Calculus 1 – Spring 2019 Section 2

Jacob Shapiro

April 15, 2019

Contents

1 Logistics

- Instructor: Jacob Shapiro shapiro@math.columbia.edu
- Course website: http://math.columbia.edu/~shapiro/teaching.html
- Location: 207 Mathematics Building
- Time: Mondays and Wednesdays 4:10pm-5:25pm (runs through January 23rd until May 6th 2019 for 28 sessions).
- Recitation sessions: Fridays 2pm-3pm in Hamilton 602.
- Office hours: Tuesdays 6pm-8pm and Wednesdays 5:30pm-7:30pm (or by appointment), in 626 Mathematics (starting Jan 29th).
- Teaching Assistants:
  - Donghan Kim dk2571@columbia.edu
  - Mat Hillman mh3691@columbia.edu
  - Ziad Saade zs2351@columbia.edu
- TAs office hours: Will be held in the help room (502 Milstein Center) http://www.math.columbia.edu/general-information/help-rooms/502-milstein/ during the following times:

  - Mat: Fridays 10am-12pm.
  - Donghan: Thursdays 4pm-6pm.
  - Ziad: Tuesdays 4pm-6pm.

- Misc. information: Calculus @ Columbia http://www.math.columbia.edu/programs-math/undergraduate-program/calculus-classes/

- Getting help: Your best bet is the office-hours, and then the help room. If you prefer impersonal communication, you may use the Piazza website to pose questions (even anonymously, if you're worried). TAs will monitor this forum regularly.

- Textbook: Lecture notes will be made available online. I will do my best to follow the Columbia Calculus curriculum so as to make sure you can go on to Calculus II smoothly. I will use material from various textbooks, some of which include: Spivak [5], Apostol [1], Courant [2] and sometimes Stewart [6] just to set the timeline (since this is what Columbia's curriculum is based upon).

- Homework assignments: Assignments will appear on this website after the Monday lecture every week and are meant to be solved before the Monday lecture of the following week. Official solutions will be published here one week before the relevant mid-term. You do not have to hand in assignments every week, but you may present your solutions to the TAs or me to gain extra credit (see below).

- Grading: There will be two midterms during lecture time (see below for dates) and one final (after the last lecture). Your grade is (automatically) the higher of the following two options: [Option A] The final carries 50% weight, the midterms each carry 25% weight for a total of 100% of your grade. [Option B] The final carries 40% weight, the midterms each carry 25% weight for a total of 90% of your grade. The remainder 10% is given to you if you succeed in accomplishing the following task at least three times during the semester: show up (no appointment necessary) to one of the office hours of the TAs or me, successfully orally present a solution to one of the homework assignments whose solution has not been published yet (should take you not more than 10 minutes max), and in the same occasion submit a neatly written down solution to the thing you just presented.

1.1 How these lecture notes are meant

These lecture notes are meant as a support for the lectures, by augmenting with a few more details and possibly enriching with more complete explanations and examples that I might not have had time to go over during the short lectures. Since they are updated on a rather weekly basis, sometimes the earlier sections which had already been covered in class, it is encouraged *not* to print out the PDF but rather read it in digital form.

Strictly speaking the material of calculus really starts in Section 4 onward (Section 2 is a philosophical motivation and Section 3 sets up the language and notation which is the basis of how we think about the various objects we deal with).

Since calculus is *not* a proof-oriented class, most of the statements in this text are *not* substantiated by a demonstration that explains *why* they are correct (i.e., a proof), be it formal or not. Sometimes I chose to ignore this principle and include the proof in the body of the text anyway, mostly because I felt the demonstration was not very much beyond what would be required of an average calculus student, and to cater to those readers who want to go a bit deeper. The reader can recognize very easily these proofs because they start with a *Proof* and are encased in a box. The contents of these proofs is *not* part of the curriculum of the class and will not be required for the midterms or final.

2 Goal of this class and some motivation

This being one of the first classes you take in mathematics–even if you are not a math major–has as its goal to expose you to mathematical thinking, which could be thought of as a way to communicate and reason about abstract notions in an efficient and precise (i.e. the opposite of vague) way. As such, it is first and foremost a language that one has to study. Like your mother tongue, you are exposed to math even before you know what a language is, or that you're undergoing the process of studying it. In this regard, one of our goals in this class is to develop (still in a very naive level) the distinction between using the language (when we use math to calculate the tip in a restaurant) and studying that language to expand our horizons of thinking.

In order to do the latter (when we're already adults–for children this is easier), we must take a step back and talk about things that might look obvious at first, or even unnecessary to discuss, simply in order to level the playing field and make sure that we all (hopefully) understand each other and mean the same thing when we use a certain 'phrase'. In this vein, the way I'm about to describe our path might seem very opaque or pedestrian. However the promise is that if you stick to this very sturdy banister you will be able to tread with confidence to whichever new uncharted territories we reach.

2.1 Naive description of math

In math we deal with certain abstract objects, we give them names, or labels, we consider various relations between them, and rules on how to manipulate them. These rules are man-made, so to speak. There is very vast philosophy on what extent this man-made universe corresponds to our physical reality in existence (e.g. the connection between physics and math: does the physical reality build or even constrain the kind of abstract mathematical structures that we can come up with?) but the fact is that we don't *need* an association with reality in order to do math, and indeed, many branches of math have nothing to do with reality at all. They are essentially studying the abstract structures that they themselves have invented.

Imagine that you are playing the popular board game called monopoly. It is loosely based on an economic system but at the end of the day it is a game with rules who were invented by people and played by people. We could spend time studying and exploring the various possibilities that arise as one plays the game of monopoly. This would be one form of mathematical activity.

One of the main "mechanisms", so to speak, of math making, is having an arc to the story: we start with the given structure, and extract out of it certain constraints that must hold given this structure. This is the basic mechanism of *logic*, where for example if we say "this person is a student" and "students attend lectures" we realize it must be the case that "this person attends lectures". It is this process that we will go through again and again, first describing the structures which we encounter and then "extracting" out of them new constraints.

Math can be done strictly with words (as I've been describing it so far) and indeed this was mostly the approach taken in previous centuries. However, more and more mathematicians realized that it is more efficient to use abbreviating graphical symbols to lay down the abstract objects, structures and relations of math. One writes down these graphical symbols on a piece of paper, on a blackboard, or increasingly, into a computer, and this is a crucial way in which we communicate about math nowadays, interlacing these graphical messages within paragraphs of text which are supposed to introduce the rationale and heuristics of what is really happening with the graphical symbols.

The graphical symbols are roughly organized as follows:

- A Latin or Greek single alphabet letter to denote the objects: $a, b, c, \cdots, x, y, z$.
- Punctuation marks, "mathematical symbols" denote structures and relations: $(), \%, *, +, /, \cdots, <, =$.
- Of course we have the numbers themselves, which for our sake can be thought of again as abstract objects but with honorary special labels: $1, 2, 3, \ldots$.

2.2 What is Calculus?

What has been described thus far could fit all of math in general. However, we will venture into one particular area of math called "calculus". Calculus means in Latin "a pebble or stone used for counting", and nowadays it is (mostly) the word used to refer to a body of knowledge developed by Newton and Leibniz around the mid 1600's in order to study continuous rates of change of quantities (for example instantaneous velocity of a physical object) or the accumulation of quantities (e.g. the distance a physical object has traversed after a given amount of time, given its acceleration).

While the main impetus to for the study of calculus comes from real life questions, we will mostly fit it into our abstract framework so that we have a "safe" way to deal with it without making conceptual mistakes.

The main tool of calculus is the mathematical concept of a limit. The limit has a stringent abstract definition using the abstract language, but intuitively it is the end result of an imagined process (i.e. a series of steps) where we specify the first few steps and imagine (as we cannot *actually*) to continue the process forever and ask what would be the end result.

2.1 Example. Let us start with 1, then go to $\frac{1}{2}$, then $\frac{1}{3}$, $\frac{1}{4}$, and so on. Now imagine that we continue taking more and more steps like this. What would be the end result? The answer is zero, even though zero is never encountered after *any* finite number of steps of this activity.

The concept of the limit is at the heart of anything we will do in this class, and in particular, taking limits of sequences of numbers, where we have given rules for generating the sequences of numbers (these are called *functions*).

Let us start now slightly more formally from the beginning.

3 Naive Naive Set Theory

Set theory is a (complex) branch of mathematics that stands at its modern heart. Naive set theory is a way to present to mathematicians who are growing up some of ideas of set theory in a way that obscures some of the hardest questions of actual set theory and allow them to get going with math. What we will do is naive "naive set theory", which means we will very informally describe what we need to setup a basic common language to allow us to study calculus in an efficient way. The advantage will be that we will then have an entry point to other branches of mathematics as well, such as actual naive set theory, but also the beginning of analysis, topology or algebra. The best source for studying more about set theory is Paul Halmos' book of the same name [3].

Set theory is a corner of mathematics where the abstract objects are collection of yet other abstract objects (you will learn math often likes to be cute like that). So a set is a collection of things, what they are–we don't have to specify. These things that a set contains could be sets themselves (this leads to nice paradoxes). One graphical way to describe sets is using curly brackets, i.e. the object (the set) which contains the objects a, b and c is graphically denoted by $\{ a, b, c \}$. Note how one uses commas to separate distinct objects. While with words it is obvious that the set that contains the objects a and b is the same thing as the set that contains the objects b and a, in graphical symbols these are a-priori two different things

$$\{ a, b \} \quad \text{versus} \quad \{ b, a \}$$

but we *declare* that these two graphical symbols refer to one and the same thing. Sometimes it is convenient to use three dots to let the reader know that a certain number of steps should continue in an obvious fashion. For instance, it is obvious that $\{ a, b, \dots, f \}$ really means $\{ a, b, c, d, e, f \}$. Other times the three dots mean a hypothetical continuation with no end, such as the case of

$\{\,\text{today}, \text{tomorrow}, \text{the day after tomorrow}, \ldots\,\}$ where it is clear that there will not be a final step to this process (setting aside fundamental questions about compactness of spacetime and the universe), and that's OK, since we actually want to consider also hypothetical procedures. Such hypothetical procedures are at the heart of *limits*, which lie at the heart of calculus. We say that such sets, whose construction is hypothetical, have *infinite* size.

We often consider sets whose elements are numbers, as numbers for us are currently just abstract mathematical objects, there should be no hindrance to consider the set $\{\,1\,\}$ if we also consider the set $\{\,a\,\}$, after all the graphical symbols 1 or a are just labels. Using the bracket notation, we *agree* that there is no "additional" meaning to the graphical symbol $\{\,a, a\,\}$, i.e., it merely means the same thing as $\{\,a\,\}$.

Since sets themselves are abstract mathematical objects, we can just write some letter, such as A or X or even a, to refer to one of them, rather than enumerating its elements every time. Since what we mostly care about when dealing with sets are their contents, i.e., the list of elements, it is convenient to also have a graphical symbol to state whether an object (an element) resides in a set or not. This is denoted via

$$\begin{array}{lll} a \in A & \text{means} & \text{The object } a \text{ lies in the set } A. \\ a \notin A & \text{means} & \text{The object } a \text{ does not lie in the set } A. \end{array}$$

Note that using this graphical notation, it is clear that whatever appears to the right of $\in$ or $\notin$ must be a set.

Since all we know about sets is that they contain things, then we can "specify" a set A by simply enumerating its contents, e.g. by using the curly-bracket graphical notation. The special way to say that is using the equal $=$ symbol:

$$\begin{array}{lll} A = \{\,a, b, c\,\} & \text{means} & A \text{ is the set } \{\,a, b, c\,\} \\ & \text{means} & A \text{ is the set whose elements are } a, b \text{ and } c. \end{array}$$

In many occasions, instead of describing a complicated situation in words, it is many times easier to simply be able to refer to a set that contains *absolutely nothing at all* (like an empty basket). This empty set is denoted by $\varnothing = \{\ \}$ (if you think about it, there is only one such set, because since all we care about sets is the list of things they contain, if both lists are equal (or empty) then both sets are equal–so there is only *one* empty set).

$$a \in \varnothing \text{ is always a false statement, regardless of what } a \text{ is.}$$

It is also convenient to have graphical notation that builds new sets out of pre-given ones. For example:

- Union with the graphical symbol $\cup$: $A \cup B$ is the set containing all elements in *either* A or B. Example: $A = \{\,x, y\,\}$, $B = \{\,\beta, ♫\,\}$, $A \cup B = \{\,x, \beta, y, ♫\,\}$ (as we said, order doesn't matter, and we are also not bound to use Latin alphabet for labels of abstract objects.

- Set difference with the symbol $\setminus$: $A \setminus B$ is the set containing elements in A which are *not* in B. Example: $A = \{ 1, 2 \}$, $B = \{ 2 \}$ gives $A \setminus B = \{ 1 \}$. But $B \setminus A = \varnothing$.

- Intersection, with $\cap$: $A \cap B$ are all elements that are in both A *and* B. Example: $A = \{ 1, 2 \}$, $B = \{ 2 \}$ gives $A \cap B = \{ 2 \}$ but $A = \{ 1 \}$ and $B = \{ 2 \}$ gives $A \cap B = \varnothing$.

Another important relation is not just between objects and sets, but also between sets and sets. Sometimes we would like to know whether all elements in one set are also in another set. This is called a *subset* and is denoted as follows

$$A \subseteq B \quad \text{means} \quad a \in B \text{ whenever } a \in A \text{ for any object } a.$$

To verify that two sets A and B are the same (since what defines them is the list of elements they contain), we must make sure of two things: $A \subseteq B$ and $B \subseteq A$. Hence

$$A = B \quad \text{means} \quad A \subseteq B \wedge B \subseteq A$$

(note $\wedge$ is the graphical symbol for the logic of 'and', and we will also substitute $\Leftrightarrow$ for 'means').

3.1 Example. $(A \cap B) \cap C = A \cap (B \cap C)$ (that is, the order of taking intersection does not change the end-result set). To really see why this is true, let us proceed step by step. Suppose that $x \in (A \cap B) \cap C$. That means $x \in A \cap B$ and $x \in C$. But $x \in A \cap B$ means $x \in A$ and $x \in B$. Hence all together we learn that the following are true: $x \in A$, $x \in B$ and $x \in C$. This would be the same end conclusion if we assumed that $x \in A \cap (B \cap C)$. What we have learnt is that $x \in A \cap (B \cap C)$ whenever $x \in (A \cap B) \cap C$ for any x. This is what we said $A \cap (B \cap C) \subseteq (A \cap B) \cap C$ means. This is half of the equality $=$ statement. The other half proceeds in the same way.

- An important structure is that of *a product of sets*. Given any two sets A and B, we can form their product set $A \times B$. This is a new set whose elements consist of sets themselves. If $A = \{ a, b, c, \ldots \}$ and $B = \{ x, y, z, \ldots \}$ then the set $A \times B$ is given by

  $$\begin{aligned} &\{\{ \{ a, 1 \}, \{ x, 2 \} \}, \{ \{ b, 1 \}, \{ x, 2 \} \}, \{ \{ c, 1 \}, \{ x, 2 \} \}, \ldots \\ &\ldots, \{ \{ a, 1 \}, \{ y, 2 \} \}, \{ \{ b, 1 \}, \{ y, 2 \} \}, \{ \{ c, 1 \}, \{ y, 2 \} \}, \ldots \\ &\ldots, \{ \{ a, 1 \}, \{ z, 2 \} \}, \{ \{ b, 1 \}, \{ z, 2 \} \}, \{ \{ c, 1 \}, \{ z, 2 \} \}, \ldots \} \end{aligned}$$

  The point here is that in addition to form pairs, we also (arbitrarily) refer to A as the first origin set, hence the 1 and B as the second origin set, hence the 2, so that we are keeping track not only of the objects in each pair, but from which origin set they've come from. In this way, $\{ \{ a, 1 \}, \{ x, 2 \} \}$ tells us immediately that a belongs to A and x belongs to B.

Because it is exhausting to write so many curly brackets, we agree on a graphical notation that $(☺, ♫)$ means $\{\,\{\,☺, 1\,\}, \{\,♫, 2\,\}\,\}$ for any $☺ \in A$ and $♫ \in B$. $(☺, ♫)$ is called an *ordered pair*. Example: $\{\,1, 2\,\} \times \{\,\blacktriangle, \blacktriangledown\,\} = \{\,(1, \blacktriangle), (2, \blacktriangle), (1, \blacktriangledown), (2, \blacktriangledown)\,\}$.
Clearly, $(☺, ♫) \neq (♫, ☺)$, because $\{\,\{\,☺, 1\,\}, \{\,♫, 2\,\}\,\} \neq \{\,\{\,☺, 2\,\}, \{\,♫, 1\,\}\,\}$– we can change orders within the curly brackets as we please, but we can't change move objects *across* curly brackets.

Another piece of notation that we want to discuss with sets involves their size. The set $\{\,a\,\}$ contains one object (no matter what a is) so we say its *size* (i.e. the number of objects it contains is 1). Graphically we write two vertical lines before and after the set in question to refer to its size, i.e.

$$\begin{aligned} |\{\,a\,\}| &= 1 \\ |\{\,a, b\,\}| &= 2 \\ &\cdots \\ |\{\,a, b, \ldots, z\,\}| &= 26 \end{aligned}$$

Of course it is useful to agree that $|\varnothing| = 0$ size $\varnothing$ contains no objects. When to enumerate a set we must continue a hypothetical process to no end, for example, the set $\{\,1, 2, 3, \ldots\,\}$, we say that the size of the set is *infinite* and graphically we write the symbol ∞:

$$|\{\,1, 2, 3, \ldots\,\}| = \infty$$

3.2 Definition. When a set is of size one, that is, when it has only one element, we call it a *singleton*. Any set of the form $\{\,a\,\}$ for any object a is a singleton.

In fact it is possible to turn the picture upside down, so to speak, and define the numbers $1, 2, 3, \ldots$ not as intrinsic abstract objects (how we used to think about them so far) but as associated with a hierarchy of sets starting from the empty one, with a natural association between the number we are naively used to and the size of the constructed set:

1. Zero is associated with the set $\varnothing$, and we have $|\varnothing| = 0$. We *define* zero to be the empty set, making the empty set (rather than zero) the more basic object and zero a *derived* object.

2. One is associated with the set $\{\,\varnothing\,\}$. It is a singleton.

3. Two is associated with the set $\{\,\varnothing, \{\,\varnothing\,\}\,\}$

4. etc.

You can find more about this in Halmos' book.

3.3 Definition. A final piece of notation that we shall sometimes use is called *set-builder notation*. This is a way to describe a subset B of a set A. Indeed, our main way to describe sets (whether they are themselves subsets of other sets or not doesn't matter) is to enumerate their contents. For example:

$$A = \text{The set of all people}$$

and the subset

$$B = \text{The set of all Americans}$$

B is indeed a subset of A because all Americans are people (excluding pets and so on). Indeed, B could be thought of as a sub-collection of the objects of A, where all objects in the sub-collection obey a *further* constraint, namely, of being American on top of being people. This is written graphically as

$$B = \{ a \in A \mid a \text{ is American} \}$$

In the above graphical notation, the symbol a is what is called a *variable*. It is a generic label used to refer to any given element in A. To "build" B (or rather to "imagine" all of its content) we must let a "run" through A.

4 Special sets of numbers

So far we have discussed sets as abstract collection of abstract objects. Let us rely (in a sly way) on our pre-existing knowledge of objects which we refer to as numbers, and imagine that we now start collecting them together into sets. We can do whatever we want, so we can for example define a new set which contains all of the following numbers

$$\{ 1, 2, 3, 4, 5, \dots \}$$

Note how now that we use the dots this means that the set has actually an infinite number of elements. This is fine–in fact this is part of what makes calculus interesting at all. This set above is called *the natural numbers* and is denoted with the special graphical symbol $\mathbb{N}$ (blackboard N):

$$\mathbb{N} = \{ 1, 2, 3, 4, 5, \dots \} .$$

Using what we learnt, we know that $0 \notin \mathbb{N}$ yet $666 \in \mathbb{N}$.

4.1 Remark. In different geographical regions of the world, $\mathbb{N}$ may or may not contain zero. In English-speaking environments, we mostly start $\mathbb{N}$ with 1, and we shall do so with no exceptions until the end of the semester.

The next special set is the same thing but extended to the negative side:

$$\{ 0, \pm 1, \pm 2, \pm 3, \dots \} = \{ \dots, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, \dots \}$$

(recall the order of enumeration does not matter)

This set is called *the integers* and is denoted by $\mathbb{Z}$. Finally we would like to include fractions as well

$$\left\{ 0, \pm 1, \pm\frac{1}{2}, \pm\frac{1}{3}, \dots, \pm 2, \pm\frac{2}{3}, \pm\frac{2}{4}, \dots, \pm 3, \pm\frac{3}{2}, \pm\frac{3}{4}, \pm\frac{3}{5}, \dots \right\}$$

i.e. any number that can be written in the form $\frac{p}{q}$ with $p \in \mathbb{Z}$ and $q \in \mathbb{Z}$. In set builder notation we would write

$$\left\{ \frac{p}{q} \;\middle|\; p \in \mathbb{Z} \text{ and } q \in \mathbb{Z} \right\}$$

These are called the rationals and denoted by $\mathbb{Q}$ (for quotient). By the way, rationals are not named such for being extra reasonable. The etymology is from the word 'ratio' which is also a quotient. There is also a decimal notation with finitely many decimal digits after the point or periodic repeating, but let us skip over that for now.

It turns out that there are certain numbers that exist (for example they come from geometry or from physics) yet they are *not* in $\mathbb{Q}$–they are irrational. All of these numbers (which turn out to be the vast majority of all numbers, where majority is meant in a certain sense, as we are trying to compare different notions of infinity) are denoted by $\mathbb{R}$ and are called *real* numbers.

4.2 Remark. We already saw schematically (though not precisely) that there is a way to build from the empty set all of $\mathbb{N}$. There is also a concrete and precise way (out of sets and manipulations of them) to construct $\mathbb{Z}$ out of $\mathbb{N}$, $\mathbb{Q}$ out of $\mathbb{Z}$ and $\mathbb{R}$ out of $\mathbb{Q}$. We will not do so in this class as this material belongs to a field of mathematics called *analysis*. If you are curious look at [4] under *Dedekind cuts*.

4.3 Example. To give example for certain numbers in $\mathbb{R} \setminus \mathbb{Q}$, consider the ratio between the circumference of a circle and its diameter. The ancient Greek realized a while back that this number cannot be written in the form $\frac{p}{q}$ for some $p, q \in \mathbb{Z}$. To see this fact requires actually some work and preparation.

4.4 Example. The square root of the a number is the answer to the question "what number do we multiply by itself to get what we started with?". So $\sqrt{4} = 2$ because $2 \times 2 = 4$, i.e. $\sqrt{4} \times \sqrt{4} = 4$. Can we express $\sqrt{2}$ in a simple way too? Clearly, $\sqrt{1} = 1$ because $1 \times 1 = 1$, so $\sqrt{2}$ must be somewhere between 1 and 2. If we take the middle $1.5 = \frac{3}{2}$ we get $1.5 \times 1.5 = \frac{9}{4} = 2.25$, so 1.5 is already too much. What about 1.4? $1.4 \times 1.4 = 1.96$, so that's already too little! It turns out that $\sqrt{2} \notin \mathbb{Q}$, i.e. $\sqrt{2} \in \mathbb{R} \setminus \mathbb{Q}$. To see this, assume otherwise. Then we have $\frac{p}{q} \times \frac{p}{q} = 2$ for some $p, q \in \mathbb{Z}$. If both p and q are even, we can divide both by 2 and get the same number, so let us assume we have done that so that now $2 = \frac{p^2}{q^2}$ with p, q integers not both even. This is the same as $p^2 = 2q^2$. That means that p^2 is even, i.e. it is of the form $2x$ for some $x \in \mathbb{Z}$. This implies that p is even, i.e., it is of the form $2y$ for some $y \in \mathbb{Z}$ (if p were odd, it would be of the form $2y + 1$, and then $p^2 = (2y + 1)^2 = 4y^2 + 4y + 1$, which is odd!)

That means that actually $p^2 = 4y^2$ for some $y \in \mathbb{Z}$, i.e. $4y^2 = 2q^2$. But then $2y^2 = q^2$, that is, q^2 is even, which implies q is even (as before). So both p and q are even!

So $\mathbb{Q}$ has some "holes", and the purpose of using $\mathbb{R}$ is to have a set that contains *everything.* Indeed the whole point of calculus is limits, and the whole point of limits is to continue procedures hypothetically with no end. These hypothetical procedures are precisely where we may suddenly find ourselves out of $\mathbb{Q}$.

It is that everything set, $\mathbb{R}$, that we *geometrically* interpret as a *continuous line,* i.e. we associated the lack of holes with the concept of *continuum.* That is why in physics when we think of a continuous time evolution, for instance, we model the set of possible times as $\mathbb{R}$. When we think of the set of possible heights a ball could take as it is thrown up the air, we model that set as $\mathbb{R}$, since we imagine physical space to be a continuum with no holes, and $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$ cannot be appropriate to describe the set of all possible physical outcomes. So you should have in your mind a picture of an infinite straight continuous line when you think of $\mathbb{R}$.

As we have seen, we also can consider products of sets, and so $\mathbb{R} \times \mathbb{R}$ could be considered the set of all possible pairs of continuum values, that is, a plane of continuum. For convenience we write $\mathbb{R}^2$ instead of $\mathbb{R} \times \mathbb{R}$. This set of pairs should be geometrically pictures as an infinite plane. Physical space, everything around us, is $\mathbb{R}^3$ (at least in Newtonian mechanics).

4.1 Intervals

Sometimes it is convenient to specify subsets of $\mathbb{R}$, which are intervals. Given any two endpoints $a \in \mathbb{R}$ and $b \in \mathbb{R}$ such that $a < b$, we define the following sets

$$\begin{aligned} [a,b] &:= \{ x \in \mathbb{R} \mid a \leq x \leq b \} \\ (a,b) &:= \{ x \in \mathbb{R} \mid a < x < b \} \\ (a,b] &:= \{ x \in \mathbb{R} \mid a < x \leq b \} \\ [a,b) &:= \{ x \in \mathbb{R} \mid a \leq x < b \} \end{aligned}$$

The first of which is called the *closed interval* between a and b, the second of which the *open interval* between a and b. The last two don't have special names.

Sometimes it is useful to have the restriction only on one side to obtain a half-infinite interval, that is, to consider the set of all numbers larger than a for some $a \in \mathbb{R}$. This is achieved in an efficient way via the ∞ symbol as follows

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

5 Functions

Given two sets A and B, we may wish to construct a rule, or a way to map, objects from A onto objects from B. For instance, if $A = \{ a, b, c \}$ and $B = \{ x, y, z \}$ then we may wish to "send" a to x, b to y and c to z. This rule defines what is called a *function,* also referred to as a *map*. We can think of various other functions from A to B, each one is distinct if it has a different way to map the objects around. For example, consider the function which "sends" a to z, b to y and c to x. It is yet another possible way to *map* the objects of A onto those of B.

If we write a table of A and B laid out together in perpendicular directions

$B \downarrow$;$A \rightarrow$	a	b	c
x			
y			
z			

then we may fill in the interior of the table with objects of the product, $A \times B$:

$A \times B$	a	b	c
x	(a, x)	(b, x)	(c, x)
y	(a, y)	(b, y)	(c, y)
z	(a, z)	(b, z)	(c, z)

Then we can think of the first function we described, i.e. that sending a to x, b to y and c to z as a way to pair elements of A and B, that is, as a subset of $A \times B$, namely, the subset $\{ (a, x), (b, y), (c, z) \}$. Looking at the table above, we can identify the function by coloring in red these pairs of the table

$A \times B$	a	b	c
x	(a, x)	(b, x)	(c, x)
y	(a, y)	(b, y)	(c, y)
z	(a, z)	(b, z)	(c, z)

Similarly, the second function, that mapping a to z, b to y and c to x can be associated with the list of pairs $\{ (a, z), (b, y), (c, z) \}$ and on the table of $A \times B$ would this subset of pairs, associated to those pairs colored red looks like:

$A \times B$	a	b	c
x	(a, x)	(b, x)	(c, x)
y	(a, y)	(b, y)	(c, y)
z	(a, z)	(b, z)	(c, z)

However, not all sets of pairs constitute function. The point is that by considering the concept of functions, we are interesting in giving a rule, or a guide to go from A to B. That means in particular that this rule should be unambiguous, so that we don't get stuck trying to decide. So the following list of ordered pairs $\{ (a, x), (b, y), (c, z), (a, z) \}$, represented in the table as

$A \times B$	a	b	c
x	(a, x)	(b, x)	(c, x)
y	(a, y)	(b, y)	(c, y)
z	(a, z)	(b, z)	(c, z)

does *not* constitute an appropriate function, because it tells us simultaneously to send a to x as well as to send it to z. So we don't know where to send a. Hence a function should send objects of a to only one place, which means that the set of pairs encoding the function shouldn't have two pairs with the same first component and different second component (e.g. (a,x) and (a,z)).

The converse, however, is perfectly fine. That is, the function $\{ (a,x), (b,x), (c,x) \}$, which sends all of the elements of A to the same spot in B, is a perfectly fine function.

Finally, we want to make sure that we *know* at all where to map any given element, so if there is some element of the origin set that doesn't exist as the first fact of some pair in the set of pairs, we won't know where to map it. We agree to exclude such scenarios from "appropriate functions".

Given these considerations, we make the

5.1 Definition. Given two sets A and B, a *function* f from A to B, written as $f : A \to B$, is a set of unambiguous rules to associate objects of A with objects of B, i.e. it is a *subset* of pairs, i.e. of $A \times B$, such that no two pairs have the same first component and different second component, and such that all elements of the origin set are covered as one enumerates all first components of all pairs. The set A is called the *domain* of f, the set B is called the *co-domain* of f. Sometimes one refers to *the graph of* f as the that subset of $A \times B$ which specifies it.

Be ware the discrepancy between the intuitive meaning of the word graph (we think of a geometric object) and the technical meaning given above (an abstract subset of pairs). This will come up again and again in math, the distinction between intuitive meanings of words from our daily lives and their actual technical definition.

Let us introduce some graphical notation which will be used throughout the course for functions:

1. Given a function $f : A \to B$, suppose that the element $a \in A$ gets mapped to $x \in B$.
   (a) Arrow notation: We write
   $$a \quad \mapsto \quad x$$
   or
   $$A \ni A \quad \mapsto \quad x \in B$$
   if we want to be slightly more explicit.
   (b) Braces notation: We write
   $$f(a) \quad = \quad x$$
   (c) Subscript notation: We write
   $$f_a \quad = \quad x$$

This is mainly used when $A \in \mathbb{N}$, or when A is of the form $A = T \times X$ for two sets T and X, and instead of writing $f((t, x))$ for some $t \in T$ and $x \in X$ one writes $f_t(x)$.

5.2 Example. If the domain of a function, A, is empty, i.e. $A = \varnothing$, then there are not many choices (since there *is* nothing to map) and so it suffices to write $f : \varnothing \to B$ (for any B), and there is just one unique function with this domain. Similarly, if B contain only one element, then there is again nothing to desscribe, because we have no choice. E.g. $A = \mathbb{N}$, $B = \{ ☹ \}$, then we know what $f : A \to B$ does. It merely converts any number into a ☹. This can graphically be written as

$$f(n) = ☹ \text{ for any } n \in \mathbb{N}$$

5.3 Example. If both the domain and co-domain are the same set, $f : A \to A$, then there is a special function which sends each element to itself. This is known as *the identity function,* and is denoted as $\mathbb{1} : A \to A$ for any set A. We have

$$\begin{aligned} \mathbb{1} : A &\to A \\ a &\mapsto a \end{aligned}$$

When the domain or codomain are rather large (think infinite), e.g. one of our special sets of numbers, it sometimes becomes easier to give a *formula* for what f does rather than specify one by one how it acts on each different element, or enumerate a list of pairs (indeed, that would be literally impossible for infinite sets). Consider the function

$$f : \mathbb{N} \to \mathbb{N}$$

which adds 1 to any given number. So $1 \mapsto 2$, $2 \mapsto 3$, $3 \mapsto 4$ and so on. An easy way to encode that is to use *variables*, i.e. objects which are placeholders for elements of a certain set[1]. A variable is thus any object we could pick from a certain set. For example, a variable n in the natural numbers is any choice $n \in \mathbb{N}$. We could have $n = 5$, $n = 200$ or $n = 6000000$ (but not $n = -1$, since $\mathbb{N}$ contains only strictly positive integers). The point is, it is convenient *not* to specify which element it is and work with a *generic* unspecified element. Once we have a variable, we can easily write down the action of f as a succession of algebraic operations, i.e. a formula *in* that variable:

$$f(n) = n + 1 \text{ for any } n \in \mathbb{N}$$

This specifies in a formula the same verbal description we gave earlier. The variable is also called *the argument of the function.*

5.4 Remark. The most common and efficient way to describe a function is to write two lines of text:

$$\begin{aligned} f : A &\to B \\ a &\mapsto \text{some formula of } a \end{aligned}$$

[1] We already encountered variables when we discussed set-builder notation

where the first line tells us the label of the function (in this case f), the domain, i.e. the origin set A, the codomain, i.e. the destination set B, and the second line tells us how to map each object of A into B. In this case the second line is in the form of a formula, but one could just as well list all possible mappings of elements in A.

