the function f is convex on the interval $a < x < b$.*

If $f''(x)$ is negative for $a < x < b$, then the function f is concave on that interval.

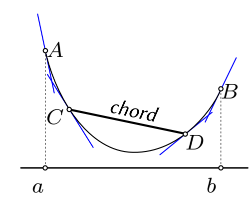
11.3. Definition. *A point on the graph of f where $f''(x)$ changes sign is called an **inflection point**.*

The third graph in Figure 9 shows an inflection point.

11.4. The chords of convex graphs. Along the graph of a convex function the slope of the tangent is increasing as you go from left to right. After you try to draw a couple of such graphs it really looks like they should always be "curved upwards." The intuitive understanding of convex and concave is that convex graphs are curved up, and concave graphs are curved down.

The phrase "curved upwards" turns out to be another one of those expressions whose mathematical definition is not clear (try explaining it with out waving your hands.) The following theorem is an attempt to make a more precise statement that essentially says the graph is curved upwards.

Theorem. *If a function f is convex on some interval $a < x < b$ then the chord connecting any two points on the graph lies above the graph.*



The chord connecting C and D lies above the graph.

While this theorem is still phrased in terms of the picture, the condition that "the chord lies above the graph" can be expressed as an inequality between the function $f(x)$ itself, and the function whose graph is the chord[1]. To prove this theorem requires invoking the Mean Value Theorem.

11.5. Example – the cubic function $f(x) = x^3 - 3x$. The second derivative of the function $f(x) = x^3 - 3x$ is

$$f''(x) = 6x$$

which is positive for $x > 0$ and negative for $x < 0$. Hence, in the graph in §6.3, 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.

11.6. The Second Derivative Test. In §7.3 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.

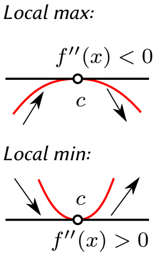
11.7. Theorem. *Suppose c is a stationary point for a function f. If $f''(c) < 0$, then f has a local maximum at $x = c$, and if $f''(c) > 0$, then f has a local minimum at c.*

The theorem doesn't say anything about what happens when $f''(c) = 0$. To determine what happens in that case, one must go back to checking the sign 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 we know that $f'(x)$ is increasing near $x = c$. But this is only possible if $f'(x) < 0$ for $x < c$ and $f'(x) > 0$ for $x > c$ (in some small interval around c). This means the

[1] The precise statement is: a function is convex on the interval $[a, b]$ if, for any t with $0 \le t \le 1$, and any c and d with $a \le c < d \le b$, it is true that

$$f(tc + (1-t)d) \le t\, f(c) + (1-t)\, f(d).$$



function f is decreasing for $x < c$ and increasing for $x > c$, which means it reaches a local minimum at $x = c$. An analogous argument applies for the $f''(c) < 0$ case.

11.8. Example – $f(x) = x^3 - 3x$, again. Consider the function $f(x) = x^3 - 3x$ from §6.3 and §11.5. We found that this function has two stationary points, at $x = \pm 1$. By looking at the sign of $f'(x) = 3x^2 - 3$ we concluded that $x = -1$ is a local maximum while $x = +1$ is a local minimum.

Instead of looking at $f'(x)$, we could also have computed $f''(x)$ at $x = \pm 1$ and then applied the Second Derivative Test: since $f''(x) = 6x$, we have

$$f''(-1) = -6 < 0 \text{ and } f''(+1) = 6 > 0.$$

Therefore f has a local maximum at -1 and a local minimum at $+1$.

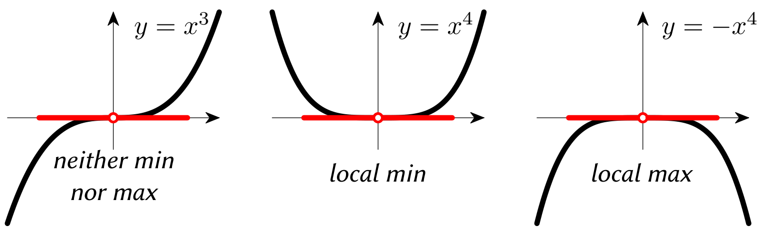


Figure 10. Three functions for which the second derivative test doesn't work.

11.9. 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. Figure 10 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.

12. Problems

1. What does the Intermediate Value Theorem say?

2. What does the Mean Value Theorem say?

3. *Rolle's Theorem* says that there is a stationary point between any pair of zeros of a function. More precisely,

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. (Hint: It's a special case of the Mean Value Theorem. You can mimic the proof that was given in the text.)

4. What is a stationary point?

5. **[Group Problem]** How can you tell if a local maximum is a global maximum?

6. **[Group Problem]** If $f''(a) = 0$ then the graph of f has an inflection point at $x = a$. ***True or False?***

7. What is an inflection point?

8. Give an example of a function for which $f'(0) = 0$ even though the function f has neither a local maximum or a local minimum at $x = 0$.

9. **[Group Problem]** Draw four graphs of functions, one for each of the following four combinations:

$$f' > 0 \text{ and } f'' > 0$$
$$f' > 0 \text{ and } f'' < 0$$
$$f' < 0 \text{ and } f'' > 0$$
$$f' < 0 \text{ and } f'' < 0$$

10. **[Group Problem]** Which of the following combinations are possible?

(a) $f'(x) > 0$ and $f''(x) = 0$ for all x

(b) $f'(x) = 0$ and $f''(x) > 0$ for all x

For each of the following functions, decide if they are increasing, decreasing, or neither on the indicated intervals:

11. $f(x) = \dfrac{x}{1+x^2}$, on $10 < x < \infty$

12. $f(x) = \dfrac{2+x^2}{x^3-x}$, on $1 < x < \infty$

13. $f(x) = \dfrac{2+x^2}{x^3-x}$, on $0 < x < 1$

14. $f(x) = \dfrac{2+x^2}{x^3-x}$, on $0 < x < \infty$

Sketch the graphs of the following functions. You should

- find where f, f' and f'' are positive or negative
- find all stationary points and classify them as local minima, maxima, or neither
- find any global maxima or minima, if they exist
- find all inflection points
- determine all intervals where the function is increasing or decreasing
- determine all intervals where the function is convex or concave
- find any horizontal, vertical, or slant asymptotes

15. $y = x^3 + 2x^2$

16. $y = x^3 - 4x^2$

17. $y = x^4 + 27x$

18. $y = x^4 - 27x$

19. $y = x^4 + 2x^2 - 3$

20. $y = x^4 - 5x^2 + 4$

21. $y = x^5 + 16x$

22. $y = x^5 - 16x$

23. $y = \dfrac{x}{x+1}$

24. $y = \dfrac{x}{1+x^2}$

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

26. $y = \dfrac{1+x^2}{1+x}$

27. $y = x + \dfrac{1}{x}$

28. $y = x - \dfrac{1}{x}$

29. $y = x^3 + 2x^2 + x$

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

31. $y = x^4 - x^3 - x$

32. $y = x^4 - 2x^3 + 2x$

33. $y = \sqrt{1+x^2}$

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

35. $y = \sqrt[4]{1+x^2}$

36. $y = \dfrac{1}{1+x^4}$

The following functions are **periodic**, meaning that they satisfy $f(x+L) = f(x)$ for all x and a constant L 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 inflection points. Sketch the graphs of the following functions as in the previous problem, but only list those “interesting points” that lie in the interval $0 \le x < 2\pi$.

37. $y = \sin x$

38. $y = \sin x + \cos x$

39. $y = \sin x + \sin^2 x$

40. $y = 2\sin x + \sin^2 x$

41. $y = 4\sin x + \sin^2 x$

42. $y = 2\cos x + \cos^2 x$

43. $y = \dfrac{4}{2+\sin x}$

44. $y = \left(2+\sin x\right)^2$

Graphing inverse trigonometric functions. Find the domains and sketch the graphs of each of the following functions

45. $y = \arcsin x$

46. $y = \arctan x$

47. $y = 2\arctan x - x$

48. $y = \arctan(x^2)$

49. $y = 3\arcsin(x) - 5x$

50. $y = 6\arcsin(x) - 10x^2$

51. Sometimes we will have functions for which we cannot solve $f'(x) = 0$, but will still allow us to say something using the 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.

52. Let $\frac{p}{q}$ be some fraction, with $p > 0$, $q > 0$. Assume also that $p < q$ so that the fraction is less than 1.

If you increase both numerator p and denominator q by the same amount $x > 0$, does the fraction increase or decrease? In other words, which of the following two inequalities holds:

$$\frac{p+x}{q+x} > \frac{p}{q} \text{ or } \frac{p+x}{q+x} < \frac{p}{q}?$$

Explain how you can answer this question by checking that a certain function is increasing or decreasing. Which function would you look at?

How does your answer change if $p > q$?

53. (a) Consider the list of numbers

$$\tfrac{1}{2}, \tfrac{2}{5}, \tfrac{3}{10}, \tfrac{4}{17}, \tfrac{5}{26}, \cdots.$$

The nth number in the list is

$$a_n = \frac{n}{n^2+1}.$$

Is it true that each number in the list is less than its predecessor? I.e., is it true that

$$a_{n+1} < a_n$$

for all $n = 1, 2, 3, 4, \ldots$?

(b) Consider the following list of numbers

$$\tfrac{1}{2}, \tfrac{2}{3}, \tfrac{3}{6}, \tfrac{4}{11}, \tfrac{5}{18}, \cdots,$$

where the nth number in the list is

$$a_n = \frac{n}{n^2-2n+3}.$$

For which n is $a_{n+1} < a_n$?

54. Show that the equation $3y + \sin y = \frac{\pi}{2}$ has exactly one solution y. (In chapter I, §4.4 we had agreed to check this.)

13. Applied Optimization

Many problems, after some amount of work, can be rephrased as

For which value of x in the interval $a \le x \le b$ is $f(x)$ the largest?

In other words, given a function f on an interval $[a, b]$, we want to determine the maximum value of f on this interval (and usually, for what values of x the function attains its maximum value).

If the function is continuous, then according to theorem 8.1 there always is at least one x in the interval $[a, b]$ which maximizes $f(x)$.

If f is differentiable then we know that any local maximum is either a stationary point or one of the end points a and b. Therefore, we can find the global maxima by following this recipe:

(1) *Find all stationary points of f;*
(2) *Compute $f(x)$ at each stationary point from step (1);*
(3) *Compute $f(a)$ and $f(b)$;*
(4) *Compare all of the computed values from steps (2) and (3), and pick the biggest one.*

Note that there is a single global maximum value on any interval $[a, b]$, although it may occur at more than one x-value in that interval.

The same procedure holds for looking for global minima, in order to *minimize* rather than *maximize* a function. The same procedure works, except of course in step (4) we want to find the smallest value rather than the biggest one.

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.

13.1. 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 we 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 $x \geq 0$, $y \geq 0$.

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

$$A = xy = x\big(\tfrac{1}{2} - x\big)$$

The values of x which we are allowed to consider are only limited by the requirements $x \geq 0$ and $y \geq 0$, i.e. $x \leq \frac{1}{2}$. So we end up with this problem:

Find the maximum of the function $f(x) = x\big(\frac{1}{2} - x\big)$ *on the interval* $0 \leq x \leq \frac{1}{2}$.

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 $0 \leq x \leq \frac{1}{2}$ 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

$$x = \tfrac{1}{4}, \text{ and } y = \tfrac{1}{2} - \tfrac{1}{4} = \tfrac{1}{4},$$

which is a square with sides of length $\frac{1}{4}$.

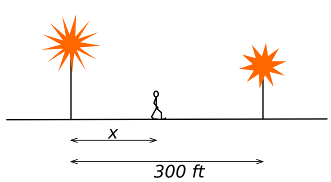
14. Problems

1. Which rectangle of area 100in^2 minimizes its height plus two times its length?

2. 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.]

3. [Group Problem] (The lamp post problem) In a street two lamp posts are 300 feet apart. The light intensity at a distance d from the first lamp post is $1000/d^2$, the light intensity at distance d from the second (weaker) lamp post is $125/d^2$ (in both cases the light intensity is inversely proportional to the square of the distance to the light source).



The *combined light intensity* is the sum of the two light intensities coming from both lamp posts.

(a) If you are in between the lamp posts, at distance x feet from the stronger light, then give a formula for the combined light intensity coming from both lamp posts as a function of x.

(b) What is the darkest spot between the two lights, i.e. where is the combined light intensity the smallest?

4. (a) You have a sheet of metal with area 100 in^2 from which you are to make a cylindrical soup can (with both a top and a bottom). 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?

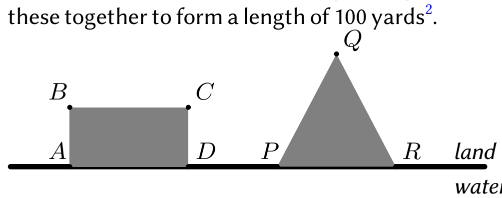
(b) If instead of making a plain cylinder you replaced the flat top and bottom of the cylinder with two hemispherical caps, then (using the same 100in^2 of sheet metal), then which choice of radius and height of the cylinder give you the container with the largest volume?

(c) Suppose you only replace the top of the cylinder with a hemispherical cap, and leave the bottom flat. Then which choice of height and radius of the cylinder result in the largest volume?

5. 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 $0 \le a < \infty$?

6. [Group Problem] *(Queen Dido's problem)*

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[2].



(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 perimeter: $PQ + QR$ is 100 yards long, and $PQ = QR$.)

[2] I made that number up. For the rest start at http://en.wikipedia.org/wiki/Dido

7. 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?

8. 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?

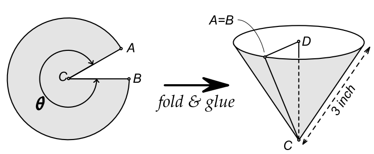
9. What are the smallest and largest values that $(\cos x)(\cos y)$ can have if $x+y=\frac{\pi}{2}$ and if x and y are both nonnegative?

10. **(a)** Find the smallest and largest values of $\tan x + \tan y$ 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?

11. If a locomotive is traveling at v miles per hour, the cost per hour of fuel $v^2/25$ dollars, and other costs are $100 per hour regardless of speed. What is the speed that minimizes cost **per mile**?

12. **[Group Problem]** 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.



(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?

13. Here are two equivalent problems (read both & choose one):

(a) Blood pressure is the pressure exerted by circulating blood on walls of the arteries. Assume the blood pressure varies periodically according to the formula

$$p(t) = 50 + 15\sin(2.5\pi t)$$

where t is the number of seconds since the beginning of a cardiac cycle. When is the blood pressure the highest for $0 \leq t \leq 1$? What is the maximum blood pressure? When is the blood pressure lowest in the interval $0 \leq t \leq 1$ and what is the corresponding blood pressure?

(b) The water level in a well is described as a function of time by

$$p(t) = 50 + 15\sin(2.5\pi t).$$

What is highest (resp. smallest) water level in the time interval $0 \leq t \leq 1$?

14. A manufacturer needs to make a cylindrical can (with top and bottom) that will hold 1 liter of liquid. Determine the dimensions of the can that will minimize the amount of material used in its construction.

15. A window is being built in the shape of semicircle on the top of a rectangle, where there is a frame between the semicircle and rectangle as well as around the boundary of the window. If there is a total of 15 meters of framing materials available, what must the dimensions of the window be to let in the most light?

16. A 5 feet piece of wire is cut into two pieces. One piece is bent into a square and the other is bent into an equilateral triangle. Where should the wire be cut so that the total area enclosed by both figures is minimal?

17. Two poles, one 5 meters tall and one 10 meters tall, are 20 meters apart. A telephone wire is attached to the top of each pole and it is also staked to the ground somewhere between the two poles.

(a) Where should the wire be staked so that the minimum amount of wire is used?

(b) Where should the wire be staked so that the angle formed by the two pieces of wire at the stake is a maximum?

18. Determine the cylinder with the largest volume that can be inscribed in a cone of height 10 cm and base radius 6 cm.

19. A straight piece of wire 10 feet long is bent into the shape of a right-angled L-shape. What is the shortest possible distance between the ends?

20. An open-top cylindrical tank with a volume of 10 cubic feet is to be made from a sheet of steel. Find the dimensions of the tank that will require as little material used in the tank as possible.

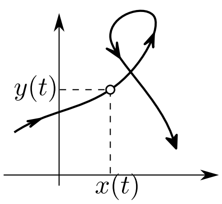
21. A box with a square base and open top must have a volume of $32,000$ cm^3. Find the dimensions of the box that minimize the amount of material used for building it.

22. Find the length of the shortest ladder that will reach over an 8-ft. high fence to the wall of a tall building which is 3 ft. behind the fence.

23. A rectangular piece of paper is 10 inches high and 6 inches wide. The lower right-hand corner is folded over so as to reach the leftmost edge of the paper. Find the minimum length of the resulting crease.

24. In the xy-plane, an observer stands at the origin $(0, 0)$. A car travels along the curve $y = 1/x^2$. How close does the car come to the observer?

15. Parametrized Curves



So far all the plane curves that we have seen were graphs of functions $y = f(x)$. A different way of describing a plane curve is to think of it as the path that is traced out by a particle moving in the plane. If you have a particle that is moving about in the plane, then you can describe its motion by specifying the coordinates (x, y) of the point as functions of time: the motion is given by two functions

$$x = x(t), \quad y = y(t).$$

As t varies ("as time goes by") the point $(x(t), y(t))$ traces out a curve. This kind of curve is called a ***parametric curve*** (or sometimes *parametrized curve*, or "a curve defined by parametric equations"). This topic reappears in third semester calculus (Math 234) where much of the material can be simplified using vectors.

15.1. Example: steady motion along a straight line. The simplest motion is where the particle moves with constant velocity. Suppose the particle starts out at (x_0, y_0) at time $t = 0$, and suppose its horizontal and vertical velocities are v_x and v_y, respectively. Then after a time t has gone by the point has moved a distance $v_x t$ in the x-direction and $v_y t$ in the y-direction. Its coordinates are therefore

$$x(t) = x_0 + v_x t, \quad y(t) = y_0 + v_y t. \tag{49}$$

One question we can ask is: how fast is the particle moving?

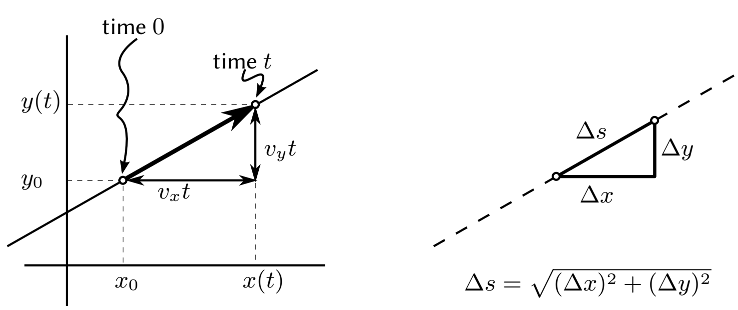


Figure 11. Motion with constant velocity along a straight line

We know that both its x-coordinate and y-coordinate increase at constant rates v_x and v_y. So, between time t and time $t + \Delta t$, the x and y coordinates increase by $\Delta x = v_x \Delta t$

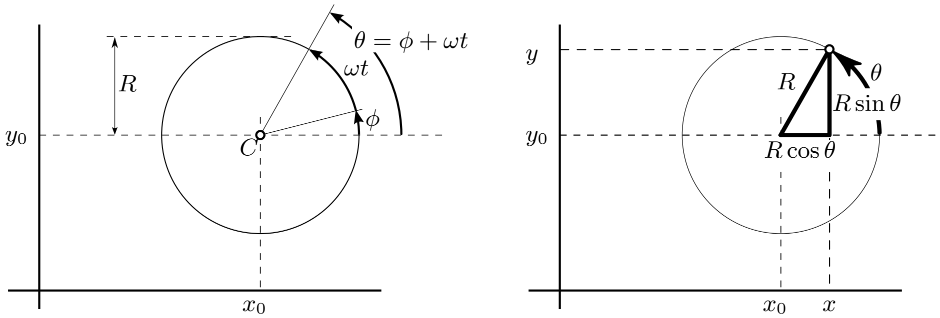


Figure 12. Constant velocity motion along a circle. The coordinates of any point on the circle with radius R and center (x_0, y_0) are completely determined by the angle θ (on the right). To describe the motion of a particle on the circle you only have to specify how the angle θ changes with time. For a motion with constant angular velocity (on the left) θ starts out at some value ϕ when $t = 0$; after time t the angle θ has increased by ωt, and thus has the value $\theta = \phi + \omega t$.

and $\Delta y = v_y \Delta t$, so therefore the distance traveled by the point is

$$\Delta s = \sqrt{(\Delta x)^2 + (\Delta y)^2} = t\sqrt{v_x^2 + v_y^2}. \tag{50}$$

Then we can get the average velocity v by dividing by Δt:

$$v = \frac{\Delta s}{\Delta t} = \sqrt{\left(\frac{\Delta x}{\Delta t}\right)^2 + \left(\frac{\Delta y}{\Delta t}\right)^2} = \sqrt{v_x^2 + v_y^2}. \tag{51}$$

Since the average velocity turns out not to depend on the length Δt of the time interval, we see that the instantaneous velocity at any time t will also be equal to the value v we have just found.

