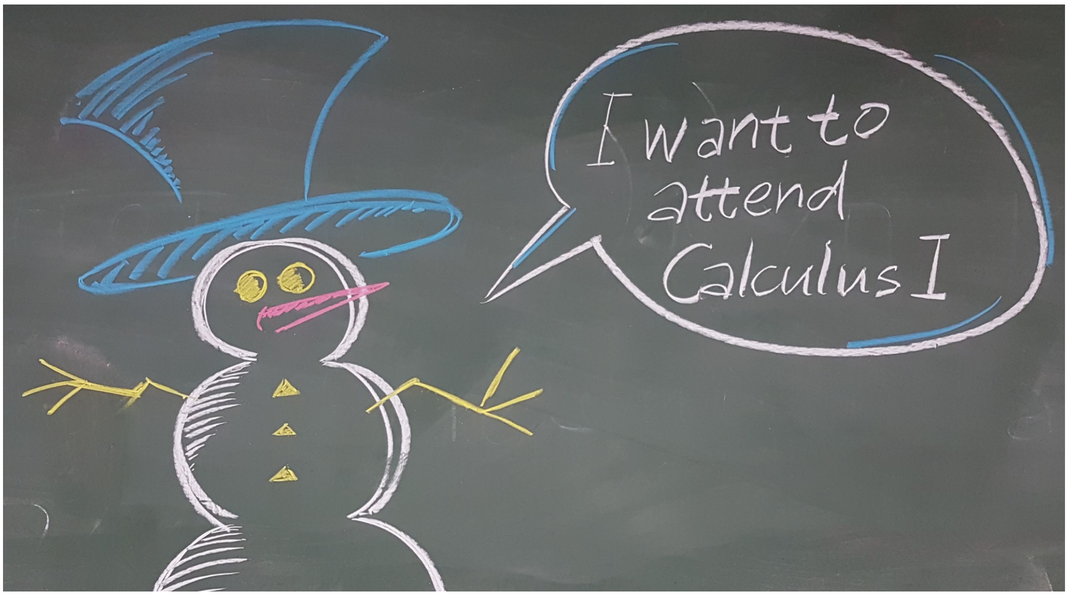


Nagoya University, G30 program

Fall Semester 2023

S. Richard

# Contents

# Chapter 1

# Numbers, limits, and functions

This chapter contains several notions which are certainly already known. However, one should pay attention to the notations: they will be used during the entire semester.

## 1.1 Reminder

Let us start by recalling some sets of numbers:

(i) The set $\mathbb{N}$ of natural numbers: $\mathbb{N} := \{1, 2, \ldots\}$,

(ii) The set $\mathbb{Z}$ of all integers: $\mathbb{Z} := \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$,

(iii) The set $\mathbb{Z}_+$ of non-negative integers: $\mathbb{Z}_+ := \{0, 1, 2, 3, \ldots\}$,

(iv) The set $\mathbb{Q}$ of all rational numbers $\mathbb{Q} := \{\frac{m}{n} \mid m, n \in \mathbb{Z} \text{ and } n \neq 0\}$,

(v) The set $\mathbb{R}$ of all real numbers: $\mathbb{R} := \{ \text{ all numbers } \}$.

Two observations have to be made, one on $\mathbb{Q}$ and one on $\mathbb{R}$. Any element of $\mathbb{Q}$ is not uniquely defined: for example one usually identifies $\frac{2}{4}$ with $\frac{1}{2}$. Thus, the set $\mathbb{Q}$ is in fact made of *equivalence class* of fractions, meaning that we identify the expressions $\frac{km}{kn}$ and $\frac{m}{n}$ for any $k \in \mathbb{Z}^* := \mathbb{Z} \setminus \{0\}$.

The construction of $\mathbb{R}$ is much more complicated. How can one define a number like $\pi$ or $\sqrt{2}$, which indeed belong to $\mathbb{R}$. How can one deal with the number $0.9999999999\ldots$ and say that this number is in fact equal to 1 ? We shall not construct the set $\mathbb{R}$ in this course, but simply say that it is the set of all (real) numbers. The lack of a precise definition of $\mathbb{R}$ will have one or two unfortunate consequences later on, they will be mentioned on due time.

There exist some natural operations on these sets of number: the addition $+$, the multiplication $\cdot$[1], taking the opposite, and taking the inverse: taking the opposite consists in the operation $x \mapsto -x$ and taking the inverse consists in the operation

[1] As usual, the notation $x \cdot y$ will very soon be replaced by $xy$.

$x \mapsto 1/x \equiv \frac{1}{x}$, whenever $x \neq 0$. Then, with these four operations one also gets the subtraction and the division, namely $x - y = x + (-y)$ and $x/y = x \cdot (1/y)$ for $y \neq 0$. Note that these operations are not all fully defined on these sets: $\mathbb{N}$ is not stable under taking the opposite and taking the inverse, while $\mathbb{Z}$ is stable under taking the opposite, but not for the inverse operation. The sets $\mathbb{Q}$ and $\mathbb{R}$ are stable under the four operations, but they will behave very differently in the near future.

These sets also contain a full (or total) order, namely $x \leq y$ or $x < y$ (and reciprocally $x \geq y$ and $x > y$). If you know already $\mathbb{C}$, the set of complex numbers, you can observe that the four operations are also defined in $\mathbb{C}$, but $\mathbb{C}$ has no full order ! In relation with the full order, let us recall a basic rule, which nevertheless leads to many mistakes: If $x \leq y$ and $\lambda \geq 0$, then $\lambda x \leq \lambda y$, while if $\lambda \leq 0$, then $\lambda x \geq \lambda y$. Similarly, if $x \leq y$ and $x, y \neq 0$, then $1/x \geq 1/y$.

In the sequel, we shall often use the notation $\mathbb{R}_+$ for $\{x \in \mathbb{R} \mid x \geq 0\}$. Depending on the authors, the number 0 is sometimes included and sometimes excluded from $\mathbb{R}_+$, it is really a matter of convention.

For any $x \in \mathbb{R}$ we also set

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}.$$

It follows that for all $x \in \mathbb{R}$, one has $|x| \geq 0$. The expression $|x|$ is called the absolute value of $x$.

The following definition is important, and exists in a much wider context:

**Definition 1.1** (Open and closed intervals)**.** *For any $a, b \in \mathbb{R}$ with $a < b$, one sets*

$$(a, b) := \{x \in \mathbb{R} \mid a < x < b\}$$

*for the* open interval $(a, b)$*, and*

$$[a, b] := \{x \in \mathbb{R} \mid a \leq x \leq b\}$$

*for the* closed interval $[a, b]$*.*

Let us come to powers. For $x \in \mathbb{R}_+$ and $n \in \mathbb{N}$ we simply set

$$x^n := \underbrace{x \cdot x \cdot x \ldots \cdot x}_{\text{n terms}},$$

while $x^0 = 1$ by convention. We also define $x^{\frac{1}{n}}$ by

$$x^{\frac{1}{n}} = y \Longleftrightarrow y \geq 0 \text{ and } y^n = x.$$

In such a case, $y$ is called *the $n^{th}$-root of $x$*. Now, if we consider also $m \in \mathbb{N}$, the definition of $x^{\frac{n}{m}}$ is quite clear, and one has

$$x^{\frac{n}{m}} = \big(x^n\big)^{\frac{1}{m}} = \big(x^{\frac{1}{m}}\big)^n.$$

In addition, one sets for any $n, m \in \mathbb{N}$ and if $x > 0$

$$x^{-\frac{n}{m}} := \frac{1}{x^{\frac{n}{m}}}.$$

With these definitions it is then easy to check that the following two rules apply for any $\alpha, \beta \in \mathbb{Q}$:

$$x^\alpha \cdot x^\beta = x^{\alpha+\beta} \qquad \text{and} \qquad \big(x^\alpha\big)^\beta = x^{\alpha\beta}. \tag{1.1}$$

So far and for any $x > 0$ we have defined $x^\alpha$ for any $\alpha \in \mathbb{Q}$, but what about $\alpha \in \mathbb{R}$ ? How can one define for example $\pi^{\sqrt{2}}$ ? Or even simply $3^\pi$ ? We have not given a definition to these expressions, and we shall be able to do it only later on.

## 1.2 Sequences and limits

In this section, we consider $X$ any subset of $\mathbb{R}$.

**Definition 1.2** (Sequence). *A collection of numbers $(a_n)_{n\in\mathbb{N}} \equiv (a_1, a_2, a_3, \dots)$, with $a_n \in X$ for any $n$, is called* a sequence in $X$.

A natural question is then: what it the limit $\lim_{n\to\infty} a_n$, if this limit exists ? What does it mean that this sequence has a limit, and how can one define it ? For example, if we set $a_n = \frac{1}{n}$, then it is quite clear than $a_n$ is approaching $0$ as $n$ tends to infinity, but how can one make this precise ?

**Definition 1.3** (Converging sequence). *A sequence $(a_n)_{n\in\mathbb{N}}$ in $X$ is converging to $a_\infty \in X$ if for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $|a_n - a_\infty| < \varepsilon$ for all $n \geq N$. In such a case one writes* $\lim_{n\to\infty} a_n = a_\infty$.

Observe that in this definition, the choice of $N$ depends on $\varepsilon$. Note also that the following equivalence holds:

$$|a_n - a_\infty| < \varepsilon \Longleftrightarrow a_\infty - \varepsilon < a_n < a_\infty + \varepsilon.$$

With this precise notion of convergence, it is easy to understand why the sequence given by $a_n = \frac{1}{n}$ converges to $0$. Indeed, for any $\varepsilon > 0$ one can choose an integer $N$ satisfying $N > \frac{1}{\varepsilon}$ and then one has for $n \geq N$

$$|a_n - 0| = \frac{1}{n} \leq \frac{1}{N} < \varepsilon,$$

which corresponds to the requirement for a convergence to $0$. However, there are plenty of sequences which do not converge. For example, the sequence defined by $a_n := (-1)^n$ is just alternating between $-1$ and $1$ and is not converging. Similarly, the sequence defined by $a_n = n$ is not converging to anything, since its values are becomming bigger and bigger (the value $\infty$ is not considered as a limiting value for a sequence).

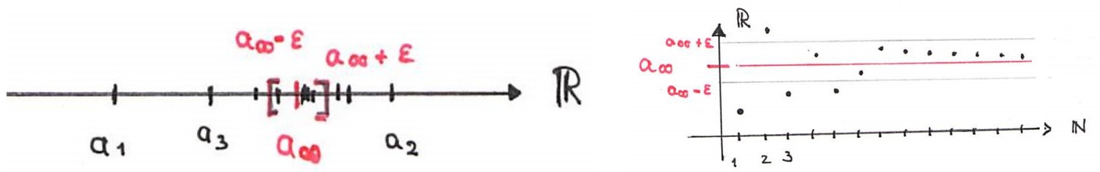


Fig. 1.1. Two representations of a converging sequence

It is possible to visualize the notion of convergence with the following two pictures, see Figure 1.1. In the first one, the horizontal axis corresponds to the possible values that $a_n$ can take. In the second picture, the horizontal axis correponds to the index $n$, and the vertical axis corresponds to the possible values taken by $a_n$.

The following statement can then be easily proved:

**Lemma 1.4.** *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ converging to $a_\infty$, and let $(b_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ converging to $b_\infty$. Let also $\lambda \in \mathbb{R}$. Then the following properties hold:*

*(i) The sequence $(\lambda a_n)_{n\in\mathbb{N}}$ converges to $\lambda a_\infty$,*

*(ii) The sequence $(a_n + b_n)_{n\in\mathbb{N}}$ converges to $a_\infty + b_\infty$,*

*(iii) The sequence $(a_n b_n)_{n\in\mathbb{N}}$ converges to $a_\infty b_\infty$.*

Let us now consider one example of a sequence in $\mathbb{Q}$. Let $a_1 = 1$, $a_2 = a_1 - \frac{1}{3}$, $a_3 = a_2 + \frac{1}{5}$, and more generally

$$a_n := \sum_{k=1}^{n} \frac{(-1)^{k+1}}{2k-1}.$$

Clearly, $a_n \in \mathbb{Q}$ for any $n \in \mathbb{N}$, but does the sequence $(a_n)_{n\in\mathbb{N}}$ converge as $n \to \infty$ ? The answer is YES, this sequence is converging and one has $\lim_{n\to\infty} a_n = \frac{\pi}{4}$ [2] [3].

Based on this example one observes an important fact: even if this sequence is taking values in $\mathbb{Q}$, its limit $\frac{\pi}{4}$ does not belong to $\mathbb{Q}$ but belongs to $\mathbb{R}$. This is one of the main difference between these two sets: A converging sequence with $a_n$ in $\mathbb{Q}$ for any $n$ might converge in $\mathbb{Q}$, but might also converge in $\mathbb{R} \setminus \mathbb{Q}$. On the other hand, a converging sequence will always converge in $\mathbb{R}$. One says that $\mathbb{R}$ is *complete*, while $\mathbb{Q}$ is not complete. In fact, it turns out that many converging sequences with all $a_n \in \mathbb{Q}$ have their limit in $\mathbb{R}$ and not in $\mathbb{Q}$.

**Exercise 1.5.** *Show that* $0.99999999\ldots = 1$. *A solution is provided in the Appendix 1.*

Let us now state a result which is going to play a very important role in the sequel. It is called the *monotone convergence theorem*, and more information on it can be found here. Note that its proof is not complicated, but relies precisely on the fact that $\mathbb{R}$ is

[2] https://en.wikipedia.org/wiki/Leibniz_formula_for_π

[3] https://math.stackexchange.com/questions/14815/why-does-this-converge-to-pi-4

complete, a notion which we do not develop in this course. In its statement we say that a sequence $(a_n)_{n\in\mathbb{N}}$ is *increasing* if $a_{n+1} \geq a_n$, or *decreasing* if $a_{n+1} \leq a_n$ for all $n \in \mathbb{N}$. We also say that the sequence is *upper bounded* if there exists $c < \infty$ such that $a_n \leq c$ for all $n \in \mathbb{N}$, or that the sequence is *lower bounded* if there exists $d > -\infty$ with $a_n \geq d$ for all $n \in \mathbb{N}$. Let us emphasize that the values of $c$ or of $d$ must be independent of $n$.

**Theorem 1.6** (Monotone convergence theorem)**.** *Any increasing sequence in $\mathbb{R}$ which is upper bounded is converging. Similarly, any decreasing sequence in $\mathbb{R}$ which is bounded from below is converging.*

## 1.3 Functions

In this section we denote by $X$ a set and by $x_1, x_2, \ldots$ its elements, which means $x_1, x_2, \ldots \in X$. For example, $X$ could be one of the sets $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ or the interval $(a, b)$ introduced before. It could also be $X = \{\text{G30 students}\}$.

**Definition 1.7** (Function)**.** *A* function *or a* map *$f$ from a set $X$ to a set $Y$ is a rule which associates to each element $x \in X$ a single element $y \in Y$. We write $y = f(x)$. The set $X$ is called* the domain of $f$*, also denoted by* $\mathrm{Dom}(f)$*. The set $Y$ is called* the codomain of $f$ *or the* target space of $f$*. The set*

$$f(X) \equiv \mathrm{Ran}(f) := \{y \in Y \mid y = f(x) \text{ for some } x \in X\}$$

*is called* the range of $f$ *or its* image*.*

The general representation of a function is provided in Figure 1.2. Quite often the following short notation will be used:

$$f : X \ni x \mapsto f(x) \in Y$$

which is a shorter version of $f : X \to Y$ with $f(x) = y$. Note that a very common mistake is to say that $f(x)$ is a function ! It is not, since $f(x)$ is a single element of $Y$. We should say that $x \mapsto f(x)$ is a function, of simply $f$.

The following notions will appear several times in this course, but also in other courses:

**Definition 1.8** (Injectivity, surjectivity, bijectivity)**.** *Consider a function $f : X \to Y$.*

*(i) $f$ is* injective *if for any $x_1, x_2 \in X$ with $x_1 \neq x_2$ one has $f(x_1) \neq f(x_2)$,*

*(ii) $f$ is* surjective *on $Y$ if for any $y \in Y$ there exists $x \in X$ with $y = f(x)$,*

*(iii) $f$ is* bijective *if $f$ is both injective and surjective.*

Observe that by definition, any function $f$ is always surjective on its image, which means surjective on $\mathrm{Ran}(f)$. Let us also look at the following example.

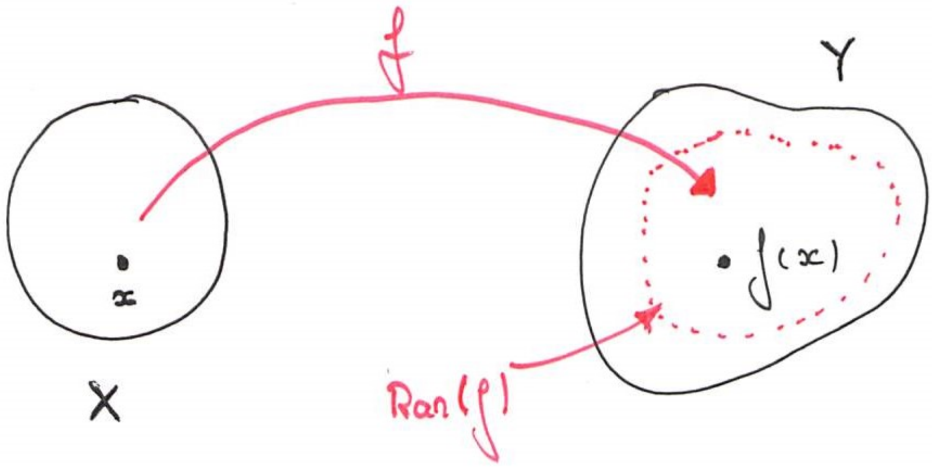


Fig. 1.2. A general function

**Example 1.9.** *(i) The function* $f : \mathbb{R} \ni x \mapsto x^2 \in \mathbb{R}$ *is neither injective nor surjective,*

*(ii) The function* $f : \mathbb{R}_+ \ni x \mapsto x^2 \in \mathbb{R}$ *is injective but not surjective,*

*(iii) The function* $f : \mathbb{R} \ni x \mapsto x^2 \in \mathbb{R}_+$ *is surjective but not injective,*

*(iv) The function* $f : \mathbb{R}_+ \ni x \mapsto x^2 \in \mathbb{R}_+$ *is bijective.*

Based this example, one infers that the property of being injective, surjective or bijective for a function $f$ depends mainly on its domain and on its codomain.

Given two functions $f_1, f_2 : X \to Y$, it is meaningful to add them, or to multiply them ? The answer is no in general, since $f_1(x) + f_2(x)$ could be meaningless, and similarly $f_1(x)f_2(x)$ could be meaningless. In fact, these operations are possible if $Y$ admits an addition or a multiplication. Once such operations are defined abstractly, it would be easy to develop the theory, but we shall not do it here. Since this course is about functions with values in $\mathbb{R}$ we shall consider only the special case $Y = \mathbb{R}$ (or sometimes $Y \subset \mathbb{R}$).

**Definition 1.10** (Addition and multiplication of functions)**.** *Consider* $f_1, f_2 : X \to \mathbb{R}$. *The* addition *of* $f_1$ *and* $f_2$, *denoted by* $f_1 + f_2$, *is the new function given by*

$$f_1 + f_2 : X \ni x \mapsto f_1(x) + f_2(x) \in \mathbb{R}.$$

*The multiplication of* $f_1$ *and* $f_2$, *denoted by* $f_1 f_2$, *is the new function given by*

$$f_1 f_2 : X \ni x \mapsto f_1(x) f_2(x) \in \mathbb{R}.$$

Another simple operation is to multiply a function by a constant. Again, this might not be well-defined for an arbitrary set $Y$, but if $Y = \mathbb{R}$ it can be easily defined, namely:

for any $f : X \to \mathbb{R}$ and for $\lambda \in \mathbb{R}$ we set $\lambda f$ for the new function $\lambda f : X \to \mathbb{R}$ defined by $[\lambda f](x) = \lambda f(x)$.

Finally, what about the composition of two functions ? This can only be defined when a certain compatibility between the two functions is imposed. More precisely, let $f : X \to Y$ and let $g : Y' \to Z$ with $\mathrm{Ran}(f) \subset Y' \subset Y$, then the composition $g \circ f$ is the new function $g \circ f : X \to Z$ defined by $[g \circ f](x) := g\big(f(x)\big)$. Clearly, the condition $\mathrm{Ran}(f) \subset \mathrm{Dom}(g)$ is necessary, otherwise the composition is not well-defined.

**Example 1.11.** *Consider* $f : \mathbb{R} \ni x \mapsto x^2 \in [0, \infty)$ *and* $g : \mathbb{R} \ni x \mapsto 3x - 1 \in \mathbb{R}$. *Then one has*

*(i)* $f + g : \mathbb{R} \ni x \mapsto x^2 + 3x - 1 \in \mathbb{R}$,

*(ii)* $fg : \mathbb{R} \ni x \mapsto x^2(3x - 1) \in \mathbb{R}$,

*(iii)* $g \circ f : \mathbb{R} \ni x \mapsto 3x^2 - 1 \in \mathbb{R}$,

*(iv)* $f \circ g : \mathbb{R} \ni x \mapsto (3x - 1)^2 \in \mathbb{R}_+$.

Note that the existence $g \circ f$ and of $f \circ g$ are not related, one could exist without the existence of the other one.

## 1.4 graphs and curves

A function $f : \mathbb{R} \to \mathbb{R}$ can be represented quite easily by a drawing in $\mathbb{R}^2$, which is usually called *the graph of* $f$ [4]. More generally, consider a function $f : \mathrm{Dom}(f) \to \mathbb{R}$, with $\mathrm{Dom}(f)$ a subset of $\mathbb{R}$. Then the set

$$\big\{\big(x, f(x)\big) \in \mathbb{R}^2 \mid x \in \mathrm{Dom}(f)\big\} \tag{1.2}$$

is called the graph of $f$. For example, the graph of the absolute value function is provided in Figure 1.3.

The graph of a few functions should be known, as the ones mentioned in the following list:

(i) The *straight line function* is defined by

$$f : \mathbb{R} \ni x \mapsto ax + b \in \mathbb{R} \tag{1.3}$$

with $a, b \in \mathbb{R}$. The coefficient $a$ is called *the slope* while the coefficient $b$ is called *the intercept*. In the special case $a = 0$, the function $f : x \mapsto b$ is called *the constant function*.

[4]Here, $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ denotes the usual plane (a 2 dimensional space), with the horizontal axis called the $x$-axis, and the vertical axis called the $y$-axis.

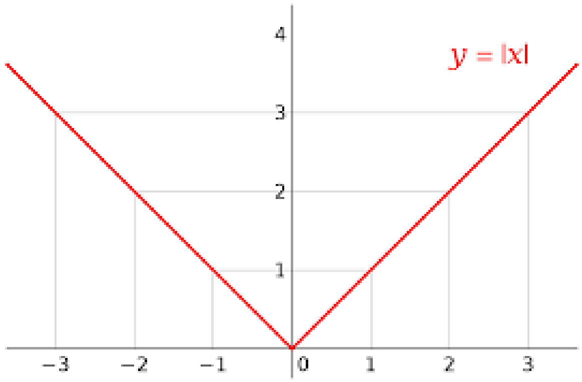


Fig. 1.3. The graph of the absolute value function

(ii) A function of the form

$$f : \mathbb{R} \ni x \mapsto (x-a)^2 + b \in \mathbb{R} \tag{1.4}$$

with $a, b \in \mathbb{R}$ is called a *parabola*. More generally one can consider a function of the form

$$f : \mathbb{R} \ni x \mapsto a_n x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_1 x + a_0 \in \mathbb{R} \tag{1.5}$$

with $a_n, a_{n-1}, \ldots, a_1, a_0 \in \mathbb{R}$. Such a function is known as a *polynomial function*. If $a_n \neq 0$, one says that the function is a polynomial of *degree* $n$.

(iii) A function of the form

$$f : \mathbb{R} \setminus \{b\} \ni x \mapsto \frac{a}{x-b} + c \in \mathbb{R} \tag{1.6}$$

with $a, b, c \in \mathbb{R}$ is called a *hyperbola*. The domain of the function can not be $\mathbb{R}$ since the value $x = b$ is not acceptable. In this case, one writes $\mathbb{R} \setminus \{b\}$ for the set of all $x \in \mathbb{R}$ with $x \neq b$.

(iv) The functions $\mathbb{R} \ni x \mapsto \sin(x) \in \mathbb{R}$ and $\mathbb{R} \ni x \mapsto \cos(x) \in \mathbb{R}$ are called *sine* and *cosine* functions.