We can quickly run into problems with math, just as we would with natural language. It doesn't make sense to write "dry rain" even though we can easily juxtapose the two words together. In the same way, if we try to write down

$$\begin{aligned} f:\mathbb{N} &\to \mathbb{N} \\ f(n) &= n-1 \text{ for any } n\in\mathbb{N} \end{aligned}$$

we quickly realize this makes no sense! The reason is that for certain $n \in \mathbb{N}$, namely, for 1, if we apply the formula, we actually land *outside* of $\mathbb{N}$, because $0 \notin \mathbb{N}$! That means that the formula-way of describing functions can be dangerous, that is, it can quickly lead us to write down nonsense. This is a manifestation of the fact that just because we have a language with rules doesn't mean that every combination of any phrase will make sense. We still must be careful, especially as we build shortcuts.

Here is another example:

$$\begin{aligned} f:\mathbb{N} &\to \mathbb{R} \\ n &\mapsto \sqrt{n} \end{aligned}$$

Note that the same formula would *not* make sense with $\mathbb{R}$ replaced by $\mathbb{N}$ or even $\mathbb{Q}$, as we just learnt (e.g. $\sqrt{3} \notin \mathbb{Q}$)!

5.5 Definition. When a function $f : A \to B$ has its domain $A = \mathbb{N}$, i.e. the natural numbers, one often calls that function a *sequence* and one writes its argument in subscript notation, i.e.

$$f_n = \sqrt{n}$$

We can write many complicated formulas. For example, we can write what is known as a *piecewise formula:*

$$\begin{aligned} a:\mathbb{R} &\to \mathbb{R} \\ x &\mapsto \begin{cases} x & x\geq 0 \\ -x & x<0 \end{cases} \end{aligned}$$

this means that before we apply the formula we must verify some conditions.

Sometimes it is helpful (though usually not unambiguous) to also *sketch* a function. A sketch is the graphical arrangement of all possible values it could take given all possible inputs. We have already seen how to do this in a rather rudimentary way using the colorings of the graph of a function within the table of $A \times B$ above. A sketch of a function is thus a way to geometrically draw the graph of the function.

Figure 1: A plot of the graph of the function $\{1,2\} \to \{1,2\}$ given by $\begin{cases} 1 & \mapsto 2 \\ 2 & \mapsto 1 \end{cases}$.

5.6 Example. $f : \{1,2\} \to \{1,2\}$ given by $1 \mapsto 2$ and $2 \mapsto 1$. Then $A \times B = \{(1,1),(1,2),(2,1),(2,2)\}$ and the graph of f may be described as $\{(1,2),(2,1)\}$. This can also be drawn graphically as in Figure 1.

5.1 Functions from $\mathbb{R} \to \mathbb{R}$

First some general notions, which rely on the fact that for any two numbers $a, b \in \mathbb{R}$, we may compare them, i.e. we necessarily have exactly one of the following: $a < b$, or $b < a$ or $a = b$ (this is not the case for any two objects from any other set, it relies on many properties of $\mathbb{R}$ which we assume but leave implicit).

5.7 Definition. A function $f : \mathbb{R} \to \mathbb{R}$ is called *monotone increasing* iff whenever $a, b \in \mathbb{R}$ and $a \leq b$ then $f(a) \leq f(b)$. It is monotonically decreasing iff $f(a) \geq f(b)$. One can add the qualifier *strictly* to change $\leq$ into $<$ and $\geq$ into $>$.

5.8 Example. Take $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ for any $x \in \mathbb{R}$. The graph of f is the subset of $\mathbb{R}\times\mathbb{R}$ given by pairs (x, x^2) for any $x \in \mathbb{R}$, i.e.

$$\left\{ \left(x, x^2\right) \in \mathbb{R}^2 \;\middle|\; x \in \mathbb{R} \right\}$$

Since $\mathbb{R}$ may be pictured as an infinite line, $\mathbb{R}\times\mathbb{R} = \mathbb{R}^2$, the set of all pairs, should be pictured as an infinite two-dimensional plane, in which case the graph of a function $f : \mathbb{R} \to \mathbb{R}$ is a *curve* in that plane. The fact it is a curve, and nothing else, is related to the fact that no two pairs have the same first component. The *shape* of that curve is what we care about when drawing the graph of the function, which is the sketch of that function. In the particular case of (x, x^2) for any $x \in \mathbb{R}$, the shape of the curve is that of the familiar parabola as in Figure 2.

5.9 Definition. The function

$$\begin{aligned} a : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto \begin{cases} x & x \geq 0 \\ -x & x < 0 \end{cases} \end{aligned}$$

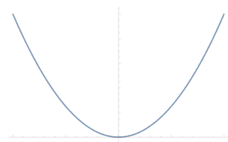


Figure 2: The parabola $x \mapsto x^2$.

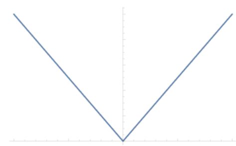


Figure 3: The sketch of the graph of the absolute value.

which we already described above to introduce the piecewise notation is called *the absolute value* function. Its graph can be sketched as in Figure 3. One way to figure out how to draw these functions is to draw a few pairs of points on the plane and then extrapolate. For example, if we want to start plotting the absolute value function, we start with a few points from the formula:

$$\begin{aligned} 0 &\mapsto 0 \\ 1 &\mapsto 1 \\ -1 &\mapsto 1 \\ 2 &\mapsto 2 \\ -2 &\mapsto 2 \end{aligned}$$

which correspond to the pairs $(0,0),(1,1),(-1,1),(2,2),(-2,2)$, and then we mark points on a grid at these coordinates. In order to extrapolate it is useful to know of a few building blocks of basic functions.

5.10 Remark. The absolute value function obeys the following properties: for any two numbers $x, y \in \mathbb{R}$:

$$|xy| = |x||y| \tag{1}$$

5.11 Definition. A function $f : \mathbb{R} \to \mathbb{R}$ is called *bounded* iff there is some $M \in \mathbb{R}$ with $M \geq 0$ such that

$$|f(x)| \leq M$$

for *all* $x \in \mathbb{R}$.

5.12 Example. The constant function $f : \mathbb{R} \to \mathbb{R}$; $x \mapsto c$ (for some constant $c \in \mathbb{R}$) is bounded. Indeed, one can pick $M := |c|$.

5.13 Example. The parabolic function $f : \mathbb{R} \to \mathbb{R}$; $x \mapsto x^2$ is *not* bounded.

5.14 Example. The parabolic function restricted to a finite interval is bounded. Foe example, $f|_{[-5,5]} : [-5,5] \to \mathbb{R}$; $x \mapsto x^2$. Indeed, one can pick $M := 25$.

Figure 4: The constant function $\mathbb{R} \ni x \mapsto c \in \mathbb{R}$ for all $x \in \mathbb{R}$, for some constant $c \in \mathbb{R}$.

5.1.1 Basic functions and their shapes

5.15 Definition. Let $c \in \mathbb{R}$ be any given number. The *constant function* $f : \mathbb{R} \to \mathbb{R}$ associated to c sends all elements of its domain to c:

$$f(x) = c \qquad (x \in \mathbb{R})$$

Graphically this function looks like a flat horizontal line at the height c as in Figure 4.

5.16 Definition. Let $a, b \in \mathbb{R}$ be given (note this is short-cut notation for $a \in R$ and $b \in \mathbb{R}$). Then *the linear function* $f : \mathbb{R} \to \mathbb{R}$ associated with a and b is given by

$$f(x) = ax + b \qquad (x \in \mathbb{R})$$

Graphically this function looks like a straight line at an angle. The parameter b sets the line's height when it meets the vertical axis, and the number $-\frac{b}{a}$ is where it meets the horizontal axis:

5.17 Definition. Let $a, b, c \in \mathbb{R}$ be given. Then *the parabolic function* $f : \mathbb{R} \to \mathbb{R}$ associated with a, b, c is given by

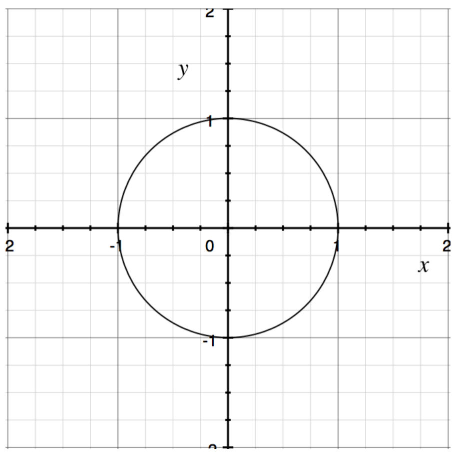
$$f(x) = ax^2 + bx + c \qquad (x \in \mathbb{R})$$

One can of course go on with these to any highest power of x, e.g. $f(x) = x^{100}$ which looks like this:

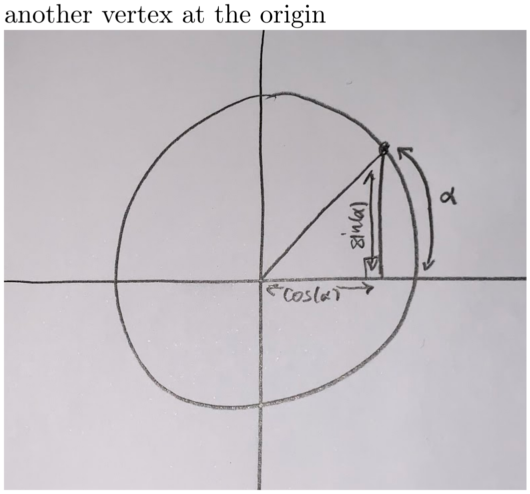
5.1.2 Trigonometric functions

We next want to introduce the trigonometric functions. These are special functions because they *do not* have special algebraic formulas which define their action. There are two possibilities: either we define them through a limiting process (to be done later on) or we can define them *pictorially* through a geometric picture. For now let us do the latter and just draw some pictures.

Let us draw a circle of radius 1 on the plane $\mathbb{R}^2$



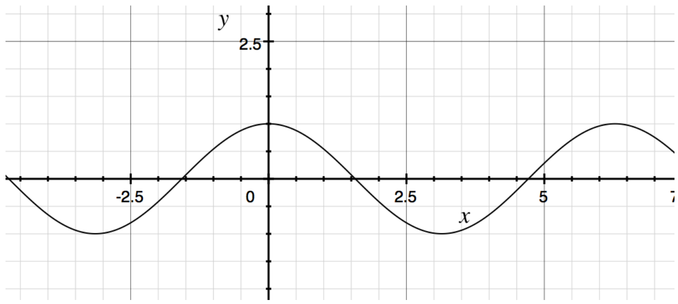
We know the entire circumference of the whole circle is 2π where π is some special irrational number equal to about 3.14 which we cannot write out explicitly. Let us traverse, along the circle, starting from the point $(1, 0)$ on the plane, an arc of arc-length α, for some $0 \leq \alpha < 2\pi$, and draw a right triangle whose base is along the horizontal axis, has a point on the circle after arc-length α and another vertex at the origin



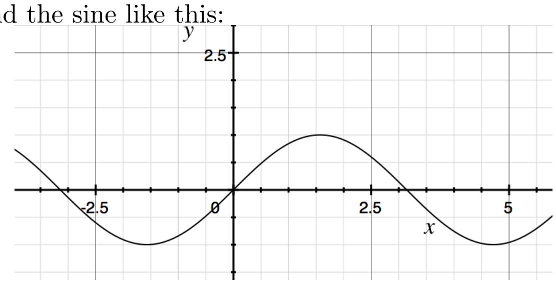
The sinus function is defined as the height of this triangle (as a function of α), and the cosine function is defined as the base length of this triangle (as a function of α). Since we are on a circle, it makes sense to agree that after $\alpha > 2\pi$ the sine and cosine functions assume the same values as if we were calculating them with $\alpha - 2\pi$, and similarly for $\alpha < 0$, so that we get a definition for the whole of $\mathbb{R}$ of a *periodic* function. Things to note:

1. $\cos(0) = 1$, $\sin(0) = 0$
2. $\cos\left(\frac{\pi}{2}\right) = 0$, $\sin\left(\frac{\pi}{2}\right) = 1$.
3. cos is decreasing on $(0, \pi)$, sin is increasing on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

The cosine looks like this:



and the sine like this:



5.2 Special sets associated with a function

Given a function $f : A \to B$, we already encountered the following sets associated with it:

1. The domain of f, which is just A.
2. The co-domain of f, which is just B.
3. The graph of f, encoding the same information as f itself, which is the subset of $A \times B$ given by

$$\text{graph}(f) \quad = \quad \{ (a, f(a)) \in A \times B \mid a \in A \}$$

We define a few more sets associated to a given function f:

5.18 Definition. The *image of a function* $f : A \to B$ is the subset of B given by the following

$$\begin{aligned} \text{im}(f) \quad &:= \quad \{ b \in B \mid \text{There is some } a \in A \text{ such that } f(a) = b \} \\ &= \quad \{ f(a) \in B \mid a \in A \} \end{aligned}$$

Note: we will not use the word "range" in this course, as it is ambiguous and sometimes conflated with either co-domain or image.

5.19 Definition. Let $f : A \to B$ be a function between two sets A and B and let $S \subseteq A$ be a given subset. Then the image of S under f is the following subset of B:

$$\begin{aligned} f(S) &:= \{ b \in B \mid \text{There is some } a \in S \text{ such that } f(a) = b \} \\ &= \{ f(a) \in B \mid a \in S \} \end{aligned}$$

Note that in this graphical notation we use the braces notation on a whole set rather than an object, and the result is then a set, rather than an object! This notation can be confusing. Using this notion we can identify

$$f(A) = \operatorname{im}(f)$$

5.20 Definition. Let $f : A \to B$ be a function between two sets A and B and let $S \subseteq B$ be a given subset. The *pre-image* of S under f is the following subset of A:

$$f^{-1}(S) = \{ a \in A \mid f(a) \in S \}$$

Note the introduction of a new notation: for the pre-image of a function f, we use the graphical symbol f^{-1}. Again this is a funny notation in the sense that we plug in a set into f^{-1} and get back a set. Despite the notation, f^{-1} is *not* a function. Of course for a function $f : A \to B$ we have $f^{-1}(B) = A$ by definition.

5.21 Definition. A function $f : A \to B$ is called *surjective* if $\operatorname{im}(f) = B$. That means there are no elements of B left "uncovered" by f.

5.22 Definition. A function $f : A \to B$ is called *injective* if

$$\left| f^{-1}(\{ b \}) \right| \leq 1 \qquad (b \in B)$$

which means that every point of B gets covered at most once (if not never) by f. In other words, no two elements of A get sent to the same element of B, that is, every destination point has a *unique* origin point, if it is in the image of f.

5.23 Definition. A function $f : A \to B$ is called bijective if it is surjective and injective. Bijective functions should be thought of as *reversible*, because they don't lose information.

5.24 Example. The constant function $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto 5$ for all $x \in \mathbb{R}$ is not surjective, since $\operatorname{im}(f) = \{ 5 \} \neq \mathbb{R}$. It is not injective because while $f^{-1}(\{ x \}) = \varnothing$ for all $x \neq 5$ and $|\varnothing| = 0$, $f^{-1}(\{ 5 \}) = \mathbb{R}$ and $|\mathbb{R}| = \infty > 1$.

5.25 Example. The linear function $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto 5x$ for all $x \in \mathbb{R}$ is bijective, because

$$\begin{aligned} \operatorname{im}(f) &= \{ 5x \mid x \in \mathbb{R} \} \\ &= \mathbb{R} \end{aligned}$$

and $f^{-1}(\{ x \}) = \{ y \in \mathbb{R} \mid 5y = x \} = \left\{ \frac{1}{5}x \right\}$ which is of size one.

What about the absolute value function?

5.3 Construction of new functions

5.26 Definition. Given two functions $f : A \to B$ and $g : B \to C$, we define their *composition*, denoted as $g \circ f$, as a new function $A \to C$ given by the formula

$$(g \circ f)(a) \quad := \quad g(f(a)) \qquad \text{for any } a \in A$$

which first applies f, and then g (considered as rules), all together passing through B but ultimately producing a route (i.e. a function) from $A \to C$. We also can compose a function itself, if its co-domain is equal to its domain: if $f : A \to A$ then

$$f^n \quad := \quad \underbrace{f \circ f \circ \cdots \circ f}_{n \text{ times}}$$

for any $n \in \mathbb{N}$.

5.27 Example. If $f : \mathbb{R} \to \mathbb{R}$ is given by $x \mapsto \sin(x)$ then $f \circ f : \mathbb{R} \to \mathbb{R}$ is the function given with the formula $x \mapsto \sin(\sin(x))$ for any $x \in \mathbb{R}$ (we don't ask what that means geometrically).

5.28 Definition. A function $f : A \to B$ is called *left-invertible* iff there is some other function $g : B \to A$ such that $g \circ f = \mathbb{1}$ where $\mathbb{1} : A \to A$ is the identity function discussed above. Conversely, f is called *right-invertible* iff there is some other function $h : B \to A$ such that $f \circ h = \mathbb{1}$ where $\mathbb{1} : B \to B$ is the identity function. If f is both left and right invertible we call it invertible, and then the left and right inverse are equal and unique $g = h$, in which case we denote that inverse by $f^{-1} = g = h$ (not to be confused with the pre-image notation, and also not to be confused as an algebraic operation–we are not dividing anything by anything else, this is merely graphical notation), so that by definition

$$\begin{aligned} f \circ f^{-1} &= \mathbb{1}_B \\ f^{-1} \circ f &= \mathbb{1}_A \end{aligned}$$

These last two equations are interesting, because they tell us that functions themselves (rather than objects, numbers, or sets) are equal. But since we have a precise way to think of functions as sets themselves, this is perfectly fine. Also we use the short-hand notation of $\mathbb{1}_A$ to denote the (unique) identity function $A \to A$ for any set A.

What kind of relationship is there between left or right invertibility and injectivity surjectivity?

5.29 Definition. Given any function $f : \mathbb{R} \to \mathbb{R}$, we can quickly define a new function by an algebraic formula on f itself. For example, the function $f + 3$ has the formula

$$(f + 3)(x) \quad = \quad f(x) + 3 \qquad (x \in \mathbb{R})$$

Sometimes these shortcuts don't always make sense and one has to be careful, for example, with $\frac{1}{f}$. Other times the notation itself becomes ambiguous, for example, f^2 could either mean $f(f(x))$ for any $x \in \mathbb{R}$ or it could mean $(f(x))^2$ for any $x \in \mathbb{R}$. So in such cases one has to write out in words what one means. Another possible confusion is with f^{-1}. Usually it means either the pre-image or the (unique) inverse of a function, as defined above, if it exists. It usually does not mean the function

$$x \quad \mapsto \quad \frac{1}{f(x)} \qquad (x \in \mathbb{R})$$

for which one usually uses the notation $\frac{1}{f}$ instead.

We can also make formulas with two or more functions, whenever that makes sense. So if $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$, then by $f + g$ (or $f - g$, fg, $\frac{f}{g}$, etc) we mean a new function $\mathbb{R} \to \mathbb{R}$ whose formula is

$$x \quad \mapsto \quad f(x) + g(x) \qquad (x \in \mathbb{R})$$

5.30 Definition. Given any function $f : A \to B$ and a subset $X \subseteq A$, we define the restriction of f to X, denoted by $f|_X : X \to B$, as

$$f|_X(a) \quad := \quad f(a) \qquad (a \in X)$$

So $f|_X$ and f have the same formula, but the former is restricted to act on a smaller subset. This is sometimes a useful notion when considering the properties of functions, some of which may only hold on a subset but not on the whole domain.

5.31 Example. Pick any number $a \in \mathbb{R}$ which is strictly positive, $a > 1$. Consider the function $\exp_a : \mathbb{R} \to \mathbb{R}$ which is given by

$$\exp_a(x) \quad := \quad a^x \qquad (x \in \mathbb{R})$$

If $x = 0$ the result is 1 (by convention). When $x = n$ for some $n \in \mathbb{N}$, we know how to perform this operation. We merely raise a to the power n, i.e. we compute $\underbrace{a \times a \times \cdots \times a}_{n \text{ times}}$. When $x = -n$ for some $n \in \mathbb{N}$, we know that this means to computer $\underbrace{a \times a \times \cdots \times a}_{n \text{ times}}$, and then we must take the reciprocal of that number, that is, $\frac{1}{a^n}$. If $x = \frac{1}{n}$ for some $n \in \mathbb{N}$, then this should be the n-th root of a, that is $\sqrt[n]{a}$, and if $x = -\frac{1}{n}$ for some $n \in \mathbb{N}$, we get $\frac{1}{\sqrt[n]{a}}$. Hence all together if $x = \frac{p}{q}$ for some $p, q \in \mathbb{Z}$ we get

$$\exp_a(x) \quad = \quad \sqrt[q]{a^p}$$

which still doesn't tell us a lot, because as we already saw, we cannot write out explicitly what $\sqrt{2}$ is, for instance, but it at least gives us some constraint on what the answer should be (i.e. $\sqrt{2}$ should be that number such that $\sqrt{2}\sqrt{2} = 2$).

It turns out that even if $x \in \mathbb{R}\backslash\mathbb{Q}$ one could proceed, via a limit procedure that makes the sketch of $\exp_a$ look smooth when plotted on $\mathbb{R}$ (using the basic fact that any element $x \in \mathbb{R}\backslash\mathbb{Q}$ has an element $y \in \mathbb{Q}$ arbitrarily close to it, so intuitively, we define $\exp_a(x)$ as $\exp_a(y)$ (which we know how to compute) where $y \in \mathbb{Q}$ is arbitrarily close to $x \in \mathbb{R}\backslash\mathbb{Q}$).

As define, $\exp_a : \mathbb{R} \to \mathbb{R}$ is not surjective, and hence not bijective. Indeed, it is *always* larger than zero. That is, we have

$$\operatorname{im}(\exp_a) = (0, \infty)$$

So we change the definition $\exp_a : \mathbb{R} \to \mathbb{R}$ by modifying the co-domain to be $(0, \infty)$:

$$\exp_a : \mathbb{R} \to (0, \infty)$$

to be defined by the same formula as before, and get a surjective function. Actually $\exp_a$ is also injective. Indeed, we can verify this by verifying that if $\exp_a(x) = \exp_a(y)$ for some $x, y \in \mathbb{R}$, then $a^x = a^y$ (in HW1 you learn this is one possible criterion for injectivity). Divide both sides of the equation by a^y to get $a^x \frac{1}{a^y} = 1$. The basic rules of exponentiation imply now that $a^{x-y} = 1$. However, we *know* that only when exponentiating some number which is strictly larger than 1 to power zero we get back 1, so that $x - y = 0$ necessarily. So that means $x = y$ and hence $\exp_a$ is indeed injective. Since $\exp_a : \mathbb{R} \to (0, \infty)$ is injective and surjective, i.e. bijective, you learn in HW1 that means it has a unique inverse $\exp_a^{-1} : (0, \infty) \to \mathbb{R}$. This inverse is called the logarithm with base a, and is denoted by $\log_a : (0, \infty) \to \mathbb{R}$.

5.32 Exercise. Both cos and sin functions when defined from $\mathbb{R} \to \mathbb{R}$ are not injective nor surjective. However, one may modify both domain and co-domain to make them bijective. How?

6 Limits

At the heart of calculus is the notion of a *limit.* The limit is a way to consider a hypothetical process that cannot actually be carried out but whose result still may have meaning. We have already encountered such hypothetical processes when we first considered the set

$$\mathbb{N} \equiv \{ 1, 2, 3, \dots \}$$

where the dots mean the hypothetical process of continuing the list with no end. Since this is not technically possible, this is merely a hypothetical notion. And yet it is useful for us to collect together in one set *all* possible natural numbers, which really just means that whatever large number one can think of, it is part of $\mathbb{N}$.

Yet another example that we already encountered was the hypothetical result of a process of enlisting fractions with increasing denominators, i.e., the sequence

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \dots$$

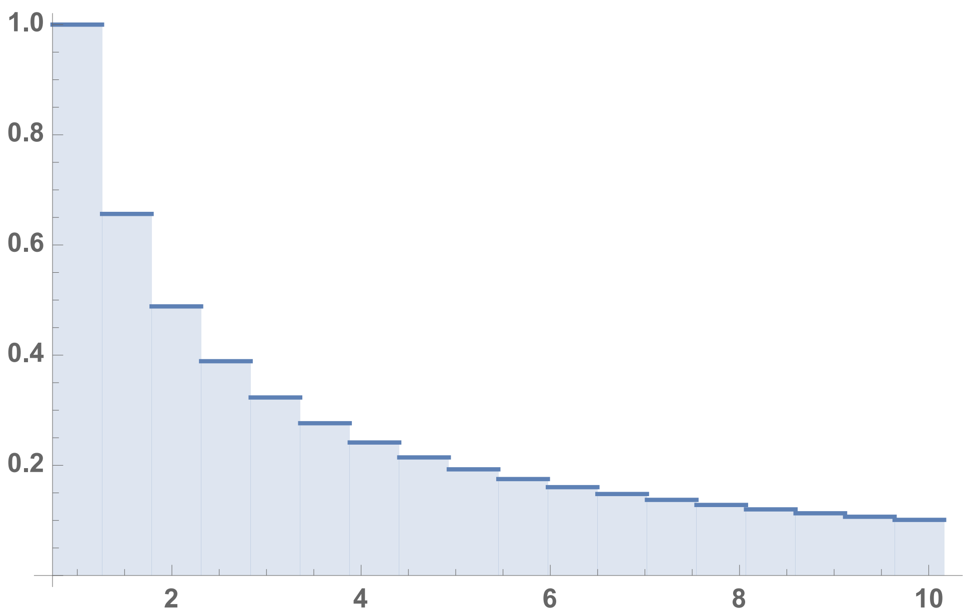


Figure 5: A plot of the graph of $\mathbb{N} \ni n \mapsto \frac{1}{n} \in \mathbb{R}$.

which has no end. Since it has no end, the final result of this process is merely hypothetical. And yet intuitively it is clear that the end result will be zero, which really just means, whatever small number you can think of, one can find a step in this process which will be smaller than that given number.

6.1 The notion of a distance on $\mathbb{R}$

We say the distance between a pair of numbers is the magnitude of their difference, i.e. we only care about how far apart they are, but not in which direction. Hence we define the distance function on *pairs* of numbers

$$\begin{aligned} d : \mathbb{R}^2 &\to [0, \infty) \\ (x, y) &\mapsto |x - y| \qquad \text{for any } (x, y) \in \mathbb{R}^2 \end{aligned}$$

The distance function has a number of important properties:

1. It is symmetric: $d(x, y) = d(y, x)$ for any $x, y \in \mathbb{R}$, because $|a| = |-a|$ for any $a \in \mathbb{R}$.

2. Distance zero implies identity: $d(x, y) = 0$ for some $x, y \in \mathbb{R}$ implies $x = y$. Indeed, the absolute value function, as we defined it, only takes the value zero at zero. So $|a| = 0$ implies $a = 0$.

3. It obeys the so-called *triangle inequality*: for any three numbers $x, y, z \in \mathbb{R}$ we have $d(x, y) \leq d(x, z) + d(z, y)$. The way to convince yourself that

this is really true is to divide the analysis into cases. The easiest case is that all three numbers are different and obey $x < z < y$. Then we have

$$\begin{aligned}
d(x,y) &\equiv |x-y| \\
&\quad \text{(From the definition of the absolute value, because } y > x\text{)} \\
&= y - x \\
&= y - z + z - x \\
&\quad \text{(From the definition of the absolute value, because } y > z \text{ and } z > x\text{)} \\
&= |y-z| + |z-x|
\end{aligned}$$

the other cases proceed similarly.

An important property of this distance is that it allows us to pinpoint exactly

6.1 Claim. Iff $d(x,y) < \alpha$ for some $x, y \in \mathbb{R}$ and some strictly positive number $\alpha > 0$ then we have

$$-\alpha < \quad x - y \quad < \alpha\,.$$

Indeed, from the definition of the absolute value, we know that $d(x,y) \equiv |x-y|$ is equal to $x-y$ if $x > y$ and $y-x$ if $y > x$. Hence, either $x > y$, $|x-y| = x-y$, and then $x - y < \alpha$, or $x < y$, $|x-y| = -(x-y)$, and then because $\alpha > 0$ and $x - y < 0$ we get $x - y < 0 < \alpha$ or just $x - y < \alpha$. This shows you that the first inequality holds. The second one proceeds similarly.

6.2 Claim. The distance function is translation invariant. That is: $d(x,y) = d(x-z, y-z)$ for any three numbers $x, y, z \in \mathbb{R}$. To see this, we write out the definition

$$\begin{aligned}
d(x,y) &= |x-y| \\
&= |x-y+z-z| \\
&= |x-z-(y-z)| \\
&= d(x-z, y-z)
\end{aligned}$$

One can also show that

6.3 Claim. (Reverse triangle inequality) We have for any $x, y \in \mathbb{R}$,

$$d(|x|, |y|) \leq d(x,y)\ .$$

To see this, note that

$$d(|x|, |y|) \quad \equiv \quad ||x| - |y||$$

and

$$\begin{aligned}
|x| &= |x-y+y| \\
&\quad \text{(Regular triangle inequality)} \\
&\leq |x-y| + |y|
\end{aligned}$$

so we have

$$|x| - |y| \quad \leq \quad |x - y|$$

By symmetry (running the same argument after having exchange x with y) we also have

$$|y| - |x| \quad \leq \quad |x - y|$$

which is equivalent to (by multiplying the inequality by minus one):

$$|x| - |y| \quad \geq \quad -|x - y|$$

so we conclude by Claim 6.1 that

$$||x| - |y|| \leq |x - y|$$

which is what we wanted to show.

6.2 Limits of sequences–functions from $\mathbb{N} \to \mathbb{R}$

More generally, let $a : \mathbb{N} \to \mathbb{R}$ be some function. Such functions whose domain is $\mathbb{N}$ have a special name: they are called *sequences*. The example above corresponds to the sequence given by the formula

$$a_n \quad = \quad \frac{1}{n} \qquad (n \in \mathbb{N})$$

but in principle this could be any formula. Here is another example

$$a_n \quad = \quad (-1)^n \qquad (n \in \mathbb{N})$$

if we list the values of this sequence we see the first few are equal to

$$-1, 1, -1, 1, -1, 1, -1, 1, \ldots$$

and intuitively it is clear that this list does not "tend" to anything as we go forward towards infinity, because it keeps jumping up and down between -1 and 1. Here is yet another example:

$$a_n \quad = \quad n^2 \qquad (n \in \mathbb{N})$$

and again we list some of the first few elements

$$1, 4, 9, 16, \ldots$$

the items on this list grow very quickly. So if "imagine" what would happen if we continued to take more and more steps of this process, one possible way to phrase the result would be to say that it tends to infinity. By that one means that whatever (large) number one could come up with, there is a sufficient number of steps of this process that may be taken so as to surpass this large number.

So we have encountered so far three possible behaviors of a sequence $a : \mathbb{N} \to \mathbb{R}$:

1. The sequence "converges" to some number $c \in \mathbb{R}$ as we plug in larger and larger arguments, as was the case in the first example.
2. The sequence keeps jumping back and forth no matter how far we go.
3. The sequence keeps growing with no bound–it *diverges*.

We formalize these considerations in the following

6.4 Definition. A sequence $a : \mathbb{N} \to \mathbb{R}$ is said to have a limit $L \in \mathbb{R}$ iff for any strictly positive number that one could pick, $\delta \in \mathbb{R}$ with $\delta > 0$, but as small as one wants, there is some number $N \in \mathbb{N}$ (this number may depend on δ) such that for all $n \in \mathbb{N}$ obeying $n \geq N$, the following condition holds:

$$d(a(n), L) < \delta$$

That is, the distance between $a(n)$ and the number L becomes as small as one wants–one merely has to go far enough into the sequence, and how far depends on how small the distance we ask for. Another way to say this is that $a(n)$ *converges* to L as $n \to \infty$. The point about this concept is that the distance between $a(n)$ and L becomes smaller and smaller and smaller. If the distance is "small", but remains fixed as we enlarge n, the notion does *not* apply.