15.2. Example: steady motion along a circle. Suppose a particle P is moving along a circle "at a constant rate", or (to be more technical) "with a constant angular velocity". Let the center C of the circle be at the point (x_0, y_0) and let its radius be R. The location of P is then completely determined by the angle θ that the radius CP makes with the horizontal line through the center; thus, we see that the coordinates of P are

$$x = x_0 + R\cos\theta, \qquad y = y_0 + R\sin\theta.$$

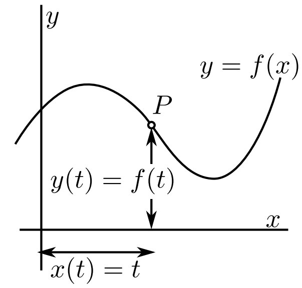
Since the particle is moving on a circle, saying that it is moving at a constant rate is equivalent to saying that the angle θ is changing in time at a constant rate. This rate, usually called ω, is the ***angular velocity*** of the point. If the angle θ starts out at time $t = 0$ with the value $\theta = \phi$, then after time t the angle θ has increased ωt, so that $\theta = \phi + \omega t$. The motion of our particle P is therefore given by

$$x(t) = x_0 + R\cos(\omega t + \phi), \quad y(t) = y_0 + R\sin(\omega t + \phi). \tag{52}$$

The initial angle ϕ is sometimes called the ***phase***.

15.3. Example: motion on a graph. If you are given a function $y = f(x)$ then we can describe its graph as a parametric curve by setting

$$x(t) = t \text{ and } y(t) = f(t).$$



For this parametric curve we always have $y(t) = f(x(t))$ so the point with coordinates $(x(t), y(t))$ always lies on the graph of the function f.

Since $x(t) = t$ the point moves from left to right in such a way that its horizontal velocity is $v_x = 1$.

15.4. General motion in the plane. In the most general motion of a point in the plane, the coordinates of the point P are given by two functions of time

$$P: \quad x = x(t), \quad y = y(t). \tag{53}$$

In this course we mostly think of t as "time" but the quantity t doesn't have to have that interpretation. The important property of t is that it is a ***parameter*** that allows us to label the points on the curve defined by (53).

To find the ***instantaneous velocity*** of the point at some given time t you follow the same strategy as in Chapter 2. Pick a (small) number Δt and compute the average velocity of the point between times t and $t + \Delta t$. Then take the limit for $\Delta t \to 0$ to get the instantaneous velocity.

Between time t and time $t + \Delta t$ the x and y coordinates of the point changed by Δx and Δy, respectively, and therefore the average velocities in the x and y directions are $\Delta x/\Delta t$ and $\Delta y/\Delta t$. The instantaneous horizontal and vertical velocities are

$$v_x(t) = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt}, \qquad v_y(t) = \lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t} = \frac{dy}{dt}.$$

The distance traveled by the point from time t to time $t + \Delta t$ is again given by (50), i.e. by $\Delta s = \sqrt{(\Delta x)^2 + (\Delta y)^2}$. Dividing by Δt, we find that the average velocity between t and $t + \Delta t$ is

$$\frac{\Delta s}{\Delta t} = \sqrt{\left(\frac{\Delta x}{\Delta t}\right)^2 + \left(\frac{\Delta y}{\Delta t}\right)^2}.$$

Let $\Delta t \to 0$ to get the instantaneous velocity at time t:

$$v(t) = \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} = \sqrt{v_x(t)^2 + v_y(t)^2}\,. \tag{54}$$

15.5. Slope of the tangent line. In most examples the curve that gets traced out by $x = x(t)$, $y = y(t)$ has a tangent line at most points. To find the slope of that tangent line we follow the idea in Chapter II, §1: first compute the slope of the line connecting two nearby points on the curve and then consider the limit as the one of the two points approaches the other. More precisely, given a time t choose a small time increment $\Delta t \neq 0$ and consider the points at times t and $t + \Delta t$. The slope of the line connecting these points is

$$\text{Slope of secant} = \frac{\Delta y}{\Delta x} \qquad \text{(see Figure 13).}$$

To relate this to derivatives of $x(t)$ and $y(t)$ with respect to time we rewrite this as

$$\text{Slope of secant} = \frac{\Delta y}{\Delta x} = \frac{\dfrac{\Delta y}{\Delta t}}{\dfrac{\Delta x}{\Delta t}}$$

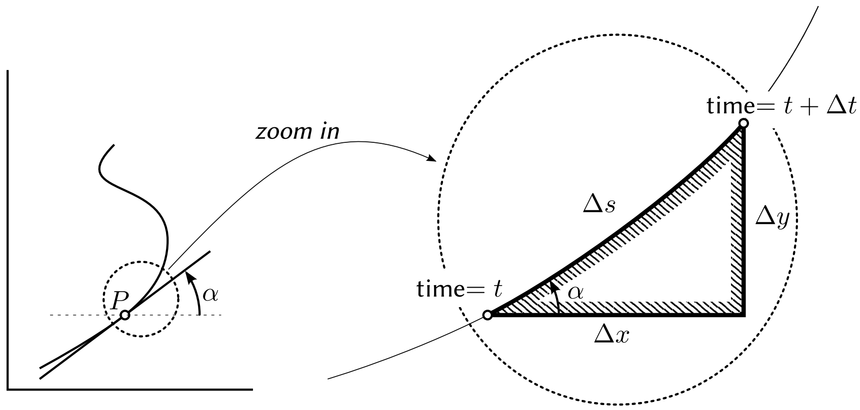


Figure 13. Motion along a curve: the figure on the right, obtained by magnifying a small piece of the curve on the left, is almost, but not quite, a triangle. It fails to be a triangle because the hypotenuse is not a straight line. The smaller you make Δt, or, the closer you "zoom in" on the curve, the more the figure on the right will seem to be a right triangle. This picture allows you to compute the velocity of the moving point P, and the slope $\tan\alpha$ of the tangent to the curve traced out by the point P during this motion

Now we let $\Delta t \to 0$ and we find

$$\text{Slope of tangent} = \lim_{\Delta t\to 0} \frac{\dfrac{\Delta y}{\Delta t}}{\dfrac{\Delta x}{\Delta t}} = \frac{\dfrac{dy}{dt}}{\dfrac{dx}{dt}} = \frac{v_y(t)}{v_x(t)}. \tag{55}$$

In Leibniz' notation this formula becomes very transparent: the slope of the tangent to the curve is dy/dx, i.e. the ratio of dy and dx where dy is the infinitely small increase of y associated with an infinitely small increase dx of x. If you pretend that infinitely small quantities exist, then you can divide numerator and denominator in the fraction dy/dx by dt, which leads to

$$\text{slope } = \frac{dy}{dx} = \frac{\dfrac{dy}{dt}}{\dfrac{dx}{dt}}.$$

This same argument can also be given without referring to infinitely small numbers, and by invoking the Chain Rule instead: if we think of y as a function of x which is itself a function of t, then the Chain Rule says

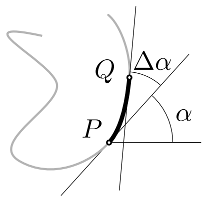
$$\frac{dy}{dt} = \frac{dy}{dx}\cdot\frac{dx}{dt}$$

and now by rearranging we obtain the formula

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{y'(t)}{x'(t)}.$$

15.6. The curvature of a curve. As the name says, curves are usually curved, and while they will touch their tangent lines, a curve normally bends away from any tangent to the curve. Some curves appear more curved than others, while any particular curve can be highly curved in some places and nearly straight in others.

To make the phrase "more curved" more precise, you can compute the ***curvature*** of the curve at any given point. The curvature measures how fast the tangent to the curve is turning as you move along the curve. To define the curvature you first consider the angle α that the tangent at any point on the curve makes with the x-axis. Given a point P on the curve choose a nearby point Q on the curve, and let $\Delta\alpha$ be the amount by which the tangent angle changes as you go from P to Q. Then you could define the "average curvature" of the segment PQ to be the ratio



$$\text{Average curvature of } PQ = \frac{\Delta\alpha}{\Delta s},$$

where Δs is the length of the arc PQ. The ***curvature*** of the curve at the point P is

$$\kappa = \lim_{Q\to P} \frac{\Delta\alpha}{\Delta s}. \tag{56}$$

15.7. A formula for the curvature of a parametrized curve. How do you compute the limit (56) that defines the curvature κ if you know the functions $x(t), y(t)$ that specify the parametric curve?

To compute κ assume that the fixed point P corresponds to a certain value of t, so P has coordinates $(x(t), y(t))$. Let Q be the point you get by changing t to $t + \Delta t$, so Q has coordinates $(x(t + \Delta t), y(t + \Delta t))$. Then we get

$$\kappa = \lim_{Q\to P} \frac{\Delta\alpha}{\Delta s} = \lim_{\Delta t\to 0} \frac{\dfrac{\Delta\alpha}{\Delta t}}{\dfrac{\Delta s}{\Delta t}}.$$

The denominator is just the velocity

$$\lim_{\Delta t\to 0} \frac{\Delta s}{\Delta t} = v = \sqrt{(x'(t))^2 + (y'(t))^2}.$$

The numerator is

$$\lim_{\Delta t\to 0} \frac{\Delta\alpha}{\Delta t} = \frac{d\alpha}{dt},$$

i.e., it is the derivative of the tangent angle with respect to t (time). You can compute it by using

$$\tan\alpha = \frac{y'(t)}{x'(t)} \implies \alpha = \arctan\frac{y'(t)}{x'(t)}.$$

Using the chain rule and the quotient rule you can differentiate this (a good exercise!), with result

$$\frac{d\alpha}{dt} = \frac{x'(t)y''(t) - y'(t)x''(t)}{x'(t)^2 + y'(t)^2}. \tag{57}$$

To get the curvature we still have to divide by v. The final result is

$$\kappa = \frac{x'(t)y''(t) - y'(t)x''(t)}{\left\{x'(t)^2 + y'(t)^2\right\}^{3/2}}. \tag{58}$$

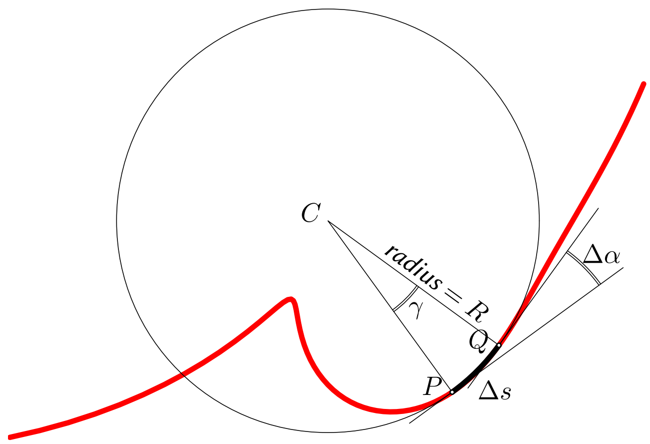
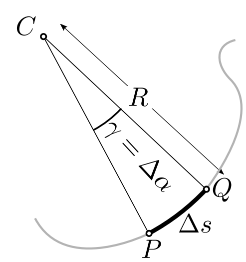


Figure 14. The Osculating Circle and the Radius of Curvature. If you assume that a small piece PQ of a curve can be approximated by a circle, then this figure shows you how to compute the radius of that circle. If the arc PQ is a part of a circle with radius R, then the length Δs of the arc is $R\Delta\alpha$, where γ is the angle between the two normals to the arc at P and Q. The angle γ is the same as the angle $\Delta\alpha$ between the two tangent lines to the curve at P and Q. The formula $\Delta s \approx R\Delta\alpha$ leads us to the definition $R = \lim_{Q\to P} \Delta s/\Delta\alpha$.

15.8. The radius of curvature and the osculating circle. Just as the tangent to a curve is the straight line that best approximates the curve at some given point, you can try to find a circle that "best approximates" the curve at some point. This best match among all circles is called the ***osculating circle*** to the curve at the particular point. Its radius is the ***radius of curvature*** of the curve.

For a more precise definition of the osculating circle, look at Figure 14, which shows a curve with a point P. To find the circle through P "that best matches the curve" we pick another point Q on the curve near P and pretend that the arc PQ is part of a circle. The center C of the circle is then found by intersecting two normals to the curve, one through P and one through Q. From basic geometry you know that the length of a circular arc is the "radius times the angle," so the radius R of the circle, the angle γ between the normals, and the length Δs of the arc PQ are related by



$$\Delta s = R \times \gamma.$$

Since you get the normals by rotating the tangents at P and Q counterclockwise by 90 degrees, the angle γ between the normals is the same as the angle $\Delta\alpha$ between the tangents. This leads to

$$R = \frac{\Delta s}{\Delta\alpha}.$$

All this is based on the assumption that PQ is a circular arc, but the curve is not really a circle, and we should only expect the segment PQ to be *approximately* a circle arc if the segment is "short enough." Therefore we *define* the radius of curvature to be the limit you get as $Q \to P$:

$$R = \lim_{Q\to P} \frac{\Delta s}{\Delta\alpha} = \frac{1}{\lim_{Q\to P} \dfrac{\Delta\alpha}{\Delta s}} = \frac{1}{\kappa}. \tag{59}$$

The ***osculating circle*** at a point P on a parametrized curve is the circle with radius R that is tangent to the curve at the point P. The center C of the osculating circle lies on the normal to the curve at P.

16. Problems

1. Describe each of the following motions in the plane. More specifically,

–find all points with a horizontal tangent,
–find all points with a vertical tangent,
–find all inflection points (i.e. points where the curvature κ changes sign).

(a) $x(t) = 1 - t$, $y(t) = 2 - t$

(b) $x(t) = 3t + 2$, $y(t) = 3t + 2$

(c) $x(t) = t$, $y(t) = t^2$

(d) $x(t) = \sin t$, $y(t) = t$

(e) $x(t) = \sin t$, $y(t) = \cos 2t$

(f) $x(t) = \sin 25t$, $y(t) = \cos 25t$

(g) $x(t) = 1 + \cos t$, $y(t) = 1 + \sin t$

(h) $x(t) = 2\cos t$, $y(t) = \sin t$

(i) $x(t) = t^2$, $y(t) = t^3$

2. Complete the steps that led to (57).

3. Find the radius of curvature and the center of the osculating circle to the curve $x(t) = t, y(t) = 1/t$ at the point with $t = 1$.

4. Find the curvature of the graph of a function $y = f(x)$. You can do this be thinking of the graph of $y = f(x)$ as a parametric curve as in § 15.3, and then using (58).

The result you should get is

$$\kappa = \frac{f''(x)}{\big(1 + f'(x)^2\big)^{3/2}}.$$

5. Use the result from the previous problem to find points with the largest or smallest curvature on the following familiar graphs. In each case make a drawing of the graph first and guess your answer before you compute it.

(a) Where is the curvature of the graph of $y = x^2$ largest?

(b) Where is the curvature of the graph of $y = x^3$ largest?

(c) Where is the curvature of the graph of $y = \frac{1}{3}x^3$ largest?

6. Compute the curvature of the parametric curve given by $x(t) = \cos^2 t$, $y(t) = \sin^2 t$.

The answer turns out to be very simple and could have been predicted without taking any derivatives – do you see why?

17. l'Hopital's rule

There is a simple way to compute limits of the form

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

that are of the form $\frac{0}{0}$, i.e. where

$$\lim_{t\to a} f(t) = 0 \text{ and } \lim_{t\to a} g(t) = 0.$$

It is given by what is called l'Hopital's rule, which says the following:

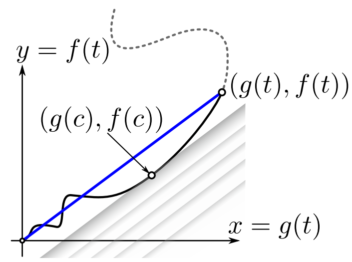
17.1. Theorem. *If f and g both are differentiable functions such that*

$$\lim_{t\to a} f(t) = 0 \textit{ and } \lim_{t\to a} g(t) = 0,$$

and if $\displaystyle\lim_{t\to a} \frac{f'(t)}{g'(t)}$ exists, then $\lim_{t\to a} \frac{f(t)}{g(t)}$ also exists. Moreover,

$$\lim_{t\to a} \frac{f(t)}{g(t)} = \lim_{t\to a} \frac{f'(t)}{g'(t)}$$

17.2. The reason why l'Hopital's rule works. The two functions f and g define a parametric curve $x = g(t)$, $y = f(t)$ that is defined for $t \geq a$. At $t = 0$ we have $x = g(a) = 0$ and $y = g(a) = 0$, so the origin lies on the curve traced out by $x = g(t)$, $y = f(t)$.



The slope of the chord connecting the origin and the point $(g(t), f(t))$ *is*

$$\frac{f(t)}{g(t)}.$$

Somewhere between this point and the origin there is a point $(g(c), f(c))$ *on the curve where the tangent has the same slope. The slope of the tangent is*

$$\frac{f'(c)}{g'(c)}$$

If we pick any value of $t > a$, then the ratio $\frac{f(t)}{g(t)}$ is the slope of the chord connecting the two points $(0, 0)$ and $(g(t), f(t))$ on the curve. By the same kind of reasoning that led to the Mean Value Theorem (see Figure 3) there has to be a point in between where the tangent to the curve is parallel to the chord: in other words, there is some c with $a < c < t$ such that

$$\frac{f(t)}{g(t)} = \frac{f'(c)}{g'(c)}.$$

The number c depends on t, but since it lies between a and t, we have $\lim_{t\to a} c = a$. Therefore

$$\lim_{t\to a} \frac{f(t)}{g(t)} = \lim_{c\to a} \frac{f'(c(t))}{g'(c(t))} = \frac{f'(a)}{g'(a)}.$$

17.3. Examples of how to use l'Hopital's rule. The limit

$$\lim_{x\to 1} \frac{3x^2 - x - 2}{x^2 - 1}$$

is of the form $\frac{0}{0}$, so we can try to apply l'Hopital's rule. We get

$$\lim_{x\to 1} \frac{3x^2 - x - 2}{x^2 - 1} = \lim_{x\to 1} \frac{6x - 1}{2x} \tag{60}$$

So far we don't know if the limit exists, so it could be that we are just saying that two things that don't exist are equal (whatever that means). But the second limit does exits:

$$\lim_{x\to 1} \frac{6x - 1}{2x} = \frac{6 - 1}{2} = \frac{5}{2}.$$

Since this exists, l'Hopital's rule tells us that the first limit in (60) also exists, and is equal to $\frac{5}{2}$.

For another example, consider the limit

$$\lim_{x\to 0} \frac{\sin x}{x}$$

which we have already computed in Chapter III, §15. It is again a limit of the form $\frac{0}{0}$, so we can try to use l'Hopital's rule:

$$\lim_{x\to 0} \frac{\sin x}{x} = \lim_{x\to 0} \frac{\cos x}{1} \overset{(*)}{=} \frac{1}{1} = 1.$$

The reasoning is that since the limit of $\dfrac{\cos x}{1}$ exists, the limit of $\dfrac{\sin x}{x}$ must also exist and be the same.

Note that evaluating the limit $\lim_{x\to 0}\frac{\sin x}{x} = 1$ in this manner is actually circular reasoning, because we used this same limit in showing that the derivative of $\sin(x)$ is $\cos(x)$.

