MATH1013 University Mathematics II
MATH1804 University Mathematics A

Dr. Ben Kane

Department of Mathematics
The University of Hong Kong

Second Semester 2013-2014

Date: April 25, 2014

1. Practical Information

- Instructor
  - Dr. Ben Kane
  - Email: bkane[at]maths.hku.hk
  - Office: Run Run Shaw Building A311
  - Consultation hours: Tuesdays 10:30–13:30
- Tutors:
  (1) Dr. Cheung Wai Shun
    - Email: cheungwaishun[at]gmail.com
    - Office: Run Run Shaw Building A316
    - Consultation hours: Tuesdays 13:30–15:30

  (2) Cheng Xiaoqing
    - Email: mechxqxiaofen[at]gmail.com
    - Office: Run Run Shaw Building A205
    - Consultation hours: Tuesdays 15:00–17:00

  (3) Lau Pan Shun
    - Email: panlau[at]hku.hk
    - Office: Run Run Shaw Building A212
    - Consultation hours: Wednesdays 10:30–12:30
- **Website:**
  http://www.hkumath.hku.hk/course/MATH1013,
  http://www.hkumath.hku.hk/course/MATH1804,
  Moodle Website:
  Section 2C: MATH1013_2C_2013, MATH1804_2C_2013,
  Section 2D: MATH1013_2D_2013, MATH1804_2D_2013.

Grade assessment

(1) The final exam is worth 50% of the final grade.
(2) There will be two midterm exams, each is worth 20%, for a total of 40% of the final grade.
(3) Homework assignments
Homework constitutes 10% of your final grade. For homework assignments, only the grades A, B, C, D, and F will be given by the tutors. The grades are cumulative throughout the semester, so the exact score will contribute to your final grade (not the letters A, B, C, D, and F, but the total scores for all assignments are kept in our records).
(4) The following guideline should be used to determine your grade:

Grade	Score
A^+	97%–100%
A	92%–97%
A^-	90%–92%
B^+	87%–90%
B	81%–87%
B^-	78%–81%
C^+	75%–78%
C	69%–75%
C^-	65%–69%
D	57%–65%
F	$< 57\%$.

Homework Assignments

(1) Please drop your work in the assignment box marked Math1013/1804 on the 4th floor Run Run Shaw Building.
(2) Homework is due weekly on Wednesdays and the assignment is to be turned in by 19:00 on the due date. No late work will be accepted.
(3) Please show your work! Answers without any of the steps shown will receive no credit.
(4) You are permitted (and even encouraged!) to discuss the homework problems with your classmates. However, you are responsible for your own work and each student is expected to write down the solutions in their own words! Photocopies of other students' solutions, combined solutions for multiple students, and plaguarized solutions will not be accepted.

Lectures

(1) Section 2C: Lectures will be held on Mondays from 9:30–11:20 and Thursdays from 9:30–10:20, except for periods of class suspension for public holidays, etc. Lecture will be held in Knowles 223.
(2) Section 2D: Lectures will be held on Tuesdays from 15:30–16:20 and Fridays from 15:30–17:20, except for periods of class suspension for public holidays, etc. Lecture will be held in Knowles 726.

Tutorials

There are 7 tutorial sessions available:

Group	Section	Tutor	Day of week	Time	Room
1C	2C	Cheng Xiaoqing	Mondays	16:30–17:20	MB122
2C	2C	Lau Pan Shun	Tuesdays	17:30–18:20	JLG04
3C	2C	Cheng Xiaoqing	Thursdays	15:30–16:20	MB103
4C	2C	Dr. Cheung Wai Shun	Thursdays	16:30–17:20	MB201
1D	2D	Dr. Cheung Wai Shun	Tuesdays	16:30–17:20	JLG04
2D	2D	Lau Pan Shun	Wednesdays	12:30–13:20	JLG04
3D	2D	Dr. Cheung Wai Shun	Thursdays	14:30–15:20	KB111

Abbrev.	Full building name
MB	Main Building
KB	Knowles Building
JLG	James Hsioung Lee Science Building

The tutorials will begin in the week of Feb. 10–14. There are 4 available timeslots for section 2C (listed as groups 1C, 2C, 3C, 4C) and 3 available timeslots for section 2D (groups 1D, 2D, and 3D). You should register for one of the available sections via Moodle. The deadline for registration is Monday, January 27, at 8:00.

Exams

There will be two midterm exams. The first midterm exam is scheduled for the week of Feb. 24–28 and the second will take place on April 4 and April 7 (for the two respective lectures). Calculators will not be permitted during the exams.

Course Description

You are expected to have a background in high school-level algebra, some experience with trigonometry. The course coveres basic calculus and some linear algebra.

More specifically: This course aims at students with Core Mathematics plus Module 1 or Core Mathematics plus Module 2 background and provides them with basic knowledge of calculus and some linear algebra that can be applied in various disciplines. It is expected to be followed by courses such as MATH2012 (Fundamental concepts of mathematics), MATH2101 (Linear Algebra I), MATH2102 (Linear Algebra II), MATH2211 (Multivariable calculus), and MATH2241 (Introduction to mathematical analysis).

Outline of Topics covered

The following will be covered in this course:

(1) Functions and their graphs
(2) Limits and continuity, the Intermediate Value Theorem
(3) Differentiation and the Mean Value Theorem
(4) Calculus with trigonometric functions
(5) Integration and partial fractions
(6) Complex numbers
(7) Basic matrix theory

(8) differential equations.

Course material

- Lecture notes: The lecture notes are the main reference material for this course. They may be downloaded from the website or through Moodle.
- Suggested reading:
  - A related textbook (with solutions to selected problems): George Thomas, Maurice Wei, and Joel Hass, Thomas' Calculus, 12th ed., Addison-Wesley, 2010.
  - Martin Anthony and Norman Biggs: Mathematics for Economics and Finance: Methods and Modelling (Cambridge University Press, 1996)
  - Adrian Banner: The Calculus Lifesaver: All the Tools You Need to Excel at Calculus (Princeton University Press, 2007)

Student responsibility

(1) It is your responsibility to attend lecture or to catch up on any missed material, should you be unable to attend. New material usually depends on earlier material, so you should be warned that missing one lecture may result in difficultly with subsequent lectures.
(2) It is highly recommended to attend tutorials and be active in the group discussion. Please come with questions that you are having about the lecture and/or the homework. We are here to help you! Moreover, students who attend tutorials perform noticably better on exams.
(3) When in doubt, please come to our consultation hours (listed above). Many times a one-on-one session can result in clearing up that small detail you couldn't quite understand the first time. Since material builds on itself, this small detail might have been holding you up on newer material, too.

2. Set Theory

A useful mathematical object is a set. One way to think of a set is to consider it a box holding certain objects. For any object, one can ask if the object is in the set/box or not in the set/box. Much in the same way that one would ask if your keys are in some box, one can ask whether the number 3 is in some set.

Example 2.1. For example, if you want to collect Bob, Jane, and Mary, then we put braces (the symbols { and }) around them to collect them together:

$$\{\text{Bob}, \text{Jane}, \text{Mary}\}\,. \tag{2.1}$$

One reads this as "The set of Bob, Jane, and Mary."

Definition 2.2. A <u>set</u> is a collection of objects with a given property.

The property is a rule used to determine whether an object is in the set or not. In Example 2.1, the property is that the object **is** Bob, Jane, or Mary.

To give a more complicated example, you could make a set of all children who are older than 7. Let's say that you have an object x and you want to know if it is in the set of all children who are older than 7. You would first ask whether x is a child. If not, then x is not in the set. If x is a child, then you would ask whether x is older than 7. If the answer is yes, then x is in the set, and otherwise it is not in the set.

Definition 2.3. One calls an object in the set an <u>element</u> of the set.

Let's now consider a mathematical way to depict the description above. Let's say that $P(x)$ is true if x is a child and $P(x)$ is false if x is not a child. Similarly, $Q(x)$ is true if x is older than 7 and $Q(x)$ is false otherwise. We could then denote the set of all children who are older than 7 by

$$\{\underbrace{x}_{\text{object}} : \underbrace{P(x) \text{ and } Q(x)}_{\text{property}}\}. \tag{2.2}$$

The symbol : is stands for "such that", so the above notation really means:

The set of x such that $P(x)$ and $Q(x)$ are both true,

which is in turn an abbreviation for

The set of all objects which are both children and older than 7.

You may consider this as an example of how mathematical notation can be used to abbreviate statements.

Sets written like equation (2.1) are given in <u>list notation.</u> This means that the property is determined by checking if the object is one of the explicitly-listed objects. It is often useful to move between the notation in equation (2.2) and an explicit list of the objects with this property. For example, if a friend asked you what languages you can speak in, they are really asking about the set

$$\{x : x \text{ is a language you can speak}\}\,. \tag{2.3}$$

Even though they may not know the answer, they are able to express the question by making a set with a property. Your answer will most likely not be in the form of equation (2.3), but rather something like

$$\{\text{Cantonese}, \text{English}, \text{German}, \text{Mandarin}\}\,. \tag{2.4}$$

Definition 2.4. One says that the sets in Equations (2.3) and (2.4) are equal because they have the same elements. We write (assuming that this is true)

$$\{x : x \text{ is a language you can speak}\} = \{\text{Cantonese}, \text{English}, \text{German}, \text{Mandarin}\}\,.$$

We use $=$ to denote equality and $:=$ to define an abbreviation for an object.

Let's try an example with mathematical content. Say

$$S := \{x : x \text{ is a positive integer and } x \leq 9\}\,. \tag{2.5}$$

This means that S abbreviates the set of all positive integers which are less than or equal to 9, and we say that S is defined to be the set of all elements which are less than or equal to 9.

Warning and Disclaimer. Some books and teachers use $=$ instead of $:=$ to make definitions.

Question. How would one write S in list notation?

Solution: We have

$$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}\,.$$

□

Returning to (2.3), instead of asking all of the languages that you speak, someone might ask you whether you speak a specific language. For example, they may ask whether you speak English. From our perspective, they are really asking whether English is an element of the set of languages which you speak. It is convenient to to be able to abbreviate this question and its answer. We use the symbol $\in$ to say that an object is an element of the set, and $\notin$ to say that the object is not an element of the set.

Example 2.5. If the set of languages you speak is really (2.4), then

$$\begin{aligned}\text{English} \in \{&\text{Cantonese}, \text{English}, \text{German}, \text{Mandarin}\} \\ &= \{x : x \text{ is a language you can speak}\}\,.\end{aligned}$$

However,

$$\text{French} \notin \{x : x \text{ is a language you can speak}\}\,.$$

To give another example (using S from (2.5)) we have $2 \in S$, but $10 \notin S$.

Certain sets occur often in mathematics and we collect their notation here.

Definition 2.6. We call the set which has no elements the empty set and denote it by $\emptyset$ or $\{\}$.

The set of all integers (both positive, negative, and zero) is written $\mathbb{Z}$, the set of all positive integers (also known as the natural numbers is written $\mathbb{N}$, the set of non-negative integers (positive integers and zero) is denoted by $\mathbb{N}_0$, the rational numbers (ratios of integers) are denoted $\mathbb{Q}$, the real numbers are denoted by $\mathbb{R}$, and the complex numbers are denoted by $\mathbb{C}$.

For a set S, it is common to write the set of pairs (x, y) with $x \in S$ and $y \in S$ as S^2. For example

$$\mathbb{R}^2 = \{(x, y) : x \in \mathbb{R} \text{ and } y \in \mathbb{R}\} .$$

It is also useful to restrict a known set by adding an extra property. Since it is bulky to write

$$\{x : x \in S \text{ and } x > 3\}$$

one commonly abbreviates this by

$$\{x \in S : x > 3\} .$$

The list notation is quite useful, but it doesn't work very well if there are too many elements. An interval of real numbers is the set of $x \in \mathbb{R}$ satisfying an inequality like

$$a \leq x \leq b$$

for some $a, b \in \mathbb{R}$ chosen beforehand (one also allows a and b "to be" ∞). We further abbreviate this by the following:

$$\begin{aligned}
(a, b) &:= \{x \in \mathbb{R} : a < x < b\} ,\\
(a, \infty) &:= \{x \in \mathbb{R} : x > a\} ,\\
(-\infty, b) &:= \{x \in \mathbb{R} : x < b\} ,\\
[a, b) &:= \{x \in \mathbb{R} : a \leq x < b\} .
\end{aligned}$$

In general, replacing a parethesis '(' or ')' with a bracket '[' or ']' changes the inequalty to include equality. The interval (a, b) is called open, while the interval $[a, b]$ is called closed (the intervals $[a, b)$ and $(a, b]$ are called half-open).

Example 2.7. Give the set of solutions to

$$-3x < 5$$

in interval notation.

Solution: Dividing by 3 on both sides gives

$$-x < \frac{5}{3}$$

We now add x to both sides and subtract $\frac{5}{3}$ to get

$$-\frac{5}{3} < x.$$

In interval notation, the solution is

$$\left(-\frac{5}{3}, \infty\right).$$

$\square$

Given two sets S and T, we use

$$S \cup T$$

to denote the set (called the union of S and T) which contains all of the elements of S and all of the elements of T. For example, if $S = \{1, 5, 7\}$ and $T = \{2, 5, 9\}$, then

$$S \cup T = \{1, 2, 5, 7, 9\}.$$

Note that 5 occurs only once in $S \cup T$. This is because elements are either inside the set or not in the set (they are only counted once).

We call the set of elements contained in both S and T the intersection of S and T and denote this by

$$S \cap T = \{5\}.$$

Example 2.8. Write the set of solutions to

$$7 < 2x < 12 \tag{2.6}$$

as an intersection and the set of solutions to

$$7 < 2x \text{ or } 3x < -12 \tag{2.7}$$

as a union.

Solution: In equation (2.6), we have

$$7 < 2x \implies x > \frac{7}{2}$$

and

$$2x < 12 \implies x < 6.$$

Thus the set of solutions is

$$\left(\frac{7}{2}, \infty\right) \cap (-\infty, 6).$$

In (2.7), we have

$$x > \frac{7}{2}$$

or

$$x < -4.$$

Thus the set of solutions is

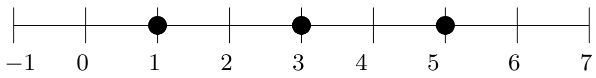
$$\left(\frac{7}{2}, \infty\right) \cup (-\infty, -4)\,.$$

$\square$

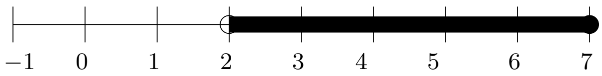
2.1. **Graphical representation.** There are a number of ways to graphically represent a set. For example, by drawing the real line as a axis, we can put a dot on each point in the set. The set

$$S := \{1, 3, 5\}$$

is graphically given by



The interval $(2, 7]$ is depicted by

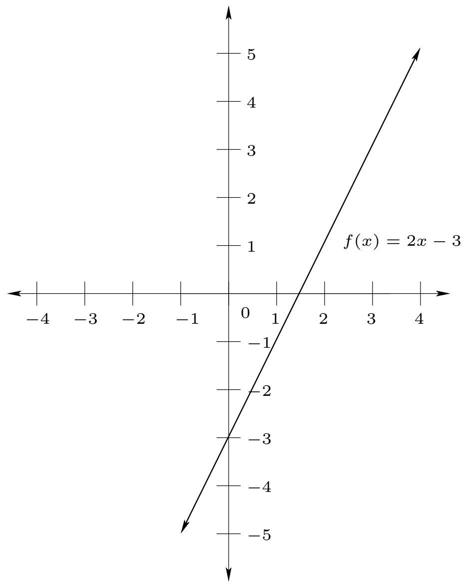


(filled in circles are included, empty circles are not)

Now consider the set

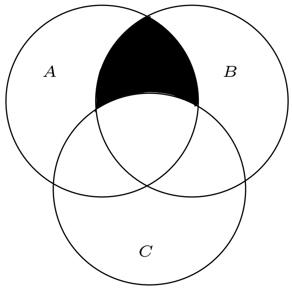
$$\left\{(x, y) \in \mathbb{R}^2 : y = 2x - 3\right\}.$$

This set (the set of solutions to $y = 2x - 3$) can be graphically represented by



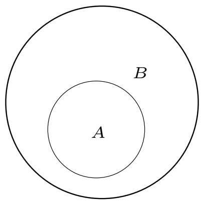
Another useful graphical representation of sets is given by Venn diagrams. This is particularly useful for unions and intersections. Say that we have three sets A, B, and C. For each set, we draw a circle to represent the box holding the collection of objects in that set. However, the circles may overlap, because there may be objects which occur in two or three of the sets simultaneously.

Example 2.9. An example of a Venn diagram is given by



In the above picture, the shaded region depicts the elements which are contained in both A and B, but are not in C.

Another Venn diagram is the following:



In the picture above, A is completely contained inside B. In other words, if $x \in A$, then $x \in B$. We say that A is a <u>subset</u> of B and write $A \subseteq B$.

3. Functions

Functions are important objects which appear throughout mathematics and beyond. One way to think of a function is to imagine a machine that has an intake and output pipe. You place an input into the intake pipe, the machine applies some rule, and then it outputs something on the other end. Consider for the moment your computer. If you type the letter 'L' on your keyboard (this is the input), your computer does something (you may or may not know what) and then something on your screen appears which looks like the letter 'L' (this is the output).

Of course, machines break, but functions are perfect machines that give the same ansswer every time that you give it the same input.

Example 3.1. The function

$$f(x) := 3x^2 + 7$$

returns $3x^2 + 7$ whenever the input is x. For example, if you give the input 3, then you get the output

$$f(3) = 3(3^2) + 7 = 34.$$

Definition 3.2. A function from a set D to a set C is a rule which for each element of D determines a unique output element of C. One calls D the domain of the function and C the co-domain.

If the domain of a function f is not explicitly written, then we assume that the domain is the largest set of real numbers for which the co-domain is real. This is called the natural domain of f.

Example 3.3. For

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

the natural domain is $[2, \infty)$, because the parameter to the square root must be non-negative for the value (the output) to be real.

Although we require a function to give a value for each element of D, it is possible that not every element of C actually occurs as an output. Furthermore, two different inputs may return the same output. The range of the function is the set of all outputs which actually occur.

Example 3.4. We return to the function

$$f(x) := 3x^2 + 7$$

with $D = \mathbb{R}$ (the natural domain). Since $x^2 \geq 0$, we have $f(x) \geq 7$ for every $x \in \mathbb{R}$. Hence the range is a subset of $[7, \infty)$.

However, if $y \geq 7$ then choose

$$x_0 = \sqrt{\frac{y - 7}{3}}$$

so that

$$f(x_0) = 3\left(\frac{y-7}{3}\right) + 7 = y.$$

Therefore y is in the range of f and we conclude that the range of f is precisely $[7, \infty)$.

Furthermore, the input which yields y is not unique because

$$f(-x_0) = y.$$

If we restrict the domain to $D = [2, \infty)$, then the range is $[19, \infty)$, since

$$3x^2 + 7 \geq 3(2)^2 + 7 = 19$$

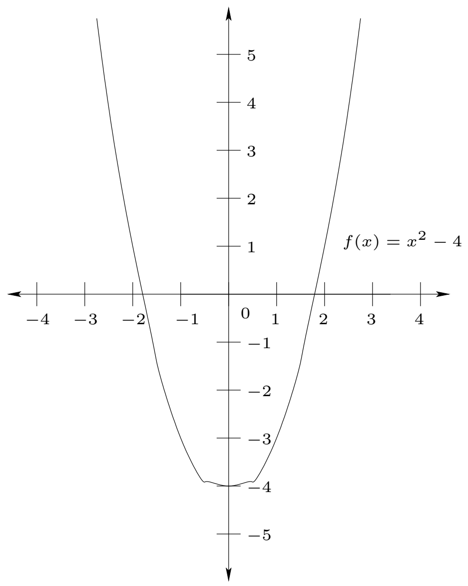
whenever $x \geq 2$.

3.1. **Graphs of functions.** There are a number of ways to represent a function. One of the most useful ways is through its graph.

Definition 3.5. The graph of a function is the set of pairs of inputs and outputs of the function.

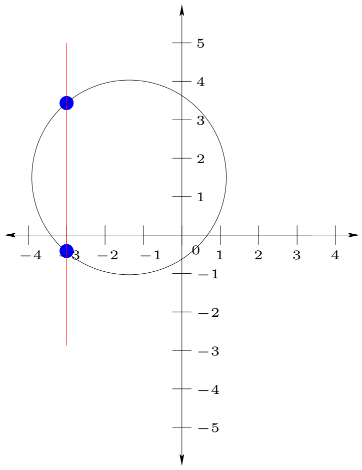
When the domain and co-domain are both $\mathbb{R}$, the graph of a function is often pictorially represented by placing dots at the pairs (x, y) which are elements of the graph. Here x is the input and y is the output.

Example 3.6. Here is a sketch of the graph of $f(x) = x^2 - 4$.

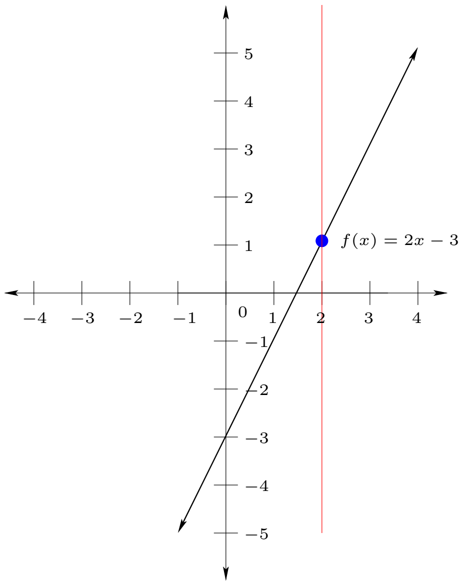


Given a graph, one can ask whether it is the graph of a function. Remember that for each input there is only one output. Looking at the graph, this means that every vertical line can only hit the graph once.

Example 3.7.



not the graph of a function



the graph of a function

The vertical line test says that a graph is the graph of a function precisely when there is no vertical line which intersects the graph more than once.

There are a number of important functions which occur often. A list of some of these is given here:

(1) Polynomial functions: These are functions of the type

$$f(x) = A_n x^n + A_{n-1} x^{n-1} + \cdots + A_1 x + A_0,$$

where $A_0, \ldots, A_n$ are some constants (this means that they do not change when x changes) and $A_n \neq 0$. The number $n \in \mathbb{N}_0$ is called the degree of the polynomial. When the degree is at least 1, one calls the x for which $f(x) = 0$ the roots or zeros of the polynomial.

There are a number of special cases of polynomials which occur in a variety of applications. For example:

- Constant functions: These are functions which have the same output for any input. They are degree 0 polynomials. One example is
$$f(x) = 7.$$
- Linear functions: These are degree 1 polynomials. An example of such a function is
$$f(x) = 3x + 4.$$
Its graph is a straight line. For a function $f(x) = mx + b$, we call m the slope of the line and b the y intercept. The slope is the number of units that the line increases in the y direction when x increases by 1.
An important special case of a linear function is the identity function
$$f(x) = x,$$
which just returns back the same output as the input.
- Quadratic functions. These are degree 2 polynomials. An example is
$$f(x) = x^2 + 3x + 4.$$
Their graph is a parabola. If $a, b, c \in \mathbb{R}$, then the quadratic function $f(x) = ax^2 + bx + c$ has either 2 real roots, one real root, or two complex (non-real) roots, depending on whether the discriminant $\Delta := b^2 - 4ac$ is positive, zero, or negative. This is because the roots are the solutions to the quadratic equation
$$x = \frac{-b \pm \sqrt{\Delta}}{2a}.$$

(2) Rational functions: These are polynomials divided by other polynomials. That is, $f(x) = \frac{g(x)}{h(x)}$, where g and h are polynomials and h is not the constant zero function. For example
$$f(x) = \frac{3x^2 + 2}{2x^3 + 3x + 7}.$$

(3) Piecewise-defined functions: These are defined for x in certain ranges, with different definitions in different ranges. For example, the piecewise linear function
$$f(x) = \begin{cases} 2x + 7 & \text{if } x < 1, \\ 9x & \text{if } x > 1, \\ 3 & \text{if } x = 1. \end{cases}$$

(4) Square-root (or m-th) function: The function

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

returns the $y \geq 0$ so that $y^2 = x$ (there are **two** choices y and $-y$, but we have to choose one to make f a function). The n-th root is the number $y \in \mathbb{R}$ ($y > 0$ if n is even) so that $y^n = x$. We write this function as

$$f(x) = \sqrt[n]{x}.$$

3.2. **Even/Odd functions.**

Definition 3.8. A function is called <u>even</u> if for every x in the domain we have

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

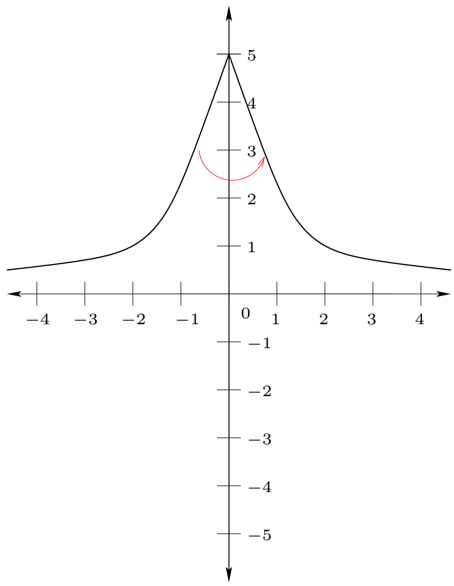
This means that $f(-7) = f(7)$, $f(-2) = f(2)$, etc. As an example, consider

$$f(x) := x^4 - 3x^2 + 1.$$

To compute $f(-x)$, we simply replace x with $-x$ everywhere (the input is $-x$). This gives (careful with parentheses here, or you'll get the wrong answer)

$$\begin{aligned} f(-x) &= (-x)^4 - 3(-x)^2 + 1 \\ &= x^4 - 3x^2 + 1 = f(x). \end{aligned}$$

The graph of an even function is symmetric about the y-axis. This means that if we spin the graph around the y-axis, we get the same graph. An example is below:



Definition 3.9. An <u>odd function</u> is a function which satisfies

$$f(-x) = -f(x)$$

for every x in its domain.