6.5 Remark. It is not possible that $a(n)$ converges to L as $n \to \infty$ and also $a(n)$ converges to L' as $n \to \infty$ if $L \neq L'$.

Proof. We have for all $n \geq \max(N, N')$, N being the threshold of distance $\delta > 0$ for the convergence of a to L and N' that to L',

$$\begin{aligned} d(L, L') &\leq d(L, a(n)) + d(a(n), L') \\ &\leq 2\delta \end{aligned}$$

That means that the distance between L and L' can be made arbitrarily small, that is, they are equal. □

6.6 Definition. A sequence $a : \mathbb{N} \to \mathbb{R}$ is said to go to ∞ (respectively $-\infty$), to *diverge* to $\pm\infty$ iff for any strictly positive number that one could pick, $M \in \mathbb{R}$, there is some number $N \in \mathbb{N}$ (this number may depend on M) such that for all $n \in \mathbb{N}$ obeying $n \geq N$, the following condition holds:

$$a(n) \geq M$$

(respectively $a(n) \leq -M$)

6.7 Definition. A sequence $a : \mathbb{N} \to \mathbb{R}$ is said to have no limit (one says the limit does not exist), if there is no $L \in \mathbb{R}$ to which it converges, and it does not go to either ∞ or $-\infty$.

Different notations for this are:

1. The limit notation: If L is a limit of a, then we write

$$\lim_{n\to\infty} a(n) = L$$

2. or sometimes

$$\lim a = L$$

3. or sometimes

$$a(n) \to L \qquad (n\to\infty)$$

4. and when it is not important what L is, but only that there is some L like that, we write that $\lim a$ *exists*.

5. When a goes to infinity, we write

$$\lim_{n\to\infty} a(n) = \infty$$

and say that the limit *diverges*.

6.8 Example. Going back to our initial example of $\mathbb{N} \ni n \mapsto \frac{1}{n} \in \mathbb{R}$, let us see why this converges to *zero* as $n \to \infty$ according to the definition. We have

$$\begin{aligned} d(a(n), 0) &\equiv |a(n) - 0| \\ &= \left|\frac{1}{n}\right| \\ &\quad (n > 0) \\ &= \frac{1}{n} \end{aligned}$$

Let us pick some number $\delta > 0$. If we want to arrange that $\frac{1}{n} < \delta$, we equivalently need $n > \frac{1}{\delta}$. So if we take N to be the smallest integer larger than $\frac{1}{\delta}$, then $n \geq N$ implies that $n > \frac{1}{\delta}$!

6.9 Example. Take $\mathbb{N} \ni n \mapsto \frac{n-1}{n+1} \in \mathbb{R}$. Does this converge or diverge? Let us list the first few elements

$$0, \frac{1}{3}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{5}{7}, \frac{3}{4}, \frac{7}{9}, \frac{4}{5}, \frac{9}{11}, \dots$$

this actually does converge to 1. The reason being that for n being very large, i.e. much larger than 1, $n-1$ and $n+1$ are not very different, so their quotient is very close to 1. To really see this, calculate

$$\begin{aligned} d\left(\frac{n-1}{n+1}, 1\right) &\equiv \left|\frac{n-1}{n+1} - 1\right| \\ &= \left|\frac{n-1-n-1}{n+1}\right| \\ &= \frac{2}{n+1} \end{aligned}$$

so if we want this distance to be arbitrarily small, we need to pick n such that $\frac{2}{n+1} < \delta$ or $n > \frac{2}{\delta} - 1$.

6.10 Example. Take $\mathbb{N} \ni n \mapsto \sqrt{n+1} - \sqrt{n} \in \mathbb{R}$. Let us write out the first few elements of this sequence (I used a computer to give approximate values of the square roots):

$$0.41, 0.31, 0.26, 0.23, 0.21, 0.19, \dots$$

it seems to be going down, but does it converge to zero? The answer is yes, again because as n is very large, the difference between $n+1$ and n becomes insignificant (essentially because n is much larger than 1!) So we try to calculate

$$\begin{aligned} d\left(\sqrt{n+1} - \sqrt{n}, 0\right) &= \left|\sqrt{n+1} - \sqrt{n}\right| \\ &= \text{(The square root is monotone increasing, so this is positive)} \\ &= \sqrt{n+1} - \sqrt{n} \\ & \quad \left(\text{Use the identity } a - b = \frac{a^2 - b^2}{a+b}\right) \\ &= \frac{n+1-n}{(n+1)^2 + n^2} \\ &= \frac{1}{2n^2 + 2n + 1} \\ & \quad \left(\text{Use } 2n^2 + 2n + 1 \geq 2n^2 \text{ for any } n\right) \\ &\leq \frac{1}{2n^2} \end{aligned}$$

and we get the same story (we can make this arbitrarily small by taking n arbitrarily large).

6.11 Example. Consider the sequence $\mathbb{N} \ni n \mapsto n \in \mathbb{R}$. Can we show it goes to infinity? Trivially, because for any big number that we can choose, $M \in \mathbb{R}$, there is some $N \in \mathbb{N}$ such that all $n \geq N$ will obey $n \geq M$. In particular, take N to be the first integer larger than M.

6.12 Claim. If $a : \mathbb{N} \to \mathbb{R}$ and $b : \mathbb{N} \to \mathbb{R}$ are two sequences which are equal except for a finite number of elements, then their limit behavior is identical. That is, if

$$a(n) = b(n)$$

for all $n \geq N$, for some $N \in \mathbb{N}$, then $\lim a = \lim b$ if this exists (that is, either both limits exist and converge to the same finite number, or both limits do not exist, or both limits diverge to infinity or minus infinity).

Proof. Assume for simplicity that $\lim a$ exists and converges to a finite number (the other cases being similar). Then we want to show $\lim b$ exists and equals $\lim a$. To show that, let us assume that $N_a(\delta) \in \mathbb{N}$ is that threshold of a such that if $n \geq N_a(\delta)$ then

$$d(a(n), \lim a) < \delta\,.$$

Then if we pick $n \geq \max(\{N, N_a(\delta)\})$ we have both $d(a(n), \lim a) < \delta$ and $a(n) = b(n)$, which implies

$$d(b(n), \lim a) \quad < \quad \delta\,.$$

Hence b converges and $\lim b = \lim a$. □

6.13 Claim. (*Algebra of limits*) If a, b are two sequences $\mathbb{N} \to \mathbb{R}$ which both have *finite* limits, then $\lim(a+b) = \lim a + \lim b$, $(\lim a)(\lim b) = \lim(ab)$. Also, if $\lim b \neq 0$, then $\lim\left(\frac{a}{b}\right) = \frac{\lim a}{\lim b}$ with the understanding of $\frac{a}{b}$ being a sequence that might be defined only after a finite number of terms.

Proof. Let us assume that both a and b have finite limits L_1 and L_2. Let us take the thresholds $N_1(\delta), N_2(\delta) \in \mathbb{N}$ for each of these limits. That means that given any $\delta > 0$, if $n \geq N_1(\delta)$ then

$$d(a(n), L_1) \quad \leq \quad \delta$$

and if $n \geq N_2(\delta)$ then

$$d(b(n), L_2) \quad \leq \quad \delta$$

Let us define $N(\delta) := \max(\{N_1(\delta), N_2(\delta)\})$ (i.e. the largest of the two thresholds, so that if $n \geq N(\delta)$ then automatically both $n \geq N_1(\delta)$ and $N_2(\delta)$). Then we have, using Claim 6.2

$$\begin{aligned} d(a(n)+b(n), L_1+L_2) &= d(a(n)-L_1, L_2-b(n)) \\ &\quad \text{(Use triangle inequality with the third point being zero)} \\ &\leq d(a(n)-L_1, 0) + d(0, L_2-b(n)) \\ &\quad \text{(Use translation invariance again invariance, twice)} \\ &= d(a(n), L_1) + d(b(n), L_2) \end{aligned}$$

Hence, if $n \geq N\left(\frac{1}{2}\delta\right)$, then

$$d(a(n)+b(n), L_1+L_2) \quad \leq \quad \frac{1}{2}\delta + \frac{1}{2}\delta = \delta$$

which proves the first statement about the sum. To get the product, we use

Remark 5.10 together with the triangle inequality below to get

$$\begin{aligned}
d\left(a\left(n\right)b\left(n\right),L_1L_2\right) &= \left|a\left(n\right)b\left(n\right)-L_1L_2\right|\\
&= \left|a\left(n\right)b\left(n\right)-L_1b\left(n\right)+L_1b\left(n\right)-L_1L_2\right|\\
&= \left|\left(a\left(n\right)-L_1\right)b\left(n\right)+L_1\left(b\left(n\right)-L_2\right)\right|\\
&\leq \left|a\left(n\right)-L_1\right|\left|b\left(n\right)\right|+\left|L_1\right|\left|b\left(n\right)-L_2\right|\\
&= \left|a\left(n\right)-L_1\right|\left|b\left(n\right)-L_1+L_1\right|+\left|L_1\right|\left|b\left(n\right)-L_2\right|\\
&\leq \left|a\left(n\right)-L_1\right|\left(\left|b\left(n\right)-L_1\right|+\left|L_1\right|\right)+\left|L_1\right|\left|b\left(n\right)-L_2\right|\\
&= \left|L_1\right|\left(\left|a\left(n\right)-L_1\right|+\left|b\left(n\right)-L_2\right|\right)+\left|a\left(n\right)-L_1\right|\left|b\left(n\right)-L_1\right|
\end{aligned}$$

so if we pick $n \geq N\left(\sqrt{\left|L_1\right|^2+\delta}-\left|L_1\right|\right)$, then

$$\begin{aligned}
d\left(a\left(n\right)b\left(n\right),L_1L_2\right) &\leq \left|L_1\right|\left(\sqrt{\left|L_1\right|^2+\delta}-\left|L_1\right|+\sqrt{\left|L_1\right|^2+\delta}-\left|L_1\right|\right)+\left(\sqrt{\left|L_1\right|^2+\delta}-\left|L_1\right|\right)^2\\
&= \left(2\left|L_1\right|+\sqrt{\left|L_1\right|^2+\delta}-\left|L_1\right|\right)\left(\sqrt{\left|L_1\right|^2+\delta}-\left|L_1\right|\right)\\
&= \left(\left|L_1\right|+\sqrt{\left|L_1\right|^2+\delta}\right)\left(\sqrt{\left|L_1\right|^2+\delta}-\left|L_1\right|\right)\\
&= \left|L_1\right|^2+\delta-\left|L_1\right|^2\\
&= \delta
\end{aligned}$$

which proves the second statement about the product.

To prove the last statement, we assume $\lim b \neq 0$. Consider first the case that $b\left(n\right) \neq 0$ for all $n \in \mathbb{N}$. Then we know what the sequence $\frac{1}{b}$ means, and by the multiplication law we just proved, we know that $\lim\left(\frac{a}{b}\right) = \left(\lim a\right)\left(\lim \frac{1}{b}\right)$, so we only have to understand that

$$\lim\left(\frac{1}{b}\right) = \frac{1}{\lim b}$$

To that end,

$$\begin{aligned}
d\left(\frac{1}{b\left(n\right)},\frac{1}{L_2}\right) &= \left|\frac{1}{b\left(n\right)}-\frac{1}{L_2}\right|\\
&= \left|\frac{L_2-b\left(n\right)}{b\left(n\right)L_2}\right|\\
&= \frac{d\left(L_2,b\left(n\right)\right)}{\left|b\left(n\right)\right|\left|L_2\right|}
\end{aligned}$$

Now, we know that $b(n) \to L_2$ as $n \to \infty$ and $L_2 \neq 0$ by hypothesis. That means that

$$d(L_2, 0) \quad \leq \quad d(L_2, b(n)) + d(b(n), 0)$$

or

$$d(b(n), 0) \quad \geq \quad d(L_2, 0) - d(L_2, b(n))$$

Since $L_2 \neq 0$, $d(L_2, 0) \equiv |L_2|$ is a nice strictly positive number. Since $b(n) \to L_2$ as $n \to \infty$, $d(b(n), L_2)$ can be made arbitrarily small by taking n above a certain threshold. For example, assume that $n \geq N_2\left(\frac{1}{2}d(L_2, 0)\right)$. Then $d(L_2, b(n)) \leq \frac{1}{2}d(L_2, 0)$ or $-d(L_2, b(n)) \geq -\frac{1}{2}d(L_2, 0)$ so that we can conclude all together $d(b(n), 0) \geq \frac{1}{2}d(L_2, 0)$ or taking the reciprocal, $\frac{1}{d(b(n),0)} \leq \frac{2}{d(L_2,0)}$. Hence we find for $n \geq N_2\left(\frac{1}{2}d(L_2, 0)\right)$,

$$\begin{aligned} d\left(\frac{1}{b(n)}, \frac{1}{L_2}\right) &\leq 2\frac{d(L_2, b(n))}{d(L_2, 0)|L_2|} \\ &= \frac{2}{|L_2|^2}d(L_2, b(n)) \end{aligned}$$

The final conclusion is that if we now take our new threshold to be

$$\max\left(\left\{ N_2\left(\frac{1}{2}d(L_2, 0)\right), N_2\left(\frac{|L_2|^2}{2}\delta\right) \right\}\right)$$

for any $\delta > 0$ and we take n to be larger than that threshold, we can conclude that

$$d\left(\frac{1}{b(n)}, \frac{1}{L_2}\right) \leq \delta$$

This way of making the proof by assuming that $b(n) \neq 0$ for all $n \in \mathbb{N}$ also tells us how to proceed in the other case. Indeed, we have just shown that due to $\lim b \neq 0$, there is a certain threshold, above which, $b(n) \neq 0$. So even if that's true in the beginning, Claim 6.12 shows it doesn't matter. □

6.14 Claim. (*The Squeeze Theorem*) If $a : \mathbb{N} \to \mathbb{R}$, $b : \mathbb{N} \to \mathbb{R}$ and $c : \mathbb{N} \to \mathbb{R}$ are sequences such that for each $n \in \mathbb{N}$, $a(n) \leq b(n) \leq c(n)$ and $\lim a = \lim c$ then $\lim a = \lim b = \lim c$.

Proof. For convenience let $l := \lim a = \lim c$. Then for any $\delta > 0$

$$|b(n) - l| \quad < \quad \delta$$

and by Claim 6.1 this is equivalent to

$$l - \delta < \quad b(n) \quad < l + \delta$$

However, we have $a \to l$ and $c \to l$, so for any $\delta > 0$ we can find N large enough such that if $n \geq n$ then $|a(n) - l|$ and $|c(n) - l|$ are both smaller than δ, that is, again by Claim 6.1, equivalent to

$$\begin{aligned} l - \delta < & \quad a(n) & < l + \delta \\ l - \delta < & \quad c(n) & < l + \delta \end{aligned}$$

so we find, $b(n) \leq c(n) < l + \delta$ and $b(n) \geq a(n) > l - \delta$ which means that $|b(n) - l| < \delta$ for all $n \geq N$ (the same threshold of both a and c). Since $\delta > 0$ was arbitrary we are finished. □

6.15 Remark. If we have two sequences $a, b : \mathbb{N} \to \mathbb{R}$ such that $a(n) < b(n)$ for any $n \in \mathbb{N}$ and such that both limits exist, we can "take the limit of the inequality" and the inequality will still hold (though it stops being strict):

$$\lim a \quad \leq \quad \lim b$$

Proof. If $\lim b = \infty$ or $\lim a = -\infty$ then there is nothing to prove. Assume first that both limits are finite. We already know then that the sequence $c := b - a$ converges to $\lim c = \lim b - \lim a$. So our goal is to show that if $c(n) > 0$ then $\lim c \geq 0$. Assume otherwise, that is, assume $\lim c < 0$. Then means that infinitely many n's have $c(n) < 0$, as $d(\lim c, c(n))$ is supposed to be small, that is

$$c(n) \quad < \quad \lim c + \delta$$

for any $\delta > 0$, for n large, and if we pick for example $\delta := -\frac{1}{2} \lim c$, we get that $c(n)$ is strictly negative, which cannot be.

The other cases follow easier reasoning. □

6.16 Claim. We have the following special sequences, where $\alpha, p \in \mathbb{R}$ and $p > 0$

1. If $a : \mathbb{N} \to \mathbb{R}$ is given by $a(n) := n^{-p}$ then

$$\begin{aligned} \lim a &= \lim_{n\to\infty} n^{-p} \\ &= 0\,. \end{aligned}$$

2. If $a : \mathbb{N} \to \mathbb{R}$ is given by $a(n) := p^{\frac{1}{n}}$ then

$$\begin{aligned} \lim a &= \lim_{n\to\infty} p^{\frac{1}{n}} \\ &= 1\,. \end{aligned}$$

3. If $a : \mathbb{N} \to \mathbb{R}$ is given by $a(n) := n^{\frac{1}{n}}$ then

$$\begin{aligned} \lim a &= \lim_{n\to\infty} n^{\frac{1}{n}} \\ &= 1\,. \end{aligned}$$

4. If $a:\mathbb{N}\to\mathbb{R}$ is given by $a(n):=\frac{n^{\alpha}}{(1+p)^{n}}$ then

$$\begin{aligned}\lim a &= \lim_{n\to\infty}\frac{n^{\alpha}}{(1+p)^{n}}\\ &= 0\,.\end{aligned}$$

5. If $a:\mathbb{N}\to\mathbb{R}$ is given by $a(n):=x^{n}$ and $x\in\mathbb{R}$ with $|x|<1$ then $\lim a=0$.

We will not include the proof for these (see [4], Theorem 3.20).

6.17 Claim. If a sequence $a:\mathbb{N}\to\mathbb{R}$ is monotone (as in Definition 5.7) then it either converges to a finite number or it diverges to ∞ or to $-\infty$. If it is both monotone and bounded (as in Definition 5.11) then it necessarily converges to a finite number.

Proof. Assume first that a is monotone increasing. This means that

$$a(n+1)\quad\geq\quad a(n)\qquad(n\in\mathbb{N})$$

If a is not bounded (as in Definition 5.11) then this fits the definition of a sequence that diverges to infinity Definition 6.6. Then assume otherwise that a is bounded by some constant $M\geq 0$. Consider the set of numbers

$$\operatorname{im}(a)\quad\equiv\quad\{\,a(1)\,,a(2)\,,a(3)\,,\dots\,\}$$

which is bounded by M from above and by $a(1)$ from below due to the monotonicity assumption. It is a fact that any bounded subset $S\subseteq\mathbb{R}$ has what is called a *least upper bound*, denoted by $\sup(S)$, which is the smallest possible upper bound on it. That is, it is an upper bound, and it is the smallest in the set of *all* upper bounds. We will show that $\lim a$ exists by showing that $\lim a=\sup(\operatorname{im}(a))$ in this case.

First we need what is called *the approximation property for the supremum.* It says the following: For any bounded set $S\subseteq\mathbb{R}$, and for any $\varepsilon>0$, there is some element $s_{\varepsilon}\in S$ such that $\sup(S)-\varepsilon<s_{\varepsilon}$. Indeed assume otherwise. Then there is some $\varepsilon_0>0$ such that for all $s\in S$, $\sup(S)-\varepsilon\geq s$. But then $\sup(S)-\varepsilon$ is an upper bound on S and since $\varepsilon>0$, $\sup(S)-\varepsilon<\sup(S)$ so that $\sup(S)$ is not the *least* upper bound. Hence we have reached a contradiction.

Using the approximation property for the supremum, let us return now to the question of existence of $\lim a$ and its equality to $\sup(\operatorname{im}(a))$. Let $\delta>0$ be given. Then we know by the approximation property that there is some $n_{\delta}\in\mathbb{N}$ such that $\sup(\operatorname{im}(a))-\delta<a(n_{\delta})$. Due to the monotonicity assumption this implies that for *all* $n\geq n_{\delta}$ we have

$$\sup(\operatorname{im}(a))-\delta\quad<\quad a(n)$$

But also, from the fact that $\sup(\operatorname{im}(a))$ is an upper bound on $\operatorname{im}(a)$ it follows that for *any* $n \in \mathbb{N}$,

$$a(n) \leq \sup(\text{im}(a)) < \sup(\text{im}(a)) + \delta$$

we conclude then that for all $n \geq n_\delta$ we have

$$|a(n) - \sup(\text{im}(a))| < \delta$$

which means that $\lim a = \sup(\text{im}(a))$. □

6.3 Limits of functions from $\mathbb{R} \to \mathbb{R}$

So far we have been dealing with limits sequence, which are functions $\mathbb{N} \to \mathbb{R}$. While there is a lot more to be said about such sequences (in particular the whole development of infinite series, which are sequences $a : \mathbb{N} \to \mathbb{R}$ of the form $a(n) = \sum_{m=1}^{n} b(m)$ for some other sequence $b : \mathbb{N} \to \mathbb{R}$), let us turn our attention to limits of other types of functions, namely, of functions $\mathbb{R} \to \mathbb{R}$. The limits of such functions are richer, since we can explore what happens as the argument approaches more than just infinity. Indeed, if before, with sequences, we had only one direction in which to probe the limit (namely, to keep going forward in the direction of $\mathbb{N}$), for functions whose domain is $\mathbb{R}$, any point can be a limit point, so to speak, which is intimately connected to the continuum property of $\mathbb{R}$ referred to earlier (indeed limits of functions $\mathbb{Q} \to \mathbb{R}$ would make less sense).

The idea is that since we are asking about the hypothetical process of what would happen if we get nearer and nearer to a certain value in the domain (for $\mathbb{N}$ that was the hypothetical value ∞, i.e. what happens to $a(n)$ if n becomes larger and larger), on $\mathbb{R}$ there is a whole continuum of values between any two given points. Hence we could, for instance, get nearer and nearer to the value zero without actually ever touching it. This brings us to our first

6.18 Example. Consider the function $f : (0, \infty) \to \mathbb{R}$ given by $f(x) := \frac{1}{x}$. There are two interesting directions for its limit that we can consider. The first one is the one analogous to what we examined already in Example 6.8. In this case we get the same result (we will define this formally shortly) that $f(x) \to 0$ as $x \to \infty$. However, now we may also approach $x \to 0$ (which for sequences was impossible since we were either *at* zero or we were always a fixed distance away from it–at least distance 1). What happens to $f(x)$ as $x \to 0$? Apparently it diverges to $+\infty$. This is obvious from looking at the sketch of the graph of the function as seen in Figure 6. We can also just plug in values smaller and smaller:

$$\begin{aligned} 1 &\mapsto 1 \\ \frac{1}{2} &\mapsto 2 \\ \frac{1}{3} &\mapsto 3 \\ &\dots \end{aligned}$$

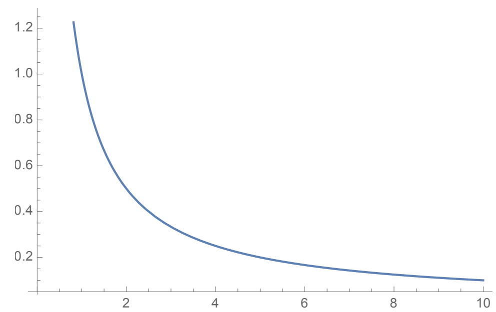


Figure 6: The graph of $(0, \infty) \ni x \mapsto \frac{1}{x}$.

Hence it is clear that now for functions whose domain is $\mathbb{R}$ or a subset of it, we need to measure the distance in the domain as well and not just let the argument go to infinity (in the language of the previous section that mean all n above a certain threshold N). This gives us the following table of options for the limit of a function $f : \mathbb{R} \to \mathbb{R}$:

1. Probe the function at some point $x \in \mathbb{R}$ (which might not lie inside its domain strictly speaking).
2. Probe the function at $+\infty$ (this was the only thing which has an analogue for sequences).
3. Probe the function at $-\infty$.

- The result may converge to a finite number.
- The result may diverge to $\pm\infty$.
- The resulting limit may not exist.

Let us consider a few more examples:

6.19 Example. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by $x \mapsto \cos(x)$. This function has no limit as $x \to \infty$ because it keeps oscillating between ± 1. Same for sin.

6.20 Example. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) := \exp_a(-x)$ for any $a > 1$ (recall Example 5.31) has a limit of zero as $x \to \infty$. (we won't show this now but we could relate it back to Claim 6.16).

6.21 Example. Sometimes the limit does not exist for a silly reason, for example, that the function is different from the left or from the right of a given point. Indeed, consider the step function $f : \mathbb{R} \to \mathbb{R}$ given by $x \mapsto \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$. Then as we approach $x \to 0$ from the right, we are always at 1, and as we approach $x \to 0$ from the left we are always at 0.

6.22 Definition. Let $f : \mathbb{R} \to \mathbb{R}$. We say that $\lim_{x\to\infty} f(x)$ exists and is equal to some $L < \infty$ iff for any $\delta > 0$ there is some $M_\delta > 0$ such that if $x > M_\delta$ then $d(f(x), L) < \delta$.

6.23 Definition. Let $f : \mathbb{R} \to \mathbb{R}$. We say that $\lim_{x\to\infty} f(x)$ diverges to ∞ iff for all $M > 0$ there is some $N > 0$ such that if $x > N$ then $f(x) > M$.

6.24 Remark. Similar definitions could be phrased concerning $-\infty$, either in the domain or in the co-domain of f.

6.25 Definition. (*Limit point of a subset of* $\mathbb{R}$) Let $A \subseteq \mathbb{R}$. The point $l \in \mathbb{R}$ is called a limit point of A iff for any $\varepsilon > 0$ there is some $a \in A \setminus \{ l \}$ such that $d(a, l) < \varepsilon$. We denote by $\overline{A}$ (called *the closure* of A) the union of A together with the set of all its limit points.

6.26 Example. If $A = \{ 1, 2, 3 \}$ then A has no limit points, since we cannot get arbitrarily close to any point from within A, as it is discrete.

6.27 Example. If $A = (0, 1)$, then 1 is a limit point of A, even though $1 \notin A$ itself. 0 is also a limit point, as well as any number in the interior of the interval $a \in (0, 1)$.

6.28 Example. If $A = (0, 1) \cup 2$, the set of limit points of A is $[0, 1]$. In particular, 2 is *not* a limit point of A. Then $\overline{A} = [0, 1] \cup 2$.

More often than not, when we talk about limit points, it will be applied when we take a set $A = (a, b)$ which is an interval and then we want to talk about a or b as limit points of A.

6.29 Definition. Let $f : A \to \mathbb{R}$ with $A \subseteq \mathbb{R}$. Let $x_0 \in \overline{A}$ be a limit point of A. Then we say that $\lim_{x\to x_0} f(x)$ exists and is equal to some $L \in \mathbb{R}$ iff for any $\varepsilon > 0$ there is some $\delta_\varepsilon > 0$ such that for any $x \in A$ such that $d(x, x_0) < \delta_\varepsilon$ we have $d(f(x), L) < \varepsilon$.

6.30 Definition. Let $f : A \to \mathbb{R}$ with $A \subseteq \mathbb{R}$. Let $x_0 \in \overline{A}$ be a limit point of A. Then we say that $\lim_{x\to x_0} f(x)$ diverges to infinity iff for any $M > 0$ there is some $\delta_M > 0$ such that for any $x \in A$ such that $d(x, x_0) < \delta_M$ we have $f(x) \geq M$.

This concept is extremely similar to the limit of a sequence. The only difference is that now we have a slightly different criterion of what "approaching" means: we need to make the distance approached in the domain small as well.

6.31 Remark. The laws of limits of sequences we derived Claim 6.13, Claim 6.14, Claim 6.17 also hold for limits of functions, and we don't repeat them in this context.

6.32 Example. Consider the limit $\lim_{x\to 0} \frac{\sin(x)}{x}$. This is related to a special function called the *sinc* function (see its sketch in Figure 7). Strictly speaking we define $\mathrm{sinc} : \mathbb{R} \to \mathbb{R}$ via the following formula

$$\mathrm{sinc}(x) \quad := \quad \begin{cases} \frac{\sin(x)}{x} & x \neq 0 \\ 1 & x = 0 \end{cases}$$

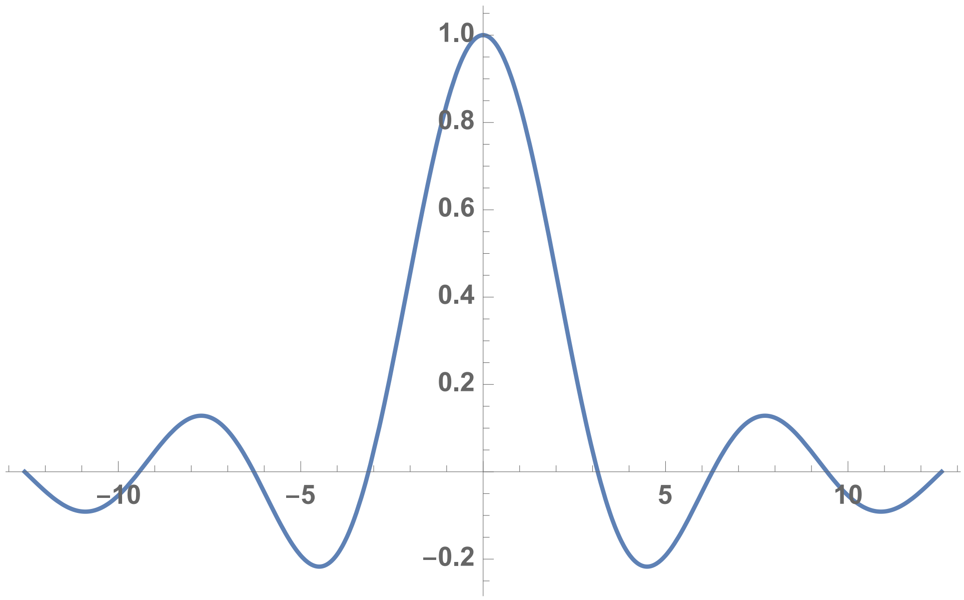


Figure 7: The sketch of the graph of the function sinc.

Note that the function $x \mapsto \frac{\sin(x)}{x}$ strictly speaking does not make sense with zero in its domain, this is the reason for the piecewise definition. Hence the limit is asking if we can stitch the two parts of the piecewise definition together in a sensible way, if we can show that

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

The first step is to show that for all $x \in \mathbb{R} \setminus \{ 0 \}$ we have:

$$\cos(x) \leq \quad \tfrac{\sin(x)}{x} \leq \quad 1\,.$$

Once we have this inequality, we simply employ the squeeze theorem Claim 6.14 since $\lim_{x \to 0} \cos(x) = 1$ (the proof of this fact is similar to related to Example 6.33).

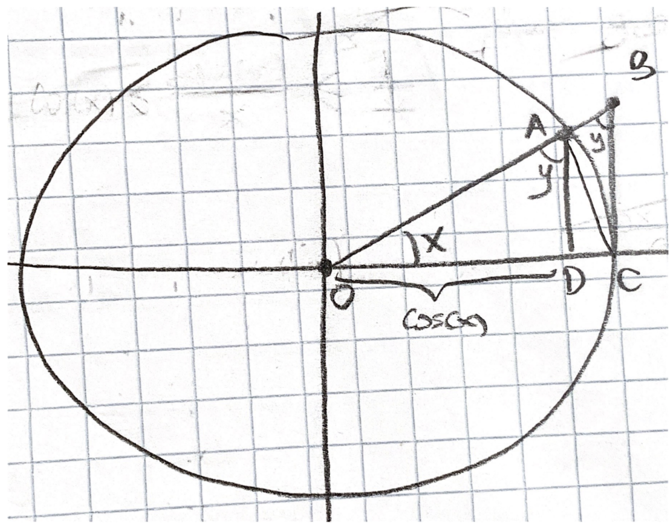
To prove the inequality we take the reciprocal:

$$1 \leq \quad \tfrac{x}{\sin(x)} \leq \quad \frac{1}{\cos(x)}\,.$$

We multiply by $\frac{1}{2}$ and $\sin(x)$ to get

$$\frac{1}{2}\sin(x) \quad \leq \tfrac{x}{2\pi}\pi \leq \quad \frac{1}{2}\frac{\sin(x)}{\cos(x)}\,.$$

Now if we picture a right-triangle inscribed inside of the unit circle as follows:



We note that the sector of the circle with vertices OCA has area $\frac{x}{2\pi}\pi$. The area of the *triangle* with vertices OCA is $\frac{1}{2}\sin(x)\times 1 = \frac{1}{2}\sin(x)$. The area of the triangle with vertices OBC can be find using the sinuses theorem: Applying it on the triangle AOD we get:

$$\frac{\sin(x)}{x} = \frac{\cos(x)}{y} = \frac{1}{\frac{\pi}{2}}$$

Applying it on the triangle BOC we get

$$\frac{BC}{x} = \frac{1}{y} = \frac{OB}{\frac{\pi}{2}}$$

Comparing the two we hence find that the area of the triangle OBC is

$$\frac{1}{2}BC \quad = \quad \frac{1}{2}\frac{x}{y} = \frac{1}{2}\frac{\sin(x)}{\cos(x)}$$

The picture then explains the inequality $\frac{1}{2}\sin(x) \leq \frac{x}{2\pi}\pi \leq \frac{1}{2}\frac{\sin(x)}{\cos(x)}$, and we're finished.

