Preliminary draft

Calculus Ia: Limits and differentiation

(BSMA1002)

Yichao Huang

March 19, 2021

University of Helsinki

Preliminary draft

Disclaimer
Any errors that remain are the author's sole responsibility.

Colophon
This document was typeset with the help of KOMA-Script and LaTeX using the kaobook class.

Publisher
First printed in Fall 2020 at the University of Helsinki

Don't just read it; fight it! Ask your own question, look for your own examples, discover your own proofs. Is the hypothesis necessary? Is the converse true? What happens in the classical special case? What about the degenerate cases? Where does the proof use the hypothesis?

– Paul Halmos

Preface

This lecture note is as boring as it gets, since it tries to be politically correct and does not contain:

1. Your own effort in trying new things in mathematics;
2. Your own taste of what is beautiful and what is not;
3. Your happiness in understanding a concept or finding a proof;
4. Your failures and experiences to refine your future choices;
5. Your interaction and teamwork with your friends;
6. Your lunch hours, roadtrips, family, dreams, and drunken moments (some by alcohol, some by art, some by people);
7. Your splendid life with all the possibilities ahead.

The goal of this lecture note is to simply provide some remainders in case one needs them. Like a photo album. In some sense, the primary goal would be for you to understand some mathematical concepts, and gradually you should be able to express your ideas in mathematical terms with ease.

It is often asked about a reference book for this course. I don't want to recommend anything in particular, but I would recommend to have at least one "classical" textbook at hand, preferably with detailed solutions to the exercises. Try the exercise yourself first, and when you really get stuck, read the solution, then try the exercise again several days later.

Instead I could recommand some "casual" books:

1. «Gödel, Escher, Bach: An Eternal Golden Braid», Hofstadter, Basic Books.
2. «Proofs from THE BOOK», Aigner-Ziegler, Springer.[1]
3. «How to solve it», Pólya, Princeton University Press.
4. «Flatland: A Romance of Many Dimensions», Abbott, Seeley & Co.

1: This one is not that casual...!

Have fun!

Yichao Huang

P.S. Although this note can be publicly distributed and reused, it is not my intention. It thus contains many personal touches and certainly does not stand the test of time.

Contents

A "gentle" introduction to the language of mathematics | 1

Since this notes is written during the Corona time, let us start by reviewing some common misunderstandings.

Do the following sentences convey the same message?[1]

1. The government has no recommandation for wearing masks.
2. The government does not recommand wearing masks.
3. The government recommand against wearing masks.

1: Does "The government now recommands wearing masks" implies "The government recommanded against wearing masks before"?

1.1 Mathematical symbols

Some symbols are specific to mathematics, such as

$\forall$	for all
$\exists$	there exist(s) [...] (such that)
$\vee$	or
$\wedge$	and
∞	infinity

and many more.[2]

2: In short, this course BSMA1002 is about the symbol $'$ and the next course BSMA1003 is about the symbol $\int$. There are subtleties that we don't want to go into in this course, for example for the symbol "=", it can refer to an "equality" or an "equation" depending on the context. Apparently in French the distinction is made, but not in English.

One uses these symbols to write statements in mathematics. For example, one can write

$$\forall x \in \mathbb{R}, (x^2 - 1 \geq 0) \vee (x^3 + 1 \geq 0).$$

This reads (from left to right!) "for all real number x, (we have) $x^2 - 1 \geq 0$ or $x^3 + 1 \geq 0$". In practice, the symbol $\vee$ is not that often used, and one encounters more often

$$\forall x \in \mathbb{R}, (x^2 - 1 \geq 0) \text{ or } (x^3 + 1 \geq 0).$$

Notice that the meaning of the word "or" is inexclusive, meaning that both alternatives can be true at the same time.[3] Also, notice that a statement does not have to be true, unfortunately it seems that my random example above is true...

3: **Exclusive or** or **exclusive disjunction** is a logical operation that outputs true only when inputs differ (one is true, the other is false). In logic, **or** by itself means the **inclusive or**, distinguished from an exclusive or, which is false when both of its arguments are true, while an "or" is true in that case. In sum, $A \vee B$ is true if A is true, or if B is true, or if both A and B are true.

One can "operate" on statements. For example, with the **negation** symbol

$$\neg \quad \text{not}$$

one can "negate" a statement:

$$\neg\left(\forall x \in \mathbb{R}, (x^2 - 1 \geq 0) \text{ or } (x^3 + 1 \geq 0)\right).$$

What does it mean? How do one write it in plain language?

(Before moving on, one could first to come up with a personal attempt. The goal is not to succeed at the first try, but to, *inter alia*, figure out some patterns and be aware of the possible difficulties.)

The general rule for negating a statement is the following: for any statements P and Q,[4]

4: Don't try to remember these sentences, but rather, do some examples and **understand the principle behind it**.

1. $\neg(P \vee Q)$ is $\neg P \wedge \neg Q$;
2. $\neg(P \wedge Q)$ is $\neg P \vee \neg Q$;
3. $\neg(\forall x, P)$ is $\exists x, \neg P$;
4. $\neg(\exists x, P)$ is $\forall x, \neg P$.

(Say these phrases with a less obscure language!)[5]

5: The first two principles are variants of the so-called “De Morgan's laws”.

For the example above, an equivalent way of writing the statement

$$\neg\left(\forall x \in \mathbb{R}, (x^2 - 1 \geq 0) \text{ or } (x^3 + 1 \geq 0)\right)$$

is

$$\exists x \in \mathbb{R}, \neg(x^2 - 1 \geq 0) \text{ and } \neg(x^3 + 1 \geq 0).$$

And if we really want to get rid of the negation symbol, we can also write it as

$$\exists x \in \mathbb{R}, (x^2 - 1 < 0) \text{ and } (x^3 + 1 < 0).$$

Remark 1.1.1 In practice, this means that if P is some property and if one wants to **disprove** a statement of type “for all x, P is true” $(\forall x, P)$, one should show the existence of some x such that P is false $(\exists x, \neg P)$. Showing that such x exists can be done by explicit construction (“pulling a rabbit out of a hat”), or by abstraction (without necessarily knowing all the properties of such x).

Implications are highly frequent statements in mathematics.[6] If P and Q are two properties (or two statements), the symbol[7]

6: Which phrase is an implication in the classical **syllogism** “All men are mortal. Socrates is a man. Therefore, Socrates is mortal.”? Or actually, how many implications are there?

7: In some Finnish textbooks it is written $\longrightarrow$. Different people use different notations, but usually they look similar and understandable by context.

$$\Longrightarrow \quad \text{implies}$$

used in the following statement

$$P \Longrightarrow Q$$

means “If P is true, then Q is true”. It does not give information on Q if P is false.

One can draw a **truth table** to understand better the symbol $\Longrightarrow$. In the table, 1 means “true” and 0 means “false”, and I leave you to figure out the rest.[8]

8: Or one writes simple T and F for “true” and “false”. In real life, you will probably soon forget about this table.

P	Q	P⟹Q
1	1	1
1	0	0
0	1	1
0	0	1

It is quite useful to realize that the implication symbol can be replaced by other symbols before. At first sight this mights seem strange, but let us draw the table of truth for

$$(\neg P) \vee Q$$

and compare it to the table before:

P	Q	¬P	(¬P) ∨Q
1	1	0	1
1	0	0	0
0	1	1	1
0	0	1	1

Proposition 1.1.1 *The statement*

$$P \implies Q$$

is equivalent to

$$(\neg P) \vee Q.$$

The **equivalence** of two statements P and Q, with the symbol

$$P \equiv Q$$

should be understood as two implications:

$$P \implies Q \quad \text{and} \quad Q \implies P.$$

It simply says that they are either both true or both false.[9]

9: Try to draw the truth table (on a paper, on an electronic device or in your head)!

Proposition 1.1.2 *The following statements are equivalent:*[a]

1.

$$P \implies Q.$$

2.

$$(\neg P) \vee Q.$$

[a] Personal dedication to my undergrad teacher Mr. Mohan: "LASSE" (Les assertions suivantes sont équivalentes).

Proposition 1.1.3 *The following statements are equivalent:*[10]

1.

$$P \equiv Q.$$

2.

$$(P \implies Q) \wedge (Q \implies P).$$

It is a little self-referencing if you take it as the definition of equivalence!

10: So, what does "$((\neg P) \vee Q) \wedge ((\neg Q) \vee P)$" mean?

Proof by contradiction is also commonly used in mathematics (and in everyday life).[11] In practice, this often means the following steps:

1. Suppose the negation of what you are proving is true;
2. Use this information to deduce something that is known to be false;
3. Therefore, you have a contradiction (since "true" cannot imply "false"), and the original statement must be true.[12]

11: Proof by contradiction is formulated as $P \equiv P \vee \bot \equiv \neg(\neg P) \vee \bot \equiv \neg P \rightarrow \bot$, where $\bot$ is a logical contradiction or a false statement (a statement which true value is false). If $\bot$ is reached via $\neg P$ via a valid logic, then $\neg P \rightarrow \bot$ is proved as true so P is proved as true. I didn't bother to read the above phrases myself (I copied it from Wikipedia), since in practice, one should seize the **idea** (which I think you all have it naturally) rather than relying on formal manipulation of symbols. It is up to you to find out what is the best way to understand a new concept!

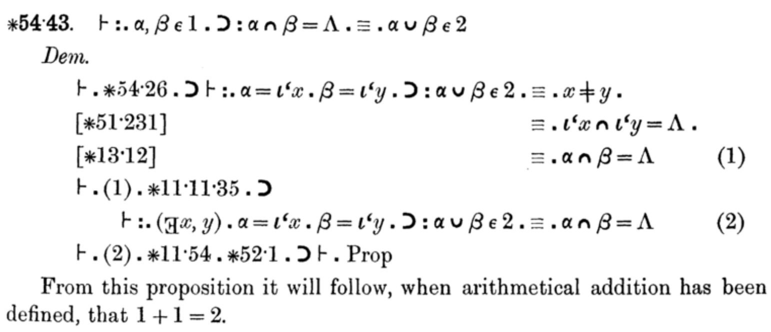
12: To be annoyingly precise, here we are assuming a basic axiom of logic called the law of noncontradiction.

Example 1.1.1 There is no smallest strictly positive real number.

Proof. Suppose the opposite and let $r > 0$ be the smallest strictly positive real number. But $r/2$ is a real number, $r/2$ is strictly smaller than r and $r/2$ is strictly positive. We have found a strictly positive real number smaller than r: contradiction. □

The above proof is very concise. In the beginning, you probably want to write a more detailed proof to make sure that it is correct and understandable. Now, as an exercise, can you write down the statement in the example with logical symbols? How would you write down its negation? What are we doing in the above proof?[13]

13: To be honest, I don't know the answer to these questions even though I wrote down the proof above: this is only because I have gathered enough experience, and interpreted subconsciously the principle in my own way.



$*54{\cdot}43.\ \vdash :.\ \alpha, \beta \in 1 . \supset : \alpha \cap \beta = \Lambda . \equiv . \alpha \cup \beta \in 2$

Dem.

$\vdash . *54{\cdot}26 . \supset \vdash :. \alpha = \iota`x . \beta = \iota`y . \supset : \alpha \cup \beta \in 2 . \equiv . x \neq y .$

$[*51{\cdot}231] \qquad \equiv . \iota`x \cap \iota`y = \Lambda .$

$[*13{\cdot}12] \qquad \equiv . \alpha \cap \beta = \Lambda \qquad (1)$

$\vdash . (1) . *11{\cdot}11{\cdot}35 . \supset$

$\vdash :. (\exists x, y) . \alpha = \iota`x . \beta = \iota`y . \supset : \alpha \cup \beta \in 2 . \equiv . \alpha \cap \beta = \Lambda \qquad (2)$

$\vdash . (2) . *11{\cdot}54 . *52{\cdot}1 . \supset \vdash . \text{Prop}$

From this proposition it will follow, when arithmetical addition has been defined, that $1 + 1 = 2$.

Figure 1.1: Whitehead and Russell proving $1 + 1 = 2$. Full story here.

Remark 1.1.2 Of course, mathematicians (or scientists) never write with symbols only, unless you are a hardcore logician. You will soon know where to draw the line: the above is just a showcase of the mathematical rigor.

One of the advantages of mathematics compared to other science, is that (almost) all proofs are reproducible and can be checked.[14] It is a good way to train your critical thinking skills: by doing mathematics (the right way!), you are living one of the rare moments where you can distinguish completely right from wrong and form a clear judgement.

14: Even this one.

1.2 Mathematical induction

One of the early difficulties of transitioning into a good undergrad student is to write mathematical sound and concise proofs. We have already seen what is proof by contradiction; let us review another classical proof technique: **proof by induction**.

Here is a learning technique: you can **start by an example before reading the theoretical descriptions**. So let us search “proof by induction” on the internet, go to the Wikipedia page, and check out the following example:

Example 1.2.1 (Sum of consecutive natural numbers) For any integer

$n \geq 0$, we have

$$0 + 1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

One can rewrite the sum using the symbol $\sum$:[15]

$$\sum_{i=0}^{n} i = 0 + 1 + 2 + \cdots + n.$$

15: The index i is called the dummy index; you can replace it with other symbols such as j or k and it only governs what happens inside the summation symbol.

A longer proof is the following. Rigorously speaking, the proof starts by defining a statement $P(n)$ for each interger $n \geq 0$:

$$P(n) : \quad 0 + 1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

For now we don't know if for a given interger n, $P(n)$ is true or not.

We then start by checking the **base case** (or "initialization"): in our case, that $P(0)$ is true. Notice that $n = 0$ is the smallest case possible. This is verified usually directly, i.e. by checking that

$$0 = \frac{0 \cdot 1}{2}.$$

Then the "inductive step" consists of checking the implication

$$P(n) \Longrightarrow P(n+1)$$

for all n greater or equal to the base case, in our case, $n \geq 0$. This means that we suppose $P(n)$ is true (this is called "induction hypothesis") for some $n \geq 0$ and from this, we show deduce that $P(n+1)$ is also true. So we suppose that $P(n)$ is true, i.e. we know that

$$0 + 1 + 2 + \cdots + n = \frac{n(n+1)}{2}$$

and we want to prove that $P(n+1)$ is true, i.e.

$$0 + 1 + 2 + \cdots + n + (n+1) = \frac{(n+1)(n+2)}{2}.$$

This follows by observing that

$$\begin{aligned}
&0 + 1 + 2 + \cdots + n + (n+1) \\
=&(0 + 1 + 2 + \cdots + n) + (n+1) \\
=&\frac{n(n+1)}{2} + (n+1) && \text{(induction hypothesis)} \\
=&\frac{n^2 + n + (2n+2)}{2} \\
=&\frac{(n+1)(n+2)}{2}.
\end{aligned}$$

In the above chain of equations, we have hightlighted the one where we used the assumption that $P(n)$ is true.

The conclusion is that, once we have checked the base case 0 and the implication $P(n) \Longrightarrow P(n+1)$ for all $n \geq 0$, we get that $P(1)$ is true (since

$P(0)$ is true and $P(0) \Longrightarrow P(1)$); and then $P(2)$ is true (since now $P(1)$ is true and $P(1) \Longrightarrow P(2)$); …; and that $P(n)$ is true for every $n \geq 0$. This argument is the principle of the mathematical induction.

Now in practice, the following (minimal writing of) proof is enough:

Proof. We prove the proposition by induction.

• Initialization: when $n = 0$, $0 = \frac{0 \cdot 1}{2}$.

• Induction: suppose that

$$0 + 1 + 2 + \cdots + n = \frac{n(n+1)}{2}$$

and prove that

$$0 + 1 + 2 + \cdots + n + (n+1) = \frac{(n+1)(n+2)}{2}.$$

By taking the difference of the above equations, it suffice to show that

$$n + 1 = \frac{(n+1)(n+2)}{2} - \frac{n(n+1)}{2} = \frac{(n+1)(n+2-n)}{2},$$

and this results from a direct algebraic calculation.

• Finally, by mathematical induction, the equation is true for all $n \in \mathbb{Z}_{\geq 0}$, thus the proposition. □

1.3 Story time: What the Tortoise Said to Achilles

See here.

1.4 Exercises

Exercise 1.1 Let $x, y \in \mathbb{R}$. Show that

$$|x + y| \leq |x| + |y|,$$

then

$$|x - y| \geq \left||x| - |y|\right|.$$

Give an interpretation of these inequalities by remembering that $|x - y|$ measures the **distance** between x and y.

Exercise 1.2 Show that the negation of

$$P \Longrightarrow Q$$

is

$$P \wedge (\neg Q).$$

Translate this exercise into human language.

Exercise 1.3 (Homework) Write the negation of the statement:

$$(\mathrm{P}) : \quad \forall \epsilon > 0, \exists \delta > 0, (|x-y| \leq \delta) \implies (||x| - |y|| \leq \epsilon).$$

Use the exercise above to determine if the statement P is true or false.[16]

16: Then, when you have time, take a (coffee) break and contemplate for a few minutes: what is this statement? You don't need to have a precise idea, but it is healthy to think about it.

Exercise 1.4 Let $x, y \in \mathbb{R}$. Show that

$$\max(x, y) = \frac{x+y}{2} + \frac{|x-y|}{2}.$$

Write a similar formula for $\min(x, y)$.

Exercise 1.5 Let $0 < q < 1$. Show by induction that

$$\sum_{k=0}^{n} q^k = 1 + q + q^2 + \cdots + q^n = \frac{1-q^{n+1}}{1-q}.$$

As an application, show that with the complex number i,[17]

17: The following expression is called the Dirichlet kernel.

$$\sum_{k=-n}^{n} e^{ikt} = \frac{\sin\left(\frac{2n+1}{2}t\right)}{\sin\left(\frac{1}{2}t\right)}.$$

Exercise 1.6 Let $n > 0$ be a positive integer. Show by induction that

$$\sum_{k=0}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

Exercise 1.7 ($*$) Show that for all integer $n \geq 4$,[18]

$$2^n < n! < n^n.$$

18: In 2021, you will learn to prove that

$$\sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} \leq n! \leq e n^{n+\frac{1}{2}} e^{-n}.$$

Exercise 1.8 ($*$) Prove that $\sqrt{2}$ is an irrational number.[19]

19: Don't hesitate to use a "canonical search engine" if you don't know what "irrational number" means.

Hint: you can start by supposing that $\sqrt{2} = \frac{p}{q}$ with p, q positive integers and try to deduce a contradiction, by studying the parity of p and of q.

Then calculate $\left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}}$ and show that there exist two irrational numbers a, b such that a^b is a rational number.

[Week I]
What are…sets?

Sets: definitions and properties | 2

A **set** is a *well-defined* collection of `distinct` OBJECTS.

What does that mean?