**Exercise 1.12.** *Sketch the graph of the above functions.*

The graph of a function has a special property: it does not contain two pairs $(x, y_1)$ and $(x, y_2)$ if $y_1 \neq y_2$. Indeed, $(x, y)$ belongs to the graph of a function $f$ if and only if $y = f(x)$. However, one can represent more general objects on $\mathbb{R}^2$, namely *curves*. We do not present here the general theory of curves, but provide a few information, and a few examples as in Figure 1.4 from Wolfram MathWorld. Note that the main difference between these curves and the graph of a function is that it is possible to have $(x, y_1)$ and $(x, y_2)$ belonging to the curve even if $y_1 \neq y_2$. Also, one of the main property of

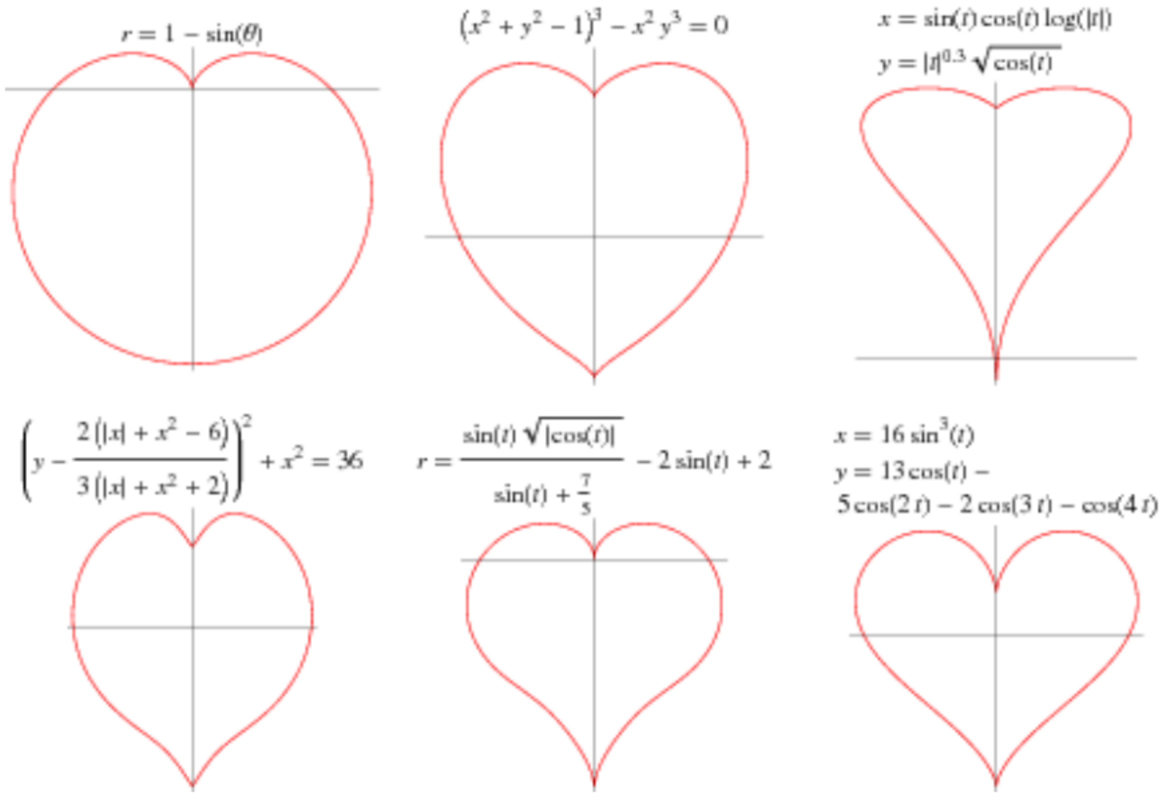


Fig. 1.4. Curves which are not graphs of functions

curves is that when one looks at them *locally*, they look like the graph of a function. For example, the curve $C$ (the unit circle in $\mathbb{R}^2$) given by

$$C := \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\} \tag{1.7}$$

is not the graph of a function since for any $x \in (-1, 1)$, the two points $\big(x, -\sqrt{1 - x^2}\big)$ and $\big(x, \sqrt{1 - x^2}\big)$ belong to $C$. However, we can also look at $C$ as the union of the graph of two functions. Indeed, we can consider

$$f_+ : [-1, 1] \ni x \mapsto \sqrt{1 - x^2} \in \mathbb{R}$$

and

$$f_- : [-1, 1] \ni x \mapsto -\sqrt{1 - x^2} \in \mathbb{R}$$

and then $C = \mathrm{Ran}(f_+) \cup \mathrm{Ran}(f_-)$.

Let us provide a second example, the folium of Descartes described by

$$\left\{(x, y) \in \mathbb{R}^2 \mid x = \frac{3t}{1 + t^3}, y = \frac{3t^2}{1 + t^3} \text{ for } t \in \mathbb{R}\right\}$$

and which is represented in Figure 1.5. It is rather clear that except at two points in $\mathbb{R}^2$ (which ones ?) this curve can be described locally by a function $f$. How can one find $f$ ?

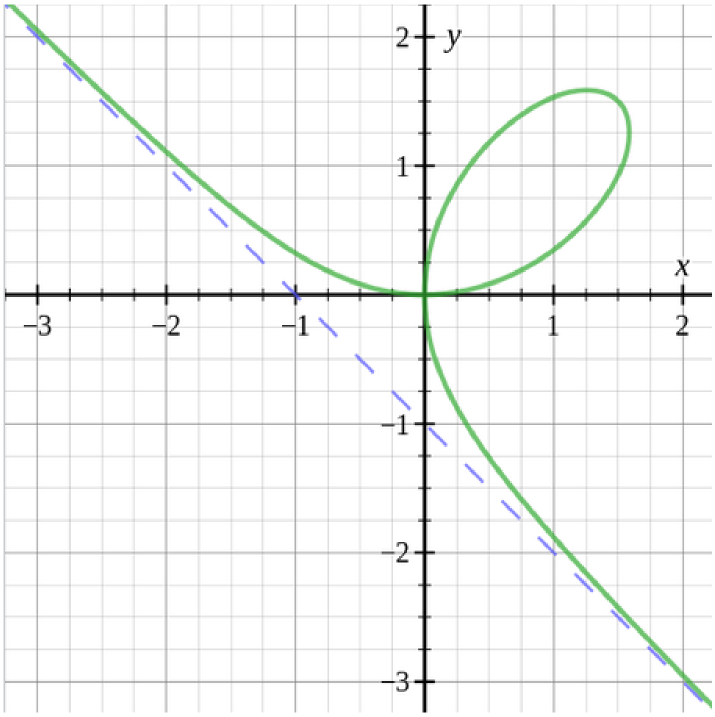


Fig. 1.5. The folium of Descartes, from Wikipedia

A priori, it would be sufficient to solve the equation $x = \frac{3t}{1+t^3}$ and express $t = g(x)$, and then insert this expression into $y = \frac{3t^2}{1+t^3} \equiv \frac{3g(x)^2}{1+g(x)^3} =: f(x)$. Usually, such a process is not simple, even if it is clear from the picture that such a function $f$ should exist locally.

# Chapter 2

# The derivative

In this chapter, we concentrate on the notion of derivative of a function. The first approach will be quite intuitive, and then an abstract definition will be provided.

## 2.1 Limits and continuity

First of all, one comes back to the notion of limit, already introduced in Section 1.2, but in the framework of a function defined on an interval of $\mathbb{R}$. As a rule, whenever we consider the interval $(a, b) \subset \mathbb{R}$ it means that $a < b$.

**Definition 2.1** (Limit from the right, limit from the left)**.** *Consider a function $f : (a, b) \to \mathbb{R}$.*

*(i) This function* has a limit at $a$ from the right *if there exists a value $f(a) \in \mathbb{R}$ such that for any $\varepsilon > 0$ there exists $\delta > 0$ with*

$$|f(x) - f(a)| < \varepsilon \quad \textit{for all} \;\; x \in (a, a + \delta).$$

*In such a case one writes $\lim_{x \searrow a} f(x) = f(a)$ or equivalently $\lim_{x \to a_+} f(x) = f(a)$.*

*(ii) This function* has a limit at $b$ from the left *if there exists a value $f(b) \in \mathbb{R}$ such that for any $\varepsilon > 0$ there exists $\delta > 0$ with*

$$|f(x) - f(b)| < \varepsilon \quad \textit{for all} \;\; x \in (b - \delta, b).$$

*In such a case one writes $\lim_{x \nearrow b} f(x) = f(b)$ or equivalently $\lim_{x \to b_-} f(x) = f(b)$.*

Note that in this definition, the letter $\delta$ reads *delta*, and that the choice of $\delta$ depends on $\varepsilon$. One should also be careful with the notation $f(a)$. At the beginning, the function $f$ is not defined at $a$, therefore $f(a)$ does not exist. On the other hand, if the function $f$ admits a limit at $a$ from the right, it is natural to denote this limit by $f(a)$, or sometimes by $f(a_+)$ or by $f(a + 0)$. Clearly, the same observation holds for $f(b)$, which is sometimes denoted by $f(b_-)$ or by $f(b - 0)$. Some examples of functions having or not having a limit are presented in Figure 2.1.

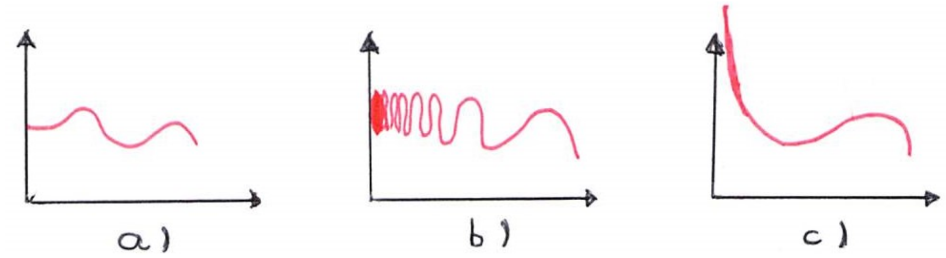


Fig. 2.1. a) has a limit at 0 from the right, b) and c) don't

It is then natural to consider simultaneously a limit from the right and a limit from the left. In the sequel and for $a < b < c$ we shall write $(a, c) \setminus \{b\}$ for the set $(a, b) \cup (b, c)$, which means the open interval $(a, c)$ with the point $b$ excluded.

**Definition 2.2** (Limit). *Consider a function $f : (a, c) \setminus \{b\}$. This function has a limit at $b$ if there exists $f(b) \in \mathbb{R}$ such that*

$$\lim_{x \to b_+} f(x) = f(b) = \lim_{x \to b_-} f(x)$$

*in which case one write $\lim_{x \to b} f(x) = f(b)$.*

In other words, the function $f$ has a limit at $b$ if it admits a limit at $b$ from the right, and the same limit at $b$ from the left. In terms of $\varepsilon$ and $\delta$ this definition is also given by: for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|f(x) - f(b)| < \varepsilon \quad \text{for all} \quad x \in (b - \delta, b + \delta) \setminus \{b\}.$$

**Exercise 2.3.** *Show as precisely as possible that the following limits exist:*

$$\begin{aligned} &\lim_{x \to 0} x^n = 0 \qquad \textit{for any } n \in \mathbb{N}, \\ &\lim_{x \to 0} \cos(x) = 1, \\ &\lim_{x \to 0} x \sin(1/x) = 0. \end{aligned}$$

*On the other hand, show that the function*

$$(0, \infty) \ni x \mapsto \sin(1/x) \in \mathbb{R}$$

*has no limit at 0 from the right. A solution is provided in the Appendix 2.*

The notion of limit is in fact very close to the notion of continuity. Examples of continuous or discontinuous functions are provided in Figure 2.2, and this notion is quite intuitive. Let us be more precise:

**Definition 2.4** (Continuity). *A function $f : (a, b) \to \mathbb{R}$ is* continuous *at $x_0 \in (a, b)$ if*

$$\lim_{x \searrow x_0} f(x) = f(x_0) = \lim_{x \nearrow x_0} f(x).$$

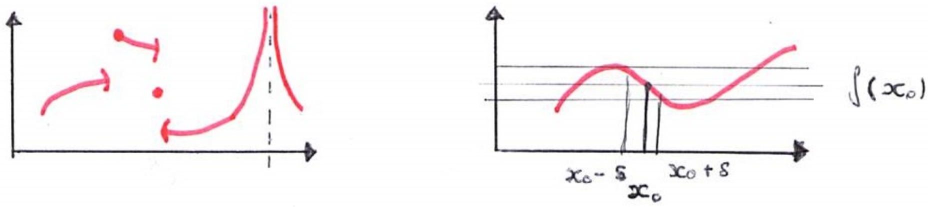


Fig. 2.2. A discontinuous and a continuous function

Note that the two equalities in the above definition are necessary. Indeed, if we consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 1$ for $x \neq 0$ and $f(0) = 0$, one has

$$\lim_{x \searrow 0} f(x) = 1 = \lim_{x \nearrow 0} f(x),$$

but these limits are not equal to $f(0) = 0$. Let us still provide some equivalent definitions of continuity. In fact, these equivalences just come from the equivalent notations for the concept of limit. One has $f : (a, b) \to \mathbb{R}$ is continuous at $x_0 \in (a, b)$ if for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|f(x) - f(x_0)| < \varepsilon \quad \text{for all} \quad x \in (x_0 - \delta, x_0 + \delta),$$

or equivalently

$$|f(x) - f(x_0)| < \varepsilon \quad \text{for all } x \text{ such that } |x - x_0| < \delta,$$

or still

$$|f(x_0 + h) - f(x_0)| < \varepsilon \quad \text{for all} \quad |h| < \delta.$$

If the function $f : (a, b) \to \mathbb{R}$ is continuous at any $x_0 \in (a, b)$ one says that $f$ is *continuous on* $(a, b)$. If in addition $f$ admits a limit at $a$ from the right, and a limit at $b$ from the left, we say that $f$ is continuous on $[a, b]$. The set of all continuous functions on $(a, b)$ is denoted by $C\big((a, b)\big)$ while the set of continuous functions on $[a, b]$ is denoted by $C\big([a, b]\big)$. In the sequel we shall often use the notation $I \subset \mathbb{R}$ either for $(a, b)$, for $[a, b]$, or for $\mathbb{R}$, whenever a statement holds for the open interval $(a, b)$, for the closed interval $[a, b]$, or for the entire set $\mathbb{R}$.

The following statement is quite similar to Lemma 1.4, and its proof can be easily obtained by using the definition of continuity.

**Lemma 2.5.** *Let $f, g \in C(I)$, and let $\lambda \in \mathbb{R}$, then*

*(i)* $\lambda f \in C(I)$,

*(ii)* $f + g \in C(I)$,

*(iii)* $fg \in C(I)$,

*(iv)* $\frac{f}{g} \in C(I)$ *if* $g(x) \neq 0$ *for any* $x \in I$.

## 2.2 The tangent and its slope

In this section we introduce the notion of the tangent to the graph of a function. This construction corresponds to the geometric interpretation of the derivative of a function provided in the next section.

Let us consider a function $f : (a, b) \to \mathbb{R}$ and let $x_0 \in (a, b)$. Let also $h > 0$ such that $x_0 + h \in (a, b)$. We can then consider the triangle in $\mathbb{R}^2$ having vertices of coordinates $\big(x_0, f(_0x)\big)$, $\big(x_0 + h, f(x_0)\big)$, and $\big(x_0 + h, f(x_0 + h)\big)$, see Figure 2.3. Clearly, the two points $\big(x_0, f(x_0)\big)$ and $\big(x_0 + h, f(x_0 + h)\big)$ belong to the graph of $f$. The slope of the straight line passing through these two points, which is also the slope one edge of the triangle, is given by

$$\frac{f(x_0 + h) - f(x_0)}{(x_0 + h) - x_0} = \frac{f(x_0 + h) - f(x_0)}{h}.$$

Our interest is to consider the limit of this expression as $h$ goes to 0. Note that the above construction also holds if $h < 0$, with the point $x_0 + h$ on the left of $x_0$.

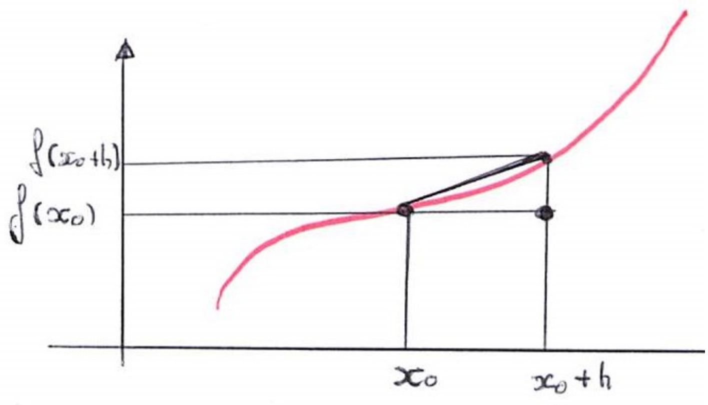


Fig. 2.3. Construction of the tangent

**Definition 2.6** (Tangent and slope of the tangent)**.** *Let $f : (a, b) \to \mathbb{R}$ and let $x_0 \in (a, b)$. Suppose that the limit $m := \lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h}$ exists. Then, the line in $\mathbb{R}^2$ having slope $m$ and passing through the point $\big(x_0, f(x_0)\big)$ is called the* tangent of $f$ at *$x_0$, and $m$ is called its* slope.

Let us emphasize that the above definition holds only if the slope exists, or more precisely if the limit $\lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h}$ exists. It is easily observed that the line passing through $\big(x_0, f(x_0)\big)$ and having slope $m$ is then given by the equation

$$y = m(x - x_0) + f(x_0).$$

If the limit $\lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h}$ does not exists, then the tangent can not be defined.

**Exercise 2.7.** *Show that in the following examples, the tangent at $x_0 = 0$ does not exist, but that it exists everywhere else:*

*(i)* $f : \mathbb{R} \to \mathbb{R}$ *with* $f(x) = |x|$,

*(ii)* $f : \mathbb{R} \to \mathbb{R}$ *with* $f(x) = -1$ *if* $x < 0$, $f(0) = 0$, *and* $f(x) = 1$ *if* $x > 0$. *This function is called the sign function.*

In Definition 2.6 we have introduced the tangent of a function. However, the tangent of a curve can be defined similarly, by looking locally at the curve as the graph of a function. Like for the tangent of a function, the tangent to a curve might not exist at some of its points.

## 2.3 The derivative of a function

In this section, we define the derivative of a function based on the geometric intuition developed in the previous section. The main idea is to extend what was done at a fixed $x_0 \in (a, b)$ to any $x \in (a, b)$

**Definition 2.8** (Derivative of a function)**.** *Let* $f : (a, b) \to \mathbb{R}$ *and let* $x \in (a, b)$. The derivative of $f$ at $x$, *denoted by* $f'(x)$, *is defined by*

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

*whenever this limit exists. If the derivative of $f$ exists at any $x \in (a, b)$, the* derivative of $f$ *is the function given by* $(a, b) \ni x \mapsto f'(x) \in \mathbb{R}$.

Note that the derivative of $f$ is also denoted by $\frac{\mathrm{d}f}{\mathrm{d}x}$. By comparing the above definition with Definition 2.6 it is clear that $f'(x)$ corresponds to the slope of the tangent of $f$ at $x$, whenever this quantity exists. If the derivative of $f$ exists, we say that $f$ is *differentiable* on $(a, b)$. Let us stress that $f'$ or $\frac{\mathrm{d}f}{\mathrm{d}x}$ represent a function, while the expressions $f'(x)$ and $\frac{\mathrm{d}f}{\mathrm{d}x}(x)$ correspond to the evaluation of this function at one point $x$, and therefore are just a number.

If a function $f : (a, b) \to \mathbb{R}$ is differentiable, and if its derivative $f'$ is a continuous function on $(a, b)$, we say that $f$ is *continuously differentiable on* $(a, b)$. In such a case one writes $f \in C^1\big((a, b)\big)$. If $a, b$ are arbitrary, one simply writes $C^1(\mathbb{R})$.

**Example 2.9.** *The function* $p_2 : \mathbb{R} \ni x \mapsto x^2 \in \mathbb{R}$ *belongs to* $C^1(\mathbb{R})$. *Indeed for any* $x \in \mathbb{R}$ *one has*

$$\frac{p_2(x+h) - p_2(x)}{h} = \frac{(x+h)^2 - x^2}{h} = \frac{2xh + h^2}{h} = 2x + h$$

*which satisfies* $\lim_{h\to 0} \frac{p_2(x+h)-p_2(x)}{h} = 2x$, *and therefore exists. In addition, the function* $x \mapsto 2x$ *is continuous. Thus, the derivative of* $p_2$ *exists and is continuous, which means that* $p_2 \in C^1(\mathbb{R})$.

We can generalize the above example, and consider the case of the polynomial $p_n$ defined for any $n \in \mathbb{N}$ by $p_n(x) := x^n$. The function $p_n$ is well-defined from $\mathbb{R}$ to $\mathbb{R}$. Let us also introduce some notations. For any $n \in \mathbb{N}$ we set

$$n! := n \cdot (n-1) \cdot (n-2) \dots 2 \cdot 1$$

for the *factorial of* $n$, and by convention, we set $0! = 1$. Also, for any two positive integers $n$ and $k$ with $n \geq k$ we set $\binom{n}{k} := \frac{n!}{k!(n-k)!}$ for the *binomial coefficient*. This number represents the number of ways to choose an (unordered) subset of $k$ elements from a fixed set of $n$ elements. This number can also be written

$$\binom{n}{k} = \frac{n \cdot (n-1) \cdot (n-2) \dots (n-k+1)}{k \cdot (k-1) \cdot (k-2) \dots 2 \cdot 1}.$$

With these notations, the following equality holds (if you have never seen it, think about it) for any $n \in \mathbb{N}$:

$$(x+h)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} h^k.$$

It then follows that

$$\begin{aligned} p_n(x+h) - p_n(x) &= \sum_{k=0}^{n} \binom{n}{k} x^{n-k} h^k - x^n \\ &= \sum_{k=1}^{n} \binom{n}{k} x^{n-k} h^k \\ &= h \sum_{k=1}^{n} \binom{n}{k} x^{n-k} h^{k-1}, \end{aligned}$$

from which one infers that

$$\lim_{h \to 0} \frac{p_n(x+h) - p_n(x)}{h} = \lim_{h \to 0} \sum_{k=1}^{n} \binom{n}{k} x^{n-k} h^{k-1} = \binom{n}{1} x^{n-1} = n x^{n-1}.$$

Thus, the derivative of $p_n$ exists, and one has

$$p_n'(x) = n x^{n-1} \qquad \text{for any } x \in \mathbb{R}. \tag{2.1}$$

Finally, since the function $x \mapsto n x^{n-1}$ is continuous, it follows that $p_n \in C^1(\mathbb{R})$.

Let us now express the notion of differentiability in terms of $\varepsilon$ and $\delta$. Indeed, Definition 2.8 is expressed in terms of a limit, but behind this limit some $\varepsilon$ and $\delta$ are hidden. If we want to be more precise one should say: a function $f : (a, b) \to \mathbb{R}$ is differentiable at $x \in (a, b)$ if there exists $f'(x) \in \mathbb{R}$ such that for any $\varepsilon > 0$ there exists $\delta > 0$ with

$$\left| \frac{f(y) - f(x)}{y - x} - f'(x) \right| < \varepsilon \quad \text{for all } \; y \in (x - \delta, x + \delta) \text{ and } y \neq x,$$

or equivalently

$$\left|\frac{f(x+h)-f(x)}{h} - f'(x)\right| < \varepsilon \quad \text{for all} \quad h \in (-\delta,\delta) \text{ and } h \neq 0.$$

A natural question is about the relation between differentiability and continuity, as introduced in Definition 2.4.

**Exercise 2.10.** *Let $f:(a,b)\to\mathbb{R}$. Show that if $f$ is differentiable on $(a,b)$, then $f$ is continuous on $(a,b)$.*

Let us stress that the converse statement is not always true, namely a continuous function is not always differentiable. For example, the function $f:\mathbb{R}\ni x\mapsto |x|\in\mathbb{R}$ is continuous on $\mathbb{R}$ but it is not differentiable at 0. On the other hand, $f$ is differentiable at any $x\neq 0$ and one has $f'(x)=-1$ if $x<0$ while $f'(x)=1$ if $x>0$.

**Exercise 2.11.** *Check precisely these statements about the absolute value function.*

**Remark 2.12.** *Since the notions of limit from the left and from the right have been introduced, we could similarly define the left and the right derivative of a function at one point. In such a case, the absolute value function would have a derivative from the right at 0 (and equal to 1) and a derivative from the left at 0 (and equal to $-1$). However, we shall not consider this refinement in this course.*

## 2.4 Back to limits

In this section we come back to the notion of a limit, and present one important result about the limit of a ratio of two functions, namely l'Hôpital's rule.

Recall first that $\mathbb{R}^*=\mathbb{R}\setminus\{0\}$, and consider two continuous functions $f,g:\mathbb{R}^*\to\mathbb{R}$. It easily follows from Lemma 2.5 and if $f$ admits a limit at 0, with $\lim_{x\to 0} f(x)=f(0)$, and similarly if $g$ admits a limit at 0, with $\lim_{x\to 0} g(x)=g(0)$, then one has

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

and

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

but what about $\lim_{x\to 0}\frac{f}{g}(x)$ ? In the good situation, one has

**Exercise 2.13.** *Suppose that $g(0)\neq 0$, show with $\varepsilon$ and $\delta$ that*

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

On the other hand, if $g(0)=0$, what can one say about $\lim_{x\to 0}\frac{f}{g}(x)$ ? If it rather clear that if $f(0)\neq 0$, then the mentioned limit does not exist, but if $f(0)=0$ the best answer is *it depends*, as one can see by considering the following examples.

**Examples 2.14.** *(i) Consider $f(x) = x^2$ and $g(x) = x^2 + x$, then*

$$\lim_{x\to 0} \frac{f}{g}(x) = \lim_{x\to 0} \frac{x^2}{x^2+x} = \lim_{x\to 0} \Big(\frac{x}{x}\frac{x}{x+1}\Big) = 1 \cdot \lim_{x\to 0} \frac{x}{x+1} = \frac{0}{1} = 0.$$

*The limist exists and is 0 in this case.*

*(ii) Consider $f(x) = x^2$ and $g(x) = x^3$, then*

$$\lim_{x\to 0} \frac{f}{g}(x) = \lim_{x\to 0} \frac{x^2}{x^3} = \lim_{x\to 0} \Big(\frac{x^2}{x^2}\frac{1}{x}\Big) = 1 \cdot \lim_{x\to 0} \frac{1}{x}$$

*which is not well-defined. Thus the limit does not exist in this case.*

*(iii) Consider $f(x) = x^2$ and $g(x) = x^2 + x^3$, then*

$$\lim_{x\to 0} \frac{f}{g}(x) = \lim_{x\to 0} \frac{x^2}{x^2+x^3} = \lim_{x\to 0} \Big(\frac{x^2}{x^2}\frac{1}{1+x}\Big) = 1 \cdot \lim_{x\to 0} \frac{1}{1+x} = \frac{1}{1} = 1.$$

*The limit exists and is 1 in this case.*

Note that in the above examples, the special role played by $0 \in \mathbb{R}$ is irrelevant, any other point $a \in \mathbb{R}$ could have been considered, as long as $\lim_{x\to a} f(x) = 0 = \lim_{x\to a} g(x)$. Let us now present an important result which could be applied in the previous situation, namely l'Hôpital's rule (also written l'Hospital's rule).