6.33 Example. $\lim_{x\to\frac{\pi}{2}}\cos(x) = 0$. We already do know that $\cos\left(\frac{\pi}{2}\right) = 0$ from the geometric picture. The question is rather can we quantify that as we approach $x \to \frac{\pi}{2}$ we really have $\cos(x) \to 0$ (later on we will see this is the definition of continuity of cos at $\frac{\pi}{2}$). The answer is yes: Given any $\varepsilon > 0$, we want

$$d(\cos(x), 0) \quad \equiv \quad |\cos(x)| \stackrel{!}{<} \varepsilon$$

to hold for all $x \in \mathbb{R}$ such that $d\left(x, \frac{\pi}{2}\right) < \delta_\varepsilon$ (that is, $\delta_\varepsilon > 0$ is the ε-dependent threshold we seek). The way we can prove this is by using the connection between cos and sin. Indeed, $\cos(x) = -\sin\left(x - \frac{\pi}{2}\right)$ for all $x \in \mathbb{R}$. Hence we need to study sin when we plug in small values of the argument. Looking at the geometric picture though, $\sin(x)$ is always smaller than the arc, which is x. Thus

$$\begin{aligned} d(\cos(x), 0) &= |\cos(x)| \\ &= \left|\sin\left(x - \frac{\pi}{2}\right)\right| \\ &\leq \left|x - \frac{\pi}{2}\right| \end{aligned}$$

so in this case we can just pick $\delta_\varepsilon := \varepsilon$!

6.34 Example. Consider $\lim_{x\to y} \frac{x^n - y^n}{x-y}$. After some algebraic manipulation we have

$$\begin{aligned} x^n - y^n &= x^n - y^n + \sum_{k=1}^{n-1} x^{n-k} y^k - \sum_{k=1}^{n-1} x^{n-k} y^k \\ &= \left(x^n + \sum_{k=1}^{n-1} x^{n-k} y^k\right) - \left(y^n + \sum_{k=1}^{n-1} x^{n-k} y^k\right) \\ &\quad \text{(Can expand the sums at no cost to simplify the expressions)} \\ &= \sum_{k=0}^{n-1} x^{n-k} y^k - \sum_{k=1}^{n} x^{n-k} y^k \\ &\quad \text{(Change the index of the second sum from } k \mapsto k+1\text{)} \\ &= \sum_{k=0}^{n-1} x^{n-k} y^k - \sum_{k=0}^{n-1} x^{n-k-1} y^{k+1} \\ &\quad \text{(Pull out a factor of } x \text{ and } y \text{ respectively (they always exist))} \\ &= x\sum_{k=0}^{n-1} x^{n-k-1} y^k - y\sum_{k=0}^{n-1} x^{n-k-1} y^k \\ &\quad \text{(Factorize again since both sums are the same)} \\ &= (x-y)\sum_{k=0}^{n-1} x^{n-k-1} y^k \end{aligned}$$

Hence we learn that

$$\frac{x^n - y^n}{x - y} = \sum_{k=0}^{n-1} x^{n-k-1} y^k$$

so that

$$\lim_{x\to y} \frac{x^n - y^n}{x - y} = \lim_{x\to y} \sum_{k=0}^{n-1} x^{n-k-1} y^k$$

Now we can use Claim 6.13 to find

$$\begin{aligned}\lim_{x\to y}\frac{x^n-y^n}{x-y} &= \sum_{k=0}^{n-1}y^{n-k-1}y^k\\ &= \sum_{k=0}^{n-1}y^{n-1}\\ &= ny^{n-1}\end{aligned}$$

6.35 Example. Consider the sequence $\lim_{\varepsilon\to 0}\frac{\sqrt{x+\varepsilon}-\sqrt{x}}{\varepsilon}$. Then as we already saw in Example 6.10, we may factorize

$$\begin{aligned}\sqrt{x+\varepsilon}-\sqrt{x} &= \frac{(x+\varepsilon)-x}{\sqrt{x+\varepsilon}+\sqrt{x}}\\ &= \frac{\varepsilon}{\sqrt{x+\varepsilon}+\sqrt{x}}\end{aligned}$$

and hence

$$\frac{\sqrt{x+\varepsilon}-\sqrt{x}}{\varepsilon} = \frac{1}{\sqrt{x+\varepsilon}+\sqrt{x}}$$

which means, using Claim 6.13, we only have to evaluate $\lim_{\varepsilon\to 0}\sqrt{x+\varepsilon}+\sqrt{x} = \lim_{\varepsilon\to 0}\sqrt{x+\varepsilon}+\lim_{\varepsilon\to 0}\sqrt{x}$. Now

$$\lim_{\varepsilon\to 0}\sqrt{x+\varepsilon} = \sqrt{x}$$

Indeed, from the equation above we get an estimate of

$$\begin{aligned}\sqrt{x+\varepsilon}-\sqrt{x} &= \frac{\varepsilon}{\sqrt{x+\varepsilon}+\sqrt{x}}\\ &\leq \frac{\varepsilon}{\sqrt{x}}\\ &\to 0 \text{ as } \varepsilon\to 0\,.\end{aligned}$$

We conclude

$$\lim_{\varepsilon\to 0}\frac{\sqrt{x+\varepsilon}-\sqrt{x}}{\varepsilon} = \frac{1}{2\sqrt{x}}\,.$$

6.36 Claim. If, for $A,B\subseteq\mathbb{R}$, $f:A\to B$ is given such that $\lim_{a\to a_0}f(a)=: L_1$ exists and $g:B\to\mathbb{R}$ such that $\lim_{b\to L_1}g(b)=:L_2$ exists then for the composition $g\circ f:A\to\mathbb{R}$, the limit $\lim_{a\to a_0}(g\circ f)(a)$ exists and equals

$$\lim_{a\to a_0}(g\circ f)(a) = \lim_{b\to L_1}g(b)$$

Proof. This proof is extremely similar to Claim 6.43.

Since $f(a) \to L_1$ as $a \to a_0$ we have, for any $\varepsilon > 0$ some $\delta_1(\varepsilon) > 0$ such that if $a \in A$ obeys $d(a, a_0) < \delta_1(\varepsilon)$ then $d(f(a), L_1) < \varepsilon$.

Since $g(b) \to L_2$ as $b \to L_1$, we have for any $\varepsilon > 0$ some $\delta_2(\varepsilon) > 0$ such that if $b \in B$ obeys $d(b, L_1) < \delta_2(\varepsilon)$ then $d(g(b), L_2) < \varepsilon$.

Hence for any $\varepsilon > 0$, if $a \in A$ obeys $d(a, a_0) < \delta_1(\delta_2(\varepsilon))$, we have $d(f(a), L_1) < \delta_2(\varepsilon)$ so that $d(g(f(a)), L_2) < \varepsilon$. But this is precisely what it means that $(g \circ f)(a) \to L_2$ as $a \to a_0$. ☐

6.37 Remark. This also gives us a reparametrization of limits. For example,

$$\lim_{x\to 0} g(x_0 + x) = \lim_{x\to x_0} g(x)$$

Proof. We define the function $f(x) := x_0 + x$. Then $\lim_{x\to 0} f(x) = x_0$ easily and we apply the claim to get

$$\begin{aligned}\lim_{x\to 0} g(x_0 + x) &= \lim_{x\to 0} g(f(x)) \\ &= \lim_{x\to x_0} g(x)\end{aligned}$$

☐

6.38 Example. Consider $\lim_{x\to 0} \frac{\sin(3x)}{x}$. Compare this to Example 6.32. We have

$$\frac{\sin(3x)}{x} = 3\frac{\sin(3x)}{3x}$$

If we define a new function $f : x \mapsto 3x$ for all x and $\operatorname{sinc} : x \mapsto \frac{\sin(x)}{x}$ for all $x \neq 0$ then we have $\frac{\sin(3x)}{3x} = (\operatorname{sinc} \circ f)(x)$. But we know that $f(x) \to 0$ as $x \to 0$ (trivially) so that $(\operatorname{sinc} \circ f)(x) \to 1$ as $x \to 0$ since $\operatorname{sinc}(x) \to 1$ as $x \to 0$. We conclude

$$\lim_{x\to 0} \frac{\sin(3x)}{x} = 3$$

based on Example 6.32.

Example 6.21 pushes us to generalize our definition of limits to *one sided limits:*

6.39 Definition. Let $f : A \to \mathbb{R}$ and $x_0 \in \overline{A}$ a limit point of A. Then the left-sided limit of f at x_0 exists and is equal to L, which is denoted by

$$\lim_{x\to x_0^-} f(x) = L$$

iff for any $\varepsilon > 0$ there is some $\delta > 0$ such that for all $x \in A$ with $x_0 - x > \delta$ we have $d(f(x), L) < \varepsilon$.

The right-sided limit of f at x_0 exists and is equal to L, which is denoted by

$$\lim_{x\to x_0^+} f(x) = L$$

iff for any $\varepsilon > 0$ there is some $\delta > 0$ such that for all $x \in A$ with $x - x_0 > \delta$ we have $d(f(x), L) < \varepsilon$.

6.40 Remark. Due to Claim 6.1, we can say that $\lim_{x\to x_0} f(x)$ exists if and only if both one-sided limits exist and are equal to each other.

6.41 Example. Going back to Example 6.21, where $f(x) := \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$ we have

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

and

$$\lim_{x\to 0^+} f(x) = 1$$

and indeed since the two limits are not equal we do not have that $\lim_{x\to 0} f(x)$ exists!

6.42 Example. If $f : (0,1) \to \mathbb{R}$ then there is no point to ask about the two-sided limits at the end points 0 or 1, since the "other side" is not part of the domain.

Actually there is a relationship between limits of sequences and limits of functions!

6.43 Claim. Let $f : A \to \mathbb{R}$ and $x_0 \in \overline{A}$ be a limit point of A. Then $\lim_{x\to x_0} f(x) = L$ for some $L \in \mathbb{R}$ if and only if for *any* sequence $a : \mathbb{N} \to A$ which converges to x_0 the new sequences $f \circ a : \mathbb{N} \to \mathbb{R}$ converges to L.

Proof. First assume that for *any* sequence $a : \mathbb{N} \to A$ which converges to x_0 the new sequences $f \circ a : \mathbb{N} \to \mathbb{R}$ converges to L. We then want to show that

$$\lim_{x\to x_0} f(x) = L\,.$$

Hence let $\delta > 0$ be given. We seek some $\varepsilon > 0$ such that if $x \in A$ is such that $d(x, x_0) < \varepsilon$ then $d(f(x), L) < \delta$. Assume the contrary. I.e. assume the limit does not converge to L. Then that means that there is some $\delta_0 > 0$ such that for each $\varepsilon > 0$ there is some $x_\varepsilon \in A$ with $d(x_\varepsilon, x_0) < \varepsilon$ yet $d(f(x_\varepsilon), L) > \delta_0$. So pick ε to be a sequence such as $n \mapsto \frac{1}{n}$. Hence there is some $\delta_0 > 0$ such that for each $n \in \mathbb{N}$, there is some $a(n) \in A$ with $d(a(n), x_0) < \frac{1}{n}$ yet $d(f(a(n)), L) > \delta_0$. But that means that $a(n) \to x_0$ yet $(f \circ a)(n)$ does *not* converge to L as assumed. Thus we arrive at a contradiction.

Assume conversely that $\lim_{x\to x_0} f(x) = L$ for some $L \in \mathbb{R}$ and let $a : \mathbb{N} \to A$ be any sequence converging to x_0. We want to show that $f \circ a : \mathbb{N} \to \mathbb{R}$

converges to L. We know by assumption that for any $\varepsilon > 0$: (1) there is some $\delta_\varepsilon > 0$ such that if $x \in A$ is such that $d(x, x_0) < \delta_\varepsilon$ then $d(f(x), L) < \varepsilon$; (2) there is some $N_\varepsilon \in \mathbb{N}$ such that if $n \geq N_\varepsilon$ then $d(a(n), x_0) < \varepsilon$. Then for $n \geq N_{\delta_\varepsilon}$,

$$d(a(n), x_0) \quad < \quad \delta_\varepsilon$$

so that

$$d(f(a(n)), L) \quad < \quad \varepsilon$$

and so we have shown that for $n \geq N_{\delta_\varepsilon}$, $d(f(a(n)), L) < \varepsilon$, i.e. $f \circ a$ is a sequence which converges to L. □

7 Continuity of functions from $\mathbb{R} \to \mathbb{R}$

7.1 Definition. The function $f : A \to \mathbb{R}$ is called continuous at a limit point $x_0 \in A$ iff $\lim_{x \to x_0} f(x) = f(x_0)$. If $x_0 \in A$ is *not* a limit point then f is defined to be continuous at x_0.

7.2 Remark. Note that now for the limit x_0 must really lie inside of A, the domain of f. This means that if x_0 is *not* a limit point of A (e.g. if $A = \{ 1, 2, 3 \}$ then any $f : A \to \mathbb{R}$ is continuous as A has no limit points, so there is nothing to verify).

In principle continuity is a property of a function *at a point of its domain.* But we can also think about continuity globally:

7.3 Definition. If $f : A \to \mathbb{R}$ is continuous for *all* $x \in A$ then f is called continuous.

In general continuity is related to the sketch of the graph of a function having no "jumps", or not being too disjunct. Another way to think about continuity is that it stipulates that the function at a point cannot be too wildly different at nearby points. That is

7.4 Claim. Let $f : A \to \mathbb{R}$ be given and pick some $a \in A$ which is a limit point and such that f is continuous at a. Then if we know the value of f at a, i.e. if we know $f(a)$, we can get an idea of what $f(a+s)$ is for $s \in \mathbb{R}$ very close to zero (i.e. $a + s$ is very close to a). Indeed, continuity at a means that $\lim_{b \to a} f(b) = f(a)$. That means that we can make $d(f(b), f(a))$ arbitrarily small if we make $d(b, a)$ arbitrarily small. More precisely, for any $\varepsilon > 0$ there exists some $\delta_\varepsilon > 0$ such that if $b \in A$ obeys $d(b, a) < \delta_\varepsilon$ then $d(f(b), f(a)) < \varepsilon$. So if we pick s as any number within $(-\delta_\varepsilon, \delta_\varepsilon)$ we find that $d(a + s, a) < \delta_\varepsilon$ so that $d(f(s + a), f(a)) < \varepsilon$. Unpacking what these distance estimates using Claim 6.1 means that:

If we pick $s \in (\delta_\varepsilon, \delta_\varepsilon)$ then

$$f(a) - \varepsilon < \quad f(a + s) \quad < f(a) + \varepsilon$$

so once we know $f(a)$ and continuity of f at a, we get a pretty good idea of what f is nearby a. Coincidentally, this also tells us now something about the meaning of δ_ε: it is the size of the neighborhoud around a for which we get estimates of size ε on $f(a+s)$.

As a general rule of thumb, any function which can be written as a sequence of algebraic manipulations (e.g. $f(x) = 5x + 3 - 8 + x^{100}$) is continuous as long as it is defined (e.g. $x \mapsto \frac{1}{x}$ is not continuous at zero as it is not *defined* at zero).

More complicated functions, such as $\cos : \mathbb{R} \to \mathbb{R}$, $\sin : \mathbb{R} \to \mathbb{R}$, $\exp_a : \mathbb{R} \to (0, \infty)$, $\log_a : (0, \infty) \to \mathbb{R}$ have to be examined and in principle their continuity should not be taken for granted (though these ones listed turn out to be indeed continuous).

Any function defined using the piecewise notation should be highly suspicious in terms of its continuity.

7.5 Example. Going back to Example 6.21, it is clear that f there is continuous on the whole of $\mathbb{R}$ except for the point zero, where it is *not* continuous.

7.6 Example. The function $\mathbb{R} \ni x \mapsto x^2 \in \mathbb{R}$ is continuous. Indeed $\lim_{y \to x} y^2 = x^2$ for any $x \in \mathbb{R}$. To see this, we calculate

$$\begin{aligned} d\left(y^2, x^2\right) &\equiv \left|y^2 - x^2\right| \\ &= \left|(y - x + x)^2 - x^2\right| \\ &= \left|(y-x)^2 + 2x(y-x) + x^2 - x^2\right| \\ &\quad \text{(Use triangle inequality)} \\ &\leq |y-x|^2 + 2x|y-x| \\ &\equiv d(y,x)^2 + 2xd(y,x) \end{aligned}$$

So if we pick $\delta_\varepsilon := \sqrt{x^2 + \varepsilon} - x$ and ask that $d(y,x) < \delta_\varepsilon$ we find that that

$$\begin{aligned} d\left(y^2, x^2\right) &\leq \left(\sqrt{x^2+\varepsilon} - x\right)^2 + 2x\left(\sqrt{x^2+\varepsilon} - x\right) \\ &= \left(\sqrt{x^2+\varepsilon} - x\right)\left(\sqrt{x^2+\varepsilon} - x + 2x\right) \\ &= \left(\sqrt{x^2+\varepsilon} - x\right)\left(\sqrt{x^2+\varepsilon} + x\right) \\ &= x^2 + \varepsilon - x^2 \\ &= \varepsilon \end{aligned}$$

Note that $\delta_\varepsilon > 0$ for *any* $x \in \mathbb{R}$. Indeed, if $x < 0$ this is obvious. If $x \geq 0$, we have

$$\begin{aligned} \sqrt{x^2+\varepsilon} - x &\overset{?}{>} 0 \\ \sqrt{x^2+\varepsilon} &\overset{?}{>} x \\ x^2 + \varepsilon &\overset{?}{>} x^2 \\ \varepsilon &> 0 \end{aligned}$$

indeed.

Actually we first started studying limits of functions $\mathbb{R} \to \mathbb{R}$ and then we introduced the concept of continuity. But now that we have continuity and are familiar with a few functions which are continuous, we may go back and use this in order to calculate limits. Indeed, we

7.7 Claim. Let $A, B \subseteq \mathbb{R}$ be two given subsets. If $f : A \to \mathbb{R}$ is continuous at some limit point $a_0 \in A$, $g : B \to A$ and has a limit at some limit point $b_0 \in B$ which equals $\lim_{b \to b_0} g(b) = a_0$ then we can "push" the limit through a continuous function:

$$\lim_{b \to b_0} f(g(b)) = f\left(\lim_{b \to b_0} g(b)\right)$$

Proof. We know that f is continuous at a_0 and it is a limit point. That means that $\lim_{a \to a_0} f(a)$ exists and equals $f(a_0)$. So we may applyClaim 6.36 (with the roles of f and g actually reversed as in that claim) to obtain that

$$\begin{aligned} \lim_{b \to b_0} f(g(b)) &= \lim_{a \to a_0} f(a) \\ &= (\text{Continuity of } f \text{ at } a_0) \\ &= f(a_0) \\ & (\text{Hypothesis on } g) \\ &= f\left(\lim_{b \to b_0} g(b)\right). \end{aligned}$$

□

7.8 Remark. Coincidentally this also shows us that the composition of continuous functions is a continuous function:

$$\begin{aligned} \lim_{x \to x_0} f(g(x)) &= f\left(\lim_{x \to x_0} g(x)\right) \\ &= f(g(x_0)). \end{aligned}$$

7.9 Example. In one of the homework exercises we had to evaluate the limit

$$\lim_{x \to 0} 2^{2^x} = ?$$

Using the fact that $y \mapsto 2^y$ is continuous, we can now "push" the limit inside twice to get that this limit exists and equals 2.

One of the important consequences of continuity is the

7.10 Theorem. *(Intermediate Value Theorem) Let $f : [a, b] \to \mathbb{R}$ be continuous and pick some $c \in [f(a), f(b)]$ (if $f(b) < f(a)$ then reverse the order of the interval). Then there is some $x \in [a, b]$ such that $f(x) = c$, that is, f takes* all *values in between the values at its end points.*

Proof. This ultimately related back to the fact that the image of an interval under a continuous map is again an interval. Since this fact requires the topological notion of connectedness, we shall not prove it here. □

7.11 Example. Suppose we are looking for the solution $x \in \mathbb{R}$ of the equation $x^{21} - 3x + 1 = 0$. Since there is this really large power of 21 we have no hope for a closed form solution (such as the formula for the quadratic equation). However, we know that the function $\mathbb{R} \ni x \mapsto x^{21} - 3x + 1 \in \mathbb{R}$ is continuous (it is just some basic arithmetic operations). Furthermore, if we plug in $x = -1$ we get

$$(-1)^{21} - 3(-1) + 1 = -1 + 3 + 1 = 3$$

and if we plug in $x = +1$ we get

$$(1)^{21} - 3 + 1 = -1$$

Since $0 \in [-1, 3]$, somewhere between 3 and -1 the continuous function must pass through zero, that is, there *is* a solution (one or more) to the equation (though we still have no ideal what it is).

7.12 Claim. On any great circle around the world, for the temperature, pressure, elevation, carbon dioxide concentration, if the simplification is taken that this varies continuously, there will always exist two antipodal points that share the same value for that variable.

Proof. Let $f : [0, 2\pi] \to \mathbb{R}$ be a continuous function on the circle (parametrized by angle), so that $f(0) = f(2\pi)$. Let us pick any $\alpha \in [0, 2\pi]$ at arbitrary. Define $d(\alpha) := f(\alpha) - f(\alpha + \pi \mod 2\pi)$. I.e., the difference between the value of f at α and the value at its anti-podal point. We know that $d(\alpha + \pi) = -d(\alpha)$ since we merely reverse the point with its anti-podal point. Hence, $d : [0, 2\pi] \to \mathbb{R}$ is a continuous function taking some value and its negative, so that there must be some value α for which $d(\alpha) = 0$, which is precisely what we want. □

7.13 Theorem. *If $f : [a, b] \to \mathbb{R}$ is continuous then there are two numbers $p, q \in [a, b]$ such that*

$$f(p) \geq f(x) \geq f(q) \qquad (x \in [a, b])$$

i.e., f attains its maximum (at p) and minimum (at q).

Proof. The proof of this theorem is again related to notions of topology and is thus outside the scope of our studies. It is related to "boundedness" and "compactness" and the fact that continuity preserves these concepts. □

7.14 Example. $f : (0, 1) \to \mathbb{R}$ given by $x \mapsto x$ does not attain its maximum within $(0, 1)$. Indeed, for any $t \in (0, 1)$, there is some $s \in (0, 1)$ such that $t < s$, because we are on the *open* interval, so we can get arbitrarily close to 1 without actually touching it.

7.15 Corollary. *Any continuous function $f : [a,b] \to \mathbb{R}$ is bounded (as in Definition 5.11).*

8 Derivatives

8.1 Definition. The derivative of $f : A \to \mathbb{R}$ at $x \in A$, denoted by $f'(x)$, is defined as the limit

$$f'(x) \quad := \quad \lim_{\varepsilon \to 0} \frac{f(x+\varepsilon) - f(x)}{\varepsilon}$$

if it exists. If it does, then f is called *differentiable* at x. If f is differentiable on the whole of A, then this defines now a new function $f' : A \to \mathbb{R}$ whose formula is $A \ni x \mapsto f'(x)$. This function is *well-defined* due to limits being unique, if they exist. If f is differentiable only on a subset of A, say, $B \subseteq A$, then f', as a function, is only defined on B.

Sometimes different notations are used for the derivative, the most common one is the Leibniz notation

$$f'(x) \quad = \quad \frac{\mathrm{d}}{\mathrm{d}x} f(x)$$

The problem with this notation (and why we will *not* be using it) is that it forces you to commit to give a name to the independent variable (x in this case), further conflating the function f (a rule on *all* numbers) with the number $f(x)$ (f evaluated at the point x). This confusion between f and $f(x)$, or between f' and $f'(x)$, we try to avoid. We prefer to think of the derivative as a function itself regardless of the name of its argument, so that we prefer to write f' with no mentioning of the name of the argument x.

There is, however, some benefit (and also danger) in the Leibniz notation, since it helps one remember what the derivative actually is: it is the *limit* of a quotient of the difference of the values of the function at near-by points divided by the distance between the nearby points. One should think of d as "Delta" (the Greek letter Δ) which stands for difference or change. Hence we are calculating the quotient "difference in f" by "difference in x". Indeed oftentimes one sees the notation

$$\lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}$$

The danger with this notation is that it (sometimes) makes one take the quotient too literally and forget that there is also a *limit* involved.

Another possible notation for the derivative is with the symbol ∂, which stands usually in math for *partial derivative* if there are several variables on which a function depends. However since for us most functions are of one variable there is no distinction. The way one uses this notation is as

$$\partial f \quad = \quad f'$$

or as

$$\partial_x f \quad = \quad f'$$

if one wishes to make explicit with respect to which variable the differentiation is happening (which in this case is x), which is sometimes useful, especially if we want to refer to a function via the formula defining it (i.e. when the domain and codomain are implicitly obvious). Then one writes conveniently, for instance

$$\partial x^n \quad = \quad n x^{n-1}$$

instead of the more cumbersome

$$(x \mapsto x^n)' \quad = \quad x \mapsto n x^{n-1}\,.$$

To summarize, for us, there is no distinction between ∂_x and $\frac{\mathrm{d}}{\mathrm{d}x}$.

8.2 Remark. Clearly this notion only makes sense if $x \in A$ is a limit point of A.

8.3 Example. The derivative of any constant function is the constant *zero* function.

Proof. The constant function f (no matter what the constant is) will always have $\frac{f(x+\varepsilon)-f(x)}{\varepsilon} = \frac{0}{\varepsilon} = 0$, so that the limit is always zero, no matter which x we plug in. □

8.4 Example. The derivative of the parabola function $f : \mathbb{R} \to \mathbb{R};\ x \mapsto x^2$ exists and equals $x \mapsto 2x$. Indeed,

$$\begin{aligned} f'(x) &= \lim_{\varepsilon\to 0} \frac{f(x+\varepsilon) - f(x)}{\varepsilon} \\ &= \lim_{\varepsilon\to 0} \frac{(x+\varepsilon)^2 - x^2}{\varepsilon} \\ &= \lim_{\varepsilon\to 0} \frac{2\varepsilon x + \varepsilon^2}{\varepsilon} \\ &= \lim_{\varepsilon\to 0} 2x + \varepsilon \\ &= \lim_{\varepsilon\to 0} 2x + \lim_{\varepsilon\to 0} \varepsilon \\ &= 2x + 0 \\ &= 2x \end{aligned}$$

8.5 Remark. Knowing the derivative of a function at a certain point gives us additional information in order to approximate it *away* from that point (but still near by), *beyond* the information already provided by continuity as in Claim 7.4. Indeed, let us say that $f : \mathbb{R} \to \mathbb{R}$ is differentiable at some $x \in \mathbb{R}$. That means that

$$f'(x) \quad = \quad \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \left(f(x+\varepsilon) - f(x) \right)$$

exists and is finite. Unpacking what the limit actually means, we get that for any $a>0$, there is some $b_a>0$ such that if $|\varepsilon|<b_a$ then

$$\left|\frac{1}{\varepsilon}\left(f\left(x+\varepsilon\right)-f\left(x\right)\right)-f'\left(x\right)\right| \quad < \quad a$$

or (upon multiplying by $|\varepsilon|$)

$$\begin{aligned} \left|f\left(x+\varepsilon\right)-f\left(x\right)-\varepsilon f'\left(x\right)\right| &< |\varepsilon|\, a \\ &\leq b_a a \\ \text{WLOG } b_a<1 & \\ &< a \end{aligned}$$

that means that

$$f\left(x\right)+\varepsilon f'\left(x\right)-a < \quad f\left(x+\varepsilon\right) \quad < f\left(x\right)+\varepsilon f'\left(x\right)+a$$

so we get even more information about the function near by, namely, *how* and in which direction it changes with ε; cf Claim 7.4.

8.6 Remark. Another, geometric interpretation of the derivative, as as the *slope* of the function at a certain point. The slope is related to the angle of a straight line which is *tangent* to the function at the given point. Recall a straight line is a function of the form

$$\mathbb{R}\ni x \quad \mapsto \quad ax+b$$

where $a,b\in\mathbb{R}$ are the parameters that define the straight line. a is called its slope and it is related to the angle that the straight line forms with the horizontal axis. Indeed, it is the tangent of that angle α: $a=\tan\left(\alpha\right)$. Hence the derivative gives us the angle of the straight line which is tangent to the function at that point, i.e., its slope at that point.

8.7 Remark. Yet another interpretation of the derivative is as *instantaneous rate of change* of the function, at the given point. What that means is, given any point, how quickly does the function increase (if its derivative is positive) or decrease (if its derivative is negative) at a given point, which is of course related to its slope at that point. In physics, if $f:\mathbb{R}\to\mathbb{R}$ denotes the function that gives a particle's position at a certain instance of time, then f' would corresponds to its instantaneous velocity. We will see this in Claim 8.49.

8.8 Example. The derivative of the absolute value $\mathbb{R}\ni x\stackrel{f}{\mapsto}|x|$ does *not* exist at zero, but otherwise exists everywhere else.

$$\begin{aligned} f'\left(0\right) &= \lim_{\varepsilon\to 0}\frac{|0+\varepsilon|-|0|}{\varepsilon} \\ &= \lim_{\varepsilon\to 0}\frac{|\varepsilon|}{\varepsilon} \end{aligned}$$

Now if $\varepsilon > 0$ we get 1 and if $\varepsilon < 0$ we get -1, i.e. the left-sided limit is -1 and the right-sided limit is $+1$, so that the double-sided limit does not exist and hence the function is not differentiable at zero. Anywhere else, e.g. if $x > 0$,

$$\begin{aligned} f'(x) &= \lim_{\varepsilon\to 0}\frac{|x+\varepsilon|-|x|}{\varepsilon} \\ &\quad (x>0 \text{ so } |x|=x;\ \text{For } |\varepsilon|<x,\ |x+\varepsilon|=x+\varepsilon) \\ &= \lim_{\varepsilon\to 0}\frac{x+\varepsilon-x}{\varepsilon} \\ &= 1 \end{aligned}$$

and similar argument if $x < 0$.