2.1 Some elements of axiomatic set theory

The axiomatic set theory starts with a list of **axioms**.[1] It is a rather complicated story to define the set theory proprement, so we assume that you already know the basics and this is more of a remainder. Actually, I must admit here that I do not have enough expertise to do a complete survey of even the basics of the axiomatic set theory, so the best we can do here is to learn by doing.

1: An axiom is a statement taken to be true; although you have to freedom to challenge it (and sometimes it could be hugely rewarding), by doing so you are basically isolating yourself from the majority of the scientific community.

Let us fix the objective here as to gradually get rid of the Venn diagram representation of sets, and start to write proper proofs; the actual study of the foundation of a theory of set would be another story. Recall some basic elements of the set theory:

Definition 2.1.1 (Elements of a set) *A **set** E is a collection of objects. For each of object x in E, we denote this inclusion by $x \in E$ and we call x an **element** of E.*

*We say a set E is **well-defined** when there is no ambiguity on which objects belong to E.[a] We also require the elements of a set to be **distinct**.[b]*

[a] For example, the collection "the world's greatest countries" is ill-defined.
[b] For example, $\{1, 1, 2\}$ is not a set, or that it is the same set as $\{1, 2\}$.

For a logician, the above definition is not precise enough, but in this course we don't require more than this.[2]

2: For a "full" introduction to a naive set theory, one can check out the book « Naive set theory » of Paul Halmos. The first chapter of this book introduces the **Axiom of extension**: two sets are equal if they have the same elements; in general Halmos is far more superior (both in language and in mathematics) than this lecture note.

Definition 2.1.2 (Empty set) *There exists a set, denoted by $\emptyset$, such that whatever object x we consider, x is not an element of $\emptyset$. We call $\emptyset$ the empty set, and the definition above can be written as $\forall x, x \notin \emptyset$.*

Notice that the set $\emptyset$ has no element, but the set $\{\emptyset\}$ has exactly one element, namely $\emptyset$ (so a set can also be an element: it suffices to be an object to be an element, and the word element sometimes implies the memebership to a set).

Definition 2.1.3 (Inclusion and subset) *Let F and E be two sets. We say that F is **included** in E, and denote it by $F \subset E$, if any element of F is an element of E. The above definition can be written as $\forall x \in F, x \in E$ and in this case, we say that F is a **subset** of E.*

Remark 2.1.1 (Axiom of specification) Most of what we do from this point on, is to define principles or operations that "creates" a set from a given set. Among them the most important one is the **Axiom of specification**, which, roughly speaking, allows one to choose elements of a given set satisfying a given property. In this way, we can create a subset of a set.

For example, given a set E: "all the students in Zoom" and the property (P): "student number ending with number 5", one can produce the subset:

$$\{x \text{ student in Zoom} \; ; \; \text{student number of } x \text{ ends with } 5\} \subset E$$

or more generally,

$$\{x \in E \; ; \; P(x)\} \subset E.$$

As before, the property (P) has to be well-defined (and not ambiguous). For example, "runs fast" is not precise enough, but "runs 100m within 10s" is well-defined.

The followings definitions are just remainders.

Definition 2.1.4 (Intersection) *Given two sets A and B, we call $A \cap B$ their intersection the set:*

$$A \cap B = \{x \; ; \; x \in A \text{ and } x \in B\}.$$

$A \cap B$ contains elements that belong to A and B at the same time.

Verify that $A \cap B$ is a subset of A and $A \cap B$ is a subset of B. Also, notice that $A \cap B = B \cap A$; this is called **commutativity**, but I don't want to introduce more terminologies at this point...I just assume that everyone knows it "subconsciously".

Definition 2.1.5 (Complement) *Given a subset A of U, we call A^c the complement of A in U:*

$$A^c = U - A = \{x \in U \; ; \; x \notin A\}.$$

Usually the set U is implicitly given by the context.

The definition also makes sense for any two sets A and B without necessarily $A \subset B$. In this case, we speak of the **relative complement** of A with respect to B and it is often denoted $B \setminus A$.

Definition 2.1.6 (Union) *Given two sets A and B, we call $A \cup B$ their union the set:*

$$A \cap B = \{x \; ; \; x \in A \text{ or } x \in B\}.$$

If I were asked to define the complement of a set A of U without specifying on the level of elements (which is more natural once you have gathered some experience on set theory, but perhaps not in the beginning), I probably want to specify it on the level of set. In other words, I would probably propose the following definition:

Definition 2.1.7 (Complement bis) *Let $A \subset U$. We call B the complement of A in U if*

1. $A \cup B = U$;
2. $A \cap B = \emptyset$.

We prove that these two definitions are equivalent. It is important to observe the proof technique: it is ultra-standard in proving that two sets are equal.[3]

3: Recall that, to prove that two statements are equivalence, we should prove two implications.

Proof. To prove that two sets are equal, we prove two inclusions. In our case, we are given two sets $A \subset U$, and we need to show that the set $U - A$ defined in the definition 2.1.5 is equal to the set B in the definition 2.1.7.

• First, we prove that $U - A \subset B$. Since $U - A$ is defined as $\{x \in U \,;\, x \notin A\}$, we need to show that for any element $x \in U$ such that $x \notin A$, we have $x \in B$. We proceed by contradiction. Suppose that there is some elements $x \in U$ such that $x \notin A$ and $x \notin B$. Then $x \notin A \cup B$, but since $A \cup B = U$ by assumption in the definition 2.1.7, it follows that $x \notin U$. This is a contradiction and we have proven that $U - A \subset B$.

• Now we prove that $B \subset U - A$. We need to show that for any element $b \in B$, we have $b \in U - A$. By definition of the set $U - A$, we need to show that $b \in U$ and $b \notin A$. The fact that $b \in U$ follows from $b \in B$ and $B \subset U$ (since $A \cup B = U$). To see that $b \notin A$, we again can proceed by contradiction: if $b \in A$, then $b \in A \cap B$, but $A \cap B = \emptyset$ by assumption, and this is impossible. We have proven that $B \subset U - A$.

• We have proven that $U - A \subset B$ and $B \subset U - A$: this implies that $B = U - A$, and that the definition 2.1.5 and the definition 2.1.7 are equivalent. □

Definition 2.1.8 (Power set) *Given a set A, we call $\mathscr{P}(A)$ the power set of A the set of all subsets of A:*

$$\mathscr{P}(A) = \{B \,;\, B \subset A\}.$$

Notice that its elements are subsets (and not elements) of A!

Quick question: what is $\mathscr{P}(\emptyset)$?[4]

4: The answer is not $\emptyset$.

We finish this section with some identities on operations on sets.

Proposition 2.1.1 (Distibutivity) *For all sets A, B, C, we have*

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

and

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Proposition 2.1.2 (De Morgan's laws) *Let A, B be subsets of U. Then*

$$(A \cup B)^c = A^c \cap B^c; \quad (A \cap B)^c = A^c \cup B^c.$$

The proofs of these propositions are left as exercises![5]

5: It means that your teacher is tired of typing and/or is having a sore back, which is bad.

2.2 Cardinal

The **cardinal** of a set A is the number of its elements. It is denoted by $|A|$ or $\mathrm{Card}(A)$, and can be finite or infinite.

Example 2.2.1 The cardinal of $\mathbb{Z}$, denoted $|\mathbb{Z}|$, is infinite.

Here are some examples of cardinals.

Example 2.2.2 The cardinal of $\emptyset$ is 0, while the cardinal of $\{\emptyset\}$ is 1.

Example 2.2.3 The cardinal of the set of prime numbers is infinity.[a]

[a] This is known as Euclid's theorem.

Example 2.2.4 The cardinal of $\{x \in \mathbb{C} \ ; \ x^2 = -1\}$ is 2.

Let A, B be two sets. We also define the **Cartesian**[6] **product** $A \times B$ in the following way:

$$A \times B = \{(a, b); a \in A, b \in B\}.$$

6: René Descartes, one of the founders of modern philosophy.

Axiom 2.2.1 (Descartes) *I think.*

Corollary 2.2.2 (Descartes) *I am.*

Proposition 2.2.3 *Let A, B be two finite sets. The cardinal of $A \times B$ is $|A| \times |B|$.*

The proof of this proposition, as well as that of the following proposition, are not object of this course, or at least not at this stage.

Proposition 2.2.4 *Let A be a finite set. The cardinal of $\mathscr{P}(A)$ is $2^{|A|}$.*[a]

[a] The idea of the proof is the following. If A is empty set, then the cardinal of $\mathscr{P}(\emptyset)$ is $1 = 2^0$. Otherwise, for any element a of A and any subset B of A, there are two possibilities, i.e. we can make one of the following choices: $a \in B$ or $a \notin B$. For every element we have two choices, and together this yields 2^n possibilities for a subset B of A. This is an example of a bad proof at this stage, since to properly write it, we need the notion of a bijection between sets, which is defined later in this course.

2.3 Story time: Russell's paradox

The naïve set theory, as opposed to the axiomatic set theory that we tried to introduce above, leads to many famous paradoxes.

Suppose that we don't have the axioms above and we define the set theory in the loosest way possible. Then one can think about the following question:

Question 2.1 (Russell's paradox) Does the set $E = \{x; x \notin x\}$ exist?

This question is related to the so-called **liar's paradoxe**: if someone says "I am a liar", is this person lying? Anyways, E cannot be well-defined, since you can ask the question: do we have $E \notin E$? If $E \notin E$, then by definition E satisfies the property defining elements of E, so E is an element of E and $E \in E$. If $E \in E$, then by definition of E, $E \notin E$. Oopsi.

Paradoxes of this type, discovered around 1900, provoked a serious crisis in set theory. One of the efforts in trying to construct a set theory free of paradoxes is called the Zermelo–Fraenkel set theory or **ZFC**.[7] However, Gödel's second incompleteness theorem shows that one cannot verify the consistency of ZFC within ZFC itself, and they are explicit examples of statement independent of ZFC (meaning they can neither be proven true or false by ZFC).[8]

For a more elaborated logic paradox of the same flavor, check out the poem on the door of Åsa Hirvonen (last retrieved: August 2020).

7: The "C" stands for "choice" or "axiom of choice". It is a famous axiom which has many consequences: people use it in mathematics all the time without even realizing it. One of the easy understanding version is probably "The Axiom of Choice is necessary to select a set from an infinite number of pairs of socks, but not an infinite number of pairs of shoes." by Bertrand Russell.

8: One of them is the "continuum hypothesis", which says that "There is no set whose cardinality is strictly between that of the integers and the real numbers."

2.4 Exercises

Exercise 2.1 Define the **symmetric difference** of two sets A and B as:

$$A \triangle B = (A \setminus B) \cup (B \setminus A).$$

1. Calculate
$$\{1, 2, 3\} \triangle \{3, 4\}.$$
2. Prove that
$$A \triangle B = (A \cup B) \setminus (A \cap B).$$
3. Sometimes we call the symmetric difference the **disjunctive union**. Do you have an explanation?

Exercise 2.2 The following questions are related.

1. Write down all subsets of the set
$$\{1, 2, 3\}.$$
How many subsets do you get?
2. Prove the formula:
$$2^3 = \binom{3}{0} + \binom{3}{1} + \binom{3}{2} + \binom{3}{3}.$$

Can you generalize the last result?

Exercise 2.3 Determine the cardinal of the following sets:

$$S_1 = \emptyset, \quad S_2 = \{S_1, \{S_1\}\}, \quad S_3 = \{S_2, \{S_2\}\}, \ldots$$

You can start by writing down explicit the first cases, e.g. $S_2 = \{\emptyset, \{\emptyset\}\}$, make a conjecture, and write a formal proof using mathematical induction.[9]

9: This is an easy exercise, but the natural numbers are defined in some system as the sets S_1, S_2, S_3 etc.

Exercise 2.4 ($*$) Let E be a finite set with n elements. Consider the set

$$\mathscr{E} = \{A, B \in \mathscr{P}(E); A \cup B = E\}.$$

What is the cardinal of $\mathscr{E}$?

Infimum and supremum | 3

"In mathematics, a small positive infinitesimal quantity, usually denoted ϵ, whose limit is usually taken as $\epsilon \longrightarrow 0$."

– Wolfram MathWorld.

All symbols are created equal, but some symbols are more equal than others. You can write $y = f(x)$ or $b = f(a)$ or $v = f(u)$ or $s = f(t)$, but at least in this course, we reserve the notation ϵ (and later δ) for special purposes.

3.1 Ordering

To define the infimum and supremum of a set, we first need to define a notion of "bigger" and "smaller". This is done by prescribing an **order** on a set E.

Example 3.1.1 (Ordering of $\mathbb{R}$) The set of real numbers $\mathbb{R}$ is equipped with a natural **total order**:

$$\leq \quad \text{less or equal to}$$

which compares two elements x, y of $\mathbb{R}$. For example,

$$3 \leq 3 < \pi \leq 4,$$

where the symbol $<$ means "strictly less than" or "less than and not equal to".

In general, we call a binary relation[1] $\leq$ a **total order** on a set E if the following properties hold for all elements $a, b, c \in E$:

1. **Antisymmetry**: if $a \leq b$ and $b \leq a$, then $a = b$;
2. **Transitivity**: if $a \leq b$ and $b \leq c$, then $a \leq c$;
3. **Connexity**: either $a \leq b$ or $b \leq a$.

On a set equipped with a total order, we can compare elements and start talking about the notions of minimum, maximum, infimum or supremum.[2]

1: I will not define what is a binary relation properly: but it is a relation between two elements as the name suggests.

2: There is also a notion of **partial order**, when the connexity property is replaced by the **reflexive property** that $a \leq a$. For example, on the set of positive integers $\mathbb{Z}_{>0}$, the relation

$$| \quad \text{"is divisible by"}$$

is only a partial order.

3.2 Minimum and maximum

Consider the set $E = \{3, 4, \pi\} \subset \mathbb{R}$. Since $\mathbb{R}$ is totally ordered[3], the elements of E can be rearranged in increasing order $3 < \pi < 4$ and it is natural to say that 3 is the smallest element and 4 is the largest element of E. A rigorous definition is the following:

3: Not to be confused with **well-ordered**, which assumes the existence of a minimum for all non-empty subsets.

Definition 3.2.1 *Let $(E, \leq)$ be a totally ordered set.*

We say that an element a is the ***minimum*** *of E if*

1. *a is an element of E, i.e. $a \in E$;*
2. *a is smaller than all other elements of E, i.e. $\forall x \in E, a \leq x$.*

We say that an element b is the ***maximum*** *of E if*

1. *b is an element of E, i.e. $b \in E$;*
2. *b is greater than all other elements of E, i.e. $\forall x \in E, x \leq b$.*

We denote respectively the minimum and the maximum of E by $\min(E)$ *and* $\max(E)$*, when they exist.*

It is a consequence of the antisymmetry property of the order $\leq$ that the minimum (or maximum), if it exists, is unique.[4] This is why we usually say "the minimum" or "the maximum".

4: Why? If say we have two candidates b and b' for the maximum of a non-empty set E, then 1) $b \in E$ and b' is greater or equal to all elements of E, so in particular $b \leq b'$ 2) $b' \in E$ and b is greater or equal to all elements of E, so $b' \leq b$. Then $b \leq b'$ and $b' \leq b$, and we have $b = b'$.

It is important to remember that the minimum or maximum of a set E must be an element of the set E itself.

Example 3.2.1 Some examples of minimum and maximum:

1. The set $\mathbb{Z}_{>0}$ of strictly positive integers has a minimum (namely $1 \in \mathbb{Z}_{>0}$) but no maximum.
2. The set $(0, 1] = \{x \in \mathbb{R}; 0 < x \leq 1\}$ has no minimum (notice that $0 \neq (0, 1]$) and is of maximum 1.

In the last example, we are tempted to say that 0 is still, in some sense, the "lower limit" of the set $(0, 1]$. To rigorously formulate this intuition, we introduce the notions of infimum and supremum.

3.3 Infimum and supremum

Let E be a subset of $\mathbb{R}$. We want to get rid of the restriction that $\min(E) \in E$ and extend the notion of minimum to something outside of the set E. For this, we first have to define a notion of comparing an arbitrary element of $\mathbb{R}$ to a set.

Definition 3.3.1 (Lower bound and upper bound) *Let $E \subset \mathbb{R}$ and $x, y \in \mathbb{R}$.*

We say that x is a ***lower bound*** *of E if for all elements $e \in E$, $x \leq e$.*

We say that y is an ***upper bound*** *of E if for all elements $e \in E$, $e \leq y$.*

Graphically, with the classical real line representation of $\mathbb{R}$, a lower bound of E is any real number "to the left" of the set E and an upper bound of E is any real number "to the right" of the set E.

Example 3.3.1 Some examples of lower bounds and upper bounds:

1. The set $\mathbb{Z}_{>0}$ has no upper bound in $\mathbb{R}$. Any real number $x \leq 0$ is

a lower bound of $\mathbb{Z}_{>0}$.
2. The set $(0, 1] = \{x \in \mathbb{R}; 0 < x \leq 1\}$ is upper bounded and lower bounded in $\mathbb{R}$. For instance, any real number $x \leq 0$ is a lower bound for $(0, 1]$.

Notice that for a given set E, different lower bounds or upper bounds might exist (in general they are never unique), so we usually say "a lower/upper bound" or "lower/upper bounds".

We are still tempted to say that for the set $(0, 1] = \{x \in \mathbb{R}; 0 < x \leq 1\}$, the real number 0 playes a special role in the set of its lower bounds. In natural language, one might characterize it by saying that it is the "best" lower bound, or "rightest" lower bound, or more precisely maybe "largest/greatest" lower bound. In a mathematical way and using the definitions we have already seen, we might propose the following definition:

Definition 3.3.2 (Infimum) *Let $E \subset \mathbb{R}$ such that E is lower bounded. In other words, the set of lower bounds of E is non-empty, or that there exists a real number $x \in \mathbb{R}$ such that x is a lower bound of E.*

We say that $a \in \mathbb{R}$ is the infimum of E if a is the maximum of the set of lower bounds of E.[a] *We denote the infimum of E by $\inf(E)$ when it exists in $\mathbb{R}$.*

[a] This is admittedly a complicated phrase to assimilate at first sight.

This is a sophisticated definition, so let us make it simpler. We write down the definition of supremum to change a little: you should be able to work out the definition for the infimum on your own.

Definition 3.3.3 (Supremum) *Let $E \subset \mathbb{R}$. We say that $b \in \mathbb{R}$ is the supremum of E if*

1. *b is a upper bound of E: for all $e \in E, e \leq b$;*
2. *b is the smallest upper bound of E: for all upper bound y of E, $b \leq y$.*

We denote by $\sup(E)$ the supremum of E when it exists in $\mathbb{R}$.