**Theorem 2.15** (L'Hôpital's rule)**.** *Let us consider $(a,b) \subset \mathbb{R}$ and $x_0 \in (a,b)$. Let $f,g : (a,b) \to \mathbb{R}$ be differentiable on $(a,b)$ and suppose that $\lim_{x\to x_0} f(x) = 0 = \lim_{x\to x_0} g(x)$. Assume also that $g(x) \neq 0$ for any $x \in (a,b) \setminus \{x_0\}$, and that $g'(x) \neq 0$ for any $x \in (a,b) \setminus \{x_0\}$. Then one has:*

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} = \lim_{x\to x_0} \frac{f'(x)}{g'(x)}$$

*if this latter limit exists.*

The proof that we provide now does not correspond exactly to the above statement. Indeed, we shall assume that $g'(x_0) \neq 0$, an assumption which is not necessary, and that $\lim_{x\to x_0} \frac{f'(x)}{g'(x)} = \frac{f'(x_0)}{g'(x_0)}$. A slightly more sophisticated proof, which does not require these assumptions, will be provided later on.

*Proof.* Let us assume that $g'(x_0) \neq 0$. Then one has

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} = \lim_{x\to x_0} \frac{\frac{f(x)-f(x_0)}{x-x_0}}{\frac{g(x)-g(x_0)}{x-x_0}} = \frac{\lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0}}{\lim_{x\to x_0} \frac{g(x)-g(x_0)}{x-x_0}} = \frac{f'(x_0)}{g'(x_0)} = \lim_{x\to x_0} \frac{f'(x)}{g'(x)},$$

where the content of Exercise 2.13 has been taken into account. □

Note that a similar statement exists if $\lim_{x\to x_0} f(x) = \pm\infty = \lim_{x\to x_0} g(x)$, but we shall not develop it here.

## 2.5 Properties of differentiation

In this section we gather and prove some important results related to the differentiation of functions. These results will be used, implicitly or explicitly, all the time during this course.

**Proposition 2.16.** *Let* $f, g : (a,b) \to \mathbb{R}$ *be differentiable at* $x \in (a,b)$*, and let also* $\lambda \in \mathbb{R}$*. Then the following equalities hold:*

*(i)* $\big(\lambda f\big)'(x) = \lambda f'(x)$,

*(ii)* $\big(f+g\big)'(x) = f'(x) + g'(x)$ (sum rule),

*(iii)* $\big(fg\big)'(x) = f'(x)g(x) + f(x)g'(x)$, (product rule),

*(iv)* $\Big(\frac{f}{g}\Big)'(x) = \frac{f'(x)g(x)-f(x)g'(x)}{g^2(x)}$ *as long as* $g(x) \neq 0$ (quotient rule).

Let us mention that we prefer writing $g^2(x)$ instead of $g(x)^2$, but that both notations are correct. These rules are often just written

$$(\lambda f)' = \lambda f', \quad (f+g)' = f' + g', \quad (fg)' = f'g + fg', \quad \Big(\frac{f}{g}\Big)' = \frac{f'g - fg'}{g^2}.$$

Also, it will become clear later on that the last rule can be partially deduced from other rules, once the derivation of a composition is introduced.

*Proof.* The proof of these statements is simply based on the definition of the derivative of functions, as provided in Definition 2.8. Note that instead of writing $\lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = f'(x)$ we shall use the notation $\frac{f(x+h)-f(x)}{h} \xrightarrow{h \to 0} f'(x)$.

$(i)$ One has

$$\frac{\lambda f(x+h) - \lambda f(x)}{h} = \lambda \frac{f(x+h)-f(x)}{h} \xrightarrow{\quad h \to 0 \quad} \lambda f'(x).$$

$(ii)$ One has

$$\begin{aligned}\frac{\big(f(x+h)+g(x+h)\big) - \big(f(x)+g(x)\big)}{h} &= \frac{f(x+h)-f(x)}{h} + \frac{g(x+h)-g(x)}{h} \\ &\xrightarrow{\quad h \to 0 \quad} f'(x) + g'(x).\end{aligned}$$

$(iii)$ One has

$$\begin{aligned}\frac{f(x+h)g(x+h) - f(x)g(x)}{h} &= \frac{f(x+h)-f(x)}{h} g(x+h) + f(x) \frac{g(x+h)-g(x)}{h} \\ &\xrightarrow{\quad h \to 0 \quad} f'(x)g(x) + f(x)g'(x).\end{aligned}$$

where we have used that a differentiable function is automatically continuous.

$(iv)$ Finally one has

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

where we have again used that a differentiable function is automatically continuous. □

If you look carefully at this proof, one small argument is missing, which one ? Now, as an application of these formulas, observe that for $f$ differentiable at $x$ with $f(x)\neq 0$ one has

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

As a consequence, if we set $p_{-n}(x):=x^{-n}=\frac{1}{x^n}$ for $x>0$ and $n\in\mathbb{N}$, one infers that

$$p'_{-n}(x)=-\frac{nx^{n-1}}{x^{2n}}=-\frac{n}{x^{n+1}}=-nx^{-n-1}=-np_{-n-1}(x).$$

If we summarize our findings about $p_n$, one has $p'_n=np_{n-1}$ for any $n\in\mathbb{Z}^*$. Clearly, if $n=0$, the function $p_0$ is the constant function equal to 1, whose derivative is the constant function equal to 0.

What about the derivative of a composition ? More precisely, if we consider the function $x\mapsto g\big(f(x)\big)$, whenever this composition is well-defined, what can we say about its derivative ?

**Proposition 2.17** (chain rule). *Consider $f:(a,b)\to\mathbb{R}$, differentiable at $x\in(a,b)$, and consider $g:(c,d)\to\mathbb{R}$, differentiable at $f(x)\in(c,d)$. Then the function $g\circ f$ is differentiable at $x$ and one has $(g\circ f)'(x)\equiv g\big(f(x)\big)'=g'\big(f(x)\big)f'(x)$.*

*Proof.* Let us set $u:=f(x)$ and $k:=f(x+h)-f(x)$, and observe that $k$ goes to 0 as $h$ goes to 0, by the continuity of the function $f$ at $x$. Since $f(x+h)=f(x)+k=u+k$ one has if $k\neq 0$

$$\begin{aligned}\frac{g\big(f(x+h)\big)-g\big(f(x)\big)}{h}&=\frac{g(u+k)-g(u)}{h}\\&=\frac{g(u+k)-g(u)}{k}\,\frac{k}{h}\\&=\frac{g(u+k)-g(u)}{k}\,\frac{f(x+h)-f(x)}{h}.\end{aligned}\tag{2.2}$$

As a consequence one has

$$\begin{aligned}\lim_{h\to 0}\frac{g\big(f(x+h)\big)-g\big(f(x)\big)}{h}&=\lim_{k\to 0}\frac{g(u+k)-g(u)}{k}\cdot\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}\\&=g'\big(u\big)f'(x)\\&=g'\big(f(x)\big)f'(x),\end{aligned}$$

as planned. Note that the only weak point in this proof is the assumption $k \neq 0$. In fact, if $f$ is a constant function near $x$, then $k = f(x+h) - f(x) = 0$, for all $h$ close enough to 0, and one gets the statement directly from the first equality in (2.2), with $(g \circ f)'(x) = 0$. On the other hand, if $f$ is not the constant function near $x$, then one can choose sufficiently many $k \neq 0$ such that the above proof is correct. □

For example, this result allows us to compute $\frac{\mathrm{d}(f^n)}{\mathrm{d}x}(x)$ for any $n \in \mathbb{Z}$. Indeed, if we set $g(x) = p_n(x) \equiv x^n$, then $f^n = p_n \circ f$, which leads to

$$\frac{\mathrm{d}(f^n)}{\mathrm{d}x}(x) = \big(p_n \circ f\big)'(x) = n\big(p_{n-1} \circ f\big)(x)f'(x) = nf^{n-1}(x)f'(x).$$

Let us now consider $f : (a,b) \to \mathbb{R}$ which is differentiable on $(a,b)$, and let $f'$ denote its derivative. If the function $f' : (a,b) \to \mathbb{R}$ is itself differentiable, one can compute $\frac{\mathrm{d}(f')}{\mathrm{d}x} \equiv f''$ which corresponds to the derivative of $f'$. A convenient notation for the derivative of $f'$ is $f^{(2)}$, but this notation should not be confused with $f^2$. The function $f'' \equiv f^{(2)} : (a,b) \to \mathbb{R}$ is called *the second derivative of $f$*. Clearly, the process can be continued, and one obtains *higher order derivatives of $f$*, or *$n$-derivatives of $f$*. The notation $f^{(n)}$ corresponds to the derivative of $f$ taken $n$ times, and should again not be confused with $f^n$ (the function $f$ at a power $n$). If the function $f^{(n)}$ is continuous, one says that the initial function $f$ is *$n$ times continuously differentiable*, and one writes $f \in C^n\big((a,b)\big)$.

**Example 2.18.** *For any $n, k \in \mathbb{N}$, the function $p_n : \mathbb{R} \ni x \mapsto x^n \in \mathbb{R}$ is $k$ times continuously differentiable with, for any $x \in \mathbb{R}$,*

$$\begin{aligned}
p_n^{(1)}(x) &= nx^{n-1}, \\
p_n^{(2)}(x) &= n(n-1)x^{n-2}, \\
p_n^{(3)}(x) &= n(n-1)(n-2)x^{n-3}, \\
\vdots\quad &= \quad\vdots \\
p_n^{(n)}(x) &= n(n-1)(n-2)\dots 2 \cdot 1 = n! \\
p_n^{(k)}(x) &= 0 \qquad \textit{for any } k > n.
\end{aligned}$$

## 2.6 Implicit differentiation

Let us come back to curves in $\mathbb{R}^2$. Quite often, curves are described by a relation of the form $F(x,y) = 0$ for $(x,y) \in \mathbb{R}^2$ and $F$ a function of two variables. For example, the unit circle described in (1.7) corresponds to the function $F(x,y) := x^2 + y^2 - 1$. As we have seen in this example, the curve can sometimes be expressed locally with some explicit functions, but in general such functions can not be exhibited. However, if we assume that such functions exist locally, then one can get quite a lot of information about them.

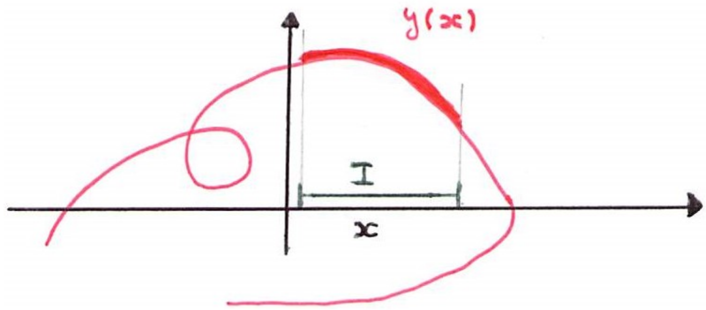


Fig. 2.4. A curve, and a local function

For example, one can get the tangent at a point of a curve without having an explicit expression of the local function. This is part of the concept of *implicit differentiation*. In fact, this subject will be studied more extensively in Calculus II, once functions of an arbitrary number of variables will be studied. Nevertheless one can still scratch the surface of the subject with our current knowledge ☺.

We shall now present the main idea of the subject on an example, but it will be clear that the special choice of the example is not important. Consider the curve defined by the function $F$ given by $F(x,y) := 3x^3y - y^4 + 5x^2 + 5$, namely

$$C = \big\{(x,y) \in \mathbb{R}^2 \mid F(x,y) = 0\big\}.$$

It seems difficult to represent this curve, and also to get a local expression for its description. Nevertheless, let us assume that locally, this curve can be described by a function. It is natural to call this function $y$, and therefore to write $y : I \ni x \mapsto y(x) \in \mathbb{R}$, for some interval $I$, see Figure 2.4. This function $y$ satisfies the relation $F\big(x, y(x)\big) = 0$ for all $x \in I$, since the corresponding points $\big(x, y(x)\big) \in \mathbb{R}^2$ belong to the curve. Thus, the function

$$I \ni x \mapsto F\big(x, y(x)\big) = 3x^3y(x) - y^4(x) + 5x^2 + 5 = 0 \tag{2.3}$$

which means that it is the 0-function on this interval. But the 0-function can be differentiate an arbitrary number of times, and all its derivatives are equal to 0. By using the properties of differentiation recalled in Propositions 2.16 and 2.17 one infers that

$$\begin{aligned}
&\frac{\mathrm{d}F\big(x, y(x)\big)}{\mathrm{d}x} = 9x^2y(x) + 3x^3y'(x) - 4y^3(x)y'(x) + 10x = 0 \\
\Longleftrightarrow\ &\Big[3x^3 - 4y^3(x)\Big]y'(x) = -9x^2y(x) - 10x \\
\Longleftrightarrow\ &y'(x) = \frac{-9x^2y(x) - 10x}{3x^3 - 4y^3(x)}
\end{aligned}$$

as long as the denominator is not equal to 0.

As an application, consider the point $(1,2)$ which satisfies $F(1,2) = 0$. As a consequence, $(1,2)$ belongs to the curve $C$, and locally around this point the curve can be described by a function $y$ as introduced above. Then, by using that $y(1) = 2$, the slope of the tangent to the curve at the point $(1, 2)$ is given by

$$y'(1) = \frac{-9 \cdot 1^2 \cdot 2 - 10 \cdot 1}{3 \cdot 1^3 - 4 \cdot 2^3} = \frac{-18 - 10}{3 - 32} = \frac{-28}{-29} = \frac{28}{29}.$$

With this information, one can also write the equation of the tangent to the curve at the point $(1, 2)$, namely

$$y = \frac{28}{29}(x - 1) + 2 = \frac{28}{29}x + \frac{30}{29}.$$

Let us emphasize that it is possible to go further in the analysis. For example, one could check if the curve is below or above the tangent, or get a local approximation of the curve by a polynomial, instead of just the tangent function. Note that the starting point of these investigations is always the trivial equality (2.3).

## 2.7 Examples with basic functions

In this final section we consider the derivative of a few basic functions, namely the sine and cosine functions, and the exponential function. For the trigonometric functions, the following identities will be used, and should be "known": for any $x, y \in \mathbb{R}$

(i) $\cos^2(x) + \sin^2(x) = 1$,

(ii) $\sin(x + y) = \sin(x)\cos(y) + \cos(x)\sin(y)$,

(iii) $\cos(x + y) = \cos(x)\cos(y) - \sin(x)\sin(y)$,

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

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

Let us now state an important lemma. Its proof will be given after two corollaries.

**Lemma 2.19.** *One has* $\lim_{h\to 0} \frac{\sin(h)}{h} = 1$.

**Corollary 2.20.** *One has* $\sin'(x) = \cos(x)$ *for any* $x \in \mathbb{R}$.

*Proof.* From its definition and by using formula (iv) one has

$$\begin{aligned}
\sin'(x) &= \lim_{h\to 0} \frac{\sin(x + h) - \sin(x)}{h} \\
&= \lim_{h\to 0} \frac{2\sin(h/2)\cos(x + h/2)}{h} \\
&= \Big(\lim_{h\to 0} \frac{\sin(h/2)}{h/2}\Big)\Big(\lim_{h\to 0} \cos(x + h/2)\Big) \\
&= 1 \cdot \cos(x) \\
&= \cos(x),
\end{aligned}$$

as stated. □

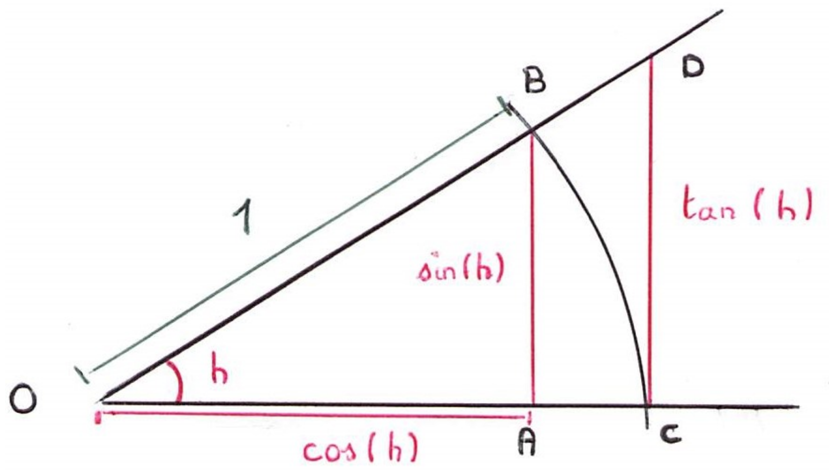


Fig. 2.5. Construction for the proof of Lemma 2.19

Other results can be deduced from Lemma 2.19. The following statements will be proved during the tutorial session:

**Corollary 2.21.** *One has*

*(i)* $\cos'(x) = -\sin(x)$ *for all* $x \in \mathbb{R}$,

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

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

*Proof of Lemma 2.19.* The proof is based on Figure 2.5, which is valid for $\frac{\pi}{2} > h > 0$. The case $-\frac{\pi}{2} < h < 0$ can be obtained similarly. In the picture, the variable $h$ should be thought as an angle in radian. Observe that on Figure 2.5 the following equalities for the lengths hold: $OA = \cos(h)$, $AB = \sin(h)$ and $CD = \tan(h) = \frac{\sin(h)}{\cos(h)}$. Based on these equalities and on inequalities about areas

$$\mathrm{Area}\big(\triangle(OAB)\big) \leq \frac{h}{2\pi}\pi 1^2 \leq \mathrm{Area}\big(\triangle(OCD)\big)$$

one infers that

$$\begin{aligned} & \frac{1}{2}\cos(h)\sin(h) \leq \frac{h}{2} \leq \frac{1}{2}1\tan(h) = \frac{1}{2}\frac{\sin(h)}{\cos(h)} \\ \iff & \cos(h) \leq \frac{h}{\sin(h)} \leq \frac{1}{\cos(h)} \\ \iff & \frac{1}{\cos(h)} \geq \frac{\sin(h)}{h} \geq \cos(h). \end{aligned}$$

Since $\lim_{h\to 0}\cos(h) = 1 = \lim_{h\to 0}\frac{1}{\cos(h)}$, one deduces [1] that

$$1 \geq \lim_{h\to 0}\frac{\sin(h)}{h} \geq 1$$

which leads to the statement. □

Let us now consider the exponential function. It has been shown during the tutorial session that the function $f:\mathbb{R}\to\mathbb{R}$ defined by $f(x) = \sum_{n=0}^{\infty}\frac{1}{n!}x^n$ for any $x\in\mathbb{R}$ satisfies the property $f' = f$. We call this function the *exponential function* and use the notation $f(x) = \mathrm{e}^x \equiv \exp(x)$. Let us stress that for the time being, this is just a notation for the function $f$, it has no other meaning. It is also easily observed that $\mathrm{e}^0 = 1$ and a numerical computation leads to $\mathrm{e}^1 = \sum_{n\in\mathbb{N}}\frac{1}{n!} \cong 2.718\ldots$.

Let us state and prove one result which says that, *modulo constant*, there exists only one function $f$ which satisfies $f' = f$. Here the expression modulo constant means that if $f$ has the mentioned property, then $\lambda f$ satisfies the same property for any $\lambda\in\mathbb{R}$.

**Lemma 2.22.** *Let $g:\mathbb{R}\to\mathbb{R}$ be a differentiable function satisfying $g' = g$. Then one has $g(x) = \lambda\mathrm{e}^x$ for some $\lambda\in\mathbb{R}$ and all $x\in\mathbb{R}$.*

*Proof.* The proof consists in showing that the ratio $\frac{g(x)}{\mathrm{e}^x}$ is a constant independent of $x$. Indeed, one has

$$\frac{\mathrm{d}}{\mathrm{d}x}\Big(\frac{g(x)}{\mathrm{e}^x}\Big) = \frac{g'(x)\mathrm{e}^x - g(x)\mathrm{e}^x}{(\mathrm{e}^x)^2} = \frac{g(x)\mathrm{e}^x - g(x)\mathrm{e}^x}{(\mathrm{e}^x)^2} = 0$$

since $(\mathrm{e}^x)' = \mathrm{e}^x$ and $g'(x) = g(x)$. Thus, one infers that $\frac{g(x)}{\mathrm{e}^x} = \lambda$ for some $\lambda\in\mathbb{R}$ and all $x\in\mathbb{R}$, which is equivalent to $g(x) = \lambda\mathrm{e}^x$. □

Let us stress that the above proof contains two facts which have not been proved yet. Firstly, we have assumed that $\mathrm{e}^x \neq 0$ for all $x\in\mathbb{R}$, which has not been proved so far. Also, we have used the fact that if a function on $\mathbb{R}$ has a derivative equal to 0, then the function is constant. This is obviously correct, but it has not been proved. We shall come back to these two preliminary statements in the near future.

[1] See Squeeze theorem.

# Chapter 3

# Mean value theorem

Our aim in this chapter is to prove the mean value theorem, a very useful tool for proving several results. Before the statement of this theorem, a few preliminary results are necessary.

## 3.1 Local maximum and minimum

In this first section we introduce the notion of local extremum (maximum or minimum), and emphasize that this notion is defined on any function, without requiring any continuity or differentiability. We recall that $I$ denotes either an open interval $(a, b)$ or a closed interval $[a, b]$.

**Definition 3.1** (Local extremum). *Let $f : I \to \mathbb{R}$ be a function. A point $x_0 \in I$ is a* local maximum *for $f$ on $I$ if there exists $\delta > 0$ such that $f(x_0) \geq f(x)$ for all $x \in (x_0 - \delta, x_0 + \delta) \cap I$. A point $x_0 \in I$ is a* local minimum *for $f$ on $I$ if there exists $\delta > 0$ such that $f(x_0) \leq f(x)$ for all $x \in (x_0 - \delta, x_0 + \delta) \cap I$.*

Let us observe that this definition corresponds to two possible situations: If $I$ is an open interval and $x_0 \in I$, then one can always choose $\delta > 0$ such that $(x_0 - \delta, x_0 + \delta) \subset I$. In such a case, the condition $(x_0 - \delta, x_0 + \delta) \cap I$ is trivially satisfied, and thus not necessary. On the other hand, if $I$ is a closed interval $[a, b]$ and if $x_0 = a$, then the condition $(x_0 - \delta, x_0 + \delta) \cap I$ is necessary, since the function is not defined for any $x < a$. A similar observation holds if $x_0 = b$.

Among the local extrema, one can often look for a global one.

**Definition 3.2** (Global extremum). *Let $f : I \to \mathbb{R}$ be a function. A point $x_0 \in I$ is a* global maximum *for $f$ on $I$ if $f(x_0) \geq f(x)$ for all $x \in I$. A point $x_0 \in I$ is a* global minimum *for $f$ on $I$ if $f(x_0) \leq f(x)$ for all $x \in I$.*

Once these notions are introduced, it is natural to wonder if they really exist ? The answer depend on $f$ and on the interval $I$. For example, the function $x \mapsto 1/x$ has no local of global maximum on $(0, 1)$, and no local or global minimum. On the other

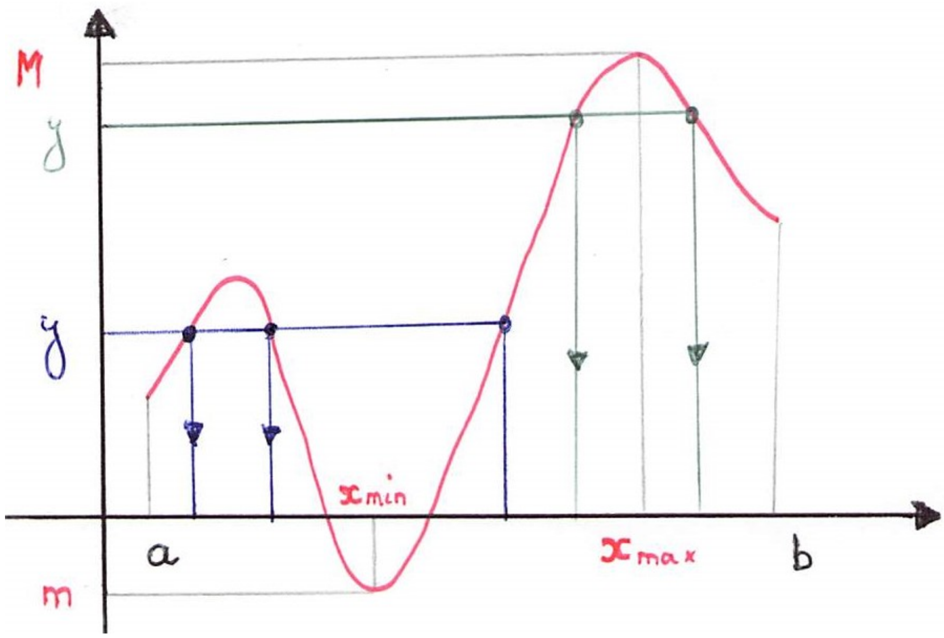


Fig. 3.1. An illustration of the Intermediate value theorem

hand, it has a global maximum and a global minimum on the interval $[1, 2]$. Another pathological example is provided by the function $f : (0, 1) \ni x \mapsto 3 \in \mathbb{R}$ with has a global maximum at any $x \in (0, 1)$ and a global minimum at any $x \in (0, 1)$, with the global maxima and the global minima equal to the same value 3. Note also that the function $x \mapsto \sin(x)$ has three global maxima on the interval $[0, 4\pi]$. These examples show that a global extremum might not exist, and if it exists it might not be unique. The definition does not require the uniqueness.

About the existence of global maxima or minima, the following result is important, and quite natural. Surprisingly, its proof is rather long and involved. We do not provide it, but refer to WIKIPEDIA for the interested reader.

**Theorem 3.3** (Extreme value theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then there exists $x_{max} \in [a, b]$ and $x_{min} \in [a, b]$ such that $x_{max}$ is a global maximum of $f$ on $[a, b]$ and $x_{min}$ is a global minimum of $f$ on $[a, b]$.*

Once again, let us emphasize that this statement does not mean that these global extrema are unique, it only says that they exist.

The following statement is very much related to the previous one. Its proof is related to the completeness of the real numbers, a subject which is slightly too advanced for this course. We refer again to WIKIPEDIA for its proof. Fortunately, the content of the theorem is very intuitive, see Figure 3.1.