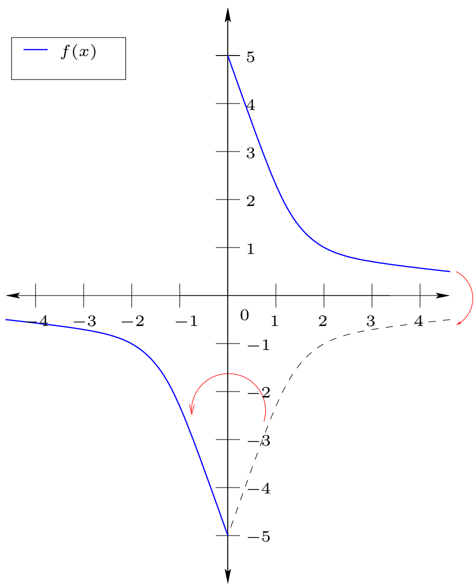
An example of an odd function is

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

To check this, we compute

$$\begin{aligned} f(-x) &= (-x)^3 - (-x) + \frac{1}{-x} \\ &= -x^3 + x - \frac{1}{x} \\ &= -\left(x^3 - x + \frac{1}{x}\right) = -f(x). \end{aligned}$$

The graph of an odd function is not symmetrical about the y-axis or the x-axis, but if you flip over **both** the y-axis and then the x-axis, the graph is the same. Here is the example of such a graph (the dashed line is only for reference, not part of the graph):



Now consider

$$f(x) := x^4 - 2x + 1.$$

To check if it is even or odd, we compute $f(-x)$. This gives

$$f(-x) = (-x)^4 - 2(-x) + 1 = x^4 + 2x + 1.$$

Question. This doesn't look like $f(x)$, and it doesn't look like $-f(x)$. But how do we know for **sure** that it's not (sometimes things look very different, but end up somehow being equal)?

Solution: To check that something is **not** even and **not** odd, all you have to do is check one value and show that $f(-x) = f(x)$ and $f(-x) = -f(x)$ are **not** satisfied. Remember that $f(-x) = f(x)$ means that I can plug in any choice of x and it should be true. Therefore, if $f(-7) \neq f(7)$, then it is not satisfied (even if it is satisfied for every other choice of x).

In the above example, we compute

$$f(-1) = (-1)^4 - (-1) + 1 = 3.$$

At the same time,

$$f(1) = 1^4 - 1 + 1 = 1.$$

Since $3 \neq 1$ and $3 \neq -1$, we know that f is not even and f is not odd. $\square$

3.3. **New functions from old functions.** For two functions f and g, there are a number of new functions that can be constructed. One can define the sum

$$h(x) := f(x) + g(x)$$

or product

$$h(x) := f(x)g(x)$$

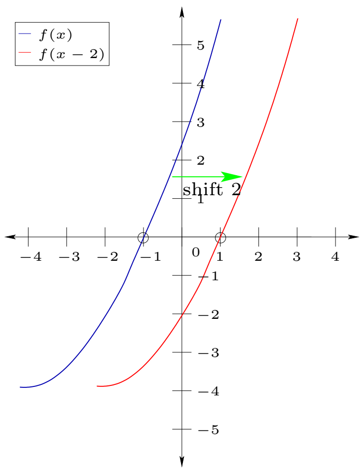
as well as the quotient

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

You can also define the horizontal shift of f by

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

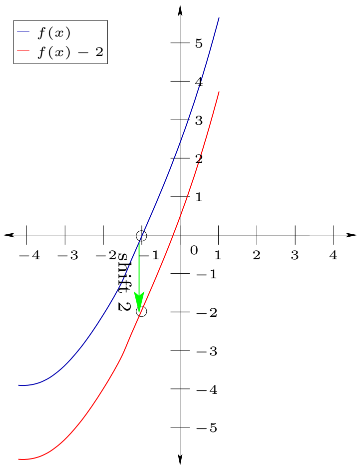
where $r \in \mathbb{R}$. The graph of h is the graph of f shifted r spaces to the right.



The vertical shift of f is given by

$$h(x) = f(x) + r.$$

This is shifted up by r.



3.4. **Composition of functions.** An important new function constructed from f and g is the composition of f and g defined by

$$f \circ g(x) := f(g(x)).$$

Example 3.10. For $f(x) = x^2 + 1$ and $g(x) = \sqrt{x}$, we have

$$f \circ g(x) = g(x)^2 + 1 = \left(\sqrt{x}\right)^2 + 1$$

and

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

Note that the natural domain of $f \circ g$ is $x \geq 0$, while the natural domain of $g \circ f$ is $\mathbb{R}$! You see directly that the order of the composition is quite important.

If $f \circ g$ and $g \circ f$ are the identity function, then we say that g is the inverse of f. It is common to use the notation $f^{-1}(x)$ for the inverse.

Warning. Do not mix this up with the inverse from multiplication, namely $\frac{1}{f(x)}$. One usually says that f^{-1} is the inverse function and $\frac{1}{f(x)}$ is the multiplicative inverse.

One way to think about the inverse function is to try to define it directly. If

$$f(a) = b,$$

then we must have

$$f^{-1}(b) = a$$

because

$$a = f^{-1}(f(a)) = f^{-1}(b).$$

The inputs of f are the outputs of f^{-1} and vice-versa. So the inverse can be thought of as taking the output from the machine and putting it in backwards to have the machine "undo" its operation.

Now suppose that $f(3) = 7$. Then we would have

$$f^{-1}(7) = 3.$$

However, if $f(2) = 7$, then

$$f^{-1}(7) = 2.$$

We've given f^{-1} the same input and gotten a different output. Our machine is broken! What did we do wrong?

The problem is that the inverse **does not always exist.** So, when does it exist? To answer this question, think about the vertical line test. The function f^{-1} must satisfy the vertical line test, so it can only have one output for each input. Since the outputs of f are the inputs of g, this means that f can only have each output at most once.

Definition 3.11. A function which has each possible output at most once is called <u>injective</u>.

The inverse of a function f exists if and only if f is injective.

Example 3.12. Consider the function

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

Does f have an inverse?

Solution: No, because $f(1) = 1$ and $f(-1) = 1$, so f is not injective. $\square$

Example 3.13. Consider the function

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

with domain $[0, \infty)$. Does f have an inverse?

Solution: Yes, in this domain f only attains each value once, so it is injective. The inverse function is

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

You can see this because

$$\left(\sqrt{x}\right)^2 = x$$

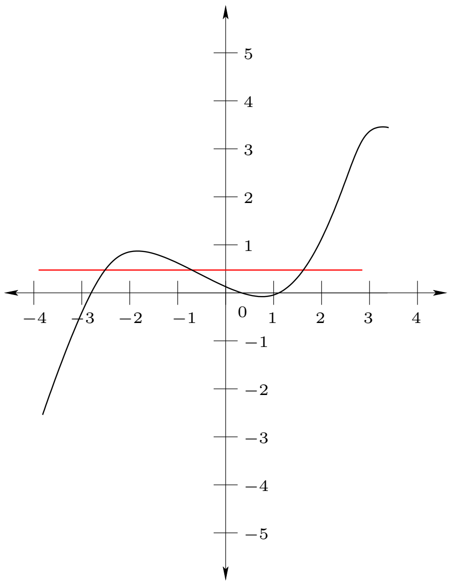
and

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

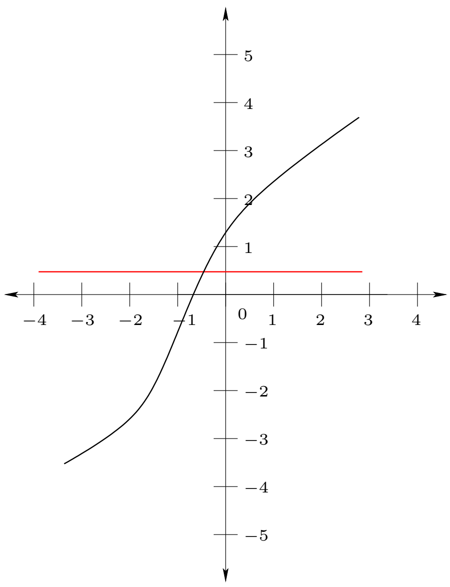
but $x \geq 0$, so $|x| = x$.

From this example, you can see that checking whether a function is injective **depends on the domain**. $\square$

An easy way to check whether a function is injective is the <u>horizontal line test</u>. If a function is injective, then every horizontal line intersections its graph at most once.

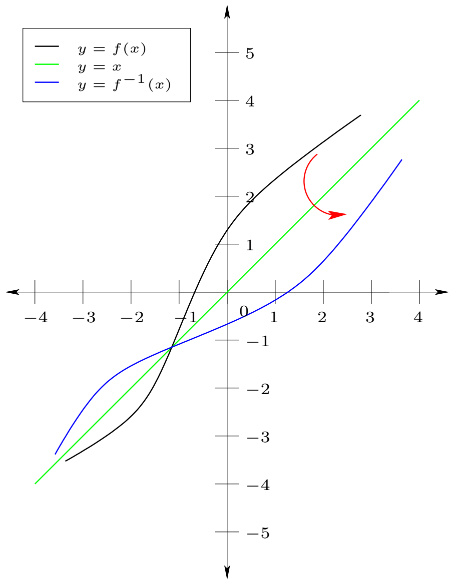


not injective



injective

The horizontal line test might remind you of the vertical line test to check whether a graph is a function. This is not an accident, because the horizontal line test is actually checking whether the graph coming from our previous attempt to define the inverse is actually a function or not. To see, this, think about what changing the input and the output does. This exchanges the roles of x and y, which is the same as flipping over the line $y = x$.



$y = f(x)$
$y = x$
$y = f^{-1}(x)$
5
4
3
2
1
0
−1
−2
−3
−4
−5
−4
−3
−2
−1
1
2
3
4

4. Limits and Continuity

The concept of a limit is fundamental in calculus. Roughly speaking, a limit of a function $f(x)$ as x approaches a is the value that you approach when you get "closer and closer" to a. This is **not** necessarily the same as the **value** of the function. Think about a magician who puts a coin in her right hand and then moves it to the right. You would say that the coin is approaching the right side. However, the magician suddenly opens her left hand to reveal that the coin has magically (and instantaneously – we'll assume that she has real magical powers). Although the coin was approaching the right side, at the exact moment that she opened her hand, it was in the left side. The **limit** is the **right side**, but the **value** is the **left side**.

Slightly more formally (for those with more interest, see me for a completely formal definition), if $f(x)$ gets closer and closer to L whenever x gets closer and closer to a, then the limit of $f(x)$ as x approaches a is L and we write

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

Let's consider an example.

Example 4.1. Consider the (piecewise-defined) function

$$f(x) := \begin{cases} \frac{x^2-4}{x-2} & \text{if } x \neq 2, \\ 1 & \text{if } x = 2. \end{cases} \tag{4.1}$$

What is the limit as x approaches 2? What is the value?

Solution: For $x \neq 2$, we have

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

In particular, as x gets really close to 2, $f(x)$ gets really close to $2+2=4$. Thus

$$\lim_{x\to 2} f(x) = 4.$$

However, directly from the definition, we have

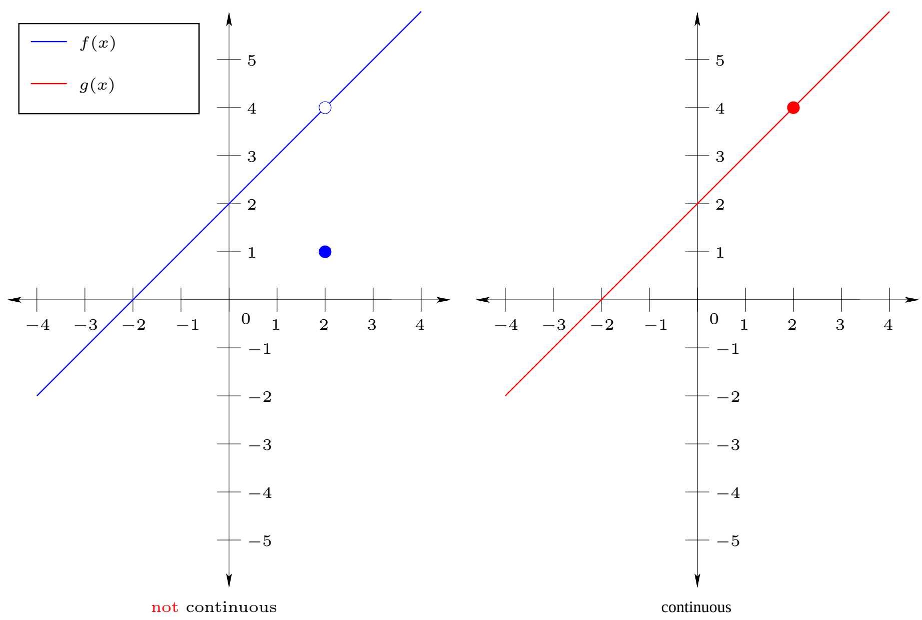
$$f(2) = 1.$$

□

Remark. To consider the limit towards a point a, the function **doesn't even need to be defined** at a! In the above example, we could have defined $f(x) = \frac{x^2-4}{x-2}$, which does not have $x = 2$ in its natural domain. However, the limit as x approaches 2 is still 4.

If you look at the graph of $f(x)$, you see that it suddenly jumps at $x = 2$, just like the magician's trick. If we instead defined $f(2)$ to be 4, then it would continue at this point. To see this difference, we graph $f(x)$ and

$$g(x) := \begin{cases} f(x) & \text{if } x \neq 2, \\ 4 & \text{if } x = 2. \end{cases}$$

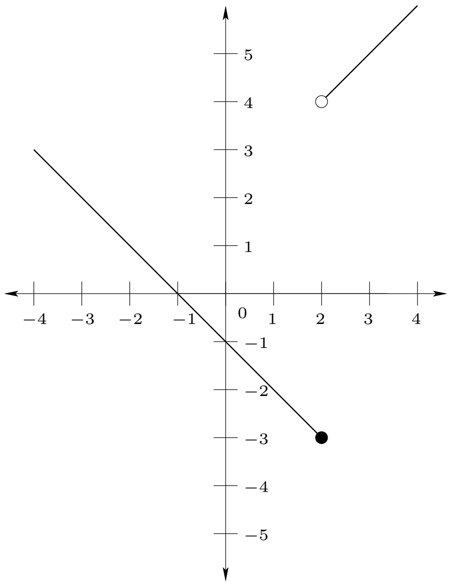


We will discuss the difference between these two phenomena when we speak about continuity later.

Before continuing, it is important to discuss the **existence** of a limit. Consider the function

$$g(x) := \begin{cases} \frac{x^2-4}{x-2} & \text{if } x > 2, \\ -x-1 & \text{if } x \leq 2. \end{cases}$$

This has the following graph:



If you approach $x = 2$ from the right, then $f(x)$ approaches 4, but if you approach $x = 2$ from the left, then $f(x)$ approaches -3. We say that the limit **does not exist**. However, one says that the limit from the right is 4 and the limit from the left is -3. This is written:

$$\lim_{x \to 2^+} f(x) = 4$$

and

$$\lim_{x \to 2^-} f(x) = -3.$$

We also simply say that $\lim_{x\to 2} f(x)$ does not exist (this is sometimes abbreviated DNE). There are even cases when the left and right limits do not exist, but we will go into that further later.

Limits satisfy many useful properties. Assume that $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ both exist. Then we have the following

(1) The limit of the sum or difference of two functions is the sum or difference of the limits of the individual functions, i.e.,

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

(2) The limit of the product of two functions is the product of the limits of the individual functions, i.e.,

$$\lim_{x \to a} [f(x)g(x)] = \left(\lim_{x \to a} f(x)\right)\left(\lim_{x \to a} g(x)\right).$$

In particular, as a special case the, limit of the product of a constant times a function is the constant times the limit of the function. That is to say, for $r \in \mathbb{R}$,

$$\lim_{x \to a} [rf(x)] = r \lim_{x \to a} f(x).$$

(3) If $\lim_{x\to a} g(x) \neq 0$, then the limit of the ratio of two functions is the ratio of the limits, i.e.,

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

(4) The limit of the n-th root of a function is the n-th root of the limit (if n is even, then we assume that $\lim_{x\to a} f(x) \geq 0$)

$$\lim_{x\to a} \sqrt[n]{f(x)} = \sqrt[n]{\lim_{x\to a} f(x)}.$$

(5) For a polynomial $f(x)$ we have

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

Example 4.2. Find the limit

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

Solution: We use the properties above to compute

$$\lim_{x\to 2} \frac{x^2-1}{x^2+2} = \frac{\lim_{x\to 2}(x^2-1)}{\lim_{x\to 2}(x^2+2)} = \frac{\lim_{x\to 2} x^2 - \lim_{x\to 2} 1}{\lim_{x\to 2} x^2 + \lim_{x\to 2} 2} = \frac{3}{6} = \frac{1}{2}.$$

We could have also computed $\lim_{x\to 2}(x^2-1) = 3$ directly by plugging in $x = 2$, since it is a polynomial. □

Example 4.3. Find the limit

$$\lim_{x\to 1} \frac{x^4-1}{x^2-3x+2}.$$

Solution: One is tempted to bring the limit to the denominator and the numerator, but this is not legal because we have to check that the limit of the denominator is non-zero. In this case, it is zero, however (if you plug in directly, you get $\frac{0}{0}$, which is not well-defined). In this case, we factor

$$x^4 - 1 = \left(x^2-1\right)\left(x^2+1\right) = (x-1)(x+1)\left(x^2+1\right)$$

and

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

Now, for $x \neq 1$, we have

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

Since we are only getting "really close" to $x = 1$, we are allowed to cancel out and then take the limit. It follows that

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

Now we can plug in directly to get

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

□

The above example illustrates another property of limits. You can **cancel a common factor** in the numerator and denominator.

Example 4.4. Find the limit

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

Solution: Again, we cannot simply plug in $x=4$. There is no obvious factor to cancel, However, by using the fact that

$$(X-r)(X+r)=X^2-r^2,$$

we can multiply and divide by $\sqrt{x}+2$ to make a common factor appear. We thusly rewrite

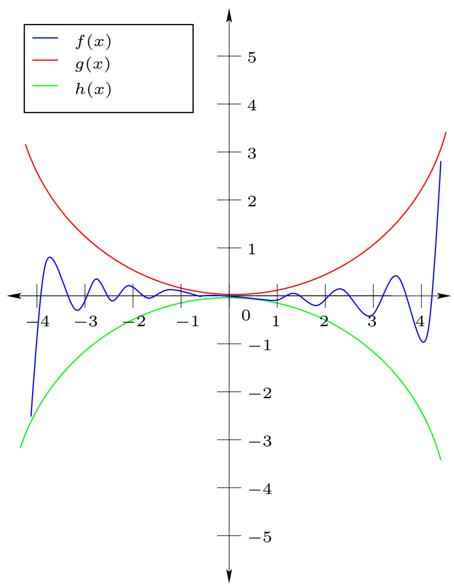
$$\frac{\sqrt{x}-2}{x-4}=\frac{(\sqrt{x}-2)(\sqrt{x}+2)}{(x-4)(\sqrt{x}+2)}=\frac{\cancel{x-4}}{\cancel{(x-4)}(\sqrt{x}+2)}=\frac{1}{\sqrt{x}+2}.$$

Therefore

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

□

In the examples above, the limits can all be directly computed, but this is not always the case. Sometimes we have to use the help of other functions. Consider the following diagram:



The functions $g(x)$ and $h(x)$ are rather simple (perhaps they are something like $g(x) = x^2$ and $h(x) = -x^2$). Therefore, it may be easy to compute the limits

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

and

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

You would guess that

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

from the picture, but it might be difficult to show this. You can then use the following theorem.

Theorem 4.5 (Squeeze theorem)**.** *Suppose that there is some interval* (A, B) *with* $a \in (A, B)$ *so that for every* $x \in (A, B)$ *with* $x \neq a$ *we have*

$$h(x) \leq f(x) \leq g(x).$$

If

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

the the limit $\lim_{x\to a} f(x)$ *exists and*

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

The idea of Theorem 4.5 is that a function squeezed between two other functions that approach the same value cannot "escape" and end up at another value.

This is a concept used even by sports stars. If you have played soccer (football), basketball, or hockey, then you would be familiar with a "box out". The idea is to use one's body to force the direction of another player to go the direction that you want them to go. They cannot go through you (this is like the inequality $f(x) \leq g(x)$). If there were 2 defenders both running towards a particular point and the other player was stuck between them, then there would be no choice other than all players ending up at the same point, much like the graph above.

Example 4.6. Compute

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

Solution: For every x, we have

$$-1 \leq \sin(x) \leq 1.$$

We also have

$$-|x| \leq x \leq |x|.$$

Therefore, for $x \neq 0$, we have

$$-|x| \leq x \sin\left(\frac{1}{x}\right) \leq |x|.$$

Since

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

and

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

we conclude that

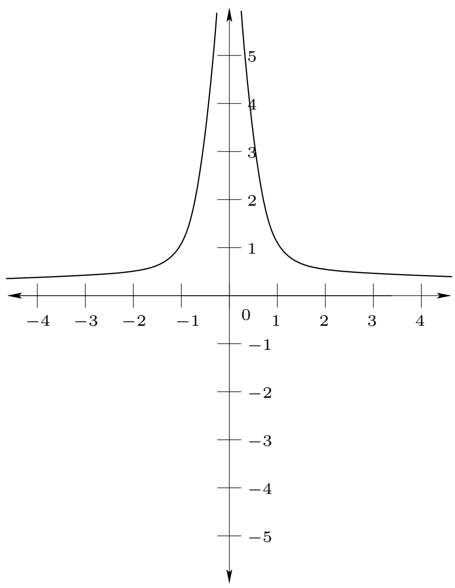
$$\lim_{x\to 0} x\sin\left(\frac{1}{x}\right) = 0.$$

$\square$

4.1. **Infinite limits and limits at infinity.** Consider the limit

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

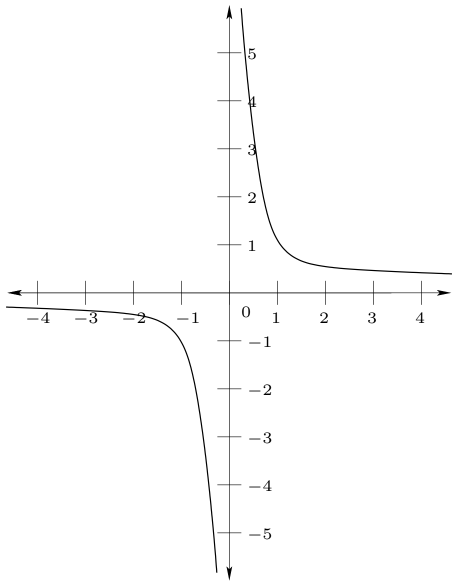
Here is the graph of $f(x) = \frac{1}{x^2}$.



As $|x|$ gets smaller and smaller, $f(x)$ gets bigger and bigger (for example, $f\left(\frac{1}{1000}\right) = 1000^2$ and also $f\left(-\frac{1}{1000}\right) = 1000^2$). If a function $f(x)$ keeps getting larger and larger as x gets closer to a, we say that <u>the limit is infinite</u>, and this is written

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

Now consider the graph of $f(x) = \frac{1}{x}$.



We see that

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

and

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

Since one goes to ∞ and one goes to $-\infty$, the limit at 0 does not exist.

In a similar way, we consider infinite limits. Here we ask the question: as x gets larger and larger, what does $f(x)$ approach? For $f(x) = \frac{1}{x}$, as x gets larger and larger, $f(x)$ gets smaller and smaller. We hence write

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

We also have

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

Example 4.7. Find

$$\lim_{x\to\infty} x^3$$

and

$$\lim_{x\to-\infty} x^3.$$

Solution: As x gets larger, x^3 blows up to infinity, so

$$\lim_{x\to\infty} x^3 = \infty.$$

If x is negative, then x^3 is also negative. At the same time

$$\left|x^3\right|$$

keeps getting larger. We conclude that

$$\lim_{x\to-\infty} x^3 = -\infty.$$

□

To find the limit towards $\pm\infty$ of a rational function, one divides the numerator and denominator by the highest power of x appearing in the denominator.