8.9 Example. The derivative of sin is cos.

Proof. This was actually a problem on the midterm. Let us see how it works. Pick any $x \in \mathbb{R}$. Then we calculate

$$\sin'(x) \equiv \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(\sin(x+\varepsilon)-\sin(x)\right)$$

We cannot evaluate this limit directly because it is of the indeterminate form $\frac{0}{0}$. So we use $\sin(a)-\sin(b) = 2\sin\left(\frac{a-b}{2}\right)\cos\left(\frac{a+b}{2}\right)$ to get

$$\sin(x+\varepsilon)-\sin(x) = 2\cos\left(x+\frac{\varepsilon}{2}\right)\sin\left(\frac{\varepsilon}{2}\right)$$

so that

$$\begin{aligned} \sin'(x) &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}2\cos\left(x+\frac{\varepsilon}{2}\right)\sin\left(\frac{\varepsilon}{2}\right) \\ &= \lim_{\varepsilon\to 0}\frac{\sin\left(\frac{\varepsilon}{2}\right)}{\left(\frac{\varepsilon}{2}\right)}\cos\left(x+\frac{\varepsilon}{2}\right) \\ &\quad \text{(Use algebraic laws of limits)} \\ &= \left(\lim_{\varepsilon\to 0}\frac{\sin\left(\frac{\varepsilon}{2}\right)}{\left(\frac{\varepsilon}{2}\right)}\right)\left(\lim_{\varepsilon\to 0}\cos\left(x+\frac{\varepsilon}{2}\right)\right) \\ &= \left(\text{Use the definition } \mathrm{sinc}(\mathrm{a}) \equiv \frac{\sin(\mathrm{a})}{\mathrm{a}}\right) \\ &= \left(\lim_{\varepsilon\to 0}\mathrm{sinc}\left(\frac{\varepsilon}{2}\right)\right)\left(\lim_{\varepsilon\to 0}\cos\left(x+\frac{\varepsilon}{2}\right)\right) \end{aligned}$$

Now both sinc and cos are continuous functions, so we may push the limit through. Recall the limit of sinc from Example 6.32: $\lim_{a\to 0}\mathrm{sinc}(a) = 1$. Thus we find

$$\sin'(x) = \cos(x)$$

Since $x \in \mathbb{R}$ was arbitrary, we conclude $\sin' = \cos$. □

8.10 Example. The derivative of cos is $-\sin$.

Proof. Pick any $x \in \mathbb{R}$. Then we have

$$\cos'(x) \equiv \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left(\cos(x+\varepsilon) - \cos(x)\right)$$

Now again we have an indeterminate form $\frac{0}{0}$, so let us use the trigonometric identity

$$\cos(a) - \cos(b) = -2\sin\left(\frac{a+b}{2}\right)\sin\left(\frac{a-b}{2}\right)$$

to get $\cos(x+\varepsilon) - \cos(x) = -2\sin\left(x + \frac{\varepsilon}{2}\right)\sin\left(\frac{\varepsilon}{2}\right)$. Using a very similar trick of identifying a sinc as in the example above we conclude that the limit converges to $-\sin(x)$. □

8.11 Claim. The derivative is *linear*. That means that if f and g are two functions which are differentiable at some $x \in \mathbb{R}$ and $\alpha, \beta \in \mathbb{R}$ then the new function

$$\alpha f + \beta g$$

is differentiable at x_0 with derivative equal to $\alpha f'(x) + \beta g'(x)$, i.e., we can write

$$(\alpha f + \beta g)' = \alpha f' + \beta g'$$

Proof. We have

$$\begin{aligned}
(\alpha f + \beta g)'(x) &\equiv \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left(\alpha f(x+\varepsilon) + \beta g(x+\varepsilon) - \alpha f(x) - \beta g(x)\right) \\
&= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left(\alpha f(x+\varepsilon) - \alpha f(x) + \beta g(x+\varepsilon) - \beta g(x)\right) \\
&= \lim_{\varepsilon \to 0}\left(\frac{1}{\varepsilon}\left(\alpha f(x+\varepsilon) - \alpha f(x)\right) + \frac{1}{\varepsilon}\left(\beta g(x+\varepsilon) - \beta g(x)\right)\right) \\
&\quad \text{(Use algebra of limits)} \\
&= \left(\lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left(\alpha f(x+\varepsilon) - \alpha f(x)\right)\right) + \left(\lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left(\beta g(x+\varepsilon) - \beta g(x)\right)\right) \\
&\quad \text{(Use algebra of limits)} \\
&= \left(\alpha \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left(f(x+\varepsilon) - f(x)\right)\right) + \left(\beta \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left(g(x+\varepsilon) - g(x)\right)\right) \\
&\quad \text{(Use definition of derivatives of } f \text{ and } g\text{)} \\
&\equiv \alpha f'(x) + \beta g'(x)
\end{aligned}$$

□

8.12 Theorem. *If f is differentiable at some $x \in \mathbb{R}$ then f is continuous at x.*

Proof. Let us pick some point $x \in \mathbb{R}$ at which f is differentiable. That means that the following limit exists and is finite

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(f(x+\varepsilon) - f(x) \right)$$

We want to show that $\lim_{\varepsilon \to 0} f(x+\varepsilon)$ exists and equals $f(x)$ (this is the definition of continuity according to Definition 7.1), that is, we want to show that the following equation holds (and the limit in it exists)

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

$$\begin{aligned}
\lim_{\varepsilon \to 0} \left(f(x+\varepsilon) - f(x) \right) &= \lim_{\varepsilon \to 0} \frac{\varepsilon}{\varepsilon} \left(f(x+\varepsilon) - f(x) \right) \\
&\quad \text{(Algebra of limits)} \\
&= \left(\lim_{\varepsilon \to 0} \varepsilon \right) \left(\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(f(x+\varepsilon) - f(x) \right) \right) \\
&= 0 \cdot f'(x) \\
&= 0
\end{aligned}$$

□

8.13 Example. It is clear that the *converse* is false, namely, continuity does *not* imply differentiability. The prime counter example is Example 8.8. The absolute value is *not* differentiable at zero yet it is continuous at zero.

8.14 Claim. The derivative obeys the so-called Leibniz rule for products. That means that if f and g are two functions differentiable at some $x \in \mathbb{R}$ then the new function fg (the product) is also differentiable at x_0 and its derivative is equal to

$$(fg)'(x) = f'(x)\, g(x) + f(x)\, g'(x)$$

i.e. one could write succinctly the Leibniz rule:

$$(fg)' = f'g + fg'$$

Proof. We have

$$\begin{aligned}
(fg)'(x) &\equiv \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(f(x+\varepsilon)g(x+\varepsilon)-f(x)g(x)\right)\\
&= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(f(x+\varepsilon)g(x+\varepsilon)-f(x+\varepsilon)g(x)+f(x+\varepsilon)g(x)-f(x)g(x)\right)\\
&= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(f(x+\varepsilon)\left(g(x+\varepsilon)-g(x)\right)+\left(f(x+\varepsilon)-f(x)\right)g(x)\right)\\
&\quad \text{(Algebra of limits)}\\
&= \left(\lim_{\varepsilon\to 0}f(x+\varepsilon)\frac{1}{\varepsilon}\left(g(x+\varepsilon)-g(x)\right)\right)+\left(\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(f(x+\varepsilon)-f(x)\right)g(x)\right)
\end{aligned}$$

Now we use again the algebra of limits, noting that because f is differentiable at x, it is also continuous at x (as proven in Theorem 8.12) so that $\lim_{\varepsilon\to 0}f(x+\varepsilon)=f(x)$. We find

$$\begin{aligned}
(fg)'(x) &= f(x)\left(\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(g(x+\varepsilon)-g(x)\right)\right)+\left(\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(f(x+\varepsilon)-f(x)\right)\right)g(x)\\
&\quad \text{(Use differentiability of } f \text{ and } g\text{)}\\
&= f(x)g'(x)+f'(x)g(x)
\end{aligned}$$

which is what we wanted to prove. □

8.15 Example. The most important example of the product rule is when one of the functions is a constant: Let $f(x)=cg(x)$ for some $c\in\mathbb{R}$, for all $x\in\mathbb{R}$, where g is a given function. Then

$$\begin{aligned}
f' &= (cg)'\\
&= c'g+cg'
\end{aligned}$$

But c is just a constant, so $c'=0$ as we saw in Example 8.3, so we find $f'=cg'$.

8.16 Claim. Actually we already saw that if $f(x)\equiv x^n$ for some $n\in\mathbb{N}$ then $f'(x)=nx^{n-1}$. Indeed, this was precisely Example 6.34! Actually this rule works for *any* $\alpha\in\mathbb{R}$ on $[0,\infty)$ and not just $n\in\mathbb{N}$: If $f(x)=x^\alpha$ for all $x\in[0,\infty)$ then $f'(x)=\alpha x^{\alpha-1}$ for all $x\in[0,\infty)$.

8.17 Example. For instance, if $f(x):=\frac{1}{x^2}$ for all $x\in\mathbb{R}\setminus\{0\}$, then since $\frac{1}{x^2}=x^{-2}$, we have $f'(x)=-2x^{-3}=-2\frac{1}{x^3}$. Of course since the function f is not defined at zero it is not differentiable there!

8.18 Example. Another example: if $f(x):=\sqrt{x}$ for all $x\geq 0$ then since $\sqrt{x}=x^{\frac{1}{2}}$ we have $f'(x)=\frac{1}{2}x^{\frac{1}{2}-1}=\frac{1}{2}x^{-\frac{1}{2}}=\frac{1}{2}\frac{1}{x^{\frac{1}{2}}}=\frac{1}{2\sqrt{x}}$.

We will see the proof for general n (i.e. not just $n\in\mathbb{N}$) further below once we understand the derivatives of log and exp.

8.19 Claim. For any $a>1$, the logarithm function is differentiable and $\log_a'(x)=\frac{1}{x}\log_a(\mathrm{e})$ where $\mathrm{e}\approx 2.718$ is the natural base of the logarithm as in Definition 10.2. In particular, $\log_{\mathrm{e}}'(x)=\frac{1}{x}$.

Proof. We have

$$\begin{aligned}
\log_a'(x) &\equiv \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(\log_a(x+\varepsilon)-\log_a(x)\right)\\
&\quad\text{(Use logarithm laws)}\\
&= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(\log_a\left(\frac{x+\varepsilon}{x}\right)\right)\\
&= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(\log_a\left(1+\frac{\varepsilon}{x}\right)\right)\\
&\quad\left(\text{Replace } y:=\frac{x}{\varepsilon}\right)\\
&= \lim_{y\to\infty}\frac{1}{x}y\log_a\left(1+\frac{1}{y}\right)\\
&\quad\text{(Use logarithm laws)}\\
&= \lim_{y\to\infty}\frac{1}{x}\log_a\left(\left(1+\frac{1}{y}\right)^y\right)\\
&= \frac{1}{x}\lim_{y\to\infty}\log_a\left(\left(1+\frac{1}{y}\right)^y\right)\\
&\quad\text{(Use continuity of log)}\\
&= \frac{1}{x}\log_a\left(\lim_{y\to\infty}\left(1+\frac{1}{y}\right)^y\right)\\
&\quad\text{(Use definition of natural logarithm base *)}\\
&= \frac{1}{x}\log_a(\mathrm{e})
\end{aligned}$$

where in the last step we used Definition 10.2. □

8.20 Claim. For any $a>1$, the exponential function is differentiable and its derivative is equal to $\exp_a' = \log_{\mathrm{e}}(a)\exp_a$ (recall e from Definition 10.2). In particular, since $\log_{\mathrm{e}}(\mathrm{e})=1$ we find that

$$\exp_{\mathrm{e}}' = \exp_{\mathrm{e}}\,.$$

Proof. We have, for any $x\in\mathbb{R}$,

$$\begin{aligned}
\exp_a'(x) &\equiv \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(\exp_a(x+\varepsilon)-\exp_a(x)\right)\\
&\quad\text{(Use exponential laws)}\\
&= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(\exp_a(x+\varepsilon)\exp_a(-x)-1\right)\exp_a(x)\\
&= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(\exp_a(\varepsilon)-1\right)\exp_a(x)
\end{aligned}$$

so we would be finished if we could show that $\lim_{\varepsilon\to0}\frac{1}{\varepsilon}\left(\exp_a(\varepsilon)-1\right) = \log_e(a)$. Let us rewrite

$$\begin{aligned}
\lim_{\varepsilon\to0}\frac{1}{\varepsilon}\left(\exp_a(\varepsilon)-1\right) &= \lim_{\varepsilon\to0}\frac{\exp_a(\varepsilon)-1}{\log_a(\exp_a(\varepsilon))} \\
&= \lim_{\varepsilon\to0}\frac{1}{\frac{1}{\exp_a(\varepsilon)-1}\log_a(\exp_a(\varepsilon)-1+1)} \\
& \quad \text{(Use logarithm laws)} \\
&= \lim_{\varepsilon\to0}\frac{1}{\log_a\left((1+\exp_a(\varepsilon)-1)^{\frac{1}{\exp_a(\varepsilon)-1}}\right)} \\
& \quad \left(\text{Use continuity of } \alpha\mapsto\frac{1}{\log_a(\alpha)} \text{ to push the limit through}\right) \\
&= \frac{1}{\log_a\left(\lim_{\varepsilon\to0}(1+\exp_a(\varepsilon)-1)^{\frac{1}{\exp_a(\varepsilon)-1}}\right)}
\end{aligned}$$

Now we use the fact that $\exp_a(\varepsilon)\to1$ as $\varepsilon\to0$, since $\exp_a$ is continuity at zero. Hence if $g(\varepsilon):=\exp_a(\varepsilon)-1$ and $f(\alpha):=(1+\alpha)^{\frac{1}{\alpha}}$, what we have is the limit $\lim_{\varepsilon\to0}(f\circ g)(\varepsilon)$, and we have already learnt in Claim 6.36 that since the limit of g at zero exists and equals zero, this limit equals $\lim_{\alpha\to0}f(\alpha)$, so we find

$$\lim_{\varepsilon\to0}\frac{1}{\varepsilon}\left(\exp_a(\varepsilon)-1\right) = \frac{1}{\log_a\left(\lim_{\alpha\to0}(1+\alpha)^{\frac{1}{\alpha}}\right)}$$

But this inner limit is precisely Definition 10.2, so that we find

$$\begin{aligned}
\lim_{\varepsilon\to0}\frac{1}{\varepsilon}\left(\exp_a(\varepsilon)-1\right) &= \frac{1}{\log_a(\mathrm{e})} \\
& \quad \text{(Use logarithm laws)} \\
&= \frac{1}{\log_e(\mathrm{e})/\log_e(a)} \\
&= \frac{1}{1/\log_e(a)} \\
&= \log_e(a)
\end{aligned}$$

which is what we were looking for. □

8.21 Claim. If $f:A\to B$ is differentiable at $x\in A$ and $g:B\to\mathbb{R}$ is differentiable at $f(x)$ then $g\circ f$ is differentiable $x\in A$ and its derivative is equal to

$$(g\circ f)'(x) = g'(f(x))f'(x)$$

or more succinctly

$$(g \circ f)' \quad = \quad (g' \circ f)\, f'$$

This is called the composition rule or the *chain rule*.

Proof. We have

$$\begin{aligned}
(g \circ f)'(x) \quad &\equiv \quad \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left(g\left(f\left(x+\varepsilon\right)\right) - g\left(f\left(x\right)\right)\right) \\
&\quad \quad \text{(Rewrite the same thing)} \\
&= \quad \lim_{\varepsilon \to 0} \frac{g\left(f\left(x+\varepsilon\right)\right) - g\left(f\left(x\right)\right)}{f\left(x+\varepsilon\right) - f\left(x\right)} \frac{1}{\varepsilon}\left(f\left(x+\varepsilon\right) - f\left(x\right)\right) \\
&\quad \quad \text{(Limit of products equal to product of limits, if both exist, and use the fact that } f \text{ is diffe} \\
&= \quad \left(\lim_{\varepsilon \to 0} \frac{g\left(f\left(x+\varepsilon\right)\right) - g\left(f\left(x\right)\right)}{f\left(x+\varepsilon\right) - f\left(x\right)}\right) f'(x)
\end{aligned}$$

If we define

$$q(y) \quad := \quad \begin{cases} \frac{g(y) - g(f(x))}{y - f(x)} & y \in \mathbb{R} \setminus \{\, f(x) \,\} \\ g'(f(x)) & y = f(x) \end{cases}$$

Then we are interested in

$$\begin{aligned}
\lim_{\varepsilon \to 0} \frac{g\left(f\left(x+\varepsilon\right)\right) - g\left(f\left(x\right)\right)}{f\left(x+\varepsilon\right) - f\left(x\right)} \quad &= \quad \lim_{\varepsilon \to 0} q\left(f\left(x+\varepsilon\right)\right) \\
&= \quad \lim_{\varepsilon \to 0} (q \circ f)(x+\varepsilon)
\end{aligned}$$

We note that q is continuous at $f(x)$. Indeed, its continuity is equivalent to the statement that the following limit exists and equals

$$\begin{aligned}
g'(f(x)) \quad &= \quad \lim_{y \to f(x)} q(y) \\
&= \quad \lim_{y \to f(x)} \frac{g(y) - g(f(x))}{y - f(x)}
\end{aligned}$$

which is the very definition of g being differentiable at $f(x)$ (which we assume). Now since q is continuous and f is continuous (as it is differentiable, by Theorem 8.12), $q \circ f$ is continuous by Remark 7.8, so that $\lim_{\varepsilon \to 0}(q \circ f)(x+\varepsilon) = (q \circ f)(x) = q(f(x)) = g'(f(x))$ and we are finished. □

8.22 Example. Let us try to evaluate the derivative of $\exp_2 \circ \sin$. This function is given by the formula $\mathbb{R} \ni x \mapsto 2^{\sin(x)}$. We have by Claim 8.21 that

$$(\exp_2 \circ \sin)' \quad = \quad (\exp_2' \circ \sin)' \sin'$$

Now we know from Example 8.9 that $\sin' = \cos$ and from Claim 8.20 we know that $\exp_2' = \log_e(2)\exp_2$. Hence

$$(\exp_2 \circ \sin)' \quad = \quad \log_e(2)(\exp_2 \circ \sin)\cos \, .$$

More explicitly, we get the formula

$$x \quad \mapsto \quad \log_e(2)\, 2^{\sin(x)} \cos(x)$$

8.23 Example. Recall that there are the so-called *hyperbolic* functions, which are analogous to the trigonometric functions (with domain and codomain $\mathbb{R}$): the hyperbolic sinus is

$$\sinh(x) \quad \equiv \quad \frac{e^x - e^{-x}}{2}$$

and the cosinus is

$$\cosh(x) \quad \equiv \quad \frac{e^x + e^{-x}}{2}$$

and the tangent is

$$\tanh(x) \quad \equiv \quad \frac{\sinh(x)}{\cosh(x)}$$

hence we can immediately calculate their derivatives and get (with $m(x) \equiv -x$ for all $x \in \mathbb{R}$)

$$\begin{aligned} \sinh' &= \frac{1}{2}\left(\exp' - (\exp \circ m)'\right) \\ &= \frac{1}{2}\left(\exp - (\exp \circ m)\, m'\right) \\ &\quad (\text{Use } m' = -1) \\ &= \frac{1}{2}\left(\exp + (\exp \circ m)\right) \\ &\equiv \cosh \end{aligned}$$

and

$$\begin{aligned} \cosh' &= \frac{1}{2}\left(\exp' + (\exp \circ m)'\right) \\ &= \frac{1}{2}\left(\exp + (\exp \circ m)\, m'\right) \\ &\quad (\text{Use } m' = -1) \\ &= \frac{1}{2}\left(\exp - (\exp \circ m)\right) \\ &\equiv \sinh \end{aligned}$$

Compare this with Example 8.9 and Example 8.10 (i.e. the lack of minus sign on $\cosh'$).

8.24 Example. Now that we have the composition, logarithm and exponential derivatives, let us prove Example 6.34 or Claim 8.16 for powers which are not necessarily natural numbers. So let $f(x) := x^\alpha$ for any $\alpha \in \mathbb{R}$ and $x \in (0, \infty)$. We want to show that $f'(x) = \alpha x^{\alpha-1}$.

Since $x > 0$, it is valid to rewrite $y = \exp(\log(y))$ for any $y > 0$ since log is the inverse of exp. So

$$\begin{aligned} f(x) &= \exp(\log(x^\alpha)) \\ &= \exp(\alpha \log(x)) \end{aligned}$$

Now let us apply the chain rule on this to get

$$\begin{aligned} f'(x) &= \exp'(\alpha \log(x))(\alpha \log(x))' \\ &= \exp(\alpha \log(x)) \alpha \frac{1}{x} \\ &= x^\alpha \alpha \frac{1}{x} \\ &= \alpha x^{\alpha-1} \end{aligned}$$

Now if $x < 0$ then x^α does not necessarily make sense, because we have no prescription to take a root of a negative number.

8.25 Example. Continuing the example above, if $\alpha = -\frac{1}{3}$, then $x^\alpha = \frac{1}{\sqrt[3]{x}}$ and we when $x < 0$, we could write $\sqrt[3]{x} = \sqrt[3]{-|x|}$. Now because taking roots is multiplicative, we have

$$\sqrt[3]{-|x|} = \sqrt[3]{-1}\sqrt[3]{|x|}$$

and $\sqrt[3]{-1} = -1$ (since $(-1)^3 = -1$) and $\sqrt[3]{|x|}$ we know how to do. More formally, let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) := \sqrt[3]{x} = \begin{cases} \sqrt[3]{x} & x > 0 \\ 0 & x = 0 \\ -\sqrt[3]{|x|} & x < 0 \end{cases}$$

For $x > 0$ we could apply Claim 8.16 and get

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

whereas for $x < 0$ we need to apply the chain rule Claim 8.21 to get

$$\begin{aligned} f'(x) &= -\frac{1}{3}|x|^{-\frac{2}{3}}(-1) \\ &= \frac{1}{3}|x|^{-\frac{2}{3}} \end{aligned}$$

At zero the situation is more delicate and the definition of f' from the limit has to be employed:

$$\begin{aligned} f'(0) &\equiv \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(f(\varepsilon)-f(0)\right) \\ &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\begin{cases}\sqrt[3]{\varepsilon} & \varepsilon>0\\ -\sqrt[3]{-\varepsilon} & \varepsilon<0\end{cases} \\ &= \lim_{\varepsilon\to 0}\begin{cases}\varepsilon^{-\frac{2}{3}} & \varepsilon>0\\ |\varepsilon|^{-\frac{2}{3}} & \varepsilon<0\end{cases} \\ &= \lim_{\varepsilon\to 0}|\varepsilon|^{-\frac{2}{3}}\end{aligned}$$

this last limit does not exist–it diverges to $+\infty$. Hence f is not differentiable at zero. The geometric meaning of this ∞ is that the slope of the tangent is actually vertical! $\tan\left(\frac{\pi}{2}\right)=\infty$.

8.26 Example. The derivative of $f:\mathbb{R}\setminus\{0\}\to\mathbb{R}$ whose formula is $f(x)=\frac{1}{x}$ is equal to $f'(x)=-\frac{1}{x^2}$.

Proof. We could either use the rule Claim 8.16 or we can appeal directly to the definition:

$$\begin{aligned} f'(x) &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(\frac{1}{x+\varepsilon}-\frac{1}{x}\right) \\ &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(\frac{x-x-\varepsilon}{x(x+\varepsilon)}\right) \\ &= \lim_{\varepsilon\to 0}\frac{-1}{x^2+\varepsilon x} \\ &\quad \left(\text{Use continuity of } x\mapsto\frac{1}{x}\right) \\ &= -\frac{1}{x^2}\end{aligned}$$

□

8.27 Example. Let $f:\mathbb{R}\to\mathbb{R}$ be defined by the formula

$$f(x) := \begin{cases} x\sin\left(\frac{1}{x}\right) & x\neq 0\\ 0 & x=0\end{cases}$$

When $x\neq 0$, we may use all of our theorems since all functions involved, including $x\mapsto\frac{1}{x}$, are differentiable. We have

$$\begin{aligned} f'(x) &= \sin\left(\frac{1}{x}\right)+x\cos\left(\frac{1}{x}\right)\left(-\frac{1}{x^2}\right) \\ &= \sin\left(\frac{1}{x}\right)-\frac{1}{x}\cos\left(\frac{1}{x}\right)\end{aligned}$$

At $x=0$, we must revert to the definition:

$$\begin{aligned} f'(0) &\equiv \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(f(\varepsilon)-f(0)\right) \\ &= \lim_{\varepsilon\to 0}\sin\left(\frac{1}{\varepsilon}\right) \end{aligned}$$

This latter limit does not exist, so f is not differentiable at zero.

8.28 Example. Let $f:\mathbb{R}\to\mathbb{R}$ be given by

$$f(x) := \begin{cases} x^2\sin\left(\frac{1}{x}\right) & x\neq 0 \\ 0 & 0 \end{cases}$$

At $x\neq 0$ we get

$$\begin{aligned} f'(x) &= 2x\sin\left(\frac{1}{x}\right)+x^2\cos\left(\frac{1}{x}\right)\left(-\frac{1}{x^2}\right) \\ &= 2x\sin\left(\frac{1}{x}\right)-\cos\left(\frac{1}{x}\right) \end{aligned}$$

To understand what happens at zero we again must revert to the definition:

$$\begin{aligned} f'(0) &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(f(\varepsilon)-f(0)\right) \\ &= \lim_{\varepsilon\to 0}\varepsilon\sin\left(\frac{1}{\varepsilon}\right) \end{aligned}$$

This limit tends to zero by the squeeze theorem for example (since $\mathrm{im}(\sin)=[-1,1]$). Hence $f'(0)$ exists and equals zero. However, f' as a function itself is *not* continuous at zero, since we do not have

$$\lim_{\varepsilon\to 0}f'(\varepsilon) = 0$$

Indeed, the left hand side does not even exist since $\cos\left(\frac{1}{\varepsilon}\right)$ does not converge to anything as $\varepsilon\to 0$. Hence f is an example of a function whose derivative exists, but that the derivative function f' is not a continuous function. The fact it's not continuous means that of course it cannot be differentiable itself, that is, f'' does not exist:

If we try to calculate the second derivative we'll find, for $x\neq 0$:

$$\begin{aligned} f''(x) &= 2\sin\left(\frac{1}{x}\right)+2x\cos\left(\frac{1}{x}\right)\left(-\frac{1}{x^2}\right)+\sin\left(\frac{1}{x}\right)\left(-\frac{1}{x^2}\right) \\ &= \left(2-\frac{1}{x^2}\right)\sin\left(\frac{1}{x}\right)-\frac{2}{x}\cos\left(\frac{1}{x}\right) \end{aligned}$$

whereas again at zero special care must be taken

$$\begin{aligned} f''(0) &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(f'(\varepsilon)-f'(0)\right) \\ &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(2\varepsilon\sin\left(\frac{1}{\varepsilon}\right)-\cos\left(\frac{1}{\varepsilon}\right)\right) \\ &= \lim_{\varepsilon\to 0}\left(2\sin\left(\frac{1}{\varepsilon}\right)-\frac{1}{\varepsilon}\cos\left(\frac{1}{\varepsilon}\right)\right) \end{aligned}$$

this last limit indeed does not exist.

8.29 Claim. If f,g are two functions such that g does not take the value zero (so that $\frac{f}{g}$ is a well-defined function–otherwise, restrict) and such that f,g are both differentiable and g' also does not take the value zero, then

$$\left(\frac{f}{g}\right)' = \frac{f'g-fg'}{g^2}\,.$$

Proof. Let us apply the various rules we have in order to figure this. Firstly, let us define the function $r(x) := \frac{1}{x}$ for all $x \neq 0$. This function takes a number and gives its reciprocal. Then we get write $\frac{1}{g} = r\circ g$, and so

$$\frac{f}{g} = f(r\circ g)$$

Hence to differentiate $\frac{f}{g}$ we need to apply both the Leibniz product rule Claim 8.14 as well as the composition rule Claim 8.21:

$$\begin{aligned} \left(\frac{f}{g}\right)' &= (f(r\circ g))' \\ & \text{(Use Leibniz rule)} \\ &= f'(r\circ g) + f\left((r\circ g)'\right) \\ & \text{(Use composition rule on second term)} \\ &= \frac{f'}{g} + f\left((r'\circ g)\,g'\right) \\ & \text{(Rewriting the same thing...)} \\ &= \frac{f'g}{g^2} + (r'\circ g)\,fg' \end{aligned}$$

Now in Example 8.26 we learn that $r' = -r^2$. (i.e. $r'(x) = -\frac{1}{x^2} = -r(x)^2$).

Hence we have $r' \circ g = -\frac{1}{g^2}$. We learn that

$$\begin{aligned} \left(\frac{f}{g}\right)' &= \frac{f'g}{g^2} - \frac{fg'}{g^2} \\ &= \frac{f'g - fg'}{g^2} \end{aligned}$$

which is the result we were looking for. □

8.30 Example. Let us try to evaluate $\tan'$. We know that $\tan \equiv \frac{\sin}{\cos}$. Hence using Claim 8.29

$$\begin{aligned} \tan' &= \left(\frac{\sin}{\cos}\right)' \\ &= \frac{\sin' \cos - \sin \cos'}{\cos^2} \end{aligned}$$

Now we use the rules Example 8.9 and Example 8.10 to find

$$\tan' = \frac{\cos^2 + \sin^2}{\cos^2}$$

However, $\cos^2 + \sin^2 \equiv 1$ (make a drawing if you need to, but this is the Pythagorian theorem). Hence

$$\tan' = \frac{1}{\cos^2}$$

This last expression, $\frac{1}{\cos}$, is sometimes called *the secant*.

Similarly, we have

$$\begin{aligned} \tanh' &= \left(\frac{\sinh}{\cosh}\right)' \\ &= \frac{\sinh' \cosh - \sinh \cosh'}{\cosh^2} \\ &= \frac{\cosh^2 - \sinh^2}{\cosh^2} \end{aligned}$$

Now there is a similar identity for the hyperbolic functions that says that $\cosh^2 - \sinh^2 = 1$ (you can verify this directly) so that

$$\tanh' = \frac{1}{\cosh^2}$$

8.31 Theorem. *(*The Hospital rule; L'Hôpital's rule*) If $f : A \to \mathbb{R}$ and $g : A \to \mathbb{R}$ are both differentiable, and if for some limit point of A, a (which is not necessarily inside of A!) we have $\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = 0$ or (both) $\pm\infty$, and $g'(x) \neq 0$ for all $x \in A$* and $\lim_{x\to a} \frac{f'(x)}{g'(x)}$ *exists, then*

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

8.32 Example. The requirement that $\lim_{x\to a}\frac{f'(x)}{g'(x)}$ exists is crucial. Consider $f(x) = x + \sin(x)$ and $g(x) = x$ at $a = \infty$. Then

$$f'(x) = 1 + \cos(x)$$

and

$$g'(x) = 1$$

so that $\frac{f'(x)}{g'(x)} = \frac{1+\cos(x)}{1} = 1 + \cos(x)$. The limit here does not exist as $x \to \infty$. But we can work with the original quotient to get

$$\begin{aligned}\lim_{x\to\infty}\frac{x+\sin(x)}{x} &= \lim_{x\to\infty} 1 + \frac{\sin(x)}{x}\\ &= 1 + \underbrace{\lim_{x\to\infty}\operatorname{sinc}(x)}_{=0}\\ &= 1\end{aligned}$$

which exists!

8.33 Example. Consider the limit $\lim_{x\to 0}\frac{\exp(x)-1}{x^2+x}$. Since $\exp(0) = 1$, we have the indeterminate form $\frac{0}{0}$. But proceeding with the hospital's rule, we get

$$\begin{aligned}\lim_{x\to 0}\frac{\exp_a(x)-1}{x^2+x} &= \lim_{x\to 0}\frac{\exp_a(x)}{2x+1}\\ &= \frac{\lim_{x\to 0}\exp_a(x)}{\lim_{x\to 0}(2x+1)}\\ &= \frac{1}{1}\\ &= 1\end{aligned}$$

8.34 Claim. The derivative of the inverse function: Let $f : A \to B$ be an *invertible* function in the sense of Definition 5.28. That means that there is some $f^{-1} : B \to A$ such that $f \circ f^{-1} = \mathbb{1}_B$ and $f^{-1} \circ f = \mathbb{1}_A$. Assume that A and B are subsets of $\mathbb{R}$ and let us further assume that f is differentiable and also that its derivative does not take the value zero. Then

$$\left(f^{-1}\right)' = \frac{1}{f'} \circ f^{-1}$$

Proof. We know that $f^{-1} \circ f = \mathbb{1}_A$ by definition of the inverse, where $(\mathbb{1}_A)(x) = x$ for all $x \in A$. Let us differentiate the RHS of the equation:

$$\mathbb{1}_A' = 1$$

i.e. the constant function which always equals 1. On the other hand if we differentiate the LHS of the equation, using Claim 8.21 we find

$$\left(f^{-1}\circ f\right)' = \left(\left(f^{-1}\right)'\circ f\right)f'$$

So we find

$$1 = \left(\left(f^{-1}\right)'\circ f\right)f'$$

or

$$\frac{1}{f'} = \left(f^{-1}\right)'\circ f\,.$$

If we now apply f^{-1} to both sides of the equation from the right (using $f\circ f^{-1} = \mathbb{1}_B$), we get

$$\frac{1}{f'}\circ f^{-1} = \left(f^{-1}\right)'$$

which is the result we were looking for. □

8.35 Example. Let us use this rule in order to find the derivative of arcsin. Recall that arcsin is defined as the inverse of sin. Since sin is in general not invertible, we must restrict its domain and codomain in order to really get an honest inverse. Then if we re-define sin as

$$\sin:\left[-\frac{\pi}{2},\frac{\pi}{2}\right] \to [-1,1]$$

then it is both injective and surjective, and so it is indeed invertible, and we denote its inverse by arcsin (whatever it is, we do not really know how to get a formula for it...) We only know that

$$\begin{aligned}(\sin\circ\arcsin)(x) &= x \qquad (x\in[-1,1])\\ (\arcsin\circ\sin)(x) &= x \qquad \left(x\in\left[-\frac{\pi}{2},\frac{\pi}{2}\right]\right)\end{aligned}$$

Using these two relations it is enough to find $\arcsin'$, even though we still have no idea what the formula for arcsin is!

$$\begin{aligned}\arcsin' &= \frac{1}{\sin'}\circ\arcsin\\ &= \frac{1}{\cos}\circ\arcsin\end{aligned}$$

You might say this is useless, since we still don't have a formula for arcsin. However, cos acting on arcsin is something we can figure out, since we can re-write (from the Pythagorian theorem $\sin^2+\cos^2 = 1$ and using the fact that on $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ the sin is always increasing, so there $\sin'$ is positive and $\cos = \sin'$, so we pick the positive version of the square root to get:

$$\cos = \sqrt{1-\sin^2}$$

so that

$$\begin{aligned}\cos(\arcsin(x)) &= \sqrt{1-(\sin(\arcsin(x)))^2} \\ &= \sqrt{1-x^2}\end{aligned}$$

We find that

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

which is quite remarkable since we *still* have no formula for what arcsin is! The sign may be found by working out when arcsin is increasing vs. decreasing.