Sometimes, the supremum is referred to as the **least upper bound**. Notice that the supremum of a set does not necessarily exist.[5] Notice also that we wrote "the supremum" in the above definition, which implicitly implies that the supremum, when it exists, is unique (of course the same applies to the infimum).

5: Do you have an example?

Lemma 3.3.1 (Uniqueness of the supremum) *The supremum $\sup(E)$ of a set $E \subset \mathbb{R}$, if it exists, is unique.*

Proof. We prove by contradiction. Suppose that E has two different suprema b, b' with $b \neq b'$. Without loss of generality, suppose that $b < b'$. Then b is an upper bound for the set E and b is strictly smaller than b', this contradicts the property that b' is the smallest upper bound of E (since we supposed that b' is a supremum of E). Therefore, E can have only one supremum. □

Try to draw a picture to represent the situation!

Proposition 3.3.2 (Maximum and supremum) *If a set $E \subset \mathbb{R}$ has a maximal element, then this maximal element is the supremum of E.*

Proof. Suppose that b is the maximal element of E. Then by definition of the maximum, b is an upper bound of E, since for all elements $e \in E$, $e \leq b$. Furthermore, b is the smallest upper bound of E, for that any upper bound of E must be bigger or equal to all elements of E, in particular bigger or equal to b. We conclude that b is indeed the supremum of E. □

Theorem 3.3.3 (On the existence of the supremum) *If a set $E \subset \mathbb{R}$ is upper bounded in $\mathbb{R}$, then the supremum of E exists in $\mathbb{R}$.*

This theorem is admitted in this course.[6]

6: This is a very special property of $\mathbb{R}$ called the Dedekind completeness property. It is often taken as an axiom in the contruction of the real numbers.

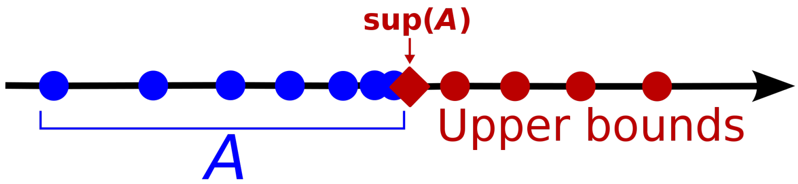


Figure 3.1: Upper bounds and supremum of a real set A.

3.4 An epsilon of room

This is an important moment of your life: you are going to see the use of ϵ in mathematical analysis.

For symmetry, let us study the infimum with the ϵ-language. I also took the liberty of changing the notations.[7]

7: One could say that we must seize the idea behind the notations, but half of the truth is that I am a little lazy.

Theorem 3.4.1 (Definition of infimum with ϵ) *Let A be a subset of $\mathbb{R}$ such that $\inf(A)$ exists. Then $p = \inf(A)$ if and only if*

1. *For every $x \in A$, $p \leq x$;*
2. *For every $\epsilon > 0$, there exists some $x \in A$ with $x < p + \epsilon$.*

Proof. The first item states that p is a lower bound of A. To see that the second item is equivalent to the fact that p is the greatest lower bound, we should write two implications. First, if p is the greatest lower bound of A, then $p + \epsilon$ is not a lower bound of A since it is strictly bigger than p, in such a way that the negation of the definition of a lower bound of A yields the existence of an element $x \in A$ with $x < p + \epsilon$. For the other implication, suppose that for every $\epsilon > 0$, there exists some $x \in A$ such that $x < p + \epsilon$. It follows that any real number $p' > p$ cannot be a lower bound of A, for that choosing $\epsilon = p' - p$ yields the existence of some $x \in A$ such that $x < p + \epsilon = p'$. So p is the largest lower bound, i.e. the infimum of the set A. □

The second item essentially expresses the idea that "there should be no gap between $\inf(A)$ and A". Again, draw a picture!

Notice also that we can change the strict $<$ inequality in the second item to a larger $\leq$ inequality (a quick way to see that is to replace ϵ by 2ϵ in the theorem).[8]

8: Try to write it out yourself!

Let us see some applications of this ϵ-formalism.

Lemma 3.4.2 (Uniqueness of the infimum, revisited) *Let E be a subset of $\mathbb{R}$. If $\inf(E)$ exists in $\mathbb{R}$, then it is unique.*

Proof. We proceed by contradiction. Suppose that $a, a' \in \mathbb{R}$ with $a < a'$ are both equal to $\inf(E)$.[9] Consider $\epsilon = \frac{a'-a}{2} > 0$. Using the ϵ-definition of infimum applied to the smaller a, we know that there exists some element $x \in E$ such that $x < a + \epsilon$. The choice of ϵ is such that $a + \epsilon < a'$, so that $x < a + \epsilon < a'$. But since a' is infimum of E, in particular a' is a lower bound of E, and for $x \in E$, we have $a' \leq x$. Thus, $x < a'$ and $a' \leq x$: this is a contradiction. □

9: The assumption $a < a'$ follows by symmetry.

Again, try to draw a picture!

Let us see an example of the above ϵ-definition in action.

Example 3.4.1 The infimum of the set $(0, 1] = \{x \in \mathbb{R}; 0 < x \leq 1\}$ is 0.

Proof. We verify the two items in the definition of infimum with ϵ.

• First, for every $x \in (0, 1]$, we have $0 < x$ by definition of the set $(0, 1]$, so that $0 \leq x$ (i.e. 0 is a lower bound of $(0, 1]$).

• Now let $\epsilon > 0$.[10] We prove that there exists some $x \in (0, 1]$, such that $x < \epsilon$. Consider $x = \frac{\epsilon}{2}$. By definition, $x > 0$ (since $\epsilon > 0$) and $x < \epsilon$. This shows that 0 is the greatest lower bound.

10: This means we can take arbitrary $\epsilon > 0$, and corresponds to the "for every $\epsilon > 0$" part of the statement.

• In conclusion, the infimum of the set $(0, 1]$ exists and is equal to 0. □

3.5 Real intervals

In the previous example, $(0, 1]$ is called an interval of $\mathbb{R}$. This interval has the following properties:

- It contains all real numbers between 0 and 1;
- It does not contain its "left end point" 0;
- It does contain its "right end point" 1;
- It has "length" equal to 1;
- etc.

In general, we call a real set $I \subset \mathbb{R}$ an interval if the "no hole" property above is true. This should be formalized better in the following way:

Definition 3.5.1 (Intermediate value property characterization of real intervals) *A real interval is a subset I of $\mathbb{R}$ such that*

$$\forall x, y \in I, \forall z \in \mathbb{R}, (x \leq z \leq y) \implies (z \in I).$$

The end points of an interval are somewhat special. There are many terminologies associated to the classification of real intervals and most of them are directed associated to the properties of the end points (since by the above, the end points characterized the real interval). Let us enumerate all types of real interval: in the following $a < b$ are real numbers.

1. Empty interval: $\emptyset$;
2. Degenerate interval or singleton: $[a, a] = \{a\}$;
3. Bounded, open interval: $(a, b) = \{x \in \mathbb{R} \,;\, a < x < b\}$;
4. Bounded, closed interval: $[a, b] = \{x \in \mathbb{R} \,;\, a \leq x \leq b\}$;
5. Bounded, left-closed, right-open: $[a, b) = \{x \in \mathbb{R} \,;\, a \leq x < b\}$;
6. Bounded, left-open, right-closed: $(a, b] = \{x \in \mathbb{R} \,;\, a < x \leq b\}$;
7. Unbounded, left-open interval: $(a, \infty) = \{x \in \mathbb{R} \,;\, a < x\}$;
8. Unbounded, left-closed interval: $[a, \infty) = \{x \in \mathbb{R} \,;\, a \leq x\}$;
9. Unbounded, right-open interval: $(-\infty, b) = \{x \in \mathbb{R} \,;\, x < b\}$;
10. Unbounded, right-closed interval: $(-\infty, b] = \{x \in \mathbb{R} \,;\, x \leq b\}$;
11. Unbounded at both ends: $(-\infty, \infty) = \mathbb{R}$.

3.6 Development: Archimedean property of the set of real numbers

We are quite familiar with real numbers. However, the set of real numbers has many interesting **structures**, which we are quite used to and taken for granted without furthur examination.[11]

One of the many features of real numbers that we take for granted is the **Archimedean property**. For example, we seldom doubt the validity of the following statement:

Proposition 3.6.1 *The set $\mathbb{Z}_{\geq 0}$ has no upper bound in $\mathbb{R}$. In other words, this set is not upper bounded and there exists arbitrarily large integer.*

Or the following one that you might have seen already:

Definition 3.6.1 (Integer part of a real number) *For every real number $x \in \mathbb{R}$, we can define its* ***integer part****, denoted $\lfloor x \rfloor$, as the unique integer k such that $k \leq x < k + 1$.*

However, although the above statements look innocently evident, it is hard (and I would say impossible for us) to prove! Indeed, what we are asking is a very deep property on the **construction** of the real numbers, otherwise put, to prove these statements require to ponder upon the question of "what are real numbers, really"?[12]

11: In a first approximation, a set is merely a collection of objects (it is in some sense the loosest structure one can imagine). What usually makes a set interesting is the additional structures it possesses: for example, if we can define a total order on a certain set, than this set becomes suddenly more interesting in many ways, since its elements are related if we have the knowledge of this structure. When you study linear algebra in the previous course, a matrix is a collection of coordinates, but for example, you have **operations** on the set of matrices, e.g. the multiplication of matrices (which can be interpreted as composition of linear functions), which makes this definition way more interesting.

12: The word "deep" is a mathematical jargon, which means approximately obscure, hard, not elementary, sophisticated, requires a long chain of non-intuitive logical deductions or things of this sort.

It is a weird feature of this course (and of mathematics as you will see), that a simple question might require a response longer than what a two-month course can cover. So let us just mention the following **axiom** for $\mathbb{R}$ (together with its absolute value $|\cdot|$):

Axiom 3.6.2 *For every non-zero real number $x \in \mathbb{R}$, there exists an integer $n > 0$ such that*

$$|x + \cdots + x| > 1$$

where there are n terms of x in the summation.

For example, from this axiom, for arbitrary integer $m > 0$, we can consider the real number $\frac{1}{m} > 0$. Applying the above axiom to $x = \frac{1}{m}$, we obtain an integer $n > 0$ such that $n \times \frac{1}{m} > 1$: that is, $n > m$.

3.7 Exercises

Exercise 3.1 True or false:[13]

1. For a finite, non-empty set A, $\sup(A) = \max(A)$.
2. For any set A, $\sup(A) = -\inf(A)$.
3. For any non-empty set A, $\inf(A) \le \sup(A)$.

13: As a good habit: always check if a set is empty. The statement "for all $x \in \emptyset$, P" is always true whatever the statement P is.

Exercise 3.2 Determine the following quantities:

1. $$\inf\{x \in \mathbb{R}\ ;\ x^2 > 2\}.$$
2. For $-1 < q < 1$,
$$\sup\left\{\sum_{k=0}^{n} q^k\ ;\ n \in \mathbb{Z}_{\ge 0}\right\}.$$
Be careful that q can be negative in this question.[14]
3. $$\inf\{x^2 - 3x + 2\ ;\ -1 < x \le 2\}.$$

14: You can use an exercise from past weeks...!

Exercise 3.3 Let $A = ([0, \pi) \cap [2, 4]) \cup (\sqrt{2}, \sqrt{10})$.

1. Determine $\inf(A)$ and $\sup(A)$.
2. Is A an interval of $\mathbb{R}$?

Exercise 3.4 Reproof all the results in this chapter using the ϵ formalism.

Exercise 3.5 (Homework) Let $A = \left\{\frac{n}{n+1}\right\}_{n \in \mathbb{Z}_{\ge 1}}$.[15]

1. Find the infimum and the supremum of A.
2. Let $\epsilon > 0$. Pick an element a of A such that $\sup(A) - \epsilon \le a$.

15: It means that $A = \left\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots\right\}$.

Notice that the element a **depends** on the choice of ϵ.[16]

16: If one wants to be very rigorous, one can write a_ϵ or $a(\epsilon)$ instead. The statement "$\forall \epsilon, \exists a, \ldots$" (implicitly) implies that a depends on ϵ.

Exercise 3.6 Define the function $\ln^+$ on $\mathbb{R}_{>0}$ as

$$\forall x > 0, \quad \ln^+(x) = \max(\ln(x), 0).$$

Show that for $x, y > 0$,

$$\ln^+(x) + \ln^+(y) \geq \ln^+(xy).$$

Exercise 3.7 Let $a, b > 0$ and $0 < p < 1$. Show that:

1. By comparing a with b,

$$\inf(a^p b^{1-p}, a^{1-p} b^p) \geq \inf(a, b).$$

2. Simplified Muirhead inequality or Hardy-Littlewood inequality:

$$a + b \geq a^p b^{1-p} + a^{1-p} b^p.$$

Exercise 3.8 (*) Let A, B be subsets of $\mathbb{R}$. Suppose that $\sup(A) = M_A \in \mathbb{R}$ and $\sup(B) = M_B \in \mathbb{R}$.

1. What can we say about $\sup(A + B)$?
2. What can we say about $\sup(A - B)$?
3. What can we say about $\sup(A - A)$?

Here, $A + B$ (resp. $A - B$) are subsets of $\mathbb{R}$ defined as

$$A + B = \{a + b; \quad a \in A, b \in B\},$$

and respectively

$$A - B = \{a - b; \quad a \in A, b \in B\}.$$

[Week II&III]
What are...functions?

4 Functions: definitions and properties

In high school, most functions are given in the form of a formula:

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

However, in full generality, a function is defined in a more abstract way. The essential idea is the **association** of an element to a given element (in the example above, for each real number x, we associate the real number $y = x^2 + 1$). The abstract definition has many advantages and covers more situations, for example, we will see that a sequence can be seen as a function (from a set of integers $\mathbb{Z}_{n\geq 0}$ to the set of real numbers $\mathbb{R}$).

4.1 Notations

Formally speaking, a function is defined with three parts:[1]

1. A set X called the **domain** of the function;
2. A set Y called the **codomain** of the function;
3. A relation that associates each element X to a single element of Y.

Sometimes, by abuse of notation,[2] we are less precise on the domain and the codomain in the definition of a function (e.g. we supposed implicitly that the function is defined where it can be defined, or that the domain of the function follows from the context). But keep in mind that two functions can be different even if they share the same expression, e.g. in the case where their domains and/or codomains differ.

1: Suppose that we have an expression in the form $y = f(x)$: X is then the set of values of x, Y the set of possible values of y (it does not mean that all elements of Y is attained), and the relation is recorded in the expression of f.

2: Sometimes this just means "The author is tired."

Here is an example of a (very useful) function defined in an abstract way.

Example 4.1.1 The Möbius function $\mu : \mathbb{Z}_{>0} \longrightarrow \{-1, 0, 1\}$ is defined depending on the factorization of a positive integer into prime factors. For any positive integer $n > 0$, the value of $\mu(n)$ is defined in the following way:

1. $\mu(n) = 1$ if n is a **square-free** positive integer with an **even** number of prime factors;
2. $\mu(n) = -1$ if n is a **square-free** positive integer with an **odd** number of prime factors;
3. $\mu(n) = 0$ if n has a squared prime factor.

The Möbius function μ has an alternative definition in terms of roots of unity (if you are interested, take a look at the extra exercise ??.)[a]

[a] Using the Möbius function μ, one defines the Mertens function $M : \mathbb{R} \longrightarrow \mathbb{R}$, $M(x) = \sum_{n\in\mathbb{Z}_{>0}, n\leq x} \mu(n)$. It is currently unknown whether $M(x) = O\left(x^{\frac{1}{2}+\epsilon}\right)$ for all $\epsilon > 0$. I will restrain myself from putting this as an extra exercise.

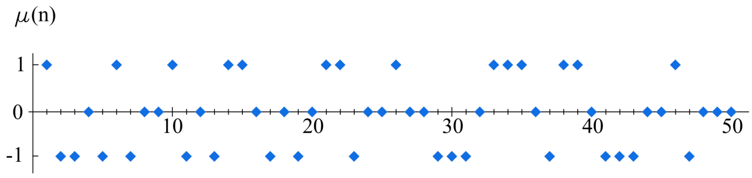


Figure 4.1: First values of the Möbius function μ.

Different authors use diffrent words for the sets X and Y: I will not make a list here since in most cases they should be quite clear from the context. In this course, we write $f : X \longrightarrow Y$ for a function between two sets X and Y, and $f : x \longmapsto y$ or $f(x) = y$ the association relation between elements of X and Y.[3]

3: Also, sometimes we speak of applications or maps instead of functions.

It is important to remember that, unless otherwise specifies, for any element of the set of departure X, one and only one element of the set of destination Y is associated.[4]

4: Sometimes one might encounter the so-called **multivariate** functions, but we do not use this notion in this course.

4.2 Injections, surjections, bijections

An important class of functions is the class of bijective functions: for a bijective function, we can define its inverse.[5]

5: The inverse function of f is **not** $\frac{1}{f}$, although the notations for the inverse function f^{-1} and for the negative power f^{-1} can be sources of confusion: one should distinguish them carefully depending on the context. The function $g = \frac{1}{f}$, when it is defined, is more commonly called the reciprocal of f. If we really want to make the distinction, we can write $(f(x))^{-1}$ for $\frac{1}{f(x)}$.

Example 4.2.1 The inverse function of the exponential function $\exp : \mathbb{R} \longrightarrow \mathbb{R}_{>0}$ is the logarithm $\ln : \mathbb{R}_{>0} \longrightarrow \mathbb{R}$.

To explain what is a bijection (i.e. bijective function), we first explain what is a injection (resp. surjection).

Definition 4.2.1 (Injection) *Let $f : X \longrightarrow Y$ be a function. We call f an **injection** if*

$$\forall x, x' \in X, (x \neq x') \Longrightarrow (f(x) \neq f(x')).$$

This reads, for different x, x' in X, their images $f(x), f(x')$ by an injective function f are different.

It is up to you to find your own formulation of this definition.[6]

6: For example, one could try "an injection seperates the elements", or "if f merges two different elements into one then f is not injective", propose and discuss with your peers to find the subtleties!

Definition 4.2.2 (Surjection) *Let $f : X \longrightarrow Y$ be a function. We call f an **surjection** if*

$$\forall y \in Y, \exists x \in X, f(x) = y.$$

This reads, for any y in Y, there exists some x in X such that y is the image of x by the function f.

Again, try to find your own formulation of this definition![7]

7: What if we change the definition to $\exists x \in X, \forall y \in Y, f(x) = y$, what does it tell you? Be careful about interchanging the positions of the quantifiers!