Example 4.8. (1) Find the limit

$$\lim_{x\to-\infty} \frac{3x^2+2x+1}{7x^2+4}.$$

(2) Find the limit

$$\lim_{x\to\infty} \frac{2x+1}{7x^2+4}.$$

(3) Find the limit

$$\lim_{x\to-\infty} \frac{3x^2+2x+1}{4x+1}.$$

Solution: (1) We divide the numerator and denominator by x^2 to compute

$$\begin{aligned}\lim_{x\to-\infty} \frac{3x^2+2x+1}{7x^2+4} &= \lim_{x\to-\infty} \frac{3+\frac{2}{x}+\frac{1}{x^2}}{7+\frac{4}{x^2}}\\ &= \frac{3+0+0}{7+0} = \frac{3}{7}.\end{aligned}$$

In the last line, we used the properties of the limit to plug in $\lim_{x\to-\infty} \frac{1}{x} = 0$ and $\lim_{x\to-\infty} \frac{1}{x^2} = 0$.
(2) We divide the numerator and denominator by x^2 to compute

$$\begin{aligned}\lim_{x\to\infty} \frac{2x+1}{7x^2+4} &= \lim_{x\to\infty} \frac{\frac{2}{x}+\frac{1}{x^2}}{7+\frac{4}{x^2}}\\ &= \frac{0+0}{7+0} = 0.\end{aligned}$$

In the last line, we used the properties of the limit to plug in $\lim_{x\to\infty} \frac{1}{x} = 0$ and $\lim_{x\to\infty} \frac{1}{x^2} = 0$.
(3) We divide the numerator and denominator by x to compute

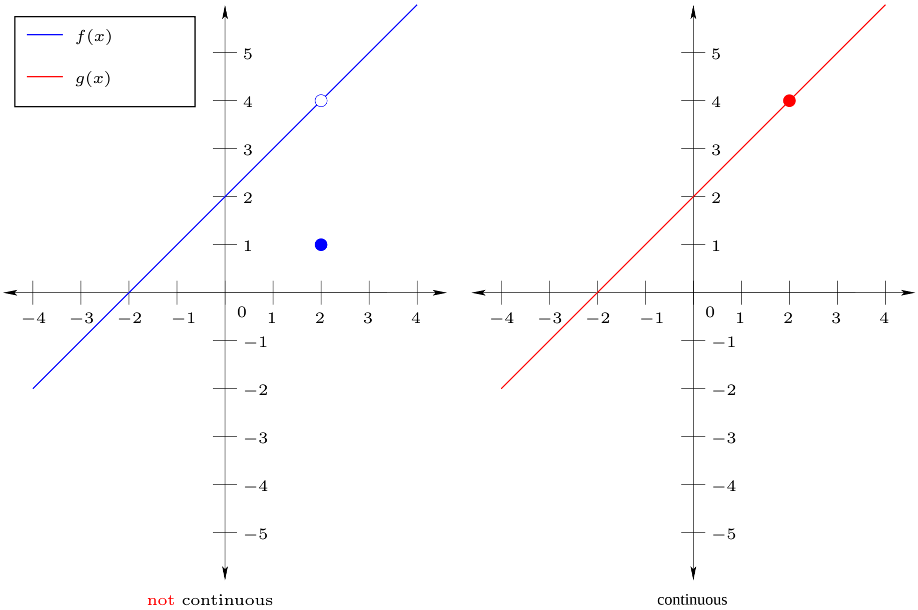
$$\lim_{x\to-\infty} \frac{3x^2+2x+1}{4x+1} = \lim_{x\to-\infty} \frac{3x+2+\frac{1}{x}}{4+\frac{1}{x}}.$$

We now note that the numerator approaches $-\infty$ as $x \to -\infty$, while the denominator approaches 4. Thus the ratio approaches $-\infty$. We conclude that

$$\lim_{x\to-\infty} \frac{3x^2+2x+1}{4x+1} = -\infty.$$

□

4.2. **Continuity.** Roughly speaking, continuity (derived from "continue") is the property of being smooth (without jumps). If a function has jumps (like the magician's trick), then we say that the function is discontinuous. The point where it has a jump we call a discontinuity (or a point of discontinuity). Recall the following graph:



The function $f(x)$ is discontinuous, and has a discontinuity at $x = 2$. The function $g(x)$ has no discontinuities, so we say that it is continuous.

We now write down the definition of continuity a little more rigorously.

Definition 4.9. We say that a function $f(x)$ is continuous at $x = a$ if:

(1) The value $f(a)$ exists.

(2) The limit
$$\lim_{x \to a} f(x)$$
exists.

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

We say that a function is continuous on the set D if $f(x)$ is continuous at every point $a \in \mathbb{D}$. If $D = \mathbb{R}$ is the (natural) domain of f, then we may abbreviate this by saying that f is continuous.

The function $f(x)$ in the diagram above satisfies the first two conditions, but the third is not satisfied, because
$$\lim_{x \to 2} f(x) = 4$$
and
$$f(2) = 1.$$

However, for the function g, we have

$$\lim_{x \to 2} g(x) = 4$$

and

$$g(2) = 4.$$

This verifies mathematically what our intuition tells us, that g is continuous.

Example 4.10. Where is the function $f(x) = \frac{1}{x}$ continuous? Where is it discontinuous?

Solution: The function is continuous for $x \neq 0$, and is discontinuous at $x = 0$, because the value does not exist (also, the limit does not exist). □

Definition 4.11. A function $f(x)$ is called right-continuous at $x = a$ if it satisfies all of the conditions of continuity with the limits replaced with right limits. If $f(x)$ satisfies all of the conditiosn with the limits are replaced with left limits, we say that $f(x)$ is left-continuous at $x = a$.

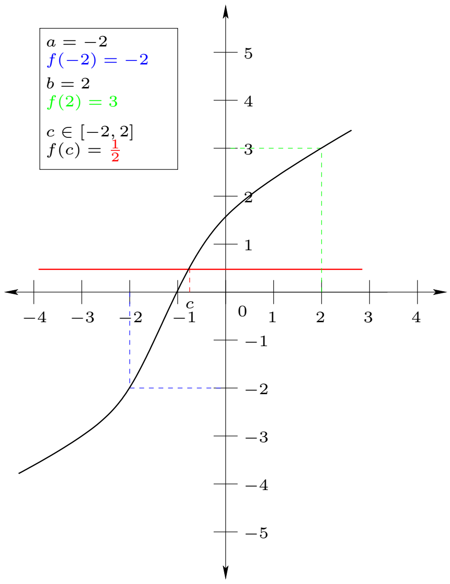
4.3. **Intermediate value theorem.** Let's return to the magician's trick. The magician managed to magically and instantaneously move the coin from the right to the left. Since it was instantaneous, the coin was never anywhere between the right and the left hand. However, if we now return the magician to mere mortal status (her magic is now only a slight of hand, and she is returned to the world of continuity), then she must physically move the coin from the right hand to the left hand. As a result, the coin must pass through all of the space between the left hand and the right hand. Therefore, at some point (we don't know when, because the magician is very good at hiding the trick), the coin had to be at every point between the two hands. This is indicative of the following important theorem about continuous functions.

Theorem 4.12 (Intermediate Value Theorem)**.** *Suppose that $f(x)$ is a continuous function on a closed interval $[a, b]$. If y_0 is between $f(a)$ and $f(b)$, then there must be some $c \in [a, b]$ for which*

$$f(c) = y_0.$$

The theorem doesn't say anything about what c is (much like we don't know the time that the coin passes through the air), but we can guarantee that it must be there sometime in between a and b.

Here is another way to visualize the intermediate value theorem. Let's say that you see someone standing outside of your room at time a. At time b, they are inside your room. Perhaps you were not paying attention and did not see him or her come in, but you can still guarantee that at some point c, they were directly in your doorway. Here is a graphical representation of the Intermediate Value Theorem.



The Intermediate Value Theorem is helpful for finding roots of a function. Suppose that f is a continuous function and $f(a) > 0$. If $f(b) < 0$ with $a < b$, then at some point between a and b, the graph must have crossed over the x-axis. Wherever it crossed (we don't know), the value of f is zero. This means that there is a c between a and b so that $f(c) = 0$.

This allows us to find roots by noticing that the graph on either side of a root changes sign. Since polynomials are always continuous, this can be quite useful in (numerically) approximating roots of polynomials. You can keep getting closer and closer to find the approximate root (where the value on each side changes sign). Together with the following theorem (which is not within the scope of this course), this allows you to find all roots of a polynomial with a computer/calculator.

Theorem 4.13. *A degree n polynomial has at most n roots. In particular, a polynomial changes sign at most $n+1$ times.*

The theorem above means that, if you manage to find n roots (or $n+1$ places where it changes sign), then you have found all of the roots!

Example 4.14. Determine the range of $f(x) := 3x^2 + 7$ using The Intermediate Value Theorem.

Solution: We have already shown that the range is $[7, \infty)$ in the first week of the class. That time, we first used $x^2 \geq 0$ to show that the range was a subset of $[7, \infty)$. After this, for each $y \geq 7$ we had to construct an x_0 so that $f(x_0) = y$. Instead of doing this, we now use the continuity of $f(x)$ (it is a polynomial). Suppose that $y \geq 7$. Since

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

there exists an $X > 0$ such that

$$f(X) = 3X^2 + 7 > y.$$

Clearly $f(0) = 7$. Thus we have $0 < X$ satisfying

$$f(0) \leq y < f(X).$$

By the Intermediate Value Theorem, there exists c such that $0 \leq c \leq X$ and

$$f(c) = y.$$

Therefore, y is in the range of f.

$\square$

5. Differentiation

5.1. **Slopes and Derivatives.** One says that a linear function $f(x) = mx + b$ has slope m. The slope is the amount that y changes whenever x changes one unit. Another way to write this is as the change of y divided by the change in x. The slope is an important piece of information. For one example, think about the stock market: every day they tell you the value of the stocks (this is $f(x)$), but they also tell you how much it has changed from the day before (this is the slope of the line connecting $f(x-1)$ and $f(x)$).

Definition 5.1. The difference quotient of a function $f(x)$ at $x = a$ with increment h is

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

The difference quotient is exactly the value given by the stock market each day. Roughly speaking, stock brokers make money on the market by deciding whether the difference quotient will be positive or negative over a given time. If it is going to be positive, then they would recommend buying the stock today, and if it is going to be negative, then they would recommend selling the stock.

With the advent of computers, buying and selling of stocks has sped up to merely be the click of a mouse button. Hence it might be quite useful to know what the difference quotient was 10 minutes ago, 1 minute ago, or even 10 seconds ago to help decide what it will be in 10 minutes, 1 minute, or 10 seconds from now. This corresponds to making the increment h smaller and smaller. This should remind you of the limits that we took in the last section.

Definition 5.2. The derivative of $f(x)$ at $x = a$ is the limit

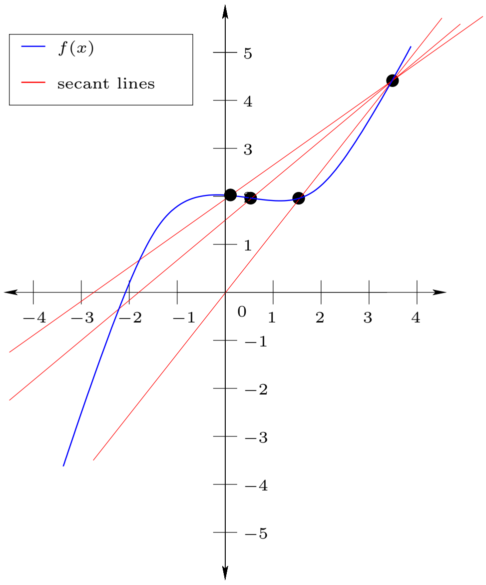
$$\lim_{h\to 0} \frac{f(a+h) - f(a)}{h}.$$

We denote the derivative at a by $f'(a)$ or $\frac{d}{dx}f(x)\Big|_{x=a}$. Rmember that limits do not always exist, so the derivative is not necessarily always defined. We say that f is differentiable at a if the limit exists. Another way to write this is

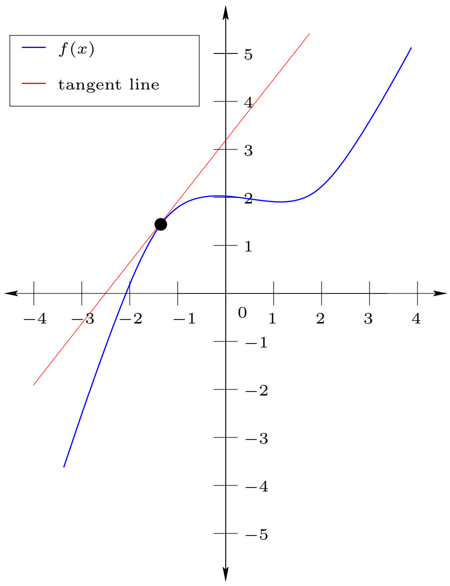
$$\lim_{x\to a} \frac{f(x) - f(a)}{x - a},$$

which may be read as the change in value between $f(a)$ and $f(x)$ divided by the change from a to x. One also simply says that f is differentiable if its derivative exists everywhere.

The difference quotients can be thought of as the slope of the secant line from $(a, f(a))$ to $(a+h, f(a+h))$, depicted below:



By taking the limit, one is computing the slope of the tangent line at the point $(a, f(a))$. Roughly speaking, the tangent line is the line which hits the graph at exactly one point (more precisely, the limit of the secant lines as the points get closer and closer, until they become one point).



Note that although there are many secant lines which go through the point $(a, f(a))$, there is only one tangent line (if it exists). The slope of the tangent line at $x = a$ is the derivative, because it is the limit of the slopes of the secant line.

Thinking of the difference quotients as the rate of change (as in the example of the stock market), one can think of the derivative as the instantaneous rate of change. In other words, at what rate does $f(x)$ change if x changes an infinitely small amount.

Remark. One often write Δ for change. Then the difference quotient is written

$$\frac{\Delta f(x)}{\Delta x}$$

When Δx becomes "infinitely small" it gets denoted dx instead of Δx. This explains the notation $\frac{d}{dx}$.

Example 5.3. To give an addition example beyond the stock market, the speed/velocity of a car is the change of distance over a given time. If you travel 100 km over a one-hour period, then your average speed (the difference quotient) is 100 km/h. However, if you look at your spedometer at a given time in between, it gives your instantaneous speed (the derivative), which might be different from 100 km/h.

5.2. **Differentiability.** As discussed above, not every function is differentiable at $x = a$.

To give an easy example of a function which is not differentiable at a point, consider

$$f(x) := |x|.$$

We have to check that the left-hand limit ($h < 0$) and the right-hand limit ($h > 0$) both exist, and that they are indeed equal.

We rewrite $f(x)$ as a piecewise-defined function, because its derivative is easier to compute then;

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

For $h > 0$, the difference quotient is

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

Therefore

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

For $h < 0$, we have

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

Thus

$$\lim_{h \to 0^-} \frac{f(h) - f(0)}{h} = -1.$$

Since the limits from each side differ, the overall limit does not exist.

Another important example comes from discontinuous functions.

Theorem 5.4. *If f is not continuous at $x = a$, then it is also not differentiable at $x = a$.*

The reason for the above theorem is the following: Take secant lines closer and closer to the point where it is discontinuous. Since f jumps at $x = a$, the lines get more and more vertical (they must jump a large distance in the y direction, but do not change much in the x direction). As a limit, the line becomes vertical, which has infinite slope.

Example 5.5. Define

$$f(x) := \begin{cases} x - 1 & \text{if } x \geq 1, \\ 2 & \text{if } x < 1. \end{cases}$$

The function $f(x)$ is discontinuous at $x = 1$, so we should find that it is not differentiable at $x = 1$ by the above theorem. Using the second definition of the derivative, we compute

$$\lim_{x \to 1^-} \frac{f(x) - f(1)}{x - 1} = \lim_{x \to 1^-} \frac{2 - 0}{x - 1} = -\infty.$$

We get $-\infty$ because the numerator is always 2, but the denominator gets smaller and smaller in absolute value but is negative.

Remark. Another way to read the theorem is the following: If f is differentiable at a, then f is continuous at a.

If the tangent line is vertical, then we also say that the derivative does not exist.

Example 5.6. The function

$$f(x) := x^{\frac{1}{3}}$$

has a veritical tangent line at $x = 0$.

5.3. **Differentiation rules.** Derivatives satisfy a number of useful rules.

(1) The derivative of a constant function is zero. That is, if there is some fixed b for which

$$f(x) := b,$$

then

$$f'(x) = 0.$$

The tangent line here is horizontal (all secant lines are also this horizontal line), and hence have slope 0.

(2) The derivative of a line

$$f(x) := mx + b$$

is

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

Again in this case, the tangent line (and all secant lines) are exactly the line $f(x)$, and hence has slope m.

(3) For every $r \in \mathbb{R}$, we have
$$\frac{d}{dx}x^r = rx^{r-1}.$$
Notice that (1) is the special case $r = 0$ and (2) (with $b = 0$) is the special case $r = 1$.

(4) If f is differentiable at $x = a$ and c is a constant, then
$$\frac{d}{dx}\left(cf(x)\right)\Big|_{x=a} = cf'(a).$$
Roughly speaking, this means that we can pull constants out of derivatives.

(5) If f and g are both differentiable at $x = a$, then
$$\frac{d}{dx}\left(f(x) + g(x)\right)\Big|_{x=a} = f'(a) + g'(a).$$
One often calls the last two properties linearity and says that derivatives are linear.

(6) If f and g are both differentiable at $x = a$, then
$$\frac{d}{dx}\left(f(x)g(x)\right)\Big|_{x=a} = f(a)g'(a) + f'(a)g(a).$$
This rule is known as the product rule.

(7) If f and g are both differentiable at $x = a$, then
$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right)\Big|_{x=a} = \frac{g(a)f'(a) - g'(a)f(a)}{g(a)^2}.$$
This rule is known as the quotient rule. One way to remember this rule is the following: The quotient rule is "Low d-high minus high d-low over low-low." This means that you take the low (low=denominator), times d-high (derivative of high/numerator) minus the high times d-low and then divide by low-squared (low-low).

Remark. If it is easier to remember, you can also write
$$\frac{f(x)}{g(x)} = f(x) \cdot \frac{1}{g(x)}$$
and then use the product rule instead. You will get the same answer.

(8) If $f(u)$ is differentiable at $u = g(a)$ and $g(x)$ is differentiable at $x = a$, then
$$\frac{d}{dx}\left(f \circ g\right)(x)\Big|_{x=a} = f'(g(a))g'(a).$$
This rule is known as the chain rule. Another way to write the chain rule is
$$\frac{dy}{dx} = \frac{dy}{dz}\frac{dz}{dx}.$$
Here we have $y = f(g(x))$ and $z = g(x)$. Roughly speaking, the chain rule says that you can take the derivative of f (ignoring what parameter is

inside) and then just plug in the parameter. However, afterwards, you have to multiply by the derivative of the parameter.

Remark. If the parameter is just x, then you get

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

Since $\frac{dx}{dx} = 1$, this is precisely $f'(x)$, as it should be.

Example 5.7. We illustrate some of the more complicated rules with an example. Find the derivatives of

$$\begin{aligned} f(x) &:= \sqrt{x}\left(x^3 - 2\right), \\ g(x) &:= \sqrt{x^4 + 2x}, \\ h(x) &:= \frac{x^3 - 2}{x^5 + 2x - 1}. \end{aligned}$$

Solution:

(1) Using the product rule (and the fact that $\sqrt{x} = x^{\frac{1}{2}}$), we have

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

(2) Using the chain rule (splitting as $\sqrt{u}$ with $u := x^4 + 2x$), we have

$$g'(x) = \frac{1}{2}\left(x^4 + 2x\right)^{-\frac{1}{2}}\left(4x^3 + 2\right).$$

(3) Using the quotient rule, we have

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

□

Definition 5.8. Repeated differentiation may be performed. Thinking of the derivative as the instantaneous change, the <u>second derivative</u> (the derivative of the derivative) is the instantaneous change in the slope. It is denoted

$$f''(x) := \frac{d}{dx}f'(x) = \frac{d}{dx}\left(\frac{d}{dx}f(x)\right) = \frac{d^2}{dx^2}f(x).$$

The <u>third derivative</u> is denoted $f'''(x)$, while higher derivatives are denoted $f^{(4)}(x)$, $f^{(5)}(x)$, ...

An example where the second derivative naturally occurs is gravity. When an object (say a ball) is thrown upwards, it has a force (gravity) pulling it back down. It is important to know the current location of the ball (the value), the speed at which the ball is travelling (derivative), and at what rate it is slowing down (second derivative). Gravity determines the rate at which it slows down. In this case, the second derivative is (essentially) constant. Slightly more complicated versions of this calculation are used to figure out how to land rockets onto planets, etc.

We will often need the derivatives of trigonometric functions. These are

$$\begin{aligned} \frac{d}{dx}\sin(x) &= \cos(x), & \frac{d}{dx}\cos(x) &= -\sin(x), \\ \frac{d}{dx}\tan(x) &= \sec^2(x), & \frac{d}{dx}\cot(x) &= -\csc^2(x), \\ \frac{d}{dx}\sec(x) &= \sec(x)\tan(x), & \frac{d}{dx}\csc(x) &= -\csc(x)\cot(x). \end{aligned}$$

5.4. **Implicit differentiation.** A useful tool for differentiation is implicit differentiation. Sometimes it is not easy to write a variable y explicitly in terms of another variable x. For example, suppose that

$$xy = \sin(y).$$

It is not very easy to determine y as a function of x. However, we may differentiate both sides of this equation:

$$\frac{d}{dx}(xy) = \frac{d}{dx}(\sin(y)).$$

Think of y as $y(x)$ (i.e., y is a function of x). We now use the product rule, chain rule, etc., as usual. This gives

$$\frac{dx}{dx}y + x\frac{dy}{dx} = \cos(y)\frac{dy}{dx}.$$

We used the product rule on the left-hand side and the chain rule on the right-hand side (together with the fact that $\frac{d}{dx}\sin(x) = \cos(x)$).

Now we plug in $\frac{dx}{dx} = 1$ and then solve for $\frac{dy}{dx}$. This is

$$\begin{aligned} & & y + x\frac{dy}{dx} &= \cos(y)\frac{dy}{dx} \\ &\Longrightarrow & \frac{dy}{dx}(\cos(y) - x) &= y \\ &\Longrightarrow & \frac{dy}{dx} &= \frac{y}{\cos(y) - x}. \end{aligned}$$

One might ask what the derivative is when $x = \frac{2}{\pi}$ and $y = \frac{\pi}{2}$ (check first that this solves the original equation!). Then we have

$$\frac{dy}{dx} = \frac{\frac{\pi}{2}}{0 - \frac{2}{\pi}} = -\frac{\pi^2}{4}.$$

Implicit differentiation is quite useful when you are doing experiments in science. One does a number of experiments (inputting an x) and determines the output (outputting a y). This might yield some relationship between x and y, but the exact formula describing this relationship might not be clear. An Engineer might then ask whether $\frac{dy}{dx}$ is ever too large (maybe something will break if it moves too quickly!). Given this data (and assuming the observed relationship is really true in general), one can determine $\frac{dy}{dx}$ using implicit differentiation.

5.5. **The Mean Value Theorem.** Let's say that you're driving on a road with a speed limit of 100 km/h. Let's say that you pass a checkpoint at 12:30pm and another 100 kilometers away at 1:20pm. You look at your spedometer as you pass the second checkpoint and it says that you are currently travelling at 80 km/h. However, a police officer pulls over your car and gives you a ticket for speeding. You argue that you were only travelling at 80 km/h, but he says that you must have been travelling over 100 km/h at some point. He explains that you travelled 100 kilometers in less than one hour, so your average speed was over 100 km/h. If you had always been travelling below 100 km/h, then you could not have gotten an average speed above 100 km/h. Unfortunately, his logic is sound and you cannot argue this.

This police officer has essentially employed The Mean Value Theorem in his explanation.

Theorem 5.9 (The Mean Value Theorem)**.** *If a function f is continuous on $[a, b]$ and differentiable on (a, b), then there exists some $c \in (a, b)$ such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Roughly speaking, the above theorem says that **at some point,** your speed was exactly your average speed. Although the police officer doesn't know **when** you were speeding, he knows that you were speeding at some point. Just like the Intermediate Value Theorem, you don't know what c is, only that it exists.

Note that there are conditions which need to hold for the Mean Value Theorem to hold. It assumes that the function is continuous (no Star Trek transporter has beamed you from one location to another). Also, the function must be differentiable. To give an example where differentiability is important, consider

$$f(x) = |x|.$$

We have $f(1) = f(-1)$, so the average of the derivative on $[-1, 1]$ is zero. However,

$$f'(x) = \begin{cases} -1 & \text{if } x < 0, \\ 1 & \text{if } x > 0. \end{cases}$$

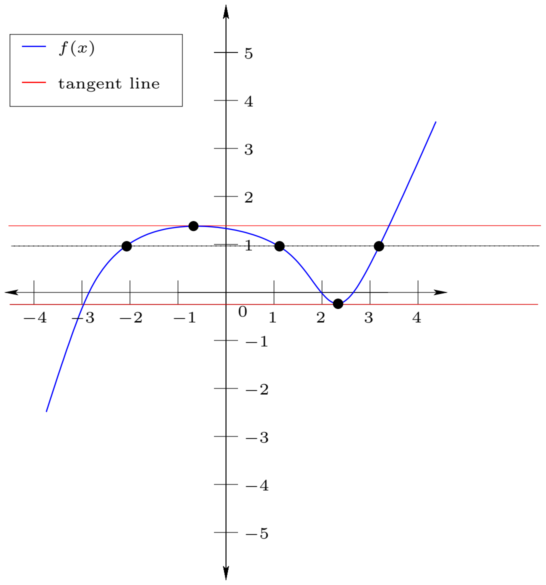
We see that there is no $c \in (-1, 1)$ for which $f'(c) = 0$.

Let's consider the special case of the Mean Value Theorem when $f(a) = f(b)$.

Theorem 5.10 (Rolle's Theorem)**.** *If a function f is continuous on $[a, b]$ and differentiable on (a, b) and $f(a) = f(b)$, then there exists some $c \in (a, b)$ such that*

$$f'(c) = 0.$$

Why might Rolle's Theorem be special enough to consider it separately? Remember that $f'(c) = 0$ means that the tangent line is horizontal (these are sometimes called critical points).



If you look at the above graph, you see that these horizontal points are (locally, meaning in some small interval around them) the maximum and minimum value for the function. Think back to the example of the stock broker. If the stock broker knew the exact points where the stock value would be lowest (at a critical point) and the exact point where it would be the highest (also at a critical point), then they would know exactly when to buy and sell. So finding critical points is important for finding important places where the direction of the stock might change.