In the same way you can do a slightly more complicated example

$$\lim_{x\to 0}\frac{\sin 5x}{\tan x} = \text{“}\tfrac{0}{0}\text{”} \stackrel{\text{l'H}}{=} \lim_{x\to 0}\frac{5\cos 5x}{1/\cos^2 x} \stackrel{(*)}{=} \frac{5\cdot 1}{1/1^2} = 5.$$

Again, the equality “$\stackrel{\text{l'H}}{=}$” follows from l'Hopital's rule and the next equality “$\stackrel{(*)}{=}$” justifies using the rule.

17.4. Examples of repeated use of l'Hopital's rule.

The limit

$$\lim_{x\to 0}\frac{x-\sin x}{x^3}$$

is again of the form $\frac{0}{0}$. When you apply l'Hopital's rule you get

$$\lim_{x\to 0}\frac{x-\sin x}{x^3} = \lim_{x\to 0}\frac{1-\cos x}{3x^2},$$

which is also of the form $\frac{0}{0}$! That means we can try to compute this new limit by again applying l'Hopital's rule. Here is the whole computation:

$$\lim_{x\to 0}\frac{x-\sin x}{x^3} = \lim_{x\to 0}\frac{1-\cos x}{3x^2} = \lim_{x\to 0}\frac{\sin x}{6x} = \lim_{x\to 0}\frac{\cos x}{6} = \tfrac{1}{6}.$$

Here we have applied l'Hopital's rule three times. When we do this we don't know if any of the limits that we run into actually exist, until at the very end we find that the last limit exists (it's $\frac{1}{6}$ in this example). The fact that the last limit exists implies that the one before that exists; the fact the second to last limit exists in turn implies that the limit before that one exists, and so on, until we conclude that the limit we started with exists. All the limits must be equal, so we find that

$$\lim_{x\to 0}\frac{x-\sin x}{x^3} = \frac{1}{6},$$

which means that for small values of x we have the approximation

$$x-\sin x \approx \tfrac{1}{6}x^3.$$

18. Problems

1. Compute the following limits using l'Hopital's rule.

(a) $\displaystyle\lim_{x\to 0}\frac{x^2-3}{x^2-8x-9}$.

(b) $\displaystyle\lim_{x\to \pi/2}\frac{\sin 2x}{\cos x}$.

(c) $\displaystyle\lim_{x\to 1/2}\frac{\cos \pi x}{1-2x}$.

2. Suppose n is some positive integer, and the limit

$$\lim_{x\to 0}\frac{\cos x - 1 - x^2/2}{x^n} = L$$

exists. Also suppose $L \neq 0$. What is n? What is the limit L?

3. What happens when you use l'Hopital's rule to compute these limits?

(a) $\displaystyle\lim_{x\to 0}\frac{x^2}{x}$.

(b) $\displaystyle\lim_{x\to 0}\frac{x^2}{x^3}$.

4. Let $f(x) = \frac{1}{2}x + x^2 \sin\frac{\pi}{x}$ be the strange function whose graph is drawn in Figure 5.

(a) Try to compute
$$\lim_{x\to 0} \frac{f(x)}{x}$$
in two ways:

- directly, and
- by using l'Hopital's rule.

(b) The following rule sounds very much like l'Hopital's:

if $\lim_{x\to a} \dfrac{f(x)}{g(x)}$ *exists, then* $\lim_{x\to a} \dfrac{f'(x)}{g'(x)}$ *also exists, and the two limits are equal.*

But this is not always true! Find a counterexample.

5. Here is a method for computing derivatives: since, by definition,
$$f'(a) = \lim_{x\to a} \frac{f(x) - f(a)}{x - a}$$
is a limit of the form $\frac{0}{0}$, we can always try to find it by using l'Hopital's rule. What happens when you do that?

6. Simplicio did the following computation
$$\begin{aligned}\lim_{x\to 0} \frac{x}{\sin x} &= \lim_{x\to 0} \frac{1}{\cos x}\\ &= \lim_{x\to 0} \frac{0}{\sin x}\\ &= \lim_{x\to 0} 0\\ &= 0 \ldots ?\end{aligned}$$

Do you see any problems with this?

CHAPTER 6

Exponentials and Logarithms (naturally)

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 rational number: for example, we know that $2^{3/4}$ is the 4th root of the third power of 2 ($\sqrt[4]{2^3}$), but what should we take as the definition of something like $2^{\sqrt{2}}$?

Next 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 discuss the notion of "exponential growth".

1. Exponents

Here we discuss 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

$$x^{p/q} = \sqrt[q]{x^p}. \tag{61}$$

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 obtained by truncating the decimal expansion of $\sqrt{2}$, namely

$$a_1 = 1, \qquad a_2 = 1.4 = \tfrac{14}{10}, \qquad a_3 = 1.41 = \tfrac{141}{100}, \qquad a_4 = 1.414 = \tfrac{1414}{1000}, \qquad \cdots.$$

Each a_n is a rational number, so we know what 2^{a_n} is; for example, $2^{a_4} = \sqrt[1000]{2^{1414}}$. Our definition of $2^{\sqrt{2}}$ then is

$$2^{\sqrt{2}} = \lim_{n\to\infty} 2^{a_n},$$

which is to say, we define $2^{\sqrt{2}}$ as the "limit of the sequence"

$$2,\ \sqrt[10]{2^{14}},\ \sqrt[100]{2^{141}},\ \sqrt[1000]{2^{1414}}, \ldots$$

(See table 1.)

$x = \dfrac{p}{q}$	$2^x = \sqrt[q]{2^p}$
$1.0000000000 = \frac{1}{1}$	**2**.000000000000
$1.4000000000 = \frac{14}{10}$	**2.6**39015821546
$1.4100000000 = \frac{141}{100}$	**2.6**57371628193
1.4140000000	**2.66**4749650184
1.4142000000	**2.6651**19088532
1.4142100000	**2.6651**37561794
1.4142130000	**2.66514**3103798
1.4142135000	**2.665144**027466
$\vdots$	$\vdots$

Table 1. Approximating $2^{\sqrt{2}}$ by computing 2^x for rational numbers x. 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 $2^{\sqrt{2}}$.

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.

In precalculus you learn that the exponential functions satisfy the following properties:

$$x^a x^b = x^{a+b}, \qquad \frac{x^a}{x^b} = x^{a-b}, \qquad \left(x^a\right)^b = x^{ab} \tag{62}$$

provided a and b are rational numbers. In fact these properties still hold if a and b are arbitrary real numbers. 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.).

1.1. 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 (61) 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 $\frac{2}{6} = \frac{1}{3}$ we might think that $(-8)^{2/6} = (-8)^{1/3}$. However our definition (61) 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: either

(1) *avoid taking fractional powers of negative numbers, or*
(2) *whenever we compute $x^{p/q}$, we first simplify the fraction by removing common divisors of p and q.*

The safest option 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[1].

2. Logarithms

Briefly, if a is a positive constant not equal to 1, then we define $y = \log_a x$ to be 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 of*** x ***to the base*** a. Note that since exponentials are always positive, the logarithm of 0 or of a negative number is not defined[2]

For instance,

$$2^3 = 8, \quad 2^{1/2} = \sqrt{2}, \quad 2^{-1} = \frac{1}{2}$$

so

$$\log_2 8 = 3, \quad \log_2\bigl(\sqrt{2}\bigr) = \frac{1}{2}, \quad \log_2 \frac{1}{2} = -1.$$

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.

3. 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, you learn the following formulas in precalculus:

$$\begin{aligned} \log_a xy &= \log_a x + \log_a y & \qquad \log_a x^y &= y \log_a x \\ \log_a \frac{x}{y} &= \log_a x - \log_a y & \qquad \log_a x &= \frac{\log_b x}{\log_b a} \end{aligned} \tag{63}$$

They follow from (62), and to review this subject it is good to write out the reason why $a^{p+q} = a^p a^q$ implies $\log_a xy = \log_a x + \log_a y$, and similarly for the other formulas in (63).

[1] There is a definition of $(-8)^\pi$ which uses complex numbers, but we won't see or need this in the calculus sequence; it will show up in math 319 or 320 where complex exponentials are treated, and you will also run into $(-8)^\pi$ if you take an electrical engineering course where "phasors" are discussed

[2] Again, there is a way to define logarithms of negative numbers using complex numbers. We will not pursue these ideas at the moment.

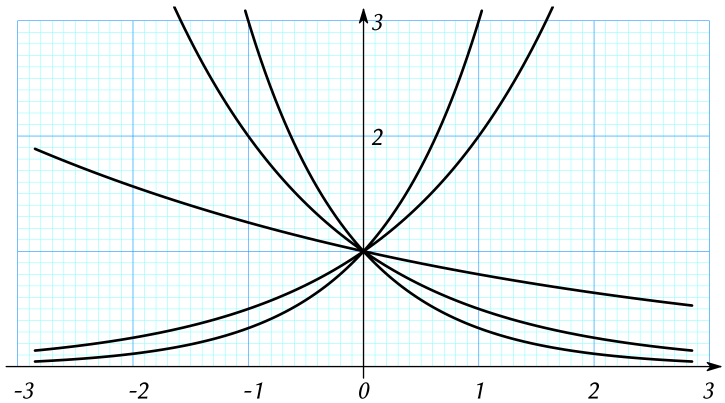


Figure 1. The graphs of $y = 2^x, 3^x, (1/2)^x, (1/3)^x$ and $y = (4/5)^x$. The graphs are purposely not labeled: can you figure out which is which?

4. Graphs of exponential functions and logarithms

Figure 1 shows the graphs of some exponential functions $y = a^x$ with different values of a, and Figure 2 shows the graphs of $y = \log_2 x$, $y = \log_3 x$, $\log_{1/2} x$, $\log_{1/3}(x)$ and $y = \log_{10} x$.

From precalculus we recall:

- If $a > 1$ then $f(x) = a^x$ is an increasing function of x.
- If $0 < a < 1$ then $f(x) = a^x$ is a decreasing function of x.

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 $a^{x_1} > a^{x_2}$.

These statements can be shown using the properties 62. However, we might also try to show these properties by computing the derivatives of these functions.

5. The derivative of a^x and the definition of e

To begin, we try to differentiate the function $y = 2^x$:

$$\begin{aligned}\frac{d}{dx}\left[2^x\right] &= \lim_{\Delta x\to 0}\frac{2^{x+\Delta x}-2^x}{\Delta x}\\ &= \lim_{\Delta x\to 0}\frac{2^x 2^{\Delta x}-2^x}{\Delta x}\\ &= \lim_{\Delta x\to 0} 2^x\frac{2^{\Delta x}-1}{\Delta x}\\ &= 2^x\lim_{\Delta x\to 0}\frac{2^{\Delta x}-1}{\Delta x}.\end{aligned}$$

So if we assume that the limit

$$\lim_{\Delta x\to 0}\frac{2^{\Delta x}-1}{\Delta x} = C$$

exists, then we have

$$\frac{d2^x}{dx} = C2^x. \tag{64}$$

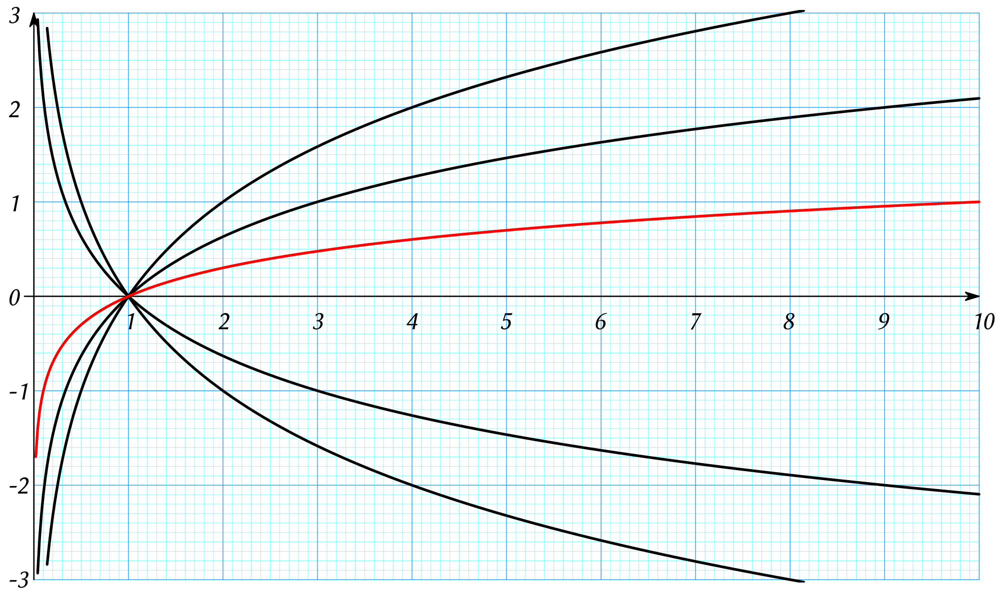


Figure 2. Graphs of some logarithms. Each curve is the graph of a function $y = \log_a x$ for various values of $a > 0$. Can you tell what a is for each graph?

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 does exist, 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 (64) 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

$$\begin{aligned} \frac{da^x}{dx} &= \frac{d}{dx}\left[2^{x\cdot\log_2 a}\right] \\ &= C\ 2^{x\cdot\log_2 a}\ \frac{d}{dx}\left[x \cdot \log_2 a\right] \\ &= C\ 2^{x\cdot\log_2 a} \cdot \log_2 a \\ &= (C \log_2 a)\ a^x. \end{aligned}$$

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

$$e = 2^{1/C} \text{ where } C = \lim_{\Delta x \to 0} \frac{2^{\Delta x} - 1}{\Delta x}.$$

One has

$$e \approx 2.718\ 281\ 828\ 459\ 045\ \cdots$$

The definition "$e = 2^{1/C}$" looks completely random and unmotivated, but here is the reason why $e = 2^{1/C}$ is a special number: 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

$$\frac{de^x}{dx} = e^x. \tag{65}$$

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

$$e^{\ln x} = x \qquad \ln e^x = x \tag{66}$$

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

$$\frac{da^x}{dx} = a^x \ln a. \tag{67}$$

6. Derivatives of Logarithms

Since the logarithm $f(x) = \log_a x$ is the inverse function of the exponential $g(x) = a^x$, we can find its derivative by implicit differentiation: we know that

$$a^{f(x)} = x.$$

Differentiating both sides, and using the Chain Rule on the left gives

$$(\ln a)a^{f(x)} f'(x) = 1.$$

Now solve for $f'(x)$ to get

$$f'(x) = \frac{1}{(\ln a)a^{f(x)}}.$$

Finally we remember that $a^{f(x)} = x$, and this 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: since $\ln e = 1$, we have

$$\frac{d}{dx}\left[\ln x\right] = \frac{1}{x}. \tag{68}$$

7. Limits involving exponentials and logarithms

7.1. Theorem. *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.

The case with general a follows after a substitution: given a we set

$$x = \frac{t}{\ln a}.$$

Then, since $a > 1$ we have $\ln a > 0$ so that as $x \to \infty$ one also has $t \to \infty$. Therefore

$$\lim_{x\to\infty} \frac{x^r}{a^x} = \lim_{x\to\infty} \frac{x^r}{e^{x\ln a}} = \lim_{t\to\infty} \frac{(t/\ln a)^r}{e^t} = \underbrace{(\ln a)^{-r}}_{\text{constant}} \cdot \underbrace{\lim_{t\to\infty} \frac{t^r}{e^t}}_{=0} = 0.$$

□

7.2. Related limits. If a, m, and r are constants, then

$$a > 1 \implies \lim_{x\to\infty} \frac{a^x}{x^r} = \infty \quad (D.N.E.) \tag{69a}$$

$$m > 0 \implies \lim_{x\to\infty} \frac{\ln x}{x^m} = 0 \tag{69b}$$

$$m > 0 \implies \lim_{x\to 0} x^m \ln x = 0 \tag{69c}$$

The second limit (69b) 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 limit (69c) follows from the second by substituting $x = 1/y$ and using $\ln \frac{1}{x} = -\ln x$.

8. Exponential growth and decay

8.1. Absolute and relative growth rates. If a quantity $X(t)$ changes with time, then the change in X during a short time interval of length Δt is

$$\Delta X = X(t + \Delta t) - X(t).$$

If X is measured in certain units, then the change ΔX has the same units as X. The change ΔX is sometimes called the ***absolute change***, to distinguish it from the ***relative change,*** which is defined as

$$\text{Relative change in } X = \frac{\Delta X}{X}.$$

No matter what units X is measured in, the relative change has no units and can be expressed as a percentage. For instance, if the population X of a city increases from $200,000$ to $250,000$ (people), then the absolute population change is $50,000$ (people), but the relative population change is

$$\frac{50,000 \text{ people}}{200,000 \text{ people}} = \frac{1}{4} = 25\%.$$

Just as we defined the rate of change of a quantity $X(t)$ to be

$$\frac{dX}{dt} = \lim_{\Delta t \to 0} \frac{\Delta X}{\Delta t},$$

we define the ***relative rate of change*** or ***relative growth rate*** to be

$$\lim_{\Delta t \to 0} \frac{\left\{ \begin{matrix} \text{Relative change in } X \\ \text{from } t \text{ to } t + \Delta t \end{matrix} \right\}}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Delta X}{X \Delta t} = \frac{1}{X} \frac{dX}{dt}.$$

8.2. Theorem on constant relative growth. *Suppose that a time-dependent quantity $X(t)$ has a constant relative growth rate, meaning that*

$$\frac{1}{X} \frac{dX}{dt} = k \text{ is constant.} \tag{70}$$

And suppose that at time t_0 the value of X is known to be $X(t_0) = X_0$. Then

$$X(t) = X_0 e^{k(t - t_0)}. \tag{71}$$

Proof. The constant relative growth rate condition (70) implies

$$\frac{dX(t)}{dt} = kX(t). \tag{72}$$

The trick is to look at the rate of change of $e^{-kt}X(t)$:

$$\begin{aligned} \frac{dX(t)e^{-kt}}{dt} &= X(t)\frac{de^{-kt}}{dt} + \frac{dX(t)}{dt}e^{-kt} \\ &= -kX(t)e^{-kt} + X'(t)e^{-kt} \\ &= (X'(t) - kX(t))e^{-kt} \\ &= 0. \end{aligned}$$

In other words $X(t)e^{-kt}$ does not depend on t. Therefore

$$X(t)e^{-kt} = X(t_0)e^{-kt_0} = X_0 e^{-kt_0}, \text{ for any } t.$$

Multiply with e^{kt} and we end up with

$$X(t) = X_0 e^{k(t-t_0)}.$$

□

The equations (70) and (72) are examples of *differential equations*. They are equations in which the unknown quantity X is not a number but a function, and in which the derivative of the unknown function appears. We have just reasoned that any function that satisfies (70) (has constant relative growth) must be of the form (71): we have solved the differential equation[3].

8.3. Examples of constant relative growth or decay. Here is a typical scenario of constant relative growth or, rather, decay. Suppose that the molecules of a certain chemical substance appear in two varieties, the "normal" kind A, and the "excited" kind A*. Left to themselves the normal kind of molecules A are stable, but the excited kind will randomly decay to the normal kind. In the long run all A* will be converted into regular A molecules. One example of this kind of reaction is radioactive decay where, say, $_{14}$C decays to $_{14}$N. There are *many* other similar examples.

How fast does the conversion from A* *to* A *happen?* A very common reasoning to find the conversion rate goes like this. Suppose $X(t)$ is the total amount of A* and suppose that during a short time interval of length Δt a certain number of conversions from A* to A take place. If the time interval is short, then the number of conversions will be small, and the total amount $X(t)$ of A* molecules will not change much. Therefore, during the next short time interval of length Δt you would expect the same number of conversions to happen. Adding these together you conclude that if you wait twice as long ($2\Delta t$) then the number of conversions doubles; more generally, you would expect the number of conversions during a short time interval of duration Δt to be proportional to Δt.

If on the other hand you fix the length of the time interval, but double the amount of A* molecules, you would expect twice as many conversions to occur, so, the number of conversions from A* to A during some short time interval Δt should also be proportional with the amount of A* present.

To say that a reaction takes place in which A* decays to A you could write:

$$\mathrm{A}^* \xrightarrow{k} \mathrm{A}.$$

The k above the arrow is the "reaction rate."