8.36 Example. Consider the function $f:[0,\infty)\to[0,\infty)$ given by $f(x):=x^2$ for all $x\in[0,\infty)$. Since we restrict the domain of f (from the naive choice of $\mathbb{R}$), f is injective. Since we restrict the codomain of f (from the naive choice of $\mathbb{R}$), f is surjective. Hence it is bijective and its inverse is given by $f^{-1}(x)=\sqrt{x}$. with the same domain and co-domain. Using Claim 8.34 we can then get make sure that our expectation of the derivative of f^{-1} goes through. Indeed, using Example 6.34 or Example 8.18 we know that

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

And from Claim 8.34 we get (Using $f(x)=x^2$ so $f'(x)=2x=2\mathbb{1}(x)$ and so $\frac{1}{f'(x)}\equiv r(x)$)

$$\begin{aligned}\left(f^{-1}\right)' &= \frac{1}{f'}\circ f^{-1} \\ &= \frac{1}{2}r\circ f^{-1} \\ &= \frac{1}{2}\frac{1}{\sqrt{\cdot}}\end{aligned}$$

8.1 Application: Minima and Maxima

8.37 Definition. Let $f:A\to\mathbb{R}$ be given. Then $a\in A$ is a *maximum point* of f iff $f(x)\le f(a)$ for all $x\in A$. Similarly, a is a *minimum point* of f iff $f(x)\ge f(a)$ for all $x\in A$.

8.38 Definition. A point is an extremum point iff it is either a maximum point or a minimum point.

8.39 Definition. Let $f:A\to\mathbb{R}$ be given. Then $a\in A$ is a *local maximum point* of f iff there is some $\varepsilon>0$ such that $f(x)\le f(a)$ for all $x\in A$ which obeys $d(x,a)<\varepsilon$. Similarly, a is a *local minimum point* of f iff there is some $\varepsilon>0$ such that $f(x)\ge f(a)$ for all $x\in A$ which obeys $d(x,a)<\varepsilon$.

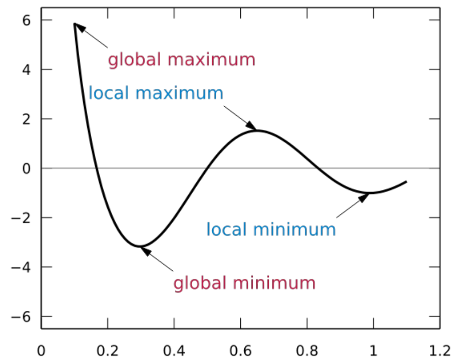


Figure 8: Local and global extrema (source: Wikipedia).

8.40 Definition. A point is a local extremum point iff it is either a local maximum or a local minimum point.

The distinction is that sometimes a function gets to a valley which is not absolutely the lowest point of a function, so we call that point a *local* minimum.

8.41 Definition. Let $f : A \to \mathbb{R}$ be a given differentiable function. Then $a \in A$ is a *stationary point* iff $f'(a) = 0$.

8.42 Theorem. *If $f : [a, b] \to \mathbb{R}$ has a local extremum point at $x \in (a, b)$ and if f is differentiable at x, then $f'(x) = 0$, i.e. x is a stationary point for f.*

Proof. Assume first that the extremum a maximum. We know that

$$f'(x) \quad \equiv \quad \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} (f(x + \varepsilon) - f(x))$$

By the definition of the limit, that means that for any $\eta > 0$ there is some $\delta_\eta > 0$ such that if $|\varepsilon| < \delta_\eta$ then $\left| \frac{1}{\varepsilon} (f(x + \varepsilon) - f(x)) - f'(x) \right| < \eta$. Without loss of generality assume that δ_η is chosen such that $x + \delta_\eta < b$ and $x - \delta_\eta > a$ (otherwise just make δ_η smaller, and the same conclusion would hold, since this should work *any* ε such that $|\varepsilon| < \delta_\eta$). Then if $-\delta_\eta < \varepsilon < 0$, $x - \delta_\eta < x + \varepsilon < x$, so we have (by assumption that x is a maximum point) that $f(x) \geq f(x + \varepsilon)$ and also $\varepsilon < 0$, that is,

$$\frac{1}{\varepsilon} (f(x + \varepsilon) - f(x)) \quad \geq \quad 0$$

since both numerator and denominator are negative. Hence

$$\begin{aligned} f'(x) &\geq \frac{1}{\varepsilon}(f(x+\varepsilon)-f(x))-\eta \\ &\geq 0-\eta \\ &= -\eta \end{aligned}$$

Since η was arbitrary with the constraint, this means that $f'(x) \geq 0$. Similarly, if $0 < \varepsilon < \delta_\eta$, then $x < x+\varepsilon < x+\delta_\eta$ so that again by the assumption of the maximum, $f(x+\varepsilon) \leq f(x)$ and so now

$$\frac{1}{\varepsilon}(f(x+\varepsilon)-f(x)) \leq 0$$

from which we learn that

$$\begin{aligned} f'(x) &\leq \frac{1}{\varepsilon}(f(x+\varepsilon)-f(x))+\eta \\ &\leq 0+\eta \\ &= \eta \end{aligned}$$

and again since $\eta > 0$ was arbitrary this implies $f'(x) \leq 0$.

Because $f'(x) \geq 0$ and $f'(x) \leq 0$, it can only be that $f'(x) = 0$.

The case where x is a local minimum proceeds similarly. □

8.43 Remark. Note the converse is false, that is, finding a stationary point does *not* mean that we found a local extremum!

8.44 Example. Consider the function $f : [-1,1] \to \mathbb{R}$ given by $f(x) = x^3$ for all $x \in [-1,1]$. The derivative is $f'(x) = x^2$. If we want to find stationary points we must solve the equation $f'(x) = 0$ for x, that is, $x^2 = 0$ in this case. The solution to this equation is just the point $x = 0$. Looking at Figure 9, it is immediately clear that the point $x = 0$ is *not* a local extremum point, so what we have is a converse of Theorem 8.42. Such stationary points which are not local extrema are called *inflection points*.

8.45 Theorem. *(Rolle's theorem) Let $f : [a,b] \to \mathbb{R}$ be continuous such that f is differentiable on (a,b). If $f(a) = f(b)$, then there must be some $x \in (a,b)$ such that $f'(x) = 0$.*

Proof. If f is constant then we are finished, because then $f' = 0$. Otherwise, there must be some $t \in (a,b)$ for which $f(t) > f(a)$ or $f(t) < f(a)$.

Assume the former. Then by Theorem 7.13, there is some $y \in [a,b]$ at which f attains its maximum. Since $f(t) > f(a)$, it must be that $y \in (a,b)$. By Theorem 8.42, $f'(y) = 0$ (the point here is that y is not one of the end points a or b).

If on the other hand the latter applies, i.e., $f(t) < f(a)$, then again by Theorem 7.13 we find some $y \in [a,b]$ at which f attains its *minimum.* Now

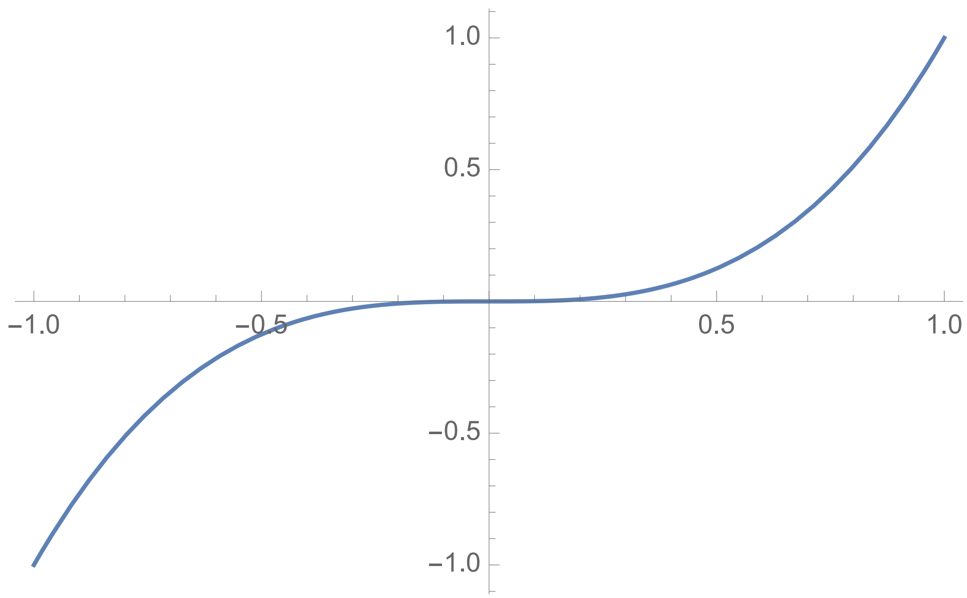


Figure 9: The sketch of the graph of $[-1,1] \ni x \mapsto x^3$.

since $f(t) < f(a)$, this means $y \in (a,b)$ necessarily, and again, $f'(y) = 0$ by Theorem 8.42.

In either case, $f'(y) = 0$ for some $y \in (a,b)$ (and not $y = a$ or $y = b$). $\square$

8.46 Example. If differentiability fails somewhere in the middle then the conclusion of Rolle's theorem fails. For instance, take $f(x) = |x|$. We know this isn't differentiable at $x = 0$ and indeed if f's domain is $[-1,1]$ then there is *no* point in $(-1,1)$ at which the derivative is zero, even though $f(-1) = f(1) = 1$. We actually already saw that $f'(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \\ \text{undefined} & x = 0 \end{cases}$.

8.47 Example. Differentiability is only required at the end points to apply Rolle's theorem. Take $f(x) := \sqrt{1-x^2}$ on $[-1,1]$. Then the function is continuous on $[-1,1]$ and differentiable on $(-1,1)$. Indeed, we have

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

for all $x \neq \pm 1$. But for $x = \pm 1$, the denominator is zero and indeed the function is not differentiable there. None the less, Rolle's theorem applies (since it only requires differentiability on the *open* interval) and so there is indeed a point at which f' is zero in the interior of $(-1,1)$, and that point is zero, as can be seen from the formula.

8.48 Theorem. *(The mean value theorem) If $f:[a,b]\to\mathbb{R}$ is continuous, and differentiable at least on (a,b), then there is some point $x\in(a,b)$ such that*

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

Proof. Let $g:[a,b]\to\mathbb{R}$ be continuous, and differentiable on (a,b). Define

$$h(x) := (f(b)-f(a))\,g(x)-(g(b)-g(a))\,f(x) \qquad (x\in[a,b])$$

Then h is continuous on $[a,b]$ and differentiable at least on (a,b) and

$$h(a) = f(b)\,g(a)-f(a)\,g(b)=h(b)$$

by direct calculation. But now by Theorem 8.45 there must be some $c\in(a,b)$ for which $h'(c)=0$. Note that

$$h'(x) = (f(b)-f(a))\,g'(x)-(g(b)-g(a))\,f'(x)$$

and so at such a point $c\in(a,b)$ we get

$$(f(b)-f(a))\,g'(c) = (g(b)-g(a))\,f'(c)$$

Now if we pick $g:=\mathbb{1}$, then $g'=1$ and $g(b)-g(a)=b-a$ and we find

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

as desired. □

8.49 Claim. Let $f:\mathbb{R}\to\mathbb{R}$ be differentiable. Let $a,b\in\mathbb{R}$ such that $a<b$. If $f'(x)\geq 0$ for all $x\in[a,b]$ then f is increasing on $[a,b]$. If $f'(x)\leq 0$ for all $x\in[a,b]$ then f is decreasing on $[a,b]$.

Proof. Assume first that $f'(x)\geq 0$ for all $x\in[a,b]$. Let $t,s\in[a,b]$ such that $t<s$. To show that f is increasing would mean to show that

$$f(t) \leq f(s)\,.$$

Using Theorem 8.48 on $[t,s]$ we learn that there must be some $\xi\in(t,s)$ such that that

$$f'(\xi) = \frac{f(t)-f(s)}{t-s}\,.$$

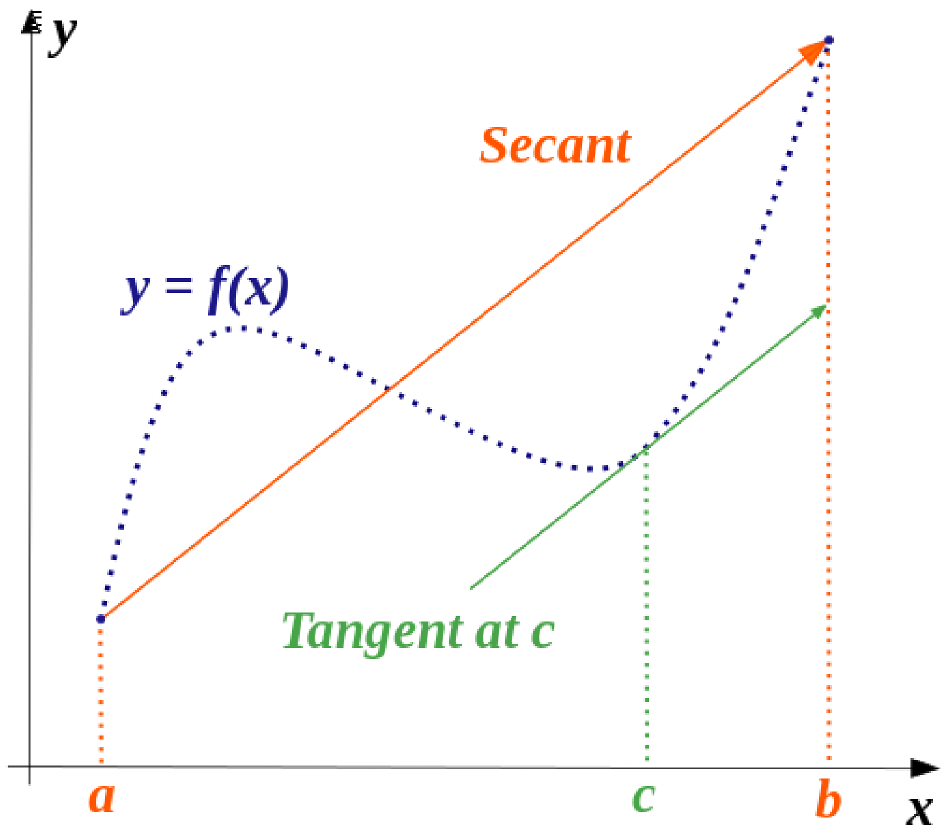


Figure 10: The mean value theorem.

However, since by assumption $f'(x) \geq 0$ for all $x \in [a, b]$, and $(t, s) \subseteq [a, b]$, we must have $f'(\xi) \geq 0$. Also, we know $t - s > 0$, so that

$$\begin{aligned} f(t) - f(s) &= \underbrace{f'(\xi)}_{\geq 0} \underbrace{(t-s)}_{>0} \\ &\geq 0\,. \end{aligned}$$

The proof follows very similar path if $f'(x) \leq 0$ is assumed. □

Claim 8.49 allows us to organize stationary points as follows:

8.50 Corollary. *If $f : \mathbb{R} \to \mathbb{R}$ is a differentiable function such that for some $x \in \mathbb{R}$, $f'(x) = 0$ (i.e., x is a stationary point) then:*

1. *If f' is negative left to x and positive right to x, i.e., it goes from decreasing to increasing, then x is a local minimum for f.*
2. *If f' is positive left to x and negative right to x, i.e., it goes from increasing to decreasing, then x is a local maximum for f.*
3. *If f' is positive left to x and positive right to x, or alternatively it is negative left to x and negative right to x, then x is an inflection point for f.*

8.51 Corollary. *If $f : [a, b] \to \mathbb{R}$ is a given function, then the global extrema points of f will be attained at one of the following points:*

1. *Stationary points.*
2. *Boundary points (i.e. either a or b).*
3. *Points of non-differentiability.*

8.52 Example. Consider the absolute value function $f : [-1, 1] \to \mathbb{R}$ given by $f(x) := |x|$ for all $x \in [-1, 1]$. As we know, f is differentiable everywhere except at zero, and

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

Figure 3 shows that zero is a global minimum for the function, but it is clearly not a stationary point since f is not differentiable there. The global maxima are the boundaries -1 and 1, again, not stationary points.

8.53 Claim. If $f : \mathbb{R} \to \mathbb{R}$ is *twice* differentiable at some $x \in \mathbb{R}$ and $f'(x) = 0$ then:

1. If $f''(x) > 0$ then x is a point of local minimum.
2. If $f''(x) < 0$ then x is a point of local maximum.

If $f''(x) = 0$ the test is not informative.

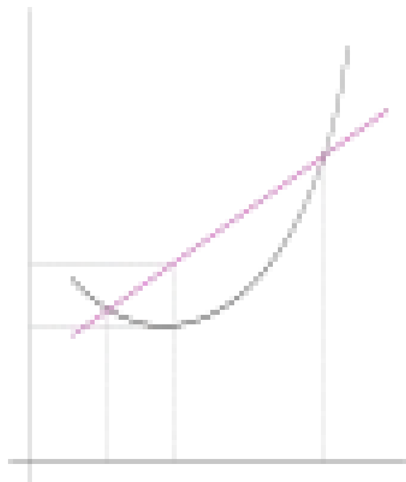


Figure 11: (Wikipedia) Convex function, because the purple line is above the black line.

Proof. Using Claim 8.49, we see that if $f''(x) > 0$, then f' is strictly increasing near x, which means that it must change from negative to positive, so that using Corollary 8.50 we learn that x is a local minimum for f.

Conversely, if $f''(x) < 0$, then by Claim 8.49 f' is strictly decreasing near x, so it changes from positive to negative, so that again by Corollary 8.50 it follows that x is a local maximum for f.

If $f''(x) = 0$, then we found a stationary point for f', which does not give further information about how it changes from being increasing to decreasing or vice versa near x, i.e., nothing about the nature of the stationary point for f. □

8.2 Convexity and concavity

8.54 Definition. Let $a, b \in \mathbb{R}$ such that $a < b$ and let $f : [a, b] \to \mathbb{R}$ be given. f is called *convex* iff for any $x, y \in [a, b]$ and for any $t \in [0, 1]$, f we have

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

f is called *strictly convex* iff for any $x, y \in [a, b]$ and for any $t \in [0, 1]$, f we have

$$f((1-t)x + ty) \quad < \quad (1-t) f(x) + t f(y) \; .$$

f is called (strictly) *concave* iff $-f$ is (strictly) convex.

8.55 Remark. If we think the function

$$[0, 1] \ni t \quad \mapsto \quad f((1-t)x + ty) \in \mathbb{R}$$

we see that it is merely the restriction of f onto the interval $[x,y]$, re-parametrized so as to elapse that interval at length one (i.e. with the variable $t\in[0,1]$ instead of $[x,y]$) so it goes between $(0,f(x))$ and $(1,f(y))$. On the other hand,

$$\begin{aligned}[0,1]\ni t &\mapsto (1-t)\,f(x)+tf(y)\\ &= f(x)+t\,(f(y)-f(x))\end{aligned}$$

corresponds to the straight line between the points $(0,f(x))\in\mathbb{R}^2$ and $(1,f(y))\in\mathbb{R}^2$ with slope $f(y)-f(x)$. Hence, the requirement of convexity is that the graph of the function between any two points always lies below the straight line between the two points, as in Figure 11.

8.56 Claim. $f:[a,b]\to\mathbb{R}$ is convex if and only if for any $s,t,u\in[a,b]$ such that

$$a<s<t<u<b$$

we have

$$\frac{f(t)-f(s)}{t-s}\quad\leq\quad\frac{f(u)-f(t)}{u-t}\,.$$

Proof. Assume f is convex. Define $\lambda:=\frac{t-s}{u-s}$. Since $s<t<u$ holds, $\lambda\in(0,1)$. Also, $1-\lambda=\frac{u-s-t+s}{u-s}=\frac{u-t}{u-s}$. Finally, note that

$$\begin{aligned}1-\lambda &= \frac{u-t}{u-s}\\ &\updownarrow\\ (1-\lambda)(u-s) &= u-t\\ &\updownarrow\\ -s-\lambda(u-s) &= -t\\ &\updownarrow\\ t &= (1-\lambda)\,s+\lambda u\,.\end{aligned}$$

Hence we may apply convexity of f to obtain

$$\begin{aligned}f(t) &= f((1-\lambda)\,s+\lambda u)\\ &\leq (1-\lambda)\,f(s)+\lambda f(u)\\ &= \frac{u-t}{u-s}f(s)+\frac{t-s}{u-s}f(u)\end{aligned}$$

which is in turn equivalent to

$$\begin{array}{rcl} (u-s)\,f(t) & \leq & (u-t)\,f(s)+(t-s)\,f(u) \\ & \updownarrow & \\ (u-s)\,f(t)-(u-s)\,f(s) & \leq & (u-t)\,f(s)+(t-s)\,f(u)-(u-s)\,f(s) \\ & \updownarrow & \\ (u-s)\,(f(t)-f(s)) & \leq & (t-s)\,(f(u)-f(s)) \\ & \updownarrow & \\ \dfrac{f(t)-f(s)}{t-s} & \leq & \dfrac{f(u)-f(s)}{u-s}\,. \end{array}$$

Now if we take

$$(u-s)\,f(t) \quad \leq \quad (u-t)\,f(s)+(t-s)\,f(u)$$

and multiply it by -1 we get

$$\begin{array}{rcl} (s-u)\,f(t) & \geq & (t-u)\,f(s)+(s-t)\,f(u) \\ & \updownarrow & \\ (s-u)\,f(t)+(u-s)\,f(u) & \geq & (t-u)\,f(s)+(s-t)\,f(u)+(u-s)\,f(u) \\ & \updownarrow & \\ (u-s)\,(f(u)-f(t)) & \geq & (u-t)\,(f(u)-f(s)) \\ & \updownarrow & \\ \dfrac{f(u)-f(s)}{u-s} & \leq & \dfrac{f(u)-f(t)}{u-t}\,. \end{array}$$

Plugging these two together we find the condition in the claim. Since all these steps have been equivalences, we can go backwards to show that if this condition is true then f is convex. □

8.57 Claim. Let $f:[a,b]\to\mathbb{R}$ be differentiable. Then f is convex iff f' is monotone non-decreasing.

Proof. Let $x,y\in[a,b]$ such that $x<y$. Pick now also some $\varepsilon>0$ and $\delta>0$. Assume f is convex. Then we have by (two applications of) Claim 8.56 on $x, x+\varepsilon, y$ and on $x+\varepsilon, y, y+\delta$, we get:

$$\frac{f(x+\varepsilon)-f(x)}{\varepsilon} \leq \frac{f(y)-f(x+\varepsilon)}{y-x-\varepsilon} \leq \frac{f(y+\delta)-f(y)}{\delta}\,.$$

Ignoring the middle part and taking the two limits $\varepsilon\to 0$, $\delta\to 0$ we find that $f'(x)\leq f'(y)$. To prove the converse we can use the mean value theorem Theorem 8.48. □

8.58 Corollary. *Let $f:[a,b]\to\mathbb{R}$ be* twice *differentiable. Then f is convex iff*

f'' is non-negative.

Proof. If f'' is non-negative, then by Claim 8.49 f' is monotone non-decreasing, so we can apply Claim 8.57. □

Connecting this with Claim 8.53, we can interpret now that if a function is strictly convex and has a stationary point, then that stationary point is necessarily a minimum!

8.59 Claim. Let $f : [a, b] \to \mathbb{R}$ be differentiable and convex. Then f' is continuous.

Proof. Since f is convex, we learn that f'' is non-negative by Corollary 8.58. □

8.60 Example. If f'' is strictly positive then f is strictly convex, but the converse is not true. Consider $f : [-1, 1] \to \mathbb{R}$ given by $f(x) = x^4$ for all $x \in \mathbb{R}^2$. Then $f''(x) = 12x^2$, which is zero at $x = 0$. However, f is strictly convex.

8.3 Application: Newton's method

Let $f : [a, b] \to \mathbb{R}$ be given such that f is differentiable *and* f' is differentiable as well (i.e. f is twice differentiable), and such that $f(a) < 0$, $f(b) > 0$, $f'(x) \geq \delta > 0$, and $0 \leq f''(x) \leq M$ for all $x \in [a, b]$. Let ξ be the *unique* point in (a, b) at which $f(\xi) = 0$ (which is guaranteed to be unique by Theorem 8.45; see HW7Q2.2).

Pick any $x_1 \in (\xi, b)$, and define a sequence $\mathbb{N} \to \mathbb{R}$, $n \mapsto x_n$ by the formula

$$x_{n+1} := x_n - \frac{f(x_n)}{f'(x_n)} \qquad (n \in \mathbb{N}) .$$

Geometrically, the point x_{n+1} is chosen so that it is where the straight line with slope $f'(x_n)$ passing through $(x_n, f(x_n))$ intercepts the horizontal axis:

$$\frac{f(x_n) - 0}{x_n - x_{n+1}} = f'(x_n) .$$

8.61 Claim. $x_{n+1} < x_n$ and

$$\lim_{n \to \infty} x_n = \xi .$$

Proof. First let us prove that $x_n \geq \xi$. We know already that $x_1 > \xi$. Assume that for all $n \leq m$ for some $m \in \mathbb{N}$, $x_n > \xi$. Check x_{m+1}. Apply the MVT

Theorem 8.48 on f' between (ξ, x_m) to get some $c \in (\xi, x_m)$ such that

$$\frac{f(x_m) - \underbrace{f(\xi)}_{=0}}{x_m - \xi} = f'(c) .$$

But since f' is increasing and $c < x_m$, we have

$$\begin{aligned} \frac{f(x_m)}{x_m - \xi} &\leq f'(x_m) \\ &\Updownarrow \\ f(x_m) &\leq f'(x_m)(x_m - \xi) \\ &\Updownarrow \\ \frac{f(x_m)}{f'(x_m)} - x_m &\leq -\xi \\ &\Updownarrow \\ x_{m+1} &\geq \xi . \end{aligned}$$

This in turn implies (by knowledge that ξ is the unique zero point of f, and that f is increasing) that $f(x_n) > 0$ for all n. We also know $f'(x_n) \geq \delta$ for all n. Hence

$$\begin{aligned} x_{n+1} - x_n &= \frac{-f(x_n)}{f'(x_n)} \\ &< 0 . \end{aligned}$$

But now Claim 6.17 implies that $\lim_{n\to\infty} x_n$ exists and equals ξ. □

8.62 Claim. [[4] Exercise 5.25 (d)] One can prove (though this is beyond the scope of this class since it uses Taylor's theorem) that

$$0 \leq \quad x_n - \xi \quad \leq \frac{2\delta}{M}\left(\frac{M}{2\delta}(x_1 - \xi)\right)^{2(n-1)} \qquad (n \in \mathbb{N}) .$$

so that we get an upper bound on how far we are from the true value ξ and our approximation x_n at any given step $n \in \mathbb{N}$.

8.4 Application: Linear Approximation

Another way to approximate a function with the derivative is via the linear approximation, i.e. going back to Remark 8.5. There, we found that if $f : \mathbb{R} \to \mathbb{R}$ is differentiable, and we know, for some $x \in \mathbb{R}$, both $f(x)$ and $f'(x)$, then if ε is a small number, it is not a bad approximation to write down

$$f(x + \varepsilon) \approx f(x) + \varepsilon f'(x) .$$

The precise meaning of the symbol $\approx$ was given in Remark 8.5, see also HW7Q2.1.

9 Integrals

9.1 The supremum and infimum

The following notions of supremum and infimum have been discussed a bit in the proof of Claim 6.17 (where also an approximation property has been presented) but we repeat them here since they are essential for the definition of the integral.

9.1 Definition. Let $S \subseteq \mathbb{R}$. Then S is called *bounded from above* iff there is some $M \in \mathbb{R}$ such that for all $s \in S$, $s \leq M$. M is then called an upper bound on S (note that many upper bounds exist once one exists).

9.2 Example. $S = \mathbb{R}$ is not bounded from above. $S = (0, 1)$ is bounded above, 1 is an upper bound, but also 2 etc.

9.3 Remark. Sets bounded from below, and lower bounds, are defined similarly.

9.4 Definition. Let $S \subseteq \mathbb{R}$ be a set bounded from above. Then a *supremum* on S is a a number $\alpha \in \mathbb{R}$ such that the following two conditions hold:

1. α is an upper bound on S, in the sense above.

2. If $\beta \in \mathbb{R}$ is any other upper bound on S, then $\alpha \leq \beta$.

Another name for the supremum is *least upper bound*. One then writes $\alpha = \sup(S)$.

9.5 Definition. Let $S \subseteq \mathbb{R}$ be a set bounded from below. Then an infimum on S is a number $\alpha \in \mathbb{R}$ such that the following two conditions hold:

1. α is a lower bound on S, in the sense above.

2. If $\beta \in \mathbb{R}$ is any other lower bound, then $\alpha \geq \beta$.

Another name for the infimum is greatest lower bound. One then writes $\alpha = \inf(S)$.

9.6 Theorem. $\mathbb{R}$ *has a completeness property that any subset of it S bounded from above has a supremum in $\mathbb{R}$, and every set T bounded from below has an infimum in $\mathbb{R}$.*

9.7 Example. Consider the set $\left(0, \sqrt{2}\right) \cap \mathbb{Q}$. This set is bounded above by $\sqrt{2}$. But the supremum, which is $\sqrt{2}$, is not in $\mathbb{Q}$. We have to embed $\mathbb{Q}$ in $\mathbb{R}$ in order to talk about the supremum of this set.

With the supremum and infimum of subsets of $\mathbb{R}$, we are finally ready to start discussing the integral.

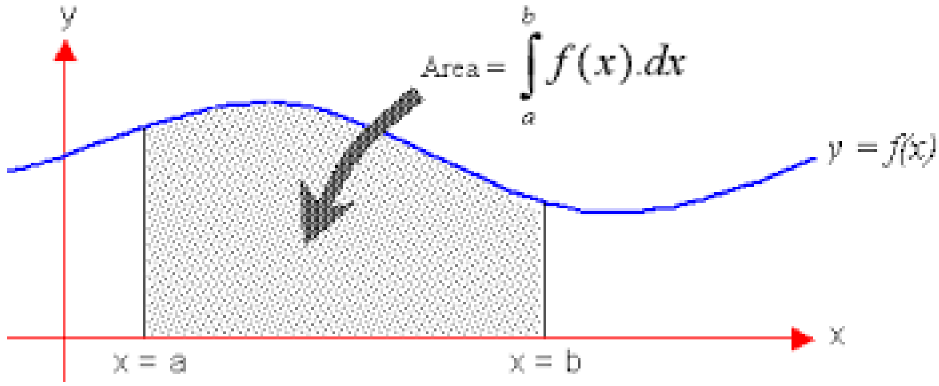


Figure 12: The integral's geometric interpretation.

9.2 The Darboux integral

The notion of the integral has many different interpretations, perhaps the most immediate one is the geometric interpretation which says that for a function $f : [a, b] \to \mathbb{R}$, its integral is the area between:

1. The vertical line defined by the set of points $\left\{ (x, y) \in \mathbb{R}^2 \,\middle|\, x = a \right\}$.
2. The vertical line defined by the set of points $\left\{ (x, y) \in \mathbb{R}^2 \,\middle|\, x = b \right\}$.
3. The horizontal line defined by the set of points $\left\{ (x, y) \in \mathbb{R}^2 \,\middle|\, y = 0 \right\}$.
4. The curve for the function defined by the set of points $\left\{ (x, y) \in \mathbb{R}^2 \,\middle|\, y = f(x) \right\}$.

Since the curve of the function may be very complicated, we want to devise a way to understand very general functions instead of restricting ourselves to simple shapes (like triangles and squares). The way we do it is via approximation by many small rectangles, and this is rigorously defined as follows.

9.8 Definition. (*The Darboux integral*) Let $f : [a, b] \to \mathbb{R}$ be a bounded function (as in Definition 5.11). We define its upper Darboux sum as the limit

$$\lim_{N\to\infty} \overline{S}_a^b (f, N)$$

with

$$\overline{S}_a^b (f, N) \quad := \quad \frac{b-a}{N} \sum_{n=0}^{N-1} \sup \left(\left\{ f(x) \in \mathbb{R} \,\middle|\, x \in a + \left[n\frac{b-a}{N}, (n+1)\frac{b-a}{N} \right] \right\} \right)$$

(if the limit exists at all) and its lower Darboux sum as

$$\lim_{N\to\infty} \underline{S}_a^b (f, N)$$

with

$$\underline{S}_a^b(f,N) \quad := \quad \frac{b-a}{N}\sum_{n=0}^{N-1}\inf\left(\left\{\, f(x)\in\mathbb{R} \,\middle|\, x\in a+\left[n\frac{b-a}{N},(n+1)\frac{b-a}{N}\right]\,\right\}\right)$$

(again, if the limit exists). If these two limits exist and are equal, i.e., if

$$\lim_{N\to\infty}\overline{S}_a^b(f,N) \quad = \quad \lim_{N\to\infty}\underline{S}_a^b(f,N)$$

then f is called integrable on $[a,b]$ and the result of these limits is called its integral on $[a,b]$, and denoted by

$$\int_a^b f$$

or sometimes by

$$\int_a^b f(x)\,\mathrm{d}x\,.$$

which is much more common and natural than the cumbersome

$$\int_a^b (x\mapsto f(x)) \quad \equiv \quad \int_a^b f(x)\,\mathrm{d}x\,.$$

Note that since f is bounded, the set

$$\left\{\, f(x)\in\mathbb{R} \,\middle|\, x\in a+\left[n\frac{b-a}{N},(n+1)\frac{b-a}{N}\right]\,\right\}$$

is bounded (from above and below) for each n, and so it necessarily has a supremum and an infimum.