5.6. **L'Hôpital's Rule.** In the last section, we discussed a number of infinite limits and limits towards infinity. Recall that

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

provided that both limits exist and that $\lim_{x\to a} g(x) \neq 0$. This begs one to question: What happens when $\lim_{x\to a} g(x) = 0$? Is there an easy way to determine the answer in this case?

Consider the simple examples

$$\frac{x^2 - x}{x-1}, \quad \frac{x-1}{x^2-2x+1}, \quad \text{and} \quad \frac{x^2-2x-1}{x-1}.$$

In each of these cases, the limit as $x \to 1$ of the denominator vanishes. We therefore cannot use the above rule. However, by cancelling a common factor, we have

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

$$\lim_{x\to 1}\frac{x-1}{x^2-2x+1}=\lim_{x\to 1}\frac{1}{x-1},$$

which does not exist, and

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

We therefore were able to construct examples where the answer was 0, 1, and even one where the limit does not exist (the left-limit and right-limit are $-\infty$ and $+\infty$, respectively).

Hence plugging in and getting $\frac{0}{0}$ doesn't tell us anything about what the limit might be. However, it there is still a useful rule which allows us to compute liits when directly plugging in yields $\frac{0}{0}$.

Theorem 5.11 (L'Hôpital's rule)**.** *Suppose that f and g are differentiable and that both $\lim_{x\to a} f(x)=0$ and $\lim_{x\to a} g(x)=0$. Then*

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

Warning. If plugging in **does not yield** $\frac{0}{0}$, then you **may not** use L'Hôpital's rule. To give an example where this can be seen, consider

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

Here the limit is zero (you can get this by plugging in directly). However, if you tried to (illegally!) use L'Hôpital's rule, you would obtain

$$\lim_{x\to 1}\frac{1}{1}=1,$$

an incorrect answer.

Example 5.12. Compute

$$\lim_{x\to 0}\frac{\sin(x)}{x}$$

and

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

Solution: Since $\sin(0)=0$, plugging $x=0$ in directly gives $\frac{0}{0}$, so we may use L'Hôpital's rule. Since $\sin'(x)=\cos(x)$, we hence have

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

Since $\cos(0)=1$ and $\frac{d}{dx}\cos(x)=-\sin(x)$, we have

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

We could again use L'Hôpital's rule or directly plug in

$$\lim_{x\to 0}\frac{\sin(x)}{x}=1,$$

which we've already computed. Therefore

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

$\square$

Example 5.13. Compute

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

Solution: We use L'Hôpital's rule twice to obtain

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

Of course, this is not the only way to compute the limit. Since

$$\lim_{x\to 0}\frac{\sin(x)}{x}=1,$$

we have

$$\lim_{x\to 0}\frac{\sin^2(x)}{x^2}=\lim_{x\to 0}\left(\frac{\sin(x)}{x}\right)^2=\left(\lim_{x\to 0}\frac{\sin(x)}{x}\right)\left(\lim_{x\to 0}\frac{\sin(x)}{x}\right)=1.$$

$\square$

Example 5.14. Compute

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

Solution: In this case, you can just plug $x=0$ in (from the original properties we learned for the limit), because

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

Therefore,

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

Had you tried to use L'Hôpital's rule (which is not allowed because the plugging in doesn't give $\frac{0}{0}$), you would have gotten

$$\lim_{x\to 0}\frac{\cos(x)}{-2\cos(x)\sin(x)}=-\lim_{x\to 0}\frac{1}{2\sin(x)},$$

which doesn't exist (the limit $x \to 0^+$ is $-\infty$ and the limit $x \to 0^-$ is $+\infty$. □

There is also a useful version of L'Hôpital's rule when the limits are both infinite.

Theorem 5.15 (L'Hôpital's rule for $\frac{\infty}{\infty}$)**.** *Suppose that f and g are differentiable and that both $\lim_{x\to a} f(x) = \pm\infty$ and $\lim_{x\to a} g(x) = \pm\infty$. Then*

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

Example 5.16. Compute

$$\lim_{x\to\infty} \frac{\sqrt{2\sqrt{x}+7}}{\sqrt{x}+1}$$

Solution: We first note that

$$\lim_{x\to\infty} \sqrt{2\sqrt{x}+7} = \infty$$

and

$$\lim_{x\to\infty} \left(\sqrt{x}+1\right) = \infty.$$

Noting that

$$\frac{d}{dx}\sqrt{x} = \frac{1}{2\sqrt{x}},$$

L'Hôpital's rule for $\frac{\infty}{\infty}$ yields

$$\lim_{x\to\infty} \frac{\sqrt{2\sqrt{x}+7}}{\sqrt{x}+1} = \lim_{x\to\infty} \frac{\frac{1}{2\sqrt{2\sqrt{x}+7}}\left(\frac{2}{2\sqrt{x}}\right)}{\frac{1}{2\sqrt{x}}} = \lim_{x\to\infty} \frac{1}{\sqrt{2\sqrt{x}+7}}.$$

Since

$$\lim_{x\to\infty} \sqrt{2\sqrt{x}+7} = \infty,$$

we conclude that

$$\lim_{x\to\infty} \frac{1}{\sqrt{2\sqrt{x}+7}} = 0.$$

Therefore

$$\lim_{x\to\infty} \frac{\sqrt{2\sqrt{x}+7}}{\sqrt{x}+1} = 0.$$

□

Example 5.17. Compute

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

Solution: Using
$$\csc(x) = \frac{1}{\sin(x)},$$
we rewrite
$$x \sin\left(\frac{1}{x}\right) = \frac{x}{\frac{1}{\sin\left(\frac{1}{x}\right)}} = \frac{x}{\csc\left(\frac{1}{x}\right)}.$$
Since $\sin(0) = 0$, we have
$$\lim_{x\to\infty} \csc\left(\frac{1}{x}\right) = +\infty$$
Furthermore, by the chain rule
$$\frac{d}{dx}\csc(x) = \frac{d}{dx}\frac{1}{\sin(x)} = -\frac{1}{\sin^2(x)}(\cos(x)) = -\csc(x)\cot(x).$$
Again using the chain rule, we obtain
$$\frac{d}{dx}\csc\left(\frac{1}{x}\right) = -\csc\left(\frac{1}{x}\right)\cot\left(\frac{1}{x}\right)\left(-\frac{1}{x^2}\right).$$
Therefore by L'Hôpital's for $\frac{\infty}{\infty}$, we have
$$\lim_{x\to\infty}\frac{x}{\csc\left(\frac{1}{x}\right)} = \lim_{x\to\infty}\frac{1}{\frac{1}{x^2}\csc^2\left(\frac{1}{x}\right)\sec\left(\frac{1}{x}\right)} = \lim_{x\to\infty}\frac{x^2}{\csc^2\left(\frac{1}{x}\right)\sec\left(\frac{1}{x}\right)}.$$
The situation seems to actually have gotten much worse (the power of x in the numerator is higher!). Usually this happens when you made the wrong choice. I'll explain what you should try next:

Maybe we want to instead write
$$x\sin\left(\frac{1}{x}\right) = \frac{\sin\left(\frac{1}{x}\right)}{\frac{1}{x}}.$$
We now use L'Hôpital's for $\frac{0}{0}$. This gives
$$\begin{aligned}\lim_{x\to\infty} x\sin\left(\frac{1}{x}\right) &= \lim_{x\to\infty}\frac{\sin\left(\frac{1}{x}\right)}{\frac{1}{x}}\\ &= \lim_{x\to\infty}\frac{\cos\left(\frac{1}{x}\right)\left(-\frac{1}{x^2}\right)}{-\frac{1}{x^2}}\\ &= \lim_{x\to\infty}\cos\left(\frac{1}{x}\right) = 1.\end{aligned}$$
□

Example 5.18. Determine
$$\lim_{x\to 0}\left(\frac{1}{x} + \frac{1}{\cos(x) - 1}\right).$$

Solution: Plugging in gives $\infty - \infty$, so we cannot directly determine the limit. We rewrite

$$\frac{1}{x} + \frac{1}{\cos(x) - 1} = \frac{\cos(x) - 1 + x}{x\left(\cos(x) - 1\right)}.$$

Thus

$$\lim_{x \to 0} \left(\frac{1}{x} + \frac{1}{\cos(x) - 1} \right) = \lim_{x \to 0} \frac{\cos(x) - 1 + x}{x\left(\cos(x) - 1\right)}.$$

This is now of the form $\frac{0}{0}$ and we may hence use L'Hôpital's rule. We compute the derivative

$$\frac{d}{dx}\left(x\left(\cos(x) - 1\right)\right) = x\sin(x) + \cos(x) - 1.$$

to obtain

$$\lim_{x \to 0} \frac{\cos(x) - 1 + x}{x(\cos(x) - 1)} = \lim_{x \to 0} \frac{-\sin(x) + 1}{-x\sin(x) + \cos(x) - 1}.$$

The limit of the numerator is 1. To determine the limit of the denominator, we plug in $x = 0$ to obtain

$$\lim_{x \to 0} x\sin(x) = 0,$$
$$\lim_{x \to 0} \cos(x) - 1 = 0.$$

The limit is hence infinite. However, how do we determine whether it is $+\infty$ or $-\infty$? This depends on whether $-x\sin(x) + \cos(x) - 1$ is positive or negative as x approaches 0. For $x \neq 0$ and

$$-\frac{\pi}{2} < x < \frac{\pi}{2},$$

we have

$$-x\sin(x) < 0,$$

while for every $x \in \mathbb{R}$ we have

$$\cos(x) - 1 \leq 0.$$

Therefore, for $-\frac{\pi}{2} < x < \frac{\pi}{2}$ and $x \neq 0$, we have

$$-x\sin(x) + \cos(x) - 1 < 0.$$

It follows that

$$\lim_{x \to 0} \left(\frac{1}{x} + \frac{1}{\cos(x) - 1} \right) = -\infty.$$

□

5.7. **Curve sketching and optimization.** We've spoken a bit about what the derivative **means** from the point of view of the graph of a function. As we described, this is the slope of the tangent line (or the instantaneous rate of change). The special case when $f'(x) = 0$ was discussed when we looked at Rolle's Theorem.

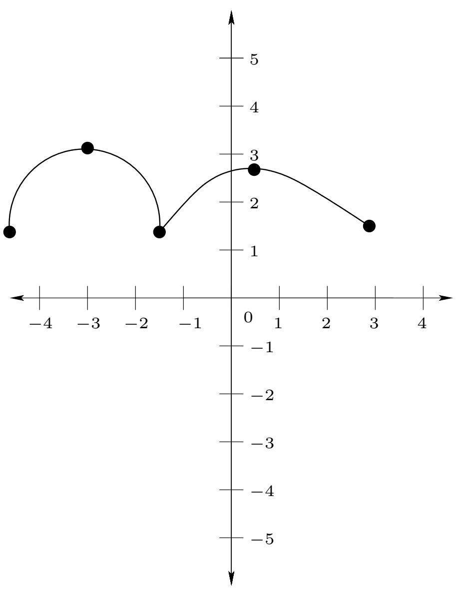
The interplay between derivatives of function and the graph of the function is an important one. Given a graph, one can identify certain special points and general properties of the function which help to understand the general structure of the function. Conversely, given a function, one can understand it by sketching its graph.

5.7.1. *Critical points and increasing/decreasing behavior.* When you look at the graph of a function, what stands out to you? The first thing that might be obvious are the points of discontinuity. You might also notice the points where the derivative does not exist (think about the graph of $f(x) = |x|$). The next important thing that you might notice are the places where the graph changes direction. Together, these points are known as the critical points.

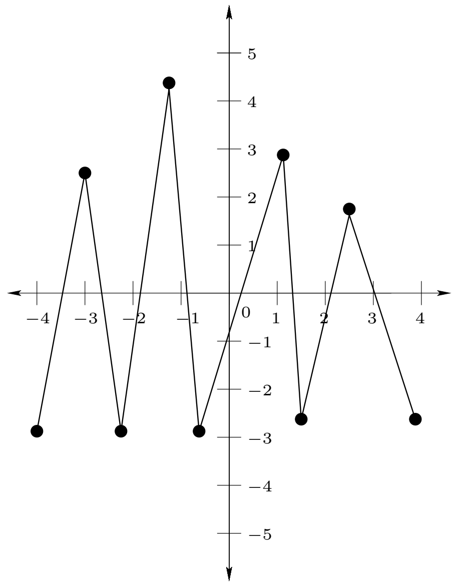
Definition 5.19. For an interval (a, b), we call $c \in (a, b)$ a critical point of f if either f is not differentiable at c or

$$f'(c) = 0.$$

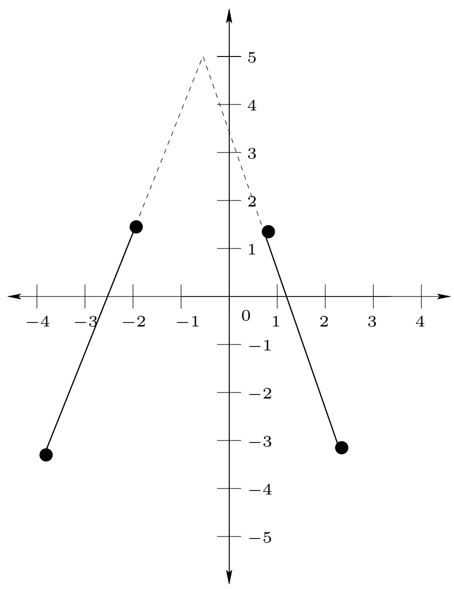
The critical points are important for finding the maximum and minimum values in an interval. Think about the top of a mountain or a hill. At the very top, the tangent line of a hill is horizontal.



The top of a mountain comes to a point (like $|x|$).



In each of these cases, the top (the highest point) is a critical point. If you have a bunch of mountains next to each other, then the top of each mountain is a critical point. To find the tallest mountain, you just have to compare the height of each of the critical points (the highest point on each mountain). The low points are also critical points, so you will find the lowest point of each mountain this way, too. Of course, if you are climbing the mountain, the highest point you've reached might not be the top. However, if you consider the part that you've climbed as an interval, the highest point will be one of the endpoints of the interval.



Therefore, if you want to find the highest (or lowest) point in an interval, you should check the endpoints and the critical points.

Definition 5.20. A function f is said to have a local maximum/minimum at the point $x = c$ if there is some small interval $[a, b]$ with $c \in (a, b)$ for which $f(c)$ is the largest/smallest value of $f(x)$ for all $x \in [a, b]$.

A function f is said to have a global maximum in its domain D at the point $x = c$ if $f(c) \geq f(x)$ for every $x \in D$. The function f is said to have a global minimum in its domain D at the point $x = c$ if $f(c) \leq f(x)$ for every $x \in D$.

The top of each mountain is a local maximum, while only the top of the tallest mountain is a global maximum.

After finding the top and bottom points where the graph changes direction, it is also notable to mark when the graph is going up and when it is going down. A function f is said to be increasing on an interval (a, b) if for any $a < r < s < b$ we have

$$f(s) \geq f(r).$$

If

$$f(s) > f(r)$$

is always satisfied, then we say that f is strictly increasing. Likewise, it is said to be decreasing if for any $a < r < s < b$ we have

$$f(s) \leq f(r).$$

Again, we say that f is strictly decreasing if $f(s) < f(r)$ always holds.

Suppose now that f is differentiable. Since

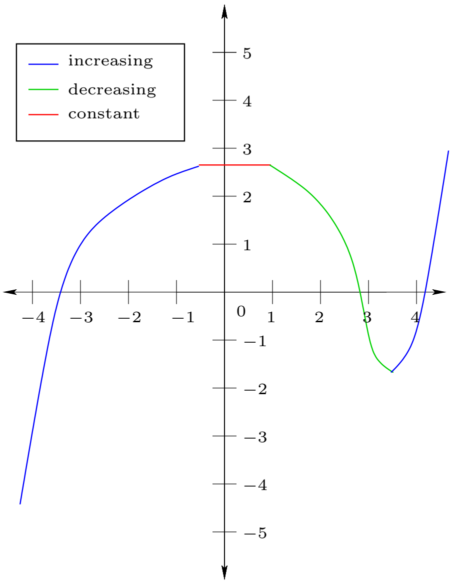
$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h},$$

if $f(x + h) > f(x)$ for every $h > 0$ (meaning that f is strictly increasing), then $f'(x) > 0$. Analyzing in this way, we obtain the following

Theorem 5.21. *Suppose that f is differentiable on an open interval (a, b).*

(1) *If $f'(x) > 0$ for every $x \in (a, b)$, then f is strictly increasing on (a, b).*
(2) *If $f'(x) < 0$ for every $x \in (a, b)$, then f is strictly decreasing on (a, b).*
(3) *If $f'(x) = 0$ for every $x \in (a, b)$, then f is constant on (a, b).*

We can hence determine the points where a function is increasing or decreasing by determining whether its derivative is positive or negative. The points where the function changes from increasing to decreasing (or decreasing to increasing) are exactly the local maxima.



Example 5.22. Find all of the critical points of $f(x) := \frac{1}{4}x^4 - x$. Also determine when it is increasing and decreasing. Determine the local maxima and minima and the global maximum and minimum, if they exist. Give a rough sketch of its graph.

Solution: Since f is a polynomial, it is always differentiable. We have

$$f'(x) = x^3 - 1.$$

The critical points are hence those x for which

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

This happens exactly when $x = 1$. For $x < 1$, we have $x^3 < 1$, so that

$$f'(x) = x^3 - 1 < 0.$$

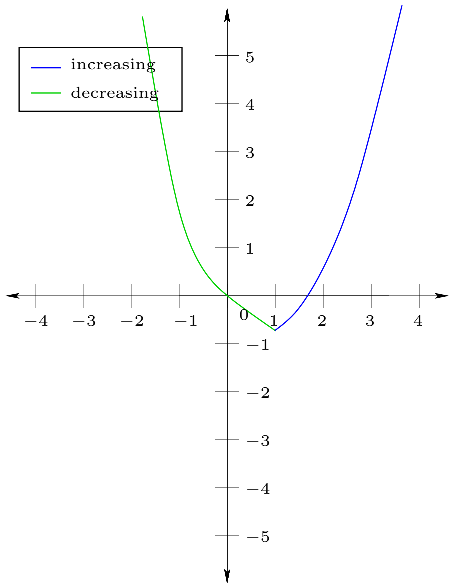
If $x > 1$, then

$$f'(x) = x^3 - 1 > 0.$$

Therefore we have that f is decreasing for $x < 1$ and increasing whenever $x > 1$. Therefore, at $x = 1$, the function f has a local minimum. Since

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

there is no global maximum. At $x = 1$, the function f has a global minimum (the minimum value is $f(1) = -\frac{3}{4}$). This yields the following sketch of the graph.



□

5.7.2. *Asymptotes.* Another property of the graph which might stick out is the limit as $x \to \pm\infty$ and any point where there is a discontinuity. This leads to the following definitions.

Definition 5.23. If $\lim_{x\to\infty} f(x) = L$ or $\lim_{x\to-\infty} f(x) = L$, then we say that the line $y = L$ is a horizontal asymptote to the curve $y = f(x)$.

If $\lim_{x\to a^+} f(x) = \pm\infty$ or $\lim_{x\to a^-} f(x) = \pm\infty$, then we call the line $x = a$ a vertical asymptote to the curve $y = f(x)$.

Example 5.24. What are the horizontal and vertical asymptotes for $f(x) := \frac{1}{x-1}$?

Solution: We have

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

and

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

so $y = 0$ is a horizontal asymptote.

Since

$$\lim_{x\to 1^-} f(x) = -\infty$$

and

$$\lim_{x\to 1^+} f(x) = +\infty,$$

we have that $x = 1$ is a vertical asymptote. □

Example 5.25. Find the vertical and horizontal asymptotes of

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

Sketch the graph.

Solution: We have

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

Therefore $y=2$ is a horizontal asymptote.

To determine the vertical asymptotes, we must determine when the denominator equals zero. This is precisely when

$$x^2-1=0,$$

or $x=\pm1$.

In order to sketch the graph, we determine the limits

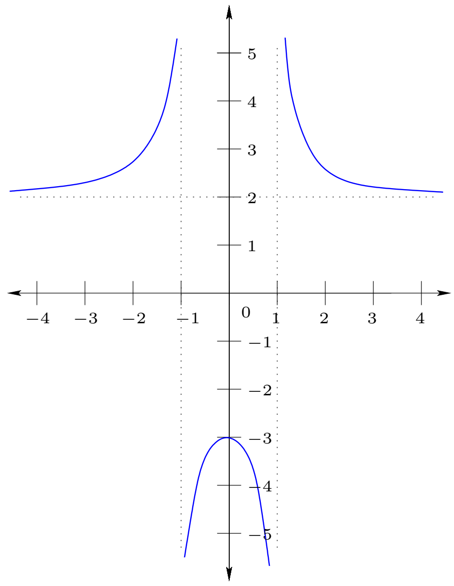
$$\begin{aligned}\lim_{x\to-1^-}\frac{2x^2+3}{x^2-1}&=+\infty,\\ \lim_{x\to-1^+}\frac{2x^2+3}{x^2-1}&=-\infty,\\ \lim_{x\to1^-}\frac{2x^2+3}{x^2-1}&=-\infty,\\ \lim_{x\to1^+}\frac{2x^2+3}{x^2-1}&=+\infty,\end{aligned}$$

because $x^2-1>0$ whenever $|x|>1$ and $x^2-1<0$ whenever $|x|<1$ (and $2x^2+3>0$ always).

To sketch the graph, it is also useful to know when the function is increasing and decreasing. We compute the derivative

$$f'(x)=\frac{(x^2-1)(4x)-(2x^2+3)(2x)}{(x^2-1)^2}=-\frac{10x}{(x^2-1)^2}.$$

Therefore $f'(x)=0$ only when $x=0$, $f'(x)>0$ for $x<0$, and $f'(x)<0$ for $x>0$. Putting this all together gives the following sketch of the graph.



□

Another kind of asymptote occurs when $f(x)$ gets closer and closer to a line as x gets bigger. We say that $y = mx + b$ is an oblique asymptote for $y = f(x)$ if

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

or

$$\lim_{x\to-\infty} (f(x) - (mx + b)) = 0.$$

Theorem 5.26. *A line $y = mx+b$ is an oblique asymptote for $y = f(x)$ if and only if*

$$\lim_{x\to\pm\infty} \frac{f(x)}{x} = m$$

and

$$\lim_{x\to\pm\infty} (f(x) - mx) = b.$$

Example 5.27. Find the oblique asymptote(s) (if they exist) for $f(x) := \frac{x^2-1}{x+2}$.

Solution: We take the limit

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

We then compute

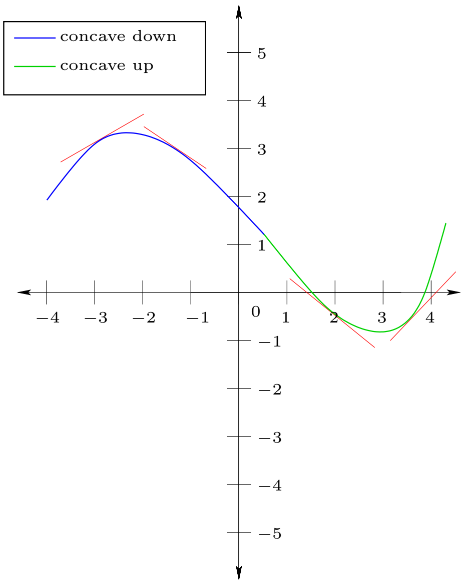
$$\lim_{x\to\infty} (f(x) - x) = \lim_{x\to\infty} \left(\frac{x^2 - 1}{x + 2} - \frac{x^2 + 2x}{x + 2} \right) = - \lim_{x\to\infty} \frac{2x + 1}{x + 2} = -2.$$

Therefore $y = x - 2$ is an oblique asymptote for $f(x)$. □

5.7.3. *Convexity and points of inflection.* We used the derivative to find interesting information about when a function increases or decreases and where its maxima and minima occur. The second derivatve also contains information about the shape of the function.

Think about the shape of a bowl on your dining room table (or a spoon, with the bottom of the spoon on the table) with some food in it. The shape of the bowl is curved upwards. This is known as concave up (also called convex). If you turn the bowl (or spoon) upside down, then it is curved downwards. This is known as concave down (in some books, this is simply called concave).

To be more precise, if the tangent line is underneath the curve in some small interval, then it is concave up. If the tangent line is above the curve in some small interval, then it is concave down. You can see this in the picture below.



Remark. Another way to write the condition for convexity is the following: A function is convex (concave up) on the interval (a, b) if for every $0 < t < 1$ and $x, y \in (a, b)$ we have

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y).$$

Roughly speaking, this inequality means that the value is less than the (weighted) average (think about $t = \frac{1}{2}$). Here "weighted" means that the closer you are to x, the closer the value should be to $f(x)$, so x has more "weight" in the average when you are close to it..

The line

$$(1-t)f(x) + tf(y)$$

is actually the secant line between the points $(x, f(x))$ and $(y, f(y))$ (plug in $t = 0$ and $t = 1$ to see that these two point are on the line), so another equivalent statement is that f is concave up (convex) in an interval if the curve is underneath all of the secant lines in the interval. It is concave down (concave) if the curve is above all of the secant lines.

Now think of the derivative as the rate of change of the function (the speed of a car, for example). Whenever a function is concave up, the rate of change is increasing (the slope of the tangent line is increasing – look at the graph above). Whenever the function is concave down, the rate of change is decreasing. In the example of a car, being concave up therefore means that the speed is increasing, or in other words, the car is accelerating. Concave down means that the speed is decreasing, i.e, the car is decelerating.