Putting this together, we see that the change in the amount of A* during a time interval of length Δt should be proportional both with X and with Δt, so

$$\Delta X \approx -k \cdot X \cdot \Delta t,$$

for some constant k. The approximation should be better as you make the time interval shorter, and therefore, dividing by $X\Delta t$ and taking the limit $\Delta t \to 0$, we see that the amount of X should satisfy

$$\frac{1}{X}\frac{dX}{dt} = -k,$$

which is (70), but with k replaced by $-k$. The conclusion is: if you know that at time t_0 the amount of A* is $X(t_0) = X_0$, then at any other time t,

$$X(t) = X_0 e^{-k(t-t_0)}.$$

[3]The trick we used here is an example of solving differential equations by multiplying with an "integrating factor". We will do this more systematically in Math 222.

8.4. 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 a 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.

8.5. Determining X_0 and k. The general exponential growth/decay function (71) 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.)

9. Problems

Sketch the graphs of the following functions. You should

- find where f, f' and f'' are positive or negative
- find all stationary points and classify them as local minima, maxima, or neither
- find any global maxima or minima, if they exist
- find all inflection points
- determine all intervals where the function is increasing or decreasing
- determine all intervals where the function is convex or concave
- find any horizontal, vertical, or slant asymptotes

(Hint for some of these: if you have to solve something like $e^{4x} - 3e^{3x} + e^x = 0$, then call $w = e^x$ to get a polynomial equation $w^4 - 3w^3 + w = 0$ for w.)

1. $y = e^x$

2. $y = e^{-x}$

3. $y = e^x + e^{-2x}$ •

4. $y = e^{3x} - 4e^x$ •

5. $y = \dfrac{e^x}{1+e^x}$

6. $y = \dfrac{2e^x}{1+e^{2x}}$

7. $y = xe^{-x}$

8. $y = \sqrt{x}e^{-x/4}$

9. $y = x^2e^{x+2}$

10. $y = e^{x/2} - x$

11. $y = \ln\sqrt{x}$

12. $y = \ln\dfrac{1}{x}$

13. $y = x\ln x$

14. $y = \dfrac{-1}{\ln x}\quad (0 < x < \infty, x \neq 1)$

15. $y = (\ln x)^2\quad (x > 0)$

16. $y = \dfrac{\ln x}{x}\quad (x > 0)$

17. $y = \ln\sqrt{\dfrac{1+x}{1-x}}\quad (|x| < 1)$

18. $y = \ln\big(1+x^2\big)$

19. $y = \ln\big(x^2 - 3x + 2\big)\quad (x > 2)$

20. $y = \ln\cos x\quad (|x| < \frac{\pi}{2})$

21. $y = \sqrt{e^{2x}}$

22. The function $f(x) = e^{-x^2}$ plays a central role in statistics and its graph is called ***the bell curve*** (because of its shape). Sketch the graph of f.

23. 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 $y = 1/x$.)

24. A ***damped oscillation*** is a function of the form

$$f(x) = e^{-ax}[\cos b(x-c)]$$

where a, b, and c are constants.

(a) $g(x) = e^{-x}\sin 10x$ is a damped oscillation. What are the constants a, b, and c?

(b) Draw the graphs of $y = e^{-x}$, $y = -e^{-x}$, and $y = g(x)$ in one sketch (with pencil and paper). Make sure you include the piece of the graph with $0 \le x \le 2\pi$.

(c) $y = g(x)$ has many local maxima and minima. *What is the ratio between the function values at two consecutive local maxima?* (Hint: the answer does not depend on which pair of consecutive local maxima you consider.)

25. Find the inflection points on the graph of $f(x) = (1+x)\ln x$, for $x > 0$.

26. **(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 $-\infty < x < \infty$)?

Find the following limits.

27. $\displaystyle\lim_{x\to\infty}\frac{e^x-1}{e^x+1}$

28. $\displaystyle\lim_{x\to\infty}\frac{e^x-x^2}{e^x+x}$

29. $\displaystyle\lim_{x\to\infty}\frac{2^x}{3^x-2^x}$

30. $\displaystyle\lim_{x\to\infty}\frac{e^x-x^2}{e^{2x}+e^{-x}}$

31. $\displaystyle\lim_{x\to\infty}\frac{e^{-x}-e^{-x/2}}{\sqrt{e^x+1}}$

32. $\displaystyle\lim_{x\to\infty}\frac{\sqrt{x+e^{4x}}}{e^{2x}+x}$

33. $\displaystyle\lim_{x\to\infty}\frac{e^{\sqrt{x}}}{\sqrt{e^x}}$

34. $\displaystyle\lim_{x\to\infty}\frac{e^{\sqrt{x}}}{\sqrt{e^x+1}}$

35. $\displaystyle\lim_{x\to\infty}\frac{\ln x}{\ln(x^2)}$

36. $\displaystyle\lim_{x\to 0} x\ln x$

37. $\displaystyle\lim_{x\to\infty}\frac{\ln x}{\sqrt{x}+\ln x}$

38. $\displaystyle\lim_{x\to 0}\frac{\ln x}{\sqrt{x}+\ln x}$

39. $\displaystyle\lim_{x\to\infty}\ln(1+x) - \ln x$

40. Find the tenth derivative of xe^x. Then find the 99th derivative of xe^x.

41. For which real number x is $2^x - 3^x$ the largest?

42. Find $\dfrac{dx^x}{dx}$, $\dfrac{dx^{x^x}}{dx}$, and $\dfrac{d(x^x)^x}{dx}$.

Hint: write x^x as $e^{\text{something}}$.

43. **[Group Problem]** *About logarithmic differentiation – see Problem 8.17.*

(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.

44. 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?

45. Polonium-210 has a half-life of 140 days.

(a) If a sample has a mass of 200mg, 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.

46. Here are three versions of the same problem (read all & choose one):

(a) The number of individuals of an endangered species decreases to half its size every 1000 days. How long does it take for the population to decrease to 75% of its current size?

(b) A radioactive substance has a half-life of 1000 years. How long does it take for 25% of it to decay?

(c) The concentration of gas released in a chamber has a half-life of 1000 seconds. How long it would take for the gas to lower its concentration to 75% of its initial level?

47. The remains of an old campfire are unearthed and it is found that there is only 80% as much radioactive carbon-14 in the charcoal samples from the campfire as there is in modern living trees. If the half-life of carbon-14 is 5730 years, how long ago did the campfire burn?

48. Max needs 1 gram of Kryptonite to perform an important experiment. Unfortunately, he doesn't have the lab set up for his experiment yet and his Kryptonite is decaying. Yesterday there were 15 grams of Kryptonite left and today there are only 12 grams left. How long does he have before he won't have enough Kryptonite left to do his experiment?

49. Two equivalent problems – read them and pick one:

(a) (*Newton's law of cooling*) Newton's law of cooling states that the rate of change of temperature T of an object is proportional to the difference between its temperature and the ambient temperature, U, that is,

$$\frac{dT}{dt} = k\left(U - T(t)\right).$$

A body is discovered at 7 am. Its temperature is then 25° C. By the time police arrives at the site, at 8 pm, the body has cooled at 20° C. Assuming the normal body temperature in humans is 37° C, and that the ambient temperature is 5° C, determine the time of death.

(b) A nuclear reactor was shut down for maintenance repairs. The rate of change of temperature in the fuel rods of the reactor is described by Newton's law of cooling:

$$\frac{dT}{dt} = k\left(U - T(t)\right),$$

where $U = 5^\circ$ C is the temperature of the water being pumped into the reactor. The normal temperature of the fuel rods in a working reactor is 37° C. If at 7 am, the temperature of fuel rods was 25° C, and an hour later the fuel rods have cooled at 20° C, when was the reactor shut down?

50. A bacteria population grows exponentially with growth rate $k = 2.5$. How long it will take for the bacteria population to double its size?

51. Three equivalent problems – read all and do one:

(a) The half-life of a certain drug in the human body is 8 hours. How long it would take for a dose of this drug to decay to 0.1% of its initial level?

(b) After a nuclear plant explosion, radioactive material is scattered in the atmosphere. Particularly dangerous for human health is the iodine-131, whose half-life is 8 days. How long it would take for the iodine-131 in the atmosphere to decay to 0.1% of its initial level?

(c) The concentration of gas released in a chamber has a half-life of 8 minutes. How long it would take for the gas to lower its concentration to 0.1% of its initial level?

52. The *hyperbolic functions* are defined by

$$\sinh x = \frac{e^x - e^{-x}}{2},$$
$$\cosh x = \frac{e^x + e^{-x}}{2},$$
$$\tanh x = \frac{\sinh x}{\cosh x}.$$

(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

$$\frac{d \sinh x}{dx} = \cosh x,$$
$$\frac{d \cosh x}{dx} = \sinh x,$$
$$\frac{d \tanh x}{dx} = \frac{1}{\cosh^2 x}.$$

(c) Sketch the graphs of the three hyperbolic functions.

53. Sketch the following parametrized curves (first review Chapter V, §15):

(a) $x(t) = e^t$, $y(t) = e^t$

(b) $x(t) = e^t$, $y(t) = t$

(c) $x(t) = e^t$, $y(t) = e^{-t}$

54. **(a)** At which point on the graph of $y = e^x$ is the curvature the largest?

(b) At which point on the graph of $y = \ln x$ is the curvature the largest?

CHAPTER 7

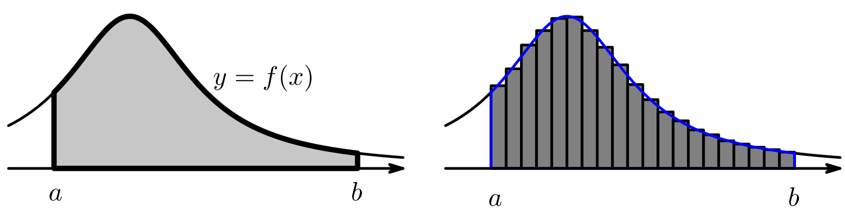
The Integral

In this chapter we define the integral of a function on an interval $[a, b]$, introduce the Fundamental Theorem of Calculus relating integration and differentiation, and develop basic techniques for computing integrals.

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.

1. Area under a Graph

Let f be a function which is defined on an interval $a \le x \le b$ and assume it is positive (so 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 $y = b$?* We can



try to compute this area (on the left here) by approximating the region with many thin rectangles (on the right). The idea is that even though we don't know how to compute the area of a region bounded by arbitrary curves, we do know how to find the area of one or more rectangles.

To write formulas for the area of those rectangles we first have to introduce some notation – have a look at Figure 1 before reading on. To make the approximating region, we choose a ***partition*** of the interval $[a, b]$: we 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 $[x_{k-1}, x_k]$, we choose a ***sample point*** c_k: thus in the first interval we choose some number c_1 such that $x_0 \le c_1 \le x_1$, in the second interval we choose some c_2 with $x_1 \le c_2 \le x_2$, and so forth. (See Figure 1.)

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).

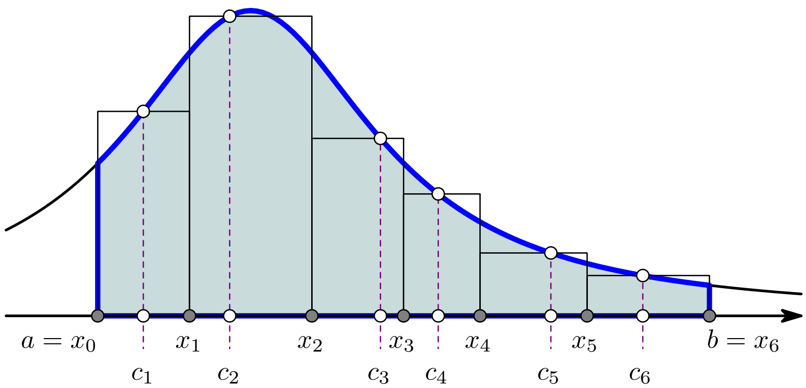


Figure 1. A partition, and a Riemann sum. Here, the interval $a < x < b$ is divided up into six smaller intervals. In each of those intervals a sample point c_i is chosen, and the resulting rectangles with heights $f(c_1), \ldots, f(c_6)$ are drawn. The total area under the graph of the function is roughly equal to the total area of the rectangles.

The area of the k^{th} rectangle is the product of its height and width, which is $f(c_k)\Delta x_k$. Adding up all the rectangles' areas yields

$$R = f(c_1)\Delta x_1 + f(c_2)\Delta x_2 + \cdots + f(c_n)\Delta x_n. \tag{73}$$

This kind of sum is called a ***Riemann sum.***

If the rectangles are all sufficiently narrow then we would expect that the total area of all rectangles should be a good approximation of the area of the region under the graph. So we would expect the "area under the curve" 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:

1.1. Definition. *If f is a function defined on an interval $[a, b]$, then we say that*

$$\int_a^b f(x)dx = I,$$

read as "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 $\Delta x_k < \delta$ with any arbitrary choice of sample points $c_1, \ldots, c_n$.

This is a perfectly good definition, but it's not even clear that the limit exists even for comparatively simple functions like $f(x) = x$, let alone for more complicated functions. It turns out that, with a fair amount of effort, one can prove that this limit always exists provided that $f(x)$ is a continuous function on the interval $[a, b]$. Even so, it is quite hard to evaluate areas this way.

2. When f changes its sign

If the function f is not necessarily positive everywhere in the interval $a \le x \le b$, 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.

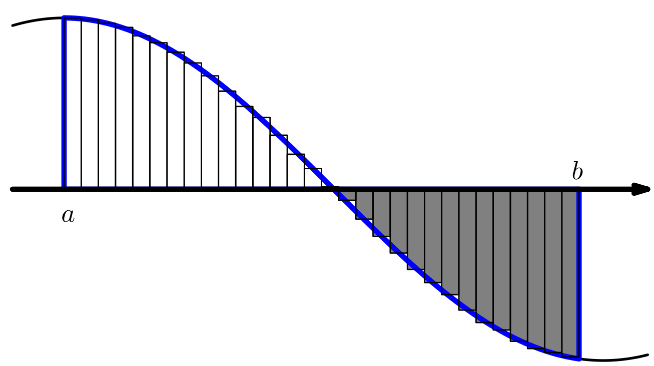


Figure 2. Riemann sum for a function whose sign changes. Always remember this:
INTEGRALS CAN BE NEGATIVE, BUT
AREAS ARE ALWAYS POSITIVE NUMBERS.
The Riemann-sum corresponding to this picture is the total area of the rectangles above the x-axis *minus* the total area of the rectangles below the x-axis.

Let f be a function on the interval $a \le x \le b$, 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 -1 times the area of the corresponding rectangle. The Riemann sum is therefore the area of the rectangles above the x-axis *minus* the area of the rectangles below the x-axis. Taking the limit over finer and finer partitions, we conclude that

$$\int_a^b f(x)dx = \begin{array}{c}\text{area above the } x\text{-axis, below the graph of } y=f(x)\\ \textbf{\textit{minus}}\\ \text{the area below the } x\text{-axis, above the graph of } y=f(x)\end{array}$$

3. The Fundamental Theorem of Calculus

3.1. Definition. *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 $a \le x \le b$.*

For instance, $F(x) = \frac{1}{2}x^2$ is an antiderivative of $f(x) = x$, but so are $G(x) = \frac{1}{2}x^2 + 1$ and $H(x) = \frac{1}{2}x^2 + 2012$.

3.2. Theorem. *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*

$$\int_a^b f(x)dx = F(b) - F(a). \tag{74}$$

(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

$$F(b) - F(a) \stackrel{\text{def}}{=} F(x)\Big]_{x=a}^{b} = F(x)\Big]_{a}^{b}.$$

3.3. Terminology. In the integral

$$\int_a^b f(x)\,dx$$

the numbers a and b are called the *limits of integration*, 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 we systematically replace it with another variable, the resulting integral will still be the same. For instance,

$$\int_0^1 x^2\,dx = \tfrac{1}{3}x^3\Big|_{x=0}^{1} = \tfrac{1}{3},$$

and if we replace x by φ we still get

$$\int_0^1 \varphi^2\,d\varphi = \tfrac{1}{3}\varphi^3\Big|_{\varphi=0}^{1} = \tfrac{1}{3}.$$

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 has absolutely nothing to do with the variable x.

4. Summation notation

A Riemann sum is a summation with very many terms: more to the point, a Riemann sum usually has "n terms" where n is allowed to vary. For such a sum we cannot write all terms in the way we can for a sum that has, say, 25 terms. This is why our formulas for Riemann sums, such as (73),

$$R = f(c_1)\Delta x_1 + f(c_2)\Delta x_2 + \cdots + f(c_n)\Delta x_n$$

contain dots ($\cdots$) to represent the terms we didn't explicitly write. Even though we did not write most of the terms in this sum it is clear what they are, just from looking at the pattern of the first two and the last term.

If we don't like the ambiguity created by the dots in the Riemann sum, then we can use summation notation: if we have a sum with n terms, and if we have a formula a_k for term number k in the sum, then we write

$$\sum_{k=1}^{n} a_k = a_1 + a_2 + \cdots + a_n.$$

The expression on the right is pronounced as "the sum from $k = 1$ to n of a_k". For example,

$$\sum_{k=3}^{7} k^2 = 3^2 + 4^2 + 5^2 + 6^2 + 7^2,$$

which adds up to $9 + 16 + 25 + 36 + 49 = 135$.

With this notation we can write the Riemann sum (73) as

$$R = f(c_1)\Delta x_1 + f(c_2)\Delta x_2 + \cdots + f(c_n)\Delta x_n = \sum_{k=1}^{n} f(c_k)\Delta x_k.$$

The advantage of Σ notation is that it takes the guesswork out of the summation: we don't have to guess what the 23^{rd} term is because there is a formula in which we can just substitute $k = 23$. If we don't care for the Σ notation (like the author of this text, but perhaps unlike your instructor) then our writing has to be so clear that the reader can easily guess what the dots represent when we write them.

Using Σ notation our definition of the integral is often written in the following form

$$\int_a^b f(x)dx = \lim_{n\to\infty} \sum_{k=1}^{n} f(c_k)\Delta x_k. \tag{75}$$

This way of writing the definition of the integral gives a hint of where the integral sign came from. On the right a Σ (the greek upper case "S") was used to denote a summation. On the left is an integral, which is not really a sum, but a limit of sums. To denote it, Leibniz chose an S, which in the typesetting of his time looked like a "long s," and which today looks more like an f than an *s*.

State ftreet, Madifon, Wifconfin

The "long s" from which the integral sign derives.

5. Problems

1. **(a)** What is a Riemann sum of a function $y = f(x)$ on an interval $a \le x \le b$?

(b) Compute $\sum_{k=1}^{4} \frac{1}{k}$.

(c) Compute $\sum_{k=4}^{7} 2^k$.

2. Let f be the function $f(x) = 1 - x^2$.

(a) Draw the graph of $f(x)$ with $0 \le x \le 2$.

(b) Compute the Riemann-sum for the partition

$$0 < \tfrac{1}{3} < 1 < \tfrac{3}{2} < 2$$

of the interval $[a, b] = [0, 2]$ if we choose each c_k to be the left endpoint of the interval it belongs to. Draw the corresponding rectangles (add them to our drawing of the graph of f).

(c) Compute the Riemann-sum you get if you choose the c_k to be the right endpoint of the interval it belongs to. Make a new drawing of the graph of f and include the rectangles corresponding to the right endpoint Riemann-sum.

3. *The analog of the Fundamental Theorem for finite sums.*In this problem you'll compute

$$1 + 3 + 5 + 7 + \cdots + 159$$

in a smart way. First note that there are 80 terms, and that the kth term is $f(k) = 2k - 1$, so you really are going to compute

$$f(1) + f(2) + \cdots + f(80).$$

(a) Show that for the function $F(k) = k^2$, it is true that $F(k) - F(k-1) = f(k)$.

(b) Compute $S = f(1) + f(2) + \cdots f(80)$ in terms of the function F.

4. **[Group Problem]** Look at figure 1 (top). Which choice of intermediate points c_1, $\ldots$, c_6 leads to the smallest Riemann sum? Which choice would give you the largest Riemann-sum?

(Note: in this problem you're not allowed to change the division points x_i, only the points c_i in between them.)

♣

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.