**Theorem 3.4** (Intermediate value theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be a continuous function, and set $m := f(x_{min})$ and $M := f(x_{max})$, with $x_{min}$ and $x_{max}$ provided by Theorem 3.3. Then, for any $y \in [m, M]$ there exists $x \in [a, b]$ with $y = f(x)$.*

In other words, the statement says that all values inside $[m, M]$ are the image by $f$ of some $x \in [a, b]$. Alternatively, one can say that the range of $f$ is $[m, M]$.

Let us emphasize that for the previous two statements, only continuity was required. It is not necessary that $f$ is differentiable. From now on and for the following statement, differentiability will be necessary.

**Definition 3.5** (Critical point). *Let $f : (a, b) \to \mathbb{R}$ be a differentiable function. A point $x_0 \in (a, b)$ is* a critical point of $f$ *if $f'(x_0) = 0$.*

Let us provide a few simple examples:

**Examples 3.6.** *(i) For $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \cos(x)$ the points $k\pi$ with $k \in \mathbb{Z}$ are critical points since $\cos'(k\pi) = -\sin(k\pi) = 0$,*

*(ii) For $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = (x-1)^2$ the point $x = 1$ is a critical point since $f'(x) = 2(x-1)$ and $f'(1) = 0$,*

*(iii) For $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^3$ the point $x = 0$ is a critical point since $f'(x) = 3x^2$ and $f'(0) = 0$.*

The main use of critical points is described in the next statement.

**Theorem 3.7.** *Let $f : (a, b) \to \mathbb{R}$ be a differentiable function and let $x_0 \in (a, b)$ be a local maximum or a local minimum. Then $x_0$ is also a critical point.*

In other words, this statement says that for a differentiable function on an open interval $(a, b)$ the local extrema can be found among the set of critical points. On the other hand, the statement does not say anything about the points $a$ and $b$.

*Proof.* Let us prove the statement for a local maximum at $x_0 \in (a, b)$, since the proof for a local minimum is completely similar. For $h > 0$ small enough one has $x_0 + h \in (a, b)$ and $\frac{f(x_0+h)-f(x_0)}{h} \leq 0$ since $x_0$ is a local maximum. On the other hand, for $h < 0$ one has $\frac{f(x_0+h)-f(x_0)}{h} \geq 0$ since both the numerator and the denominator are negative. Since the function $f$ is differentiable at $x_0$ one has

$$0 \leq \lim_{h \nearrow 0} \frac{f(x_0+h) - f(x_0)}{h} = f'(x_0) = \lim_{h \searrow 0} \frac{f(x_0+h) - f(x_0)}{h} \leq 0$$

which implies that $f'(x_0) = 0$, as claimed. □

## 3.2 The mean value theorem

The first result is about the existence of critical points for a function taking the same value at two distinct points. Recall that a function $f$ is continuous on a close interval $[a, b]$ if it is continuous on $(a, b)$ and if $\lim_{x \searrow a} f(x) = f(a)$ and $\lim_{x \nearrow b} f(x) = f(b)$.

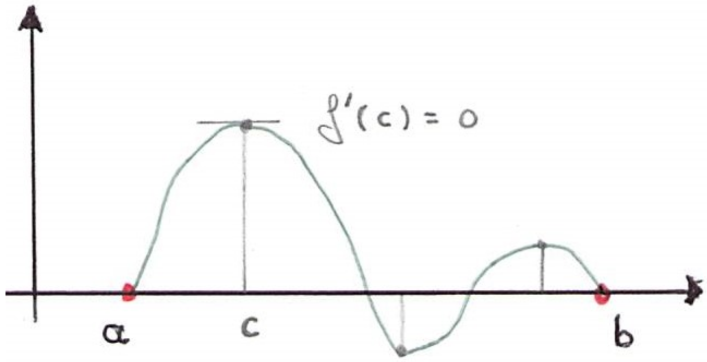


Fig. 3.2. An illustration of Rolle's theorem

**Theorem 3.8** (Rolle's theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on $(a, b)$. Assume also that $f(a) = 0 = f(b)$. Then there exists $c \in (a, b)$ with $f'(c) = 0$.*

An illustration of this statement is provided in Figure 3.2.

*Proof.* If $f = 0$ (the constant function equal to 0), then $f'(x) = 0$ for any $x \in (a, b)$, and thus the statement is proved. So, let us assume that $f \neq 0$, which means that there exists $x \in (a, b)$ with $f(x) \neq 0$. Suppose that $f(x) > 0$. Then, by the Extreme value theorem 3.3 there exists a global maximum $c$ on $(a, b)$. Note that $c \neq a$ and $c \neq b$ since $f(a) = f(b) = 0$ while $f(x) > 0$. By Theorem 3.7 one infers that $f'(c) = 0$, as claimed. If $f(x) < 0$, the proof is similar with a global minimum instead of a global maximum. □

Let us now consider a slightly different setting, with $f(a) \neq f(b)$, see Figure 3.3. In this figure, we have assumed that $f(b) > f(a)$. If we consider the right triangle in $\mathbb{R}^2$ with coordinates $\big(a, f(a)\big)$, $\big(b, f(a)\big)$, and $\big(b, f(b)\big)$, the slope of its hypothenuse is given by $\frac{f(b)-f(a)}{b-a}$. The mean value theorem states that there exists $c \in (a, b)$ with $f'(c) = \frac{f(b)-f(a)}{b-a}$. More precisely:

**Theorem 3.9** (Mean value theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on $(a, b)$. Then there exists $c \in (a, b)$ with $f'(c) = \frac{f(b)-f(a)}{b-a}$.*

*Proof.* Consider first the equation of the straight line passing through the points of coordinates $\big(a, f(a)\big)$ and $\big(b, f(b)\big)$:

$$y = \frac{f(b) - f(a)}{b - a}(x - a) + f(a).$$

We now define a new function $h : [a, b] \to \mathbb{R}$ given by the difference between $f$ and the straight line defined above, namely:

$$h(x) = f(x) - \Big(\frac{f(b) - f(a)}{b - a}(x - a) + f(a)\Big).$$

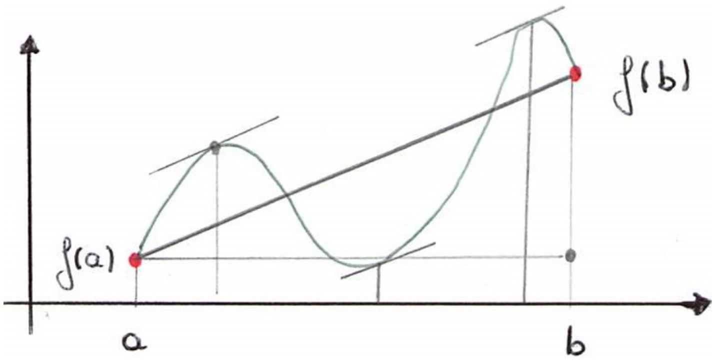


Fig. 3.3. An illustration of the mean value theorem

One easily checks that $h(a) = 0 = h(b)$, and that $h$ is differentiable on the interval $(a, b)$. Thus, by Rolle's theorem, there exists $c \in (a, b)$ with $h'(c) = 0$. By computing explicitly $h'$ and by evaluating this derivative at $c$ it then follows that

$$f'(c) - \frac{f(b) - f(a)}{b - a} = 0,$$

which corresponds to the statement of the theorem. □

As already mentioned, this result has several applications. We now provide two of them. The following result is related to a missing argument in the proof of Lemma 2.22.

**Corollary 3.10.** *Let $f : (a, b) \to \mathbb{R}$ be differentiable with $f'(x) = 0$ for any $x \in (a, b)$. Then $f$ is a constant function.*

*Proof.* Choose any $a', b' \in (a, b)$ with $a < a' < b' < b$. The restriction of the function $f$ to the interval $[a', b']$ is continuous, and differentiable on $(a', b')$. Thus, by the mean value theorem, there exists $c \in (a', b')$ with $f'(c) = \frac{f(b')-f(a')}{b'-a'}$. However, since $f'(c) = 0$, it follows that $f(b') = f(a')$. Since the two points $a', b'$ are arbitrary, one infers that $f$ takes the same value on any point of $(a, b)$, and therefore $f$ is a constant function. □

For the second statement, let us introduce the concept of an increasing (or decreasing) function.

**Definition 3.11** (Increasing, decreasing). *A function $f : I \to \mathbb{R}$ is* increasing *if $f(x_1) \leq f(x_2)$ whenever $x_1, x_2 \in I$ satisfy $x_1 < x_2$. The function is* strictly increasing *if $f(x_1) < f(x_2)$ whenever $x_1, x_2 \in I$ satisfy $x_1 < x_2$. Conversely, the function $f$ is* decreasing *if $f(x_1) \geq f(x_2)$ whenever $x_1, x_2 \in I$ satisfy $x_1 < x_2$. The function is* strictly decreasing *if $f(x_1) > f(x_2)$ whenever $x_1, x_2 \in I$ satisfy $x_1 < x_2$.*

Observe that for this definition, it is not necessary that the function is differentiable, or even continuous. However, if the function is differentiable, one has:

**Lemma 3.12.** *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on $(a, b)$. If $f'(x) \geq 0$ for any $x \in (a, b)$, then the function is increasing on $[a, b]$, and strictly increasing if $f'(x) > 0$. Conversely, if $f'(x) \leq 0$ for any $x \in (a, b)$, then the function is decreasing on $[a, b]$, and strictly decreasing if $f'(x) < 0$.*

*Proof.* We only consider the first situation, namely $f'(x) \geq 0$ or $f'(x) > 0$. Choose $x_1, x_2 \in (a, b)$ with $x_1 < x_2$. By the mean value theorem applied to $f$ restricted to $[x_1, x_2]$, there exists $c \in (x_1, x_2)$ such that $f'(c) = \frac{f(x_2)-f(x_1)}{x_2-x_1}$. If $f'(x) \geq 0$, it follows that $f(x_2) \geq f(x_1)$, while if $f'(x) > 0$ one infers that $f(x_2) > f(x_1)$. Since $x_1$ and $x_2$ are arbitrary, it follows that $f$ is either increasing or strictly increasing, depending on the inequality (not strict or strict). Note that the decreasing situation can be proved with a similar argument. □

Let us still derive a slightly stronger version of the mean value theorem. It will then be used for a precise proof of L'Hôpital's rule.

**Theorem 3.13** (Cauchy mean value theorem)**.** *Let $f, g : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on $(a, b)$. Then there exists $c \in (a, b)$ such that*

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

*Proof.* We shall consider two cases. Firstly, if $g(b) = g(a)$, then the existence of $c$ such that $g'(c) = 0$ can be easily deduced from Rolle's theorem by considering the shifted function $g - g(a)$.

We can now assume that $g(b) \neq g(a)$. Let us define the function $h : [a, b] \to \mathbb{R}$ by

$$h(x) := f(x) - \frac{f(b) - f(a)}{g(b) - g(a)} g(x) - \frac{f(a)g(b) - f(b)g(a)}{g(b) - g(a)}, \quad \forall x \in [a, b].$$

Clearly, the function $h$ is differentiable on $(a, b)$, and satisfies $h(a) = h(b) = 0$. By Rolle's theorem applied to $h$ one infers that there exists $c \in (a, b)$ with $h'(c) = 0$. However, this is equivalent to

$$f'(c) - \frac{f(b) - f(a)}{g(b) - g(a)} g'(c) = 0 \Longleftrightarrow f'(c)\big(g(b) - g(a)\big) = g'(c)\big(f(b) - f(a)\big),$$

as stated. □

*Proof of L'Hôpital's rule, Theorem 2.15.* In the framework of this statement one has

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(x_0)}{g(x) - g(x_0)} = \frac{f'(c)}{g'(c)}$$

for some $c \in (x_0, x)$, as provided by Theorem 3.13. Since $x$ approaching $x_0$ implies that $c$ is also approaching $x_0$, one infers that

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = \lim_{c \to x_0} \frac{f'(c)}{g'(c)} = \lim_{x \to x_0} \frac{f'(x)}{g'(x)}$$

which exists, by the assumption of the theorem. □

## 3.3 Sketching graphs

With the notions introduced so far, the graph of a function can be drawn quite precisely. We shall not develop this subject here, but just list the main points, and refer to Chapter VI of the book of Lang for more details. The key points for drawing the graph of a differentiable function $f$ are:

(i) Compute the intersections of the function with the coordinate axes (when $x = 0$ or when $f(x) = 0$),

(ii) Locate the critical points, see Definition 3.5,

(iii) Among the critical points, determine the local maxima and local minima, see Definition 3.1,

(iv) Determine the asymptotic behavior, as for example $\lim_{x\to\pm\infty} f(x)$, $\lim_{x\searrow x_0} f(x)$, or $\lim_{x\nearrow x_0} f(x)$ whenever these expressions are meaningful, and when $x_0$ is a singular point for $f$,

(v) Determine the region of increase or decrease for $f$, see Definition 3.11.

Let us finally mention that if the function $f$ is twice differentiable, then the second derivative also provides some information. For example, if $x_0$ is a critical point for $f$, the following properties hold: if $f''(x_0) > 0$ then $x_0$ is a local minimum, while if $f''(x_0) < 0$ then $x_0$ is a local maximum. If $f''(x_0) = 0$, one can not conclude if $x_0$ is a local maximum, a local minimum, or none of them. In this situation, further analysis is necessary. As an example of this situation, the functions $x \mapsto x^3$ and $x \mapsto x^4$ possess a critical point at $x_0 = 0$ but the behaviors of these functions at this point are quite different.

# Chapter 4

# Inverse functions

In this chapter we shall deal with the following important question: given $f : I \to \mathbb{R}$ and any $y \in \mathbb{R}$, is it possible to find explicitly $x$ such that $f(x) = y$, or more precisely can we solve $x = g(y)$ for some function $g$ ? Clearly, the answer is sometimes *yes* and sometimes *no*, depending on $f$, on its domain $\mathrm{Dom}(f)$, on its codomain and on its range $\mathrm{Ran}(f)$, see Definition 1.7 for these concepts. In addition, if the inverse function exists, we shall also derive some of its properties. A few examples will be considered.

## 4.1 The inverse function

Let us be more precise about the definition of the inverse function. We are looking for a function $g$ having the following properties:

$$g\big(f(x)\big) = x \text{ for all } x \in \mathrm{Dom}(f) \quad \text{and} \quad f\big(g(y)\big) = y \text{ for all } y \in \mathrm{Ran}(f). \tag{4.1}$$

Note that the function $g$ has to be defined on $\mathrm{Ran}(f)$ in order to give a meaning to these conditions. Before looking at a necessary condition for the existence of $g$, let us consider two examples:

**Examples 4.1.** *(i) Consider* $f : [0, \infty) \to \mathbb{R}$ *with* $f(x) := x^2$*. Then the function* $g : [0, \infty) \to [0, \infty)$ *given by* $g(x) := \sqrt{x}$ *satisfies the two conditions stated in* (4.1) *since* $g\big(f(x)\big) = \sqrt{x^2} = x$ *for any* $x \in [0, \infty)$ *and* $\big(\sqrt{y}\big)^2 = y$ *for any* $y \in [0, \infty)$*. Here, we have used that* $\mathrm{Ran}(f) = [0, \infty)$*.*

*(ii) Consider* $f : \mathbb{R} \to \mathbb{R}$ *with* $f(x) := x^2$*. In this case, the previous function* $g$ *does not verify the first condition of* (4.1) *since one has* $\sqrt{(-2)^2} = 2 \neq -2$*. In fact, in this example it is impossible to find a function* $g$ *verifying the two conditions of* (4.1)*.*

With the notions of injectivity, surjectivity and bijectivity introduced in Definition 1.8 it is rather easy to understand the problem with the second example. Indeed, as already mentioned a function is always surjective on its image, which means that for any $y \in \mathrm{Ran}(f)$ one can always find $x \in \mathrm{Dom}(f)$ such that $f(x) = y$. Because of this, a function $g : \mathrm{Ran}(f) \to \mathrm{Dom}(f)$ given for $y \in \mathrm{Ran}(f)$ by $g(y) := x$ (the $x$ mentioned in the previous sentence) can always be defined, and satisfy the relation $f\big(g(y)\big) = y$. Thus, the second condition of (4.1) can always be satisfied for a certain function $g$. As a consequence, the main problem is the first condition in (4.1), which is related to the notion of injectivity. Indeed, if there exist $x_1, x_2 \in \mathrm{Dom}(f)$ with $x_1 \neq x_2$ but which satisfy $f(x_1) = f(x_2)$, then it is impossible to define a function $g : \mathrm{Ran}(f) \to \mathrm{Dom}(f)$ verifying $g\big(f(x_1)\big) = x_1$ and $g\big(f(x_2)\big) = x_2$, since this would contradict that the function $g$ associates to any point on its domain *a single value* in its codomain. By putting the two observations together one infers that there exists a function $g : \mathrm{Ran}(f) \to \mathrm{Dom}(f)$ satisfying (4.1) if and only if $f$ is bijective. In fact, whenever such a function $g$ exists one easily proves that this function is unique. As a consequence, one sets:

**Definition 4.2** (Inverse of a function)**.** *For any function $f : \mathrm{Dom}(f) \to \mathrm{Ran}(f)$ which is bijective, the only function $g : \mathrm{Ran}(f) \to \mathrm{Dom}(f)$ which satisfies (4.1) is called* the inverse of $f$ *and is denoted by $f^{-1}$.*

Let us immediately stress that the notation is ambiguous and that one has to be careful. Indeed, $f^{-1}$ denotes the inverse of $f$, but depending on the context it can also denote the function $\frac{1}{f}$. Unfortunately, both notations are commonly used, but usually the context provides sufficient information for deciding the correct meaning of the expression. In fact, the notation $f^{-1}$ for $\frac{1}{f}$ corresponds to the inverse function with respect to the multiplication operation (since $f \cdot \frac{1}{f} = 1$) while the notation $f^{-1}$ for the function $g$ satisfying (4.1) corresponds to the inverse for the composition of functions since $f \circ f^{-1} = \mathrm{id}$ and $f^{-1} \circ f = \mathrm{id}$ with id the identity function defined by $\mathrm{id}(x) = x$. Note that with this function the two relations (4.1) simply read

$$f^{-1} \circ f = \mathrm{id} \qquad f \circ f^{-1} = \mathrm{id}. \tag{4.2}$$

In the next statement, we provide a sufficient condition for the existence of an inverse. Recall that the notion of a strictly increasing or strictly decreasing function has been introduced in Definition 3.11.

**Proposition 4.3.** *Let $f : [a, b] \to \mathbb{R}$ be a continuous and strictly increasing function, and set $\alpha := f(a)$ and $\beta := f(b)$. Then $f$ is bijective, and therefore there exists an inverse function $f^{-1} : [\alpha, \beta] \to [a, b]$ such that the conditions (4.2) are satisfied. Similarly, if $f$ is continuous and strictly decreasing, then $f$ is bijective, and therefore there exists an inverse function $f^{-1} : [\beta, \alpha] \to [a, b]$ such that the conditions (4.2) are satisfied.*

The proof of this statement will be studied during the tutorial session. In fact, the previous statement can be strengthened a little bit:

**Exercise 4.4.** *In the framework of the previous proposition, show that the inverse is a continuous function.*

Let us still provide one geometric characterization of the inverse function. For this we recall that the graph of a function has been introduced in (1.2). In the sequel, we shall write $\mathrm{graph}(f)$ for the graph of a function.

**Theorem 4.5** (Reflection theorem)**.** *Let $f : [a, b] \to [\alpha, \beta]$ be a bijective function, with inverse $f^{-1} : [\alpha, \beta] \to [a, b]$. Then, for any $(x, y) \in [a, b] \times [\alpha, \beta]$ one has*

$$(x, y) \in \mathrm{graph}(f) \iff (y, x) \in \mathrm{graph}(f^{-1}).$$

By using the bijectivity of $f$ the proof of this statement is rather easy and is left as an exercise.

## 4.2 Derivative of the inverse function

We shall now study the derivative of the inverse function, in the framework of Proposition 4.3. In the next statement, we shall assume that $f$ is strictly increasing, but a similar statement holds if $f$ is strictly decreasing.

**Theorem 4.6.** *Let $f : [a, b] \to \mathbb{R}$ be continuous and differentiable on $(a, b)$ with $f'(x) > 0$ for all $x \in (a, b)$. Set $\alpha := f(a)$ and $\beta := f(b)$, and let $f^{-1} : [\alpha, \beta] \to [a, b]$ be the inverse function. Then $f^{-1}$ is continuous on $[\alpha, \beta]$ and differentiable on $(\alpha, \beta)$ with*

$$\big(f^{-1}\big)'(y) = \frac{1}{f'\big(f^{-1}(y)\big)}, \qquad \forall y \in (\alpha, \beta). \tag{4.3}$$

*Proof.* Let us fix $y \in (\alpha, \beta)$ and $\delta > 0$ such that $y + h \in (\alpha, \beta)$ for all $|h| < \delta$. We also set $x := f^{-1}(y)$ and $x + k := f^{-1}(y + h)$, which exist by Proposition 4.3. Since $f^{-1}$ is continuous, by Exercise 4.4 it follows that

$$\lim_{h\to 0} k = \lim_{h\to 0} \big(f^{-1}(y + h) - f^{-1}(y)\big) = 0$$

and therefore $k$ goes to 0 as $h$ goes to 0. Note also that $h = (y+h)-y = f(x+k)-f(x)$, and therefore one has

$$\begin{aligned}\lim_{h\to 0} \frac{f^{-1}(y + h) - f^{-1}(y)}{h} &= \lim_{h\to 0} \frac{k}{f(x + k) - f(x)} = \lim_{k\to 0} \frac{1}{\frac{f(x+k)-f(x)}{k}} \\ &= \frac{1}{f'(x)} = \frac{1}{f'\big(f^{-1}(y)\big)},\end{aligned}$$

as expected. □

Let us now consider a few examples.

**Examples 4.7.** *(i) Consider* $\sin : \big[-\frac{\pi}{2},\frac{\pi}{2}\big] \ni x \mapsto \sin(x) \in [-1,1]$*. For any* $x \in \big(-\frac{\pi}{2},\frac{\pi}{2}\big)$ *one has* $\sin'(x) = \cos(x) > 0$*, which implies by Theorem 4.6 that* $\sin$ *(defined on the domain* $\big[-\frac{\pi}{2},\frac{\pi}{2}\big]$*) is invertible, with inverse denoted by* $\sin^{-1}$*, or more commonly by* $\arcsin$*. By the previous theorem one also has for any* $y \in (-1,1)$

$$\arcsin'(y) = \frac{1}{\sin'(\arcsin(y))} = \frac{1}{\cos(\arcsin(y))} = \frac{1}{\sqrt{1-\sin(\arcsin(y))^2}} = \frac{1}{\sqrt{1-y^2}}.$$

*(ii) Consider* $\cos : [0,\pi] \ni x \mapsto \cos(x) \in [-1,1]$*. For any* $x \in (0,\pi)$ *one has* $\cos'(x) = -\sin(x) < 0$*, which implies by an adaptation of Theorem 4.6 to strictly decreasing functions that* $\cos$ *(defined on the domain* $[0,\pi]$*) is invertible, with inverse denoted by* $\cos^{-1}$*, or more commonly by* $\arccos$*. By the previous theorem one also has for any* $y \in (-1,1)$

$$\arccos'(y) = -\frac{1}{\sin(\arccos(y))} = -\frac{1}{\sqrt{1-\cos(\arccos(y))^2}} = -\frac{1}{\sqrt{1-y^2}}.$$

*(iii) Consider* $\tan : \big(-\frac{\pi}{2},\frac{\pi}{2}\big) \ni x \mapsto \tan(x) \in \mathbb{R}$*. For any* $x \in \big(-\frac{\pi}{2},\frac{\pi}{2}\big)$ *one has* $\tan'(x) = 1+\tan^2(x) > 1$*, which implies by Theorem 4.6 that* $\tan$ *(defined on the domain* $\big(-\frac{\pi}{2},\frac{\pi}{2}\big)$*) is invertible, with inverse denoted by* $\tan^{-1}$*, or more commonly by* $\arctan$*. By the previous theorem one also has for any* $y \in \mathbb{R}$

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

## 4.3 Exponential and logarithm functions

In this section we thoroughly introduce the exponential function and its inverse, the logarithm function. Recall that the exponential function has been introduced as the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) := \sum_{n=0}^{\infty} \frac{1}{n!} x^n \quad \text{for any } x \in \mathbb{R}. \tag{4.4}$$

It has been shown during the tutorial session that this function satisfies $f' = f$. In Lemma 2.22, it has also been proved that *modulo a multiplicative constant* there exists only one function satisfying the relation $f' = f$. Thus, the exponential function defined in (4.4) corresponds to the only function satisfying $f' = f$ together with the normalization $f(0) = 1$. As already mentioned, we shall use the notation $\mathrm{e}^x \equiv \exp(x)$ for this function, which means that $\exp : \mathbb{R} \to \mathbb{R}$ with $\mathrm{e}^x$ given by the r.h.s. of (4.4). However, so far the notation $\mathrm{e}^x$ has no special meaning: it is just the function defined by the r.h.s. of (4.4).

The main properties of this function are gathered in the next statement.

**Lemma 4.8.** *One has for any $x, y \in \mathbb{R}$*

*(i)* $\mathrm{e}^x \mathrm{e}^{-x} = 1$,

*(ii)* $\mathrm{e}^x > 0$,

*(iii) The function* $\mathbb{R} \ni x \mapsto \mathrm{e}^x \in \mathbb{R}$ *is strictly increasing,*

*(iv)* $\mathrm{e}^{x+y} = \mathrm{e}^x \mathrm{e}^y$.

Note that property $(ii)$ has already been used in the proof of Lemma 2.22. Also, property $(i)$ gives a justification of the equality $\mathrm{e}^{-x} = \frac{1}{\mathrm{e}^x} = (\mathrm{e}^x)^{-1}$.

*Proof.* $(i)$ By the product rule together with the composition rule one has

$$\big(\mathrm{e}^x \mathrm{e}^{-x}\big)' = (\mathrm{e}^x)'\mathrm{e}^{-x} + \mathrm{e}^x(\mathrm{e}^{-x})' = \mathrm{e}^x\mathrm{e}^{-x} - \mathrm{e}^x\mathrm{e}^{-x} = 0 \quad \forall x \in \mathbb{R}.$$

It then follows that the function $x \mapsto \mathrm{e}^x\mathrm{e}^{-x}$ is a constant function, see Corollary 3.10. Now, this constant can be determined since $\mathrm{e}^0\mathrm{e}^{-0} = \big(\mathrm{e}^0\big)^2 = 1^2 = 1$.