Definition 4.2.3 (Bijection) *A function $f : X \longrightarrow Y$ is a bijection if and only if it is an injection and a surjection.*

In some sense, a bijection associates an element of X to an element of Y in a unique way. Another way of formulating the above definitions is via the notion of **preimage** (or sometimes **inverse image**).

Definition 4.2.4 (Preimage) *Let $f : X \longrightarrow Y$ be a function and let B be a subset of Y. We define the **preimage** of B by f, denotes as $f^{-1}(B)$, as the following subset of X:*

$$f^{-1}(B) = \{x \in X\ ;\ f(x) \in B\}.$$

In particular, if B is a singleton, i.e. $B = \{y\}$ with $y \in Y$, we can speak of the preimage of the element y as the subset $f^{-1}(\{y\})$ of X. Notice that this set is not necessarily a singleton (it might be any subset of X from the empty set $\emptyset$ to the whole set X).

The notation $f^{-1}(B)$ does not in general refer to the inverse function, since we do not suppose anything on the function f: in particular the notion of preimage is always well-defined even for non-bijective functions.

Now one can reformulate the above definitions about injective/surjective/bijective functions:

Definition 4.2.5 (Preimage and injection/sujection/bijection) *Let $f : X \longrightarrow Y$ be a function.*

- *The function f is injective if for any singleton $\{y\} \subset Y$, the preimage $f^{-1}(\{y\})$ has **at most one** element.*[8]
- *The function f is surjective if for any singleton $\{y\} \subset Y$, the preimage $f^{-1}(\{y\})$ has **at least one** element.*[9]
- *The function f is bijective if for any singleton $\{y\} \subset Y$, the preimage $f^{-1}(\{y\})$ has **exactly one** element.*[10]

8: An injective function is called a **one-to-one** function.

9: A surjective function is called an onto function.

10: We speak of **one-to-one correspondence** in this case. Personally I don't like these terminologies, sometimes they cause confusions for me.

To finish this section, we define the image of a function. Although it might sound weird for many as this moment, you will see that in practise, the notion of the preimage is more convenient to use compared to the notion of the image.

Definition 4.2.6 (Image of a function) *Let $f : X \longrightarrow Y$ be a function and $A \subset X$ a subset of X. The image of the set A by the function f is defined as the following subset of Y, denoted $f(A)$:*

$$f(A) = \{y \in B\ ;\ \exists x \in A, y = f(x)\}.$$

If $f : X \longrightarrow Y$ is a function, then the image of the whole set X, as a subset $f(X)$ of Y, is called the **range** of the function f.[11]

11: Sometimes, the range of a function can refer to the codomain of the function: it depends on the author and the context. To avoid this possible confusion, we can write $f(X)$ for the range of $f : X \longrightarrow Y$.

4.3 Composition and inverse

Let us define the composition of functions properly.

Definition 4.3.1 (Composition of functions) *Let X, Y and Z be three sets and let $f : X \longrightarrow Y$, $G : Y \longrightarrow Z$ be two functions. We define the composition of f and g to be the function, denoted by $g \circ f$, in the following way:*[12]

$$\forall x \in X, \quad (g \circ f)(x) = g(f(x)) \in Z.$$

It follows that $g \circ f$ is a function from the set X to the set Z.

12: Notice the order in the notation!

It is important to verify before using the notation $g \circ f$ that it is indeed well-defined (in particular, the function g should be at least defined on the range of the function f).

Similarly to what you have already seen in the linear algebra course, the inverse of a function is defined in the following way:

Definition 4.3.2 (Inverse of a function) *Let $f : A \longrightarrow B$ be a bijective function. A function $g : B \longrightarrow A$ is called the inverse of the function f if and only if $f \circ g = Id$ and $g \circ f = Id$.*

Let us make a list of remarks here.

1. The identity function Id is such that $\mathrm{Id}(x) = x$ for all x; it has the property that for any function f, $f \circ \mathrm{Id} = \mathrm{Id} \circ f = f$ (when the domain and codomain are suitably chosen).
2. The fact that $f \circ g$ and $g \circ f$ are well-defined gives some information. For example, it shows that the range of f should be contained in the domain of g (and the range of g should be contained in the domain of f).
3. If we require further, for example that $g \circ f = \mathrm{Id}$, then more information can be deduced. This means that for all $x \in A$ (why it is not B here?), we have $g(f(x)) = x$. It shows that the range of g is at least A: for each $x \in A$, if we take $t = f(x)$ then $g(t) = x$. This shows the surjectivity of g if we require $g : B \longrightarrow A$. What is perhaps a little bit less trivial is that f is then injective, since for different $x \neq y$ in A, if $f(x) = f(y)$ then $g(f(x)) = g(f(y))$, but means that $x = y$ and is impossible.
4. If we incorporate the other equation $f(g(x)) = x$, then we see that imposing these two requirements actually implies the bijectivity of f (and that of g)!

A final remark on an important concept (unfortunately, we don't have time to develop much here in this course) called the **restriction** of a function.

Definition 4.3.3 (Restriction) *Let $f : S \longrightarrow \mathbb{R}$ be a function and let R a subset of S. The* ***restriction*** *of f to R, denoted $f|_R$, is the function defined on R such that*[a]

$$\forall x \in R; \quad f|_R(x) = f(x).$$

[a] Or sometimes denoted $\tilde{f}$, or even just f when the author is lazy: of course this goes against the principle of our course, but it is commonly accepted if you declare "by abuse of notation" aforehand.

In a similar fashion, one can restrict the domain instead of the codomain, as long as the function is still well-defined! This can be useful for turning an injective function into a bijective function.

4.4 Real functions and monotonicity

In this course, we will be mainly interested in functions defined on a subset S of $\mathbb{R}$ and with values in $\mathbb{R}$: they will be called **real functions**.

Example 4.4.1 The (natural) logarithm function ln is a real function. Its domain of definition is $\mathbb{R}_{>0}$.[a]

[a] There are some differences between the notations log and ln. If you are interested, see here.

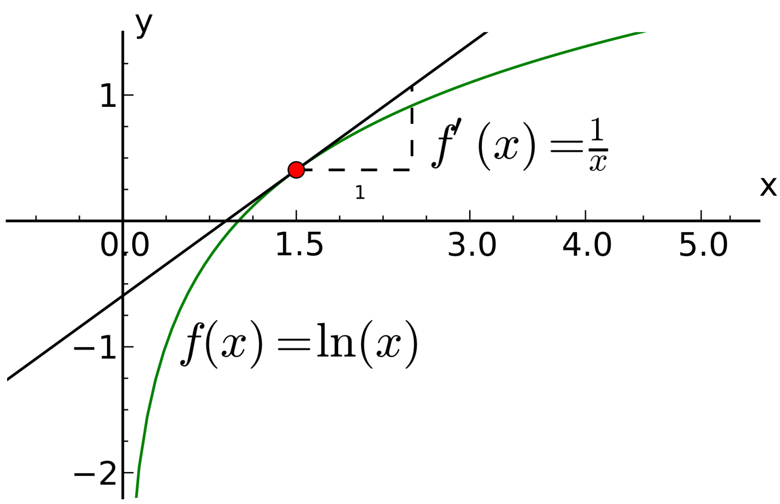


Figure 4.2: The graph of the natural logarithm function.

For a real function, we can draw its **graph**. More precisely,

Definition 4.4.1 *The graph of a real function $f : S \to \mathbb{R}$ is the set of points $(x, f(x))$ in the plane $\mathbb{R} \times \mathbb{R}$. In other words, a point $(x, y) \in \mathbb{R}^2$ is in the graph of f if and only if $y = f(x)$.*

Some properties on the function can be checked from its graph. For example:

1. A function f is well-defined at x if the vertical line crossing the point $(x, 0)$ intersects the graph of f at one and only one point;
2. A function f is injective, if for every y, the horizontal line crossing the point $(0, y)$ intersects the graph of f at at most one point;
3. Monotonicity of a function: see below!

However, for some more complicated properties[13], it is hard to see them on a graph. Sometimes it is even impossible to draw the graph of a given function!

13: A mathematician would say "fine" properties.

Sometimes it is important to point out if a function is increasing or decreasing.[14]

14: Well, for example, in this corona time, it is important to know if the situation is getting better or worse...

Definition 4.4.2 *A function $f : S \longrightarrow \mathbb{R}$ is called* ***increasing*** *or* ***non-decreasing****, if for all $x, y \in S$ with $x \le y$, we have $f(x) \le f(y)$. We say that f is* ***strictly increasing*** *if for all $x, y \in S$ with $x < y$, we have $f(x) < f(y)$.*

Similarly, a function $f : S \longrightarrow \mathbb{R}$ is called ***decreasing*** *or* ***non-increasing****, if for all $x, y \in S$ with $x \le y$, we have $f(x) \ge f(y)$. We say that f is* ***strictly decreasing*** *if for all $x, y \in S$ with $x < y$, we have $f(x) > f(y)$.*

It is a good habit to create some simple exercises to check the definition. For example, you can test that the identity function $f(x) = x$ is always (strictly) increasing, and that if g is increasing, then $-g$ is decreasing. Everything seems coherent.

Proposition 4.4.1 (Strictly monotonicity implies injectivity) *Let $f : S \longrightarrow \mathbb{R}$ be a strictly increasing function. Then f is injective.*

Proof. Let $x \neq y$ be in S. We can assume $x < y$ by symmetry.[15] Then $f(x) < f(y)$ by strict monotonicity, and in particular $f(x) \neq f(y)$: this shows the injectivity of f. □

15: Or some might write **w.l.o.g.** for "without loss of generality", for that if $x > y$, we can just change x into y and y into x.

Sometimes, a function can be defined implicitly. For example, the equation

$$xy = 1$$

defines a real function from $\mathbb{R} \setminus \{0\}$ to $\mathbb{R} \setminus \{0\}$ which sends x to y, such that (x, y) is a solution to the equation.[16] In the above example you can work out an explicit expression of y in the form of $y = f(x)$, but in more complicated situations, we will work with the equation instead of the explicit expression. However, if the equation has several solutions y for the same x, one has to specify one of them – otherwise the function is ill-defined!

16: The set $\mathbb{R} \setminus \{0\}$ is often denoted $\mathbb{R}^*$.

4.5 Story time: Cantor's theorem

We have already seen that for a finite set X with n elements, the cardinal of the power set $\mathscr{P}(X)$, i.e. the set of subsets of X, contains exactly 2^n elements.[17] It is easy to verify that $n < 2^n$ for all $n \in \mathbb{Z}_{>0}$: therefore, for a finite set X, we have

$$|X| < |\mathscr{P}(X)|$$

with a strict inequality.

17: Quick remainder: for each element of X, they have 2 choices, to be or not to be, in a subset of X.

What is the situation for infinite sets? We first use the following rule for comparing the cardinals of sets in general:

Definition 4.5.1 (Comparison of cardinals) *Let X and Y be two sets. We write*

$$|X| \le |Y|$$

if there exists an injection $f : X \longrightarrow Y$.

Intuitively, this definition says that via the injection f, we obtain a copy of the set X in the set Y, namely $f(X) \subset Y$ (since by a theorem before, $f : X \longrightarrow f(X)$ is a bijection if f is injective).[18] This is coherent with our intuition on finite sets.

We will prove a theorem of Cantor, that

Theorem 4.5.1 *There is no injection from $\mathscr{P}(X)$ to X for any set X. (Almost) equivalently, there is no surjection from X to $\mathscr{P}(X)$.*[a]

[a] See an exercise of this section for this equivalence.

A direct but confusing at first sight consequence: there exists different levels of infinity, and actually infinite[19] levels of infinity, since for any infinite set X,

$$|X| < |\mathscr{P}(X)| < |\mathscr{P}(\mathscr{P}(X))| < \dots$$

Let us prove this theorem. The way it is formulated suggests a proof by contradiction.

Proof. Suppose that there is a surjection $f : X \longrightarrow \mathscr{P}(X)$. Remember that f sends an element of X to a subset of X. We are going to construct a subset of X which it is impossible to be in the image of f.

Consider the subset E of X defined as[20]

$$E = \{x \in X; x \notin f(x)\}.$$

Since f is a surjection, there exists $x \in X$ such that $E = f(x)$.

Now the magical question: does x belong to E? If $x \in E$, by definition of E, $x \notin f(x) = E$. But if $x \notin E$, since $E = f(x)$, it means that $x \notin f(x)$ and by definition of E, $x \in E$. So $x \in E \iff x \notin E$: contradiction. □

It is pretty clear that the map $x \mapsto \{x\}$ from X to $\mathscr{P}(X)$ is an injection. So the cardinal of $\mathscr{P}(X)$ is strictly bigger than the cardinal of X.

To finish the story, let us mention some highlights of mathematics:

Question 4.1 Which one is bigger, $|\mathbb{N}|$ or $|\mathbb{R}|$?

This question is answered by Cantor, e.g. by using his diagonal argument.

Question 4.2 Which one is bigger, $|\mathscr{P}(\mathbb{N})|$ or $|\mathbb{R}|$?

You might be able to answer this question: try to use the binary representation of a real number.

Question 4.3 Is there any level of infinity strictly between $|\mathbb{N}|$ and $|\mathbb{R}|$?

This question is first settled by Gödel. In a highly heuristical and popular way of putting it into words: this question is undecidable.

18: However it is horrendously difficult for at this stage for you to prove that $\leq$ is antisymmetric. Some might call it the Bendixon-Bernstein-Borel-Cantor-Dedekind-Schröder-Zermelo theorem.

19: So you might ask, what is this infinity? That's a good question to ask a logician of the department!

20: To this day it still amazes me how one come up with such a beautiful construction.

4.6 Exercises

Exercise 4.1 Let $f, g : \mathbb{R} \longrightarrow \mathbb{R}$ be two (strictly) increasing real functions. Define the real functions

$$\begin{aligned} f+g &: \mathbb{R} \longrightarrow \mathbb{R}, & (f+g)(x) &= f(x)+g(x); \\ f \cdot g &: \mathbb{R} \longrightarrow \mathbb{R}, & (f \cdot g)(x) &= f(x) \cdot g(x); \\ g \circ f &: \mathbb{R} \longrightarrow \mathbb{R}, & (g \circ f)(x) &= g(f(x)); \\ \max(f,g) &: \mathbb{R} \longrightarrow \mathbb{R}, & (\max(f,g))(x) &= \max(f(x), g(x)); \\ \min(f,g) &: \mathbb{R} \longrightarrow \mathbb{R}, & (\min(f,g))(x) &= \min(f(x), g(x)). \end{aligned}$$

Prove or disprove:

1. The function $f + g$ is increasing;
2. The function $f - g$ is increasing;
3. The function $f \cdot g$ is increasing;
4. The function $g \circ f$ is increasing;[21]
5. The function $\max(f, g)$ is increasing;
6. The function $\min(f, g)$ is increasing;
7. The function f^{-1} exists and f^{-1} is increasing.

21: Is this function well-defined?

Exercise 4.2 Let $A = \{1, 2, 3\}$ and $B = \{1, 2\}$. Consider a function $f : A \longrightarrow B$.

1. Can f be a surjection? If yes, give an example.
2. Can f be an injection? If yes, give an example.
3. Can the preimage of $\{1\}$ be A? If yes, give an example.
4. Can the image of $\{2\}$ be $\emptyset$? If yes, give an example.

Exercise 4.3 Let $A = \{1, 2\}$. Consider a function $f : A \longrightarrow A$.

1. If the range of f is $\{1\}$, can you determine the function f?
2. If the range of f is $\{1, 2\}$, can you determine the function f? What are the possibilities?
3. Show that $\underbrace{f \circ \cdots \circ f}_{k-1 \text{ times } \circ} = \underbrace{f \circ \cdots \circ f}_{l-1 \text{ times } \circ}$ for some integers $k \neq l$.

Exercise 4.4 (Homework) Consider the functions $f : \mathbb{R} \longrightarrow \mathbb{R}$ with $f(x) = x^2 + x + 1$ and $g : \mathbb{R} \longrightarrow \mathbb{R}$ with $g(x) = x - 1$.

1. Find the image of the interval $[-1, 1]$ by f.
2. Find the preimage of the interval $[2, 7]$ by f.
3. Calculate $g \cdot f$, $f \cdot g$, $f \circ g$, $g \circ f$.

Exercise 4.5 (Homework) Consider the functions $\ln : \mathbb{R}_{>0} \longrightarrow \mathbb{R}$ and $\sin : \mathbb{R} \longrightarrow [-1, 1]$.

1. Is ln a bijection? Is sin a bijection? If any of these answers is yes, give the corresponding inverse function.
2. Is $\sin \circ \ln$ well-defined on the domain of ln? Is $\ln \circ \sin$ well-defined on the domain of sin? Is $\ln \circ \sin$ well-defined on the interval $(0, 1)$?
3. What is the preimage of $(0, 1)$ by ln? What is the preimage of $(0, 1)$ by sin?
4. Is ln monotonic? Is sin monotonic?
5. Find all solutions to the equation $\ln(\sin(x)) = 0$.

Exercise 4.6 Suppose that $f : X \longrightarrow Y$ is an injection. Show that there exists a surjection $g : Y \longrightarrow X$.

Suppose that $f : X \longrightarrow Y$ is a surjection. Show that there exists an injection $g : Y \longrightarrow X$.[22]

22: If you feel that something is wrong, that's an extremely good sign and totally justified. You might want to check out the axiom of choice.

Exercise 4.7 (*) Let $f : [0,1] \longrightarrow \mathbb{R}$ be a bounded function. We define a new function $g : [0,1] \longrightarrow \mathbb{R}$ by the following:

$$\forall x \in [0,1], \quad g(x) = \sup_{t\in[0,x]} f(t) = \sup\{f(t)\ ;\ t \in [0,x]\}.$$

Verify that g is well-defined and that g is increasing on $[0,1]$.

Limits and continuity of functions 5

This chapter is probably the most essential part of this course: we introduce the rigorous definition of limit using the (ϵ, δ)-formalism.[1]

> The first encounter with a limit marks the dividing line between the elementary and the advanced parts of a school course. Here we have not a new manipulation of old operations, but a new operation; not a new trick, but a new idea.
>
> – Alan Broadbent

1: Of course, it does not really mean that it is the "best" formalism, which in general does not exist. You can try to challenge yourself by proposing another formalism after finishing this chapter: it seems very hard to propose a better alternative. That's why in this course, and for most scientists, we stick to this formalism.

5.1 Limit of a function

Let us first look at two examples.

Example 5.1.1 Consider the function $|\cdot| : x \mapsto |x|$ from $\mathbb{R}$ to $\mathbb{R}$. This function has a limit at the point $x = 0$, and the limit is 0.

We denote this by $\lim\limits_{x \to 0} |x| = 0$.