5. $f(x) = 2x + 1$

6. $f(x) = 1 - 3x$

7. $f(x) = x^2 - x + 11$

8. $f(x) = x^4 - x^2$

9. $f(x) = 1 + x + \dfrac{x^2}{2} + \dfrac{x^3}{3} + \dfrac{x^4}{4}$

10. $f(x) = \dfrac{1}{x}$

11. $f(x) = e^x$

12. $f(x) = \dfrac{2}{x}$

13. $f(x) = e^{2x}$

14. $f(x) = \dfrac{1}{2+x}$

15. $f(x) = \dfrac{e^x - e^{-x}}{2}$

16. $f(x) = \dfrac{1}{1+x^2}$

17. $f(x) = \dfrac{e^x + e^{-x}}{2}$

18. $f(x) = \dfrac{1}{\sqrt{1-x^2}}$

19. $f(x) = \sin x$

20. $f(x) = \dfrac{2}{1-x}$

21. $f(x) = \cos x$

22. $f(x) = \cos 2x$

23. $f(x) = \sin(x - \pi/3)$

24. $f(x) = \sin x + \sin 2x$

25. $f(x) = 2x(1 + x^2)^5$

♡

In each of the following exercises, draw the indicated region and compute its area.

26. The region between the vertical lines $x = 0$ and $x = 1$, and between the x-axis and the graph of $y = x^3$.

27. 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 $n = \frac{1}{2}, 1, 2, 3, 4$).

28. The region above the graph of $y = \sqrt{x}$, below the line $y = 2$, and between the vertical lines $x = 0$, $x = 4$.

29. The bounded region above the x-axis and below the graph of $f(x) = x^2 - x^3$.

30. The bounded region above the x-axis and below the graph of $f(x) = 4x^2 - x^4$.

31. The region above the x-axis and below the graph of $f(x) = 1 - x^4$.

32. The region above the x-axis, below the graph of $f(x) = \sin x$, and between $x = 0$ and $x = \pi$.

33. The region above the x-axis, below the graph of $f(x) = 1/(1 + x^2)$ (a curve known as the *witch of Maria Agnesi*), and between $x = 0$ and $x = 1$.

34. 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.)

35. The bounded region below the x-axis and above the graph of

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

36. Compute

$$\int_0^1 \sqrt{1-x^2}dx$$

without finding 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?)

37. **[Group Problem]** Compute these integrals without finding antiderivatives.

$$I = \int_0^{1/2} \sqrt{1-x^2}\, dx$$

$$J = \int_{-1}^{1} |1 - x|\, dx$$

$$K = \int_{-1}^{1} |2 - x|\, dx$$

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 $f(x)$*.	$\int_a^b f(x)dx$ is defined in terms of Riemann sums and can be interpreted as the "area under the graph of $y = f(x)$" 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 different 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 $\int_0^1 2xdx = \int_0^1 2tdt$.

Table 1. The differences between ***definite*** and ***indefinite*** integrals

6. 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 an antiderivative. In order to simplify the computation of the integral

$$\int_a^b f(x)dx = F(b) - F(a) \tag{76}$$

the following notation is commonly used for the antiderivative:

$$F(x) = \int f(x)dx. \tag{77}$$

For instance,

$$\int x^2\, dx = \tfrac{1}{3}x^3 + C, \qquad \int \sin 5x\, dx = -\tfrac{1}{5}\cos 5x + C, \qquad \text{etc}\ldots$$

The integral which appears here does not have the limits of integration a and b. It is called an ***indefinite integral***, as opposed to the integral in (76) which is called a ***definite integral***.

It is important to distinguish between the two kinds of integrals. The main differences are listed in Table 1.

6.1. About "$+C$".

Let $f(x)$ be a function defined on the interval $[a, b]$. 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. *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. $\square$

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

$$\int 2\sin x\cos x\,dx = \sin^2 x + C$$
$$\int 2\sin x\cos x\,dx = -\cos^2 x + C$$

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: $\sin^2 x = -\cos^2 x + 1$, and the "$+1$" gets absorbed into the $+C$ term.

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.

6.2. 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 you get an "answer", $F(x)$, but you're not sure if you trust that it's right. Fortunately, it is easy to check: you need only differentiate the $F(x)$ you found, and if $F'(x)$ turns out to be equal to $f(x)$, then your $F(x)$ is indeed an antiderivative.

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:

$$\frac{d}{dx}\bigl(x\ln x - x\bigr) = x\cdot\frac{1}{x} + 1\cdot\ln x - 1 = \ln x.$$

Who knows how Louie thought of this[1], but it doesn't matter: he's right! We now know that $\int \ln x dx = x\ln x - x + C$.

Table 2 lists a number of antiderivatives which you should know. All of these integrals should be familiar from the differentiation rules we have learned so far, except for the integrals of $\tan x$ and of $\dfrac{1}{\cos x}$. You can check those by differentiation (using $\ln\frac{a}{b} = \ln a - \ln b$ simplifies things a bit).

7. 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

$$\frac{d}{dx}\bigl(F(x) + G(x)\bigr) = F'(x) + G'(x) = f(x) + g(x).$$

[1]He took math 222 and learned to integrate by parts.

In words, $F+G$ is an antiderivative of $f+g$. Using indefinite integrals we can write this as

$$\int \big(f(x)+g(x)\big)\,dx = \int f(x)\,dx + \int g(x)\,dx. \tag{78}$$

$f(x) = \dfrac{dF(x)}{dx}$	$\displaystyle\int f(x)\,dx = F(x) + C$	
$(n+1)x^n = \dfrac{dx^{n+1}}{dx}$	$\displaystyle\int x^n\,dx = \frac{x^{n+1}}{n+1} + C$	$n \neq -1$
$\dfrac{1}{x} = \dfrac{d\ln\lvert x\rvert}{dx}$	$\displaystyle\int \frac{1}{x}\,dx = \ln\lvert x\rvert + C$	absolute values!!
$e^x = \dfrac{de^x}{dx}$	$\displaystyle\int e^x\,dx = e^x + C$	
$-\sin x = \dfrac{d\cos x}{dx}$	$\displaystyle\int \sin x\,dx = -\cos x + C$	
$\cos x = \dfrac{d\sin x}{dx}$	$\displaystyle\int \cos x\,dx = \sin x + C$	
$\tan x = -\dfrac{d\ln\lvert\cos x\rvert}{dx}$	$\displaystyle\int \tan x = -\ln\lvert\cos x\rvert + C$	absolute values!!
$\dfrac{1}{1+x^2} = \dfrac{d\arctan x}{dx}$	$\displaystyle\int \frac{1}{1+x^2}\,dx = \arctan x + C$	
$\dfrac{1}{\sqrt{1-x^2}} = \dfrac{d\arcsin x}{dx}$	$\displaystyle\int \frac{1}{\sqrt{1-x^2}}\,dx = \arcsin x + C$	
$f(x)+g(x) = \dfrac{dF(x)+G(x)}{dx}$	$\displaystyle\int \{f(x)+g(x)\}\,dx = F(x)+G(x)+C$	See (80)
$cf(x) = \dfrac{d\,cF(x)}{dx}$	$\displaystyle\int cf(x)\,dx = cF(x)+C$	See (81)

To find derivatives and integrals involving a^x instead of e^x use $a = e^{\ln a}$, and thus $a^x = e^{x\ln a}$, to rewrite all exponentials as $e^{\cdots}$.

The following integral is also useful, but not as important as the ones above:

$$\int \frac{dx}{\cos x} = \frac{1}{2}\ln\left(\frac{1+\sin x}{1-\sin x}\right) + C \text{ for } \cos x \neq 0.$$

Table 2. The list of the standard integrals everyone should know

Similarly, $\frac{d}{dx}\big(cF(x)\big) = cF'(x) = cf(x)$ implies that

$$\int cf(x)\,dx = c\int f(x)\,dx \tag{79}$$

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

$$\int_a^b \big[f(x) + g(x)\big]dx = \int_a^b f(x)dx + \int_a^b g(x)dx, \tag{80}$$

and for any function f and constant c one has

$$\int_a^b cf(x)dx = c\int_a^b f(x)dx. \tag{81}$$

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.

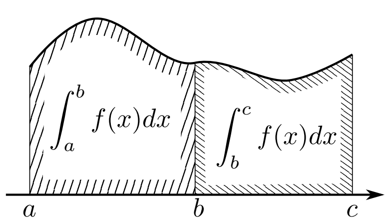
7.1. Theorem. *Given $a < b < c$, and a function on the interval $[a, b]$ then*

$$\int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx. \tag{82}$$

A picture proof, assuming $f(x) > 0$. The drawing shows the graph of a positive function on some interval $a \le x \le c$. The region below the graph consists of two pieces, namely the part with $a \le x \le b$ and the part with $b \le x \le c$. The total area under the graph is the sum of the areas of the two pieces. Since the function is positive, the areas are given by the corresponding integrals. The total area is

$$\int_a^c f(x)dx.$$

and it is the sum of the areas of the two smaller regions. This implies (82).



A proof using the Fundamental Theorem. Let F be an antiderivative of f. Then

$$\int_a^b f(x)dx = F(b) - F(a) \text{ and } \int_b^c f(x)dx = F(c) - F(b),$$

so that

$$\begin{aligned}\int_a^b f(x)dx + \int_b^c f(x)dx &= F(b) - F(a) + F(c) - F(b)\\ &= F(c) - F(a)\\ &= \int_a^c f(x)dx.\end{aligned}$$

7.2. When the upper integration bound is less than the lower bound. So far we have always assumed that $a < b$ in all indefinite integrals $\int_a^b \ldots$. The Fundamental Theorem suggests that when $b < a$, we should define the integral as

$$\int_a^b f(x)dx = F(b) - F(a) = -\big(F(a) - F(b)\big) = -\int_b^a f(x)dx. \tag{83}$$

For instance,

$$\int_1^0 x dx = -\int_0^1 x dx = -\tfrac{1}{2}.$$

8. The definite integral as a function of its integration bounds

8.1. A function defined by an integral. Consider the expression

$$I = \int_0^x t^2\, dt.$$

What does I depend on? To find out, we calculate the integral

$$I = \left[\tfrac{1}{3}t^3\right]_0^x = \tfrac{1}{3}x^3 - \tfrac{1}{3}0^3 = \tfrac{1}{3}x^3.$$

So the integral depends on x. It does not depend on t, since t is a "dummy variable" (see §3.3, where we already discussed this point).

In this way we 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, $I(x) = \frac{1}{3}x^3$).

This example does not really define a new function, in the sense that we already had a much simpler way of defining the same function, namely "$I(x) = x^3/3$." An example of a *new* function defined by an integral is the so called ***error function*** from statistics:

$$\operatorname{erf}(x) \stackrel{\text{def}}{=} \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt, \tag{84}$$

so that $\operatorname{erf}(x)$ is the area of the shaded region in figure 3. The integral in (84) cannot

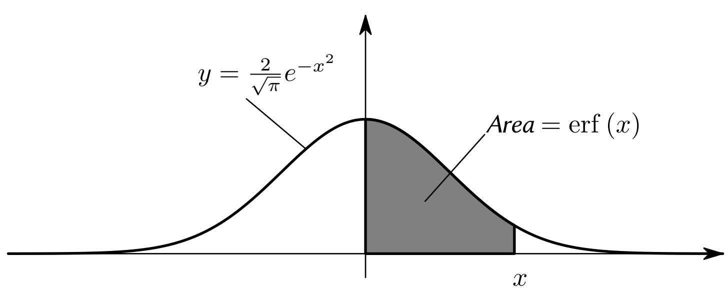


Figure 3. Definition of the Error function. The graph is known as the "Bell curve" or the Gaussian curve.

be computed in terms of the standard functions (square and higher roots, sine, cosine, exponential and logarithms). Since the integral in (84) occurs very often in statistics (in relation with the so-called normal distribution) it has been given its own name, "erf (x)".

8.2. 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

$$\frac{d}{dx}\int_a^x f(t)\,dt = \frac{d}{dx}\big(F(x) - F(a)\big) = F'(x) = f(x),$$

Note that here we use operator notation, as in Chapter IV, § 9.3, and that, by definition,

$$\frac{d}{dx}\int_a^x f(t)dt = \frac{d\int_a^x f(t)dt}{dx}$$

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? It is

$$\begin{aligned}\text{erf}\,'(x) &= \frac{d}{dx}\left[\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt\right]\\ &= \frac{2}{\sqrt{\pi}}\frac{d}{dx}\left[\int_0^x e^{-t^2}\,dt\right]\\ &= \frac{2}{\sqrt{\pi}}e^{-x^2}.\end{aligned}$$

9. Substitution in Integrals

The Chain Rule says that

$$\frac{dF(G(x))}{dx} = F'(G(x))\cdot G'(x),$$

so that

$$\int F'(G(x))\cdot G'(x)\,dx = F(G(x)) + C.$$

We can use this result to evaluate new integrals by making substitutions.

9.1. 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 that $2x = \frac{d}{dx}(x^2+3)$. So let's call $G(x) = x^2+3$, and $F(u) = -\cos u$, then

With practice it gets easier to recognize what substitution could help you compute a specific integral, but even with a lot of practice, substitution is still a matter of trial and error.

$$F(G(x)) = -\cos(x^2+3)$$

and

$$\frac{dF(G(x))}{dx} = \underbrace{\sin(x^2+3)}_{F'(G(x))}\cdot\underbrace{2x}_{G'(x)} = f(x),$$

so that

$$\int 2x\sin(x^2+3)\,dx = -\cos(x^2+3) + C. \tag{85}$$

9.2. Leibniz notation for substitution. The most transparent way of computing an integral by substitution is by following Leibniz and introduce new variables. Thus to evaluate 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 compute the integral (85) 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

$$\int 2x\sin(x^2+3)\,dx = -\cos\big(x^2+3\big) + C.$$

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.

9.3. Substitution for definite integrals. Substitution in definite integrals is essentially the same as substitution in indefinite integrals, except we must remember that the limits of integration will change as well when we change variables. Explicitly: for definite integrals the Chain Rule

$$\frac{d}{dx}\big(F(G(x))\big) = F'(G(x))G'(x) = f(G(x))G'(x)$$

implies

$$\int_a^b f(G(x))G'(x)\,dx = F(G(b)) - F(G(a)).$$

which you can also write as

$$\int_{x=a}^{b} f(G(x))G'(x)\,dx = \int_{u=G(a)}^{G(b)} f(u)\,du. \tag{86}$$

9.4. 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 we must compute the new integration bounds $G(0)$ and $G(1)$ (see equation (86).) 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

$$\int_0^1 \frac{x}{1+x^2}\,dx = \tfrac{1}{2}\int_1^2 \frac{1}{u}\,du, \tag{87}$$

which is in our list of memorable integrals. So we find

$$\int_0^1 \frac{x}{1+x^2}\,dx = \tfrac{1}{2}\int_1^2 \frac{1}{u}\,du = \tfrac{1}{2}\ln u\Big|_{u=1}^{2} = \tfrac{1}{2}\ln 2.$$

Sometimes the integrals in (87) 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.

10. Problems

1. **(a)** Simplicio has read the Fundamental Theorem of Calculus. Now he has a function F, and he says that

$$F(x) = \int_0^x F'(t)dt.$$

Is he right?

(b) He also claims that

$$F(x) = \int F'(x)dx.$$

Is this right?

Compute these derivatives:

2. $\displaystyle \frac{d}{dx}\int_0^x \left(1+t^2\right)^4\,dt$

3. $\displaystyle \frac{d}{dx}\int_x^1 \ln z\,dz$

4. $\displaystyle \frac{d}{dt}\int_0^t \frac{dx}{1+x^2}$

5. $\displaystyle \frac{d}{dt}\int_0^{1/t} \frac{dx}{1+x^2}$

6. $\displaystyle \frac{d}{dx}\int_x^{2x} s^2 ds$

7. **(a)** $\displaystyle \frac{d}{dq}\int_{-q}^{q} \frac{dx}{1-x^2}$

(b) For which values of q is your answer to part (a) valid?

8. $\displaystyle \frac{d}{dt}\int_0^{t^2} e^{2x}dx$

9. **[Group Problem]** You can see the graph of the error function at

wikipedia.org/wiki/Error_function

(a) Compute the second derivative of the error function. How many inflection points does the graph of the error function have?

(b) The graph of the error function on Wikipedia shows that $\operatorname{erf}(x)$ is negative when $x < 0$. But the error function is defined as an integral of a positive function so it should be positive. Is Wikipedia wrong? Explain.

10. Suppose $f(x)$ is any function, and consider a new function $g(x)$ given by

$$g(x) = \int_0^x (x-t)f(t)dt.$$

This kind of integral is an example of what is called a *convolution integral.* It often shows up as the solution to linear differential equations (take math 319 or 320 to learn more) and can describe the response (g) of a linear electronic circuit to some given input (f).

This problem tests how careful you are with dummy variables.

(a) To warm up, compute $g(x)$ in the case that $f(x) = x$.

From here on we drop the assumption $f(x) = x$, i.e. suppose $f(x)$ could be any function.

(b) Simplicio rewrote the integral like this:

$$\begin{aligned} g(x) &= \int_0^x (x-t)f(t)dt \\ &= \int_0^x \{xf(t) - tf(t)\}dt \\ &= \int_0^x xf(t)dt - \int_0^x tf(t)dt \\ &= x\int_0^x f(t)dt - \int_0^x tf(t)dt. \end{aligned}$$

Now Simplicio says that t is a dummy variable, so you can replace it with any other variable you like. He chooses x:

$$\begin{aligned} g(x) &= x\int_0^x f(x)dx - \int_0^x xf(x)dx \\ &= 0. \end{aligned}$$

In particular, he predicts that your answer in part **(a)** should have been $g(x) = 0$. Where did Simplicio go wrong?

(c) Find $g'(x)$ and $g''(x)$. (Hint: part of Simplicio's computation is correct, and will help you compute $g'(x)$ and $g''(x)$.)

