1RA – Differentiation

lecture notes

Yuzhao Wang[1]

University of Birmingham
Autumn Term 2022

---

[1]Based on the notes by Alessio Martini

Contents

1 Sets and functions

Informally speaking, a *set* is a (finite or infinite) collection of objects (typically some kind of mathematical entities), called *elements* of the set.

We mainly use latin capital letters, such as A, B, C, ..., to denote sets. We write $x \in A$ to say that "x is an element of A"; correspondingly, we write $x \notin A$ to say that "x is not an element of A".

In mathematics we work with objects of various kinds and correspondingly we may consider sets of such objects. A typical example is given by sets of *numbers*; we write

- $\mathbb{N}$ for the set of *natural numbers*, i.e., the positive integers $1, 2, 3, \dots$;

- $\mathbb{N}_0$ for the set of *nonnegative integers*, that is, the natural numbers[1] including 0;

- $\mathbb{Z}$ for the set of *integers* (or *whole numbers*), including 0, the natural numbers $1, 2, 3, \dots$, as well as their negatives $-1, -2, -3, \dots$;

- $\mathbb{Q}$ for the set of *rational numbers*, i.e., those that can be written as a fraction p/q of integers $p, q \in \mathbb{Z}$;

- $\mathbb{R}$ for the set of *real numbers*, which include the rational numbers as well as other (irrational) numbers such as $\sqrt{2}$ and π.

Note however that elements of a set need not be numbers! For example, in geometry, we can consider sets of points (the points of a line, or a plane, or a circle, ...).

1.1 Denoting and defining sets

Sets can be defined by listing their elements: for example, we can write

$$A = \{1, 2, 3\}$$

to say that A is the set whose elements are 1, 2, 3. This mainly makes sense for finite sets, where a complete list of elements can actually be written. Sometimes we write

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

however this is not entirely precise, since the dots $\dots$ do not clearly specify what comes after $1, 2, 3$. Another way of defining sets is through the *set-builder notation*, i.e., by specifying a property that the elements of the set must satisfy; for example,

$$P = \{x \in \mathbb{Z} : x = 2y \text{ for some } y \in \mathbb{Z}\}$$

says that P is the set of all integers x that are a multiple of 2, that is, the set of *even* integers. We also can construct new sets by using images of other sets under a function, for example,

$$Q = \{2y + 1; \text{ for } y \in \mathbb{Z}\}$$

---

[1]<u>Warning</u>: other texts may use the symbol $\mathbb{N}$ to denote the set of nonnegative integers (i.e., 0 is considered to be a "natural number"). This is mainly a matter of convention or taste; the important thing is to be consistent throughout.

says that Q is the set of odd integers. See Definition 1.4 for more details. We use the symbol

$$\emptyset$$

to denote the *empty set*, that is, the set that has no elements.

1.2   Subsets and equality of sets

Given two sets A and B, we say that:

- A is a *subset* of B (in formulas: $A \subseteq B$, or alternatively $B \supseteq A$) if every element of A is also an element of B;

- A is *equal* to B (in formulas: $A = B$) if A and B have the the same elements, that is, every element of A is also an element of B, and every element of B is also an element of A.

Note that $A = B$ if and only if both $A \subseteq B$ and $B \subseteq A$. In particular, every set A is a subset of itself; any subset of A that is not equal to A is called a *proper subset* of A. Note also that, according to the above definition of equality,

$$\{1, 2, 3\} = \{3, 1, 2\};$$

in other words, the order in which we list the elements of a set does not matter, and different reorderings of the same elements define the same set. We also have

$$\{1, 2, 2\} = \{1, 2\},$$

which means that sets can't have repeated elements.

The empty set $\emptyset$ is a subset of any set and a proper subset of any set except itself. In particular, the empty set is a subset of itself, which implies the uniqueness of empty set.

1.3   Set operations

It is convenient to define the following *set operations*, that allow us to construct new sets starting from given ones.

- The *union* of two sets A and B, denoted by $A \cup B$, is the set defined by

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

- The *intersection* of A and B, written $A \cap B$, is the set defined by

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

If $A \cap B = \emptyset$, then A and B are said to be disjoint; we say that A intersects (meets) B if they are not disjoint.