$(ii)$ It follows from $(i)$ that $\mathrm{e}^x \neq 0$ for any $x \in \mathbb{R}$. By the Intermediate value theorem 3.4 one infers that $\mathrm{e}^x > 0$ for all $x \in \mathbb{R}$, or $\mathrm{e}^x < 0$ for all $x \in \mathbb{R}$. However, since $\mathrm{e}^0 = 1$, the second alternative can be ruled out, and therefore $\mathrm{e}^x > 0$ for all $x \in \mathbb{R}$.

$(iii)$ Since $\big(\mathrm{e}^x\big)' = \mathrm{e}^x > 0$ (by $(ii)$), it follows that the exponential function is strictly increasing.

$(iv)$ For fixed $x \in \mathbb{R}$, let us consider the function $\phi : \mathbb{R} \ni y \mapsto \phi(y) := \frac{\mathrm{e}^{x+y}}{\mathrm{e}^y} \in \mathbb{R}$. By computing its derivative (with respect to $y$) one has

$$\phi'(y) = \frac{(\mathrm{e}^{x+y})'\mathrm{e}^y - \mathrm{e}^{x+y}(\mathrm{e}^y)'}{(\mathrm{e}^y)^2} = \frac{\mathrm{e}^{x+y}\mathrm{e}^y - \mathrm{e}^{x+y}\mathrm{e}^y}{(\mathrm{e}^y)^2} = 0.$$

Thus, one infers again from Corollary 3.10 that $\phi$ is a constant function, and therefore $\phi(y) = \phi(0)$ for any $y \in \mathbb{R}$. Since $\phi(0) = \mathrm{e}^x$ one deduces that $\frac{\mathrm{e}^{x+y}}{\mathrm{e}^y} = \mathrm{e}^x$, as planned. ♡

From the above property $(iii)$ and from Theorem 4.6 one infers that the exponential function is invertible. In order to determine its inverse, the range of the exponential function has to be computed first. In fact, one easily observes that $x \mapsto \mathrm{e}^x$ can take arbitrary large values, which means that for any large $M \in \mathbb{R}$, there exists $x \in \mathbb{R}$ such that $\mathrm{e}^x > M$. Then, from the above property (i), it follows that $\mathrm{e}^{-x}$ can be made arbitrarily close to 0. From this observation one infers that $\mathrm{Ran}(\exp) = (0, \infty)$. Thus, the inverse of the exponential function is a function from $(0, \infty)$ to $\mathbb{R}$, and this function is called *the logarithm function*. More precisely, one has $\ln : (0, \infty) \to \mathbb{R}$ with $\mathrm{e}^{\ln(y)} = y$ for any $y \in (0, \infty)$, and $\ln\big(\mathrm{e}^x\big) = x$ for any $x \in \mathbb{R}$. From Theorem 4.6, it also follows that the logarithm function is a differentiable function on $(0, \infty)$.

Let us now state a few properties of the logarithm function. These relations can be deduced from Lemma 4.8, and they will be proved at the tutorial session.

**Lemma 4.9.** *For any $x, y \in (0, \infty)$ and $q \in \mathbb{Q}$ one has* *(i)* $\ln(x)' = \frac{1}{x}$,

*(ii)* $\ln(xy) = \ln(x) + \ln(y)$,

*(iii)* $\ln(x^q) = q\ln(x)$.

The property *(iii)* is going to play an important role for answering a question already raised in Chapter 1, namely how to define $\pi^{\sqrt{2}}$ or $3^\pi$ ? However, in order to do it properly, one property of continuous functions has to stated. We present it in an exercise:

**Exercise 4.10.** *Prove the following (closely related) statements:*

*(i) Let $f : (a, c) \to \mathbb{R}$ be a continuous function at $b \in (a, c)$, and let $(x_n)_{n\in\mathbb{N}}$, with $x_n \in (a, c)$, be a convergent sequence with $\lim_{n\to\infty} x_n = b$. Show that the following equality holds:*

$$\lim_{n\to\infty} f(x_n) = f(b).$$

*(ii) Let $f : (a, c) \to \mathbb{R}$ be a continuous function at $b \in (a, c)$, and let $g : (\alpha, \gamma) \to (a, c)$ be another function satisfying $\lim_{x\to\beta} g(x) = b$ for some $\beta \in (\alpha, \gamma)$. Show that the following equality holds:*

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

*With a rough notation, these statements read*

$$\lim_{n\to\infty} f(x_n) = f\big(\lim_{n\to\infty} x_n\big) \qquad \textit{and} \qquad \lim_{x\to\beta} f\big(g(x)\big) = f\big(\lim_{x\to\beta} g(x)\big). \tag{4.5}$$

*Let us stress that these properties hold because $f$ is continuous at $b$, for arbitrary functions these equalities are not correct.*

Let us now recall that for any $\alpha \in \mathbb{R}$, there exists a sequence $(q_n)_{n\in\mathbb{N}}$ with $q_n \in \mathbb{Q}$ such that $\lim_{n\to\infty} q_n = \alpha$. This sequence corresponds to an approximation of any real number by rational numbers. Clearly, this sequence is not unique, but this non-uniqueness does not play any role. Thus, if we use the rough notation introduced above together with the property *(iii)* of Lemma 4.9 one has for $x \in (0, \infty)$:

$$x^\alpha := \lim_{n\to\infty} x^{q_n} = \mathrm{e}^{\ln(\lim_{n\to\infty} x^{q_n})} = \mathrm{e}^{\lim_{n\to\infty} \ln(x^{q_n})} = \mathrm{e}^{\lim_{n\to\infty} q_n \ln(x)} = \mathrm{e}^{\alpha\ln(x)}$$

where we have used (4.5) in the third equality, since the logarithm function is continuous. Based on these equalities, it is now natural to define precisely $x^\alpha$:

**Definition 4.11.** *For any $x \in (0, \infty)$ and any $\alpha \in \mathbb{R}$ one sets*

$$x^\alpha := \mathrm{e}^{\alpha\ln(x)}. \tag{4.6}$$

Thus, with this expression we can give a meaning to $x^\alpha$ for any $\alpha \in \mathbb{R}$, and not only in $\mathbb{Q}$. Clearly, if $\alpha \in \mathbb{Q}$ the above definition corresponds to the previous one, since $\mathrm{e}^{\alpha\ln(x)} = \mathrm{e}^{\ln(x^\alpha)} = x^\alpha$, by property *(iii)* of Lemma 4.9. As a consequence, the above definition is an extension of the one seen in Chapter 1. Note that a few additional properties of the exponential function will be studied at the tutorial session.

# Chapter 5

# Integration

The aim of this chapter is to provide an operation which is essentially the inverse of taking a derivative. We shall also use this operation for computing some areas.

## 5.1 The indefinite integral

We start with the main definition of this section:

**Definition 5.1** (Indefinite integral)**.** *Let $f : (a, b) \to \mathbb{R}$ be a function. An* indefinite integral for *$f$ is a differentiable function $F : (a, b) \to \mathbb{R}$ which satisfies $F' = f$. By extension, for $f : [a, b] \to \mathbb{R}$, an* indefinite integral *is a function $F : [a, b] \to \mathbb{R}$ differentiable on $(a, b)$ and satisfying $F'(x) = f(x)$ for any $x \in (a, b)$. The notations $\int f$ of $\int f(x)\mathrm{d}x$ are used for representing an indefinite integral for $f$.*

An indefinite integral is also called an *antiderivative* or a *primitive function* for $f$. Note that this definition does not imply the existence or the uniqueness of an indefinite integral for an arbitrary function $f$.

**Examples 5.2.** *(i) For $n \in \mathbb{N}$ and $f : \mathbb{R} \ni x \mapsto x^n \in \mathbb{R}$, an indefinite integral is given by the function $F : \mathbb{R} \ni x \mapsto \frac{1}{n+1}x^{n+1}$. Note that the function $F + c$, where $c$ is any constant, is also an indefinite integral for $f$.*

*(ii) For $f : [0, 2\pi] \ni x \mapsto \cos(x) \in \mathbb{R}$, an indefinite integral is given by the function $F : [0, 2\pi] \ni x \mapsto \sin(x) + 3 \in \mathbb{R}$. Note that $3$ could be replaced by any number, and the new function would also be an indefinite integral for $f$.*

*(iii) For $f : (0, \infty) \ni x \mapsto \frac{1}{x} \in \mathbb{R}$, the function $F : (0, \infty) \ni x \mapsto \ln(x) + \mathrm{cst} \in \mathbb{R}$, where* cst *means any constant, is an indefinite integral for $f$.*

It clearly follows from these examples that if $F$ is an indefinite integral for $f$, then the function $F + \mathrm{cst}$ is also an indefinite integral for $f$. In fact, on an interval, all indefinite integrals for $f$ differ by a constant. Indeed, suppose that $F_1$ and $F_2$ are indefinite integrals for $f$ on $(a, b)$. By assumption $F_1' = f$ and $F_2' = f$, from which one infers that $(F_1 - F_2)'(x) = 0$ for all $x \in (a, b)$. It then follows from Corollary 3.10 that $F_1 - F_2 = \text{cst}$, which means $F_1 = F_2 + \text{cst}$. However, if $f$ is defined on two disjoint intervals, then two indefinite integrals for $f$ might differ more. For example, consider $f : (-\pi, 0) \cup (0, \pi) \ni x \mapsto \cos(x) \in \mathbb{R}$, then

$$F : (-\pi, 0) \cup (0, \pi) \ni x \mapsto \begin{cases} \sin(x) + \text{cst}_1 & \text{if } x \in (-\pi, 0) \\ \sin(x) + \text{cst}_2 & \text{if } x \in (0, \pi). \end{cases}$$

On this example and because the two intervals are disjoint, one can choose two constants. Thus, the indefinite integrals for $f$ are not indexed by one constant, but by two of them. Another example is provided by the function $f : \mathbb{R}^* \ni x \mapsto \frac{1}{x} \in \mathbb{R}$, what is an indefinite integral for $f$ ? In fact, one easily observes that the following function $F$ is the most general indefinite integral for $f$:

$$F : \mathbb{R}^* \ni x \mapsto \begin{cases} \ln(x) + \text{cst}_1 & \text{if } x > 0 \\ \ln(-x) + \text{cst}_2 & \text{if } x < 0. \end{cases}$$

Here, *most general* means that any indefinite integral for $f$ is obtained from this general expression by fixing the constants $\text{cst}_1$ and $\text{cst}_2$.

## 5.2 Areas

Given a function $f : [a, b] \to \mathbb{R}$, it is not clear how an indefinite integral for $f$ can be found. In this section, we show the relation between an indefinite integral, and the area below a graph. However, this section is a little bit sloppy ☹, but the following one will contain a more rigorous approach ☺.

Let us consider a function $f : [a, b] \to \mathbb{R}_+$ which is continuous. We recall that $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$. Let us also define $F : [a, b] \to \mathbb{R}_+$, given for $x \in [a, b]$ by

$$F(x) = \text{ the area below the graph of } f \text{ between } a \text{ and } x. \tag{5.1}$$

A representation of the function $F$ is provided in Figure 5.1. Clearly, one has $F(a) = 0$ and $F(b)$ is equal to the area below the graph of $f$ between $a$ and $b$. Our main interest for $F$ is explained by the following statement:

**Theorem 5.3.** *Let $f : [a, b] \to \mathbb{R}_+$ be continuous and let $F$ be defined as explained above. Then $F : [a, b] \to \mathbb{R}_+$ is differentiable on $(a, b)$ and satisfies $F'(x) = f(x)$ for any $x \in (a, b)$.*

In other words, the function $F$ is one indefinite integral for $f$. In fact, it is not an arbitrary one, it is the indefinite integral which satisfies $F(a) = 0$. Before providing the proof of this Theorem, let us point out the weakness of the statement: How is *the area below the graph of $f$* defined ? We haven't described this concept very precisely, and it is more or less an intuitive definition. The role of the next section will be to give a precise meaning to this concept. However, we can already give a proof of this theorem, which will be correct once the good definition of area is provided.

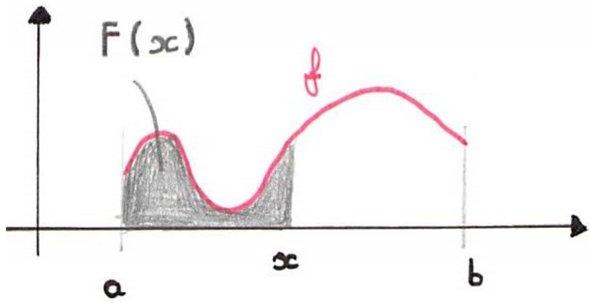


Fig. 5.1. A representation of the function $F$

*Proof.* Let $x \in (a, b)$ and consider $h > 0$ small enough such that $x + h \in (a, b)$ (a similar argument holds for $h < 0$ small enough). One has to consider $\frac{F(x+h)-F(x)}{h}$. For that purpose, let us recall from the Extreme value theorem 3.3 that there exists $x_{max}(h) \in [x, x + h]$ and $x_{min}(h) \in [x, x + h]$ such that

$$f\big(x_{min}(h)\big) \leq f(y) \leq f\big(x_{max}(h)\big) \qquad \forall y \in [x, x + h].$$

Note that we have indicated that $x_{max}$ and $x_{min}$ depend on $h$ because they are related to the interval $[x, x + h]$. Then by looking at the areas in Figure 5.2 one infers that

$$hf\big(x_{min}(h)\big) \leq F(x + h) - F(x) \leq hf\big(x_{max}(h)\big) \tag{5.2}$$

where the left and the right terms represent the areas of some rectangles, while the term in the middle represent the area below the graph of $f$ between $x$ and $x+h$. These inequalities are equivalent to

$$f\big(x_{min}(h)\big) \leq \frac{F(x + h) - F(x)}{h} \leq f\big(x_{max}(h)\big).$$

Since $f$ is continuous, and since $x_{min}(h)$ and $x_{max}(h)$ belong to the inteval $[x, x + h]$ one infers that

$$\lim_{h\to 0} f\big(x_{min}(h)\big) = f(x) = \lim_{h\to 0} f\big(x_{max}(h)\big).$$

Finally, by the Squeeze theorem one gets

$$f(x) = \lim_{h\to 0} f\big(x_{min}(h)\big) \leq \lim_{h\to 0} \frac{F(x + h) - F(x)}{h} \leq \lim_{h\to 0} f\big(x_{max}(h)\big) = f(x),$$

which means that $F'(x) := \lim_{h\to 0} \frac{F(x+h)-F(x)}{h}$ exists, and is equal to $f(x)$. Since $x$ is arbitrary in $(a, b)$ this proves the statement. □

Consider again the continuous function $f : [a, b] \to \mathbb{R}$, and let $G$ be an arbitrary indefinite integral for $f$. By the discussion of the previous section, one has $G = F + c$, where $F$ is the indefinite integral defined in (5.1) and $c \in \mathbb{R}$ is a constant. Since $F$ is related to the area below the graph of $f$, one can easily deduce the following result:

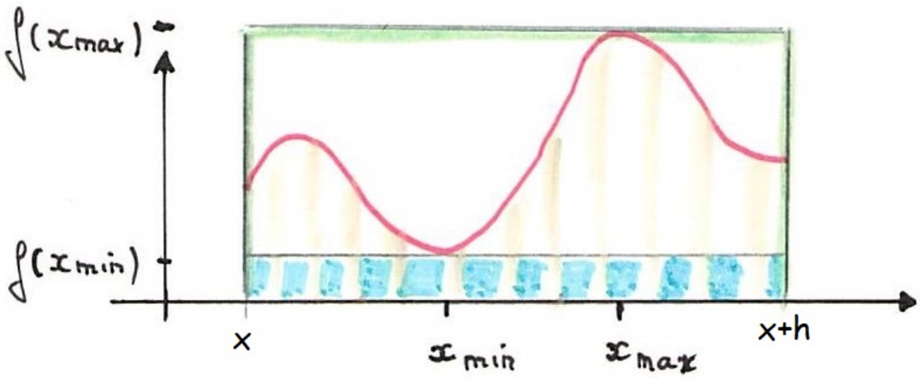


Fig. 5.2. The three areas appearing in (5.2)

**Corollary 5.4.** *Let $f : [a, b] \to \mathbb{R}_+$ be continuous, and let $G$ be an indefinite integral for $f$. Then the area below the graph of $f$ between $a$ and $b$ is given by $G(b) - G(a)$.*

*Proof.* Since $G = F + c$ for some constant $c$ and the function $F$ discussed in Theorem 5.3, one has

$$G(b) - G(a) = \big(F(b) + c\big) - \big(F(a) + c\big) = F(b) - F(a) = F(b) - 0 = F(b),$$

where $F(b)$ corresponds to the area below the graph of $f$. □

As already mentioned, the main problem with this section is that the area is not a well-defined notion. In the next section, we formalize this concept.

## 5.3 Riemann sums

In this section we shall consider how to make the construction of the previous section more precise.

**Definition 5.5** (Partition of an interval)**.** *For any interval $[a, b]$ and any $n \in \mathbb{N}$ we call a $n$-*partition $\mathcal{P}$ of $[a, b]$ *a set of $n + 1$ points $\{x_0, x_1, \dots, x_n\}$ with $x_0 = a$, $x_n = b$ and $x_{j-1} < x_j$ for any $j = 1, 2, \dots, n$.*

Let us mention that the notion of partition is more general than this. It corresponds to the division of an object into smaller parts. Here, since we consider an interval, the smaller parts consist also in smaller intervals. However, since intervals are fully determined by their endpoints, it is sufficient to provide a list of these endpoints for defining the partition. This is the content of Definition 5.5. Clearly, the simplest example is the so-called *regular partition* with $x_j = a + j\frac{b-a}{n}$. For this partition, each subinterval is of the same length, namely $\frac{b-a}{n}$.

In the next definition and in addition to the partition, we shall have to fix one point in each subinterval. Later on, we shall see that the precise choice of these points is not important. We also recall that a function $f : [a, b] \to \mathbb{R}$ is *bounded* if there exists $M < \infty$ such that $|f(x)| < M$ for all $x \in [a, b]$.

**Definition 5.6** (Riemann sum)**.** *Let $f : [a, b] \to \mathbb{R}$ be bounded, let $n \in \mathbb{N}$, and let $\mathcal{P}$ be a $n$-partition of $[a, b]$. For any family $\mathcal{C} := \{c_1, c_2, \dots, c_n\}$ with $x_{j-1} \le c_j \le x_j$ one sets*

$$\mathcal{R}_{\mathcal{C}}(f, \mathcal{P}) := \sum_{j=1}^{n} f(c_j)(x_j - x_{j-1})$$

*and call it* the Riemann sum for $f$*, depending on the partition $\mathcal{P}$ and on the family $\mathcal{C}$.*

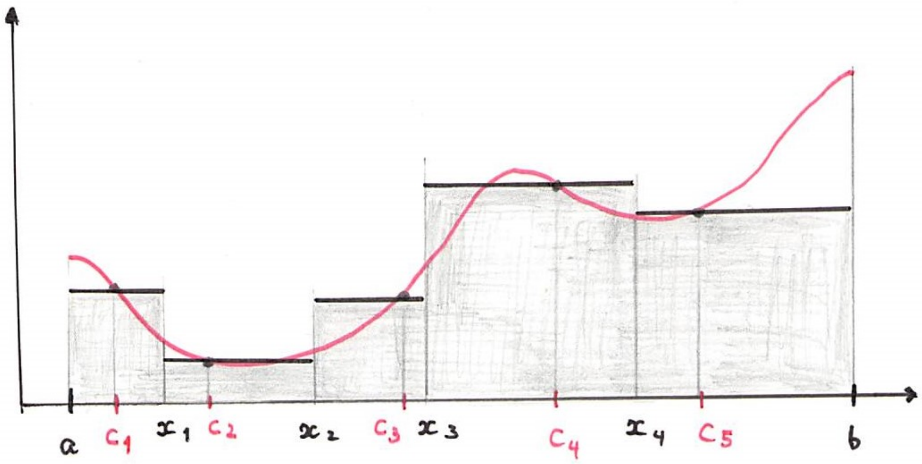


Fig. 5.3. One Riemann sum for a partition in 5 subintervals

One example of a Riemann sum is provided in Figure 5.3. Compared to the previous section, note that we do not assume that $f$ is a positive function anymore. However, if the function is positive, it appears from this picture that the Riemann sum gives an approximation of the area below the graph of $f$ between $a$ and $b$. The main idea now will be to consider finer and finer partitions of $[a, b]$, which means partitions with more subintervals and of smaller lengths.

In addition to choosing finer partitions, the arbitrary choice of the points $c_j$ will also stop playing a role. For any $j \in \{1, \dots, n\}$ let us fix the value $\sup_{x\in[x_{j-1},x_j]} f(x)$ and define the *upper Riemann sum* (in fact Darboux sum)

$$\mathcal{R}_{\max}(f, \mathcal{P}) := \sum_{j=1}^{n} \sup_{x\in[x_{j-1},x_j]} f(x)(x_j - x_{j-1}).$$

Similarly, one can choose $\inf_{x\in[x_{j-1},x_j]} f(x)$ and define the *lower Riemann sum* by

$$\mathcal{R}_{\min}(f, \mathcal{P}) := \sum_{j=1}^{n} \inf_{x\in[x_{j-1},x_j]} f(x)(x_j - x_{j-1}).$$

Note that these sums are well defined thanks to the boundedness condition on $f$. If $f$ is not bounded, one can not be sure that these expressions are well defined. Also, observe that if $f$ is continuous, these infimum and supremum are realized by some points $c_j$, but if $f$ is not continous, these $c_j$ might not exist. Then, one also observes that the following inequalities hold:

$$-M(b-a) \leq \mathcal{R}_{min}(f,\mathcal{P}) \leq \mathcal{R}_{\mathcal{C}}(f,\mathcal{P}) \leq \mathcal{R}_{max}(f,\mathcal{P}) \leq M(b-a),$$

for any choice of $\mathcal{C}$ as specified above. Note that in these inequalities, the partition $\mathcal{P}$ has been kept constant. What about considering different partitions ?

In order to compare partitions, let us give a precise definition to a finer partition. Consider $\mathcal{P}$ a $n$-partition of $[a,b]$ and let $\mathcal{P}'$ be a $n'$-partition of $[a,b]$ with $n' > n$. We say that $\mathcal{P}' = \{x_0', x_1', \dots, x_{n'}'\}$ with $x_0' = a$ and $x_{n'}' = b$ is *finer* than $\mathcal{P} = \{x_0, x_1, \dots, x_n\}$ with $x_0 = a$ and $x_n = b$ if $\mathcal{P} \subset \mathcal{P}'$. In other words, the points of $\mathcal{P}$ are included inside the set of points corresponding to the partition $\mathcal{P}'$. This means that some of the intervals of the partition given by $\mathcal{P}$ are divided into subintervals in the partition defined by $\mathcal{P}'$, see Figure 5.4.

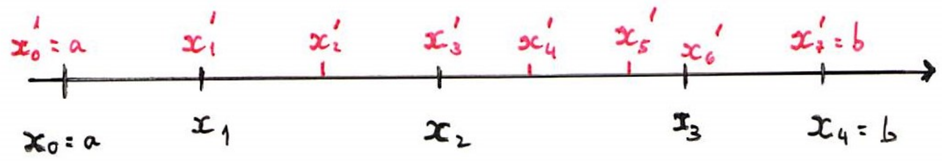


Fig. 5.4. One partition and a finer one

When considering a finer partition $\mathcal{P}'$, one infers that

$$\mathcal{R}_{min}(f,\mathcal{P}) \leq \mathcal{R}_{min}(f,\mathcal{P}') \leq \mathcal{R}_{max}(f,\mathcal{P}') \leq \mathcal{R}_{max}(f,\mathcal{P})$$

as it clearly appears in Figure 5.5.

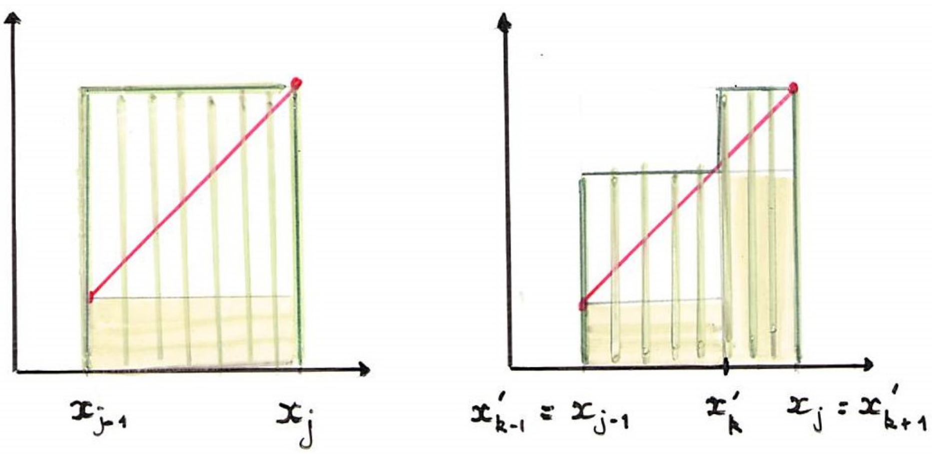


Fig. 5.5. The lower contribution in yellow and the upper contribution in green

It means that whenever one considers a finer partition, the lower Riemann sum increases, while the upper Riemann sum decreases. Thus, by considering a family of finer and finer partitions one ends up with an increasing family of lower Riemann sums, and a decreasing family of upper Riemann sums. Since the increasing family is upper bounded (by $M(b-a)$ for example) while the decreasing family is lower bounded (by $-M(b-a)$ for example), these two families are converging. Note that strictly speaking we are using here a slightly stronger argument than the Monotone convergence theorem presented in Theorem. Thus, the remaining question is about the equality of the two limits, namely

$$\lim_{\text{finer } \mathcal{P}} \mathcal{R}_{min}(f, \mathcal{P}) \stackrel{?}{=} \lim_{\text{finer } \mathcal{P}} \mathcal{R}_{max}(f, \mathcal{P}). \tag{5.3}$$

But each good question deserves a definition, isn't it ?