♣

Compute the following indefinite integrals:

11. $\displaystyle\int (6x^5 - 2x^{-4} - 7x)\,dx$

12. $\displaystyle\int (3/x - 5 + 4e^x + 7^x)\,dx$

13. $\displaystyle\int (x/a + a/x + x^a + a^x + ax)\,dx$

14. $\displaystyle\int \left(\sqrt{x} - \sqrt[3]{x^4} + \frac{7}{\sqrt[3]{x^2}} - 6e^x + 1\right)\,dx$

15. $\displaystyle\int \left(2^x + \left(\tfrac{1}{2}\right)^x\right)\,dx$

16. $\displaystyle\int_{-2}^4 (3x-5)\,dx$

17. $\displaystyle\int_1^4 x^{-2}\,dx$

18. $\displaystyle\int_1^4 t^{-2}\,dt$

19. $\displaystyle\int_1^4 x^{-2}\,dt$

20. $\displaystyle\int_0^1 (1-2x-3x^2)\,dx$

21. $\displaystyle\int_1^2 (5x^2-4x+3)\,dx$

22. $\displaystyle\int_{-3}^0 (5y^4-6y^2+14)\,dy$

23. $\displaystyle\int_0^1 (y^9-2y^5+3y)\,dy$

24. $\displaystyle\int_0^4 \sqrt{x}\,dx$

25. $\displaystyle\int_0^1 x^{3/7}\,dx$

26. $\displaystyle\int_1^3 \left(\frac{1}{t^2} - \frac{1}{t^4}\right)\,dt$

27. $\displaystyle\int_1^2 \frac{t^6-t^2}{t^4}\,dt$

28. $\displaystyle\int_1^2 \frac{x^2+1}{\sqrt{x}}\,dx$

29. $\displaystyle\int_0^2 (x^3-1)^2\,dx$

30. $\displaystyle\int_0^1 u(\sqrt{u}+\sqrt[3]{u})\,du$

31. $\displaystyle\int_1^2 (x+1/x)^2\,dx$

32. $\displaystyle\int_3^3 \sqrt{x^5+2}\,dx$

33. $\displaystyle\int_1^{-1} (x-1)(3x+2)\,dx$

34. $\displaystyle\int_1^4 (\sqrt{t} - 2/\sqrt{t})\,dt$

35. $\displaystyle\int_1^8 \left(\sqrt[3]{r} + \frac{1}{\sqrt[3]{r}}\right) dr$

36. $\displaystyle\int_{-1}^0 (x+1)^3\, dx$

37. $\displaystyle\int_{-5}^{-2} \frac{x^4-1}{x^2+1}\, dx$

38. $\displaystyle\int_1^e \frac{x^2+x+1}{x}\, dx$

39. $\displaystyle\int_4^9 \left(\sqrt{x} + \frac{1}{\sqrt{x}}\right)^2 dx$

40. $\displaystyle\int_0^1 \left(\sqrt[4]{x^5} + \sqrt[5]{x^4}\right) dx$

41. $\displaystyle\int_1^8 \frac{x-1}{\sqrt[3]{x^2}}\, dx$

42. $\displaystyle\int_{\pi/4}^{\pi/3} \sin t\, dt$

43. $\displaystyle\int_0^{\pi/2} (\cos\theta + 2\sin\theta)\, d\theta$

44. $\displaystyle\int_0^{\pi/2} (\cos\theta + \sin 2\theta)\, d\theta$

45. $\displaystyle\int_{2\pi/3}^{\pi} \frac{\tan x}{\cos x}\, dx$

46. $\displaystyle\int_{\pi/3}^{\pi/2} \frac{\cot x}{\sin x}\, dx$

47. $\displaystyle\int_1^{\sqrt{3}} \frac{6}{1+x^2}\, dx$

48. $\displaystyle\int_0^{0.5} \frac{dx}{\sqrt{1-x^2}}$

49. $\displaystyle\int_4^8 (1/x)\, dx$

50. $\displaystyle\int_{\ln 3}^{\ln 6} 8e^x\, dx$

51. $\displaystyle\int_8^9 2^t\, dt$

52. $\displaystyle\int_{-e^2}^{-e} \frac{3}{x}\, dx$

53. $\displaystyle\int_{-2}^3 |x^2-1|\, dx$

54. $\displaystyle\int_{-1}^2 |x-x^2|\, dx$

55. $\displaystyle\int_{-1}^2 (x-2|x|)\, dx$

56. $\displaystyle\int_0^2 (x^2-|x-1|)\, dx$

57. $\displaystyle\int_0^2 f(x)\, dx$ where

$$f(x) = \begin{cases} x^4 & \text{if } 0 \le x < 1, \\ x^5, & \text{if } 1 \le x \le 2. \end{cases}$$

58. $\displaystyle\int_{-\pi}^{\pi} f(x)\, dx$ where

$$f(x) = \begin{cases} x, & \text{if } -\pi \le x \le 0, \\ \sin x, & \text{if } 0 < x \le \pi. \end{cases}$$

59. 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 $u = 1+x^2$.

60. Compute

$$I_n = \int 2x\big(1+x^2\big)^n\, dx.$$

61. If $f'(x) = x - 1/x^2$ and $f(1) = 1/2$, find $f(x)$.

62. 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 $x=0$, $x=4$.

63. 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.

64. Graph the curve $y = 2\sqrt{1-x^2}$, for $0 \le x \le 1$, 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 $y = \sqrt{1-x^2}$.)

65. Determine the area under the curve $y = \sqrt{a^2-x^2}$ and between the lines $x=0$ and $x=a$.

66. Graph the curve $y = 2\sqrt{9-x^2}$ and determine the area enclosed between the curve and the x-axis.

67. Graph the area between the curve $y^2 = 4x$ and the line $x = 3$. Find the area of this region.

68. Find the area bounded by the curve $y = 4 - x^2$ and the lines $y = 0$ and $y = 3$.

69. Find the area enclosed between the curve $y = \sin 2x$, $0 \le x \le \pi/4$ and the axes.

70. Find the area enclosed between the curve $y = \cos 2x$, $0 \le x \le \pi/4$ and the axes.

71. Graph $y^2 + 1 = x$, and find the area enclosed by the curve and the line $x = 2$.

72. Find the area of the region bounded by the parabola $y^2 = 4x$ and the line $y = 2x$.

73. Find the area bounded by the curve $y = x(2 - x)$ and the line $x = 2y$.

74. Find the area bounded by the curve $x^2 = 4y$ and the line $x = 4y - 2$.

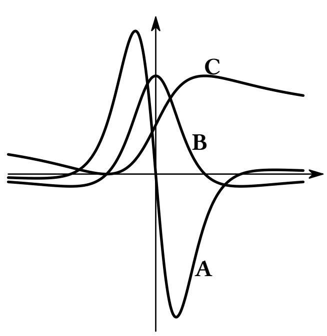
75. Calculate the area of the region bounded by the parabolas $y = x^2$ and $x = y^2$.

76. Find the area of the region included between the parabola $y^2 = x$ and the line $x + y = 2$.

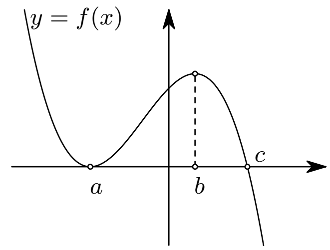
77. Find the area of the region bounded by the curves $y = \sqrt{x}$ and $y = x$.

78. **[Group Problem]** You asked your assistant Joe to produce graphs of a function $f(x)$, its derivative $f'(x)$ and an antiderivative $F(x)$ of $f(x)$.

Unfortunately Joe simply labelled the graphs “A,” ”B,” and “C,” and now he doesn’t remember which graph is f, which is f' and which is F. Identify which graph is which ***and explain your answer.***



79. **[Group Problem]** Below is the graph of a function $y = f(x)$.



The function $F(x)$ (graph not shown) is an antiderivative of $f(x)$. Which of the following statements are true?

(a) $F(a) = F(c)$

(b) $F(b) = 0$

(c) $F(b) > F(c)$

(d) $F(x)$ has exactly 2 inflection points.

♠

Use a substitution to evaluate the following integrals:

80. $\displaystyle\int_1^2 \frac{u\,du}{1+u^2}$

81. $\displaystyle\int_0^5 \frac{x\,dx}{\sqrt{x+1}}$

82. $\displaystyle\int_1^2 \frac{x^2\,dx}{\sqrt{2x+1}}$

83. $\displaystyle\int_0^5 \frac{s\,ds}{\sqrt[3]{s+2}}$

84. $\displaystyle\int_1^2 \frac{x\,dx}{1+x^2}$

85. $\displaystyle\int_0^\pi \cos\left(\theta + \frac{\pi}{3}\right) d\theta$

86. $\displaystyle\int \sin\left(\frac{\pi + x}{5}\right) dx$

87. $\displaystyle\int \frac{\sin 2x}{\sqrt{1+\cos 2x}}\,dx$

88. $\displaystyle\int_{\pi/4}^{\pi/3} \sin^2\theta\cos\theta\,d\theta$

89. $\displaystyle\int_2^3 \frac{1}{r\ln r}\,dr$

90. $\displaystyle\int \frac{\sin 2x}{1+\cos^2 x}\,dx$

91. $\displaystyle\int \frac{\sin 2x}{1+\sin x}\,dx$

92. $\displaystyle\int_0^1 z\sqrt{1-z^2}\,dz$

93. $\displaystyle\int_1^2 \frac{\ln 2x}{x}\,dx$

94. $\displaystyle\int_0^{\sqrt{2}} \xi(1+2\xi^2)^{10}\,d\xi$

95. $\displaystyle\int_2^3 \sin\rho\big(\cos 2\rho\big)^4\,d\rho$

96. $\displaystyle\int \alpha e^{-\alpha^2}\,d\alpha$

97. $\displaystyle\int \frac{e^{\frac{1}{t}}}{t^2}\,dt$

CHAPTER 8

Applications of the integral

The integral appears as the answer to many different questions. In this chapter we will describe a number of "things that are an integral". In each example, there is a quantity we want to compute 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.

One could read this chapter as a list of formulas to compute areas, volumes, lengths and other things, but that would be a mistake. In each case the important thing to learn is the reasoning that tells us how to turn a question about areas, volumes, etc into an integral.

1. Areas between graphs

1.1. The problem. Suppose you have two functions f and g on an interval $[a, b]$ such that $f(x) \leq g(x)$ for all x in the interval $[a, b]$. Then the area of the region between the graphs of the two functions is

$$\text{Area} = \int_a^b \big(g(x) - f(x)\big)dx. \tag{88}$$

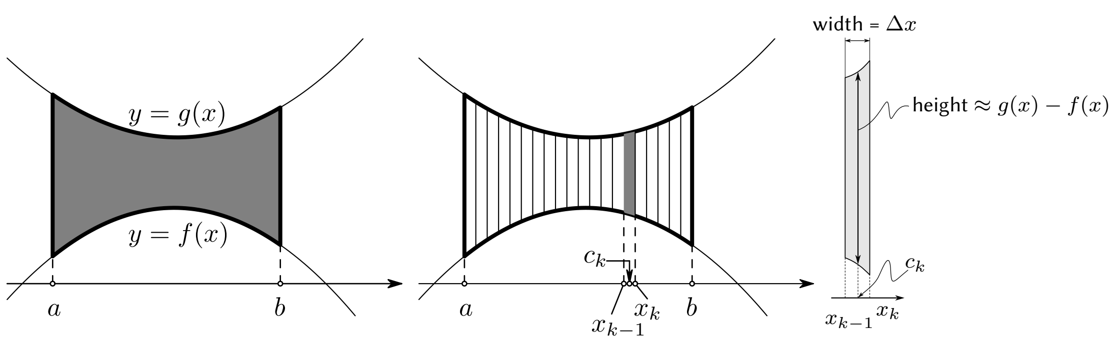


Figure 1. The area between two graphs. To compute the area on the left you slice it into many thin vertical strips. Each strip is approximately a rectangle so its area is "height×width."

1.2. Derivation using Riemann sums. To get this formula, we 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 $[a, b]$ and choose a number c_k in each interval $[x_{k-1}, x_k]$. Then we form the rectangles

$$x_{k-1} \leq x \leq x_k, \qquad f(c_k) \leq y \leq g(c_k).$$

The area of this rectangle is

$$\text{width} \times \text{height} = \Delta x_k \cdot \big(g(c_k) - f(c_k)\big).$$

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 on the right hand side in (88). Therefore,

- 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,
- and since the Riemann sums R converge to the integral I,

we conclude that the area between the graphs of f and g is exactly the integral in (88).

1.3. Summary of the derivation. Since most applications of the integral follow the pattern in the derivation above, it is important to understand its general features. Here we can summarize the above calculation using Riemann sums by the following (rather sketchy) argument: if we slice up the area between the two curves $y = f(x)$ and $y = g(x)$ into n strips all of width $\Delta x = \dfrac{b-a}{n}$, then if we write the corresponding Riemann sum with Σ notation, we have

$$R = \sum_{j=1}^{n} \big(g(c_j) - f(c_j)\big)\Delta x.$$

If we abuse the notation slightly and write the indices of summation as values of x, this becomes

$$R = \sum_{\text{“}x\,=\,a\text{”}}^{\text{“}x\,=\,b\text{”}} \big(g(x_j) - f(x_j)\big)\Delta x.$$

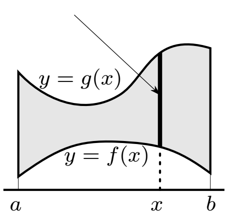
Now, if we take the limit as $n \to \infty$, then the result is exactly the integral

$$I = \int_{x=a}^{x=b} \big(g(x) - f(x)\big)\, dx.$$

Essentially, the idea is that as we take the limit, the sum becomes an integral and the “Δ”s become “d”s.

Infinitely thin strip with:

height $= g(x) - f(x)$
infinitely small width $= dx$
infinitely small area
$= (g(x) - f(x)) \times dx$



1.4. Leibniz’ derivation. Many scientists and engineers think of integrals in terms of Leibniz’s “sums of infinitely many infinitely small quantities,” and will prefer the following description of the derivation of the area formula (88)

> *…to find the area of the region in the drawing on the right, divide this region into infinitely many infinitely thin strips. Each strip has height $g(x) - f(x)$, where x is the x-coordinate of the strip, and its width is the infinitely small number dx. The area of one of these strips is therefore*
>
> $$\text{\textit{area of one infinitely thin strip}} = \big(g(x) - f(x)\big) \times dx.$$
>
> *To get the total area add the areas of all these infinitely thin strips together, which we (i.e. Leibniz) write as follows:*
>
> $$\text{\textit{Area}} = \int_{x=a}^{b} \big(g(x) - f(x)\big)dx.$$

The "long s" indicates that Leibniz is summing a lot of numbers. This is of course a very strange kind of sum, since the "numbers" he is adding are infinitely small, and since there are infinitely many of them (there is one term for each value of x between a and b!). Because there are infinitely many terms in the sum you can't write it out using "$+$"signs, and for this reason Leibniz invented the notation we use for integration today. The integral sign is nothing but an elongated "S" and its shape should remind us that, at least in Leibniz' interpretation, an integral comes about by adding infinitely many infinitely small numbers.

2. Problems

1. Find the area of the finite region bounded by the parabola $y^2 = 4x$ and the line $y = 2x$.

2. Find the area bounded by the curve $y = x(2-x)$ and the line $x = 2y$.

3. Find the area bounded by the curve $x^2 = 4y$ and the line $x = 4y - 2$.

4. Calculate the area of the finite region bounded by the parabolas $y = x^2$ and $x = y^2$.

5. Find the area of the finite region included between the parabola $y^2 = x$ and the line $x + y = 2$.

6. Find the area of the finite region bounded by the curves $y = \sqrt{x}$ and $y = x$.

7. Use integration to find the area of the triangular region bounded by the lines $y = 2x + 1$, $y = 3x + 1$ and $x = 4$.

8. Find the finite area bounded by the parabola $x^2 - 2 = y$ and the line $x + y = 0$.

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 $y = f(x)$ and below the graph of $y = g(x)$. (Hint: if you need to integrate $1/(2 + x^2)$ you could substitute $u = x/\sqrt{2}$.)

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.

11. Find the area of the finite region between the parabolas $y^2 = x$ and $x^2 = 16y$.

12. Find the area of the finite region enclosed by the parabola $y^2 = 4ax$ and the line $y = mx$, where a and m are positive constants.

13. Find a so that the curves $y = x^2$ and $y = a\cos x$ intersect at the point $(x, y) = (\frac{\pi}{4}, \frac{\pi^2}{16})$. Then find the finite area between these curves.

14. **[Group Problem]** 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. [Hint: The area can be found using geometry, using the observation that the part of a circle cut off by a line is a circular sector with a triangle removed.]

3. Cavalieri's principle and volumes of solids

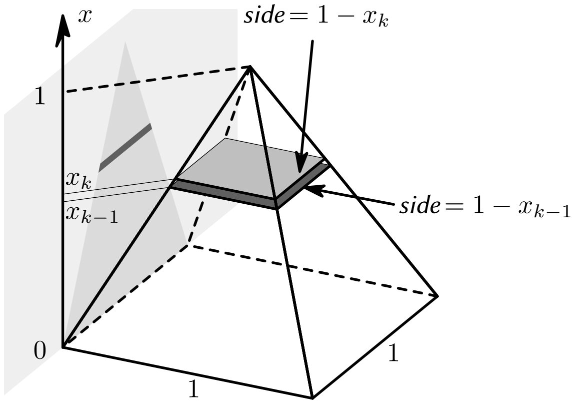
In this section we'll discuss the "method of slicing" and use integration to derive the formulas for volumes of spheres, cylinders, cones, and other solid objects in a systematic way.

3.1. Example – Volume of a pyramid.

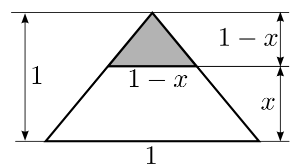
As an example let's compute the volume of a pyramid whose base is a square of side 1, and whose height is 1. Our strategy will be to divide the pyramid into thin horizontal slices whose volumes we can compute, and to add the volumes of the slices to get the volume of the pyramid.

To construct the slices we choose a partition of the (height) interval $[0, 1]$ into N subintervals, i.e., we pick numbers

$$0 = x_0 < x_1 < x_2 < \cdots < x_{N-1} < x_N = 1,$$

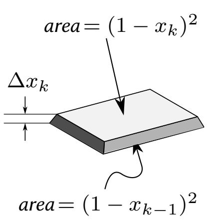


Side view of the pyramid



A side view of the pyramid shows a triangle with base 1 *and height* 1*. The slice at height* x *cuts off a smaller triangle at the top. The small and large triangles are similar.*
The height of the small triangle is $1 - x$ *times the height of the large triangle, so the base of the small triangle is* $1-x$ *times the base of the large triangle.*

Figure 2. Volume of a pyramid.



and as usual we set $\Delta x_k = x_k - x_{k-1}$.

The kth slice consists of those points on the pyramid whose height is between x_{k-1} and x_k. The top of this slice is a square with side $1 - x_k$ and the bottom is a square with side $1 - x_{k-1}$ (see Figure 2). The height, or thickness, of the slice is $x_k - x_{k-1} = \Delta x_k$.

If you ignore the fact that the sides of the kth slice are slanted and not vertical, then the volume of the kth slice is the product of the area of its top and its thickness, i.e.

$$\text{volume of } k\text{th slice} \approx (1 - x_k)^2 \Delta x_k.$$

On the other hand, you could also estimate the volume of the slice by multiplying the thickness and the area of the bottom instead of the top. This leads to

$$\text{volume of } k\text{th slice} \approx (1 - x_{k-1})^2 \Delta x_k.$$

Neither formula is exactly right and we might as well have chosen any c_k in the interval $[x_{k-1}, x_k]$ and estimated

$$\text{volume of } k\text{th slice} \approx (1 - c_k)^2 \Delta x_k.$$

The idea is that as we make the partition finer, each of these approximations for the volume of a slice will get better, and that it won't matter which of them we use.

Adding the volumes of the slices, we find that the volume V of the pyramid is given by

$$V \approx (1 - c_1)^2 \Delta x_1 + \cdots + (1 - c_N)^2 \Delta x_N,$$

where we expect the approximation to get better as we choose more and more partition points. 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

$$I = \lim_{\cdots} \left\{ (1 - c_1)^2 \Delta x_1 + \cdots + (1 - c_N)^2 \Delta x_N \right\} = V.$$

Computing the integral then yields that the volume of the pyramid is

$$V = \frac{1}{3}.$$

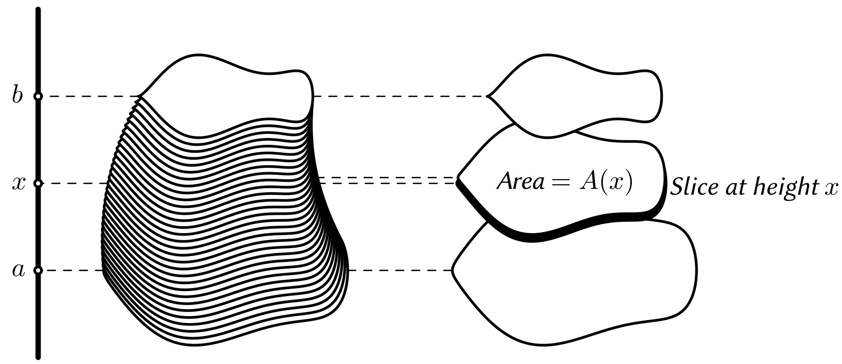


Figure 3. Slicing a solid to compute its volume. The volume of one slice is approximately the product of its thickness (Δx) and the area $A(x)$ of its top. Summing the volume $A(x)\Delta x$ over all slices leads approximately to the integral $\int_a^b A(x)dx$ (see (90).)

3.2. Computing volumes by slicing – the 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, computing the volume of each slice, summing the volumes of the slices, and then recognizing these sums as Riemann sums for some integral. That integral then is the volume of the solid.

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 kth 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 kth 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

$$V = \lim_{\cdots}\big(A(c_1)\Delta x_1 + \cdots + A(c_N)\Delta x_N\big). \tag{89}$$

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

$$V = \int_a^b A(x)dx. \tag{90}$$

In summary, to compute the volume of a solid, we pick an axis, slice up the solid perpendicular to that axis, and then integrate (along the axis) the cross-sectional areas of each of the slices.

3.3. Cavalieri's principle. The formula (90) 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 then the volume of the pile will still be the sum of the volumes of the coins, meaning that it doesn't change.