- The *difference* of A and B, written $A \setminus B$, is the set given by

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

Suppose that a *universal set* U (a set containing all elements being discussed) has been fixed, then the *complement* of B is, written B^c, is the set given by

$$B^c = U \setminus B.$$

1.4   Ordered pairs and cartesian product

We have seen above that, for any two given objects x and y, the sets $\{x, y\}$ and $\{y, x\}$ are the same. If we want to "keep track" of the order, instead of the set $\{x, y\}$ we must use the *ordered pair* (x, y) — note the different kind of parentheses. Given an ordered pair (x, y), we say that x is the *first element* (or *first component*) of the pair and y is the *second element* (or *second component*) of the pair. For two ordered pairs (x, y) and (z, w) we have

$$(x, y) = (z, w) \text{ if and only if } x = z \text{ and } y = w;$$

in other words, two ordered pairs are the same if and only if they have the same first element and the same second element. Here the order of the elements matters! In particular $(x, y) \neq (y, x)$, unless $x = y$.

A typical example where order matters is when speaking of points of a plane: the point with coordinates $(1, 2)$ is not the same as the point of coordinates $(2, 1)$. Using cartesian coordinates, it is actually natural to think of the plane as the set of all ordered pairs of real numbers: here we are identifying each point of the plane with its cartesian coordinates.

This example leads us to the definition of the *cartesian product* of two sets A and B, written $A \times B$, which is defined as the set of all ordered pairs (x, y) where $x \in A$ and $y \in B$. When $A = B$, we also write A^2 instead of $A \times A$.

1.5   Intervals

It is convenient to introduce notation for particular subsets of $\mathbb{R}$, called *intervals*. First we consider *bounded intervals*, whose elements are those real numbers between two given numbers $a, b \in \mathbb{R}$, called *endpoints* of the interval. There are several types of such intervals, depending on whether each endpoint is included or not included:

- the open interval[2]

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

- the closed interval

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

- the half-open intervals

$$(a, b] = \{x \in \mathbb{R} : a < x \leq b\},$$
$$[a, b) = \{x \in \mathbb{R} : a \leq x < b\}.$$

---

[2]<u>Warning</u>: there is an unfortunate notation clash between open intervals and ordered pairs. For example, the expression

$$(2, 3)$$

may denote

- either the ordered pair with components 2 and 3,

- or the open interval with endpoints 2 and 3.

Usually the context will determine which of the two is meant. Other texts may avoid this clash by using different notation; e.g., sometimes the notation $]a, b[$ is used for the open interval with endpoints a and b.

In addition, we consider *unbounded intervals*:

- the open half-lines

$$(a, \infty) = \{x \in \mathbb{R} : x > a\},$$
$$(-\infty, a) = \{x \in \mathbb{R} : x < a\};$$

- the closed half-lines

$$[a, \infty) = \{x \in \mathbb{R} : x \geq a\},$$
$$(-\infty, a] = \{x \in \mathbb{R} : x \leq a\};$$

- the real line

$$(-\infty, \infty) = \mathbb{R}.$$

The symbol ∞ reads "infinity". It is suggestive to think of $-\infty$ and ∞ as "endpoints" of the real line $\mathbb{R}$. Indeed we may define the *extended real line* $\overline{\mathbb{R}}$ by formally adding two new points to the real line $\mathbb{R}$:

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}.$$

We can extend to $\overline{\mathbb{R}}$ the ordering of $\mathbb{R}$ by stipulating, as it is natural, that

$$-\infty < x < \infty$$

for all real numbers $x \in \mathbb{R}$. In this way, the above definitions of unbounded intervals are consistent with those of bounded intervals.

While it is convenient to use the symbols ∞ and $-\infty$ in real analysis, it is crucial to remember that

$$\boxed{\infty \text{ and } -\infty \text{ are } \mathbf{NOT} \text{ real numbers,}}$$

that is, they do not belong to $\mathbb{R}$, and cannot be treated as if they did. In Section 3.5 below we will extend some of the operations on real numbers (addition and multiplication) in order to deal with certain expressions involving $\pm\infty$; however not all the rules that commonly apply when operating with real numbers extend to the case of $\pm\infty$, so one must be careful when dealing with these symbols.