9.9 Remark. Another common name for this construction is the Riemann integral, which is defined in a slightly different way, but the end result can be proven to be equivalent to our *Darboux integral*.

9.10 Example. The simplest example is the integral of the constant function. Indeed let

$$f:[a,b] \quad \to \quad \mathbb{R}$$

be given by

$$f(x) \quad = \quad c$$

for some $c\in\mathbb{R}$. Then we always get

$$\left\{\, f(x)\in\mathbb{R} \,\middle|\, x\in a+\left[n\frac{b-a}{N},(n+1)\frac{b-a}{N}\right]\,\right\} \quad = \quad \{\,c\,\}$$

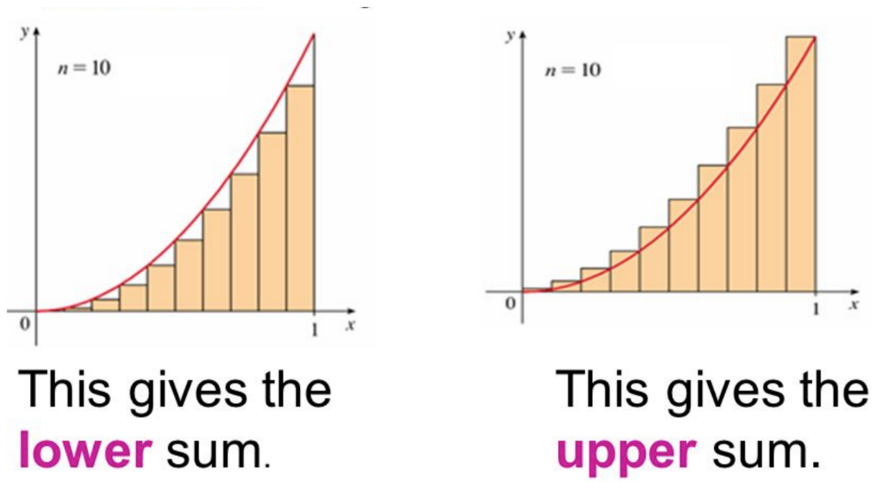


Figure 13: The two Darboux sum limits approximating the integral.

regardless of n, so that the upper Darboux sums are both equal to

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\sup\left(\{\,c\,\}\right) &= \lim_{N\to\infty}\frac{b-a}{N}\sum_{n=0}^{N-1}c \\ &= c\lim_{N\to\infty}\frac{b-a}{N}N \\ &= c\lim_{N\to\infty}(b-a) \\ &= c\,(b-a)\ .\end{aligned}$$

The result is the same for the lower Darboux sum, and so the constant function is integrable and equal to the constant times the length of the interval, i.e. the area of the rectangle trapped between the constant, the horizontal axis, and the limits a and b.

9.11 Example. Let us now take the linear function, $f:[a,b]\to\mathbb{R}$ defined by

$$f(x) \quad = \quad \alpha x+\beta$$

for some $\alpha,\beta\in\mathbb{R}$. Let us assume that $\alpha>0$, so that f is monotonically increasing (otherwise one has to flip all the logic). Note that while this function is not bounded if the domain were $\mathbb{R}$, it is bounded on $[a,b]$. On each set

$$\left\{\,f(x)\in\mathbb{R}\;\middle|\;x\in a+\left[n\frac{b-a}{N},(n+1)\frac{b-a}{N}\right]\,\right\}$$

we have due to the monotonically increasing property, that the supremum is equal to

$$f\left(a+(n+1)\frac{b-a}{N}\right) \quad = \quad \alpha\left(a+(n+1)\frac{b-a}{N}\right)+\beta$$

and the infimum to

$$f\left(a+n\frac{b-a}{N}\right) \quad = \quad \alpha\left(a+n\frac{b-a}{N}\right)+\beta$$

Let us attempt to calculate the upper Darboux sum:

$$\begin{aligned}
\lim_{N\to\infty}\overline{S}_a^b\left(f,N\right) &= \lim_{N\to\infty}\frac{b-a}{N}\sum_{n=0}^{N-1}\sup\left(\left\{\,f\left(x\right)\in\mathbb{R}\,\middle|\,x\in a+\left[n\frac{b-a}{N},\left(n+1\right)\frac{b-a}{N}\right]\,\right\}\right)\\
&= \lim_{N\to\infty}\frac{b-a}{N}\sum_{n=0}^{N-1}\left(\alpha\left(a+\left(n+1\right)\frac{b-a}{N}\right)+\beta\right)\\
&= \lim_{N\to\infty}\left(\alpha\left(\left(b-a\right)a+\left(\frac{b-a}{N}\right)^2\sum_{n=0}^{N-1}\left(n+1\right)\right)+\beta\left(b-a\right)\right)\\
&= \left(b-a\right)\left(\alpha\left(a+\left(b-a\right)\lim_{N\to\infty}\frac{\sum_{n=1}^{N}n}{N^2}\right)+\beta\right)
\end{aligned}$$

Now we use the formula $\sum_{n=1}^{N}n=\frac{1}{2}N\left(N+1\right)$ (see Claim 11.4) to get that this equals

$$\begin{aligned}
&\left(b-a\right)\left(\alpha\left(a+\left(b-a\right)\lim_{N\to\infty}\frac{\frac{1}{2}N\left(N+1\right)}{N^2}\right)+\beta\right)\\
=\ &\left(b-a\right)\left(\alpha\left(a+\frac{1}{2}\left(b-a\right)\lim_{N\to\infty}\left(1+\frac{1}{N}\right)\right)+\beta\right)\\
=\ &\left(b-a\right)\left(\alpha\left(a+\frac{1}{2}\left(b-a\right)\right)+\beta\right)\\
=\ &\left(b-a\right)\left(\frac{\alpha}{2}\left(b+a\right)+\beta\right)\\
=\ &\frac{\alpha}{2}\left(b^2-a^2\right)+\beta\left(b-a\right)\,.
\end{aligned}$$

Similarly, let us verify that the lower Darboux sum is equal to this same thing:

$$\begin{aligned}
\lim_{N\to\infty}\underline{S}_a^b\left(f,N\right) &= \lim_{N\to\infty}\frac{b-a}{N}\sum_{n=0}^{N-1}\inf\left(\left\{\,f\left(x\right)\in\mathbb{R}\,\middle|\,x\in a+\left[n\frac{b-a}{N},\left(n+1\right)\frac{b-a}{N}\right]\,\right\}\right)\\
&= \lim_{N\to\infty}\frac{b-a}{N}\sum_{n=0}^{N-1}\left(\alpha\left(a+n\frac{b-a}{N}\right)+\beta\right)\\
&= \left(\alpha\left(b-a\right)a+\alpha\left(b-a\right)^2\lim_{N\to\infty}\frac{\sum_{n=0}^{N-1}n}{N^2}\right)+\beta\left(b-a\right)
\end{aligned}$$

now

$$\begin{aligned}\lim_{N\to\infty}\frac{\sum_{n=0}^{N-1} n}{N^2} &= \lim_{N\to\infty}\frac{\sum_{n=1}^{N}(n-1)}{N^2}\\ &= \lim_{N\to\infty}\frac{\sum_{n=1}^{N} n-\sum_{n=1}^{N}}{N^2}\\ &= \lim_{N\to\infty}\frac{\frac{1}{2}N(N+1)-N}{N^2}\\ &= \frac{1}{2}\end{aligned}$$

so we indeed get the same result, as

$$\begin{aligned}&\alpha(b-a)a+\frac{1}{2}\alpha(b-a)^2\\ =\;&\alpha(b-a)\left(a+\frac{1}{2}(b-a)\right)\\ =\;&\alpha(b-a)\frac{1}{2}(b+a)\\ =\;&\frac{\alpha}{2}\left(b^2-a^2\right)\,.\end{aligned}$$

In conclusion, the two Darboux sum limits equal, so that f is integrable and we conclude

$$\begin{aligned}\int_a^b f &\equiv \int_a^b(\alpha x+\beta)\,\mathrm{d}x\\ &= \alpha\frac{1}{2}\left(b^2-a^2\right)+\beta(b-a)\,.\end{aligned}$$

This makes perfect sense thinking about the meaning of the integral geometrically, as it is precisely the area of the trapezoid defined by the straight line f.

9.12 Example. Here is an example of when a function is *not* integrable. Define

$$f:[a,b]\quad\to\quad\mathbb{R}$$

by the formula

$$f(x)\quad:=\quad\begin{cases}1 & x\in\mathbb{Q}\\ 0 & x\notin\mathbb{Q}\end{cases}$$

Then f keeps jumping between zero and 1. The lower Darboux sum will be zero, but the upper Darboux sum will be 1 and so we'll get that

$$b-a\quad\neq\quad 0$$

hence the two limits both do exist, but are not equal, and so the function is not integrable.

9.13 Claim. For any $N \in \mathbb{N}$, we have

$$\underline{S}_a^b(f, N) \quad \leq \quad \overline{S}_a^b(f, N) .$$

Proof. The infimum of a set is always smaller than the supremum of that same set. □

9.14 Claim. The upper Darboux sums define a monotone decreasing *sub*sequence in N and the lower Darboux sums define a monotone increasing *sub*sequence in N:

$$\begin{aligned} \underline{S}_a^b(f, N) &\leq \underline{S}_a^b(f, 2N) \qquad (N \in \mathbb{N}) \\ \overline{S}_a^b(f, N) &\geq \overline{S}_a^b(f, 2N) \qquad (N \in \mathbb{N}) . \end{aligned}$$

Proof. Let us consider just the upper sums (the lower sums follow a similar argument). To make the notation a bit shorter and the argument clearer, let us (without loss of generality assume that $a = 0$ and $b = 1$; otherwise one may rescale the function afterwards). Then we have (with $\overline{f}_{\left[\frac{n}{N}, \frac{n+1}{N}\right]} := \sup\left(\left\{ f(x) \,\middle|\, x \in \left[\frac{n}{N}, \frac{n+1}{N}\right] \right\}\right)$ for the sake of brevity)

$$\overline{S}_0^1(f, N) \quad \equiv \quad \frac{1}{N} \sum_{n=0}^{N-1} \overline{f}_{\left[\frac{n}{N}, \frac{n+1}{N}\right]} .$$

This calculation represents a division of the interval $[0, 1]$ into N subintervals, each of length $\frac{1}{N}$. I.e. the boundary points for the sub-intervals are

$$0, \frac{1}{N}, \frac{2}{N}, \frac{3}{N}, \dots, \frac{N-1}{N}, 1 .$$

For $\overline{S}_0^1(f, 2N)$, we divide $[0, 1]$ into $2N$ intervals, each of length $\frac{1}{2N}$, so this is actually a subdivision of the previous one, since now the boundary points of the sub-intervals are

$$0\frac{1}{2N}, \frac{1}{N}, \frac{3}{2N}, \frac{2}{N}, \dots, 1 .$$

Thus we can re-write $\overline{S}_0^1(f,N)$

$$\begin{aligned}
\overline{S}_0^1(f,N) &= \frac{1}{N}\sum_{n=0}^{N-1}\overline{f}_{\left[\frac{n}{N},\frac{n+1}{N}\right]} \\
&= \frac{1}{2N}\sum_{n=0}^{N-1}\overline{f}_{\left[\frac{n}{N},\frac{n}{N}+\frac{1}{2N}\right]\cup\left[\frac{n}{N}+\frac{1}{2N},\frac{n+1}{N}\right]}+\frac{1}{2N}\sum_{n=0}^{N-1}\overline{f}_{\left[\frac{n}{N},\frac{n}{N}+\frac{1}{2N}\right]\cup\left[\frac{n}{N}+\frac{1}{2N},\frac{n+1}{N}\right]} \\
& \quad (\sup(A\cup B)\geq\sup(A)) \\
&\geq \frac{1}{2N}\sum_{n=0}^{N-1}\overline{f}_{\left[\frac{n}{N},\frac{n}{N}+\frac{1}{2N}\right]}+\frac{1}{2N}\sum_{n=0}^{N-1}\overline{f}_{\left[\frac{n}{N}+\frac{1}{2N},\frac{n+1}{N}\right]} \\
&= \frac{1}{2N}\sum_{n=0}^{N-1}\overline{f}_{\left[\frac{2n}{2N},\frac{2n+1}{2N}\right]}+\frac{1}{2N}\sum_{n=0}^{N-1}\overline{f}_{\left[\frac{2n+1}{2N},\frac{2n+2}{2N}\right]} \\
&= \frac{1}{2N}\sum_{n=0}^{2N-1}\overline{f}_{\left[\frac{n}{2N},\frac{n+1}{2N}\right]} \\
&\equiv \overline{S}_0^1(f,2N)
\end{aligned}$$

□

9.15 Theorem. *A function $f:[a,b]\to\mathbb{R}$ is integrable if and only if for any $\varepsilon>0$ there is some $N_\varepsilon\in\mathbb{N}$ such that the upper and lower Darboux sums at N_ε are at most ε far away from each other, that is, that*

$$\overline{S}_a^b(f,N_\varepsilon)-\underline{S}_a^b(f,N_\varepsilon) < \varepsilon\,.$$

Proof. In order to prove this, we need the notions of lim inf and lim sup which we have not yet introduced. □

9.16 Theorem. *A function $f:[a,b]\to\mathbb{R}$ is integrable if it is continuous.*

9.17 Remark. Note that since $f:[a,b]\to\mathbb{R}$ is continuous and defined on a closed interval, it is automatically bounded via Corollary 7.15, so that it was valid to ask if it is integrable at all.

9.18 Example. The converse is false. Consider the function

$$\begin{aligned} f:[0,1] &\to \mathbb{R} \\ x &\mapsto \begin{cases} 5 & x=\frac{1}{2} \\ 0 & x\neq\frac{1}{2}\end{cases}\,. \end{aligned}$$

This function is clearly not continuous, because $\lim_{x\to\frac{1}{2}}f(x)=0$ yet $f\left(\frac{1}{2}\right)=5$. However, it is integrable, and

$$\int_a^b f = 0\,.$$

9.19 Theorem. *A function $f : [a, b] \to \mathbb{R}$ is integrable if it is monotone (increasing or decreasing).*

Proof. Assume that f is monotone increasing without loss of generality. Then

$$\sup\left(\left\{\, f(x) \in \mathbb{R} \;\middle|\; x \in a + \left[n\frac{b-a}{N}, (n+1)\frac{b-a}{N}\right] \,\right\}\right) \quad = \quad f\left(a + (n+1)\frac{b-a}{N}\right)$$

and

$$\inf\left(\left\{\, f(x) \in \mathbb{R} \;\middle|\; x \in a + \left[n\frac{b-a}{N}, (n+1)\frac{b-a}{N}\right] \,\right\}\right) \quad = \quad f\left(a + n\frac{b-a}{N}\right)$$

Hence we have

$$\begin{aligned}
\overline{S}_a^b(f, N) - \underline{S}_a^b(f, N) &= \frac{b-a}{N}\sum_{n=0}^{N-1} f\left(a + (n+1)\frac{b-a}{N}\right) - f\left(a + n\frac{b-a}{N}\right) \\
&= \frac{b-a}{N}\left[\sum_{n=0}^{N-1} f\left(a + (n+1)\frac{b-a}{N}\right) - \sum_{n=0}^{N-1} f\left(a + n\frac{b-a}{N}\right)\right] \\
&= \frac{b-a}{N}\left[\sum_{n=1}^{N} f\left(a + n\frac{b-a}{N}\right) - \sum_{n=0}^{N-1} f\left(a + n\frac{b-a}{N}\right)\right] \\
&= \frac{b-a}{N}\left[f\left(a + N\frac{b-a}{N}\right) + \sum_{n=1}^{N-1} f\left(a + n\frac{b-a}{N}\right) - f(a) - \sum_{n=1}^{N-1} f\left(a + n\frac{b-a}{N}\right)\right] \\
&= \frac{b-a}{N}\left(f(b) - f(a)\right)
\end{aligned}$$

But this can be made arbitrarily small, so that by Theorem 9.15 we conclude that f is integrable. □

9.20 Remark. Again the assumption of monotonicity for a function on a closed interval implies boundedness immediately. To see this, let us assume (without loss of generality) that f is monotone increasing. That means that f attains its maximum at b and its minimum at a so that

$$f(a) \leq \quad f(x) \leq \quad f(b) \qquad (x \in [a, b])$$

and so $\max(\{\, |f(b)|, |f(a)| \,\})$ is a bound on f. Hence it was legitimate to ask whether f was integrable at all.

In light of Example 9.18, it is not surprising that the following is true:

9.21 Theorem. *A function $f : [a, b] \to \mathbb{R}$ is integrable if it is bounded and has only finitely many points of discontinuity.*

Note this last theorem stands in no contrast to Example 9.12, since in that example, f had infinitely many points of discontinuity.

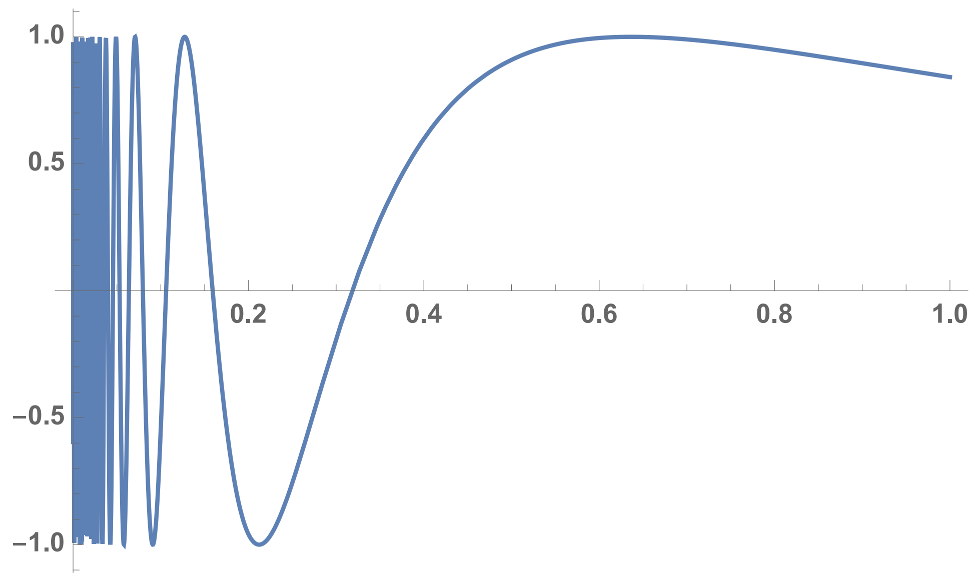


Figure 14: The function $x \mapsto \sin\left(\frac{1}{x}\right)$ is integrable.

9.22 Example. Let $f : [0,1] \to \mathbb{R}$ be given by

$$f(x) \quad := \quad \begin{cases} \sin\left(\frac{1}{x}\right) & x \in (0,1] \\ 0 & x = 0 \end{cases} .$$

This function is certainly bounded, and in fact, it is continuous on $(0,1)$ and has one point of discontinuity, namely, at zero. Hence by Theorem 9.21 it is integrable! By the way,

$$\int_0^1 \sin\left(\frac{1}{x}\right) \mathrm{d}x$$

has no explicit formula.

9.3 Properties of the integral

9.23 Theorem. *Let $f : [a,b] \to \mathbb{R}$ and $g : [a,b] \to \mathbb{R}$ be two integrable functions. Then $f+g$ is integrable, and*

$$\int_a^b (f+b) \quad = \quad \int_a^b f + \int_a^b g .$$

Furthermore, if $\alpha \in \mathbb{R}$ then αf is integrable and

$$\int_a^b \alpha f \quad = \quad \alpha \int_a^b f .$$

9.24 Theorem. *(Monotonicity) Let* $f : [a, b] \to \mathbb{R}$ *and* $g : [a, b] \to \mathbb{R}$ *be two integrable functions. If* $f(x) \leq g(x)$ *for all* $x \in [a, b]$ *then*

$$\int_a^b f \quad \leq \quad \int_a^b g \, .$$

9.25 Theorem. *Let* $f : [a, b] \to \mathbb{R}$ *be integrable and* $c \in (a, b)$*. Then* $f|_{[a,c]} : [a, c] \to \mathbb{R}$ *and* $f|_{[c,b]} : [c, b] \to \mathbb{R}$ *are integrable, and*

$$\int_a^c f + \int_c^b f \quad = \quad \int_a^b f \, .$$

A good way to remember this result is that if $A, B \subseteq \mathbb{R}$ *are two sets with no intersection, then*

$$\int_{A \cup B} f \quad = \quad \int_A f + \int_B f \, ,$$

i.e. the integral on the union is the sum of the integrals.

9.26 Theorem. *Let* $f : [a, b] \to \mathbb{R}$ *be integrable and bounded by* $M \in \mathbb{R}$*, i.e.,* $|f(x)| \leq M$ *for all* $x \in \mathbb{R}$*. Then*

$$\left| \int_a^b f \right| \quad \leq \quad M(b - a) \, .$$

9.27 Theorem. *Let* $f : [a, b] \to \mathbb{R}$ *and* $g : [a, b] \to \mathbb{R}$ *be two integrable functions. Then* $fg : [a, b] \to \mathbb{R}$ *is integrable.*

9.28 Remark. Note that even though this result says that fg is integrable, there is *nothing* like the Leibniz rule (Claim 8.14) for integrals! It is in principle very hard to calculate the integral of fg given the integrals of f and g separately.

9.29 Theorem. *Let* $f : [a, b] \to \mathbb{R}$ *be integrable. Then* $|f| : [a, b] \to \mathbb{R}$ *is integrable, and*

$$\left| \int_a^b f \right| \quad \leq \quad \int_a^b |f| \, .$$

9.30 Theorem. *(Leibniz integral rule) Let* $f : [a, b] \to \mathbb{R}$ *be integrable. Define* $F : [a, b] \to \mathbb{R}$ *by the formula*

$$F(x) \quad := \quad \int_a^x f \qquad (x \in [a, b]) \, .$$

Then F *is continuous, and if* f *is continuous at some* $x \in [a, b]$*, then* F *is differentiable at* x*, and*

$$F'(x) \quad = \quad f(x) \, .$$

Proof. For continuity, let us check that

$$\begin{aligned} F(x) &\stackrel{?}{=} \lim_{\varepsilon\to 0} F(x+\varepsilon) \\ &= \lim_{\varepsilon\to 0}\int_a^{x+\varepsilon} f\,. \end{aligned}$$

Now using Theorem 9.25 we have

$$\begin{aligned} \int_a^{x+\varepsilon} f &= \int_a^x f + \int_x^{x+\varepsilon} f \\ &= F(x) + \int_x^{x+\varepsilon} f\,. \end{aligned}$$

so that we have to prove

$$\lim_{\varepsilon\to 0}\int_x^{x+\varepsilon} f = 0\,.$$

Since f is bounded (as it is integrable!), say, by $M \geq 0$, we have by Theorem 9.26

$$\left|\int_x^{x+\varepsilon} f\right| \leq \varepsilon M$$

and so

$$-\varepsilon M \leq \int_x^{x+\varepsilon} f \leq \varepsilon M$$

and so by the squeeze theorem Claim 6.14 the limit is zero. We learn that F is indeed continuous.

Let us now further assume that f is continuous at some $x \in [a, b]$, and verify that F is differentiable at that same x. We have

$$\begin{aligned} F'(x) &\stackrel{?}{=} \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(F(x+\varepsilon) - F(x)\right) \\ &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(\int_a^{x+\varepsilon} f - \int_a^x f\right)\,. \end{aligned}$$

Now using Theorem 9.25 again we have

$$\int_a^{x+\varepsilon} f = \int_a^x f + \int_x^{x+\varepsilon} f$$

so that

$$F'(x) \stackrel{?}{=} \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_x^{x+\varepsilon} f\,.$$

Since f is continuous at x, for any $e>0$ there is some $d_e>0$ such that if $y\in[a,b]$ is such that $|y-x|<d_e$ then $|f(y)-f(x)|<e$. So when $|\varepsilon|<d_e$, we have for any $y\in[x,x+\varepsilon]$ that $|x-y|<d_e$ and so

$$f(x)-e \leq f(y) \leq f(x)+e$$

Hence we find for such ε using Theorem 9.24 that

$$\int_x^{x+\varepsilon}(f(x)-e)\,\mathrm{d}y \leq \int_x^{x+\varepsilon} f(y)\,\mathrm{d}y \leq \int_x^{x+\varepsilon}(f(x)+e)\,\mathrm{d}y$$

but the outer integrals are of constant functions (in y) so that (upon dividing by $\frac{1}{\varepsilon}$ we learn)

$$f(x)-e \leq \frac{1}{\varepsilon}\int_x^{x+\varepsilon} f(y)\,\mathrm{d}y \leq f(x)+e$$

which is equivalent to $\left|\frac{1}{\varepsilon}\int_x^{x+\varepsilon} f(y)\,\mathrm{d}y - f(x)\right| < e$. Since $e>0$ was arbitrary, we learn that $\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_x^{x+\varepsilon} f$ exists and equals $f(x)$, i.e.

$$F'(x) = f(x)\,.$$

□

9.31 Remark. If one thinks of integration as a map from functions to functions, taking a function f into a new function $x\mapsto\int_a^x f$, i.e.,

$$f \mapsto \left(x\mapsto\int_a^x f\right)$$

and as differentiation as a map from functions to functions f to their derivatives (using the notation in the end of Definition 8.1 for the derivative)

$$f \mapsto \partial f$$

then we learn that in a certain sense, Theorem 9.30 says that, ∂ is the left inverse to $\int$:

$$\partial\circ\int = \mathbb{1}_{\text{functions}}\,.$$

9.32 Theorem. *(The fundamental theorem of calculus) If $f:[a,b]\to\mathbb{R}$ is differentiable and if $f':[a,b]\to\mathbb{R}$ is integrable, then*

$$\int_a^b f' = f(b)-f(a) =: f(x)\big|_{x=a}^{b}\,.$$

Proof. We know by Definition 9.8 that $\int_a^b f'$ is going to be approximated from below by $\underline{S}_a^b(f', N)$ and from above by $\overline{S}_a^b(f', N)$ for some finite N, and that these approximations become better as N grows larger. Consider the upper approximation,

$$\overline{S}_a^b(f', N) \equiv \frac{b-a}{N}\sum_{n=0}^{N-1} \sup\left(\left\{ f'(x) \in \mathbb{R} \,\middle|\, x \in a + \left[n\frac{b-a}{N}, (n+1)\frac{b-a}{N}\right] \right\}\right).$$

Since f is differentiable on each subinterval, we may apply the mean value theorem Theorem 8.48 to get that (for fixed N), for each $n = 0, \ldots, N-1$, there is some $x_n \in \left(a + \left[n\frac{b-a}{N}, (n+1)\frac{b-a}{N}\right]\right)$ such that

$$\begin{aligned} f'(x_n) &= \frac{f\left(a+(n+1)\frac{b-a}{N}\right) - f\left(a+n\frac{b-a}{N}\right)}{a+(n+1)\frac{b-a}{N} - a - n\frac{b-a}{N}} \\ &= \frac{f\left(a+(n+1)\frac{b-a}{N}\right) - f\left(a+n\frac{b-a}{N}\right)}{\frac{b-a}{N}} \end{aligned}$$

Now when we calculate

$$\frac{b-a}{N}\sum_{n=0}^{N-1} f'(x_n) = \frac{b-a}{N}\sum_{n=0}^{N-1} \frac{f\left(a+(n+1)\frac{b-a}{N}\right) - f\left(a+n\frac{b-a}{N}\right)}{\frac{b-a}{N}}$$

the sum telescopes and only the first and last terms survive:

$$\begin{aligned} &= \frac{b-a}{N}\sum_{n=0}^{N-1} \frac{f\left(a+(n+1)\frac{b-a}{N}\right) - f\left(a+n\frac{b-a}{N}\right)}{\frac{b-a}{N}} \\ &= \sum_{n=0}^{N-1} f\left(a+(n+1)\frac{b-a}{N}\right) - f\left(a+n\frac{b-a}{N}\right) \\ &= \left(f\left(a+\frac{b-a}{N}\right) - f(a)\right) + \left(f\left(a+2\frac{b-a}{N}\right) - f\left(a+\frac{b-a}{N}\right)\right) + \cdots + \\ &\quad + \left(f\left(a+N\frac{b-a}{N}\right) - f\left(a+(N-1)\frac{b-a}{N}\right)\right) \\ &= f(b) - f(a)\,. \end{aligned}$$

However, trivially, since $x_n \in \left(a + \left[n\frac{b-a}{N}, (n+1)\frac{b-a}{N}\right]\right)$,

$$\inf\left(\left\{ f'(x) \in \mathbb{R} \,\middle|\, x \in a + \left[n\frac{b-a}{N}, (n+1)\frac{b-a}{N}\right] \right\}\right) \leq f'(x_n) \leq \sup\left(\left\{ f'(x) \in \mathbb{R} \,\middle|\, x \in a + \left[n\frac{b-a}{N},\right.\right.\right.$$

and hence when we sum up,

$$\underline{S}_a^b(f', N) \leq \quad \tfrac{b-a}{N}\textstyle\sum_{n=0}^{N-1} f'(x_n) \quad \leq \overline{S}_a^b(f', N)$$

But we just learnt above that the inner term is independent of N (due to the telescoping) and simply equals $f(b) - f(a)$, hence,

$$\underline{S}_a^b(f', N) \leq \quad f(b) - f(a) \quad \leq \overline{S}_a^b(f', N)\ .$$

Taking the limit $N \to \infty$ of this inequality (Remark 6.15) we learn that

$$\int_a^b f' \leq \quad f(b) - f(a) \quad \leq \int_a^b f'$$

which is equivalent to

$$\int_a^b f' \quad = \quad f(b) - f(a)\ .$$

□

9.33 Remark. Theorem 9.32, which is a culmination of our entire effort in this course, says that differentiation is the right inverse of integration, i.e.

$$\int \circ\partial \quad = \quad \mathbb{1}_{\text{functions}}\ .$$

9.34 Remark. Theorem 9.32 will be our main hammer or work horse to "solve" integrals, rather than the explicit definition Definition 9.8 which is explicitly worked out in Example 9.11 (i.e. one almost never follows the procedure in Example 9.11, which is really presented here more for illustration than an actual computational tool). Using Theorem 9.32 we learn that if we can re-write a function as the derivative of another function, then we can immediately integrate. This is easier said than done, and many many functions for which one can prove (by brute-force) integrability, one still cannot write down an explicit formula for the result of the integral.

At any rate, this tells us immediately the following rules, by essentially undoing Section 8:

1. Derivative of power law: for any $\alpha \in \mathbb{R} \setminus \{-1\}$ from Claim 8.16,

$$\begin{aligned}\int_a^b x^\alpha \mathrm{d}x \quad &= \quad \int_a^b \partial \frac{1}{\alpha+1} x^{\alpha+1}\mathrm{d}x \\ &= \quad \frac{1}{\alpha+1}x^{\alpha+1}\Big|_{x=a}^{b}\ .\end{aligned}$$

   (Note indeed this formula would *not* make sense for $\alpha = -1$ due to $\alpha + 1$ in the denominator).