Here the discussion is about the rate of change of the rate of change (the change of the speed). But this is then the **second derivative** (the derivative of the derivative is the rate of change of the derivative).

Theorem 5.28. *Suppose that f is twice differentiable on the interval (a, b).*

(1) *If $f''(x) \geq 0$ for every $x \in (a, b)$, then f is concave up on (a, b).*
(2) *If $f''(x) \leq 0$ for every $x \in (a, b)$, then f is concave down on (a, b).*

The points where the convexity changes are also important.

Definition 5.29. If f is continuous at $c \in \mathbb{R}$ and f changes convexity (i.e., from concave up to concave down or concave down to concave up) between $x < c$ and $x > c$, then we call c a point of inflection of f.

By Theorem 5.28 the convexity changes precisely when the sign of the second derivative changes.

Example 5.30. Find the points of inflection points of

$$f(x) := x^3 - 2x$$

and

$$g(x) := x^4.$$

Solution: We take the second derivatives

$$f''(x) = 6x$$

and

$$g''(x) = 12x^2.$$

When the sign of the the second derivative changes, we must pass through a point where the second derivative equals zero. Thus for f, this could only happen when

$$6x = 0 \implies x = 0.$$

It is easy to check that for $x < 0$ we have $f''(x) < 0$ and for $x > 0$ we have $f''(x) > 0$, so f indeed has a point of inflection at $x = 0$.

For $g''(x)$, we can only have a point of inflection at $x = 0$. However, for $x < 0$ we have $g''(x) > 0$ and for $x > 0$ we also have $g''(x) > 0$. Therefore, the sign of the second derivative doesn't change and hence g has no points of inflection. $\square$

We are now ready to sketch a graph. The following process will help to accurately sketch the graph:

(1) Determine the natural domain of f and determine where f is continuous/not continuous.
(2) Determine the asymptotes of f, if any.
(3) Determine the y-intercept (and any x-intercepts, if possible).
(4) Find $f'(x)$ and $f''(x)$.
(5) Find critical points (i.e., those points where the derivative doesn't exist or the derivative is zero) by setting $f'(x) = 0$.
(6) Between each two adjacent critical points, determine whether the graph is increasing/decreasing. Also determine whether the function is increasing or decreasing as $x \to -\infty$ and $x \to \infty$.
(7) Determine whether critical points are local minima, local maxima or neither.
(8) Find possible points of inflection (where the second derivative doesn't exist or $f''(x) = 0$.
(9) Compute the convexity between any two possible points of inflection.

Example 5.31. Use these techniques to sketch a graph of

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

Solution:

(1) The natural domain is $(-\infty, -1) \cup (-1, \infty)$, since the denominator equals zero when $x = -1$. There is a point of discontinuity at $x = -1$ and no other points of discontinuity.
(2) There is a vertical asymptote at $x = -1$ because the denominator is zero but the numerator is not. To compute the horizontal asymptote, we take the limits

$$\lim_{x\to\infty} f(x) = \lim_{x\to\infty} \frac{1}{1 + \frac{1}{x^3}} = 1$$

and

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

Therefore, there is a horizontal asymptote at $y = 1$.
(3) The y-intercept is $f(0) = 0$. The only x-intercept is when the numerator is zero, which occurs for $x = 0$.
(4) To easier compute the derivatives, we first rewrite

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

We compute (using the chain rule, but one could also do this with the quotient rule)

$$f'(x) = \frac{3x^2}{\left(x^3+1\right)^2}$$

We now use the quotient rule to compute

$$\begin{aligned} f''(x) &= \frac{6x\left(x^3+1\right)^2 - 2\left(x^3+1\right)\left(3x^2\right)^2}{\left(x^3+1\right)^4} \\ &= \frac{6x\left(x^3+1\right) - 18x^4}{\left(x^3+1\right)^3} \\ &= \frac{6x\left(-2x^3+1\right)}{\left(x^3+1\right)^3}. \end{aligned}$$

(5) The critical points are $x = -1$ (where the function is discontinuous) and whenever $f'(x) = 0$. The derivative is zero whenever

$$\frac{3x^2}{\left(x^3+1\right)^2} = 0 \Leftrightarrow x = 0.$$

(6) We look at the intervals $(-\infty, -1)$, $(-1, 0)$, and $(0, \infty)$. To determine whether the function is increasing or decreasing, we only need to plug in one point **(the sign doesn't change in each interval, because otherwise there would be another critical point!)**. For the interval $(-\infty, -1)$ we plug in $x = -2$ to obtain

$$f'(-2) = \frac{3(-2)^2}{\left((-2)^3+1\right)^2} = 12 > 0.$$

For the interval $(-1, 0)$ we plug in $x = -\frac{1}{2}$ to obtain

$$f'\left(-\frac{1}{2}\right) = \frac{3\left(-\frac{1}{2}\right)^2}{\left(\left(-\frac{1}{2}\right)^3+1\right)^2} = \frac{\frac{3}{4}}{\left(\frac{7}{8}\right)^2} > 0.$$

In the interval from 0 to ∞, we plug in $x = 2$ to obtain

$$f'(2) = \frac{3(2)^2}{\left((2)^3+1\right)^2} = \frac{12}{81} > 0.$$

We collect these in an small table:

	$(-\infty, -1)$	$(-1, 0)$	$(0, \infty)$
Sign of f'	$+$	$+$	$+$
Behavior of f	increasing	increasing	increasing

(7) The point $x = 0$ is not a local maximum or minimum, because the function is increasing both in the intervals $(-1, 0)$ and $(0, \infty)$.

(8) We set

$$f''(x) = 0$$

to determine that the possible points of inflection are at $x = -1$ and whenever

$$6x\left(-2x^3 + 1\right) = 0 \Leftrightarrow x \in \left\{0, \sqrt[3]{\frac{1}{2}}\right\}.$$

(9) We compute the convexity in the intervals $(-\infty, -1)$, $(-1, 0)$, $\left(0, \sqrt[3]{\frac{1}{2}}\right)$, and $\left(\sqrt[3]{\frac{1}{2}}, \infty\right)$.

In the interval $(-\infty, -1)$, we plug in $x = -2$ to obtain

$$f''(-2) = \frac{-12 \cdot 17}{\left((-2)^3 + 1\right)^3} = \frac{204}{7^3} > 0.$$

In the interval $(-1, 0)$, we plug in $x = -\frac{1}{2}$ to obtain

$$f''\left(-\frac{1}{2}\right) = \frac{-3\left(\frac{5}{4}\right)}{\left(\frac{7}{8}\right)^3} < 0.$$

In the interval $\left(0, \sqrt[3]{\frac{1}{2}}\right)$, we plug in $\sqrt[3]{\frac{1}{1000}} = \frac{1}{10}$ to obtain

$$f''\left(\frac{1}{10}\right) = \frac{\frac{3}{5}\left(\frac{499}{500}\right)}{\left(\frac{1001}{1000}\right)^3} > 0.$$

Finally, for the interval $\left(\sqrt[3]{\frac{1}{2}}, \infty\right)$, we plug in $x = 1$ to obtain

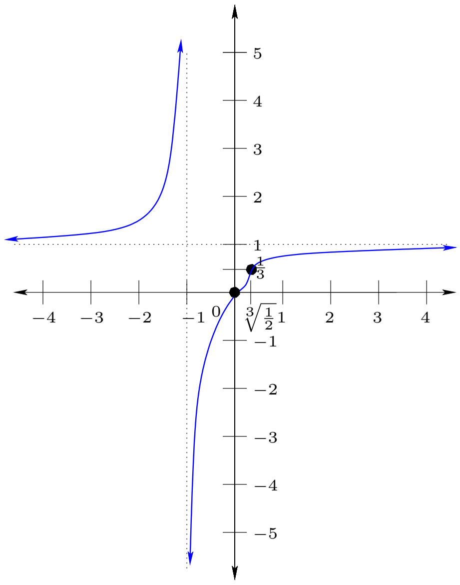
$$f''(1) = -\frac{3}{4} < 0.$$

We collect these in an small table:

	$(-\infty, -1)$	$(-1, 0)$	$\left(0, \sqrt[3]{\frac{1}{2}}\right)$	$\left(\sqrt[3]{\frac{1}{2}}, \infty\right)$
Sign of f''	$+$	$-$	$+$	$-$
Convexity of f	concave up	concave down	concave up	concave down

We see that there are changes in concavity at $x = -1$, $x = 0$, and $x = \sqrt[3]{\frac{1}{2}}$.

We are now ready to sketch the graph, making sure that all of the above data is on the sketch.



Notice that the convexity changes when we pass the asymptote at $x = -1$ and at the points $(0, 0)$ and $\left(\sqrt[3]{\frac{1}{2}}, \frac{1}{3}\right)$. It is not very important that $\sqrt[3]{\frac{1}{2}}$ is at precisely the right place, but it is better to mark this point explicitly on the axis.

Also notice that the graph never passes the asymptote $y = 1$. One could determine this in one of two ways: Firstly, you could try to solve for $y = 1$ and see directly that it has no solution. Secondly, you could use the fact that the function is always increasing. Since it is always increasing and $\lim_{x\to\infty} f(x) = 1$, if f went above the line, then it would have to come back down to have the limit equalling one. As the function would then have to decrease, this is impossible. Similarly, the limit $\lim_{x\to-\infty} f(x) = 1$ may be used to determine that it cannot start below the asymptote and then the fact that it is increasing precludes the possibility of the graph crossing the line $y = 1$. $\square$

5.8. **Taylor approximation and error estimation.** Consider again the difference quotient

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

For h very small, this is extremely close to $f'(a)$, assuming that the limit exists. Now recall that $f'(a)$ is the slope of the tangent line at the point $(a, f(a))$.

Recall that to determine the equation for a line it is enough to know the slope and one point on the line. In particular, if the slope is m and the point (A, B) lies on the line, then

$$y = m(x - A) + B.$$

Therefore, the tangent line $y = P_1(x)$ can be written

$$P_1(x) = f'(a)(x - a) + f(a).$$

For $x = a + h$ we have

$$P_1(x) = f'(a)(h) + f(a).$$

If h is very small, then since

$$f'(a) \approx \frac{f(a+h) - f(a)}{h},$$

we have

$$\begin{aligned} P_1(x) &\approx \frac{f(a+h) - f(a)}{h}(h) + f(a) \\ &= f(a+h). \end{aligned}$$

Therefore the line $y = P_1(x)$ closely approximates $f(a+h)$ whenever h is very small.

The idea to use this as an approximation is what is known as linear approximation.

Example 5.32. Find the linear approximation $P_1(x)$ for $f(x) := x^3 - 2x^2$ near $x = 2$.

Solution: We have

$$f'(x) = 3x^2 - 4x,$$

so that in particular

$$f'(2) = 12 - 8 = 4.$$

Moreover,

$$f(2) = 0.$$

Therefore, for x near 2 we have

$$f(x) \approx 4\,(x - 2)\,.$$

To see that this really does closely approximate f, let's try $x = 2.01$. For this, we have

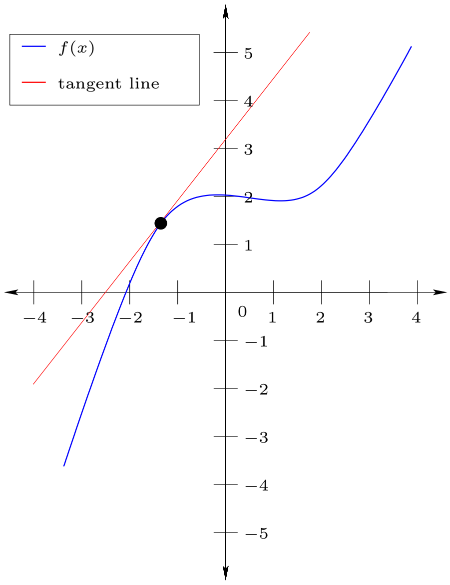
$$f(2.01) = 0.0404001.$$

The approximation we obtained by the tangent line is

$$f(2.01) \approx 0.04.$$

We see that the approximation is pretty good. □

Of course, for h very large, this approximation is not very good, as evidenced in the following diagram:



It is natural to ask whether one can get better approximations. Since the second derivative also contains information about the concavity of f maybe packaging this information into a second approximation would be helpful (a line has no concavity because the second derivative is always zero). In particular, one obtains a quadratic approximation by

$$P_2(x) := f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2$$

This is approximately equal to $P_1(x)$ near $x = a$, but also matches the concavity of f nearby $x = a$, so it approximates the function better. Of course, a quadratic function has fixed concavity (it is either always concave up or always concave down, because the second derivative is constant). For the concavity to change, one would need a higher order polynomial. Continuing in this way, one is naturally led to define the Taylor polynomial of order n of f at $x = a$ by

$$P_n(x) := f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n .$$

Here $n!$ is the factorial, which is recursively defined by $0! = 1$ and

$$(n+1)! := n!(n+1).$$

It may also be written

$$n! = 1 \cdot 2 \cdot 3 \cdots n.$$

Example 5.33. Find the Taylor polynomial of order 3 of $f(x) := \sin(x)$ at $x = 0$.

Solution: We compute

$$\begin{aligned}
f(0) &= 0, \\
f'(x) &= \cos(x) \\
\implies f'(0) &= 1, \\
f''(x) &= -\sin(x) \\
\implies f''(0) &= 0, \\
f'''(x) &= -\cos(x) \\
\implies f'''(0) &= -1.
\end{aligned}$$

Therefore

$$P_3(x) = x - \frac{x^3}{6}.$$

□

Example 5.34. Find the second Taylor polynomial for $f(x) := \sqrt{x}$ at $x = 1$.

Solution: We have

$$\begin{aligned}
f(1) &= 1, \\
f'(x) &= \frac{1}{2\sqrt{x}} \\
\implies f'(1) &= \frac{1}{2}, \\
f''(x) &= -\frac{1}{4x^{\frac{3}{2}}} \\
\implies f''(1) &= -\frac{1}{4}.
\end{aligned}$$

Therefore

$$\begin{aligned}
P_2(x) &= 1 + \frac{1}{2}(x-1) + \frac{-\frac{1}{4}}{2}(x-1)^2 \\
&= 1 + \frac{1}{2}(x-1) - \frac{1}{8}(x-1)^2.
\end{aligned}$$

□

As alluded to earlier, the Taylor polynomial at $x = a$ gives a pretty good approximation to $f(x)$ for x near a. It is useful to get an idea about the error between the approximation $P_n(x)$ and $f(x)$. We hence define the error of the nth Taylor approximation (also known as the remainder)

$$R_n(x) := f(x) - P_n(x).$$

The following theorem allows us to get a good bound on how large $R_n(x)$ can be.

Theorem 5.35 (Taylor's Theorem)**.** *Suppose that the* $(n+1)$*th derivative exists in an open interval* (A,B) *containing* a*. Then for every* $x \in (A,B)$*, there exists* c *between* x *and* a *such that*

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}.$$

In other words, there is some c *between* a *and* x *such that*

$$f(x) = f(a)+f'(a)(x-a)+\frac{f''(a)}{2}(x-a)^2+\cdots+\frac{f^{(n)}(a)}{n!}(x-a)^n+\frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}.$$

By Taylor's Theorem, if we get a good bound on $\frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$ somehow, then we can get a very good approximation for $f(x)$!

Example 5.36. Use the second Taylor approximation to approximate $\sqrt{4.01}$. Estimate the absolute value of the error from the actual value.

Solution: We define

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

Since $\sqrt{4} = 2$, we compute the second Taylor approximation at $x = 4$. We have

$$\begin{aligned} f(4) &= 2, \\ f'(x) &= \frac{1}{2\sqrt{x}} \\ \implies f'(4) &= \frac{1}{4}, \\ f''(x) &= -\frac{1}{4x^{\frac{3}{2}}} \\ \implies f''(4) &= -\frac{1}{32}. \end{aligned}$$

Therefore

$$\begin{aligned} P_2(x) &= 2 + \frac{1}{4}(x-4) - \frac{-\frac{1}{32}}{2}(x-4)^2 \\ &= 2 + \frac{1}{4}(x-4) - \frac{1}{64}(x-4)^2. \end{aligned}$$

To approximate $\sqrt{4.01}$, we plug in $x = 4.01$. This gives

$$f(x) \approx 2 + \frac{1}{4}(0.01) - \frac{1}{64}(0.01)^2 = 2 + \frac{1}{400} - \frac{1}{640000} = 2 + \frac{1601}{640000} = \frac{1281601}{640000}.$$

By Taylor's Theorem, the error is

$$R_2(4.01) = \frac{f'''(c)}{3!}(4.01-4)^3$$

for some $c \in (4, 4.01)$.

We compute

$$f'''(x) = \frac{3}{8x^{\frac{5}{2}}},$$

Therefore, for $c \in (4, 4.01)$, we have

$$|f'''(c)| < \frac{3}{8\,(4)^{\frac{5}{2}}}.$$

Here we used the fact that a larger x will give a smaller $f'''(x)$ ($f'''(x)$ is decreasing), so the value is always smaller than plugging $x = 4$ in directly. Since $4^{\frac{5}{2}} = 2^5 = 32$, we have

$$|f'''(c)| < \frac{3}{256}.$$

Hence the error may be bounded by

$$|R_2(4.01)| < \frac{\frac{3}{256}}{3!}\,(0.01)^3 = \frac{1}{512\,(100^3)} = \frac{1}{512000000}.$$

We see that the approximation is pretty good. $\square$

6. Exponential and logarithm functions

6.1. **Exponential function.** Consider a population of bacteria. These reproduce by splitting in half into make two copies of themself (a process known as binary fission). Each time that the bacteria reproduce, we call the new bacteria a new generation. If you start with one bacteria, then in the second generation there would be 2 bacteria. Each of those would split in half again, making 4 bacteria in the third generation. Repeating this process, in the nth generation there would be

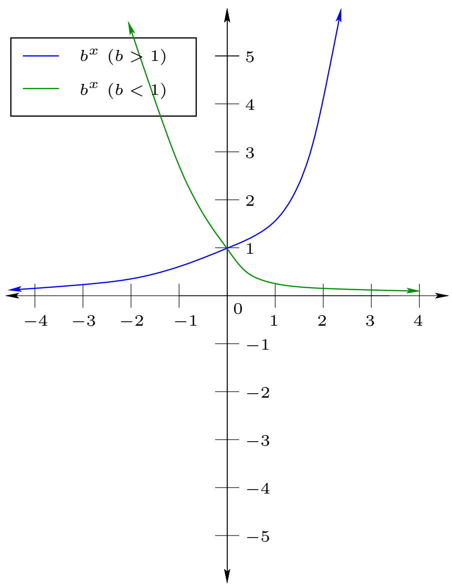
$$2^{n-1}$$

bacteria. This is what is known as exponential growth.

If $b > 0$ and $b \neq 1$, then we call the function

$$f(x) := b^x$$

the exponential function with base b. For $b > 1$, this grows large very fast as x gets bigger, while for $b < 1$ it very quickly gets small (we say that it decays fast). Its natural domain is $\mathbb{R}$ and its range is $(0, \infty)$. This is shown in the graph below.



The exponential functions satisfy a number of useful properties:

(1)
$$b^x b^y = b^{x+y}$$

(2)
$$(b^x)^y = b^{xy}$$

(3)
$$(ab)^x = a^x b^x$$

(4)
$$b^{-x} = \frac{1}{b^x}$$
(5)
$$\underbrace{\frac{b^x}{b^y}}_{=b^x \cdot b^{-y}} = b^{x-y}$$
(6)
$$\underbrace{\left(\frac{a}{b}\right)^x}_{=\left(a \cdot \frac{1}{b}\right)^x} = \frac{a^x}{b^x}.$$
(7)
$$b^0 = 1.$$

Consider now the derivative
$$\frac{d}{dx}\left(b^x\right) = \lim_{h \to 0} \frac{b^{x+h} - b^x}{h} = \lim_{h \to 0} \left(b^x \frac{b^h - 1}{h}\right) = b^x \lim_{h \to 0} \frac{b^h - 1}{h}.$$
Notice that the last limit does not depend on the input x. Therefore $\frac{d}{dx}\left(b^x\right)$ is essentially b^x (up to a constant). Defining
$$f(x) := b^x,$$
the constant is precisely
$$f'(0) = \lim_{h \to 0} \frac{b^h - 1}{h}.$$
Thus
$$f'(x) = f'(0) b^x.$$

It turns out that there is a unique number $e \approx 2.7182818284509$ such that
$$\lim_{h \to 0} \frac{e^h - 1}{h} = 1.$$
Thus
$$\frac{d}{dx} e^x = e^x.$$

Since $b^x > 0$ for every x and
$$f'(x) = f'(0) b^x,$$
we conclude that either $f'(x) > 0$ for every x, or $f'(x) < 0$ for every x (depending on whether $f'(0) > 0$ or $f'(0) < 0$). Since
$$\lim_{x \to \infty} b^x = \begin{cases} \infty & \text{if } b > 1, \\ 0 & \text{if } b < 1, \end{cases}$$
we have the following (which is consistent with the graph given above).

Theorem 6.1. *The function $f(x) := b^x$ is increasing for all x if $b > 1$ and decreasing for all x if $b < 1$.*

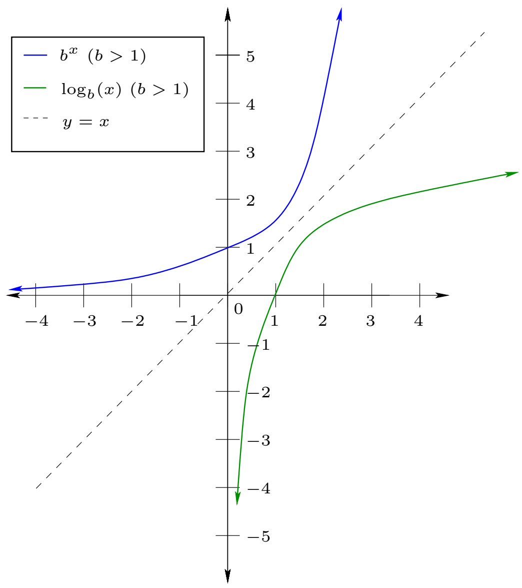
6.2. **Logarithms.** Since the function f is always increasing or always decreasing, it is also injective. Therefore f is invertible. The inverse function is known as the logarithm with base b and is written

$$\log_b(x).$$

In the special case that $b = e$, we write

$$\ln(x) := \log_e(x).$$

This is known as the natural logarithm. The graph of the logarithm is given below:



Recall that the inverse reverses the roles of x and y. Therefore, if

$$b^x = y \Leftrightarrow \log_b(y) = x.$$

This means that $\log_b(y)$ is the power to which you exponentiate b to get y.

Example 6.2. What is $\log_2(16)$?

Solution: Since $2^4 = 16$, we have

$$\log_2(16) = 4.$$

□

Recall that the range of the inverse function is the domain of the original function and vice-versa. Therefore, the natural domain of $\log_b(x)$ is $(0, \infty)$ and the range is $\mathbb{R}$.

By the definition of the inverse, for every $x \in \mathbb{R}$ we have

$$\log_b(b^x) = x.$$

For every $x \in (0, \infty)$, we also have

$$b^{\log_b(x)} = x.$$

The logarithm satisfies a number of properties.

(1) For every $x, y > 0$, we have

$$\log_b(xy) = \log_b(x) + \log_b(y).$$

(2) For every $x, y > 0$, we have

$$\log_b\left(\frac{x}{y}\right) = \log_b(x) - \log_b(y).$$

(3) For every $r \geq 0$, we have

$$\log_b\left(x^r\right) = r\log_b(x).$$

(4) (Change of base formula) For every a, b, we have

$$\log_b(x) = \frac{\log_a(x)}{\log_a(b)}.$$

In particular,

$$\log_b(x) = \frac{\ln(x)}{\ln(b)}.$$

Example 6.3. Radioactive elements decay at an exponential rate. The half-life of a radioactive element is the amount of time necessary for the amount of the element remaining to be half of the amount at the starting time. The half-life of Plutonium is about 24,100 years. If you begin with 100 grams of Plutonium, how much would be remaining after 10,000 years?

Solution: Suppose that $P(t)$ is the number of grams of plutonium left after t years. We have

$$P(t) = P(0)e^{kt},$$

for some constant k. We know that

$$P(24100) = \frac{1}{2}P(0).$$

Therefore

$$P(0)e^{24100k} = \frac{1}{2}P(0),$$

which implies that

$$k = \frac{1}{24100}\ln\left(\frac{1}{2}\right) = -\frac{\ln(2)}{24100}.$$

Therefore

$$P(10000) = P(0)e^{-\frac{10000}{24100}\ln(2)} = 100e^{-\frac{10000}{24100}\ln(2)} \approx 75.005185.$$

Hence there are approximately 75.005185 grams of plutonium remaining after $10,000$ years. $\square$

6.3. **Derivatives of the exponential and logarithm functions.** We were almost able to compute the derivative of the exponential function before. We now use the logarithm to help us finish this calculation. Since

$$e^{\ln(y)} = y$$

for every $y > 0$ and $y = a^x > 0$, we have

$$b^x = e^{\ln(b^x)}.$$

However, since

$$\ln(x^r) = r\ln(x),$$

we have

$$\ln(b^x) = x\ln(b).$$

Therefore,

$$b^x = e^{x\ln(b)}.$$

Since

$$\frac{d}{dx}(e^x) = e^x,$$

the chain rule implies that

$$\frac{d}{dx}(b^x) = e^{x\ln(b)}(\ln(b)) = b^x(\ln(b)).$$

This gives the derivative of a general exponential function.