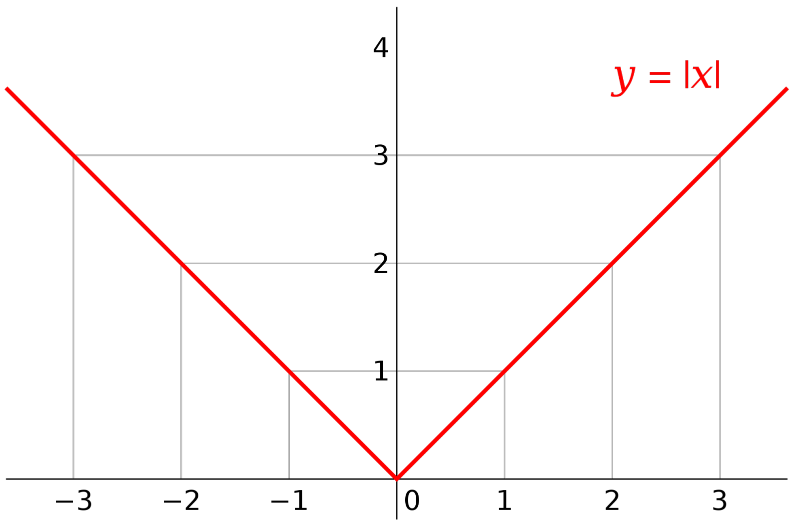


Figure 5.1: The graph of the absolute value function.

Example 5.1.2 Consider the function $H : \mathbb{R} \longrightarrow \mathbb{R}$ such that $H(x) = 0$ if $x < 0$, $H(0) = \frac{1}{2}$ and $H(x) = 1$ if $x > 0$. This function does not have a limit at the point $x = 0$.

In this case, the notation $\lim\limits_{x \to 0} H(x)$ makes no sense (and its usage is forbidden in this course), and we write simple "the limit of the function H at 0 does not exist".[2]

2: The convention in this course is that, $\lim\limits_{x \to 0} H(x) \neq 0$ means "the limit of the function H at 0 **exists**, but it is not 0". By writing $\lim\limits_{x \to 0} H(x)$, we assume implicitly that the limit already exists.

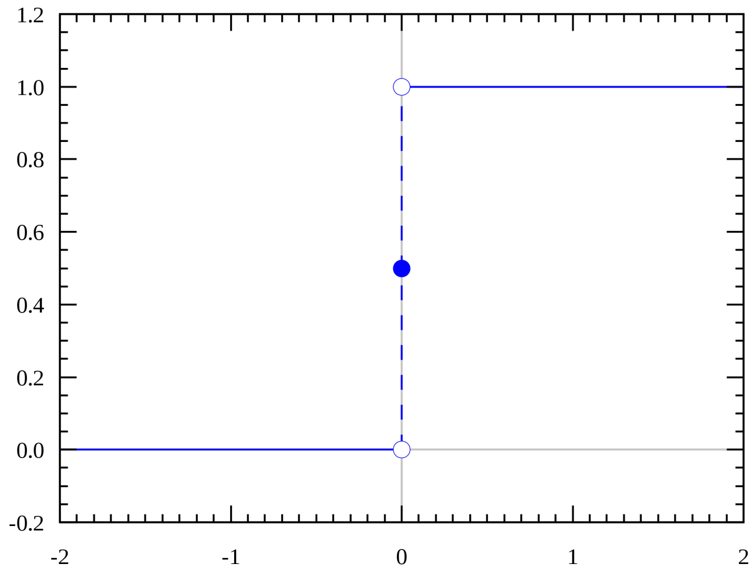


Figure 5.2: The graph of the Heaviside function.

How do we rigorously define the limit of a function? We introduce the **epsilon-delta definition of limit**.[3]

3: Since this is an important definition, I'm following the section II.2 of the classical textbook « Undergraduate Analysis » by Serge Lang.

Definition 5.1.1 *Let* $f : S \longrightarrow \mathbb{R}$ *and* $a \in \mathbb{R}$ *such that for all* $\epsilon > 0$*, there exists some* $x \in S$ *such that* $|x - a| < \epsilon$*.*

We say that the limit of $f(x)$ *as* x *approaches* a *exists, if there exists a number* L *having the following property. Given* ϵ*, there exists a number* $\delta > 0$ *such that, for all* $x \in S$ *verifying* $|x - a| < \delta$*, we have* $|f(x) - L| < \epsilon$*.*

If that is the case, we write[a]

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

[a] *A priori*, we should specify that the limit is taken with respect to S, but it can be shown that this is not necessary. We don't want to go into this now.

We also say that the function f is continuous at the point a.[4] In this course, except for the study of sequences, we will be looking at limits of functions defined on an interval I instead of the general setting with arbitrary set S.

4: At a first glance, one might wonder what the requirement on a is in the above definition. It is called **adherence** and without it, we might get an empty set for x in the above, which by the "false implies everything" logic convention, every function is automatically continuous a point around which the function is not even defined. It does not create contradiction, but it is rather boring.

The first time one comes across the above definition, there might be a moment of doubt: is it really necessary to have such a long definition with these symbols (ϵ, δ)? Why does this definition work? How useful it is? Although these are certainly good questions, it is impossible to make a list of answers to everything. In this case, a safer way to proceed is to **trail and error**; you should try to form your own ideas about this definition and test it out on examples to check if it is coherent with the reality.

So I recommand at this point to do some exercises.[5] I'll start with the above two examples.

5: I want to say, quite some number of exercises, or a lot of exercises.

Example 5.1.3 (The absolute value function is continuous at 0) By looking at the graph of the absolute value, we guess that $|x|$ goes to 0 as x goes to 0, that is

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

To show this, we must show that, for all $\epsilon > 0$, there exists a $\delta > 0$ such that, for all $x \in \mathbb{R}$ such that $|x - 0| < \delta$, we have $||x| - L| < \epsilon$ with $L = 0$. The last part rewrites as for all $|x| < \delta$, we have $|x| < \epsilon$. Then for any given $\epsilon > 0$, the choice $\delta = \epsilon$ suffices: in this case, we have $|x| < \epsilon$ implies $|x| < \epsilon$, which is rather obvious.

The following is an (overdetailed) illustration of a proof of discontinuity.

Example 5.1.4 (The Heaviside function is **not** continuous at 0) By looking at the graph of the Heaviside function, we want to prove that the limit of $H(x)$ as x approaches 0 does not exist. Indeed, it is hard to decide between the values $0, \frac{1}{2}, 1$ as the limit and we will see that none of them works.

To prove this, we must show that, for all $L \in \mathbb{R}$, L cannot be the limit of $H(x)$ as x goes to 0. So for any $L \in \mathbb{R}$, we want to show that there exists an $\epsilon > 0$, such that for all $\delta > 0$, we have, for some $x \in \mathbb{R}$, $|x - 0| < \delta$ and $|H(x) - L| > \epsilon$.

The difficulty is to choose the ϵ. Let us do an analysis of the situation. For any $\delta > 0$, the set of values that $H(x)$ takes, with $|x| < \delta$, is **exactly** $\{0, \frac{1}{2}, 1\}$. The quantity $|H(x) - L|$ measures the distance between $H(x)$ and L. But for whatever L is, one of the numbers in $\{0, \frac{1}{2}, 1\}$ must be at least $\frac{1}{3}$ away from L. In other words, ϵ should be of the order of the "jump" one observes at the point of discontinuity.

So take $\epsilon = \frac{1}{4}$ whatever L is. If $L \geq \frac{1}{2}$, for all $\delta > 0$, consider $x = -\frac{\delta}{2}$: we have $|x| < \delta$ and $H(x) = 0$, so that $|H(x) - L| = |L| \geq \frac{1}{2} > \epsilon$. If $L < \frac{1}{2}$, for all $\delta > 0$, consider $x = \frac{\delta}{2}$: we have $|x| < \delta$ and $H(x) = 1$, so that $|H(x) - L| = |1 - L| > \frac{1}{2} > \epsilon$. So in all cases, we have proven the discontinuity of H at $x = 0$, namely:

$$\forall L \in \mathbb{R}, \exists \epsilon > 0, \forall \delta > 0, \exists x \in \mathbb{R}, |x - 0| < \delta \text{ and } |H(x) - L| > \epsilon.$$

Now, a little bit of logic. Let us write the definition above as

$$\exists L \in \mathbb{R}, \forall \epsilon > 0, \exists \delta > 0, \forall x \in S, (|x - a| < \delta) \Longrightarrow (|f(x) - L| < \epsilon).$$

This means that, if you want to prove some function f has a limit at some point a, the first step to do is usually...guess what the limit L is![6] For this part, you should just trust your intuition and experience. Very often, when the function is defined at the point a, it would be natural to consider first $L = f(a)$.

6: You first have to guess if the function f has a limit at a or not...

Once you have made up your mind about the value L, it becomes quite automatic after some practise: you should write down the above definition, which is the same in any situation except for the last term $|f(x) - L| < \epsilon$. The game consists of giving an upper bound of $|f(x) - L|$ using expressions involving $|x - a|$, since we restrict the value of $|x - a|$ to be smaller than δ. We will see more examples in the following.

Now, if you want to show that a function does **not** have a limit at a, you should consider the negation of the above definition, which requires you to prove something that starts with

$$\forall L \in \mathbb{R}, \exists \epsilon > 0, \forall \delta > 0, \exists x \in S, \ldots$$

The only choice to make here is ϵ: it should be the magnitude of "jump" that one observes at the discontinuity (to be cautious, take the jump as 2ϵ).

5.2 Operations on limits

First, we show that if the limit of a function f at some point x exists, then it is **unique**.

Theorem 5.2.1 (Uniqueness of the limit) *Suppose that the limit of a function f at point a, as defined in the previous section, exists. Then this limit is unique.*

Proof. Suppose that we have

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

and prove that $L = L'$.

Let $\epsilon > 0$ be arbitrary. By definition, there exists $\delta > 0$ such that, for all $|x-a| < \delta$, $|f(x)-L| < \epsilon$. Also, there exists $\delta' > 0$ such that, for all $|x-a| < \delta$, $|f(x) - L| < \epsilon$.[7]

7: Notice that, *a priori*, δ and δ' are not equal. But the ϵ can be the same for L and L'.

Consider $\min(\delta, \delta') > 0$ and any x such that $|x - a| < \min(\delta, \delta')$. We have $|x - a| < \delta$, so $|f(x) - L| < \epsilon$. Also, $|x - a| < \delta'$, so $|f(x) - L'| < \epsilon$ as well. By the triangular inequality, $|L - L'| \leq |L - f(x)| + |f(x) - L'| < 2\epsilon$. Since $\epsilon > 0$ can be arbitrary small, necessarily, $L = L'$. ☐

As you suspected: the definition of the limit is "compatible" with usual operations on functions, such as addition, multiplication, or composition.[8]

8: When one faces a division $\frac{f}{g}$, one should think of it as the multiplication of f with $\frac{1}{g}$. To define $\frac{1}{g}$ at a point a, you need to first make sure that the function g is non-zero at a. Indeed, the expression $\frac{1}{g}$ is a composition of g by the function $x \mapsto \frac{1}{x}$: the latter is not defined at $x = 0$.

Proposition 5.2.2 *Let $S \subset \mathbb{R}$, $a \in \mathbb{R}$ adherent to S and $f : S \longrightarrow \mathbb{R}$, $g : S \longrightarrow \mathbb{R}$ two real functions. Suppose that*

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

Then

1. $\lim_{x \to a} (f + g)(x) = M + L;$

2. $\lim_{x \to a} (f \cdot g)(x) = M \cdot L$.

We will explain this proposition during the lecture. Some simple corollaries:

1. $\lim_{x \to a} (f - g)(x) = M - L$;
2. $\lim_{x \to a} (\lambda \cdot f)(x) = \lambda M, \quad \forall \lambda \in \mathbb{R}$;
3. If $M = 0$ or $L = 0$, then $\lim_{x \to a} (f \cdot g)(x) = 0$.

We end this section with the composition of limits.[9]

9: We don't require to prove these results in this course. However, it goes without saying that it is a healthy exercise to try to prove all these basic results! Otherwise, you can open a textbook and try to understand the proofs yourself, or discuss on the forum.

Proposition 5.2.3 *Let $f : S \longrightarrow T$ and $g : T \longrightarrow \mathbb{R}$ be two real functions. Let a be adherent to S. Assume that*

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

and that b is adherent to T. Assume that

$$\lim_{y \to b} g(y) = L.$$

Then,

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

5.3 Continuous functions

Once we have the epsilon-delta definition of limit, it is relatively easy to define a continuous real function. Essentially:

Definition 5.3.1 *Let I be a real interval. A function $f : I \longrightarrow \mathbb{R}$ defined on I is said to be* **continuous** *on the interval I, if f has a limit at all points $a \in I$ and such that*

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

In this case, we say that f is of class $\mathscr{C}^0$ on the interval I.

Most of the elementary functions you know are continuous. By a proposition in the above section, we can also obtain continuous functions by composing these elementary functions.

Proposition 5.3.1 *A composite of continuous function is continuous. In particular, the following expression makes sense if f and g are continuous real functions (and when the composition is well-defined)*

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

We can also "glue" continuous functions together: this is the concept of piecewise continuous functions. If you are interested, click on the link to see some examples. Let us mention that the absolute value function $|\cdot|$

for example can be defined as a piecewise continuous function, but also as an elementary function using $|x| = \sqrt{x^2}$.

5.4 One-sided limit

In the previous section, we have not properly define the limit for a point a at the extremity of I, so we are going to specify those cases as well.

Definition 5.4.1 (One-sided limit) *Let $I \subset \mathbb{R}$ be within the domain of definition of f, and $a \in I$ some point.*

The right-sided limit can be rigorously defined as:

$$\forall \epsilon > 0, \exists \delta > 0, \forall x \in I, x - a \in (0, \delta) \implies |f(x) - L| < \epsilon.$$

Similarly for the left-sided limit:

$$\forall \epsilon > 0, \exists \delta > 0, \forall x \in I, x - a \in (-\delta, 0) \implies |f(x) - L| < \epsilon.$$

The above definition is particularly useful for defining limits at the end points of a real interval. Notice that the difference is that we replace the inequality $|x - a| < \delta$ by a one-sided inequality $x - a < \delta$ or $x - a > -\delta$.

There is another important case: one can define the **limit of a function at infinity**.[10] However, we decided to see this properly with the convergence of real sequences, which is the topic of the next chapter.

10: Actually, this might be more important to know then the limit at a point, and in some textbooks this comes before the current chapter.

5.5 Elementary functions

Before we start to state the main theorems concerning continuous functions, let us admit an important fact:

Lemma 5.5.1 *All elementary functions are continuous at any point where they are defined.*

Elementary functions include constant functions, logarithmic functions, exponential functions, trigonometric functions (and their inverses), hyperbolic functions (and their inverses). Sum, product and composition of finitely many elementary functions are also elementary functions.

Example 5.5.1 For all $p \in \mathbb{R}$, the function $x \mapsto x^p$ is an elementary function defined on $\mathbb{R}_{>0}$.

Proof. Write $x^p = e^{p \ln(x)}$. □

For example, if you see an expression like

$$f(x) = \frac{e^{\tan x}}{1 + x^2} \sin\left(\sqrt{1 + (\ln x)^2}\right),$$

you can say that f is continuous at its (largest possible) domain of definition (but you must study then the domain of definition of f).

5.6 Squeeze theorem

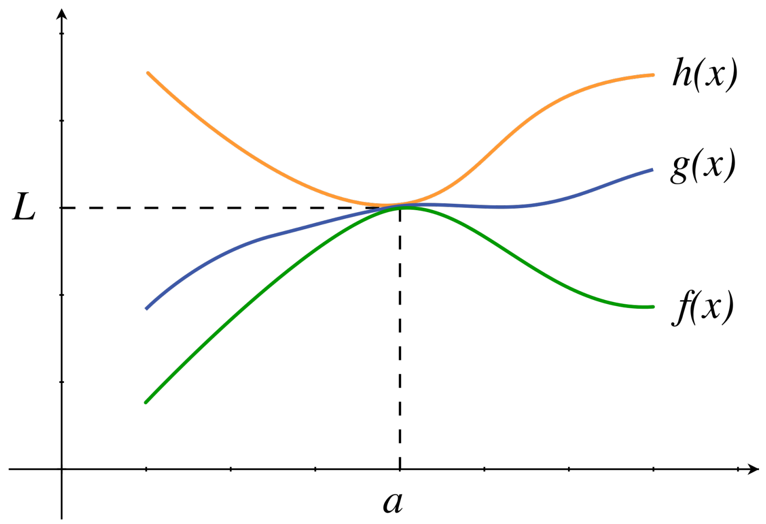


Figure 5.3: The squeeze theorem, a.k.a. the sandwich theorem or even the police theorem.

Theorem 5.6.1 (Squeeze theorem) *Let $f, g, h : S \longrightarrow \mathbb{R}$ real functions and $a \in S$. Suppose that for all x sufficiently close to a, we have*

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

Suppose that the limits of f and h at point a exist and are equal; that is,

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

Then the limit of g at point a exists and is equal to L; that is

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

In practise, we can use this theorem to

1. Establish the existence of a limit;
2. Calculate the value of this limit.

Example 5.6.1 Compute $\lim_{x \longrightarrow 0} x^2 \cos\left(\frac{2}{x}\right)$.

Solution. Let $g(x) = x^2 \cos\left(\frac{2}{x}\right)$ and $f(x) = x^2, h(x) = -x^2$. Since $|\cos(z)| \leq 1$ for all $z \in \mathbb{R}$, we have $f(x) \leq g(x) \leq h(x)$ for x sufficiently close to 0.

Since $\lim_{x \longrightarrow 0} x^2 = 0$, we have $\lim_{x \longrightarrow 0} f(x) = \lim_{x \longrightarrow 0} h(x) = 0$. By the Squeeze theorem, the limit of $g(x)$ as x goes to 0 exists and is equal to 0. □

Notice that the limit of $\cos\left(\frac{2}{x}\right)$ at the point 0 does not exist (why?).

5.7 Bolzano's theorem

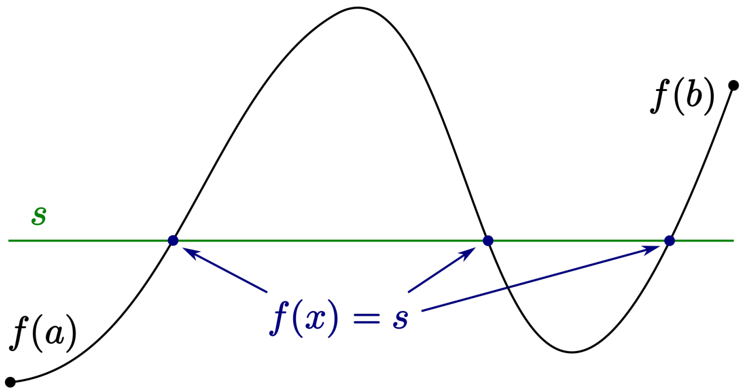


Figure 5.4: Bolzano's theorem, a.k.a. the intermediate value theorem.