3.4. Solids of revolution. In principle, formula (90) 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.***

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 that is defined on an interval $[a, b]$ and that 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 $a \le x \le b$, 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 with centers $(x, 0)$ that 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 (90) 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

$$V = \pi \int_a^b r^2 \, dx = \pi \int_a^b f(x)^2 \, dx. \tag{91}$$

3.5. Volume of a sphere. As an example we compute the volume of a sphere with radius R. You can get that sphere by revolving a semicircle with radius R around the

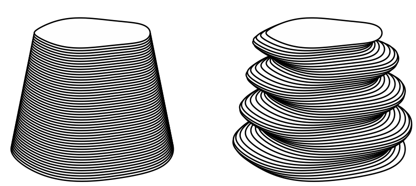


Figure 4. Cavalieri's principle. Both solids consist of a pile of horizontal slices. The solid on the right was obtained from the solid on the left by sliding some of the slices to the left and others to the right. This operation does not affect the volumes of the slices, and hence both solids have the same volume.

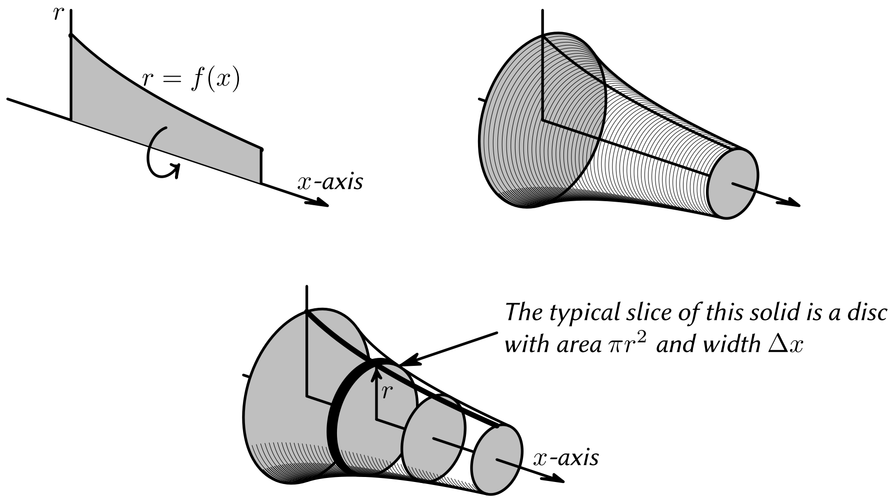


Figure 5. A solid of revolution consists of all points in three-dimensional space whose distance r to the x-axis satisfies $r \le f(x)$.

x-axis: this gives $r(x) = \sqrt{R^2 - x^2}$. Here is a drawing (don't get confused by the fact that the x-axis is vertical in this picture):

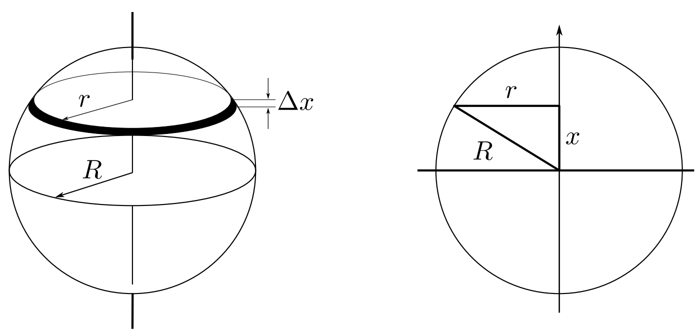


Figure 6. On the left: a sphere of radius R with a slice of thickness Δx at height x. **On the right:** a cross section which allows you to compute how the radius r depends on x.

Since $r = \sqrt{R^2 - x^2}$ we see that the area of a slice at height x is

$$A(x) = \pi r^2 = \left(\sqrt{R^2 - x^2}\right)^2 = \pi\left(R^2 - x^2\right).$$

Furthermore, we only get slices if $-R \le x \le +R$, so the x coordinate ranges between $-R$ and $+R$. The volume is therefore

$$V = \pi \int_{-R}^{R} \left(R^2 - x^2\right) dx = \pi\left[R^2 x - \tfrac{1}{3}x^3\right]_{-R}^{R} = \tfrac{4}{3}\pi R^3,$$

which is the standard formula we have seen before.

3.6. Volume of a solid with a core removed – the method of "washers." Another class of solids whose volume you can compute by slicing arises when you have one solid of revolution, and remove a smaller solid of revolution. Here we'll derive the formula for the volume.

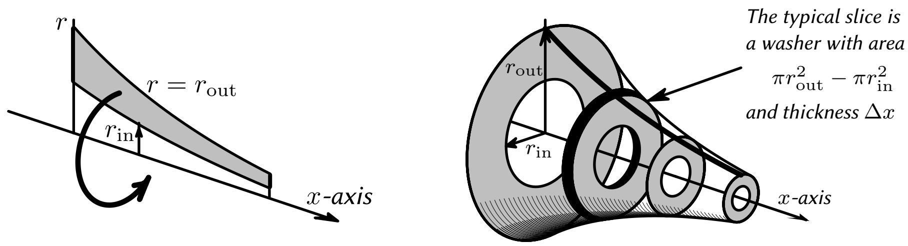


Figure 7. The method of washers. If you remove one smaller solid of revolution from a larger solid of revolution you get the region that consists of all points in three-dimensional space whose distance r to the x-axis satisfies $r_{\text{in}} \le r \le r_{\text{out}}$. A thin slice of this solid looks like a washer.

If the two solids have the same axis of rotation, and if the outer solid has radius $r = r_{\text{out}} = f(x)$, while the inner solid has radius $r = r_{\text{in}} = g(x)$, then the slice you get by intersecting the solid with a plane perpendicular to the x-axis looks like a "washer," or an ***annulus*** (ring shaped region). The area of each "infinitely thin" slice is the difference between the areas enclosed by the outer circle and the inner circle, thus,

$$A(x) = \pi r_{\text{out}}(x)^2 - \pi r_{\text{in}}(x)^2.$$

The volume of such a slice is the product of its area $A(x)$ and its infinitely small thickness dx. "Adding" the volumes of the thin washers together we arrive at

$$V = \pi \int_a^b \left(r_{\text{out}}(x)^2 - r_{\text{in}}(x)^2\right) dx. \tag{92}$$

4. Three examples of volume computations of solids of revolution

4.1. Problem 1: Revolve $\mathcal{R}$ around the y-axis. Consider the solid obtained by revolving the region

$$\mathcal{R} = \left\{(x, y) \mid 0 \le x \le 2, \quad (x-1)^2 \le y \le 1\right\}$$

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 $y = 1$.

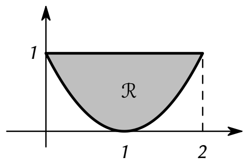
If we intersect the solid with a plane at height y then we get a ring shaped region, or "annulus", i.e. a large disc with a smaller disc removed. You can see it in the figure below: if you cut the region $\mathcal{R}$ horizontally at height y you get the line segment AB, and if you rotate this segment 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

$$y = (1-r)^2$$

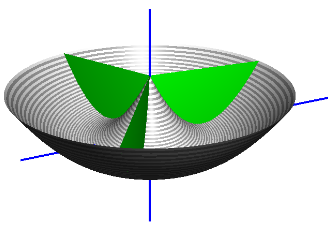
so they are

$$r_{\text{in}} = 1 - \sqrt{y}, \qquad r_{\text{out}} = 1 + \sqrt{y}.$$

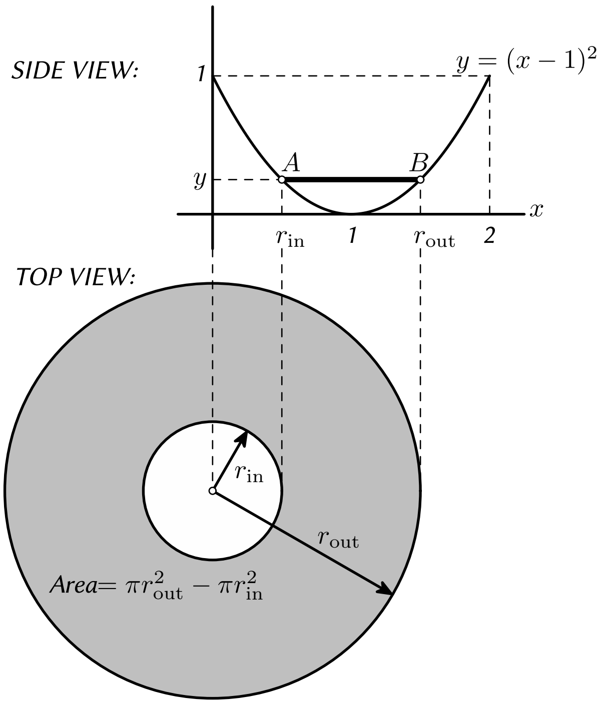
Problem 1: Computing the volume of the solid you get when you revolve this region around the y-axis.



The solid you get, with its top removed so you can look inside, and with a few cross sections (rotated copies of the region $\mathcal{R}$) looks like this:



On the right you see the drawing that is used in the computation of the volume. At the top is the cross section again. When you rotate the line segment AB around the y-axis it sweeps out a "washer," pictured below.



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

$$V = \int_0^1 A(y)dy = 4\pi \int_0^1 \sqrt{y}\, dy = 4\pi \cdot \tfrac{2}{3} = \frac{8\pi}{3}.$$

4.2. 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 $x = -1$.

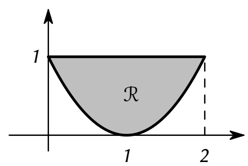
Solution: The line $x = -1$ is vertical, so we slice the solid with horizontal planes. The height of each plane will be called y.

As before the slices are ring shaped regions ("washers") but the inner and outer radii are now given by

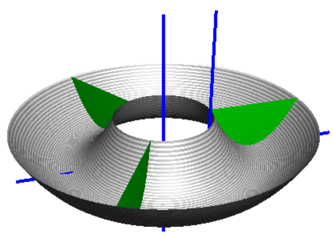
$$r_{\text{in}} = 1 + x_{\text{in}} = 2 - \sqrt{y}, \qquad r_{\text{out}} = 1 + x_{\text{out}} = 2 + \sqrt{y}.$$

The volume is therefore given by

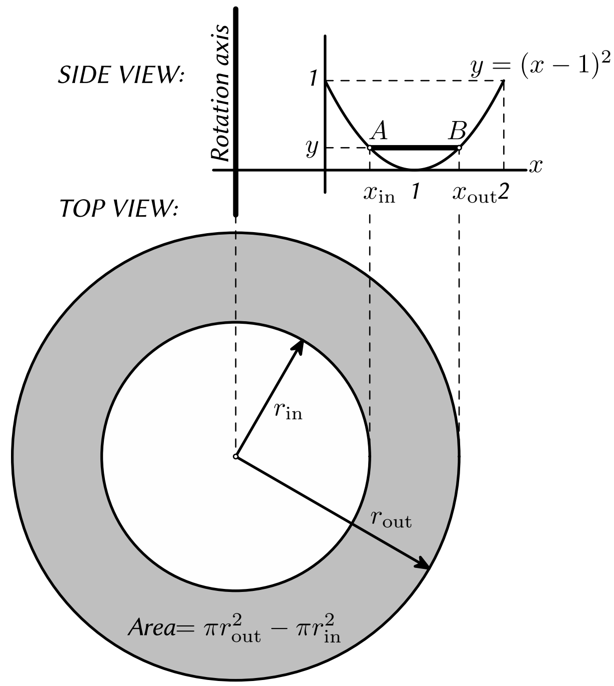
$$V = \int_0^1 \big(\pi r_{\text{out}}^2 - \pi r_{\text{in}}^2\big)dy = \pi \int_0^1 8\sqrt{y}\, dy = \frac{16\pi}{3}.$$



Problem 2: In this problem we revolve the region $\mathcal{R}$ around the line $x = -1$ instead of around the y-axis.
The solid you get, again with its top removed so you can look inside, and with a few cross sections (rotated copies of the region $\mathcal{R}$) looks like this:



On the right you see the drawing that is used in the computation of the volume, with the cross section again at the. Rotating the line segment AB around the line $x = 1$ produces the "washer," pictured below.



4.3. 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.

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

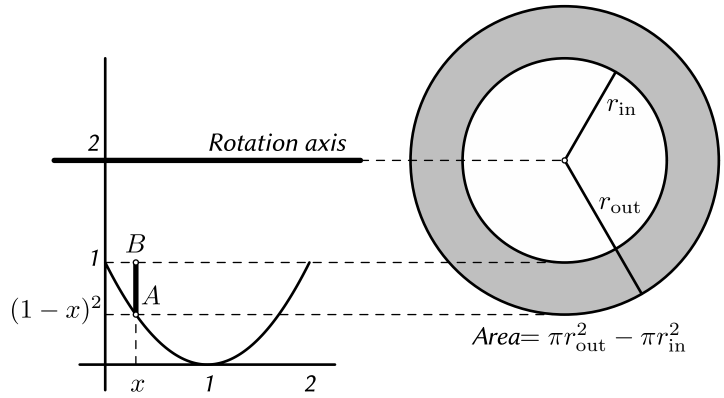
$$r_{\text{in}} = 1, \qquad r_{\text{out}} = 2 - (1-x)^2.$$

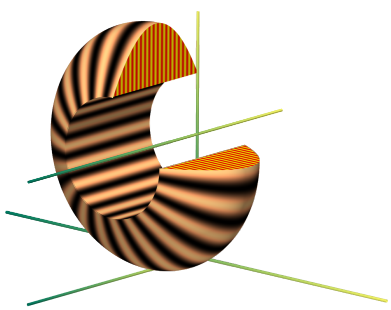
The area of the slice is therefore

$$A(x) = \pi\big(2 - (1-x)^2\big)^2 - \pi \cdot 1^2 = \pi\left(3 - 4(1-x)^2 + (1-x)^4\right).$$

The x values which occur in the solid are $0 \le x \le 2$, and so its volume is

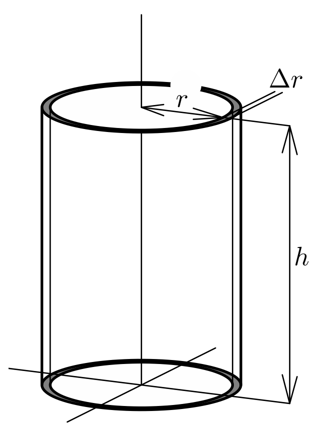
$$\begin{aligned} V &= \pi \int_0^2 \left(3 - 4(1-x)^2 + (1-x)^4\right) dx \\ &= \pi \left[3x + \tfrac{4}{3}(1-x)^3 - \tfrac{1}{5}(1-x)^5\right|_0^2 \\ &= \tfrac{56}{15}\pi \end{aligned}$$





Problem 3: We revolve the same region $\mathcal{R}$ as in the previous two problems about the line $y = 2$. The solid you get is shown on the left (cut open to show that the cross section is the region $\mathcal{R}$).
Above is the drawing that is used to find the inner and outer radii of the washers you get when you rotate the line segment AB around the horizontal line $y = 2$.

5. Volumes by cylindrical shells



Finding the volume of a thin cylindrical shell

Instead of slicing a solid with parallel planes you can also compute its volume by splitting it into thin cylindrical shells and adding the volumes of those shells. To find the volume of a thin cylindrical shell, recall that the volume of a cylinder of height h and radius r is $\pi r^2 h$ (height times the area of the 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

$$\mathcal{R} = \{(r, h) \mid a \le r \le b, 0 \le h \le f(r)\}$$

around the h-axis (instead of using x and y we put r on the horizontal axis and h on the vertical axis as in Figure 8). By partitioning the interval $a \le r \le b$ into many small intervals we can decompose the solid into many thin shells. The volume of each shell will approximately be given by $2\pi r f(r)\Delta r$. 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

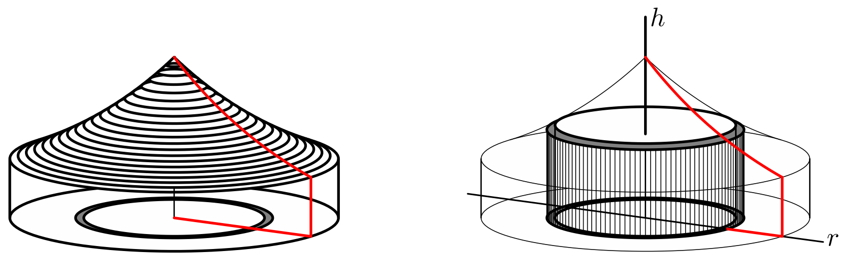


Figure 8. Computing the volume under a circus tent using cylindrical shells. This particular tent is obtained by rotating the graph of $y = e^{-x}$, $0 \le x \le 1$ around the y-axis.

of revolution:

$$V = 2\pi \int_a^b r f(r)\, dr. \tag{93}$$

If the region $\mathcal{R}$ is not the region under the graph, but rather the region between the graphs of two functions $f(r) \le g(r)$, then we get

$$V = 2\pi \int_a^b r\big(g(r) - f(r)\big)\, dr.$$

5.1. Example – The solid obtained by rotating $\mathcal{R}$ about the y-axis, again. The region $\mathcal{R}$ from §4.1 can also be described as

$$\mathcal{R} = \big\{(x, y) \mid 0 \le x \le 2, f(x) \le y \le g(x)\big\},$$

where

$$f(x) = (x-1)^2 \text{ and } g(x) = 1.$$

The volume of the solid which we already computed in §4.1 is thus given by

$$\begin{aligned} V &= 2\pi \int_0^1 x\big\{1 - (x-1)^2\big\}\, dx \\ &= 2\pi \int_0^2 \big\{-x^3 + 2x^2\big\}\, dx \\ &= 2\pi \big[-\tfrac{1}{4}x^4 + \tfrac{2}{3}x^3\big]_0^2 \\ &= 8\pi/3, \end{aligned}$$

which coincides with the answer we found in §4.1.

6. Problems

1. [Group Problem] What do the dots in "lim…" in equation (89) stand for? (In other words, what approaches what in this limit?)

♣

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.

♡

Find the volume of the solids you get by rotating each of the following graphs around the x-axis:

3. $f(x) = x, 0 \le x \le 2$

4. $f(x) = \sqrt{2-x}, 0 \le x \le 2$

5. $f(x) = \left(1+x^2\right)^{-1/2}, |x| \le 1$

6. $f(x) = \sin x, 0 \le x \le \pi$

7. $f(x) = 1 - x^2, |x| \le 1$

8. $f(x) = \cos x, 0 \le x \le \pi$ [Note that this function crosses the x-axis: is this a problem?]

9. $f(x) = 1/\cos x, 0 \le x \le \pi/4$

10. Find the volume that results by rotating the semicircle $y = \sqrt{R^2 - x^2}$ about the x-axis, for $-R \le x \le R$.

11. Let $\mathcal{T}$ be the triangle $1 \le x \le 2, 0 \le y \le 3x - 3$.

Find the volume of the solid obtained by …

(a) …rotating $\mathcal{T}$ around the x-axis.

(b) …rotating $\mathcal{T}$ around the y axis.

(c) …rotating $\mathcal{T}$ around the line $x = -1$.

(d) …rotating $\mathcal{T}$ around the line $y = -1$.

12. A spherical bowl of radius a contains water to a depth $h < 2a$.

(a) Find the volume of the water in the bowl in terms of h and a. (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?

13. What we have called Cavalieri's principle was already known in some form to Archimedes. He used this idea to compute the volumes of various solids. One of his more famous discoveries is depicted below in Figure 9. He found that the ratios of the volumes of a cone of height and radius R, half a sphere of radius R, and a cylinder of height and radius R are given by this very simple rule:

$$\textbf{cone} : \textbf{sphere} : \textbf{cylinder} = \mathbf{1 : 2 : 3}$$

In other words, the half sphere has twice the volume of the cone, and the cylinder has three times the volume of the cone.

(a) Prove Archimedes is right by computing the volumes of these solids of revolution.

This computation will give you practice in setting up and doing integrals, which you compute by using a fair amount of algebra[1].

(b) Show that the volume of the cone plus the volume of the half-sphere is exactly the volume of the cylinder by comparing the areas of slices and using Cavalieri's principle (and thus without doing any integrals).