**Definition 5.7** (Riemann integrable function, Riemann integral)**.** *A bounded function* $f : [a, b] \to \mathbb{R}$ *is called* Riemann integrable *if*

$$\sup_{\text{partition } \mathcal{P}} \mathcal{R}_{min}(f, \mathcal{P}) = \inf_{\text{partition } \mathcal{P}} \mathcal{R}_{max}(f, \mathcal{P}). \tag{5.4}$$

*If $f$ is Riemann integrable, we write $\int_a^b f(x)\mathrm{d}x$ for the value given by* (5.4)*, and call it the* Riemann integral *of $f$ on $[a, b]$. The set $[a, b]$ is also called* the domain of integration.

Let us just emphasize that the notation in (5.4) is slightly more general than the one presented in (5.3). Indeed, in (5.3) the idea was to start with one partition, and then to consider finer ones. In (5.4) one does not prescribe any initial partition, but one considers all of them through sequences (or more precisely nets) of finer and finer partitions.

A natural attitude now is to wonder if any function is Riemann integrable ? Let us first check that not all functions are Riemann integrable.

**Example 5.8.** *Consider the function $f : [0, 1] \to \mathbb{R}$ given for $x \in [0, 1]$ by*

$$f(x) := \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}.$$

*Then, one easily observes that for any partition $\mathcal{P}$ of $[0, 1]$ one has $\mathcal{R}_{min}(f, \mathcal{P}) = 0$ while $\mathcal{R}_{max}(f, \mathcal{P}) = 1$. In such a situation, it is clear that* (5.4) *will not be satisfied, which means that this function $f$ is not Riemann integrable.*

Fortunately, there exist quite many functions which are Riemann integrable. The following statement is rather standard and the proof can be found in any serious textbook about calculus or Riemann integrals. Before stating it, let us mention a rather easy fact about continuous functions on $[a, b]$: such functions are bounded. This can be inferred from the Extreme value theorem 3.3.

**Theorem 5.9.** *Any continuous function $f : [a, b] \to \mathbb{R}$ is Riemann integrable.*

This theorem asserts that the expression $\int_a^b f(x)\mathrm{d}x$ is well defined for any continuous function on $[a, b]$ and corresponds to the quantity defined by (5.4). Let us emphasize that (5.4) is only a real number, it is not a function, and that this number, or the expression $\int_a^b f(x)\mathrm{d}x$, is called *a definite integral for $f$ on the interval* $[a, b]$. If the function $f$ satisfies $f(x) \geq 0$ for all $x \in [a, b]$, then the expression $\int_a^b f(x)\mathrm{d}x$ corresponds to the construction of Section 5.2. On the other hand, if the sign of $f$ is not prescribed, the expression $\int_a^b f(x)\mathrm{d}x$ computes the *oriented area*, as shown in Figure 5.6

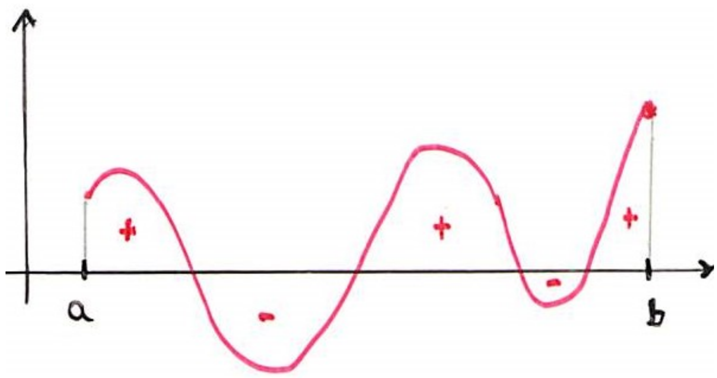


Fig. 5.6. Oriented area

**Remark 5.10.** *The theory of Riemann integrals is only sketched here, a complete and very well written approach can be found in the lecture notes of Prof. Hunter (UC Davies). In particular, these notes contain the proof of the properties of the Riemann integral stated in the next section. Let us also mention that other types of integrals exist, like the Lebesgue integral, which allow us to integrate more general functions.*

## 5.4 Properties of the Riemann integral

Let us state a few properties of the Riemann integral. Intuitively, these properties are rather natural, but a precise proof involves the Riemann sums introduced in the previous section, and some limiting procedures.

**Lemma 5.11.** *Let $f, g$ be two bounded and Riemann integrable functions on the interval $[a, b]$, and let $\lambda \in \mathbb{R}$. Then the following properties hold:*

*(i)* *If $c \in (a, b)$, then* $\int_a^b f(x)\mathrm{d}x = \int_a^c f(x)\mathrm{d}x + \int_c^b f(x)\mathrm{d}x$,

*(ii)* $\int_a^b [\lambda f](x)\mathrm{d}x = \lambda \int_a^b f(x)\mathrm{d}x$,

*(iii)* $\int_a^b [f + g](x)\mathrm{d}x = \int_a^b f(x)\mathrm{d}x + \int_a^b g(x)\mathrm{d}x$,

*(iv)* *If $g(x) \geq f(x)$ for all $x \in [a, b]$, then* $\int_a^b g(x)\mathrm{d}x \geq \int_a^b f(x)\mathrm{d}x$,

*(v) The product $fg$ is a Riemann integrable function.*

Let us just mention that for the proof of $(iv)$ one can consider the function $g-f$ which satisfies $[g-f](x) \geq 0$ for all $x \in [a,b]$. Then, one has $\int_a^b [g-f](x)\mathrm{d}x \geq 0$, since this integral corresponds to the surface below the graph of $g-f$, as introduced in the previous section. One finally deduces from $(iii)$ that

$$\int_a^b [g-f](x)\mathrm{d}x \geq 0 \Longleftrightarrow \int_a^b g(x)\mathrm{d}x - \int_a^b f(x)\mathrm{d}x \geq 0 \Longleftrightarrow \int_a^b g(x)\mathrm{d}x \geq \int_a^b f(x)\mathrm{d}x.$$

So far, we have consider Riemann integrals on intervals $[a,b]$ with $a < b$. It is then natural to set

$$\int_b^a f(x)\mathrm{d}x := -\int_a^b f(x)\mathrm{d}x$$

which allows us to write $\int_c^d f(x)\mathrm{d}x$ without imposing that $c < d$ or $d < c$. In particular, the above property $(i)$ holds for any $a, b$, and $c$, as long as the corresponding integrals exist. We also set $\int_a^a f(x)\mathrm{d}x := 0$.

Let us now stress an important consequence of the point $(v)$. Consider first the function $\chi_{[c,d]} : \mathbb{R} \to \mathbb{R}$ given by $\chi_{[c,d]}(y) = 1$ if $y \in [c,d]$ and $\chi_{[c,d]}(y) = 0$ if $y \notin [c,d]$. Such a function is called the *characteristic function on the interval* $[c,d]$. Clearly, this function is not continuous, but nevertheless it is Riemann integrable, as it can be checked rather easily. Thus, instead of the interval $[c,d]$, one can consider the interval $[a,b]$, a point $x \in [a,b]$, and deduce that the characteristic function $\chi_{[a,x]} : [a,b] \to \mathbb{R}$ is Riemann integrable. Then, as a consequence of $(v)$ of Lemma 5.11, for any Riemann integrable function $f$ on $[a,b]$, the new function $f\chi_{[a,x]}$ is Riemann integrable, and one has

$$\int_a^b \big(f\chi_{[a,x]}\big)(y)\mathrm{d}y = \int_a^x f(y)\mathrm{d}y.$$

In summary, if the function $f$ is Riemann integrable on $[a,b]$ and if $x \in [a,b]$, the expression $\int_a^x f(y)\mathrm{d}y$ is well-defined. Since $x$ is an arbitrary value in $[a,b]$, the correspondence $x \mapsto \int_a^x f(y)\mathrm{d}y$ defines a function on the interval $[a,b]$. In fact, if $f$ is positive valued, then this expression corresponds to the function $F$ introduced in Section 5.2, or more precisely

$$F(x) := \int_a^x f(y)\mathrm{d}y.$$

This definition replaces the approximate definition mentioned in (5.1).

Let us now provide a more precise version of Theorem 5.3. In fact, the statement is more general since the functions are not assumed to be positive valued. In the statement, we shall use the content of Theorem 5.9, namely that any continuous function is Riemann integrable. Note that this theorem plays a very important role for calculus.

**Theorem 5.12** (Fundamental theorem of calculus)**.** *Let $f : [a,b] \to \mathbb{R}$ be a continuous function, let $x \in [a,b]$ and set $F(x) := \int_a^x f(y)\mathrm{d}y$. Then $F : [a,b] \to \mathbb{R}$ is continuous on $[a,b]$ and differentiable on $(a,b)$ with $F'(x) = f(x)$ for any $x \in (a,b)$.*

*Proof.* Consider $x \in (a,b)$ and let $h > 0$ be small enough such that $x+h \in (a,b)$ (a similar argument holds for $h < 0$ small enough). Then one has by the properties of the Riemann integral presented in Lemma 5.11

$$\begin{aligned}\frac{F(x+h)-F(x)}{h} &= \frac{1}{h}\Big(\int_a^{x+h} f(y)\mathrm{d}y - \int_a^x f(y)\mathrm{d}y\Big)\\ &= \frac{1}{h}\Big(\int_a^{x+h} f(y)\mathrm{d}y + \int_x^a f(y)\mathrm{d}y\Big)\\ &= \frac{1}{h}\int_x^{x+h} f(y)\mathrm{d}y.\end{aligned}$$

Since $f$ is continuous on $[x, x+h]$ let $x_{min}(h)$ be the global minimum of $f$ on this interval, and let $x_{max}(h)$ be the global minimum of $f$ on this interval, which exist by the Extreme value theorem 3.3. One then infers that

$$f\big(x_{min}(h)\big) \leq f(y) \leq f\big(x_{max}(h)\big) \tag{5.5}$$

for any $y \in [a,b]$. Since $\int_x^{x+h} f\big(x_{min}(h)\big)\mathrm{d}x = hf\big(x_{min}(h)\big)$, and similarly $\int_x^{x+h} f\big(x_{max}(h)\big)\mathrm{d}x = hf\big(x_{max}(h)\big)$, it follows from $(iv)$ of Lemma 5.11 that

$$hf\big(x_{min}(h)\big) \leq \int_x^{x+h} f(y)\mathrm{d}y \leq hf\big(x_{max}(h)\big)$$

or equivalently

$$f\big(x_{min}(h)\big) \leq \frac{1}{h}\int_x^{x+h} f(y)\mathrm{d}y \leq f\big(x_{max}(h)\big).$$

By the Squeeze theorem and since $\lim_{h\to 0} f\big(x_{min}(h)\big) = f(x) = \lim_{h\to 0} f\big(x_{max}(h)\big)$, one gets the statement. □

Note that if the function $f$ is not continuous, then the last step does not hold. For this reason, the continuity assumption in the statement of the theorem is essential. We also infer from the statement that the function $F$ is an indefinite integral for $f$, as defined in Definition 5.1. Thus, any continuous function on an interval $[a,b]$ has an indefinite integral, given by the function $F : x \mapsto \int_a^x f(y)\mathrm{d}y$. In such a case, it follows from the discussion in Section 5.1 that any other indefinite integral $G$ for $f$ is equal $F + c$, for some constant $c$, and therefore

$$G(b) - G(a) = F(b) - F(a) = \int_a^b f(y)\mathrm{d}y.$$

These equalities extend the result of Corollary 5.4 to functions which are not positive valued. Of course, the interpretation in terms of area is lost for general functions. Note that we shall also use the notations

$$\int_a^b f(y)\mathrm{d}y = F(y)\big|_a^b \equiv F(b) - F(a).$$

Let us now prove a few additional results about Riemann integrals for continuous functions. For the next statement, one has to observe that if $f$ is a continuous function, then $|f|$ is also a continuous function.

**Proposition 5.13.** *Let $f : [a,b] \to \mathbb{R}$ be continuous. Then one has*

$$-\int_a^b |f(y)|\mathrm{d}y \leq \int_a^b f(y)\mathrm{d}y \leq \int_a^b |f(y)|\mathrm{d}y.$$

*Proof.* Observe that $f(y) \leq |f(y)|$ for any $y \in [a,b]$. It then follows from point $(iv)$ of Lemma 5.11 that

$$\int_a^b f(y)\mathrm{d}y \leq \int_a^b |f(y)|\mathrm{d}y.$$

Since the inequality $-|f(y)| \leq f(y)$ also holds for any $y \in [a,b]$, one infers from the same lemma that

$$-\int_a^b |f(y)|\mathrm{d}y = \int_a^b \big(-|f(y)|\big)\mathrm{d}y \leq \int_a^b f(y)\mathrm{d}y,$$

which leads to the statement. □

As direct corollary, one also gets an upper bound on these integrals, namely:

**Corollary 5.14.** *Let $f : [a,b] \to \mathbb{R}$ be continuous, and assume that there exists $M > 0$ such that $|f(x)| \leq M$ for all $x \in [a,b]$. Then one has*

$$\left|\int_a^b f(x)\mathrm{d}x\right| \leq M(b-a). \tag{5.6}$$

## 5.5 Improper Riemann integrals

So far, Riemann integrals have been defined for bounded functions on intervals of the form $[a,b]$. What about unbounded functions, or about integrals on $[a,\infty)$, or $(-\infty,b]$ or on $\mathbb{R}$ ? Such integrals can be defined by an additional limiting process, and are usually called *improper Riemann integrals*.

As a preliminary example, consider the function $(0,\infty) \ni x \mapsto \frac{1}{x^{3/2}} \in \mathbb{R}$. If we consider $[a,b] \subset (0,\infty)$ then the function $f : [a,b] \ni x \mapsto \frac{1}{x^{3/2}}$ is continuous and one has

$$\int_a^b f(y)\mathrm{d}y = \int_a^b \frac{1}{y^{3/2}}\mathrm{d}y = -2\frac{1}{y^{1/2}}\Big|_a^b = -2\left(\frac{1}{b^{1/2}} - \frac{1}{a^{1/2}}\right).$$

Clearly, this expression is well defined even when $b \to \infty$, and therefore one can set

$$\int_a^\infty f(y)\mathrm{d}y := \lim_{b\to\infty}\int_a^b f(y)\mathrm{d}y = \lim_{b\to\infty} -2\left(\frac{1}{b^{1/2}} - \frac{1}{a^{1/2}}\right) = \frac{2}{a^{1/2}}.$$

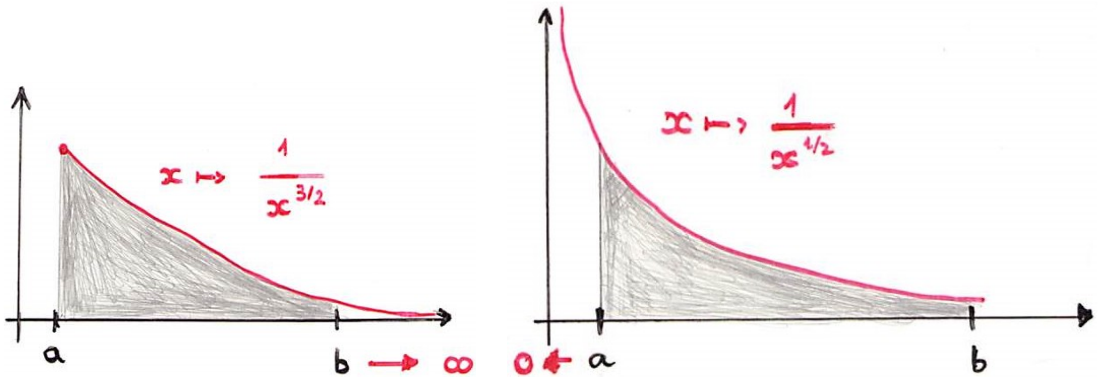


Fig. 5.7. Representations of the two improper integrals

In this example, the domain of integration is extended to an infinite interval, which was not covered by our initial definition of a Riemann integral.

Consider now the function $(0, \infty) \ni x \mapsto \frac{1}{x^{1/2}} \in \mathbb{R}$. If we consider $[a, b] \subset (0, \infty)$ then the function $f : [a, b] \ni x \mapsto \frac{1}{x^{1/2}}$ is continuous, and the integral $\int_a^b f(y)\mathrm{d}y$ is well-defined and given by

$$\int_a^b f(y)\mathrm{d}y = \int_a^b \frac{1}{y^{1/2}}\mathrm{d}y = 2y^{1/2}\big|_a^b = 2\left(b^{1/2} - a^{1/2}\right).$$

Clearly, this expression is well-defined even when $a \searrow 0$, and therefore one can set

$$\int_0^b f(y)\mathrm{d}y := \lim_{a \searrow 0} \int_a^b f(y)\mathrm{d}y = \lim_{a \searrow 0} 2\left(b^{1/2} - a^{1/2}\right) = 2b^{1/2}.$$

In this example, the domain of integration is still the interval $[0, b]$, but the function $f$ is not bounded on this interval, and even not defined at 0. Such a situation is again not covered by our initial definition of a Riemann integral. An illustration of these two situations is presented in Figure 5.7.

Since these two integrals are well defined and rather natural, it is necessary to extend the definition of the Riemann integral. In the following statement, the special cases $a = -\infty$ and $b = \infty$ are accepted.

**Definition 5.15** (Improper Riemann integral). *Consider $f : (a, b) \to \mathbb{R}$, and suppose that for any $a < a' < b' < b$ the restricted function $f : [a', b'] \to \mathbb{R}$ is bounded and Riemann integrable. Then, if $\lim_{a' \searrow a} \int_{a'}^{b'} f(y)\mathrm{d}y$ exists, one writes $\int_a^{b'} f(y)\mathrm{d}y$ for this limit. Similarly, if $\lim_{b' \nearrow b} \int_{a'}^{b'} f(y)\mathrm{d}y$ exists, one writes $\int_{a'}^{b} f(y)\mathrm{d}y$ for this limit. If both limits exist separately, then one sets:*

$$\int_a^b f(y)\mathrm{d}y := \lim_{a' \searrow a}\left(\lim_{b' \nearrow b} \int_{a'}^{b'} f(y)\mathrm{d}y\right) = \lim_{b' \nearrow b}\left(\lim_{a' \searrow a} \int_{a'}^{b'} f(y)\mathrm{d}y\right).$$

*Such integrals are called* improper Riemann integrals.

The interest of the above definition is that the domain of integration can be unbounded, like a half-line or the entire line, and that the function can be unbounded on its domain of definition. However, when dealing with improper Riemann integral, one has to be a little bit cautious. Indeed, according to this definition, an integral of the form $\int_{\mathbb{R}} f(y)\mathrm{d}y$ has to be understood as

$$\int_{\mathbb{R}} f(y)\mathrm{d}y := \lim_{b\nearrow\infty}\Big(\lim_{a\searrow-\infty}\int_a^b f(y)\mathrm{d}y\Big). \tag{5.7}$$

A common mistake is to say that $f$ admits an improper Riemann integral if the following limit exists: $\lim_{M\nearrow\infty}\int_{-M}^M f(y)\mathrm{d}y$. Indeed, such a definition would lead to some rather unpleasant surprises ↯. For example, one has

$$\lim_{M\nearrow\infty}\int_{-M}^M \sin(y)\mathrm{d}y = 0$$

since $\int_{-M}^M \sin(y)\mathrm{d}y = 0$ for any $M > 0$, but by considering a slightly asymmetric domain of integration, one has for $\epsilon \in (0,\pi)$

$$\begin{aligned}\int_{-M-\epsilon}^{M-\epsilon}\sin(y)\mathrm{d}y &= \cos(-M-\epsilon) - \cos(M-\epsilon)\\ &= \cos(M+\epsilon) - \cos(M-\epsilon)\\ &= -2\sin(\epsilon)\sin(M),\end{aligned}$$

which has no limit as $M \nearrow \infty$. Clearly, by considering separately the two limits, as in (5.7), one would easily deduce than none of these limits exists: the function $x \mapsto \sin(x)$ is not Riemann integrable, even in the improper sense, on $\mathbb{R}$.

## 5.6 Techniques of integration

In this section we review some classical tricks for finding indefinite integrals. There exist several other tricks or methods, but this subject becomes rapidly rather technical, and boring !

### 5.6.1 Substitution

Let $g$ and $u$ be two functions such that $g \circ u$ is well-defined, and such that $u$ is differentiable. For any indefinite integral $G$ for $g$ the following equality holds:

$$\int g\big(u(x)\big)u'(x)\mathrm{d}x = G\big(u(x)\big). \tag{5.8}$$

This equality is easily checked, since by the composition rule one has $G\big(u(x)\big)' = g\big(u(x)\big)u'(x)$. Let us emphasize that *strictly speaking* the equality (5.8) is not correct:

indeed the l.h.s. corresponds to a function, while the r.h.s. represents the evaluation of a function at a point $x$. For this reason, we should write $G\big(u(\cdot)\big)$ instead of $G\big(u(x)\big)$. However, by adopting such a strict rule, the gain in rigor would be smaller than the loss in understanding, and therefore we shall keep this common notation. The following two examples show that the notation used in (5.8) is quite natural:

**Examples 5.16.**

*(i)* $\int \big(x^3+x\big)^9(3x^2+1)\mathrm{d}x = \frac{1}{10}\big(x^3+x\big)^{10}$,

*(ii)* $\int x\sin\big(x^2\big)\mathrm{d}x = -\frac{1}{2}\cos\big(x^2\big)$.

Let us still list some equalities which are useful for justifying a mnemonic trick:

$$\begin{aligned}\int_a^b g\big(u(x)\big)\frac{\mathrm{d}u}{\mathrm{d}x}(x)\mathrm{d}x &\equiv \int_a^b g\big(u(x)\big)u'(x)\mathrm{d}x \\ &= G\big(u(x)\big)\Big|_{x=a}^{x=b} \\ &= G\big(u(b)\big) - G\big(u(a)\big) \\ &= G(u)\Big|_{u=u(a)}^{u=u(b)} \\ &= \int_{u(a)}^{u(b)} g(u)\mathrm{d}u.\end{aligned} \tag{5.9}$$

Thus, if one compares the first and the last expression, one ends up with the following equality

$$\frac{\mathrm{d}u}{\mathrm{d}x}(x)\mathrm{d}x = \mathrm{d}u \quad \text{or equivalently} \quad u'(x)\mathrm{d}x = \mathrm{d}u, \tag{5.10}$$

but clearly, these relations have to be understood with a grain of salt: the only precise meaning of (5.10) is the one provided by the equalities (5.9). Even though the equalities provided in (5.10) are not really precise, they are useful, as illustrated in the next example.

**Example 5.17.** *Let us consider* $\int x^5\sqrt{1-x^2}\mathrm{d}x$*, and set* $u := 1-x^2$*, which is equivalent to* $x^2 = 1-u$*. Then, by the mentioned trick one has* $\mathrm{d}u = -2x\mathrm{d}x$ *and therefore:*

$$\begin{aligned}\int x^5\sqrt{1-x^2}\mathrm{d}x &= -\frac{1}{2}\int x^4\sqrt{1-x^2}(-2x)\mathrm{d}x \\ &= -\frac{1}{2}\int (1-u)^2\sqrt{u}\mathrm{d}u \\ &= -\frac{1}{2}\int \big(u^{1/2}-2u^{3/2}+u^{5/2}\big)\mathrm{d}u \\ &= -\frac{1}{2}\Big(\frac{u^{3/2}}{3/2}-\frac{2u^{5/2}}{5/2}+\frac{u^{7/2}}{7/2}\Big) \\ &= -\frac{1}{3}(1-x^2)^{3/2}+\frac{2}{5}(1-x^2)^{5/2}-\frac{1}{7}(1-x^2)^{7/2}.\end{aligned}$$

### 5.6.2 Integration by parts

For this trick, let us first recall the product rule, namely $(fg)' = f'g + fg'$, which is equivalent to the equality $fg' = (fg)' - f'g$. By taking an indefinite integral for this equality and by recalling that $\int [fg]'(x)\mathrm{d}x = fg$ (which is a restatement of the Fundamental theorem of calculus 5.12) one infers that

$$\int fg' = fg - \int f'g \tag{5.11}$$

or equivalently

$$\int f(x)g'(x)\mathrm{d}x = f(x)g(x) - \int f'(x)g(x)\mathrm{d}x. \tag{5.12}$$

Note again that the r.h.s. contains a slight abuse of notation with the term $f(x)g(x)$, and therefore the equality (5.11) is preferable in this case. On the other hand, if we evaluate these equalities at $b$ and subtract their values at $a$ one deduces the correct equality

$$\int_a^b f(x)g'(x)\mathrm{d}x = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)\mathrm{d}x. \tag{5.13}$$

Typical applications of these equalities are:

**Examples 5.18.**

*(i)* $\int x\mathrm{e}^x\mathrm{d}x = x\mathrm{e}^x - \int 1\mathrm{e}^x\mathrm{d}x = x\mathrm{e}^x - \mathrm{e}^x = \mathrm{e}^x(x-1)$, *where we have used* (5.11) *with* $f(x) = x$ *and* $g'(x) = \mathrm{e}^x$,

*(ii)* $\int \ln(x)\mathrm{d}x = x\ln(x) - \int x\frac{1}{x}\mathrm{d}x = x\ln(x) - x$, *where we have used* (5.11) *with* $f(x) = \ln(x)$ *and* $g'(x) = 1$.

### 5.6.3 Trigonometric integrals

For integrals involving trigonometric functions, a few equalities are really useful, as for example

(i) $\sin^2(x) + \cos^2(x) = 1$,

(ii) $\sin^2(x) = \frac{1-\cos(2x)}{2}$ and $\cos^2(x) = \frac{1+\cos(2x)}{2}$.

As an example of their use, one has:

**Examples 5.19.**

*(i)* $\int \sin^2(x)\mathrm{d}x = \frac{1}{2}\int \big(1-\cos(2x)\big)\mathrm{d}x = \frac{1}{2}\big(x - \frac{1}{2}\sin(2x)\big) = \frac{1}{2}x - \frac{1}{4}\sin(2x)$,

*(ii)* $\int \cos^3(x)\mathrm{d}x = \int \big(1-\sin^2(x)\big)\cos(x)\mathrm{d}x = \sin(x) - \frac{1}{3}\sin^3(x)$.