Theorem 5.7.1 *Let f be a continuous function on a closed interval $[a, b]$. Let $\alpha = f(a)$ and $\beta = f(b)$. Let s be a number such that $\alpha < s < \beta$. Then there exists $c \in (a, b)$ such that $f(c) = s$.*

Notice the topological nature of this theorem: a continuous curve in the plane $\mathbb{R}^2$ can very well go from $(0, 0)$ to $(1, 0)$ without passing the point $(\frac{1}{2}, 0)$. What should be the analogy of Bolzano's theorem in this case? In some sense, in $\mathbb{R}^1$, the point $\frac{1}{2}$ seperates the real line into a "left" part and a "right" part. In $\mathbb{R}^2$, a curve that divides the plane into an "interior" part and an "exterior" part is called a Jordan curve. Bolzano's theorem for a Jordan curve might seem intuitive, but it is actually not easy to prove.

Example 5.7.1 The equation $x^3 + x - 1 = 0$ has at least one solution in the interval $(0, 1)$.

Solution. Let $f(x) = x^3 + x - 1$. Then $f(0) = -1$ and $f(1) = 1$: the function f has opposite sign at points 0 and 1. By continuity of the function f on the interval $[0, 1]$ and Bolzano's theorem, there exists $c \in (0, 1)$ such that $f(c) = 0$. The equation then has at least one solution c in the interval $(0, 1)$. □

Let us mention an important consequence of Bolzano's theorem.

Corollary 5.7.2 *If f is a strictly monotonic continuous function on an interval I, then $f^{-1} : f(I) \longrightarrow I$ is continuous.*

We don't require to prove this result in the course.[11]

11: One of the exercises of this chapter has the following result: that a surjective, increasing real function from $[a, b]$ to $[f(a), f(b)]$ is continuous. This will be a main input in the proof of this corollary.

5.8 Weierstrass' theorem

Theorem 5.8.1 *Let f be a continuous function on a closed interval $[a, b]$.*

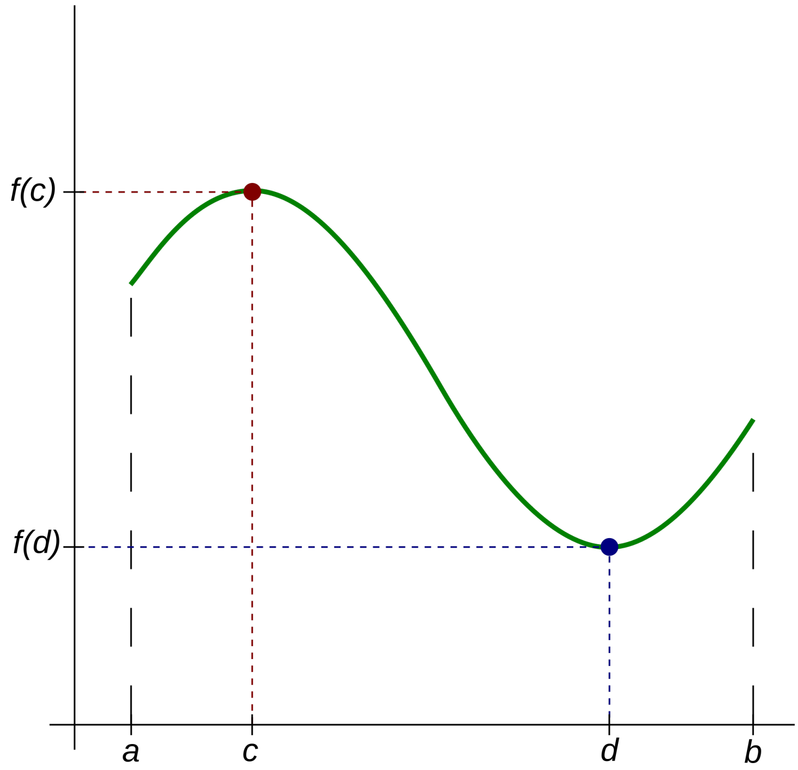


Figure 5.5: Weierstrass' theorem, a.k.a. the extreme value theorem.

Then there exists an element $c \in [a, b]$ such that $f(c)$ is the maximum of $f([a, b])$ and an element $d \in [a, b]$ such that $f(d)$ is the minimum of $f([a, b])$.

Let us give an example of combining two theorems together.

Example 5.8.1 The image of a closed bounded interval by a continuous function is a closed bounded interval.

Solution. Let $f : [a, b] \longrightarrow \mathbb{R}$ be a continuous function. By Bolzano's theorem, the image $f([a, b])$ verifies the intermediate value property, so it is an interval of $\mathbb{R}$. By Weierstrass' theorem, it has a maximum and a minimum, so it is bounded and closed on both end points. □

5.9 Story time: the origins of rigorous Calculus

> All truth passes through three stages. First, it is ridiculed. Second, it is violently opposed. Third, it is accepted as being self-evident.
>
> – Arthur Schopenhauer

This chapter is already too long: stories of this type are best summerized in SMBC.[12]

12: SMBC=Saturday Morning Breakfast Cereal. Similar comics: xkcd, phdcomics, abstruse goose.

5.10 Exercises

Exercise 5.1 (Homework) Show by the (ϵ, δ)-definition that

1. The function $x \longmapsto x^2$ is continuous at -1;
2. The function $x \longmapsto \lfloor x \rfloor$ is discontinuous at 1.[13]

13: Recall that the **floor function** is defined as $\lfloor x \rfloor = k$ where $k \in \mathbb{Z}$ is the unique integer such that $k \leq x < k+1$.

Exercise 5.2 Consider the function $f : x \longmapsto \frac{1}{x}$ defined on $\mathbb{R} \setminus \{0\}$. Show that

1. The function f is continuous on $\mathbb{R} \setminus \{0\}$;
2. The function f is not continuous at 0.

Exercise 5.3 Let $f : \mathbb{R} \longrightarrow \mathbb{R}$, $g : \mathbb{R} \longrightarrow \mathbb{R}$ be two continuous functions. Show that $\max(f(x), g(x))$ is continuous on $\mathbb{R}$.

Exercise 5.4 Some exercises on the Squeeze theorem.

You can use the inequalities $\cos(x) \leq \frac{\sin(x)}{x} \leq 1$ for x close enough to 0.

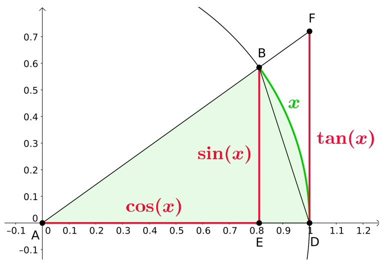


Figure 5.6: Comparison between $\sin(x)$, x and $\tan(x)$.

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

Exercise 5.5 Some exercises on the intermediate value theorem.

1. Let f be a continuous function such that $f(0) < 0$ and $f(1) > 1$. By considering the function $g(x) = f(x) - x$, show that $f(z) = z$ for some $z \in (0,1)$.
2. Let $f : (0,1) \longrightarrow \mathbb{R}$ be a continuous function such that $\lim_{x \longrightarrow 0} f(x) = 0$ and $\lim_{x \longrightarrow 1} f(x) = 1$. Show that $f(z) = \frac{1}{2}$ for some $z \in (0,1)$.

Exercise 5.6 Some exercises on the extreme value theorem.

1. Let $f : [a,b] \longrightarrow \mathbb{R}$ be a continuous function. Show that
$$\sup f([a,b]) = \sup f((a,b)).$$
2. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a continuous, 1-periodic function, that is, for all $x \in \mathbb{R}$, $f(x+1) = f(x)$. Show that f is bounded.

Exercise 5.7 (Homework) Let $f : [a,b] \longrightarrow \mathbb{R}$ be increasing and surjective from $[a,b]$ to $[f(a), f(b)]$. Show that f is continuous.[14]

14: This is not an easy exercise at first; if you cannot solve it, try to come up with some pictures or arguments that support the claim!

Exercise 5.8 ($\star$) In this exercises, we consider several notions stronger than the local continuity. If I is an interval of $\mathbb{R}$ and $f : I \longrightarrow \mathbb{R}$ is a real function, we say respectively that f is uniformly continuous, Lipschitz continuous or α-Hölder continuous with $\alpha \in (0,1)$ on I if

1. (Uniform continuity)
$$\forall \epsilon > 0, \exists \delta > 0, \forall x, y \in I, |x - y| < \delta \Longrightarrow |f(x) - f(y)| < \epsilon.$$
2. (Lipschitz continuity)
$$\exists K > 0, \forall x, y \in I, |f(x) - f(y)| \leq K|x - y|.$$

3. (Hölder continuity)

$$\exists C, \alpha > 0, \forall x, y \in I, |f(x) - f(y)| \leq C|x - y|^{\alpha}.$$

Prove that any of the above implies the local continuity.

Exercise 5.9 (*) Historically, the intermediate value property (i.e. Bolzano's theorem) was suggested (and rejected) as the definition of a continuous function. In this exercise, we will study a counter-example: that is, we study a function that satisfies the intermediate value property, but is not continuous in the (ϵ, δ)-sense.

Consider the function $f : [0, \infty) \longrightarrow [-1, 1]$ defined as

$$f(x) = \begin{cases} \sin\left(\frac{1}{x}\right) & \text{if } x > 0; \\ 0 & \text{if } x = 0. \end{cases}$$

Prove that the limit of $f(x)$ as x goes to 0 does not exist; then prove that f satisfies the intermediate value property.[15]

15: A function that satisfies the intermediate value property is called a **Darboux function**.

[Week IV]
What are...sequences?

Sequences: definitions and properties

6

A sequence is a collection of objects ordered by integers. In this course, we are mostly interested in real sequences: most often they appear in the form of $(u_0, u_1, u_2, \dots)$, but we can give a more formal definition.[1]

1: The definition is not very useful for this course: it is only a remainder that a sequence is a function defined on an ordered discret set.

Definition 6.0.1 *A real sequence is formally a function $u : \mathbb{Z} \longrightarrow \mathbb{R}$. We will use the notation u_n instead of $u(n)$:*

$$\forall n, \quad u_n = u(n).$$

Usually, we restrict the function u to the set of departure $\mathbb{Z}_{>0}$ or $\mathbb{Z}_{\geq 0}$.

6.1 Classical sequences

We recall some classical sequences.[2]

2: If one is really serious about notations, it is better to distinguish $\{u_n\}_{n\geq 0}$, which is the notation for a sequence, with u_n, which is just one term of the sequence. However, in practice, we are used to interchange them for simplicity.

Definition 6.1.1 (Constant sequence) *A sequence u_n such that $u_n \equiv c$ for all $n \geq 0$ is called a constant sequence.*

Definition 6.1.2 (Sequence with finitely many non-zero terms) *A infinite sequence u_n such that only finitely indices n satisfy $u_n \neq 0$ is called a sequence with finitely many non-zero terms. We can identify those sequences with finite sequences.*

In a more mathematical way, those sequences are characterized by the property that there exists some $N_0 \in \mathbb{N}$ such that for all $n \geq N_0$, we have $u_n = 0$. In a compact way,

$$\exists N_0 \geq 0, \forall n \geq N_0, u_n = 0.$$

The above sequences is also called **almost zero sequences**.

Definition 6.1.3 (Arithmetic sequence) *Let $a, b \in \mathbb{R}$. The sequence u_n such that $u_0 = a$ and $u_{n+1} = u_n + b$ is called an arithmetic sequence.*

One can prove by induction that for all $n \geq 0$, $u_n = a + bn$.

Definition 6.1.4 (Geometric sequence) *Let $q, r \in \mathbb{R} \setminus \{0\}$. The sequence u_n such that $u_0 = q$ and $u_{n+1} = q \cdot u_n$ is called a geometric sequence.*

One can prove by induction that for all $n \geq 0$, $u_n = q \cdot r^n$.

6.2 Limit of a sequence

Of course, it would be difficult to say that a sequence is continuous. However, we can still define the limit of a sequence **at infinity**, using the epsilon-delta formalism.

Heuristically, we want to say that a sequence converges to a limit L if this sequence gets closer and closer to L and not to any other number. Some might say that the sequence approaches L as the index increases. A precise definition that translates the above phrases is the following:

Definition 6.2.1 (Convergence of a sequence) *We say a sequence $\{u_n\}_{n\geq 0}$ converges to $L \in \mathbb{R}$ as n goes to infinity, if for all $\epsilon > 0$, there exists some $N_0 \in \mathbb{N}$, such that for all $n \geq N_0$, we have $|u_n - L| < \epsilon$.*

In a compact way,

$$\exists L \in \mathbb{R}, \forall \epsilon > 0, \exists N_0 \in \mathbb{N}, \forall n \geq N_0, |u_n - L| < \epsilon.$$

We denote the above, as in the case of function, by[a]

$$\lim_{n \to \infty} u_n = L.$$

[a] Or more loosely, $u_n \longrightarrow L$.

Compare this to the definition of the limit of a function f at a point a:

$$\exists L \in \mathbb{R}, \forall \epsilon > 0, \exists \delta > 0, \forall x \in (a-\delta, a+\delta), |f(x) - L| < \epsilon.$$

Here is a quick application of the above definition:

Proposition 6.2.1 (A convergent sequence is bounded) *Let $\{u_n\}_{n\geq 0}$ be a convergent sequence of limit $L \in \mathbb{R}$. Then $\{u_n\}_{n\geq 0}$ is bounded (i.e. bounded from above and from below).*

Proof. Let us prove that $\{u_n\}$ is upper bounded: the lower-boundedness will be similar. Let L be the limit of u_n and take $\epsilon = 1$. Then there exists some $N \in \mathbb{N}$ such that for all $n \geq N$, $u_n \in (L-1, L+1)$. In particular, the part of u_n with $n \geq N$ is upper bounded by $L + 1$.

Before the term u_N, there are only finitely many terms: let M be their maximum (which exists by finiteness). Then the sequence u_n is upper bounded by the maximum of the above bounds $\max(M, L+1)$. □

A remark on the notation: we call a sequence **divergent** if it is not convergent (even if the sequence is bounded). In this course, we avoid using the imprecise phrase that "a sequence converges to ∞"; when we say that a sequence converges, it means that its limit is a real number.

The Squeeze theorem for functions survives for a sequence (notice that in the statement of this theorem for a function, no continuity is assumed: it only concerns the definition of the limit).

Theorem 6.2.2 (Squeeze theorem for sequences) *Suppose that we have three sequences* $u_n \leq v_n \leq w_n$ *and that*

$$\lim_{n \to \infty} u_n = \lim_{n \to \infty} w_n = L.$$

Then $\lim_{n \to \infty} v_n = L$ *(recall that this notation implicitely says that the limit exists).*

Of course, this theorem works if the comparison is only valid after "time" N_0, that is, for all $n \geq N_0$ with some $N_0 \in \mathbb{N}$.[3]

We add a new theorem to our arsenal: it has similar forms for the function case, which we will review in the next section.

3: In some sense, the question of convergence is independent of the first terms of the sequence, just like the limit of a function at some point a is independent of what happens outside the interval $(a-\delta, a+\delta)$ for any $\delta > 0$.

Theorem 6.2.3 (Monotone convergence theorem for sequences) *Suppose that* u_n *is a increasing sequence, that is, for all* $n \in \mathbb{N}$, $u_n \leq u_{n+1}$. *Then if* u_n *is upper bounded, that is, if there exists some* $M \in \mathbb{R}$ *such that* $u_n \leq M$ *for all* n, *then*

$$\lim_{n \to \infty} u_n$$

exists. Furthermore, this limit is smaller or equal to M.

Again, the theorem holds even if the monotonicity is only present by removing a finite number of terms.

6.3 The extended real line

Similarly to the convergence of sequences, one can define the limit of a real function at infinity:

Definition 6.3.1 *Let* f *be a function defined on some interval* $[a, \infty)$. *We say that* $f(x)$ *converge* L *when* x *goes to infinity, if for all* $\epsilon > 0$, *there exists some* $A \in \mathbb{R}$ *such that, for all* $x \geq A$, *we have* $|f(x) - L| < \epsilon$.

Otherwise said, we write

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

if and only if

$$\forall \epsilon > 0, \exists A \in \mathbb{R}, \forall x \geq A, |f(x) - L| < \epsilon.$$

It enjoys the same property as the limit of a function at a point $a \in \mathbb{R}$, except for the composition since f is only defined at each $x \in \mathbb{R}$. We will return to this particular point in a section below.

Theorem 6.3.1 (Monotone convergence theorem for function) *Let* f *be an upper bounded increasing function defined on some interval* $[a, \infty)$.

Then the following limit exists:

$$\lim_{x \longrightarrow \infty} f(x).$$

Notice that no requirement on the continuity is needed in the above theorem. The same idea applies when infinity is replaced by the end point of an interval in the above:

Theorem 6.3.2 *Let f be an upper bounded increasing function defined on some interval $[a, b)$. Then the following (one-sided) limit exists:*

$$\lim_{x \longrightarrow b^-} f(x).$$

I invite you to write a proper version for the limit at the left end-point of the interval, since the signs can get tricky!

We add another theorem to the list:

Theorem 6.3.3 (Comparison theorem) *Let $f, g : S \longrightarrow \mathbb{R}$ and let a be adherent to S. Then if $g(x) \leq f(x)$ for all x sufficiently close to a, and the $\lim_{x \longrightarrow a} f(x) = L$ and $\lim_{x \longrightarrow a} g(x) = M$, then $M \leq L$.*

It is very important to observe that the inequality between the limit is always a **large** inequality.

The comparison theorem also holds in the case of a limit at infinity, and in particular, for sequences.[4]

4: I leave it as an exercise for you to write down the statement, since I believe that you already get the idea.

Remark 6.3.1 Sometimes, we will see expressions of type

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

A formal interpretation of this is that

$$\forall A \in \mathbb{R}, \exists \delta > 0, \forall x \in]a - \delta, a + \delta[, f(x) \geq A.$$

The case where a is replace by ∞

$$\lim_{x \longrightarrow \infty} f(x) = \infty$$

is defined similarly:

$$\forall A \in \mathbb{R}, \exists M \in \mathrm{r}, \forall x \geq M, f(x) \geq A.$$

6.4 Subsequence

Intuitively, a subsequence of a sequence is obtained by removing certain terms without changing the order of the remaining terms.

Theorem 6.4.1 *A sequence is convergent if and only if every subsequence is convergent. In this case, the limits of all subsequences are the same.*

This criterion gives a simple way of proving divergence of sequences, without using the (ϵ, δ)-definition.