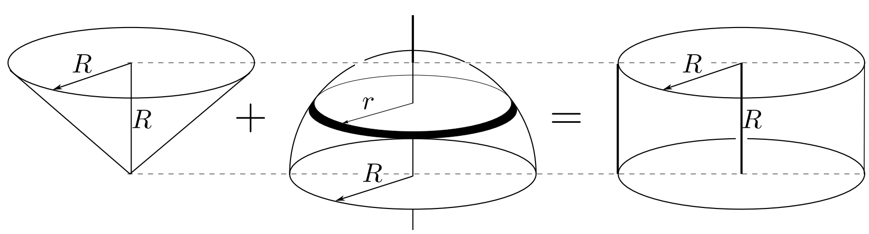


Figure 9. Archimedes' volume computation. See problem 6.13.

7. Distance from velocity

7.1. 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

[1]In the time of Archimedes there were no integrals nor even algebra, so we might wonder: how did Archimedes do this? The answer is: a combination of his "lever principle" and a complicated geometric argument by contradiction showing that a cone has one-third the area of the cylinder with the same base and height

is an antiderivative of the velocity, and the fundamental theorem of calculus says that

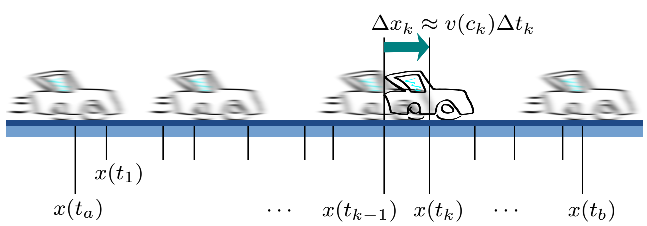
$$\int_{t_a}^{t_b} v(t)\;dt = x(t_b) - x(t_a), \tag{94}$$

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 by the object (during the interval of integration).

You can also get equation (94) by using Riemann sums. Namely, to see how far the



object moved between times t_a and t_b you 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 you can estimate the velocity by $v(c_k)$ where c_k is some number between t_{k-1} and t_k. You then have

$$\Delta s_k \approx v(c_k)\Delta t_k$$

Here the approximation will be more accurate as you make the Δt_k's smaller. The total distance travelled is the sum of the travel distances for all time intervals $t_{k-1} < t < t_k$, i.e.

$$\text{Distance travelled} \approx \Delta s_1 + \cdots + \Delta s_N \approx v(c_1)\Delta t_1 + \cdots + v(c_N)\Delta t_N.$$

The right hand side is again a Riemann sum for the integral in (94). As you make the partition finer and finer you therefore get

$$\text{Distance travelled} = \int_{t_a}^{t_b} v(t)\;dt.$$

Leibniz would have said it even shorter:

> *. . . to see how far the object travelled between time t_a and time t_b, I, Leibniz, divide the interval $t_a < t < t_b$ into infinitely many intervals, each of length dt. This length dt has to be infinitely small. During the infinitely short time interval from t to $t + dt$ the velocity is constant, so the distance travelled is $v(t)\,dt$. When I um all these infinitely short displacements over all the time intervals I get*
>
> $$\textit{Ditance travelled} = {}_{t=t_a}^{t_b}\, v(t)\;dt.$$

Again, we have to wonder what "adding infinitely many infinitely small quantities" actually means. Leibniz never explained that. Fortunately the derivation with Riemann sums leads to the same answer.

7.2. 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 (94), 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

$$\boldsymbol{x(t) = x(0) + \int_0^t v(t)\, dt} \quad \longleftarrow \quad \textbf{BAD FORMULA}$$

because the variable t gets used with two different meanings: 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 Chapter III, §10 and Chapter VIII, §3.3). 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

$$x(t) = x(0) + \int_0^t v(\bar{t})\, d\bar{t}.$$

7.3. Motion along a curve in the plane. If an object is moving in the plane, and if its position is given by

$$x = x(t), \quad y = y(t)$$

then we can also compute the distance travelled during any time interval. The derivation is very similar to the discussion of the velocity of a parametric curve in Chapter V, §15.

To find the distance travelled between $t = t_a$ and $t = t_b$, divide the time interval into many short intervals

$$t_a = t_0 < t_1 < \cdots < t_N = t_b.$$

You then get a sequence of points

$$P_0(x(t_0), y(t_0)),\ P_1(x(t_1), y(t_1)),\ \ldots,\ P_N(x(t_N), y(t_N)),$$

and after "connecting the dots" you get a polygon. You could approximate the distance

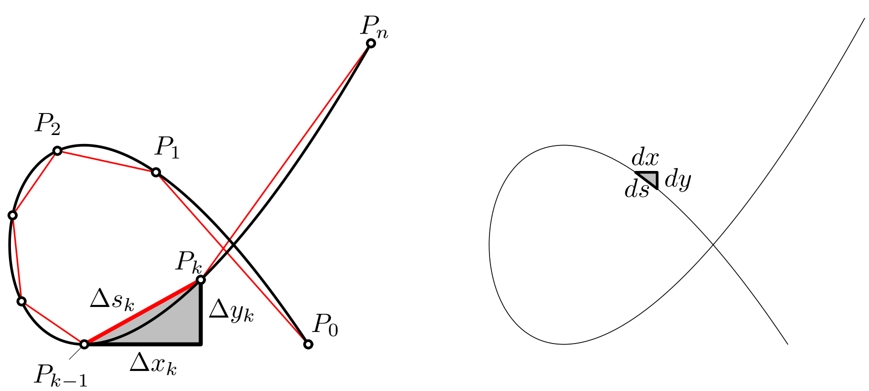


Figure 10. The distance travelled. On the left the picture describing the derivation with Riemann sums, on the right Leibniz' version of the picture.

travelled by finding the length of this polygon. The distance between two consecutive points P_{k-1} and P_k is

$$\begin{aligned}\Delta s_k &= \sqrt{(\Delta x_k)^2 + (\Delta y_k)^2}\\ &= \sqrt{\Big(\frac{\Delta x_k}{\Delta t_k}\Big)^2 + \Big(\frac{\Delta y_k}{\Delta t_k}\Big)^2}\ \ \Delta t_k\\ &\approx \sqrt{x'(c_k)^2 + y'(c_k)^2}\ \Delta t_k\end{aligned}$$

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 sample time 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 that the distance travelled is

$$s = \int_{t_a}^{t_b} \sqrt{x'(t)^2 + y'(t)^2}\ dt. \tag{95}$$

As usual, Leibniz has a shorter version of this derivation: *during an infinitely small time interval of length dt, the coordinates of the point change by dx and dy, respectively. The (infinitely short) distance ds travelled during this time interval follows from Pythagoras (see Figure 10), so the point moves a distance*

$$ds = \sqrt{(dx)^2 + (dy)^2} = \sqrt{\Big(\frac{dx}{dt}\Big)^2 + \Big(\frac{dy}{dt}\Big)^2}\ dt.$$

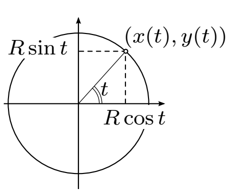
The total distance travelled is the sum of all these infinitely small displacements:

$$s = \int_{t_a}^{t_b} ds = \int_{t_a}^{t_b} \sqrt{\Big(\frac{dx}{dt}\Big)^2 + \Big(\frac{dy}{dt}\Big)^2}\ dt.$$

Of course, you have to believe in infinitely small numbers when you use this argument.

8. The arclength of a curve

We introduced the integral as a tool to compute areas of plane regions. You can also use it to compute arclengths of plane curves. In fact, if the curve is a parametric curve, given by $x = x(t), y = y(t)$ then we already have the formula. The arclength of the curve traced out by the point $\big(x(t), y(t)\big)$ as the parameter t varies from t_a to t_b is nothing but the distance travelled by the point. We can use (95) to find the length. In this section we'll do three examples of length computations. One point that these examples will show is that the formula (95) is a good source of impossible or very difficult integrals.



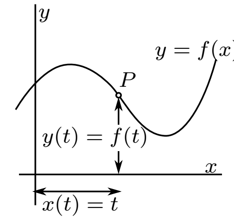
8.1. The circumference of a circle. You can parametrize the circle with radius R by

$$x(t) = R\cos t, \quad y(t) = R\sin t, \quad (0 \le t \le 2\pi)$$

The velocity is $v(t) = \sqrt{x'(t)^2 + y'(t)^2} = R$. Therefore (95) tells us that the length of the circle is

$$L = \int_0^{2\pi} \sqrt{x'(t)^2 + y'(t)^2}\ dt = \int_0^{2\pi} R\ dt = 2\pi R.$$

The length of a circle is $2\pi R$ – we knew that. This cannot be a PROOF that the circle has length $2\pi R$ 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 (95) is at least consistent with what we already knew.

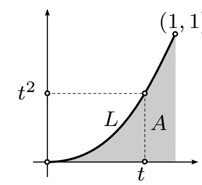


8.2. The length of the graph of a function $y = f(x)$. In Chapter V, § 15.3, we saw that you can parametrize the graph of a function $y = f(x)$, by

$$x(t) = t, \qquad y(t) = f(t).$$

For such a parametric curve you have $x'(t) = 1$ and $y'(t) = t$, so the length of the segment with $a \le t \le b$ is $\int_a^b \sqrt{1 + f'(t)^2} dt$. In that integral t is a dummy variable and we can replace it with x, which leads to a formula for the length of the graph of a function $y = f(x)$:

$$L = \int_a^b \sqrt{1 + f'(x)^2}\, dx \tag{96}$$



Area: $1/3$
Length:
$\frac{1}{2}\sqrt{5} + \frac{1}{4}\ln\bigl(2 + \sqrt{5}\bigr).$

8.3. Arclength 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 ($\frac{1}{3}$), its arclength turns out to be much more complicated. Our length formula (96) says that the arclength of the parabola is given by

$$L = \int_0^1 \sqrt{1 + \bigl(\frac{dx^2}{dx}\bigr)^2}\, dx = \int_0^1 \sqrt{1 + 4x^2}\, dx. \tag{97}$$

To find this integral you would have to use one of the following (not at all obvious) substitutions[2]

$$x = \tfrac{1}{4}\left(z - \frac{1}{z}\right) \qquad \text{(then } 1 + 4t^2 = \frac{1}{4}(z + 1/z)^2 \text{ so you can simplify the } \sqrt{\cdots}\text{)} \tag{98}$$

or (if you like hyperbolic functions)

$$x = \tfrac{1}{2}\sinh w \qquad \text{(in which case } \sqrt{1 + 4x^2} = \cosh w.\text{)}$$

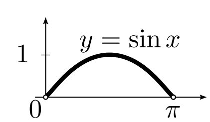
The answer is :

$$L = \tfrac{1}{2}\sqrt{5} + \tfrac{1}{4}\ln\bigl(2 + \sqrt{5}\bigr) \approx 1.47894\cdots$$

The computation is left as a challenging exercise (Problem 9.8).

8.4. Arclength 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:

$$L = \int_0^\pi \sqrt{1 + \left(\frac{d\sin x}{dx}\right)^2}\, dx = \int_0^\pi \sqrt{1 + \cos^2 x}\, dx. \tag{99}$$



Unfortunately this is not an integral that 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 arclength of a curve. In spite of the fact that we get stuck when we try to compute the integral in (99), 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

$$\int_0^\pi 1 dx \;\le\; \int_0^\pi \sqrt{1 + \cos^2 x}\, dx \;\le\; \int_0^\pi \sqrt{2}\, dx,$$

[2] Many calculus textbooks will tell you to substitute $x = \frac{1}{2}\tan\theta$, but the resulting integral is still not easy.

i.e.

$$\pi \le L \le \pi\sqrt{2}.$$

9. Problems

1. The velocity of a particle moving along the x-axis is

$$v(t) = -(t-1)^2 + 2$$

for $0 \le t \le 6$. If at time $t = 0$ the particle is at the origin, find the location of the particle at time $t = 3$.

2. A snail travels at a constant speed of 0.001 units per hour on a straight path from point $A(1, 1)$ to point $B(2, 3)$.

(a) Find a parametrization of the snail's motion and use it to compute the distance traveled by the snail.

(b) Is there an easier way to find the distance travelled, and does it lead to the same result?

3. Find the length of the piece of the graph of $y = \sqrt{1-x^2}$ where $0 \le x \le \frac{1}{2}$.

The graph is a circle, so there are two ways of computing this length. One uses geometry (length of a circular arc = radius times angle), the other uses an integral.

Use both methods and check that you get the same answer.

4. Compute the length of the part of the *evolute of the circle*, given by

$$x(t) = \cos t + t\sin t, \quad y(t) = \sin t - t\cos t$$

where $0 \le t \le \pi$. (This problem has a nice answer, but any small algebra error can lead you to an impossible integral.)

5. **[Group Problem]** Show that the *Archimedean spiral*, given by

$$x(\theta) = \theta\cos\theta,\ y(\theta) = \theta\sin\theta,\ 0 \le \theta \le \pi$$

has the same length as the parabola given by

$$y = \tfrac{1}{2}x^2, \quad 0 \le x \le \pi.$$

Hint: you can set up integrals for both lengths. If you get the same integral in both cases, then you know the two curves have the same length (even if you don't try to compute the integrals).

6. Three equivalent problems, pick one:

(a) Let $N(t)$ denote the size of a bacteria population at time t. The dynamics of the population is given by

$$\frac{dN}{dt} = 3e^{-2t}$$

with $N(0) = 25$. Express the change in population size between time 0 and time t as an integral. Compute the change in population size between $t = 0$ and $t = 10$.

(b) The amount $X(t)$ of oil spilled into the ocean waters by a faulty ship varies in time according to the rule

$$\frac{dX}{dt} = 3e^{-2t}.$$

Time is measured in hours. If at time $t = 0$ of the first measurement, the amount of oil in the water was already $X(0) = 25$ liters, how much oil was spilled in the water during the first 10 hours of measurements?

(c) A particle is moving along the x-axis with velocity $v(t) = 3e^{-2t}$. If at time $t = 0$ the particle was positioned at $x = 25$, what is the distance traveled by the particle during the time interval $0 \le t \le 10$?

7. A rope that hangs between two poles at $x = -\ln 2$ and $x = \ln 2$ takes the shape of a *Catenary*, with equation

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

(a) Find an integral for the length of the rope.

(b) Compute the integral you got in part **(a)**.

8. Find the length of the parabola L in (97) by using the substitution (98).

10. Velocity from acceleration

The acceleration of the object is by definition the rate of change of its velocity,

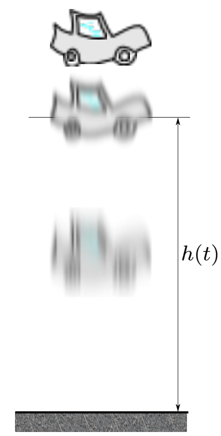
$$a(t) = v'(t),$$

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.*

10.1. 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, at sea level on Earth, it is about $9.8\text{m}/\text{sec}^2 \approx 32\text{ft}/\text{sec}^2$. If we designate 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:

$$\begin{aligned}
h(t) &= h(0) + \int_0^t v(\bar{t})\, d\bar{t} && \text{(Note the use of the dummy } \bar{t}\text{)}\\
&= h(0) + \int_0^t \big[v(0) - g\bar{t}\big]\, d\bar{t} && \text{(use } v(\bar{t}) = v(0) - g\bar{t}\text{)}\\
&= h(0) + \big[v(0)\bar{t} - \tfrac{1}{2}g\bar{t}^2\big]_{\bar{t}=0}^{\bar{t}=t}\\
&= h(0) + v(0)t - \tfrac{1}{2}gt^2.
\end{aligned}$$

For instance, if you launch the object upwards with velocity $5\text{ft}/\text{sec}$ from a height of 10ft, then you have

$$h(0) = 10\text{ft}, \quad v(0) = +5\text{ft}/\text{sec},$$

and thus

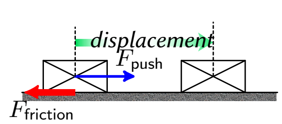
$$h(t) = 10 + 5t - 32t^2/2 = 10 + 5t - 16t^2.$$

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

$$h_{\max} = h(\tfrac{5}{32}) = 10 + \tfrac{25}{32} - \tfrac{25}{64} = 10\tfrac{25}{64}\text{ft}.$$

11. Work done by a force

11.1. Work as an integral. According to Newtonian mechanics any force that acts on an object in motion performs a certain amount of work, i.e. it transfers a certain amount of energy. (In Physics "work" is a form of energy.)



If the acting force is constant then the work done by this force is

$$\text{Work} = \text{Force} \cdot \text{Distance of displacement}.$$

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.

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* $\cdot$ *displacement* does not apply directly since it assumes that the force is constant. To compute the work done by the varying force $F(t)$ during some time interval $t_a < t < t_b$, we partition the time interval into lots of short intervals,

$$t_a = t_0 < t_1 < \cdots < t_{N-1} < t_N = t_b.$$

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 sample time $t_{k-1} \le c_k \le t_k$. 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

$$W = \int_{t_a}^{t_b} F(t)v(t)dt, \tag{100}$$

in which $v(t) = x'(t)$ is the velocity of the object.

11.2. Kinetic energy. Newton's famous law relating force and acceleration says

$$F = ma.$$

In this formula, F is the sum of all the forces exerted on some object and m is its mass, and a is the acceleration of the object. For instance, in the box example above, F is the sum of F_{push} and F_{friction}. Newton's law predicts that if the pushing force and the friction force don't cancel precisely, then its velocity must be changing.

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

$$W = \int_{t_a}^{t_b} F(t)v(t)dt = \int_{t_a}^{t_b} m\frac{dv(t)}{dt}\ v(t)\ dt. \tag{101}$$

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,

$$m\frac{dv(t)}{dt}\ v(t) = \frac{d}{dt}\left[\frac{1}{2}mv(t)^2\right].$$

(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 (101). 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.*

12. Problems

1. A ball is thrown upward from a height of 6 ft, with an initial velocity of 32 ft/sec. Is that enough speed for the ball to clear a fence 17 ft high?

2. A particle with mass m on a straight line (the x-axis) is being pushed and pulled by a force of magnitude $F(t) = F_0 \sin(\omega t)$. Here F_0 is the amplitude of the force, and ω is the frequency at which it oscillates. m, F_0, and ω are constants.

(a) What is the acceleration of the particle? If the initial velocity was v_0, and if the particle starts at $x = 0$, then where is the particle at time t?

(b) What effect does an increase in the frequency ω have on the velocity $v(t)$ of the particle?

3. Give a shorter derivation *a la* Leibniz of the integral for work in equation (100).

4. As an object falls, the force of gravitation transfers energy to the object. Let $h(t)$ be the height of the object at time t.

(a) Use the integral (100), i.e.

$$W = \int_{t_a}^{t_b} Fv\,dt$$

to show that the work done by gravitation during a time interval $t_a < t < t_b$ is

$$W = -mg\big(h_b - h_a\big),$$

where $h_a = h(t_a)$ and $h_b = h(t_b)$.

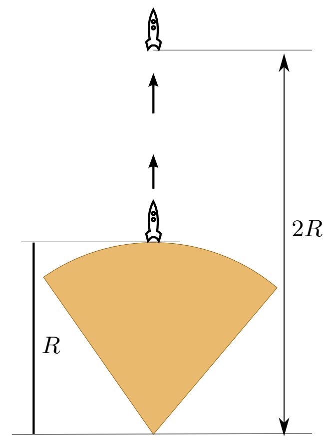
(Hint: $v(t) = \frac{dh}{dt}$.)

(b) If you throw a ball in the air, and catch it when it returns to the same height, how much work did the gravitational force do on the ball?

5. (*Rocket Science.*) A rocket is shot straight up into space. Its height, measured from the center of the Earth, is $r(t)$. The gravitational force on the rocket gets less as it gets further away from the Earth. According to another of Newton's laws the force is

$$F_{\mathsf{grav}} = -mg\frac{R^2}{r(t)^2},$$

where m is the mass of the rocket, $g = 9.81\mathrm{m}/\mathrm{sec}^2$, and R is the radius of the Earth.



If the rocket starts at radius R (the Earth's surface) and ends at radius $2R$, how much work did the gravitational force apply to the rocket? Use (100) again. In this problem $v = \frac{dr}{dt}$.