To take the derivative of the logarithm, we use a trick involving implicit differentiation on the equation

$$e^{\ln(x)} = x.$$

Set $y := \ln(x)$ (note that $e^y = x$). We want to compute $\frac{dy}{dx}$. Taking the derivative on both sides, we have

$$e^y\frac{dy}{dx} = 1.$$

Therefore

$$\frac{dy}{dx} = \frac{1}{e^y}.$$

Since $e^y = x$, we have

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

More generally, by the change of base formula we have

$$\frac{d}{dx}(\log_b(x)) = \frac{d}{dx}\left(\frac{\ln(x)}{\ln(b)}\right) = \frac{1}{x\ln(b)}.$$

The function $f(x) := e^x$ is strictly increasing and strictly concave up, since

$$f'(x) = e^x > 0$$

and

$$f''(x) = e^x > 0$$

for every $x \in \mathbb{R}$. The function $g(x) := \ln(x)$ is strictly increasing because

$$g'(x) = \frac{1}{x} > 0$$

for every $x \in (0, \infty)$ and g is strictly concave down because

$$g''(x) = -\frac{1}{x^2} < 0.$$

Example 6.4. Compute the derivative of

$$f(x) := 3^{\cos(x)}.$$

Solution: By the chain rule, we have

$$f'(x) = 3^{\cos(x)} \ln(3) \frac{d}{dx} cos(x) = 3^{\cos(x)} \ln(3) \left(-\sin(x)\right).$$

$\square$

Example 6.5. Compute the derivative of

$$f(x) := \sqrt{\ln\left(\sqrt{x}\right)}.$$

What is the natural domain of f?

Solution: By the chain rule, we have

$$f'(x) = \frac{1}{2\sqrt{\ln\left(\sqrt{x}\right)}} \left(\frac{1}{\sqrt{x}}\right) \left(\frac{1}{2\sqrt{x}}\right).$$

The natural domain is $[1, \infty)$.

Here is another way to compute the derivative. Notice also that since $\sqrt{x} = x^{\frac{1}{2}}$, we could rewrite

$$f(x) = \sqrt{\ln\left(x^{\frac{1}{2}}\right)}.$$

Using the properties of the logarithm, we hence have

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

The chain rule then gives

$$f'(x) = \frac{1}{2\sqrt{\frac{1}{2}\ln(x)}} \left(\frac{1}{2x}\right).$$

To determine the natural domain, recall that the domain of $\sqrt{x}$ is $[0, \infty)$ and the domain of $\ln(x)$ is $(0, \infty)$. In particular, we need

$$\ln\left(\sqrt{x}\right) \geq 0$$

Since the function $\ln(x)$ is strictly increasing and $\ln(1) = 0$, this happens if and only if

$$\sqrt{x} \geq 1.$$

This is true if and only if $x \geq 1$. The domain is hence $[1, \infty)$. $\square$

6.4. **Logarithmic differentiation.** Sometimes, taking the logarithm of a function and using implicit differentiation can help to compute the derivatives of certain functions. This is called logarithmic differentiation.

Example 6.6. For $f(x) := x^x$, compute $f'(x)$.

Solution: Write $y := f(x)$. Taking the logarithm of both sides, we have

$$\ln(y) = \ln\left(x^x\right) = x \ln(x).$$

We now take the derivative of both sides:

$$\frac{1}{y}\left(\frac{dy}{dx}\right) = x\left(\frac{1}{x}\right) + \ln(x) = 1 + \ln(x).$$

Therefore

$$\frac{dy}{dx} = (1 + \ln(x))\, y = (1 + \ln(x))\, x^x.$$

$\square$

Another useful application of logarithmic differentiation is an alternative to the quotient rule.

Example 6.7. Find the derivative of

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

We again write $y := f(x)$ and implicitly differentiate. We have (using the properties of the logarithm

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

Therefore

$$\frac{1}{y}\frac{dy}{dx} = \frac{1}{3\left(x^2+1\right)}(2x) - \frac{2}{x+2}.$$

Thus

$$\frac{dy}{dx} = y\left(\frac{1}{3\left(x^2+1\right)}(2x) - \frac{2}{x+2}\right) = \frac{\left(x^2+1\right)^{\frac{1}{3}}}{(x+2)^2}\left(\frac{2x}{3\left(x^2+1\right)} - \frac{2}{x+2}\right).$$

6.5. **Computation of limits.** The logarithm and exponential function are also useful for computing certain limits.

The logarithm and exponential functions are both continuous, meaning that for any function $f(x)$ for which $\lim_{x\to a} f(x)$ exists and $\lim_{x\to a} f(x) > 0$, we have

$$\lim_{x\to a} \ln(f(x)) = \ln\left(\lim_{x\to a} f(x)\right)$$

and if $\lim_{x\to a} f(x)$ exists (the limit does not need to be positive)

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

Using this, one is able to compute limits where plugging in directly you would get ∞^0, 0^0, and 1^∞.

Example 6.8. Compute

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

Solution: Plugging in directly would give ∞^0. Set $y := x^{\frac{1}{x}}$. Since

$$y = e^{\ln(y)},$$

we have

$$\lim_{x\to\infty} x^{\frac{1}{x}} = e^{\lim_{x\to\infty} \ln(y)}.$$

However,

$$\ln(y) = \frac{\ln(x)}{x}.$$

We compute

$$\begin{aligned}\lim_{x\to\infty} \ln(y) &= \lim_{x\to\infty} \frac{\ln(x)}{x} \\ &\overset{(\frac{\infty}{\infty})}{\underset{\downarrow}{=}} \lim_{x\to\infty} \frac{\frac{1}{x}}{1} = 0.\end{aligned}$$

Here we used L'Hôpital's rule for $\frac{\infty}{\infty}$.

Above, we had shown that

$$\lim_{x\to\infty} x^{\frac{1}{x}} = e^{\lim_{x\to\infty} \ln(y)},$$

so that

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

□

Example 6.9. Find

$$\lim_{x\to\infty} \left(1 + \frac{.01}{x}\right)^x$$

Remark. The limit above naturally occurs when computing interest. Consider what is known as compound interest (this is the actual interest system used by most banks). Let's say that you get n interest payments every year at a yearly interest rate of 1%. The 1% is split into n pieces.

For simplicity, suppose that $n = 12$ (you get monthly interest). At the end of the first month, you receive $\frac{1}{12}$% interest (since the 1% is split into 12 equal pieces). If you started with \$1, then you'd now have

$$1 + \frac{.01}{12}$$

dollars. But now you also get interest on the interest in the next month. So after two months, you'd have

$$\underbrace{1 + \frac{.01}{12}}_{\text{already in bank}} + \underbrace{\left(1 + \frac{.01}{12}\right)\frac{.01}{12}}_{\text{interest earned}} = \left(1 + \frac{.01}{12}\right)^2.$$

Continuing like this, after 12 months you'd have

$$\left(1 + \frac{.01}{12}\right)^{12}$$

dollars. If you instead split into n pieces, you'd have

$$\left(1 + \frac{.01}{n}\right)^{n}$$

dollars. Taking $n \to \infty$ gives what is known as continuously compounded interest (which is what is being asked for in the question). Some banks do indeed compute interest in this way.

Solution: Set

$$y := \left(1 + \frac{.01}{x}\right)^{x}.$$

Plugging $x\infty$ in directly would give 1^∞. We take the logarithm to obtain

$$\ln(y) = x \ln\left(1 + \frac{.01}{x}\right) = \frac{\ln\left(1 + \frac{.01}{x}\right)}{\frac{1}{x}}.$$

Therefore

$$\lim_{x\to\infty} (\ln(y)) = \lim_{x\to\infty} \frac{\ln\left(1 + \frac{.01}{x}\right)}{\frac{1}{x}}.$$

This is now of the form $\frac{0}{0}$, so that L'Hôpital's rule gives

$$\lim_{x\to\infty} \frac{\frac{1}{1+\frac{.01}{x}}\left(-.01x^{-2}\right)}{-\frac{1}{x^2}} = \lim_{x\to\infty} \frac{.01}{1 + \frac{.01}{x}} = .01.$$

Hence

$$\lim_{x\to\infty} y = e^{\lim_{x\to\infty} \ln(y)} = e^{.01}.$$

□

6.6. **Taylor polynomials of e^x and $\ln(x)$.** Define $f(x) := e^x$. Since

$$f'(x) = e^x,$$

repeated differentiation for $m \geq 0$ gives

$$f^{(m)}(x) = e^x.$$

Therefore $f^{(m)}(0) = 1$.

Therefore, the order n Taylor polynomial at $x = 0$ for f is

$$\begin{aligned} P_n(x) &= f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n \\ &= 1 + x + \frac{x^2}{2} + \cdots + \frac{x^n}{n!}. \end{aligned}$$

Example 6.10. Find the order n Taylor polynomial of $\ln(x)$ at $x = 1$.

Solution: Set $g(x) := \ln(x)$. Then $g'(x) = \frac{1}{x}$. The second derivative is

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

The next derivative is

$$g'''(x) = 2x^{-3}.$$

Repeated differentiation gives (for $m > 0$)

$$g^{(m)}(x) = (-1)^{m-1}(m-1)!x^{-m}$$

(This is true for $m = 1$, so to check this, take the derivative inductively/recursively). Therefore

$$P_n(x) = (x-1) - \frac{(x-1)^2}{2} + \cdots + \frac{(-1)^{n-1}}{n}(x-1)^n.$$

□

7. Integration and antiderivatives

Given the speed of a car at any moment for the first hour after starting out. Is it possible to figure out the position of the car at all times? Roughly speaking, this question asks whether one can, given only the derivative of a function, determine the original function.

Consider the functions

$$f(x) := 3x + 5$$

and

$$g(x) := 3x + 7.$$

Then $f'(x) = 3$ and $g'(x) = 3$. Both f and g have the same dervative, but are different functions. The derivative is hence not enough information by itself to figure out the original function. Another way to think about this is as follows: if you do not know the starting position of then car, then you won't know the overall location. You must therefore know the starting position as well.

Definition 7.1. A function F is called an <u>antiderivative</u> of f if

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

Remark. You should notice that we only say that F is **an** antiderivative, not **the** antiderivative. This is because antiderivatives of f are not unique; there are many antiderivatives, as you see in the above example where we find 2 antiderivatives of the constant function 3.

Suppose that $F_1(x)$ and $F_2(x)$ are both antiderivatives of $f(x)$. Then the slopes of the tangent lines of F_1 and F_2 are always the same. This would be like another car starting behind you and always matching your speed exactly. They would then always be the same distance away from you, or in other words there is a constant $C \in \mathbb{R}$ such that

$$F_1(x) - F_2(x) = C.$$

Knowing the constant C is like knowing the starting position of the car. Another way to see this is the following. If F_1 and F_2 both have the antiderivative f, then for $G(x) := F_1(x) - F_2(x)$, we have

$$G'(x) = 0.$$

The only functions which have $G'(x) = 0$ for every x are the constant functions $G(x) = C$.

Example 7.2. The following is a list of functions and their antiderivatives (up to the constant C, it is unique):

function $f(x)$	x^2	$\sin(x)$	$\frac{1}{x}$	$x + \cos(x)$	$\frac{1}{x^2}$
antiderivative $F(x)$	$\frac{1}{3}x^3$	$-\cos(x)$	$\ln(x)$	$\frac{1}{2}x^2 + \sin(x)$	$-\frac{1}{x}$

You can check each of these by taking the derivative of the second row. For example, for $F(x) := -\cos(x)$, we have

$$F'(x) = -(-\sin(x)) = \sin(x).$$

Definition 7.3. One also calls the (set of) antiderivative $F(x) + C$ of $f(x)$ the indefinite integral of $f(x)$. One writes

$$\int f(x)dx = F(x) + C,$$

where C is an unknown constant. The function f is called the integrand of the indefinite integral.

Indefinite integrals have certain helpful properties.

(1) For every constant $k \in \mathbb{R}$, we have

$$\int kdx = kx + C.$$

(2) For every function f and constant $k \in \mathbb{R}$, we have

$$\int kf(x)dx = k \int f(x)dx.$$

(3) for two functions f and g, we have

$$\int (f(x) \pm g(x))\, dx = \int f(x)dx \pm \int g(x)dx.$$

(4) For $r \in \mathbb{R}$ with $r \neq -1$, we have

$$\int x^r dx = \frac{x^{r+1}}{r+1} + C.$$

(5) We have (when $x \neq 0$)

$$\int \frac{1}{x} dx = \frac{dx}{x} = \ln|x| + C.$$

(6) We have

$$\int e^x dx = e^x + C.$$

(7) The antiderivatives of certain trigonometric functions are given below:

$$\int \sin(x)dx = -\cos(x) + C,$$
$$\int \cos(x)dx = \sin(x) + C,$$
$$\int \sec^2(x)dx = \tan(x) + C,$$
$$\int \csc^2(x)dx = -\cot(x) + C,$$
$$\int \sec(x)\tan(x)dx = \sec(x) + C,$$
$$\int \csc(x)\cot(x)dx = -\csc(x) + C.$$

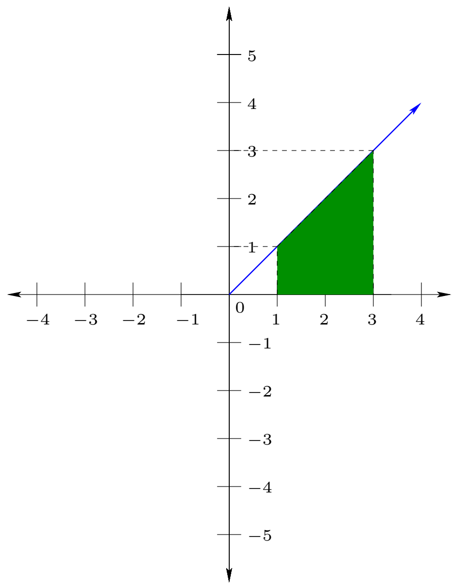
We're now going to look at the antiderivative in another light. Consider the function

$$f(x) := x.$$

The antiderivative of this function is

$$\int f(x)dx = \frac{1}{2}x^2 + C.$$

Now draw $f(x)$ and consider the area under the line from $x = 1$ to $x = 3$.



The area under the line $f(x)$ from $x = 1$ to $x = 3$ is the area under the triangle with the endpoints $(0, 0)$, $(3, 0)$, and $(3, 3)$ minus the area under the triangle with the endpoints $(0, 0)$, $(1, 0)$ and $(1, 1)$. The area of a right triangle is half the area of the rectangle which shares two of its sides.

The area under the line from $x = 1$ to $x = 3$ is hence

$$\frac{1}{2}3^2 - \frac{1}{2}1^2 = \frac{9}{2} - \frac{1}{2} = 4.$$

You may now notice that The area under $f(x)$ from $x = 1$ to $x = 3$ is therefore

$$F(3) - F(1),$$

where

$$F(x) := \frac{1}{2}x^2 + C = \int f(x)dx.$$

The constant C does not matter, because we add it for $F(3)$ and subtract it together with $F(1)$. This is an example of a central theme of calculus (known as the Fundamental Theorem of Calculus), which we first need some notation to investigate.

Definition 7.4. We call the area under the curve $f(x)$ from $x = a$ to $x = b$ the definite integral of f from a to b. It is denoted by

$$\int_a^b f(x)dx.$$

The number a is called the lower limit of the integral and b is called the upper limit of the integral.

By "area under the curve" we mean the area between the curve $y = f(x)$ and the x-axis. If the curve is below the x-axis, then the area is counted negatively. If it is above the x-axis, then the area is counted positively.

Theorem 7.5 (First Fundamental Theorem of Calculus)**.** *If F is an antiderivative of f on $[a, b]$, then*

$$\int_a^b f(x)dx = F(b) - F(a).$$

Remark. It is also useful to sometimes abbreviate this as

$$\int_a^b f(x)dx = F(x)\Big|_a^b.$$

The notation

$$F(x)|_a^b := F(b) - F(a)$$

means to substitute $x = b$ in and then subtract with $x = a$ plugged in.

Another way to write the Fundamental Theorem of Calculus is

$$\int_a^x F'(t)dt = F(x) - F(a).$$

Essentially, this means that integration reverses differentiation. Differentiation also reverse the action of differentiation.

Theorem 7.6 (Second Fundamental Theorem of Calculus)**.** *If f is continuous on $[a,b]$, then*

$$F(x) := \int_a^x f(t)dt$$

is an antiderivative of f on $[a,b]$, that is to say for every $x \in [a,b]$

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

We see that integration and differentiation are inverses of each other. From the definition of the indefinite integral, this may not be a big surprise (an antiderivative is defined to reverse this operation). That this happens to also be the area under the curve $f(x)$ is the main substance of this statement.

Example 7.7. Compute

$$\int_0^3 \left(x^2 - 7x\right) dx.$$

Solution: By the first Fundamental Theorem of Calculus, we can take any antiderivative and subtract the values. Since

$$F(x) := \frac{1}{3}x^3 - \frac{7}{2}x^2$$

is an antiderivative of $x^2 - 7x$, the value is

$$\left(\frac{x^3}{3} - \frac{7}{2}x^2\right)\Big|_0^3 = \frac{3^3}{3} - \frac{7 \cdot 3^2}{2} = 9 - \frac{63}{2} = -\frac{45}{2}.$$

Note that the area is negative, which is accounted for by the fact that the curve is below the x-axis for $x \in (0,7)$.

□

Example 7.8. Find

$$\frac{d}{dx}\int_0^x \frac{t}{1+t^5}dt.$$

Solution: Define

$$f(x) := \frac{x}{1+x^5}$$

and

$$F(x) := \int_0^x f(t)dt.$$

Then by the second Fundamental Theorem of Calculus, we have

$$F'(x) = f(x) = \frac{x}{1+x^5}..$$

□

Example 7.9. Find

$$\frac{d}{dx}\int_2^{x^3+4} \cos\left(t^3\right) dt.$$

Solution: Define

$$f(x) := \cos\left(x^3\right)$$

and

$$F(x) := \int_2^x f(t)dt.$$

By the second Fundamental Theorem, we have

$$F'(x) = f(x) = \cos\left(x^3\right).$$

Applying the chain rule, we hence obtain

$$\frac{d}{dx}\int_2^{x^3+4} \cos\left(t^3\right) dt = 3x^2 \cos\left(\left(x^3+4\right)^3\right).$$

□

Example 7.10. Find a function f and a constant c such that

$$\int_c^x f(t)dt = \tan(x) - 1$$

for $x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

Solution: Taking the derivative of both sides yields

$$\frac{d}{dx}\int_c^x f(t)dt = \frac{d}{dx}(\tan(x) - 1) = \sec^2(x).$$

By the second fundamental Theorem of Calculus (note that $\sec^2(x)$ is continuous for $x \in \left(\frac{\pi}{2}, \frac{\pi}{2}\right)$), the left-hand side of this equation is

$$\frac{d}{dx}\int_c^x f(t)dt = f(x).$$

Hence we have

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

We also have the antiderivative $F(x) = \tan(x)$. Therefore

$$\int_c^x f(t)dt = F(x) - F(c) = \tan(x) - \tan(c)$$

We conclude that $\tan(c) = 1$. Therefore $c = \frac{\pi}{4}$. □

7.1. **Integration by substitution.** We next consider the reverse of the chain rule. Remember that for

$$h(x) := f(g(x)),$$

the chain rule states that

$$h'(x) = f'(g(x))g'(x)dx.$$

However, by the second Fundamental Theorem of Calculus, we have

$$h(x) + C = \int h'(x)dx.$$

Putting these together, we have

$$h(x) + C = \int f'(g(x))g'(x)dx.$$

This leads to a rule called integration by substitution, which we next describe.

Let's call

$$u := g(x).$$

Then

$$\frac{du}{dx} = g'(x).$$

Pretending that du and dx are "very small changes in u and x", this would give

$$du = g'(x)dx.$$

One calls du and dx differentials. If you have an integral of the form

$$\int f'(g(x))g'(x)dx,$$

then it is easier to write

$$\int f'(u)du = f(u) + C.$$

Let's do a few examples.

Example 7.11. Find the integral

$$\int \frac{\ln(x)}{x}dx$$

Solution: Set $u := g(x) := \ln(x)$ and recall that

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

Then we have the integral

$$\int \frac{\ln(x)}{x}dx = \int udu = \frac{u^2}{2} + C$$

We substitute back in $u = \ln(x)$ to obtain

$$\int \frac{\ln(x)}{x} dx = \frac{\ln(x)^2}{2} + C.$$

□

You can also do substitution with definite integrals, but you have to be careful about the **endpoints**!

Example 7.12. Find the value of the definite integral

$$\int_0^{\sqrt{\pi}} x \sin\left(x^2\right) dx.$$

Solution: We make the substitution $u = x^2$. Then

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

so

$$du = 2xdx.$$

Note that for $x = 0$ we have $u = 0$ and for $x = \sqrt{\pi}$ we have $u = \pi$. This changes the endpoints of the integral. The integral then becomes

$$\begin{aligned}\int_0^{\sqrt{\pi}} x \sin\left(x^2\right) dx &= \frac{1}{2}\int_0^{\pi} \sin(u)du \\ &= -\frac{1}{2}\cos(u)\Big|_0^{\pi} \\ &= -\frac{1}{2}(\cos(\pi) - \cos(0)) = 1.\end{aligned}$$

Remark. You could also first compute the indefinite integral

$$\begin{aligned}\int x \sin\left(x^2\right) dx &= \frac{1}{2}\int \sin(u)du \\ &= -\frac{1}{2}\cos(u) \\ &= -\frac{1}{2}\cos\left(x^2\right).\end{aligned}$$

Then, by the first Fundamental Theorem of Calculus, you have

$$\int_0^{\sqrt{\pi}} x \sin\left(x^2\right) dx = -\frac{1}{2}\cos\left(x^2\right)\Big|_0^{\sqrt{\pi}} = 1.$$

□

7.2. **Trigonometric Substitution.** Sometimes it is useful to make a substitution such as

$$x = \sin(\theta).$$

Then

$$dx = \cos(\theta) d\theta.$$

Trigonometric identities may they be used to find the value of the integral. This is know as integration by trigonometric substitution.

Example 7.13. Find

$$\int \frac{dx}{\sqrt{1-x^2}}.$$

Solution: Set $x = \sin(\theta)$. Then we have

$$\begin{aligned}\int \frac{dx}{\sqrt{1-x^2}} &= \int \frac{\cos(\theta)}{\sqrt{1-\sin^2(\theta)}} d\theta \\ &= \int \frac{\cos(\theta)}{\sqrt{\cos^2(\theta)}} d\theta,\end{aligned}$$

where in the last equality we used

$$\cos^2(\theta) + \sin^2(\theta) = 1.$$

Note that $\theta = \arcsin(x)$ and $-1 < x < 1$ because this is the natural domain of the function

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

Therefore $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$. However, $\cos(\theta) > 0$ for $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$, so

$$\sqrt{\cos^2(\theta)} = \cos(\theta).$$

We hence obtain

$$\int \frac{dx}{\sqrt{1-x^2}} = \int \frac{\cos(\theta)}{\cos(\theta)} d\theta = \theta + C = \arcsin(x) + C.$$

Remark. Since integration and differentiation are inverses of each other, one concludes that

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

□

Example 7.14. Compute

$$\int_0^1 \frac{dx}{x^2+4}.$$

Solution: We make the change of variables

$$x = 2\tan(\theta).$$

Then

$$\theta = \arctan\left(\frac{x}{2}\right)$$

and

$$dx = 2\sec^2(\theta)d\theta$$

Moreover,

$$\begin{aligned}\frac{1}{x^2+4} &= \frac{1}{4\tan^2(\theta)+4}\\ &= \frac{1}{4}\left(\frac{1}{\frac{\sin^2(\theta)}{\cos^2(\theta)}+1}\right)\\ &= \frac{1}{4}\left(\frac{1}{\frac{\sin^2(\theta)+\cos^2(\theta)}{\cos^2(\theta)}}\right)\\ &= \frac{1}{4}\cos^2(\theta).\end{aligned}$$

Thus

$$\int_0^1 \frac{dx}{x^2+4} = \frac{1}{2}\int_{\arctan(0)}^{\arctan\left(\frac{1}{2}\right)} \sec^2(\theta)\cos^2(\theta)d\theta = \frac{1}{2}\theta\Big|_{\arctan(0)}^{\arctan\left(\frac{1}{2}\right)} = \frac{\arctan\left(\frac{1}{2}\right)}{2}.$$

□

7.3. **Integration by parts.** The reverse of the product rule is known as <u>integration by parts</u>. Recall that the product rule is

$$\frac{d}{dx}\left(f(x)g(x)\right) = f'(x)g(x) + f(x)g'(x).$$

Taking the (indefinite) integral of both sides, this is

$$\begin{aligned}\int \frac{d}{dx}\left(f(x)g(x)\right)dx &= \int \left(f'(x)g(x) + f(x)g'(x)\right)dx\\ &= \int g(x)f'(x)dx + \int f(x)g'(x)dx.\end{aligned}$$

The left-hand side of this equation is (by the second Fundamental Theorem of Calculus)

$$f(x)g(x) = \int \frac{d}{dx}\left(f(x)g(x)\right)dx.$$

Therefore

$$f(x)g(x) = \int g(x)f'(x)dx + \int f(x)g'(x)dx.$$

Rearranging, one obtains the <u>integrations by parts formula</u>

$$\int f(x)g'(x)dx = f(x)g(x) - \int g(x)f'(x)dx.$$

If $u = f(x)$ and $v = g(x)$, then this may be written

$$\int udv = uv - \int vdu.$$

The point is that sometimes we don't know how to compute

$$\int udv$$

but we do know how to compute

$$\int vdu.$$

Integration by parts thus gives us a way to compute $\int udv$.