Example 6.4.1 The sequence $u_n = (-1)^n$ is not convergent.

Proof. The sequence has two convergent subsequences: $u_{2k} \equiv 1$ and $u_{2k+1} \equiv -1$. They are both convergent but have different limits: this shows that the sequence u_n cannot be convergent. □

Another absolutely fundamental result, not required in this course, is the Bolzano-Weierstrass theorem:[5]

5: Historically, it was used by Bolzano in his proof of the intermediate value theorem. It has applications to economics.

Theorem 6.4.2 (Bolzano-Weierstrass) *Each bounded real sequence has a convergent subsequence.*

There are different nice proofs of this result that you can find on the internet.

6.5 Sequential characterization of continuity

There are a lot of interactions between sequences and functions, too many that for this course, we can only give a sneap peak into it.

Theorem 6.5.1 *Let $f : I \longrightarrow \mathbb{R}$ be a function defined on some interval I, and let $a \in \mathbb{R}$ be adherent to I. Then f is continuous at a if and only if, for all sequence u_n converging to a, $f(u_n)$ converges to $f(a)$.*

This is extremely useful for someone familiar with sequences by not with functions. More seriously, in most applications on sequences, we will use it in the following form:

Corollary 6.5.2 *If u_n converges to a and if a function f is continuous at a, then $f(u_n)$ converges to $f(a)$.*

Remark 6.5.1 This theorem is important in the study of convergence of recurrent sequences. A recurrent sequence is of the form $u_{n+1} = f(u_n)$ for all $n \in \mathbb{N}$ with some initial condition $u_0 = a$. We are interested in understanding if the limit of u_n exists, and if yes, can we calculate it.

In the case where f is continuous, we can apply the above theorem to get the following: if u_n converges to $L \in \mathbb{r}$, then $f(L) = L$. This shows that the potential limits of u_n are fix points of the function f. Of course, u_n can also be divergent even if f has some fix points!

6.6 Indeterminate forms at infinity

Let me state it in bold: this is a very important section for other areas of science![6]

6: For example, with the Covid, this sections explains why a linear growth in cases is a significantly more controllable situation than an exponential growth in cases.

Let me just state the result informally and discuss it:

Proposition 6.6.1 *As x goes to infinity, we have, for all $p > 0$,*

$$\ln(x) \ll x^p \ll e^x.$$

In other words, for example,[7],

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

7: However, in this course, it is forbidden to write

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

These results can be properly proven, but we skip the details. By a change of variables $u = \frac{1}{x}$, we can also restate it near 0:

Proposition 6.6.2 *As x goes to 0^+ (we are taking the right-limit at 0 because of the $\ln$ function), we have, for all $p > 0$,*

$$-\ln(x) \ll x^{-p} \ll e^{-\frac{1}{x}}.$$

In other words, for example,

$$\lim_{x \to 0^+} x \ln(x) = 0.$$

In case of doubt, a useful of recovering these formulas is via the **L'Hôpital's rule**.[8]

8: At this stage, it is more of a mnemotechnique.

Proposition 6.6.3 (L'Hôpital's rule) *In a simplified manner, L'Hôpital's rule states that when you have to calculate*

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

but both the limit of f and g at point a are 0 (or ∞), then you can replace it by

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

In principle, apart from some pathological cases, you can reiterate this procedure until you find a valid limit.

6.7 Exercises

Exercise 6.1 Determine the limits of the following sequences:

1. $u_n = \frac{2+(-1)^n}{n}$;

2. $v_n = \frac{3n+2}{n^2-5}$;
3. $w_n = n^{\frac{1}{n}}$.

Exercise 6.2 (Homework) Let $\{u_n\}_{n\geq 0}$ be a sequence such that for all $n \geq 0$, $u_n \in \{-1, 1\}$. Prove that it is impossible to have

$$\lim_{n\longrightarrow\infty} u_n = 0.$$

Exercise 6.3 Prove that none of the following sequences has a limit:

1. $u_n = \sqrt{n}$
2. $v_n = \cos(\frac{n\pi}{2})$;
3. $w_n = \sin(n)$.

You can apply (ϵ, δ) or find a quicker way. The last one is difficult.

Exercise 6.4 (Homework) We want to show that if $\{a_n\}_{n\geq 0}$ is convergent to 0 if and only if $\{|a_n|\}_{n\geq 0}$ is convergent to 0.

1. Use the (ϵ, δ)-definition;
2. Use the Squeeze theorem;
3. Use yet another method.

Exercise 6.5 In this exercise, we study a sequence s_n defined in the following way.[9] Let a_n denote the geometric sequence $a_n = q^n$ for $|q| < 1$, and define

$$s_n = \sum_{k=0}^{n-1} a_k = 1 + q + \cdots + q^{n-1}.$$

Show that s_n converges to the limit $\frac{1}{1-q}$.

You can use an exercise from the first week.

9: More precisely, as it is defined as a partial sum, S_n is usually called a series.

Exercise 6.6 (*) In this exercise, we study the Cesàro summation. Let $\{a_n\}$ be an arbitrary real sequence and define its partial sum $\{s_n\}$ as

$$s_n = \sum_{k=0}^{n-1} a_k.$$

We call the sequence a_n **Cesàro summable** if the sequence $\{M_n\}_{n\geq 1}$ of the arithmetic means of s_n,

$$M_n = \frac{1}{n}\sum_{k=0}^{n-1} s_k = \frac{1}{n}(s_0 + s_1 + \cdots + s_{n-1})$$

converges.

Show that

1. The sequence $b_n = (-1)^n$ is not convergent, but Cesàro summable;
2. A convergence sequence is always Cesàro summable.

Exercise 6.7 (*) Show that a sequence is convergent toward L if and only if every subsequence has its own subsubsequence that converges to the same limit L. Is the requirement that all subsubsequences converge to the same limit necessary, or it is sufficient that everyone of them converges?

Exercise 6.8 (*) Let $\{a_n\}_{n\geq 0}$ be a real sequence such that

1. For all $n > 0$, $a_n \geq 0$;
2. $\lim_{n\to\infty} a_n = 0$;
3. For all $n > 0$,

$$a_{n-1} + a_{n+1} - 2a_n \geq 0.$$

Show that $\lim_{n\to\infty} n(a_n - a_{n+1}) = 0$.

[Week V&VI]
What are...derivatives?

7 Calculation of derivatives

Roughly speaking, the derivative of a function at a given point is defined by looking at its infinitesimal rate of change, otherwise said, the slope of its tangent at this point. We first review some basic formulas and applications of this concept.

N.B. This chapter is **informal** and the focus here is the applications of derivative; rigorous justifications will be reviewed in the next chapter.

7.1 Product rule and chain rule

Let I, J be two intervals of $\mathbb{R}$, and let $f : I \longrightarrow J$ and $g : J \longrightarrow \mathbb{R}$ be two derivable functions.[1] Recall the formulae:

1: For now, a derivable function just means that you know how to calculate the derivative of the function, and that the result makes sense (e.g. the derivative is not ∞).

Proposition 7.1.1 *The following formulas hold as long as they make sense:*

1. *"Product rule", or **Leibniz rule**:*

$$(f \cdot g)'(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x);$$

2. *"Chain rule",*[2]

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

where the last term is the derivative of g evaluated at the point $f(x)$.

3. *"Quotient rule", which is rather a combinaison of the two previous rules:*[a]

$$\left(\frac{g}{h}\right)'(x) = \frac{g'(x)h(x) - g(x)h'(x)}{h(x)^2}.$$

[a] Personally, I never remembers this one and I derive it everytime from the two others...

2: Or "change of variables", but backward. Anyways, this is very important for the theory of integration.

Let us start by seeing an application of the **product rule**.

Example 7.1.1 Let us apply the quotient rule to calculate the derivative of the tangent function $\tan = \frac{\sin}{\cos}$. Recall that the tangent function is defined whenever $\cos(x) \neq 0$, i.e. if

We have

$$\tan'(x) = \left(\frac{\sin}{\cos}\right)'(x) = \frac{\sin'(x)\cos(x) - \sin(x)\cos'(x)}{\cos^2(x)} = \frac{1}{\cos^2(x)}.$$

It is more convenient to remember that

$$\tan'(x) = 1 + \tan^2(x).$$

We can apply the **chain rule** to get a useful formula relating the derivatives of a function and its inverse. Although the proof will only be reviewed in the next chapter, here's a quick way of recovering the formula.

Corollary 7.1.2 *From the relation $f(f^{-1}(x)) = x$, we get, by deriving both sides and the chain rule,*[a]

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

When $(f^{-1})'(x) \neq 0$, we have the formula[b]

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

[a] In a formal proof, one has to justify the derivability of both sides.
[b] Again, notice that $f'(f^{-1}(x))$ is the derivative of f evaluated at the point $f^{-1}(x)$.

Here's an explicit application of the corollary above.

Example 7.1.2 Consider the inverse function arctan of tan. Recall that arctan : $(-\pi, \pi) \longrightarrow \mathbb{R}$.

Using two formulas above, we have that, for all $x \in (-\pi, \pi)$,

$$\arctan'(x) = \frac{1}{\tan'(\arctan(x))} = \frac{1}{1 + \tan^2(\arctan(x))} = \frac{1}{1 + x^2}.$$

7.2 Calculation of limits

The derivative $f'(a)$ is defined as the infinitesimal rate of change of a function f at the point a. Formally,

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

given that the limit exists.

We can use this definition in the other direction.

Example 7.2.1 We have

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

Indeed, by definition of the derivative of sin at point 0,

$$\sin'(0) = \lim_{x \longrightarrow 0} \frac{\sin(x) - \sin(0)}{x - 0} = \lim_{x \longrightarrow 0} \frac{\sin(x)}{x}.$$

The result follows from

$$\sin'(0) = \cos(0) = 1.$$

A more sophisticated version is the **L'Hôpital's rule**: we postpone its proof until the next chapter. Just to recall how to apply this rule:

Example 7.2.2 We can calculate

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

Indeed, since this is an indeterminate form, by (successive applications of) L'Hôpital's rule, we have

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

There will be another explanation of the above example by **Taylor expansion**, although it is very unlikely that we will have time to cover it in this course.

7.3 Monotone functions

We use the following result:[3]

3: This can be seen as a corollary of the mean value theorem, see next chapter.

Theorem 7.3.1 *Let $f : I \longrightarrow \mathbb{R}$ be a derivable function defined on an interval I. If $f'(x) \geq 0$ for all $x \in I$, then f is increasing.*

Also, if $f'(x) > 0$ everywhere, then f is strictly increasing.

This theorem can be very helpful to compare functions. Notice that it only gives a sufficient condition at this stage: a more general situation will be discussed later.

Example 7.3.1 Show that for all $x \in [0, \frac{\pi}{2})$,

$$\sin(x) \leq x \leq \tan(x).$$

(Simplified) solution. We already know that $\sin(0) = 0 = \tan(0)$. Furthermore,

$$\cos(x) = \sin'(x) \leq 1 \leq \tan'(x) = 1 + \tan^2(x).$$

Since all functions are derivable on $[0, \pi)$, this yields the result. □

Maybe a little bit more surprisingly, this theorem can be very helpful to establish **equality** between two expressions.

Example 7.3.2 Consider the hyperbolic tangent function

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

We denote by artanh : $(-1, 1) \longrightarrow \mathbb{R}$ its inverse function. Show that

$$\operatorname{artanh}(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right).$$

(Simplified) solution. One can calculate the derivatives of both sides: they both yield $x \mapsto \frac{1}{1-x^2}$.

Since both sides are derivable on $(-1, 1)$ and that they are equal when $x = 0$, they must coincide over the whole interval $(-1, 1)$. □

7.4 Local minimum and local maximum

We use the following result:[4]

4: This can be seen as a corollary of Rolle's theorem, see next chapter.

Theorem 7.4.1 *Let $f : I \longrightarrow \mathbb{R}$ be a derivable function defined on an interval I. If $a \in I$ is a local extremum of f, then $f'(a) = 0$.*

A precise definition of the local maximum (idem for the minimum) is the following: we call a is a local maximum of f if there exists some $\delta > 0$, such that for all $x \in (a - \delta, a + \delta)$, we have $f(x) \leq f(a)$.

Example 7.4.1 Let $f(x) = x^2 + x + 1$. Find the image of $[-1, 1]$ by f.

Solution. Already, by Bolzano's theorem and Weierstrass' theorem, the image of $[-1, 1]$ by f is a closed bounded interval $[a, b]$ of $\mathbb{R}$.

Let us study the minimum and the maximum of f on $[-1, 1]$. For this, we first study the equation $f'(x) = 0$. Since $f'(x) = 2x + 1$, $f'(x) = 0$ if and only if $x = -\frac{1}{2}$. So the potential minimum and maximum of f happens at $a \in \left\{-1, -\frac{1}{2}, 1\right\}$ (notice that we include the end points).

Explicitly, $f(-1) = 1$, $f\left(-\frac{1}{2}\right) = \frac{3}{4}$ and $f(1) = 3$. So we can conclude that the image of f is $\left[\frac{3}{4}, 3\right]$. □

An interesting application of Rolle's theorem is on the number of real roots of a real polynomial function.

Example 7.4.2 Let $n \geq 1$ and $P(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ be a real polynomial function of degree n, i.e. $a_n \neq 0$. Then P has at most n different roots in $\mathbb{R}$, i.e. the equation $P(x) = 0$ has at most n different solutions.

Simplified proof. We can do this by induction.

For $n = 1$, this is true since $P(x)$ is then a linear function with non-zero slope: it has one solution $x = -\frac{a_0}{a_1}$.

Suppose this for $n \geq 1$ and prove it for rank $n+1$ by contradiction. Suppose that $P(x)$, of degree $n+1$, has (at least) $n+2$ different solutions, arranged in increasing order as $x_1 < x_2 < \cdots < x_{n+2}$. Since a polynomial function is derivable everywhere and that for all $k \in [1, n+1]$, $P(x_k) = P(x_{k+1}) = 0$, Rolle's theorem tells us that P' takes the value 0 at least once at y_k in (x_k, x_{k+1}). This is true for every $k \in [1, n+1]$, so we have found $n+1$ different solutions for $P'(y) = 0$, namely $y_1 < y_2 < \cdots < y_{n+1}$. But P' is a polynomial of degree n, by induction hypothesis, it can have at most n diffrent solutions. This is a contradiction.

By induction, a real polynomial function of degree n has at most n different real roots. □

It is a good place to finish this chapter by spoiling you about the **fundamental theorem of algebra**:[5]

5: In some sense, the mean value theorem is intimately related to the **fundamental theorem of analysis**.

Theorem 7.4.2 (D'Alembert-Gauss) *Every non-constant single-variable polynomial with complex coefficients has at least one complex root. Equivalently, every non-zero, single-variable, degree n polynomial with complex coefficients has, counted with multiplicity, exactly n complex roots.*

7.5 Exercises

Exercise 7.1 Calculate the first derivatives of the following functions:

1. $\mathbb{R}_{>0} \longrightarrow \mathbb{R}, \quad x \longmapsto (x\ln(x) - x)$;
2. For $c \neq 0$, $\quad \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto e^{cx}\left(\frac{x^2}{c} - \frac{2x}{c^2} + \frac{2}{c^3}\right)$;
3. $\mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto 2^x$.

Exercise 7.2 Give the domain and codomain of the function arccos, then calculate its derivative. Then, calculate the limit

$$\lim_{x \longrightarrow 0} \frac{\frac{\pi}{2} - \arccos(x)}{x}.$$

Exercise 7.3 Show that for all $x > 0$,

$$\operatorname{artanh}(x) \geq x + \frac{x^3}{3}.$$

Can you replace $\geq$ by $>$ in the above equation?

Exercise 7.4 Use derivatives to prove that, for all $x \in [-1, 1]$,

$$\arccos(x) + \arcsin(x) = \frac{\pi}{2}.$$

Exercise 7.5 Show that $f(x) = \frac{x}{2} + \sin(x)$ on the interval $[0, 4]$ is positive. What is the range of f?

Exercise 7.6 Let $n \geq 2$ and $p, q \in \mathbb{R}$. Show that the polynomial function

$$g(x) = x^n + px + q$$

has at most three real roots.

Exercise 7.7 (∗) Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a positive $\mathscr{C}^1$ function such that for all $x \in \mathbb{R}$, $f'(x) \leq -f(x)$. Prove that[6]

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

6: This is a very special case of **Grönwall's lemma**.

Exercise 7.8 (∗) Let f be derivable on the interval $[0, 1]$. Suppose that

$$f(0) = f(1) = f'(0) = 0.$$

Show that the tangent of (the graph of) f at some point $x \in (0, 1)$ passes by the origin $(0, 0)$.

Hint: you can consider the function $\frac{f(x)}{x}$.

Differentiation: rigorous definition 8

In physics, velocity of an object is defined as the rate of change of its position, and acceleration is the rate of change of its velocity. The term "rate of change" can be rigorously defined by the concept of derivation in mathematics. In this chapter, we prove the formulas used previously.

8.1 Derivable functions

We first give two different definitions of a differentiable function.[1] The first one is the classical one; the second one is less useful for now, but it will become important later.

1: Sometimes, I will say derivable functions as well, but that is not the most common terminology.

Definition 8.1.1 *Suppose that we have a real function f defined on a non-trivial interval I. We say that f is* ***differentiable*** *at $x \in I$ if the following limit (called* ***Newton quotient****) exists*

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

The limit is then called the ***derivative*** *of the function f at x, and is denoted $f'(x)$.*

This definition is a rigorous formulation of the expression "infinitesimal rate of change". Let us formulate it a little differently.

Definition 8.1.2 *The function f is differentiable at point $x \in I$ if and only if there exists some number L, a function $\varphi(h)$ defined for arbitrary small value of h with $\varphi(0) = 0$ and*

$$\lim_{h \to 0} \frac{\varphi(h)}{h} = 0$$

and such that

$$f(x+h) = f(x) + Lh + \varphi(h)$$

for all h sufficiently close to 0.

With the so-called **Landau's notation**,[2] any function $\varphi(h)$ as in the above definition is written as $o(h)$. So the above equation becomes

2: We will not discuss this in detail in this course, but in general, $f(x) \underset{x \to a}{=} o(g(x))$ means that $\lim_{x \to a} \frac{f(x)}{g(x)} = 0$. Similar definition holds when one replaces a by ∞.

$$f(x+h) = f(x) + Lh + o(h), \quad \text{or} \quad \frac{f(x+h) - f(x)}{h} = L + \frac{o(h)}{h}.$$

By definition, $\lim_{h \to 0} \frac{o(h)}{h} = 0$, so we can identify L with $f'(x)$.