### 5.6.4 Partial fractions

Let us start with a few standard examples:

**Examples 5.20.**

*(i)* $\int \frac{1}{x-\alpha} \mathrm{d}x = \ln(x-\alpha)$ *for any* $\alpha \in \mathbb{R}$ *and* $x > \alpha$,

*(ii)* $\int \frac{1}{(x-\alpha)^n} \mathrm{d}x = \frac{1}{-n+1}(x-\alpha)^{-n+1}$ *for any* $n \neq 1$, *any* $\alpha \in \mathbb{R}$ *and* $x > \alpha$,

*(iii)* $\int \frac{1}{x^2+1} \mathrm{d}x = \arctan(x)$,

*(iv)* $\int \frac{1}{(x^2+1)^2} \mathrm{d}x = \frac{1}{2}\frac{x}{x^2+1} + \frac{1}{2}\arctan(x)$.

Note that these equalities can be checked by taking the derivative of the r.h.s., and by comparing the result with the l.h.s.. The last equality is however quite tricky, let us show how to obtain it. One has, by using an integration by parts in the second equality with $f(x) = \frac{1}{x^2+1}$ and $g'(x) = 1$,

$$\begin{aligned}
\arctan(x) &= \int \frac{1}{x^2+1} \mathrm{d}x \\
&= \frac{x}{x^2+1} + 2\int \frac{x^2+1-1}{(x^2+1)^2} \mathrm{d}x \\
&= \frac{x}{x^2+1} + 2\int \frac{1}{x^2+1} \mathrm{d}x - 2\int \frac{1}{(x^2+1)^2} \mathrm{d}x \\
&= \frac{x}{x^2+1} + 2\arctan(x) - 2\int \frac{1}{(x^2+1)^2} \mathrm{d}x,
\end{aligned}$$

from which the equality $(iv)$ can be easily deduced. Note that a similar approach can be used for computing iteratively all integrals of the form $\int \frac{1}{(x^2+1)^n} \mathrm{d}x$, see for example here.

Consider now an expression of the form $\frac{f}{g}$, with $f$ and $g$ some polynomials, and with the degree of the polynomial $f$ strictly smaller than the degree of the polynomial $g$. In order to compute an indefinite integral for such a ratio, the leading idea is to rewrite this in terms of simpler expressions, like:

$$\begin{aligned}
\frac{f(x)}{g(x)} = &\frac{c_1}{(x-\alpha_1)^{p_1}} + \frac{c_2}{(x-\alpha_2)^{p_2}} + \ldots + \frac{c_n}{(x-\alpha_n)^{p_n}} \\
&+ \frac{d_1 x + e_1}{\big((x-\beta_1)^2 + \gamma_1^2\big)^{q_1}} + \ldots + \frac{d_m x + e_m}{\big((x-\beta_m)^2 + \gamma_m^2\big)^{q_m}},
\end{aligned}$$

where $c_j, d_j, e_j, \alpha_j, \beta_j$, and $\gamma_j$ are real numbers, and $p_\ell, q_\ell$ belong to $\mathbb{N}$. Then, each term can be treated separately.

Let us look at one concrete example. For $\frac{f(x)}{g(x)} = \frac{2x+5}{(x^2+1)^2(x-3)}$, one rewrites it as

$$\begin{aligned}\frac{2x+5}{(x^2+1)^2(x-3)} &= \frac{c_1}{x-3} + \frac{d_1x+e_1}{x^2+1} + \frac{d_2x+e_2}{(x^2+1)^2} \\ &= \frac{c_1}{x-3} + \frac{d_1x}{x^2+1} + \frac{e_1}{x^2+1} + \frac{d_2x}{(x^2+1)^2} + \frac{e_2}{(x^2+1)^2}.\end{aligned}$$

In order to determine the coefficients, it is necessary to put all expressions on the same denominator, and to set a system of equations. Namely one gets:

$$\begin{aligned}\frac{2x+5}{(x^2+1)^2(x-3)} &= \frac{c_1(x^2+1)^2 + (d_1x+e_1)(x^2+1)(x-3) + (d_2x+e_2)(x-3)}{(x^2+1)^2(x-3)} \\ &= \frac{(c_1+d_1)x^4 + (-3d_1+e_1)x^3 + (2c_1+d_1-3e_1+d_2)x^2}{(x^2+1)^2(x-3)} \\ &\quad + \frac{(-3d_1+e_1+e_2-3d_2)x + (c_1-3e_1-3e_2)}{(x^2+1)^2(x-3)},\end{aligned}$$

and then

$$\begin{cases} c_1 + d_1 = 0 \\ -3d_1 + e_1 = 0 \\ 2c_1 + d_1 - 3e_1 + d_2 = 0 \\ -3d_1 + e_1 + e_2 - 3d_2 = 2 \\ c_1 - 3e_1 - 3e_2 = 5. \end{cases}$$

The solutions can then be obtained by some linear manipulations, and one gets

$$c_1 = \frac{11}{100}, \quad d_1 = -\frac{11}{100}, \quad d_2 = -\frac{110}{100}, \quad e_1 = -\frac{33}{100}, \quad e_2 = -\frac{130}{100}.$$

By collecting all information obtained so far, one finally gets

$$\begin{aligned}\int \frac{2x+5}{(x^2+1)^2(x-3)}\mathrm{d}x &= \int \Big(\frac{c_1}{x-3} + \frac{d_1x}{x^2+1} + \frac{e_1}{x^2+1} + \frac{d_2x}{(x^2+1)^2} + \frac{e_2}{(x^2+1)^2}\Big)\mathrm{d}x \\ &= \frac{11}{100}\ln(x-3) - \frac{11}{200}\ln(x^2+1) - \frac{33}{100}\arctan(x) \\ &\quad + \frac{110}{200}\frac{1}{x^2+1} - \frac{130}{100}\Big(\frac{1}{2}\frac{x}{x^2+1} + \frac{1}{2}\arctan(x)\Big) \\ &= \frac{11}{100}\ln(x-3) - \frac{11}{200}\ln(x^2+1) + \frac{11}{20}\frac{1}{x^2+1} \\ &\quad - \frac{13}{20}\frac{x}{x^2+1} - \frac{49}{50}\arctan(x).\end{aligned}$$

# Chapter 6

# Taylor's formula

The aim of this chapter is to construct local approximations of a function by polynomials. The difference between the original function and the approximation has to studied and kept under control. Let us mention that in this section, we will sometimes write $[a, b]$ without knowing if $a < b$ or if $a > b$. In the former case, the notation $[a, b]$ has the usual meaning, while in the latter case it should be understood as $[b, a]$.

## 6.1 Taylor's expansion

Before starting the construction, let us recall that if $f : (a, b) \to \mathbb{R}$ is sufficiently many times differentiable, then one writes $f' \equiv f^{(1)}$ for its first derivative, $f'' \equiv f^{(2)}$ for its second derivative, ..., $f^{(n)}$ for its $n^{\text{th}}$-derivative, which means for the derivative of $f^{(n-1)}$.

**Definition 6.1** (Taylor polynomial). *Let $n \in \mathbb{N}$, and let $f : (a, b) \to \mathbb{R}$ be a $n$-times differentiable function. Let also $x_0 \in (a, b)$. The polynomial $p_n(\cdot, x_0)$ of degree $n$ defined for any $x \in [a, b]$ by*

$$\begin{aligned}p_n(x, x_0) :=& \sum_{j=0}^{n} \frac{1}{j!} f^{(j)}(x_0)(x - x_0)^j \\ =& f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f^{(2)}(x_0)(x - x_0)^2 + \ldots + \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n\end{aligned}$$

*is called* the Taylor polynomial of $f$ at $x$ with respect to $x_0$ and of degree $n$.

What about the evaluation of this polynomial at $x_0$ ? Clearly, one has $p_n(x_0, x_0) =$

$f(x_0)$. More generally one observes that

$$\begin{aligned}
p_n(x_0,x_0) &= f(x_0),\\
p_n'(x_0,x_0) &= f'(x_0),\\
p_n^{(2)}(x_0,x_0) &= f^{(2)}(x_0),\\
\vdots \quad &\vdots \quad \vdots\\
p_n^{(n)}(x_0,x_0) &= f^{(n)}(x_0).
\end{aligned}$$

In other words, the function $f$ and the polynomial $p_n(\cdot,x_0)$ have the same derivatives up to order $n$ when evaluated at $x_0$. In this sense, the polynomial $p_n(\cdot,x_0)$ corresponds to the best approximation of $f$ at $x_0$ by a polynomial of degree $n$.

**Examples 6.2.**

*(i) Consider* $f:\mathbb{R}\ni x\mapsto \sin(x)\in\mathbb{R}$, *and let us fix* $x_0=0$ *and* $n=5$. *Then one has*

$$\begin{aligned}
p_5(x,0) &= \sin(0)+\cos(0)x-\frac{1}{2}\sin(0)x^2-\frac{1}{3!}\cos(0)x^3+\frac{1}{4!}\sin(0)x^4+\frac{1}{5!}\cos(0)x^5\\
&= x-\frac{1}{6}x^3+\frac{1}{120}x^5.
\end{aligned}$$

*(ii) Consider* $\mathbb{R}\ni x\mapsto \mathrm{e}^x\in\mathbb{R}$, *and let us fix* $x_0=0$ *and* $n\in\mathbb{N}$ *arbitrary. Then one has*

$$p_n(x,0)=1+x+\frac{1}{2!}x^2+\frac{1}{3!}x^3+\ldots+\frac{1}{n!}x^n=\sum_{j=0}^{n}\frac{1}{j!}x^j.$$

So far, we have produced a polynomial approximating a function at a certain point $x_0$. But what about the difference between the initial function and the polynomial, or more precisely how can one estimate the difference $f-p_n(\cdot,x_0)$ ? This is the content of the next statement. For this statement, let us introduce one more notation: $g\in C^n\big([a,b]\big)$ which means that $g$ is $n$-times continuously differentiable on $(a,b)$, and that $g$ and all its $n$ derivatives have well-defined values at $a$ and at $b$. In particular, it means that all these derivatives are bounded on $[a,b]$.

**Theorem 6.3** (Taylor's expansion theorem)**.** *Let* $n\in\mathbb{N}$, *and let* $f\in C^{n+1}\big([a,b]\big)$. *Let also* $x_0\in(a,b)$. *For any* $x\in[a,b]$ *the following equality holds:*

$$f(x)=p_n(x,x_0)+R_{n+1}(x,x_0) \tag{6.1}$$

*with*

$$R_{n+1}(x,x_0)=\int_{x_0}^{x}\frac{(x-t)^n}{n!}f^{(n+1)}(t)\mathrm{d}t. \tag{6.2}$$

*In addition,* $R_{n+1}(x,x_0)=\frac{1}{(n+1)!}f^{(n+1)}(c)(x-x_0)^{n+1}$ *for some* $c\in(x_0,x)$.

Before the proof, let us just stress that the complicated assumption on $f$, namely $f \in C^{n+1}\big([a,b]\big)$ allows us to give a meaning to the integral in (6.2) and to the evaluation $f^{(n+1)}(c)$.

*Proof.* i) The proof is performed by induction over the index $n$. For $n = 0$ one has $p_0(x, x_0) = f(x_0)$ for any $x \in [a, b]$, and (6.1) is equal to

$$f(x) = f(x_0) + \int_{x_0}^{x} \frac{(x-t)^0}{0!} f'(t)\mathrm{d}t = f(x_0) + \int_{x_0}^{x} f'(t)\mathrm{d}t.$$

Since these equalities hold, the initial statement is correct for $n = 0$. Thus, we shall assume that the statement is true for some $n-1$ with $n \geq 1$, and show it for $n$.

Assume then that the statement is true for $n-1$, namely $f(x) = p_{n-1}(x, x_0) + R_n(x, x_0)$ with $R_n(x, x_0)$ given by $\int_{x_0}^{x} \frac{(x-t)^{n-1}}{(n-1)!} f^{(n)}(t)\mathrm{d}t$. By an integration by parts (with respect to the variable $t$) one infers that

$$\begin{aligned}
&\int_{x_0}^{x} \frac{(x-t)^{n-1}}{(n-1)!} f^{(n)}(t)\mathrm{d}t \\
&= -\frac{1}{(n-1)!}\frac{1}{n}(x-t)^n f^{(n)}(t)\Big|_{x_0}^{x} + \int_{x_0}^{x} \frac{(x-t)^n}{n!} f^{(n+1)}(t)\mathrm{d}t \\
&= \frac{1}{n!}(x-x_0)^n f^{(n)}(x_0) + R_{n+1}(x_0, x).
\end{aligned}$$

Since

$$p_{n-1}(x, x_0) + \frac{1}{n!}(x-x_0)^n f^{(n)}(x_0) = p_n(x, x_0),$$

it follows that one has obtained the equality (6.1) for $n$ instead of for $n-1$. Finally, since $n$ is arbitrary, it follows by induction that the statement is true for any $n$.

ii) It remains to prove the second expression for the remainder term. Since $f^{(n+1)}$ is continuous on the subset $[x_0, x]$, there exists a $t_{min} \in [x_0, x]$ and $t_{max} \in [x_0, x]$ such that

$$f^{(n+1)}(t_{min}) \leq f^{(n+1)}(t) \leq f^{(n+1)}(t_{max})$$

for any $t \in [x_0, x]$. Note that we assume that $x > x_0$, but a similar argument holds for $x < x_0$. Thus one has for $t \in [x_0, x]$

$$\frac{(x-t)^n}{n!} f^{(n+1)}(t_{min}) \leq \frac{(x-t)^n}{n!} f^{(n+1)}(t) \leq \frac{(x-t)^n}{n!} f^{(n+1)}(t_{max})$$

and by taking the integrals

$$\begin{aligned}
f^{(n+1)}(t_{min}) \int_{x_0}^{x} \frac{(x-t)^n}{n!}\mathrm{d}t &\leq \int_{x_0}^{x} \frac{(x-t)^n}{n!} f^{(n+1)}(t)\mathrm{d}t \\
&\leq f^{(n+1)}(t_{max}) \int_{x_0}^{x} \frac{(x-t)^n}{n!}\mathrm{d}t
\end{aligned}$$

from which one infers that

$$f^{(n+1)}(t_{min})\frac{(x-x_0)^{n+1}}{(n+1)!} \leq R_{n+1}(x,x_0) \leq f^{(n+1)}(t_{max})\frac{(x-x_0)^{n+1}}{(n+1)!}. \tag{6.3}$$

By the Intermediate value theorem 3.4 applied to $f^{(n+1)}$, this function takes all the values between $f(x_{min})$ and $f(x_{max})$. Therefore, there exists $c \in [x_0, x]$ such that

$$f^{(n+1)}(c)\frac{(x-x_0)^{n+1}}{(n+1)!} = R_{n+1}(x,x_0),$$

which corresponds to the second statement.

Let us observe that (6.3) provides an estimate on the remainder term. Indeed, one easily infers that

$$\begin{aligned}|R_{n+1}(x,x_0)| &\leq \frac{|x-x_0|^{n+1}}{(n+1)!}\sup_{t\in[x_0,x]}|f^{(n+1)}(t)| \\ &= \frac{|x-x_0|^{n+1}}{(n+1)!}\max\big\{|f^{(n+1)}(x_{min})|, |f^{(n+1)}(x_{max})|\big\}.\end{aligned}$$

This estimate is sometimes very useful, as shown in the next examples:

**Examples 6.4.**

(i) *For the sine function and for $x_0 = 0$ one has*

$$\sin(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 + \ldots + R_{n+1}(x,0)$$

*with $|R_{n+1}(x,0)| \leq \frac{|x|^{n+1}}{(n+1)!}$, since $|\pm\sin(c)| \leq 1$ and $|\pm\cos(c)| \leq 1$.*

(ii) *For the exponential function and for $x_0 = 0$ one has*

$$\mathrm{e}^x = \sum_{j=0}^{n}\frac{1}{j!}x^j + R_{n+1}(x,0)$$

*with $|R_{n+1}(x,0)| \leq \frac{|x|^{n+1}}{(n+1)!}\max\{1,\mathrm{e}^x\}$ since the exponential function is an increasing function. More precisely, if $x > 0$ one has $\sup_{t\in[0,x]}\mathrm{e}^t = \mathrm{e}^x$ while if $x < 0$ one has $\sup_{t\in[0,x]}\mathrm{e}^t \equiv \sup_{t\in[x,0]}\mathrm{e}^t = 1$.*

Let us add one more information about the ratio $\frac{|x-x_0|^n}{n!}$ when $n$ goes to infinity. Clearly, it is equivalent to study $\frac{p^n}{n!}$ for any real number $p > 0$. This estimate is useful for evaluating the remainder term in a Taylor's expansion.

**Lemma 6.5.** *For any $p > 0$ one has $\lim_{n\to\infty} \frac{p^n}{n!} = 0$.*

*Proof.* For a fixed $p > 0$ let $n_0 \in \mathbb{N}$ with $n_0 > 2p$. Clearly, this condition is equivalent to $\frac{p}{n_0} < \frac{1}{2}$. Then, for any $n > n_0$ one has

$$\begin{aligned}
\frac{p^n}{n!} &= \frac{p}{1}\frac{p}{2}\frac{p}{3}\cdots\frac{p}{n_0}\frac{p}{n_0+1}\frac{p}{n_0+2}\cdots\frac{p}{n} \\
&\leq \frac{p^{n_0}}{n_0!}\frac{1}{2}\frac{1}{2}\frac{1}{2}\cdots\frac{1}{2} \\
&= \frac{p^{n_0}}{n_0!}\Big(\frac{1}{2}\Big)^{n-n_0}
\end{aligned}$$

and this expression goes to 0 as $n \to \infty$. □

## 6.2 Taylor's expansion at 0 for even or odd functions

In this section, we consider the special case of even or odd functions, and consider their Taylor's expansion around 0. Recall that an even function $f : I \to \mathbb{R}$ satisfies $f(-x) = f(x)$ for any $x \in I$ with $-x \in I$, while for an odd function $f$ one has $f(-x) = -f(x)$ for any $x \in I$ with $-x \in I$. For simplicity, we shall consider even or odd function on $\mathbb{R}$, which means $I = \mathbb{R}$, but a local version of what is presented below can be easily obtained.

The first observation is that the derivative of an even function or of an odd function is not arbitrary. More precisely one has:

**Lemma 6.6.** *If $f : \mathbb{R} \to \mathbb{R}$ is an even differentiable function, then its derivative is odd. Similarly, if $f : \mathbb{R} \to \mathbb{R}$ is an odd differentiable function, then its derivative is even.*

*Proof.* Let $f$ be even and differentiable, let $x \in \mathbb{R}$, then one has

$$\begin{array}{l}
\dfrac{f(x+h)-f(x)}{h} \xrightarrow{\ h\to 0\ } f'(x) \\
\qquad\quad \| \\
\dfrac{f\big(-(x+h)\big)-f(-x)}{h} = -\dfrac{f(-x-h)-f(-x)}{-h} \xrightarrow{\ h\to 0\ } -f'(-x).
\end{array}$$

One thus infers that $f'(x) = -f'(-x)$, which shows that $f'$ is an odd function. Similarly, if $f$ is odd and differentiable one has

$$\begin{array}{l}
\dfrac{f(x+h)-f(x)}{h} \xrightarrow{\ h\to 0\ } f'(x) \\
\qquad\quad \| \\
-\dfrac{f\big(-(x+h)\big)-f(-x)}{h} = \dfrac{f(-x-h)-f(-x)}{-h} \xrightarrow{\ h\to 0\ } f'(-x)
\end{array}$$

which shows that $f'$ is an even function. □

Before the main result of this section, let us do one more observation about odd functions: whenever $f$ is an odd function, one has $f(0) = 0$. Indeed, since the equation $f(-x) = -f(x)$ has to be satisfied for any $x \in \mathbb{R}$, it follows that for $x = 0$ one has $f(-0) = -f(0)$, which means $f(0) = -f(0)$. The only solution of this equation is clearly $f(0) = 0$.

**Theorem 6.7.** *Let $n \in \mathbb{N}$, and let $f : \mathbb{R} \to \mathbb{R}$ be $n$ times differentiable, and consider the Taylor polynomial $x \mapsto p_n(x, 0) := \sum_{j=0}^{n} \frac{1}{j!} f^{(j)}(0) x^j$. If $f$ is even, then the coefficient $f^{(j)}(0) = 0$ for any $j$ odd, while if $f$ is odd the coefficient $f^{(j)}(0) = 0$ for any $j$ even.*

*Proof.* If $f$ is even, then the functions $f'$, $f^{(3)}$, $f^{(5)}, \dots$ are odd, by Lemma 6.6. It thus follows from the observation made before that $f'(0)$, $f^{(3)}(0)$, $f^{(5)}(0), \dots$ are all equal to 0. On the other hand, if $f$ is odd, then $f$, $f^{(2)}$, $f^{(4)}, \dots$ are odd, also by Lemma 6.6. It then follows again from the above observation that $f(0)$, $f^{(2)}(0)$, $f^{(4)}(0), \dots$ are also equal to 0. The statement follows from these observations. ☐

# Chapter 7

# Series

A natural question triggered by Taylor's expansions is the following one: can we consider $n = \infty$, or more precisely, do we have

$$f(x) = \sum_{j=0}^{\infty} \frac{1}{j!} f^{(j)}(x_0)(x - x_0)^j \quad ? \tag{7.1}$$

If we think about the exponential function initially introduced in Section 2.7 by $\mathrm{e}^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n$, and if we fix $x_0 = 0$, it seems that the answer to the question is "yes". Unfortunately (or fortunately, it depends on the point of view), it is not always the case. In this chapter, we shall just provide a few information about this subject, and refer to a course on... complex analysis for deeper and beautiful results. But first of all, series, convergente series, and absolutely convergent series have to be introduced.

## 7.1 Convergent series

Let us consider a sequence of real numbers $(a_j)_{j\in\mathbb{N}}$, as already introduced in Definition 1.2. We would like to define the infinite sum of these numbers, namely $a_1 + a_2 + a_3 + \ldots$. Such an infinite sum is called *a series*. However, one has to be careful because this formal expression is often not well defined. For example, if $a_j = 1$ for all $j$, then one has

$$a_1 + a_2 + a_3 + \ldots = 1 + 1 + 1 + \ldots = \infty,$$

but if $a_j = (-1)^j$, then one has

$$a_1 + a_2 + a_3 + \ldots = -1 + 1 - 1 + \ldots = ?$$

One way to solve these problems (at least partially) is to work with finite sums, and to look at the limit of these finite sums. More precisely, for any $n \in \mathbb{N}$ we define the *partial sum*:

$$s_n := \sum_{j=1}^{n} a_j = a_1 + a_2 + \ldots + a_{n-1} + a_n.$$

Clearly, such a finite sum is always well defined, but what happens when $n \to \infty$ ? The following definition takes care of this situation. However, let us stress that a convergent series is very different from a convergent sequence, as introduced in Definition 1.3. However, the definition of a convergent series involves the notion of a convergent sequence...

**Definition 7.1** (Convergent series)**.** *A series of real numbers $a_1 + a_2 + a_3 + \dots$ is* convergent *if the sequence of partial sums is convergent, or more precisely if $\lim_{n\to\infty} s_n$ exists. If the series is convergent one writes $\sum_{j=1}^{\infty} a_j$ for $\lim_{n\to\infty} s_n$. If the limit of the sequence of partial sums does not exist, we say that the series is* not convergent *or is* divergent*.*

Starting from a sequence of real numbers $(a_j)_{j\in\mathbb{N}}$ we say that the *corresponding series* is convergent if the series $a_1 + a_j + a_3 + \dots$ is convergent. Equivalently, one says that the sequence $(a_j)_{j\in\mathbb{N}}$ *defines a convergent series*. In this course, we shall mainly deal with convergent series. However let us just mention that divergent sequences are in fact very interesting and can be studied with more advanced tools, see for example here. Divergent series have even applications in physics, and are linked to the notions of regularization or renormalization.

**Examples 7.2.**

*(i) The geometric series is one of the most commonly used convergent series, and reads for $|a| < 1$*

$$\sum_{j=1}^{\infty} a^j = a\sum_{j=0}^{\infty} a^j = a\frac{1}{1-a} = \frac{a}{1-a}.$$

*Indeed, the partial sums are given for any $a \neq 1$ by $s_n = \frac{a(1-a^n)}{1-a}$, and the sequence $(s_n)_{n\in\mathbb{N}}$ has a limit if and only if $\lim_{n\to\infty} a^n = 0$, and this takes place if and only if $|a| < 1$.*

*(ii) The harmonic series is another famous series, but this series is not convergent. It reads*

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \dots$$

*One way to see that the series does not converge is to consider together some successive terms whose sum is bigger than $\frac{1}{2}$ as for example $\frac{1}{3} + \frac{1}{4} > \frac{1}{2}$, then $\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} > \frac{1}{2}$, and then $\frac{1}{9} + \frac{1}{10} + \frac{1}{11} + \frac{1}{12} + \frac{1}{13} + \frac{1}{14} + \frac{1}{15} + \frac{1}{16} > \frac{1}{2}, \dots$. The number of terms in each block is increasing, but their sum is always bigger than $\frac{1}{2}$. One then ends up with an infinite sequence of $\frac{1}{2}$ which sum up to $\infty$.*

Let us provide one result about manipulations of convergent series. In fact, this statement (and its proof) are quite similar to Lemma 5.11 about Riemann integrals.

**Lemma 7.3.** *Let $(a_j)_{j\in\mathbb{N}}$ and $(b_j)_{j\in\mathbb{N}}$ be two sequences of real numbers whose corresponding series are convergent. Let also $\lambda \in \mathbb{R}$.*

*(i)* $(\lambda a_j)_{j\in\mathbb{N}}$ *defines is a convergent series, with* $\sum_{j=1}^{\infty}\lambda a_j = \lambda\sum_{j=1}^{\infty} a_j$,

*(ii)* $(a_j+b_j)_{j\in\mathbb{N}}$ *defines a convergent series, and the following equality holds:*

$$\sum_{j=1}^{\infty}(a_j+b_j) = \sum_{j=1}^{\infty} a_j + \sum_{j=1}^{\infty} b_j,$$

*(iii) If* $s_n := \sum_{j=1}^{n} a_j$ *and* $s'_n := \sum_{j=1}^{n} b_j$ *one has*

$$\sum_{j=1}^{\infty} a_j \cdot \sum_{j=1}^{\infty} b_j = \lim_{n\to\infty} s_n\, s'_n,$$

*but this is NOT equal to* $\sum_{j=1}^{\infty} a_j b_j$.