Example 7.15. Find

$$\int \ln(x)dx.$$

Solution: We set $u = \ln(x)$ and $dv = dx$. Then

$$\begin{array}{rcl|rcl} u & = & \ln(x), & dv & = & dx, \\ du & = & \frac{1}{x}dx, & v & = & x. \end{array}$$

Therefore

$$\begin{aligned} \int \ln(x)dx &= x\ln(x) - \int x\left(\frac{1}{x}\right)dx \\ &= x\ln(x) - \int dx = x\ln(x) - x + C. \end{aligned}$$

□

Example 7.16. Compute

$$\int_0^1 x^2e^x dx.$$

Solution: Setting $u = x^2$ and $dv = e^x dx$, we have

$$\begin{array}{rcl|rcl} u & = & x^2, & dv & = & e^x dx, \\ du & = & 2xdx, & v & = & e^x. \end{array}$$

Therefore

$$\int x^2e^x dx = x^2e^x - 2\int xe^x dx.$$

We now do integration by parts again to compute $\int xe^x dx$. Setting $u = x$ and $dv = e^x dx$, we have

$$\begin{array}{rcl|rcl} u & = & x, & dv & = & e^x dx, \\ du & = & dx, & v & = & e^x. \end{array}$$

Therefore

$$\int xe^x dx = xe^x - \int e^x dx = xe^x - e^x + C.$$

We conclude that

$$\int x^2 e^x dx = x^2 e^x - 2xe^x + 2e^x + C.$$

Using the first Fundamental Theorem of Calculus, we then compute

$$\int_0^1 x^2 e^x dx = \left(x^2 e^x - 2xe^x + 2e^x\right)\Big|_0^1 = e - 2e + 2e - 2 = e - 2.$$

□

Example 7.17. Compute

$$\int e^x \sin(x) dx.$$

Solution: We set $u = \sin(x)$ and $dv = e^x dx$. We then have

$$\begin{array}{lll|lll} u & = & \sin(x), & dv & = & e^x dx, \\ du & = & \cos(x)dx, & v & = & e^x. \end{array}$$

We then obtain

$$\int e^x \sin(x) dx = e^x \sin(x) - \int e^x \cos(x) dx.$$

We again apply integration by parts with $u = \cos(x)$ and $dv = e^x dx$. This yields

$$\begin{array}{lll|lll} u & = & \cos(x), & dv & = & e^x dx, \\ du & = & -\sin(x)dx, & v & = & e^x. \end{array}$$

Therefore

$$\int e^x \cos(x) dx = e^x \cos(x) + \int \sin(x) e^x dx.$$

Hence

$$\int e^x \sin(x) dx = e^x \sin(x) - e^x \cos(x) - \int \sin(x) e^x dx.$$

It follows that

$$2 \int e^x \sin(x) dx = e^x \sin(x) - e^x \cos(x) + C$$

Therefore

$$\int e^x \sin(x) dx = \frac{1}{2} e^x \sin(x) - \frac{1}{2} e^x \cos(x) + C.$$

□

7.4. **Further properties of indefinite integrals.** Indefinite integrals satisfy a few additional properties which are worth mentioning.

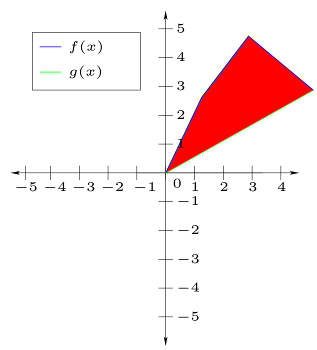
(1) We have
$$\int_a^a f(x)dx = 0.$$
The area under the curve is zero because there is no width.

(2) If $a < c < b$, then we have
$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.$$
You can think of this as splitting the interval from a to b into two pieces and computing the area underneath the curve for each one.

(3) If $a > b$, then we define
$$\int_a^b f(x)dx := -\int_b^a f(x)dx.$$
The intuitive reason that the area is negative is because the order is reversed. The fact that the area is negative if the curve is below the x-axis means that area has some sort of ordering. Hence, if you reverse things, the area should be negative.

(4) As a result of the last definition, for any $a, b, c \in \mathbb{R}$ we have
$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.$$
This is one motivation for defining things as above. If you had the above identity for all c and $a = b$, then
$$0 = \int_a^a f(x)dx = \int_a^c f(x)dx + \int_c^a f(x)dx.$$
This gives the previous definition.

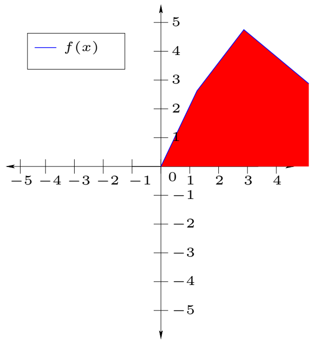
(5) Suppose that you want to compute the area **between** two functions f and g. Namely, you'd like to compute shaded red area in the following diagram:



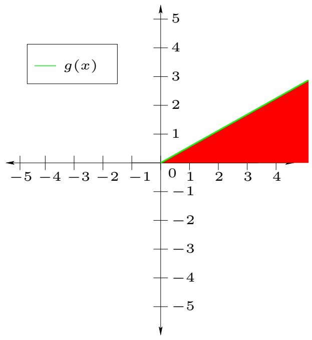
Then you would take the difference

$$\int_a^b f(x)dx - \int_a^b g(x)dx.$$

The first integral is the area between f and the x-axis:



The second integral is the area between g and the x-axis:



The difference is hence the area between f and g.

8. Matrices

Definition 8.1. Suppose that $n, m \in \mathbb{N}$. A $m \times n$ matrix with real coefficients is an array

$$A := \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix},$$

where $A_{11}, A_{12}, \ldots, A_{mn} \in \mathbb{R}$. One calls A_{ij} the (i,j)th entry of A. This is sometimes abbreviated $A := (A_{ij})_{m\times n}$ or $A := (A_{ij})$ if m and n are clear from the context. We call $m \times n$ the size or dimensions of the matrix.

If $n = 1$, then we call A a column vector of length m, and if $m = 1$, then we call A a row vector of length n. More generally, for a matrix A and $1 \leq j \leq n$, we call $(A_{ij})_{m\times 1}$ the jth column and for $1 \leq i \leq m$ we call $(A_{ij})_{1\times n}$ the ith row.

In the case that $n = m$, we call A a square matrix of size n.

Example 8.2. The matrix

$$A := \begin{pmatrix} 1 \\ 2 \\ 7 \\ 3 \end{pmatrix}$$

is a column vector of length 4.

The matrix

$$B := \begin{pmatrix} -1 & 2 & -7 & 3 & 1 \end{pmatrix}$$

is a row vector of length 5.

The matrix

$$C := \begin{pmatrix} 0 & 1 & 4 \\ -1 & 3 & 6 \end{pmatrix}$$

is a 2×3 matrix.

The matrix

$$D := \begin{pmatrix} 1 & -3 & 5 & -2 \\ -7 & 8 & 4 & 1 \\ -3 & -2 & 1 & 6 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

is a square matrix of size 4 (or, in other words, a 4×4 matrix). The column vector

$$\begin{pmatrix} -3 \\ 8 \\ -2 \\ 1 \end{pmatrix}$$

is the second column of D.

There are a number of very special matrices which occur throughout mathematics and applications to engineering and the sciences. The first such matrix is the zero matrix $(0)_{m\times n}$, all of whose entries are zero ($A_{ij} = 0$ for every i, j). This is often simply written 0 if m and n are known.

The entries A_{ii} are called the diagonal entries (or simply diagonal) of the matrix A. The terms A_{ij} with $i \neq j$ are called the off-diagonal elements. The square matrix of size n which has all diagonal entries equal to 1 and all off-diagonal entries equal to 0 is called the identity matrix of size n.

Example 8.3. The identity matrix of size 4 is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The identity matrix of size 5 is

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

8.1. **Arithmetic of matrices.** There are a number of operations on matrices which can be performed. The first is addition and subtraction. The sum is simply defined by

$$\left(A_{ij}\right)_{m\times n} + \left(B_{ij}\right)_{m\times n} := \left(A_{ij} + B_{ij}\right)_{m\times n}.$$

That is, we add the (i, j)th entry of A to the (i, j)th entry of B. The difference of A and B is simply defined by

$$\left(A_{ij}\right)_{m\times n} - \left(B_{ij}\right)_{m\times n} := \left(A_{ij} - B_{ij}\right)_{m\times n}.$$

That is to say, we take the (i, j)th entry of A minus the (i, j)th entry of B.

Example 8.4. For

$$A := \begin{pmatrix} 1 & 0 & -3 \\ 2 & -5 & 4 \end{pmatrix}$$

and

$$B := \begin{pmatrix} -2 & 3 & -4 \\ 1 & -1 & 2 \end{pmatrix},$$

we have

$$A + B = \begin{pmatrix} 1-2 & 0+3 & -3-4 \\ 2+1 & -5-1 & 4+2 \end{pmatrix} = \begin{pmatrix} -1 & 3 & -7 \\ 3 & -6 & 6 \end{pmatrix}.$$

We also have

$$A - B = \begin{pmatrix} 1+2 & 0-3 & -3+4 \\ 2-1 & -5+1 & 4-2 \end{pmatrix} = \begin{pmatrix} 3 & -3 & 1 \\ 1 & -4 & 2 \end{pmatrix}.$$

The zero matrix is the precisely the matrix which satisfies

$$A + 0 = A$$

for every A. One also has

$$A - A = 0.$$

For the next operation on matrices, we take a matrix A and $\lambda \in \mathbb{R}$. The number λ is called a <u>scalar</u>. We define <u>scalar multiplication</u> by multiplying the (i, j)th entry of A by λ. That is to say, scalar multiplication is defined by

$$\lambda A = \lambda \left(A_{ij}\right)_{m\times n} := \left(\lambda A_{ij}\right)_{m\times n}.$$

Example 8.5. For $\lambda := 4$ and

$$A := \begin{pmatrix} 1 & -3 & 2 \\ 0 & 1 & 5 \end{pmatrix},$$

we have

$$\lambda A = 4A = \begin{pmatrix} 4\cdot 1 & 4(-3) & 4\cdot 2 \\ 4\cdot 0 & 4\cdot 1 & 4\cdot 5 \end{pmatrix} = \begin{pmatrix} 4 & -12 & 8 \\ 0 & 4 & 20 \end{pmatrix}.$$

Since addition and scalar multiplication for matrices are defined by usual addition and multiplication entry-wise, they satisfy properties of usual addition.

Theorem 8.6. *Suppose that A, B, and C are matrices and $\alpha, \lambda \in \mathbb{R}$.*

(1) *(Commutative law for addition) We have*

$$A + B = B + A.$$

(2) *(Associative law for addition) We have*

$$(A + B) + C = A + (B + C).$$

(3) *We have*

$$A + 0 = A = 0 + A.$$

(4) *We have*

$$\alpha\left(\lambda A\right) = (\alpha\lambda)A$$

(5) *We have*

$$(\alpha + \lambda)A = \alpha A + \lambda A.$$

(6) *We have*

$$\lambda(A + B) = \lambda A + \lambda B.$$

There is also a multiplication defined on matrices. This is however not defined entry-wise. Moreover, the size of the matrices have to satisfy certain relations to even define multiplication. For $A = (A_{ij})_{m\times n}$ and $B = (B_{ij})_{n\times r}$, we have

$$AB := (A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{in}B_{nj})_{m\times r}\,.$$

Note that the number of columns of A must be the number of rows of B, since otherwise

$$A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{in}B_{nj}$$

does not make sense. The (i, j)th entry of the product AB is the product of the ith row of A times the jth column of B, i.e., adding the sum of the products of the kth element of the row with the kth element of the column.

Example 8.7. For

$$A := \begin{pmatrix} 1 & -1 & 2 \\ 1 & -3 & 0 \\ -4 & 1 & 7 \\ 0 & -2 & 3 \end{pmatrix}$$

and

$$B := \begin{pmatrix} 0 & -4 \\ -2 & 1 \\ 6 & 3 \end{pmatrix},$$

what is AB (if it exists)? What is BA (if it exists)?

Solution: We have

$$\begin{aligned} AB &= \begin{pmatrix} 1\cdot 0 + (-1)(-2) + 2\cdot 6 & 1(-4) + (-1)\cdot 1 + 2\cdot 3 \\ 1\cdot 0 + (-3)(-2) + 0\cdot 6 & 1(-4) + (-3)\cdot 1 + 0\cdot 3 \\ -4\cdot 0 + 1(-2) + 7\cdot 6 & (-4)(-4) + 1\cdot 1 + 7\cdot 3 \\ 0\cdot 0 + (-2)(-2) + 3\cdot 6 & 0(-4) + (-2)\cdot 1 + 3\cdot 3 \end{pmatrix} \\ &= \begin{pmatrix} 14 & 1 \\ 6 & -7 \\ 40 & 38 \\ 22 & 7 \end{pmatrix}. \end{aligned}$$

The product BA simply isn't defined, because B has 2 columns while A has 4 rows. These must be the same to multiply the matrices. □

As evidenced in the above example, AB may be defined sometimes when BA is not. Actually, AB and BA are defined if and only if A is an $m \times n$ matrix and B is a $n \times m$ matrix, for some $m, n \in \mathbb{N}$. The matrices AB and BA do not need to be the same (they don't even have to be the same size!).

Example 8.8. For

$$A := \begin{pmatrix} 1 & -2 \end{pmatrix}$$

and

$$B := \begin{pmatrix} 3 \\ -1 \end{pmatrix},$$

find AB and BA. What are the dimensions of these matrices?

Solution: We have

$$AB = \big(1 \cdot 3 + -2(-1)\big) = (5).$$

This is a 1×1 matrix.

Multiplying the other way, we have

$$AB = \begin{pmatrix} 3 \cdot 1 & 3(-2) \\ -1 \cdot 1 & (-1)(-2) \end{pmatrix} = \begin{pmatrix} 3 & -6 \\ -1 & 2 \end{pmatrix}.$$

This is a 2×2 matrix. Clearly they are not equal, since they have different sizes. □

When A and B are both square matrices of size n, then both AB and BA are square matrices of size n. However, even in this case they are not necessarily equal.

Example 8.9. For

$$A := \begin{pmatrix} 1 & -1 \\ 3 & -2 \end{pmatrix}$$

and

$$B := \begin{pmatrix} 2 & -4 \\ -2 & 1 \end{pmatrix},$$

find AB and BA.

Solution: We have

$$AB = \begin{pmatrix} 1 \cdot 2 - 1(-2) & 1(-4) - 1 \cdot 1 \\ 3 \cdot 2 - 2(-2) & 3(-4) - 2 \cdot 1 \end{pmatrix} = \begin{pmatrix} 4 & -5 \\ 10 & -14 \end{pmatrix}.$$

Moreover,

$$BA = \begin{pmatrix} 2 \cdot 1 - 4 \cdot 3 & 2(-1) - 4(-2) \\ -2 \cdot 1 + 1 \cdot 3 & -2(-1) + 1(-2) \end{pmatrix} = \begin{pmatrix} -10 & 6 \\ 1 & 0 \end{pmatrix}.$$

□

The $n \times n$ identity matrix I_n is the special matrix which for any $n \times n$ matrix A satisfies

$$AI_n = A$$

and

$$I_n A = A.$$

For example, with

$$I_2 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$A := \begin{pmatrix} 1 & 7 \\ 4 & 3 \end{pmatrix}$$

we have
$$AI_2 = \begin{pmatrix} 1\cdot 1 + 0\cdot 4 & 1\cdot 7 + 0\cdot 3 \\ 0\cdot 1 + 1\cdot 4 & 0\cdot 7 + 1\cdot 3 \end{pmatrix} = A$$
and
$$I_2A = \begin{pmatrix} 1\cdot 1 + 7\cdot 0 & 1\cdot 0 + 1\cdot 7 \\ 4\cdot 1 + 3\cdot 0 & 4\cdot 0 + 3\cdot 1 \end{pmatrix} = A.$$
Multiplication also satisfies a number of useful properties.

Theorem 8.10.

(1) *If AB and BC are both well-defined, then we have*
$$(AB)C = A(BC).$$
(2) *If AB and AC are both well-defined, then we have*
$$A\,(B + C) = AB + AC.$$
(3) *If BA and CA are both well-defined, then we have*
$$(B + C)\,A = BA + CA.$$
(4) *For any $\lambda \in \mathbb{R}$, if AB is well-defined, then we have*
$$\lambda\,(AB) = (\lambda A)\,B = A\,(\lambda B)\,.$$

Sometimes it is useful to construct matrices from other known matrices. For example, given a matrix $A = (A_{ij})_{n\times m}$, we define the transpose matrix A^T by
$$A^T := (A_{ji})_{m\times n}\,.$$
In other words, the rows and the columns of the matrix are reversed.

Example 8.11. Consider the matrix
$$A := \begin{pmatrix} 1 & 3 & 7 \\ -2 & 1 & 0 \end{pmatrix}.$$
The transpose of this matrix is
$$A^T := \begin{pmatrix} 1 & -2 \\ 3 & 1 \\ 7 & 0 \end{pmatrix}.$$

We call a square matrix A symmetric if
$$A^T = A.$$

Example 8.12. The matrix
$$A = \begin{pmatrix} 2 & 5 & -3 \\ 5 & 1 & 0 \\ -3 & 0 & 2 \end{pmatrix}$$
is symmetric.

In addition to returning the matrix A when multiplying against it, the identity matrix I_n appears in one additional important place. If A and B are both square matrices of size n and

$$BA = AB = I_n,$$

then we say that B is the (multiplicative) inverse of A. We write $B = A^{-1}$ and call A invertible. The inverse does not always exist, however. The existence of an inverse is determined by something called the determinant of the matrix.

8.2. **Determinants.** Checking whether the inverse of a square matrix A exists involves a number called the determinant, which is usually denoted $\det(A)$ or $|A|$.

Theorem 8.13. *A square matrix A is invertible if and only if its determinant is non-zero.*

In these notes, we won't give the full definition of the determinant (there is a systematic definition, however), but will give it in the special case that the size of the square matrix is $n = 1$, $n = 2$, or $n = 3$.

Definition 8.14.

(1) If $n = 1$, then the determinant of

$$A := (a)$$

is

$$\det(A) = a.$$

(2) If $n = 2$, then the determinant of

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is

$$\det(A) = ad - bc.$$

(3) If $n = 3$, then the determinant of

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & j \end{pmatrix}$$

is

$$\det(A) = aej + bfg + cdh - gec - hfa - jdb.$$

Example 8.15. For

$$A := \begin{pmatrix} 2 & -1 \\ 3 & -2 \end{pmatrix},$$

we have

$$\det(A) = 2(-2) - (-1)(3) = -4 + 3 = -1.$$

Since $\det(A) \neq 0$, the matrix A is invertible. In particular,

$$A^{-1} = \begin{pmatrix} 2 & -1 \\ 3 & -2 \end{pmatrix}.$$

To check this, we multiply

$$\begin{pmatrix} 2 & -1 \\ 3 & -2 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 3 & -2 \end{pmatrix} = \begin{pmatrix} 4-3 & -2+2 \\ 6-6 & -3+4 \end{pmatrix} = I_2.$$

In general, the inverse of a 2×2 matrix has the following simple shape.

Theorem 8.16. *If*

$$A := \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$$

has non-zero determinant $ad - bc$, then

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

The determinant satisfies a number of useful properties.

(1) If any row or column of A is zero, then $\det(A) = 0$.
(2) Multiplying one row or column by a constant $\lambda \in \mathbb{R}$ multiplies the determinant by λ.
(3) We have

$$\det\left(A^T\right) = \det(A).$$

(4) For any two matrices A and B, we have

$$\det(AB) = \det(A)\det(B).$$

8.3. **Solving systems of linear equations.** Recall that a linear equation (with two variables x, y) is an equation of the form

$$ay + bx = c,$$

where $a, b, c \in \mathbb{R}$ are fixed constants. The solution to a linear equation with two variables is a line.

More generally, if $x_1, \ldots, x_n$ are variables and $a_1, \ldots, a_n$, and b are constants, then we call

$$a_1x_1 + \cdots + a_nx_n = b$$

a linear equation.

It is often useful to be able to find choices of $x_1, \ldots, x_n$ which solve many linear equations at the same time. When we want to find simultaneous solutions to a number of equations, we call the set of these equations a system of linear equations.

Example 8.17. The following is a system of linear equations in two variables:

$$\begin{aligned} 5x + 2y &= 9, \\ 3x + y &= 5. \end{aligned}$$

The solution to the above system of linear equations is the unique (x, y) where the two lines intersect. For example, this might represent the supply and demand curves in an economics problem. Their intersection point is then the optimal number supplied. In the above example, this is

$$(x, y) = (1, 2).$$

One way to solve a system of linear equations is to do substitution. For example, in the above example you could take the first equation and subtract twice the second equation. This gives

$$\begin{aligned} 5x + 2y - 2\,(3x + y) &= 9 - 2 \cdot 5 \\ -x &= -1 \\ x &= 1 \\ 3(1) + y &= 5 \\ y &= 2. \end{aligned}$$

This works fine enough for 2 variables and two equations, but might get quite difficult when you have more than 2 variables or more than two equations.

We are next going to use matrices to package the system of linear equations together nicely and also give a way to solve the system sometimes. Suppose that you have a system of linear equations

$$\begin{aligned} A_{11}x_1 + \cdots + A_{1n}x_n &= b_1 \\ A_{21}x_1 + \cdots + A_{2n}x_n &= b_2 \\ \vdots \ddots + \vdots &= \vdots \\ A_{m1}x_1 + \cdots + A_{mn}x_n &= b_m \end{aligned}$$

Then we can package the linear system of equations as the matrix equation

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

For simplicity, one often simply writes

$$Ax = b,$$

where A is the matrix

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix},$$

x is the column vector

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

and b is the column vector

$$\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

Example 8.18. Consider

$$5x + 2y = 9,$$
$$3x + y = 5.$$

We may then write this

$$\begin{pmatrix} 5 & 2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 9 \\ 5 \end{pmatrix}.$$

Now suppose that A is invertible. In this case, consider the equation

$$Ax = b.$$

We may multiply on both sides by A^{-1} to obtain

$$A^{-1}Ax = A^{-1}b.$$

But $A^{-1}A$ is the identity, so $A^{-1}Ax = x$. The solution is then

$$x = A^{-1}b.$$

Example 8.19. The inverse of

$$\begin{pmatrix} 5 & 2 \\ 3 & 1 \end{pmatrix}$$

is

$$\begin{pmatrix} -1 & 2 \\ 3 & -5 \end{pmatrix}.$$

Therefore, the solution to

$$\begin{pmatrix} 5 & 2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 9 \\ 5 \end{pmatrix}$$

is

$$
\begin{aligned}
\begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} -1 & 2 \\ 3 & -5 \end{pmatrix} \begin{pmatrix} 9 \\ 5 \end{pmatrix} \\
&= \begin{pmatrix} -9+10 \\ 27-25 \end{pmatrix} \\
&= \begin{pmatrix} 1 \\ 2 \end{pmatrix}.
\end{aligned}
$$

This is precisely the same answer that we got with substitution.

9. Complex numbers

Recall the definitions of $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{R}$. If you ask for the solutions to

$$\left(x^2-2\right)\left(x^2-4\right)=0$$

in $\mathbb{N}$, $\mathbb{Z}$, or $\mathbb{R}$, you would get a different answer.

Specifically, the solutions in $\mathbb{N}$ are $x=2$, the solutions in $\mathbb{Z}$ are $x=\pm 2$, and the solutions in $\mathbb{R}$ are $\left\{\pm 2, \pm\sqrt{2}\right\}$. One sees from this that the solutions can really depend on the set that the solutions are coming from. Now consider the equation

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

This has no solutions in $\mathbb{N}$, $\mathbb{Z}$, or $\mathbb{R}$. Throughout this course, we've been assuming that the underlying set is $\mathbb{R}$, so we've been essentially been avoiding the equation

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

However, it is perfectly natural to generalize $\mathbb{R}$ to a larger set which includes a solution to the above equation (when you first started out in school, $x^2-2=0$ also wouldn't have a solution, because you didn't know about real numbers yet). We add a new number i which satisfies $i^2=-1$ (essentially, we can call $i=\sqrt{-1}$ – although there is an issue with choosing i versus $-i$ because up to now we always choose $\sqrt{d}$ to be the positive choice and there is no obvious idea of positive versus negative when we're not in the set $\mathbb{R}$).

Definition 9.1. The complex numbers are defined by

$$\mathbb{C}:=\{a+bi : a, b\in\mathbb{R}\}.$$

For a complex number $z:=a+bi\in\mathbb{C}$ $(a,b\in\mathbb{R})$, we call a the real part of z an b the imaginary part of z. One write $a=\operatorname{Re}(z)$ and $b=\operatorname{Im}(z)$. If $a=0$, then we call $z=bi$ purely imaginary. If $b=0$, then $z\in\mathbb{R}$ (that is, $\mathbb{R}\subset\mathbb{C}$). One also often writes $a+ib$ instead of $a+bi$.

9.1. The algebra of complex numbers.

We next investigate the algebra of complex numbers.