2. Derivative of reciprocal from Claim 8.19:

$$\begin{aligned}\int_a^b \frac{1}{x}\mathrm{d}x &= \int_a^b \log' \\ &= \log|_a^b .\end{aligned}$$

3. Derivative of trigonometric functions Example 8.10:

$$\begin{aligned}\int_a^b \sin &= \int_a^b -\cos' \\ &= -\int_a^b \cos' \\ &= -\cos|_a^b\end{aligned}$$

   and Example 8.9 yields

$$\begin{aligned}\int_a^b \cos &= \int_a^b \sin' \\ &= \sin|_a^b .\end{aligned}$$

4. Derivative of exponential from Claim 8.20:

$$\begin{aligned}\int_a^b \exp &= \int_a^b \exp' \\ &= \exp|_a^b .\end{aligned}$$

5. Derivative of hyperbolic trigonometric functions from Example 8.23:

$$\begin{aligned}\int_a^b \sinh &= \int_a^b \cosh' \\ &= \cosh|_a^b\end{aligned}$$

   and

$$\begin{aligned}\int_a^b \cosh &= \int_a^b \sinh' \\ &= \sinh|_a^b .\end{aligned}$$

6. More complicated functions, e.g., from Example 8.30

$$\begin{aligned}\int_a^b \frac{1}{\cos^2} &= \int_a^b \tan' \\ &= \tan|_a^b\end{aligned}$$

and

$$\int_a^b \frac{1}{\cosh^2} = \int_a^b \tanh' = \tanh|_a^b .$$

From Example 8.35 we get

$$\int_a^b \frac{1}{\sqrt{1-x^2}}\mathrm{d}x = \int_a^b \arcsin' = \arcsin|_a^b .$$

9.35 Theorem. *(Change of variables) Let $f : [a,b] \to \mathbb{R}$ be continuous (and hence integrable by Theorem 9.16) and $\varphi : [A,B] \to [a,b]$ be continuous such that $\varphi' : [A,B] \to \mathbb{R}$ is continuous. Then*

$$\int_{\varphi(A)}^{\varphi(B)} f = \int_A^B (f \circ \varphi)\,\varphi' .$$

Proof. Define $F : [a,b] \to \mathbb{R}$ by $F(x) := \int_a^x f$. Then by Theorem 9.30, since f is continuous, $F' = f$. Let us calculate, using Theorem 9.32

$$\begin{aligned} \int_A^B (f \circ \varphi)\,\varphi' &= \int_A^B (F' \circ \varphi)\,\varphi' \\ &= \int_A^B (F \circ \varphi)' \\ &\quad \text{(Use the fundamental theorem of calc.)} \\ &= F(\varphi(B)) - F(\varphi(A)) \\ &= \int_{\varphi(A)}^{\varphi(B)} F' \\ &= \int_{\varphi(A)}^{\varphi(B)} f . \end{aligned}$$

□

9.36 Example. Consider the function

$$\begin{aligned} g : [0,2] &\to \mathbb{R} \\ x &\mapsto 2x\sqrt{x^2+1} . \end{aligned}$$

We are interested in

$$\int_0^2 g \equiv \int_0^2 2x\sqrt{x^2+1}\mathrm{d}x .$$

To this end, let us try to use the change of variables theorem above. Define $\varphi : [0,2] \to [1,5]$ by

$$\varphi(x) \quad := \quad x^2 + 1$$

Then φ is continuously differentiable, and $\varphi'(x) = 2x$, which is a linear function, which we already saw was integrable (Example 9.11). Also note that

$$g(x) \quad = \quad \varphi'(x)\sqrt{\varphi(x)}$$

and so applying the change of variables theorem with $f : [1,5] \to \mathbb{R}$ defined via $f(x) := \sqrt{x}$ for all $x \in [1,5]$ we find that

$$\int_0^2 2x\sqrt{x^2+1}\mathrm{d}x \quad = \quad \int_1^5 \sqrt{x}\mathrm{d}x\,.$$

The point being, it is *much* easier to integrable $x \mapsto \sqrt{x}$ than $x \mapsto 2x\sqrt{x^2+1}$.

The key here was to define φ and to observe that $g = \varphi'\sqrt{\varphi}$.

Now using Remark 9.34 we find that

$$\begin{aligned} \int_1^5 \sqrt{x}\mathrm{d}x \quad &= \quad \frac{2}{3}x^{\frac{3}{2}}\bigg|_1^5 \\ &= \quad \frac{2}{3}\left(5^{\frac{3}{2}} - 1\right)\,. \end{aligned}$$

9.37 Example. Let us consider the integral

$$\int_0^4 \frac{x}{(1+x^2)^2}\mathrm{d}x.$$

Let us define $\varphi(x) := 1 + x^2$, in which case $\varphi'(x) = 2x$ so that if $f(x) := \frac{1}{2x^2}$ then

$$\begin{aligned} \frac{x}{(1+x^2)} \quad &= \quad \frac{1}{2}\frac{\varphi'(x)}{\varphi(x)^2} \\ &= \quad f(\varphi(x))\,\varphi'(x) \end{aligned}$$

then the change of variables Theorem 9.35 says

$$\begin{aligned} \int_0^4 \frac{x}{(1+x^2)^2}\mathrm{d}x \quad &= \quad \int_0^4 \varphi'(f \circ \varphi) \\ &= \quad \int_{\varphi(0)}^{\varphi(4)} f \\ &= \quad \int_1^{17} \frac{1}{2x^2}\mathrm{d}x \\ &= \quad -\frac{1}{2}x^{-1}\bigg|_{x=1}^{x=17} \\ &= \quad \frac{1}{2}\left(1 - \frac{1}{17}\right)\,. \end{aligned}$$

9.38 Example. Consider

$$\int_0^4 \frac{1}{1+x^2}\mathrm{d}x\,.$$

Define $\varphi(x) := \tan(x)$. Then $\varphi'(x) = \frac{1}{\cos(x)^2}$ and so if $f(x) := \frac{1}{1+x^2}$ then change of variables says

$$\begin{aligned}\int_0^4 \frac{1}{1+x^2}\mathrm{d}x &= \int_0^4 f \\ &= \int_{\varphi(\varphi^{-1}(0))}^{\varphi\left(\varphi^{-1}(4)\right)} f \\ &= \int_{\varphi^{-1}(0)}^{\varphi^{-1}(4)} (f\circ\varphi)\,\varphi' \\ &= \int_{\varphi^{-1}(0)}^{\varphi^{-1}(4)} \frac{1}{1+\tan^2}\frac{1}{\cos^2}\,.\end{aligned}$$

But now, $\frac{1}{1+\tan^2} = \frac{1}{1+\frac{\sin^2}{\cos^2}} = \frac{\cos^2}{\cos^2+\sin^2} = \cos^2$ and hence we find

$$\frac{1}{1+\tan^2}\frac{1}{\cos^2} = 1\,.$$

Since $\int_c^d = d-c$ we have

$$\int_0^4 \frac{1}{1+x^2}\mathrm{d}x = \varphi^{-1}(4) - \varphi^{-1}(0)$$

But $\varphi^{-1} \equiv \tan^{-1} \equiv \arctan^{-1}$. So the integral equals

$$\int_0^4 \frac{1}{1+x^2}\mathrm{d}x = \arctan(4) - \arctan(0)\,.$$

9.39 Remark. By the way, Example 9.38 raises an interesting point: since

$$\begin{aligned}\lim_{x\to\frac{\pi}{2}}\tan(x) &= \lim_{x\to\frac{\pi}{2}}\frac{\sin(x)}{\cos(x)} \\ &= \infty\end{aligned}$$

we must have (via the definition $\arctan \equiv \tan^{-1}$)

$$\lim_{x\to\infty}\arctan(x) = \frac{\pi}{2}$$

and similarly

$$\lim_{x\to-\infty}\arctan(x) = -\frac{\pi}{2}.$$

Hence it makes sense to think of

$$\int_a^b \frac{1}{1+x^2}\mathrm{d}x \quad = \quad \arctan(b) - \arctan(a)$$

with $a \to -\infty$ and $b \to +\infty$, and the result is called *an improper integral*. I.e.,

$$\begin{aligned} \lim_{a\to-\infty}\lim_{b\to\infty}\int_a^b \frac{1}{1+x^2}\mathrm{d}x &= \lim_{b\to\infty}\arctan(b) - \lim_{a\to-\infty}\arctan(a) \\ &= \frac{\pi}{2} - \left(-\frac{\pi}{2}\right) \\ &= \pi\,. \end{aligned}$$

One writes this in compact form as

$$\int_{-\infty}^{\infty} \frac{1}{1+x^2}\mathrm{d}x \quad = \quad \pi\,.$$

Since this result involves a limit of the end points a and b, it is called *an improper integral*.

9.40 Definition. An improper integral is the result of the limit (if it exists) after integrating

$$\lim_{a\to a_0}\lim_{b\to b_0}\int_a^b f\,,$$

where $a_0 \in \mathbb{R}$ or $a_0 = -\infty$ and $b_0 \in \mathbb{R}$ or $b_0 = \infty$.

9.41 Theorem. *(Integration by parts) If $f, g : [a, b] \to \mathbb{R}$ are differentiable such that f', g' are integrable, then*

$$\int_a^b fg' \quad = \quad f(b)\,g(b) - f(a)\,g(a) - \int_a^b f'g\,.$$

Proof. Using Claim 8.14, we have

$$(fg)' \quad = \quad f'g + fg'\,.$$

But Theorem 9.32 says that

$$\int_a^b (fg)' \quad = \quad fg|_a^b\,.$$

We learn that

$$\int_a^b f'g + \int_a^b fg' \quad = \quad fg|_a^b\,.$$

which is essentially the result. □

9.42 Example. Consider the function $\mathbb{R} \ni x \mapsto x\sin(x) \in \mathbb{R}$. Let us define $f(x) := x$ (with the same domains and co-domains). Then

$$\begin{aligned}
\int_a^b x\sin(x)\,\mathrm{d}x &= \int_a^b f\sin \\
&= -\int_a^b f\cos' \\
&\quad \text{(integration by parts)} \\
&= -f\cos|_a^b + \int_a^b f'\cos \\
&\quad \text{(Use } f' = 1\text{)} \\
&= -f\cos|_a^b + \int_a^b \cos \\
&= -f\cos|_a^b + \int_a^b \sin' \\
&\quad \text{(Use fundamental thm. of calc.)} \\
&= -f\cos + \sin|_a^b \,.
\end{aligned}$$

The point is we know how to integrate trigonometric functions, but we don't know how to integrate $x \mapsto x\sin(x)$, and that's when integration by parts could help.

9.43 Example. [Courant] We have

$$\begin{aligned}
\int_a^b x\mathrm{e}^x\mathrm{d}x &= \int_a^b x\,(x \mapsto \mathrm{e}^x)'\,\mathrm{d}x \\
&= x\mathrm{e}^x|_a^b - \int_a^b \mathrm{e}^x\mathrm{d}x \\
&= \mathrm{e}^x\,(x-1)|_a^b \,.
\end{aligned}$$

9.44 Example. More generally,

$$\int_a^b xf'(x)\,\mathrm{d}x = xf(x)|_a^b - \int_a^b f\,.$$

9.45 Example. [Courant] We have

$$\begin{aligned}
\int_a^b \arcsin &= \int_a^b \arcsin\,(x \mapsto x)' \\
&= x\arcsin(x)|_a^b - \int_a^b x\arcsin' \\
&= x\arcsin(x)|_a^b - \int_a^b \frac{x}{\sqrt{1-x^2}}\mathrm{d}x
\end{aligned}$$

for this last integral we do a change of variables with $\varphi(x) := 1 - x^2$ so that $\varphi'(x) = -2x$ and hence with $f(x) := \frac{1}{\sqrt{x}}$ we find

$$\begin{aligned}
\int_a^b \frac{x}{\sqrt{1-x^2}}\mathrm{d}x &= \int_a^b \frac{1}{-2}\varphi'(x) f(\varphi(x))\,\mathrm{d}x \\
&= -\frac{1}{2}\int_{\varphi(a)}^{\varphi(b)} f \\
&= -\frac{1}{2}\, 2\sqrt{x}\Big|_{\varphi(a)}^{\varphi(b)} .
\end{aligned}$$

9.46 Example. [Courant] Consider $\int_a^b \mathrm{e}^{\alpha x}\sin(\beta x)\,\mathrm{d}x$. In this example repeated integration by parts will result in an algebraic equation:

$$\begin{aligned}
\int_a^b \mathrm{e}^{\alpha x}\sin(\beta x)\,\mathrm{d}x &= -\int_a^b \mathrm{e}^{\alpha x}\left(x\mapsto \frac{1}{\beta}\cos(\beta x)\right)'\mathrm{d}x \\
&= -\mathrm{e}^{\alpha x}\frac{1}{\beta}\cos(\beta x)\Big|_a^b + \int_a^b \alpha\mathrm{e}^{\alpha x}\frac{1}{\beta}\cos(\beta x)\,\mathrm{d}x \\
&= -\mathrm{e}^{\alpha x}\frac{1}{\beta}\cos(\beta x)\Big|_a^b + \frac{\alpha}{\beta}\int_a^b \mathrm{e}^{\alpha x}\left(x\mapsto \frac{1}{\beta}\sin(\beta x)\right)'\mathrm{d}x \\
&= -\mathrm{e}^{\alpha x}\frac{1}{\beta}\cos(\beta x)\Big|_a^b + \frac{\alpha}{\beta}\left(\mathrm{e}^{\alpha x}\frac{1}{\beta}\sin(\beta x)\Big|_a^b - \frac{\alpha}{\beta}\int_a^b \mathrm{e}^{\alpha x}\sin(\beta x)\,\mathrm{d}x\right) \\
&= \left[-\frac{1}{\beta}\mathrm{e}^{\alpha x}\cos(\beta x) + \frac{\alpha}{\beta^2}\mathrm{e}^{\alpha x}\sin(\beta x)\right]\Bigg|_a^b - \frac{\alpha^2}{\beta^2}\int_a^b \mathrm{e}^{\alpha x}\sin(\beta x)\,\mathrm{d}x\,.
\end{aligned}$$

We solve this for $\int_a^b \mathrm{e}^{\alpha x}\sin(\beta x)\,\mathrm{d}x$ to get

$$\begin{aligned}
\int_a^b \mathrm{e}^{\alpha x}\sin(\beta x)\,\mathrm{d}x &= \frac{1}{1+\frac{\alpha^2}{\beta^2}}\left[-\frac{1}{\beta}\mathrm{e}^{\alpha x}\cos(\beta x) + \frac{\alpha}{\beta^2}\mathrm{e}^{\alpha x}\sin(\beta x)\right]\Bigg|_a^b \\
&= \frac{1}{\beta^2+\alpha^2}\,\mathrm{e}^{\alpha x}\left[\alpha\sin(\beta x) - \beta\cos(\beta x)\right]\Big|_a^b .
\end{aligned}$$

9.47 Remark. We summarize that our main tools to evaluate integrals is combine Remark 9.34 together with Theorem 9.23, Theorem 9.35 and Theorem 9.41. This is not a lot, and indeed most functions which can be integrated don't admit an explicit formula for their integral.

10 Important functions

10.1 The trigonometric functions

Recall the definitions and properties of sin, cos, tan, cot etcetera discussed in Section 5.1.2.

- The tangent function is defined as the quotient $\tan \equiv \frac{\sin}{\cos}$ whenever cosine is non-zero (so one must restrict its domain of definition).
- The cotangent function is defined as the quotient $\cot \equiv \frac{\cos}{\sin}$ whenever sine is non-zero (so one must restrict its domain of definition).
- These two definitions mean that one must mainly keep in mine sin and cos whereas properties of tan and cot may be inferred by the quotient definition.
- There is a special (irrational) number, denoted by π, equal to approximately 3.1415. Geometrically it is the ratio of a circle's circumference to its diameter. It is also a convenient way to measure arc lengths on the circle of radius 1 for that reason: an arc-length of 2π is the entire circle, π is half the circle, $\frac{\pi}{2}$ is one-quarter of it, etc. Naturally the trigonometric functions, which related to arc-lengths of the unit circle, have special values corresponding to special multiples of π:
  - $\sin(n\pi) = 0$ for all $n \in \mathbb{Z}$.
  - $\cos(n\pi) = (-1)^n$ for all $n \in \mathbb{Z}$.
  - $\sin\left(n\frac{\pi}{2}\right) = -(-1)^n$ for all $n \in \mathbb{Z}$.
  - $\cos\left(n\frac{\pi}{2}\right) = 0$ for all $n \in \mathbb{Z}$.
- The trigonometric functions are *periodic* of period 2π:

$$\begin{aligned} \cos(x+2\pi) &= \cos(x) \qquad (x \in \mathbb{R}) \\ \sin(x+2\pi) &= \sin(x) \qquad (x \in \mathbb{R}) \end{aligned}$$

- The sine and cosine are related by a *shift* of the angle in $\frac{\pi}{2}$:

$$\sin(x) = \cos\left(x - \frac{\pi}{2}\right) \qquad (x \in \mathbb{R})$$

- They obey the pythagoras law:

$$\cos(x)^2 + \sin(x)^2 = 1 \qquad (x \in \mathbb{R})$$

- Their image is in $[-1, 1]$ and they are continuous throughout their domain $\mathbb{R}$. They do not have limits at $\pm\infty$ as they keep oscillating.

10.2 The exponential and logarithmic functions

Recall the definitions and properties of $\exp_a$ and $\log_a$ for $a > 1$ discussed in Example 5.31:

1. $\log_a$ is defined on $(0, \infty)$ and has as its image the whole of $\mathbb{R}$. It is strictly monotone increasing. It is a continuous function.

2. $\log_a(1) = 0$.
3. $\log_a(a) = 1$.
4. $\lim_{x\to 0} \log_a(x) = -\infty$.
5. $\lim_{x\to\infty} \log_a(x) = +\infty$.
6. $\log_a(xy) = \log_a(x) + \log_a(y)$.
7. $\log_a\left(\frac{x}{y}\right) = \log_a(x) - \log_a(y)$.
8. $\log_a(x^p) = p \log_a(x)$ for all $p \in \mathbb{R}$.
9. $\log_a(x) \leq C_a(x-1)$ for some strictly positive constant (independent of x) C_a.
10. $\exp_a$ is defined on $\mathbb{R}$, and has as its image $(0, \infty)$ (so it is always strictly positive). It is strictly monotone increasing. It is a continuous function.
11. $\exp_a(0) = 1$.
12. $\lim_{x\to-\infty} \exp_a(x) = 0$.
13. $\lim_{x\to\infty} \exp_a(x) = \infty$.
14. $\exp_a(1) = a$.
15. $\exp_a(x+y) = \exp_a(x)\exp_a(y)$.
16. $\exp_a(x-y) = \frac{\exp_a(x)}{\exp_a(y)}$.
17. $\exp_a(px) = (\exp_a(x))^p$ for all $p \in \mathbb{R}$.

- In the context of evaluating limits, it is useful to know that

$$\begin{aligned} x &= \exp_a(\log_a(x)) \qquad (x > 0) \\ x &= \log_a(\exp_a(x)) \qquad (x \in \mathbb{R}) \end{aligned}$$

10.2.1 The natural base for the logarithm

10.1 Claim. The limit $\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$ exists and equals some number between 2 and 3.

Proof. We have by the binomial theorem (see below)

$$\begin{aligned}
\left(1+\frac{1}{n}\right)^n &= \sum_{j=0}^{n} \frac{n!}{j!\,(n-j)!}\frac{1}{n^j} \\
&= \sum_{j=0}^{n} \frac{n\,(n-1)\cdots(n-j+1)\,(n-j)!}{j!\,(n-j)!}\frac{1}{n^j} \\
&= \sum_{j=0}^{n} \frac{1}{j!}\frac{n\,(n-1)\cdots(n-j+1)}{n^j} \\
&= \sum_{j=0}^{n} \frac{1}{j!}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{j-1}{n}\right) \\
&= 1+1+\frac{1}{2!}\left(1-\frac{1}{n}\right)+\frac{1}{3!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)+\cdots+\frac{1}{n!}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{n-1}{n}\right)
\end{aligned}$$

So that

$$\begin{aligned}
&\left(1+\frac{1}{n}\right)^n-\left(1+\frac{1}{n+1}\right)^{n+1} \\
= {} & \sum_{j=0}^{n}\frac{1}{j!}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{j-1}{n}\right)-\sum_{j=0}^{n+1}\frac{1}{j!}\left(1-\frac{1}{n+1}\right)\cdots\left(1-\frac{j-1}{n+1}\right) \\
= {} & \sum_{j=0}^{n}\frac{1}{j!}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{j-1}{n}\right)-\sum_{j=0}^{n}\frac{1}{j!}\left(1-\frac{1}{n+1}\right)\cdots\left(1-\frac{j-1}{n+1}\right)- \\
& -\frac{1}{(n+1)!}\left(1-\frac{1}{n+1}\right)\cdots\left(1-\frac{n}{n+1}\right) \\
= {} & \sum_{j=0}^{n}\frac{1}{j!}\left(\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{j-1}{n}\right)-\left(1-\frac{1}{n+1}\right)\cdots\left(1-\frac{j-1}{n+1}\right)\right)- \\
& -\frac{1}{(n+1)!}\left(1-\frac{1}{n+1}\right)\cdots\left(1-\frac{n}{n+1}\right) \\
& \left(\text{Using } \frac{1}{n+1}\le\frac{1}{n}\right) \\
\le {} & \sum_{j=0}^{n}\frac{1}{j!}\left(\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{j-1}{n}\right)-\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{j-1}{n}\right)\right)- \\
& -\frac{1}{(n+1)!}\left(1-\frac{1}{n+1}\right)\cdots\left(1-\frac{n}{n+1}\right) \\
= {} & -\frac{1}{(n+1)!}\left(1-\frac{1}{n+1}\right)\cdots\left(1-\frac{n}{n+1}\right) \\
\le {} & 0
\end{aligned}$$

This shows that the sequence $\mathbb{N} \ni n \mapsto \left(1 + \frac{1}{n}\right)^n$ is increasing.

It is also bounded, since

$$\begin{aligned}\left(1+\frac{1}{n}\right)^n &= \sum_{j=0}^{n} \frac{1}{j!}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{j-1}{n}\right) \\ &\quad \left(\frac{j-1}{n} \geq \frac{1}{2} \text{ for all } j\right) \\ &\leq 1+\sum_{j=0}^{n}\frac{1}{2^j}\end{aligned}$$

The latter sequence, $\mathbb{N} \ni n \mapsto 1 + \sum_{j=0}^{n} \frac{1}{2^j}$ actually converges to 3. Being monotone increasing and bounded, $\mathbb{N} \ni n \mapsto \left(1 + \frac{1}{n}\right)^n$ converges by Claim 6.17. Since it is increasing and the first term is larger than 2, we find that

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \text{ exists and } \in (2,3)$$

□

10.2 Definition. We define $\mathrm{e} := \lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n$ and call it *the natural base for the logarithm.* It turns out that $\mathrm{e} \approx 2.718$ and is irrational. When we don't write the subscript a for $\exp_a$ or $\log_a$, we mean $a = \mathrm{e}$. The reason why it is called the natural base for the logarithm is because its derivative is given by

$$\log'(x) = \frac{1}{x}$$

and also that

$$\log(x) \leq x - 1$$

(i.e. the constant $C_{\mathrm{e}} = 1$).

10.3 Claim. We have $\log(x) \leq x - 1$ for any $x \in (0, \infty)$.

Proof. Let us first assume that $x > 1$. Then we may apply the mean value theorem Theorem 8.48 to get that there must be some $y \in (1, x)$ such that

$$\log'(y) = \frac{\log(x) - \log(1)}{x-1}$$

Now

$$\log'(y) = \frac{1}{y}$$

and $\log(1) = 0$, so we find

$$\frac{1}{y} = \frac{\log(x)}{x-1}$$

Since $y \in (1, x)$, $y > 1$ so $\frac{1}{y} < 1$, and so we find

$$1 < \frac{\log(x)}{x-1}$$

moving the denominator to the other side of the inequality we get the result.

When $x < 1$, we have again by an application of Theorem 8.48 that there is some $y \in (x, 1)$ such that

$$\frac{1}{y} = \frac{\log(1) - \log(x)}{1-x}$$

However now $\frac{1}{y} > 1$ so that

$$1 > \frac{-\log(x)}{1-x}$$

and we find the same result.

If $x = 1$, then $\log(x) = 0$ and $x - 1 = 0$ so the inequality is actually an equality. □

10.3 The hyperbolic functions

Recall from Example 8.23 where we introduced the hyperbolic trigonometric functions, which are analogous to the trigonometric functions. This follows from the Euler identities

$$\sin(x) = \frac{1}{2\mathrm{i}}\left(\mathrm{e}^{\mathrm{i}x} - \mathrm{e}^{-\mathrm{i}x}\right)$$

and

$$\cos(x) = \frac{1}{2}\left(\mathrm{e}^{\mathrm{i}x} + \mathrm{e}^{-\mathrm{i}x}\right)$$

(If that means nothing to you because of the i, then ignore this short introduction). By analogy we define, with $\mathbb{R}$ being the domain and codomain,

$$\sinh(x) := \frac{1}{2}\left(\mathrm{e}^{x} - \mathrm{e}^{-x}\right)$$

and

$$\cosh(x) := \frac{1}{2}\left(\mathrm{e}^{x} + \mathrm{e}^{-x}\right)$$

10.4 Special functions

10.4.1 The sinc function

The sinc function, which is discussed in Example 6.32, is defined as

$$\mathbb{R} \ni x \mapsto \operatorname{sinc}(x) \quad := \quad \begin{cases} 1 & x = 0 \\ \frac{\sin(x)}{x} & x \neq 0 \end{cases} \in \mathbb{R}$$

The example cited shows that this function is continuous at the origin. It converges to zero at $\pm\infty$.

10.4.2 The square root

The square-root function $\sqrt{\cdot} : [0, \infty) \to [0, \infty)$ is continuous, positive, and monotone increasing. Furthermore it obeys

$$\sqrt{x+y} \quad \leq \quad \sqrt{x} + \sqrt{y} \qquad (x, y \geq 0)$$

11 Useful algebraic formulas to recall

11.1 Factorizations

11.1 Claim. Newton's binomial theorem: If $a, b \in \mathbb{R}$ and $n \in \mathbb{N}$ then

$$(a+b)^n \quad = \quad \sum_{j=0}^{n} \frac{n!}{j!\,(n-j)!} a^j b^{n-j}$$

11.2 Claim. If $a, b \in \mathbb{R}$ then

$$a^2 - b^2 \quad = \quad (a-b)(a+b)$$

In fact, we have the more general formula for $n \in \mathbb{N}$:

$$a^n - b^n \quad = \quad (a-b) \sum_{k=0}^{n-1} a^{n-k-1} b^k$$

which is proven in Example 6.34.

These formulas are especially useful when one is dealing with square roots, e.g.

$$\sqrt{a} - \sqrt{b} \quad = \quad \frac{a-b}{\sqrt{a} + \sqrt{b}}\,.$$

11.3 Claim. Recall if $ax^2 + bx + c = 0$ for some $a, b, c \in \mathbb{R}$ then the two solutions $x_1, x_2 \in \mathbb{R}$ can be written as

$$x_{1,2} \quad = \quad \frac{1}{2a}\left(-b \pm \sqrt{b^2 - 4ac}\right)$$

given that $b^2 - 4ac > 0$. If $b^2 - 4ac = 0$ then there is only one solution. If $b^2 - 4ac < 0$ then there are no solutions. If there are two solutions, then we may factorize

$$ax^2 + bx + c = \left(x - \frac{1}{2a}\left(-b - \sqrt{b^2 - 4ac}\right)\right)\left(x - \frac{1}{2a}\left(-b + \sqrt{b^2 - 4ac}\right)\right)$$

and if there is only one solution then we factorize

$$ax^2 + bx + c = \left(x - \frac{1}{2a}(-b)\right)^2$$

11.4 Claim. We have $\sum_{n=1}^{N} n = \frac{1}{2}N(N+1)$ for all $N \in \mathbb{N}$.

11.2 Inequalities

- If $a < b$ and $a, b > 0$ then $\frac{1}{a} > \frac{1}{b}$.
- If $a < b$ then $-a > -b$.

First of all it is important to note that, as in Claim 6.1

11.5 Claim. For any $\varepsilon > 0$, the inequality (for the unknown $x \in \mathbb{R}$)

$$|x| < \varepsilon$$

is equivalent to the *two* simultaneous inequalities (for the unknown $x \in \mathbb{R}$)

$$-\varepsilon < x < \varepsilon$$

which is sometimes useful in algebraic manipulations of limits.

11.6 Claim. The inequality $ax^2 + bx + c < \alpha$ for some $a, b, c, \alpha \in \mathbb{R}$ has to be solved as follows: If $a > 0$, this is an upwards parabola. Hence to be smaller than some value (α), the unknown x must be constrained to the interior between two points x_1, x_2. These two points are given by the solution to the equation $ax^2 + bx + c - \alpha = 0$, assuming $b^2 - 4a(c - \alpha) > 0$, so that there are two unique solutions. If on the other hand $a < 0$, this is a downwards parabola, and so the unknown x must be constrained to the left or right of x_1 and x_2 respectively, assuming $x_1 < x_2$.

If $x_1 = x_2$ more care has to be taken: If $a > 0$, then any x solves the inequality and if $a < 0$ then no x solves the inequality.

11.7 Claim. If f is monotone increasing and $a < b$ then $f(a) \leq f(b)$. If f is strictly monotone increasing then $f(a) < f(b)$. If f is monotone decreasing then $f(a) \geq f(b)$ and if f is strictly monotone decreasing then $f(a) > f(b)$.

This can be useful, for example, the function $x \mapsto x^2$ is monotone increasing for $x > 0$, so that if $0 < a < b$ then $a^2 < b^2$.

12 Dictionary of graphical symbols and acronyms

12.1 Graphical symbols

1. The equivalent symbol, $\equiv$, means the following: if a and b are any two objects, then $a \equiv b$ means we have agreed, at some earlier point, that a and b are two different labels for one and the same thing. Example: $\{ a, a \} \equiv \{ a \}$.

2. The definition symbol, $:=$, means the following: if a and b are any two objects, then $a := b$ means that right now, through this very equation, we are agreeing that a is a new label for the pre-existing object b. The main difference to $\equiv$ is about when this agreement happens: The $\equiv$ symbol is a reminder about our previous conventions, whereas the $:=$ symbol is an event of establishing a new convention. These both should be contrasted with the equal sign $=$, which merely says that two things (turn out) to be equal, whether by convention or not is not specified.

3. ∞ means the size of a set whose number of elements is unbounded.

4. $x \mapsto |x|$ means the absolute value function, i.e. $|x| \equiv \begin{cases} x & x \geq 0 \\ -x & x < 0 \end{cases}$.

5. The maximum or minimum of a set of numbers is the largest element of that set. Note this only works when we take the maximum or minimum or elements in $\mathbb{R}$ (or its subsets)!

$$\max (\{ 1, 2, 100 \}) = 100$$

$$\min (\{ 1, 2, 100 \}) = 1$$

6. Sometimes it is useful to write out the sum of numbers in a compact way. The graphical symbol for that is the capital Greek letter Sigma, written as $\sum$. We use it as follows: Let us assume we have a sequence $a : \mathbb{N} \to \mathbb{R}$ and $S \subseteq \mathbb{N}$ is given. For example $S = \{ n_1, n_2, \dots, n_N \}$. Then

$$\sum_{n \in S} a(n) \equiv a(n_1) + a(n_2) + \dots + a(n_N)$$

For instance if $S = \{ 1, 2, \dots, N \}$ then we usually write

$$\sum_{n=1}^{N} a(n) \equiv a(1) + a(2) + \dots + a(N)$$

7. The factorial of a number $n!$ is a short-hand notation for the following arithmetic operation:

$$n! \equiv n(n-1)(n-2)\dots 2$$

so that

$$\begin{aligned} 2! &= 2 \\ 3! &= 3 \cdot 2 = 6 \\ 4! &= 4 \cdot 3 \cdot 2 = 24 \\ 5! &= 5 \cdot 4 \cdot 3 \cdot 2 = 120 \\ &\cdots \end{aligned}$$

From this we see immediately that $n! = n\,(n-1)!$ for any $n \in \mathbb{N}$.

12.2 Acronyms

1. s.t. means "such that".
2. w.r.t. means "with respect to".
3. l.h.s. means "left hand side", usually of an equation. r.h.s. means "right hand side".
4. iff means "if and only if", which is a relationship between two statements, meaning the first implies the second and the second implies the first.
5. "The origin" means either $0 \in \mathbb{R}$ or $(0,0) \in \mathbb{R}^2$.
6. WLOG means "without loss of generality".

References

[1] Tom M. Apostol. *Calculus, Vol. 1: One-Variable Calculus, with an Introduction to Linear Algebra*. Wiley, 1991.

[2] Richard Courant and Fritz John. *Introduction to Calculus and Analysis, Vol. 1 (Classics in Mathematics)*. Springer, 1998.

[3] Paul R. Halmos. *Naive Set Theory*. Martino Fine Books, 2011.

[4] Walter Rudin. *Principles of Mathematical Analysis (International Series in Pure and Applied Mathematics)*. McGraw-Hill Education, 1976.

[5] Michael Spivak. *Calculus, 4th edition*. Publish or Perish, 2008.

[6] James Stewart. *Calculus: Early Transcendentals 8th Edition*. Cengage Learning, 2015.