1.6 Some logical symbols

We record here a few symbols that are commonly used when writing mathematics.

- $\Rightarrow$ is the *logical implication* symbol: the expression "$P \Rightarrow Q$" reads "P implies Q" or "if P then Q".

- $\Leftarrow$ is the *reverse logical implication* symbol: the expression "$P \Leftarrow Q$" is another way of writing "$Q \Rightarrow P$".

- $\Leftrightarrow$ is the *logical equivalence* symbol: the expression "$P \Leftrightarrow Q$" reads "P if and only if Q", and is shorthand for "$P \Rightarrow Q$ and $Q \Rightarrow P$".

- $\forall$ is the *universal quantifier* symbol, and reads "for all" or "for each".

- $\exists$ is the *existential quantifier* symbol, and reads "there exists".

1.7   Functions

The notion of *function* is a fundamental notion in mathematics.

Definition 1.1. Given two sets A and B, a *function from A to B* is a rule that associates to every element of A exactly one element of B. We write $f : A \to B$ to say that f is a function from A to B; in this case A is called the *domain* of f and B is called the *codomain* of f.

There is a fair amount of notation and terminology related to functions. Some of it is described in the following definitions.

Definition 1.2. If $f : A \to B$ and $x \in A$, the element of B associated to x by f is called the *image* of x (via f) and is denoted by $f(x)$. We also say that:

- f maps x to $f(x)$;

- $f(x)$ is the result of applying f to x;

- $f(x)$ is the value of f at x.

Definition 1.3. If $f : A \to B$ and $y \in B$, any element x of A such that $f(x) = y$ is called a *preimage* of y (via f).

It is worth noticing that, according to Definition 1.1, if $f : A \to B$, then every element x of the domain A has exactly one image $f(x)$ in the codomain B; however an element y of the codomain B may have no preimage in A at all, or multiple preimages in A. We discuss this further in Section 1.9 below.

Definition 1.4. The *image* (or *range*) of a function $f : A \to B$ is the subset of B whose elements are the images via f of all the elements of A, and is denoted by $f(A)$. In other words,

$$f(A) = \{f(x) : x \in A\}.$$

A common way is using a formula, describing $f(x)$ in terms of x. For example, we can define a function $f : \mathbb{R} \to \mathbb{R}$ by the formula

$$f(x) = x^3 - 3x^2 + 2. \tag{1.1}$$

In place of the previous formula, sometimes the following notation is also used:

$$f : x \mapsto x^3 - 3x^2 + 2.$$

We may use more involved descriptions, such as a *definition by cases*: for example, the function $s : \mathbb{R} \to \mathbb{R}$ given by

$$s(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0 \end{cases} \tag{1.2}$$

is often called the *sign function*. Another important example is the *modulus function*, which is the function from $\mathbb{R}$ to $\mathbb{R}$ that associates to every $x \in \mathbb{R}$ its modulus (or absolute value) $|x|$, given by

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

For a function $f : \mathbb{R} \to \mathbb{R}$ such as the one defined in (1.1), we can consider "plotting its graph", that is, the curve in the xy-plane defined by the equation $y = f(x)$. In other words, the graph of a function $f : \mathbb{R} \to \mathbb{R}$ can be thought of a particular subset of the plane $\mathbb{R} \times \mathbb{R}$. This idea can be extended to more general functions.

Definition 1.5. The *graph* of a function $f : A \to B$ is the subset Γ_f of the cartesian product $A \times B$ defined by

$$\Gamma_f = \{(x, y) \in A \times B : y = f(x)\}.$$

In case $A, B \subseteq \mathbb{R}$, then the graph of a function $f : A \to B$ is a subset of $\mathbb{R}^2$ and therefore we can think of plotting it (as a set of points of the plane). Higher dimensional variants are conceivable: for example, the graph of the function $g : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$g(x, y) = x^2 + y^2 \tag{1.3}$$

can be thought of as a subset of $\mathbb{R}^3$,

$$\Gamma_g = \{(x, y, z) \in \mathbb{R}^3 : z = x^2 + y^2\},$$

and can be plotted as a surface in 3-dimensional space.