## 7.2 Series with positive terms only

In this section, let us concentrate on series with positive terms only, which means that the corresponding sequence $(a_j)_{j\in\mathbb{N}}$ satisfies $a_j \geq 0$ for all $j \in \mathbb{N}$. The special feature of such series can be seen in the partial sums: they are increasing. More precisely, if $a_j \geq 0$ for all $j \in \mathbb{N}$, then one has $s_{n+1} \geq s_n$ for all $n \in \mathbb{N}$. In such a situation, the Monotone convergence theorem 1.6 can be applied, and one deduces that the series is convergent whenever the partial sums are upper bounded. In other terms, the sequence $(a_j)_{j\in\mathbb{N}}$ with $a_j \geq 0$ defines a convergent series if and only if there exists $M \in \mathbb{R}$ with $s_n \leq M$ for all $n \in \mathbb{N}$.

**Example 7.4.** *Consider the sequence* $(a_j)_{j\in\mathbb{N}}$ *with* $a_j = \frac{1}{j^2}$ *for all* $j \in \mathbb{N}$, *and let us show that the corresponding series is convergent. Indeed, one has for n large enough*

$$\begin{aligned}
s_n &= 1 + \underbrace{\frac{1}{2^2}+\frac{1}{3^2}}+\underbrace{\frac{1}{4^2}+\frac{1}{5^2}+\frac{1}{6^2}+\frac{1}{7^2}}+\underbrace{\frac{1}{8^2}+\frac{1}{9^2}+\ldots}+\underbrace{\ldots+\frac{1}{n^2}}\\
&\leq 1 + \underbrace{\frac{1}{2^2}+\frac{1}{2^2}}+\underbrace{\frac{1}{4^2}+\frac{1}{4^2}+\frac{1}{4^2}+\frac{1}{4^2}}+\underbrace{\frac{1}{8^2}+\frac{1}{8^2}+\ldots}+\underbrace{\ldots+\frac{1}{n^2}}\\
&\leq 1 + \frac{2}{2^2}+\frac{4}{4^2}+\frac{8}{8^2}+\ldots\\
&= \sum_{j=0}^{\infty}\frac{1}{2^j}\\
&= \sum_{j=0}^{\infty}\Big(\frac{1}{2}\Big)^j\\
&= 2.
\end{aligned}$$

*Thus, all partial sums are bounded by* 2*, and therefore the series is convergent. Note that* 2 *is only an upper bound, the following equality holds and is called the Basel problem:*

$$\sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{\pi^2}{6}.$$

Let us add one statement which is quite natural and simple, and which can be used on a daily basis.

**Lemma 7.5** (Comparison lemma). *Let $(a_j)_{j\in\mathbb{N}}$ and $(b_j)_{j\in\mathbb{N}}$ be two sequences of positive numbers, and assume that the series defined by $(b_j)_{j\in\mathbb{N}}$ is convergent. If there exists $d > 0$ such that $a_j \leq db_j$ for all $j \in \mathbb{N}$, then the series defined by $(a_j)_{j\in\mathbb{N}}$ is convergent as well. In addition one has*

$$\sum_{j=1}^{\infty} a_j \leq d \sum_{j=1}^{\infty} b_j.$$

*Proof.* Let $s_n := \sum_{j=1}^{n} b_j$, and observe that since $\lim_{n\to\infty} s_n$ exists, the increasing sequence of partial sums $(s_n)_{n\in\mathbb{N}}$ is bounded. As a consequence, there exists $M < \infty$ such that $s_n \leq M$ for all $n \in \mathbb{N}$. Thus, one has for any $n \in \mathbb{N}$

$$\sum_{j=1}^{n} a_j \leq \sum_{j=1}^{n} (db_j) = d \sum_{j=1}^{n} b_j = d\, s_n \leq dM, \tag{7.2}$$

which means that the partial sums related to the sequence $(a_j)_{j\in\mathbb{N}}$ are also upper bounded. It then follows from the Monotone convergence theorem 1.6 that this sequence is convergent, which implies the convergence of the series defined by $(a_j)_{j\in\mathbb{N}}$. The last part of the statement follows then easily from (7.2). ☐

Note that a converse statement also holds: if there exists $d > 0$ and $a_j \geq db_j$ for all $j \in \mathbb{N}$, and if the series defined by the sequence $(b_j)_{j\in\mathbb{N}}$ does not converge, then the series defined by $(a_j)_{j\in\mathbb{N}}$ will also not converge.

Two additional criteria for the convergence of positive series are presented in the following exercises.

**Exercise 7.6** (Ratio test). *Prove the following statement :*
Let $(a_j)_{j\in\mathbb{N}}$ be a sequence with positive terms only. Assume that there exists $c \in (0, 1)$ and $N \in \mathbb{N}$ such that

$$\frac{a_{j+1}}{a_j} \leq c \qquad \forall j \geq N.$$

Then the corresponding series is convergent.

**Exercise 7.7** (Integral test). *For $f : [1,\infty) \to \mathbb{R}_+$ decreasing, prove the following statement :*
The series $f(1) + f(2) + f(3) + \dots$ is convergent if and only if $\lim_{M\to\infty} \int_1^M f(x)\,\mathrm{d}x$ is

convergent,
*or in a simpler form show that*

$$\sum_{j=1}^{\infty} f(j) < \infty \quad \Longleftrightarrow \quad \int_1^{\infty} f(x)\,\mathrm{d}x < \infty.$$

**Remark 7.8.** *Let us emphasize a rather basic fact but which has deep consequences. The sum of a finite number of terms is always finite, but series are troublesome because it involves the sum of an infinite number of terms. Thus, for any series, one can always divide it into two parts: one part with a finite number of terms, and one part which contains an infinite number of terms. If a criterion of convergence applies for the infinite part, then this is enough for the convergence of the series. Thus, if one can show that the sequence $(a_j)_{j=N}^{\infty}$ defines a convergent series (by applying some of the criteria mentioned above), then the sequence $(a_j)_{j=1}^{\infty}$ also defines a convergent series, no matter what the first $N-1$ terms are.*

## 7.3 Absolute convergence

In the previous section, all terms in the series were positive. This prevents any cancellation, namely a positive and a negative term summing to 0. When the sign is not fixed, one has to be much more careful. For example, consider again the series defined by $a_j = (-1)^j$. One has

$$-1+1-1+1-1+1-1+\ldots = ?$$

but if we pair the contributions one can obtain

$$-1+\underbrace{1-1}_{0}+\underbrace{1-1}_{0}+\underbrace{1-1}_{0}+\ldots = -1+0+0+0+\ldots = -1,$$

but equivalently one can also obtain

$$\underbrace{-1+1}_{0}\underbrace{-1+1}_{0}\underbrace{-1+1}_{0}+\ldots = 0+0+0+0+\ldots = 0.$$

Clearly, these expressions are misleading, and in fact the only correct approach is to consider the partial sums and to observe that they do not converge. Roughly speaking, one can say that for series with terms of arbitrary signs, the order of summation matters, while for series with positive terms only, the order of summation does not matter.

For series with terms of arbitrary sign, one can also consider the summation process of the absolute values. This is the content of the next definition.

**Definition 7.9** (Absolute convergence of a series)**.** *A sequence of real number $(a_j)_{j\in\mathbb{N}}$ defines an* absolutely convergent series *if the series defined by the positive terms $(|a_j|)_{j\in\mathbb{N}}$ is convergent. In other terms, a series is absolutely convergent if $\sum_{j=1}^{\infty}|a_j| < \infty$.*

It is clear that for series with positive terms only, the previous definition coincides with the convergence of the series. On the other hand, if $a_j$ takes an arbitrary sign, then this new notion of convergence is stronger, as shown in the next statement.

**Theorem 7.10.** *Any series which is absolutely convergent is also convergent in the sense of Definition 7.1.*

Before giving the proof, let us stress that the converse statement is not true: there are converging series which are not absolutely convergent. One rather famous example will be presented soon.

*Proof.* Consider a sequence $(a_j)_{j\in\mathbb{N}}$ which defines an absolutely convergent series, which means $\sum_{j=1}^{\infty}|a_j|<\infty$. Let us set $b_j:=a_j$ whenever $a_j\geq 0$ and $b_j:=0$ whenever $a_j<0$. Similarly, let us set $c_j:=-a_j$ whenever $a_j<0$ and $c_j=0$ whenever $a_j\geq 0$. Note that $b_j\geq 0$ and $c_j\geq 0$ for all $j\in\mathbb{N}$. Clearly, one has for any $j\in\mathbb{N}$

$$a_j=b_j-c_j \qquad \text{and} \qquad |a_j|=b_j+c_j,$$

where the sum is a sum of two positive numbers. Since $b_j\leq|a_j|$ and since $c_j\leq|a_j|$ it follows from the Comparison lemma 7.5 that

$$\sum_{j=1}^{\infty}b_j\leq\sum_{j=1}^{\infty}|a_j|<\infty,$$

and similarly

$$\sum_{j=1}^{\infty}c_j\leq\sum_{j=1}^{\infty}|a_j|<\infty.$$

As a consequence one has

$$s_n:=\sum_{j=1}^{n}a_j=\sum_{j=1}^{n}(b_j-c_j)=\sum_{j=1}^{n}b_j-\sum_{j=1}^{n}c_j$$

and then

$$\lim_{n\to\infty}s_n=\lim_{n\to\infty}\Big(\sum_{j=1}^{n}b_j-\sum_{j=1}^{n}c_j\Big)=\lim_{n\to\infty}\sum_{j=1}^{n}b_j-\lim_{n\to\infty}\sum_{j=1}^{n}c_j=\sum_{j=1}^{\infty}b_j-\sum_{j=1}^{\infty}c_j$$

which means that the partial sums converge. Note that the second equality holds precisely because the two limits exist separately, otherwise we could not separate the limit into two limits. □

Thus, if a series is absolutely convergent, the summation process can be done in any order, and the series will converge. But if the series is not absolutely convergent, the order of summation matters. However, several criteria also exist for such series, and we present one which is rather simple.

**Theorem 7.11** (Alternating series). *Let $(a_j)_{j\in\mathbb{N}}$ be a sequence of real numbers satisfying*

*(i)* $\lim_{j\to\infty} a_j = 0$,

*(ii)* $a_j a_{j+1} \leq 0$ *for all* $j \in \mathbb{N}$,

*(iii)* $|a_{j+1}| \leq |a_j|$ *for all* $j \in \mathbb{N}$.

*Then the corresponding series is convergent.*

It is rather clear that the condition *(ii)* gives the name to such series: one has alternatively a positive term and a negative one. Before providing the proof, let us look at one example. For $a_j := (-1)^j \frac{1}{j}$, one easily observes that the three conditions are satisfied, which means that the corresponding series is convergent. Thus one has $\sum_{j=1}^{\infty}(-1)^j \frac{1}{j} < \infty$. On the other hand, recall from Examples 7.2 that the harmonic series does not converge, namely $\sum_{j=1}^{\infty} \frac{1}{j} = \infty$. Thus, the series defined by $a_j := (-1)^j \frac{1}{j}$ is convergent, but is NOT absolutely convergent.

*Proof.* We suppose in this proof that $a_1 \geq 0$, but the case $a_1 \leq 0$ can be treated similarly. Let us set $b_j := a_{2j-1}$ and $c_j := -a_{2j}$, which means that $a_1 + a_2 + a_3 + a_4 + \dots$ can be rewritten as $b_1 - c_1 + b_2 - c_2 + \dots$, where $b_j \geq 0$ and $c_j \geq 0$ for any $j \in \mathbb{N}$. Let us also set

$$s_n := b_1 - c_1 + b_2 - c_2 + \dots + b_{n-1} - c_{n-1} + b_n$$

and

$$t_n := b_1 - c_1 + b_2 - c_2 + \dots + b_{n-1} - c_{n-1} + b_n - c_n.$$

Clearly one has

$$s_{n+1} = s_n \underbrace{-c_n + b_{n+1}}_{\leq 0} \leq s_n$$

while

$$t_{n+1} = t_n \underbrace{+b_{n+1} - c_{n+1}}_{\geq 0} \geq t_n.$$

Also , $s_n \geq t_n$ by their definition. Thus, one ends up with the following inequalities

$$\begin{array}{ccccc} s_n & \geq & s_{n+1} & \geq & s_{n+2} & \dots \\ \vee\!\!\!\mid\!\!\!\vee & & \vee\!\!\!\mid\!\!\!\vee & & \vee\!\!\!\mid\!\!\!\vee & \\ t_n & \leq & t_{n+1} & \leq & t_{n+2} & \dots \end{array}$$

As a consequence, $(s_n)_{n\in\mathbb{N}}$ is a decreasing sequence, bounded from below by $t_1$, while $(t_n)_{n\in\mathbb{N}}$ is an increasing sequence, bounded from above by $s_1$. It follows by the Monotone convergence theorem 1.6 that these two sequences are converging. Let us then set $\lim_{n\to\infty} s_n = s_\infty \in \mathbb{R}$ and $\lim_{n\to\infty} t_n = t_\infty \in \mathbb{R}$. Clearly, for any $n \in \mathbb{N}$

$$s_n \geq s_\infty \geq t_\infty \geq t_n = s_n - c_n.$$

Since $\lim_{n\to\infty} c_n = 0$, one deduces by the Squeeze theorem that

$$s_\infty = \lim_{n\to\infty} s_n = \lim_{n\to\infty} t_n = t_\infty$$

and these limits imply the existence of the limit of the partial sums. Thus, the series defined by $(a_j)_{j\in\mathbb{N}}$ is convergent, and one has $\sum_{j=1}^\infty a_j = s_\infty = t_\infty$. □

As we already know, the series defined by $a_j = (-1)^j \frac{1}{j}$ is convergent but not absolutely convergent. Let us observe that the summation procedure has to be done carefully. For example if we think that

$$\begin{aligned} -1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \frac{1}{5} + \frac{1}{6} - \frac{1}{7} + \frac{1}{8} - \ldots = &- 1 - \frac{1}{3} - \frac{1}{5} - \frac{1}{7} - \ldots \\ &+ \frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \frac{1}{8} + \ldots \end{aligned} \tag{7.3}$$

then we do a big mistake. Indeed one has

$$-1 - \frac{1}{3} - \frac{1}{5} - \frac{1}{7} - \ldots < -\frac{1}{2}\Big(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \ldots\Big) = -\infty$$

and similarly

$$+\frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \frac{1}{8} + \ldots > \frac{1}{2}\Big(\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \ldots\Big) = \infty.$$

As a consequence, the r.h.s. of (7.3) is really ill-defined, while the l.h.s. is well defined as an alternating series.

## 7.4 Power series

In this section, we come back to the question raised in (7.1).

**Definition 7.12** (Power series)**.** *Let $(a_j)_{j=0}^\infty$ be a sequence of real numbers. A* power series *is a "formal" function defined for $x \in \mathbb{R}$ by*

$$\sum_{j=0}^\infty a_j x^j. \tag{7.4}$$

This definition is only formal because a function is defined with its domain. Here, it is not clear what is the domain of this function, or in other words which $x$ can be considered for this infinite sum. In fact, for $x = 0$, this expression is well-defined, but can it be defined for $x \neq 0$? Clearly, the answer will depend on the coefficients $a_j$. For example, if $a_j = \frac{1}{j!}$, then the power series (7.4) corresponds to the definition of the exponential function, and it has been shown that this sum is finite for any $x \in \mathbb{R}$. It means that for this choice of the coefficients $a_j$, this formal power series corresponds to a well-defined function on $\mathbb{R}$. But what about a more general situation? The following statement provides already some information.

**Theorem 7.13.** *Let $(a_j)_{j=0}^\infty$ be a sequence of real numbers, and assume that there exists $r \geq 0$ such that the series with positive terms $\sum_{j=0}^\infty |a_j| r^j$ converges. Then, for any $x \in \mathbb{R}$ with $|x| \leq r$, the series $\sum_{j=0}^\infty a_j x^j$ is absolutely convergent.*

*Proof.* Since $|a_j x^j| = |a_j||x|^j \leq |a_j| r^j$, one infers from the Comparison lemma 7.5 with $d = 1$ that the series $\sum_{j=0}^\infty |a_j x^j|$ is convergent. This correspond to the absolute convergence of the power series $\sum_{j=0}^\infty a_j x^j$. □

The meaning of the above statement is rather clear: Whenever $\sum_{j=0}^\infty |a_j| r^j < \infty$ the corresponding power series $\sum_{j=0}^\infty a_j x^j$ is well-defined for all $|x| \leq r$. In this sense, for $|x| \leq r$, the power series defines a function on the interval $[-r, r]$. It is thus interesting to look for the largest value of $r$, and hence for the maximal domain of definition of the function:

**Definition 7.14** (Radius of convergence)**.** *The least upper bound of $r \geq 0$ such that $\sum_{j=0}^\infty |a_j| r^j < \infty$ is called* the radius of convergence *of the power series $\sum_{j=0}^\infty a_j x^j$.*

Note that the radius of convergence can be 0, a finite positive number, or equal to $\infty$. It all depends on the coefficients $(a_j)_{j=0}^\infty$, as shown in the following examples.

**Examples 7.15.** *(i) If $a_j = \frac{1}{j!}$, then one has to consider $\sum_{j=0}^\infty \frac{1}{j!} r^j$. For the application of the ratio test presented in Exercise 7.6, one computes*

$$\frac{|a_{j+1}| r^{j+1}}{|a_j| r^j} = \frac{\frac{1}{(j+1)!} r^{j+1}}{\frac{1}{j!} r^j} = \frac{j!}{(j+1)!} \frac{r^{j+1}}{r^j} = \frac{1}{j+1} r,$$

*which means that for any fixed $r > 0$ and for any $j > 2r$, one has*

$$\frac{r}{j+1} < \frac{r}{2r+1} < \frac{r}{2r} = \frac{1}{2}.$$

*Thus, the ratio test implies that the series $\sum_{j=0}^\infty a_j r^j$ is absolutely convergent for any $r > 0$. As a consequence, the radius of convergence of the corresponding power series is $\infty$, which means that the power series $\sum_{j=0}^\infty \frac{1}{j!} x^j$ is well-defined for all $x \in \mathbb{R}$. Obviously, we already knew it because this power series corresponds to the exponential function.*

*(ii) If $a_j = j!$, then for an application of the ratio test one considers*

$$\frac{|a_{j+1}| r^{j+1}}{|a_j| r^j} = \frac{(j+1)! r^{j+1}}{j! r^j} = (j+1) r.$$

*Thus, for any fixed $r > 0$ there exists $j_0 \in \mathbb{N}$ such that $(j+1)r > 1$ for any $\geq j_0$. It means that the ratio test fails for any $r > 0$. However, be aware that the failure of ONE test does not imply that the series is not convergent, you should try others. In the present situation, it turns out that the power series $\sum_{j=0}^\infty j! x^j$ has really a radius of convergence equal to 0. One can not define a function by this power series except at $x = 0$.*

*(iii)* *If $a_j = 1$ for all $j$, then the corresponding power series has radius of convergence equal to 1. Indeed, the power series $\sum_{j=0}^{\infty} x^j$ corresponds to the geometric series, and we have seen in Examples 7.2 that this series converges absolutely for $|x| < 1$ but does not converge absolutely for $|x| \geq 1$.*

*(iv)* *A slightly more surprising example is provided by the power series with coefficients $a_j := \ln(j)$, since these coefficients are growing. If we want to apply the ratio test to the corresponding series one observes that $\frac{|a_{j+1}|r^{j+1}}{|a_j|r^j} = \frac{\ln(j+1)}{\ln(j)}r$. For fixed $r$, is it possible to have $\frac{\ln(j+1)}{\ln(j)}r \leq c < 1$ for some constant $c$ and all $j$ large enough? The answer is YES. Indeed, consider a fixed $r < 1$ and the equation*

$$\frac{\ln(j+1)}{\ln(j)}r \leq c \Longleftrightarrow \ln(j+1) \leq \frac{c}{r}\ln(j) \Longleftrightarrow \mathrm{e}^{\ln(j+1)} \leq \mathrm{e}^{\ln(j^{c/r})} \Longleftrightarrow j+1 \leq j^{c/r}.$$

*Thus, if we fix $c$ with $r < c < 1$, then one has $c/r > 1$, and the the inequality $j + 1 \leq j^{c/r}$ is satisfied for $j$ large enough. In summary, it means that the ratio test can be applied to this example, and that for any $r < 1$ there exists $c < 1$ and $j_r > 0$ such that $\frac{\ln(j+1)}{\ln(j)}r \leq c$ for all $j \geq j_r$. The corresponding power series is then absolutely convergent. Since this argument holds for any $r < 1$, but does not hold for $r \geq 1$ we infer that the radius of convergence of this series is equal to 1.*

As already said several times, the radius of convergence for the exponential function is $\infty$, and this is why the exponential function is defined for all real numbers. It turns out that the radius of convergence for a few other common functions is also $\infty$. For example, the power series

$$\sum_{j=0}^{\infty} \frac{(-1)^j x^{2j}}{(2j)!} \qquad \text{and} \qquad \sum_{j=0}^{\infty} \frac{(-1)^j x^{2j+1}}{(2j+1)!}$$

have a radius of convergence equal to $\infty$, and correspond to the expressions $\cos(x)$ and $\sin(x)$, respectively. In the general situation, whenever we define a power series $\sum_{j=0}^{\infty} a_j x^j$ one should provide its radius of convergence $r$, and then the function $f : (-r, r) \to \mathbb{R}$ given by $f(x) := \sum_{j=0}^{\infty} a_j x^j$ is a well-defined function. Let's study it!

## 7.5 Differentiation and integration of power series

In this section we consider a function $f : (-r, r) \to \mathbb{R}$ given by $f(x) = \sum_{j=0}^{\infty} a_j x^j$, where $r$ denotes the radius of convergence of the power series. Our aim is to study the differentiation of $f$, and its integration. We start with a technical lemma.

**Lemma 7.16.** *Let $(a_j)_{j=0}^{\infty}$ be a sequence of real numbers, and assume that the power series $\sum_{j=0}^{\infty} a_j x^j$ has a radius of convergence $r > 0$. Then the two series defined by*

$$\sum_{j=1}^{\infty} j a_j x^{j-1} \qquad \textit{and} \qquad \sum_{j=0}^{\infty} \frac{1}{j+1} a_j x^{j+1}$$

*converge absolutely for any $x$ with $|x| < r$.*

*Proof.* Let us fix $x \in \mathbb{R}$ with $x \neq 0$ and $|x| < r$. Let us also fix $c \in \mathbb{R}$ with $|x| < c < r$. Since $\lim_{j\to\infty} j^{1/j} = 1$ (as proved during the tutorial session) one has for $j$ large enough,

$$|ja_j x^j| = |a_j| \big|j^{1/j} x\big|^j \leq |a_j| c^j$$

and thus

$$|ja_j x^{j-1}| = \frac{1}{|x|} |ja_j x^j| \leq \frac{1}{|x|} |a_j| c^j.$$

By assumption, the series $\sum_{j=0}^{\infty} |a_j| c^j$ is convergent, and by the Comparison lemma 7.5 with $d = \frac{1}{|x|}$ one infers that the series $\sum_{j=0}^{\infty} ja_j x^{j-1}$ is absolutely convergent.

Similarly one has

$$\frac{|a_j|}{j+1} |x|^{j+1} \leq |x| |a_j| |x|^j \leq |x| |a_j| c^j,$$

and by one more application of the Comparison lemma with $d = |x|$ one infers that $\sum_{j=0}^{\infty} \frac{1}{j+1} a_j x^{j+1}$ converges absolutely. ☐

The role of the series appearing in the previous statement is quite clear: they correspond to the derivative and to an indefinite integral for the initial power series. In the next Theorem, we state it properly. However, we do not provide a proof since it is a little bit tricky: a limit has to be exchanged with an infinite sum. On the other hand, if we accept this exchange, then it follows from the previous lemma that the two new functions have a radius of convergence at least equal to the radius of convergence of the initial function.

**Theorem 7.17.** *Let $(a_j)_{j=0}^{\infty}$ be a sequence of real numbers, and assume that the power series $\sum_{j=0}^{\infty} a_j x^j$ has a radius of convergence $r > 0$. Then the corresponding function $f : (-r, r) \to \mathbb{R}$ given by $f(x) = \sum_{j=0}^{\infty} a_j x^j$ is differentiable with derivative $f'(x) = \sum_{j=1}^{\infty} ja_j x^{j-1}$ for any $x \in (-r, r)$. In addition, an indefinite integral for $f$ is provided by the function $F : (-r, r) \to \mathbb{R}$ with $F(x) := \sum_{j=0}^{\infty} \frac{1}{j+1} a_j x^{j+1}$ for any $x \in (-r, r)$.*

One of the interesting consequences of this theorem is about higher derivatives: since the derivative of $f$ is also a power series which is absolutely convergent for any $x$ with $|x| < r$, then the theorem can be applied again, and one gets that the second derivative of $f$ also exists. Thus, any function defined by a power series and with a radius of convergence $r > 0$ is in fact a smooth function, which means that it belongs to $C^k\big((-r, r)\big)$ for any $k \in \mathbb{N}$. In other words, any power series with a non-zero radius of convergence defines a function which is infinitely many times differentiable.

With this theorem, we have fully proved our initial approach for the exponential function, namely as a function $f$ given by a power series and which satisfies $f' = f$.

As a final remark, let us mention that a function $f$ defined by a power series and with a radius of convergence $r > 0$ can sometime be defined on a domain bigger than

$(-r, r)$. For example, the function $f : (-1, 1) \to \mathbb{R}$ given by $f(x) := \sum_{j=1}^{\infty} \frac{1}{j} x^j$ is not well defined at 1, but it is well-defined for $x = -1$. However, for $x = -1$ it can not be defined as an absolutely convergent series, but it is an alternating series, as seen in Theorem 7.11. Note that this kind of extensions are much more interesting once the function are defined on the complex plane $\mathbb{C}$. In fact, power series are an essential part of complex analysis, and several nice constructions can be studied in this framework. If the opportunity is given, don't miss a complex analysis course.

The end