Two complex numbers $z:=x+iy$ and $\tau:=u+iv$ are equal if and only if

$$x=u,$$
$$y=v.$$

One can add z and τ by

$$z+\tau=(x+u)+i\,(y+v)\,.$$

Multiplication is defined the same way as usual multiplication in algebra, except that whenever we have i^2 we replace it with -1. That is to say,

$$\begin{aligned}(x+iy)(u+iv) &= xu + ixv + iyu + i^2yv \\ &= xu - yv + i(xv + yu).\end{aligned}$$

One can also easily check that

$$z\tau = \tau z.$$

For $x + iy \neq 0$, we can also define an inverse under multiplication. To compute the inverse, recall that

$$z^{-1}z = 1.$$

Suppose $\tau = z^{-1}$. Then

$$\begin{aligned}\tau z &= 1 \\ \implies xu - yv + i(xv + yu) &= 1 \\ \implies xu - yv &= 1 \\ \text{and } xv + yu &= 0.\end{aligned}$$

Solving the simultaneous system (for example, using matrices)

$$\begin{aligned}xu - yv &= 1, \\ xv + yu &= 0,\end{aligned}$$

we obtain

$$\begin{aligned}u &= \frac{x}{x^2+y^2}, \\ v &= -\frac{y}{x^2+y^2}.\end{aligned}$$

Therefore, for $z = x + iy$, we have

$$z^{-1} = \frac{x}{x^2+y^2} - \frac{y}{x^2+y^2}i = \frac{x-iy}{x^2+y^2}.$$

You can also use this to define division by

$$\frac{\tau}{z} := \tau z^{-1} = (u+iv)\left(\frac{x-iy}{x^2+y^2}\right) = \frac{ux+vy}{x^2+y^2} + \left(\frac{vx-uy}{x^2+y^2}\right)i.$$

Another natural way to see division is by using something called conjugation. Given $z = x + iy \in \mathbb{C}$, we define the complex conjugate of z to be

$$\overline{z} := x - iy.$$

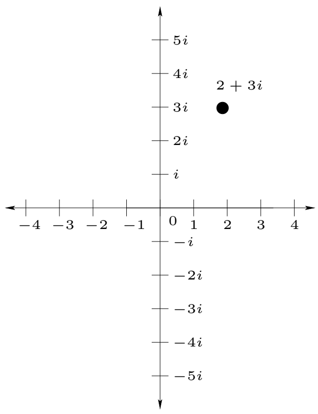
The reason that this is somewhat natural is as follows. Consider

$$z\overline{z} = (x+iy)(x-iy) = x^2 - (iy)^2 = x^2 + y^2.$$

This is now a real number (actually, it is an important real number, but we will get there soon).

Then
$$\frac{1}{z} = \frac{1}{z} \cdot \frac{\overline{z}}{\overline{z}} = \frac{\overline{z}}{z\overline{z}} = \frac{x - iy}{x^2 + y^2}.$$
Conjugation satisfies a number of useful properties. For $z, \tau \in \mathbb{C}$, we have:
$$\overline{z \pm \tau} = \overline{z} \pm \overline{\tau},$$
$$\overline{z\tau} = \overline{z\tau},$$
$$\overline{\frac{z}{\tau}} = \frac{\overline{z}}{\overline{\tau}}.$$

9.2. **Graphical representation of $\mathbb{C}$.** Now consider the following graphical representation of $\mathbb{C}$. For $x+iy \in \mathbb{C}$ (with $x, y \in \mathbb{R}$), we simply make a grid with x- and y-axes. This plane of these x/y-axes is called the complex plane. We put a point at (x, iy) on the complex plane. The point $2 + 3i$ is marked in the graph below:



The x and iy plane is known as the Cartesian plane, and the representation (x, y) is known as the Cartesian coordinates of z. With this graphical interpretation, it is natural to consider the distance between two points. For $z_1 = x_1 + iy_1 \in \mathbb{C}$ and $z_2 = x_2 + iy_2 \in \mathbb{C}$ (with $x_1, x_2, y_1, y_2 \in \mathbb{R}$), we define the distance between z_1 and z_2 by
$$|z_1 - z_2| := \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$
Notice that if $z_1 = x_1, z_2 = x_2 \in \mathbb{R}$, then this simply becomes
$$|z_1 - z_2| = \sqrt{(x_1 - x_2)^2} = |x_1 - x_2|.$$
In particular, we define the absolute value of $z = x + iy \in \mathbb{C}$ by
$$|z| := \sqrt{x^2 + y^2}.$$
Now notice that
$$|z|^2 = x^2 + y^2 = z\overline{z}.$$

Remark. The number $|z|$ is also sometimes called the modulus of z. It satisfies (for $z, \tau \in \mathbb{C}$)
$$|z\tau| = |z| \cdot |\tau|$$

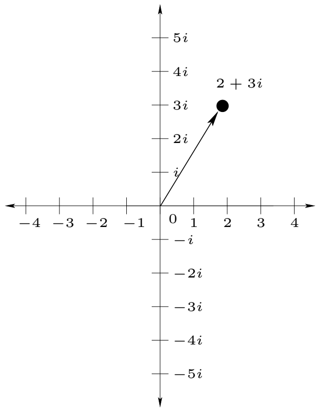
and (for $\tau \neq 0$)

$$\left|\frac{z}{\tau}\right| = \frac{|z|}{|\tau|}.$$

Another important inequality satisfied by the absolute value is the Triangle Inequality, which states that for $z, \tau \in \mathbb{C}$

$$|z + \tau| \leq |z| + |\tau|.$$

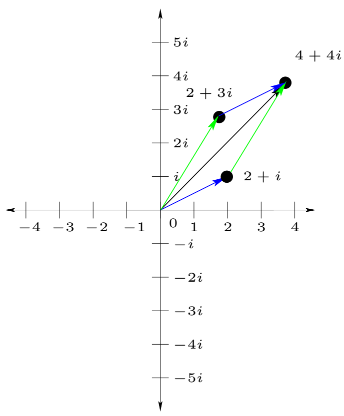
Another way to interpret a complex number is to think of it as a line starting at the origin 0 and going to the point on the complex plane, instead of just a point on the complex plane. This is depicted below:



The line is called a vector. This also gives a new interpretation of addition. If one thinks of these vectors as moveable (so that they don't have to start at the origin), then you can put one arrow starting at the end of another arrow. The result is the addition of these two vectors (or equivalently, the addition of the two complex numbers that these vectors represent). An example of

$$(2 + 3i) + (2 + i) = 4 + 4i$$

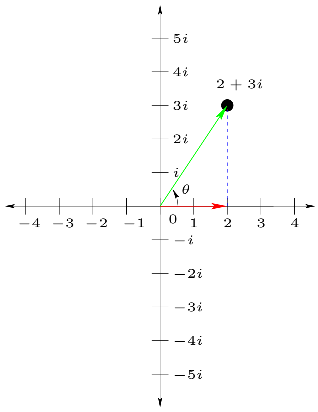
is depicted in the graph below:



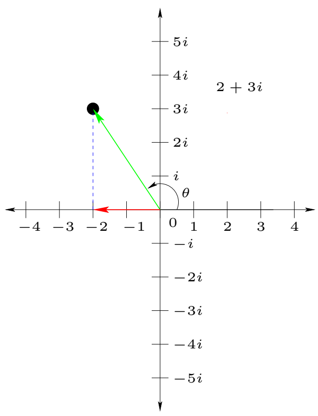
9.3. **Polar coordinates.** We next consider another geometric representation of complex numbers. A complex number $z \in \mathbb{C}$ is uniquely determined by the absolute value

$$r := |z|$$

of z and the angle θ of the counter-clockwise-oriented arc beginning from the positive half of the x-axis and ending at z (this will be further explained below). This representation is known as polar coordinates. For z in the first quadrant of the x and iy plane, the angle, then just make the right triangle which is formed by the x axis and the complex number (considered as a vector). That is to say, the endpoints are 0, z, and $\mathrm{Re}(z)$. The angle meant above is precisely the interior angle of the triangle at the corner at the endpoint 0:



If z is in the other quadrants, then we start at the same point on the x-axis and make an arc to the vector representing z. The angle of this arc is then θ:



The number θ is sometimes called the argument of z. This number is only really unique up to addition by multiples of 2π. In other words, i has argument $\frac{\pi}{2}$, but we could also say that i has argument

$$\frac{\pi}{2} + 2\pi = \frac{5\pi}{2}.$$

This occurs by wrapping one time around before getting to z. Throughout we choose to always pick the choice of θ for which $0 \leq \theta < 2\pi$.

Write $z = x + iy$ with $x = \mathrm{Re}(z) \in \mathbb{R}$ and $y = \mathrm{Im}(z) \in \mathbb{R}$ (the real and imaginary parts). Again denoting $r := |z|$, trigonometry on the triangle in the above diagram yields

$$x = r\cos(\theta),$$
$$y = r\sin(\theta).$$

Using

$$r = \sqrt{x^2 + y^2},$$

we also get the formula

$$\cos(\theta) = \frac{x}{\sqrt{x^2+y^2}},$$
$$\sin(\theta) = \frac{y}{\sqrt{x^2+y^2}}.$$

The above two sets of formulas allow us to go back and forth between Cartesian coordinates and polar coordinates.

Example 9.2. Consider the complex number $z := \frac{1+i\sqrt{3}}{2}$. Give z in Cartesian coordinates and polar coordinates.

Solution: The Cartesian coordinates are precisely given above by

$$\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right).$$

To obtain the polar coordinates, we have to take the absolute value

$$|z| = \sqrt{\left(\frac{1}{2}\right)^2 + \left(\frac{\sqrt{3}}{2}\right)^2} = \sqrt{\frac{1+3}{4}} = 1.$$

We also compute

$$\cos(\theta) = \frac{x}{|z|} = \frac{\frac{1}{2}}{1} = \frac{1}{2}.$$

We now recall that

$$\cos\left(\frac{\pi}{3}\right) = \frac{1}{2},$$

yielding (note that this is only unique up to adding multiples of 2π, but we have restricted the possible choices of θ above)

$$\theta = \frac{\pi}{3}.$$

Hence the polar coordinates are

$$(r, \theta) = \left(1, \frac{\pi}{3}\right).$$

□

Since $\sin(-\theta) = i\sin(\theta)$ and $\cos(-\theta) = \cos(\theta)$, the polar coordinates for $\overline{z}$ (if the polar coordinates of $z = x + iy$ are (r, θ)) are given by

$$\begin{aligned}\overline{z} &= x - iy \\ &= r\cos(\theta) - ir\sin(\theta) \\ &= r\cos(-\theta) + ir\sin(-\theta).\end{aligned}$$

Hence, up to adding multiples of 2π to the angle, the polar coordinates of $\overline{z}$ are

$$(r, -\theta).$$

Multiplication has an interesting form when we take polar coordinates. Suppose that the polar coordinates for z_1 is (r_1, θ_1) and the polar coordinates for z_2 is (r_2, θ_2). Using the formula for x_1, y_1, x_2, and y_2 (going back to Cartesian coordinates), we have

$$\begin{aligned}z_1 &= x_1 + iy_1 = r_1 (\cos(\theta_1) + i\sin(\theta_1)), \\ z_2 &= x_2 + iy_2 = r_2 (\cos(\theta_2) + i\sin(\theta_2)).\end{aligned}$$

Then

$$\begin{aligned}z_1 z_2 =& r_1 r_2 (\cos(\theta_1) + i\sin(\theta_1))(\cos(\theta_2) + i\sin(\theta_2)) \\ =& r_1 r_2 \Big(\cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2) \\ &+ i(\sin(\theta_1)\cos(\theta_2) + \sin(\theta_2)\cos(\theta_1)) \Big).\end{aligned}$$

Now recall that

$$\cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2) = \cos(\theta_1 + \theta_2)$$

and

$$\sin(\theta_1)\cos(\theta_2) + \sin(\theta_2)\cos(\theta_1) = \sin(\theta_1 + \theta_2).$$

Therefore, we have

$$z_1 z_2 = r_1 r_2 (\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)).$$

Hence the polar coordinates of $z_1 z_2$ is (up to possibly subtracting a multiples of 2π to get θ between 0 and 2π)

$$(r_1 r_2, \theta_1 + \theta_2).$$

The angles nicely add together and the absolute values multiply. This makes multiplying in polar coordinates rather easy.

9.4. **De Moivre's Theorem.** In the case that $z_1 = z_2 = z$ with polar coordinates (r, θ), we get a nice formula

$$z^2 = r^2 \left(\cos(2\theta) + i\sin(2\theta)\right).$$

Multiplying this by z, we get multiply by r again and add θ. Hence

$$z^3 = r^3 \left(\cos(3\theta) + i\sin(3\theta)\right).$$

Repeating this, for $n \in \mathbb{N}$ we have

$$z^n = r^n \left(\cos(n\theta) + i\sin(n\theta)\right).$$

For $n = 0$, we have $z^0 = 1$, so this also holds for $n = 0$. For $n = -1$, we have

$$z^{-1} = \frac{\overline{z}}{|z|^2}.$$

Recall that $|z| = r$ and (by the polar coordinates of $\overline{z}$

$$\overline{z} = r\left(\cos(-\theta) + i\sin(-\theta)\right),$$

we have

$$z^{-1} = \frac{r\left(\cos(-\theta) + i\sin(-\theta)\right)}{r^2} = \frac{1}{r}\left(\cos(-\theta) + i\sin(-\theta)\right).$$

Thus

$$z^n = r^n \left(\cos(n\theta) + i\sin(n\theta)\right)$$

holds for $n = -1$ as well.

Using the fact that

$$z^{-n} = \left(z^{-1}\right)^n$$

and then the computation of the polar coordinates for the product, we obtain

$$z^{-n} = \left(r^{-1}\left(\cos(-\theta) + i\sin(-\theta)\right)\right)^n = r^{-n}\left(\cos(-n\theta) + i\sin(-n\theta)\right).$$

We have hence concluded the following.

Theorem 9.3 (De Moivre's Theorem)**.** *For every $n \in \mathbb{Z}$ and $z \in \mathbb{C}$ with $z \neq 0$, we have*

$$z^n = r^n \left(\cos(n\theta) + i\sin(n\theta)\right).$$

9.5. **Roots of unity.** For $n \in \mathbb{N}$, we call z an <u>nth root of unity</u> (or <u>nth root of 1</u>) if

$$z^n = 1.$$

By De Moivre's formula, we have

$$z^n = r^n \left(\cos(n\theta) + i\sin(n\theta)\right).$$

We immediately conclude that

$$\begin{aligned} r^n &= 1, \\ \cos(n\theta) &= 1, \\ \sin(n\theta) &= 0. \end{aligned}$$

Since $r \in \mathbb{R}$ and $r^n = 1$, we have

$$|r| = 1.$$

Since $r \geq 0$, we obtain $r = 1$. Now recall that

$$\cos(\delta) = 1$$

if and only if

$$\delta = 2\pi m$$

for some m. Moreover $\sin(2\pi m) = 0$. But then

$$n\theta = 2\pi m.$$

We conclude that

$$\theta = \frac{2\pi m}{n}.$$

The restriction

$$0 \leq \theta < 2\pi$$

is equivalent to

$$0 \leq m < n.$$

Setting

$$\omega_n := \cos\left(\frac{2\pi}{n}\right) + i\sin\left(\frac{2\pi}{n}\right),$$

the polar coordinates of ω_n^m (again using De Moivre's Theorem) are

$$\left(1, \frac{2\pi m}{n}\right).$$

Therefore

$$1, \omega_n, \omega_n^2, \ldots, \omega_n^{n-1}$$

are all of the nth roots of unity.

More generally, for $w \in \mathbb{C}$, we call z an <u>nth root of w</u> if

$$z^n = w.$$

De Moivre's Theorem also give us a method to find nth roots of other complex numbers.

Example 9.4. Find all nth roots of $w := -\sqrt{2} + \sqrt{2}i$

Solution: We first compute the polar coordinates for w. We have

$$r = |w| = \sqrt{\left(-\sqrt{2}\right)^2 + \left(\sqrt{2}\right)^2} = \sqrt{4} = 2.$$

Moreover, we have

$$\cos(\theta) = -\frac{\sqrt{2}}{2}$$

and

$$\sin(\theta) = \frac{\sqrt{2}}{2}.$$

Thus

$$\theta = \frac{3\pi}{4} + 2\pi m,$$

for some $m \in \mathbb{Z}$. Following the same argument given for finding the nth roots of unity, if z (with polar coordinates r and θ) is an nth root of w, then

$$r^n = |z|^n = 2$$
$$\theta = \frac{3\pi}{4n} + \frac{2\pi m}{n}.$$

Here $m \in \mathbb{Z}$. Dividing by the nth root of unity ω_n, we have

$$\alpha := \frac{z}{\omega_n^m} = z\left(\omega_n^{-m}\right).$$

But then $|\alpha| = |z|$ and the θ corresponding to α is

$$\frac{3\pi}{4n}.$$

It follows that every nth root of w can be written as

$$\alpha\omega_n^m,$$

with

$$m = 0, 1, \ldots, n-1.$$

□

9.6. **The Fundamental Theorem of Algebra.** Notice that by using De Moivre's Theorem, we found exactly n solutions in $\mathbb{C}$ to the equation

$$z^n - w = 0.$$

A theorem that we saw earlier says that there are at most n solutions to a polynomial of order n. This is a special case of a more general theorem.

Theorem 9.5 (The Fundamental Theorem of Algebra)**.** *Any polynomial equation*

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$$

where $a_0, \ldots, a_n \in \mathbb{C}$ *with* $a_n \neq 0$ *and* $n \geq 1$ *has a solution in* $\mathbb{C}$.

Remark. An alternate version of the Fundamental Theorem of Algebra states that every polynomial equation

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0$$

where $a_0, \ldots, a_n \in \mathbb{C}$ with $a_n \neq 0$ and $n \geq 1$ has exactly n solutions in $\mathbb{C}$.

If z is a root from the theorem, then the proof basically follows by factoring $x - z$ out of the polynomial (using polynomial long division). This yields a polynomial with complex coefficients and degree one smaller. But this new polynomial must also have a root. Continuing in this direction, we get exactly n roots.

10. First and Second Order Differential Equations

A differential equation is an equation which has both a function and its derivative. For example, you could have an equation

$$f''(x) + f'(x) + xf(x) = \sin(x).$$

This is often written

$$y'' + y' + xy = \sin(x)$$

or

$$\frac{d^2y}{dx^2} + \frac{dy}{dx} + xy = \sin(x).$$

In a sense, we have already seen certain differential equations. Consider

$$y' = x^2 y,$$

or, equivalently,

$$\frac{y'}{y} = x^2.$$

In order to solve this differential equation, we integrate on both sides to obtain

$$\ln(y) = \frac{x^3}{3} + c.$$

Then

$$y = e^{\frac{x^3}{3}+c} = e^c e^{\frac{x^3}{3}}.$$

It is usual to rewrite

$$e^c$$

as C or y_0, so

$$y = Ce^{\frac{x^3}{3}}.$$

The notation y_0 is sometimes used because $y = y_0$ when $x = 0$.

A first-order linear differential equation is an equation of the form

$$y' + f(x)y = g(x),$$

where f and g are functions. The above example would be the case $f(x) = -x^2$ and $g(x) = 0$, since

$$\frac{y'}{y} = -f(x).$$

These are called linear because both y and y' only occur to the first power and they are not parameters of a function. As a function of y, the coefficients $f(x)$ and $g(x)$ are constants, so it is like a linear function where y' is the other variable.

In the first example that we gave (with $g(x) = 0$), the variables x and y could be separated. In general, if $g(x) = 0$, then we have

$$\frac{y'}{y} = f(x).$$

Therefore, integrating on both sides,

$$\ln(y) = \int f(x)dx + C.$$

Thus

$$y = Ce^{\int f(x)dx}.$$

Therefore, if we know how to integrate $f(x)$, then we know how to solve the differential equation.

Now let's try to solve a general linear differential equation. Multiply the left-hand side by a function $h(x)$

$$h(x)y' + f(x)h(x)y.$$

This now looks a bit like the derivative from the product rule. Indeed, if

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

then

$$(h(x)y)' = h'(x)y + h(x)y'.$$

The point is that by the second Fundamental Theorem of Calculus, we now know the integral

$$\int (h(x)y)'\, dx = h(x)y + C.$$

We then get y showing up so that we can solve for it.

We now find out what $h(x)$ should be. Since

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

we have

$$\frac{h'(x)}{h(x)} = f(x).$$

This is exactly like the separable differential equation that we solved initially. Integrating, we obtain

$$\ln(h(x)) = \int f(x)dx + C.$$

Thus we have

$$h(x) = Ce^{\int f(x)dx}.$$

For this particular choice of h, we now obtain

$$h(x)y' + f(x)h(x)y = g(x)h(x).$$

We rewrite the left-hand side to obtain

$$(h(x)y)' = g(x)h(x).$$

Integrating both sides yields

$$h(x)y = \int g(x)h(x)dx + C.$$

Example 10.1. Solve the linear differential equation

$$x^2y' + xy = x^3.$$

Solution: We first rewrite this in the standard form (with y' by itself)

$$y' = \frac{1}{x}y = x.$$

We now integrate

$$\int \frac{1}{x}dx = \ln(x) + C.$$

Thus

$$h(x) = e^{\int \frac{dx}{x}} = e^{\ln(x)} = x.$$

Therefore, multiplying by x gives the differential equation

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

The left hand side is $(xy)'$, so we have

$$(xy)' = x^2.$$

Integrating on both sides yields

$$xy = \frac{x^3}{3} + C.$$

Therefore

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

We can now check that y satisfies the differential equation directly. We have

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

Thus

$$x^2y' + xy = x^2\left(\frac{2x}{3} - \frac{C}{x^2}\right) + x\left(\frac{x^2}{3} + \frac{C}{x}\right) = \frac{2x^3}{3} - C + \frac{x^3}{3} + C = x^3.$$

This is exactly the differential equation that we wanted to show. □

Example 10.2. Suppose that

$$\frac{dy}{dt} + 7y = 5t$$

and $y(0) = 2$ (this means that $y = 2$ when $t = 0$). Solve the differential equation.

Remark. Such a differential equation is known as an initial value problem, because the initial value $y(0)$ is given. The variable t usually stands for **time** and hence the initial value is the value at the starting time.

Solution: We first compute

$$e^{\int 7dt} = e^{7t}.$$

Multiplying by e^{7t} then yields

$$\frac{d}{dt}\left(e^{7t}y\right) = e^{7t}\frac{dy}{dt} + 7e^{7t}y = 5te^{7t}$$

Integrating on both sides gives

$$e^{7t}y = 5\int te^{7t}dt.$$

Using integration by parts (with $u = t$ and $dv = e^{7t}$), the integral is

$$\int te^{7t}dt = \frac{t}{7}e^{7t} - \frac{1}{7}\int e^{7t}dt = \frac{t}{7}e^{7t} - \frac{1}{49}e^{7t} + C.$$

We then obtain

$$e^{7t}y = \frac{t}{7}e^{7t} - \frac{1}{49}e^{7t} + C$$

Thus

$$y = \frac{5t}{7} - \frac{5}{49} + Ce^{-7t}.$$

Since

$$y(0) = -\frac{5}{49} + C$$

and $y(0) = 2$ by assumption, we obtain

$$2 + \frac{5}{49} = C.$$

Thus

$$C = \frac{103}{49}.$$

It follows that

$$y = \frac{5t}{7} - \frac{5}{49} + \frac{103}{49}e^{-7t}.$$

□

We next consider an application of differential equations. Suppose that you are mixing two solutions of water mixed with salt. One is already in a large tank and the other solution is being pumped in. However, some of the mixed solution from the tank is simultaneously being pumped out (for use somewhere else). Let's say that you begin with a 1000 liters of water with 200 grams of salt mixed in it. You then add a solution which has 1 gram per liter. The solution is being pumped in at 110 liters per minute and the mixture in the tank is being pumped out at 120 liters per minute. The question is, what is the amount of salt in the container at any time t? We first find the volume of water in the tank at time t (measured in minutes). To do so, note that the change in volume is the rate of the water coming in minus the rate of the water going out, or in other words

$$V'(t) = 110 - 120 = -10.$$

But then (integrating, or solving this simple differential equation), we have

$$V(t) = C - 10t.$$

However, we know that

$$C = V(0) = 1000.$$

Therefore

$$V(t) = 1000 - 10t.$$

If $y(t)$ is the amount of salt (in grams) in the water after t minutes, then the mixture in the tank has

$$\frac{y(t)}{V(t)}$$

grams of salt per liter of mixture. Since the mixture is leaving the tank at a speed of 120 liters per minute, so salt is leaving the tank at the speed of

$$120 \cdot \frac{y(t)}{V(t)}$$

grams per minute.

Recall now that the water coming in added 1 gram per liter at a rate of 110 liters per minute. Therefore, salt is entering the tank at a rate of 110 grams per minute. Overall, the change in the amount of salt is

$$110 - \frac{120}{V(t)} y(t)$$

grams per minute. Written as a differential equation, this is

$$\frac{dy}{dt} = 110 - \frac{120}{V(t)} y = 110 - \frac{120}{1000 - 10t} y.$$

This is a first-order linear differential equation. We now solve the differential equation to determine the amount of salt y in the water at time t.

We first rewrite the differential equation as

$$\frac{dy}{dt} + \frac{12}{100 - t} y = 110.$$

We multiply on both sides by

$$e^{\int \frac{12}{100-t} dt} = e^{-12 \ln(100-t)} = (100 - t)^{-12}.$$

Hence

$$(100 - t)^{-12} \frac{dy}{dt} + \frac{12}{100 - t} (100 - t)^{-12} y = 110 (100 - t)^{-12},$$

which can be rewritten as

$$\frac{d}{dt} \left((100 - t)^{-12} y \right) = 110 (100 - t)^{-12}.$$

Integrating both sides yields

$$(100-t)^{-12}\,y = 110\int (100-t)^{-12}\,dt = \frac{110}{11}\,(100-t)^{-11} + C.$$

In other words

$$y = 10(100-t) + C\,(100-t)^{12}\,.$$

Finally, we use the fact that there was originally 200 grams of salt (namely, $y(0) = 200$ to conclude that

$$C(100)^{12} = 200 - 1000 = -800.$$

Therefore

$$C = -8\cdot 10^{-22},$$

so

$$y = 10(100-t) - 8\cdot 10^{-22}\,(100-t)^{12}\,.$$