When we look at an expression such as (1.3), we usually say that x and y are the *variables* of the function g. We can also say that g is function of two *real* variables, to indicate that x and y take values from the set $\mathbb{R}$ of real numbers. Further, we say that g is *real-valued*, since it takes its values in $\mathbb{R}$.

In this series of lectures, we will be mainly concerned with real-valued functions of one real variable, that is, functions whose domain and codomain are subsets of $\mathbb{R}$, such as the function $f : \mathbb{R} \to \mathbb{R}$ defined in (1.1). We will say more on this in Section 2 below.

Definition 1.6. We say that two functions $f : A \to B$ and $g : C \to D$ are *equal*, and we write $f = g$, if

- $A = C$ (that is, f and g have the same domain),

- $B = D$ (that is, f and g have the same codomain), and

- $f(x) = g(x)$ for all $x \in A$.

As a matter of fact, equality of functions can be reduced to equality of sets: indeed Definition 1.6 can be reworded by saying that two functions are equal if they have the same domain, codomain and graph.[3]

1.8   Restriction, composition, identity function, constant functions

We now discuss some simple, yet very important, ways of constructing new functions starting from existing ones.

---

[3]For this reason, some texts define a function from A to B as a triple (A, B, Γ), where A and B are sets and Γ is a subset of $A \times B$ such that for all $x \in A$, there exists exactly one $y \in B$ such that $(x, y) \in X$; in our language, Γ is the graph of the function.

Definition 1.7. Let $f : A \to B$ and $S \subseteq A$. The *restriction* of f to S, denoted by $f|_S$, is the function from S to B defined by

$$f|_S(x) = f(x)$$

for all $x \in S$.

Example 1.8. Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = x \qquad \text{and} \qquad g(x) = |x|.$$

Then $f \neq g$; indeed $f(-1) = -1 \neq 1 = g(-1)$. However, for all $x \in [0, \infty)$,

$$g(x) = |x| = x = f(x),$$

so $g|_{[0,\infty)} = f|_{[0,\infty)}$.

Definition 1.9. Given two functions $f : A \to B$ and $g : C \to D$ such that $f(A) \subseteq C$, we define the *composition* of f and g as the function $g \circ f : A \to D$ such that

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

for all $x \in A$.

In other words, the image via the composition $g \circ f$ of an element $x \in A$ is obtained by first applying f to x, thus obtaining an element $y = f(x) \in C$, and then applying g to y.

Note that, for the composition $g \circ f$ of f and g to be defined, the image of f must be contained in the domain of g; this happens, for example, when the codomain of f is equal to the domain of g. The resulting function $g \circ f$ has then the same domain as f and the same codomain as g.

Example 1.10. Let the functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = x + 1, \qquad g(x) = x^2$$

for all $x \in \mathbb{R}$. Then, for all $x \in \mathbb{R}$,

$$(g \circ f)(x) = g(f(x)) = (x + 1)^2 = x^2 + 2x + 1,$$

while

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

From this we can see that $g \circ f \neq f \circ g$; indeed

$$(g \circ f)(1) = 4 \neq 2 = (f \circ g)(1).$$

In other words, composition of functions is *not commutative*: the result of the composition may depend on the order in which the functions are composed.

Definition 1.11. Given any set A, the function from A to A that associates to every $x \in A$ the element x itself is called the *identity function* of A and is denoted by id_A. In other words, $\mathrm{id}_A : A \to A$ is defined by

$$\mathrm{id}_A(x) = x$$

for all $x \in A$.

From the above definitions, it is immediately checked that, for all functions $f : A \to B$,

$$f \circ \mathrm{id}_A = \mathrm{id}_B \circ f = f. \tag{1.4}$$

In other words, identity functions behave as "identity elements" for the operation of composition. Moreover, if $f : A \to B$ and $S \subseteq A$, then

$$f|_S = f \circ \mathrm{id}_S.$$

Definition 1.12. A function $f : A \to B$ is said to be *constant* if there exists $b \in B$ such that $f(x) = b$ for all $x \in A$.

In other words, a function $f : A \to B$ is constant if its image $f(A)$ has exactly one element.