The idea of the second definition is that, near a point a where f is differentiable, the function looks like a linear function $x \mapsto f(x) + f'(x)(x - a)$

3: This is actually a first example of a Taylor expansion.

with an error of order $o(|x - a|)$.[3] Another way of writing the function $\varphi(h)$ is by putting $\varphi(h) = h \cdot \psi(h)$ with $\lim_{h \to 0} \psi(h) = 0$.

Proposition 8.1.1 *If a function f is differentiable at some point x, it is continuous at the point x.*

The proposition follows from the second definition by taking the $h \to 0$ limit. In particular, the function f is also bounded near a point x where f is differentiable.

8.2 Operations (with proofs)

It is not hard to see from any of the above definition that if f and g are both differentiable at some point x, then the sum function $f + g$ is differentiable at x and

$$(f + g)'(x) = f'(x) + g'(x).$$

It is a little harder for the **product rule**: there is a nice trick to remember in the proof.

Proposition 8.2.1 (Product rule) *If f and g are both differentiable at some point x, then the product function $f \cdot g$ is differentiable at x and*

$$(f \cdot g)'(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x).$$

Proof. To study the derivative of the function $f \cdot g$ at point x, we should look at the expression (i.e. infinitesimal rate of change)

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

Now the trick is to separate the numerator into two parts: write

$$f(x+h)g(x+h) - f(x)g(x) = f(x+h)(g(x+h) - g(x)) + g(x)(f(x+h) - f(x)).$$

Now, when h goes to 0, $f(x+h) \to f(x)$, $\frac{g(x+h)-g(x)}{h} \to g'(x)$, $\frac{f(x+h)-f(x)}{h} \to f'(x)$, so that

$$\frac{f(x + h)g(x + h) - f(x)g(x)}{h} \to f(x)g'(x) + f'(x)g(x).$$

This finished the proof. □

It is a litte harder still for the **chain rule**: for an easier proof, we use the second definition.

Proposition 8.2.2 (Chain rule) *Let f be defined on I and g be defined on J. Suppose that $x \in I$, $f(x) \in J$ and that f is differentiable at x and g is* *differentiable at $f(x)$. Then the composite function $g \circ f$ is differentiable at x and*

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

Proof. Let h be arbitrarily small and consider $g(f(x+h)) - g(f(x))$. Setting $k(h) = f(x+h) - f(x)$, this can be written as

$$g(f(x+h)) - g(f(x)) = g(f(x)+k) - g(f(x)) = g'(f(x)) \cdot k + k\psi(k)$$

with $\lim_{k \to 0} \psi(k) = 0$. We can also write k as

$$k(h) = f(x+h) - f(x) = f'(x) \cdot h + h\phi(h)$$

with $\lim_{h \to 0} \phi(h) = 0$. Combining we get that

$$\frac{g(f(x+h)) - g(f(x))}{h} = g'(f(x)) \cdot f'(x) + g'(f(x))\phi(h) + h \cdot o(1)$$

and the result follows by taking the $h \to 0$ limit. □

The rule for the derivative of the inverse function is harder to state, but easier to prove.

Proposition 8.2.3 *Assume that f is differentiable on some open interval (a, b), and that $f'(x) > 0$ on this interval. Then if the inverse function g of f is defined on some interval (α, β), then g is differentiable for any $y \in (\alpha, \beta)$, and*

$$g'(y) = \frac{1}{f'(g(y))}.$$

Proof. Let y_0 be close to y and study the expression

$$\frac{g(y_0) - g(y)}{y_0 - y}.$$

We can write $y_0 = f(x_0)$, $y = f(x)$ with $x_0 = g(y_0)$ and $x = g(y)$. Then the above expression becomes

$$\frac{x_0 - x}{f(x_0) - f(x)} = \frac{1}{\frac{f(x_0) - f(x)}{x_0 - x}}.$$

As y_0 goes to y, x_0 goes to x since g is continuous (as the inverse of a continuous function). Therefore, as y_0 goes to y, the above expression converges to $\frac{1}{f'(x)}$ (this limit is well-defined since $f'(x) \neq 0$, so we can compose it with the inverse function). We finish the proof by writing $x = g(y)$, so that $g'(y)$ exists and is equal to $\frac{1}{f'(g(y))}$. □

The requirement $f'(x) > 0$ is to ensure that $f'(x) \neq 0$, in such a way that the expression of $g'(y)$ makes sense.[4]

4: Although the derivative of a differentiable function is not necessarily continuous, it verifies the intermediate value theorem: this is known as Darboux's theorem. Therefore, it f' changes sign on the interval (a, b), then at some point, g will not be differentiable: this is why we restrict the proposition to the constant sign case for f'.

8.3 Rolle's theorem

A prelude to the mean value theorem is called Rolle's theorem.[5] Graphically, it is quite easy to understand its validity. The difficulty, in some sense, is to convince ourselves that the definition of the derivative above is strong enough to provide a rigorous proof.

5: Although historically, Rolle's theorem is a corollary of the mean value theorem...

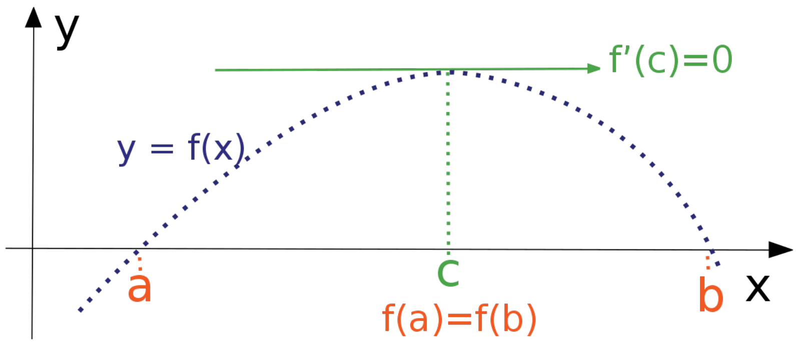


Figure 8.1: Rolle's theorem.

Theorem 8.3.1 (Rolle's theorem) *Let f be differentiable on some interval (a, b). Let $c \in (a, b)$ such that $f(c)$ is a maximum, i.e. for all $x \in (a, b)$, $f(x) \leq f(c)$. Then*

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

Proof. Since we need a result on $f'(c)$, it is natural to go back to its definition and consider

$$\frac{f(x) - f(c)}{x - c}$$

for x close to c and $x \in (a, b)$ (this is possible since the interval (a, b) does not include its end points).

Notice that $f(x) - f(c) \leq 0$ for all $x \in (a, b)$. Now, if $x > c$, then $x - c > 0$ and by the comparison theorem on limits, we have

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

Similarly, if $x < c$, then $x - c < 0$ and by the compaison theorem on limits,

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c} = \lim_{x \to c^-} \frac{f(x) - f(c)}{x - c} \geq 0.$$

Therefore, we are left with the only possibility that $f'(c) = 0$. □

Notice that the assumption of differentiability on f is crucial, otherwise we cannot conclude anything particularly useful only from the observation on the sign of[6]

$$\frac{f(x) - f(c)}{x - c}.$$

6: Although it is still worth looking at it in any case!

Via Weierstrass' theorem, we obtain the more convenient version:

Theorem 8.3.2 (Rolle's theorem bis) *Let f be continuous on some non-*

trivial interval $[a, b]$ *and differentiable on* (a, b). *Suppose that* $f(a) = f(b)$. *Then there exists* $c \in (a, b)$ *such that*

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

This is a consequence of the previous form, since by Weierstrass' theorem, there exists $c \in (a, b)$ such that $f(c)$ is the maximum of f on $[a, b]$, and that there exists $d \in (a, b)$ such that $f(d)$ is the minimum of f on $[a, b]$. To fit into the hypothesis of the theorem, we just need to show that we can choose c or d different from a or b. Indeed, if $\{c, d\} \subset \{a, b\}$, then since $f(a) = f(b)$, it means that f is constant on $[a, b]$ and Rolle's theorem is true. If f is not constant on $[a, b]$, then one of c or d is in (a, b), and applying the previous form of Rolle's theorem yields the result.

8.4 The main value theorem

The main value theorem is one of the most important results in analysis. Roughly speaking, under a condition of differentiability, one is able to compare the function on an interval to a linear function. Thus, if we are lucky, we can replace the problem about a general function to a problem of...linear algebra!

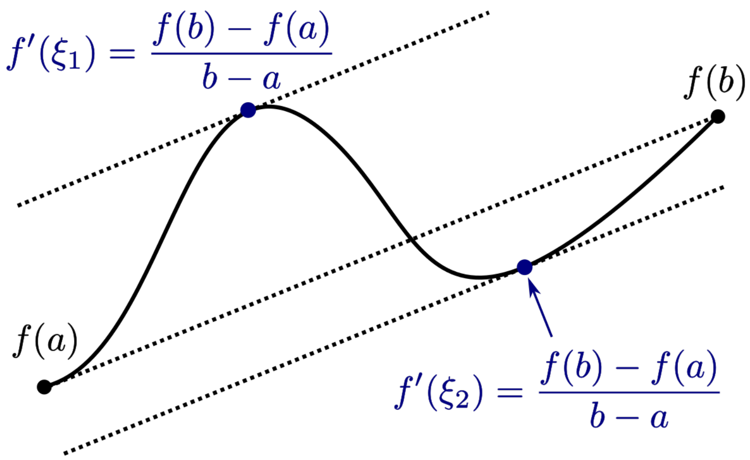


Figure 8.2: Mean value theorem.

Theorem 8.4.1 (Main value theorem) *Let* f *be continuous on some non-trivial interval* $[a, b]$ *and derivable on* (a, b). *Then there exists* $c \in (a, b)$ *such that*

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

Let us discuss how to reduce the main value theorem to Rolle's theorem: this also serves as a prelude to the theory of Taylor expansion.

When we compare the conditions in these two theorems, we see that the only difference is that in Rolle's theorem, there is an extra condition $f(a) = f(b)$. The idea then is to consider "linearly drifted version" of f in such a way that we can force this condition to appear.

More precisely, let f be continuous on some non-trivial interval $[a, b]$ and derivable on (a, b) with possibly different values $f(a)$ and $f(b)$. We consider the auxiliary function

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

and we want to choose α and β in such a way that $g(a) = g(b)$. Notice that adding a linear drift does not change the continuity and differentiability of the function, so we can apply Rolle's theorem on the auxiliary function g to get some information on the original function f.

Observe that the value of β does not really matter. We can then search for g in the form

$$g(x) = f(x) + \alpha(x - a),$$

which has the extra property that $g(a) = f(a)$. Solving $g(a) = g(b)$, we find $\alpha = -\frac{f(b)-f(a)}{b-a}$, in such a way that the good auxiliary function is

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

Now we can apply Rolle's theorem to g, and find $c \in (a, b)$ such that $g'(c) = 0$. Writing this in terms of f, we find

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

When writing a proof or reading a textbook, the order is usually reversed:

Proof. Consider the auxiliary function

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

We verify that g is continuous on $[a, b]$, derivable on (a, b) and $g(a) = g(b)$. Applying Rolle's theorem to g yields the existence of some $c \in (a, b)$ such that $g'(c) = 0$. This writes

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

□

An important application, already announced in the previous chapter, is a criteria for establishing the monotonicity of a derivable function. The second part of the criteria is somewhat technical to read and is not required in this course.

Corollary 8.4.2 *Let $I \subset \mathbb{R}$ be a real interval and $f : I \longrightarrow \mathbb{R}$ a differentiable function.*

1. *The function f is increasing if and only if for all $x \in I$, $f'(x) \geq 0$.*
2. *The function f is strictly increasing if and only if for all $x \in I$, $f'(x) \geq 0$ and such that the set of x where $f'(x) = 0$ contains no*

non-trivial intervals (i.e. is of empty interior).

Proof. We only prove the first part of this corollary.

First we prove that if f is differentiable on I and f is increasing, then $f'(x) \geq 0$ for all $x \in I$. Consider the expression

$$\frac{f(y) - f(x)}{y - x}$$

for y close to x. If $y \geq x$, then $f(y) \geq f(x)$ by monotonicity of f, and this expression is always positive. Passing to the limit, we have

$$f'(x) = \lim_{y \to x} \frac{f(y) - f(x)}{y - x} = \lim_{y \to x^+} \frac{f(y) - f(x)}{y - x} \geq 0$$

by comparison of limits. The case where x is one of the end points of I is similar (but you have to pick the right side for y).

Now we study the other direction: suppose that f is differentiable on I and $f'(x) \geq 0$ for all $x \in I$. To prove that f is increasing, pick $x < y$ with $x, y \in I$ and show that $f(x) \leq f(y)$. Again, we look at the expression

$$\frac{f(y) - f(x)}{y - x}.$$

Since f is continuous on the interval $[x, y]$ (being differentiable) and differentiable on (x, y), we can apply the mean value theorem to exhibit a number $c \in (x, y)$ such that

$$f'(c) = \frac{f(y) - f(x)}{y - x}.$$

Notice that $c \in (x, y) \subset I$, so that $f'(c) \geq 0$ by assumption. Together with the assumption $x < y$, this shows that $f(y) - f(x) \geq 0$, as expected. □

In this course, we will only use a weaker form of the second part of the corollary: if $f'(x) > 0$ for all $x \in I$, then f is strictly increasing on I. The proof is similar.

Moreover, if we know that at some point x, we have $f'(x) > 0$, then locally, f is injective (being strictly increasing), and it is thus locally invertible.[7]

7: This is known as a special case of the local inversion theorem.

8.5 The main value inequality

We say that a function f is of class $\mathscr{C}^1$ if f is derivable and that f' is continuous. In this case, we can mix the mean value theorem with Weierstrass' theorem to get a cooler version.

Corollary 8.5.1 (Main value inequality) *Let f be a $\mathscr{C}^1$ defined on some* *non-trivial interval* $[a,b]$. *Then there exists a constant* $M \geq 0$, *such that*

$$\left|\frac{f(b)-f(a)}{b-a}\right| \leq M.$$

In fact, one can choose $M = \sup_{x\in[a,b]} |f'(x)|$.

We leave the proof as an exercise.[8]

8: Post your question on the forum if it is not clear! Remember, there is no bad question in mathematics.

8.6 An extension theorem

This is sort of a "French" theorem, in the sense that it is mentioned in many lecture notes in France, but hard to find in a English textbook. A form of this theorem that one encounters often in practice is the following:

Theorem 8.6.1 *Let* $f : [a,b] \longrightarrow \mathbb{R}$ *be continuous and differentiable on* $(a,b]$. *Suppose that* f' *has a right limit* L *at the point* a. *Then* f *is differentiable at* a, *and* f' *is continuous at the point* a. *In particular,*

$$\lim_{x\longrightarrow a^+} \frac{f(x)-f(a)}{x-a} = L.$$

Actually, this theorem works even if f is only differentiable on $(a,b]$ without assuming the continuity of f at the point a.[9]

9: But the proof then involves the so-called Cauchy sequence, so we leave this part under the rug.

Proof. It should be clear at this point that, the main character of this chapter is not the symbol f', but rather the expression

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

To show that f is differentiable at a, we should show that the above expression has a limit when x goes to a from above (i.e. when x goes to a^+). We can apply the mean value theorem to f on the interval $[a,x]$, since f is continuous on $[a,x]$ and differentiable on (a,x) by assumption. So we can find $c \in (a,x)$ such that

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

As x goes to a^+, c goes to a^+ and $f'(c)$ goes to L by assumption. This shows that

$$\lim_{x\longrightarrow a^+} \frac{f(x)-f(a)}{x-a} = L.$$

So f' is differentiable at a. Moreover, since L is the limit of f' at a, this shows the continuity of f' at the point a as well. □

8.7 Taylor expansion

Normally, this part requires a new chapter of its own. We will be content with having only an impressionistic view of the theory.

Figure 8.3: Claude Monet, «Waterloo Bridge» (Hermitage), 1903.

Recall that the idea of the mean value theorem is the add a linear function to the original function f in such a way that the new function g verifies $g(a) = f(a)$ and $g(b) = g(a)$. In other words, we add a polynomial function P_1 to f such that $g = f + P_1$ satisfies $g(a) = f(a)$ and $g(b) = g(a)$.

We want to generalize it in the following way: how to add a polynomial function P_n of degree n to f such that $g = f + P_n$ satisfies $g(a) = f(a)$, $g'(a) = f'(a)$, ..., $g^{(n-1)}(a) = f^{(n-1)}(a)$ and such that $g(b) = g(a)$?

I strongly suggest you to continue independently in this direction, to create statements on your own and to try to prove them. And then, one can check this excellent blog post of Gowers to see a well-explained argument.

8.8 Exercises

Exercise 8.1 Consider the sequence of functions

$$f_n(x) = \frac{1}{1 + n^2 x}$$

with $x \in \mathbb{R}_{\geq 0}$ and $n \in \mathbb{Z}_{>0}$.

1. For a fixed $x \in \mathbb{R}_{\geq 0}$, what is the limit of $f_n(x)$ as n goes to infinity?
2. For a fixed $n \in \mathbb{Z}_{>0}$, what is the limit of $f_n(x)$ as x goes to 0?
3. Do we have $\lim_{x \to 0} \lim_{n \to \infty} f_n(x) = \lim_{n \to \infty} \lim_{x \to 0} f_n(x)$?[10]

10: In particular, the pointwise limit of a sequence of continuous function is not necessarily a continuous function. What is lacking here is the notion of "uniform convergence".

Exercise 8.2 Using the mean value theorem, find the limit

$$\lim_{n \to \infty} \left(n^{1/3} - (n+1)^{1/3} \right).$$

Exercise 8.3 Show that for every positive integer n,

$$\frac{1}{n+1} < \ln\left(1+\frac{1}{n}\right) < \frac{1}{n}.$$

Deduce that

$$\frac{1}{2} + \cdots + \frac{1}{n} < \ln(n) < \frac{1}{1} + \cdots + \frac{1}{n-1}.$$

Exercise 8.4 (*) Deduce from the previous question that

$$\sum_{n=1}^{\infty} \frac{1}{n} = \infty.$$

Modifying some of the arguments and show that

$$\sum_{n=1}^{\infty} \frac{1}{n\ln(n)} = \infty.$$

Exercise 8.5 Let $f : [0, 100] \longrightarrow \mathbb{R}$ be a real $\mathscr{C}^1$ function such that $f' \geq 1$ uniformly on $[0, 100]$. That is, for all $x \in [0, 100]$, $f'(x) \geq 1$. Show that

1. The set $f^{-1}([-1, 1])$ is a real interval;
2. The length of $f^{-1}([-1, 1])$ is upper bounded by 2.

Exercise 8.6 (*) Determine the Taylor expansion of $x \longmapsto \ln(1 + x)$ at 0 at any order.