1.9   Invertibility: injective, surjective and bijective functions

Definition 1.13. A function $f : A \to B$ is said to be *invertible* if there exists a function $g : B \to A$ such that, for all $x \in A$ and $y \in B$,

$$f(x) = y \quad \text{if and only if} \quad g(y) = x;$$

if it exists, such a function g is uniquely determined by f and is called the *inverse* of f. We use the notation f^{-1} for the inverse of f.

From the definition it is clear that, if $g : B \to A$ is the inverse of $f : A \to B$, then conversely f is the inverse of g; in other words, if f is invertible, then its inverse f^{-1} is invertible too, and $(f^{-1})^{-1} = f$.

We have already observed in (1.4) that identity functions can be thought of as "identity elements" for the operation of composition of functions; similarly, the next proposition shows that the inverse of a function is indeed the inverse with respect to composition.

Proposition 1.14. *The function $g : B \to A$ is the inverse of $f : A \to B$ if and only if*

$$g \circ f = \mathrm{id}_A \qquad and \qquad f \circ g = \mathrm{id}_B. \tag{1.5}$$

Example 1.15. Please note that <u>both</u> conditions in (1.5) must be satisfied for g to be the inverse of f. For example, let $f : [0, \infty) \to \mathbb{R}$ and $g : \mathbb{R} \to [0, \infty)$ be defined by

$$f(x) = \sqrt{x}, \qquad g(x) = x^2.$$

Then, for all $x \in [0, \infty)$,

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

that is, $g \circ f = \mathrm{id}_{[0,\infty)}$. However g is not the inverse of f: indeed, for all $y \in \mathbb{R}$,

$$f(g(y)) = \sqrt{y^2} = |y|;$$

in particular, $(f \circ g)(-1) = 1 \neq -1$, so $f \circ g \neq \mathrm{id}_{\mathbb{R}}$.

One may ask what conditions one needs to ask of a function $f : A \to B$ for it to be invertible. The following properties turn out to be relevant.

Definition 1.16. We say that $f : A \to B$ is *injective* (or *one-to-one*, or an *injection*) if distinct elements of A are mapped by f to distinct elements of B; in other words, for all $x, x' \in A$,

$$x \neq x' \qquad \Longrightarrow \qquad f(x) \neq f(x').$$

Definition 1.17. We say that $f : A \to B$ is *surjective* (or *onto*, or a *surjection*) if

$$f(A) = B,$$

that is, the range of f is equal to the codomain of f.

Definition 1.18. We say that a function is *bijective* if it is both injective and surjective.

Proposition 1.19. *Let $f : A \to B$.*

(i) *The function f is injective if and only if every element of B has at most one preimage in A.*

(ii) *The function f is surjective if and only if every element of B has at least one preimage in A.*

(iii) *The function f is invertible if and only if f is bijective.*

Example 1.20. Let $f : [0, \infty) \to \mathbb{R}$ and $g : \mathbb{R} \to [0, \infty)$ be the functions from Example 1.15. Then f is not surjective, since $f([0, \infty)) = [0, \infty) \neq \mathbb{R}$, while g is not injective, since $g(-1) = 1 = g(1)$. From Proposition 1.19 we conclude that f and g are not invertible.

1.10   Images and preimages of sets

For a function $f : A \to B$, we have defined the image $f(A)$ in Definition 1.4. More generally, we can give a meaning to the expression $f(S)$ for all subsets S of the domain A.

Definition 1.21. Let $f : A \to B$.

(i) For all $S \subseteq A$, the *image* of S via f is the subset $f(S)$ of B whose elements are the images via f of the elements of S. In other words,

$$f(S) = \{f(x) : x \in S\}.$$

(ii) For all $T \subseteq B$, the *preimage* of T via f is the subset $f^{-1}(T)$ of A whose elements are the preimages via f of the elements of T. In other words,

$$f^{-1}(T) = \{x \in A : f(x) \in T\}.$$

The following proposition shows that "taking the preimage of a set" in general is not the inverse operation to "taking the image of a set", unless the function satisfies some additional properties. Its proof is left as an exercise.

Proposition 1.22. *Let $f : A \to B$.*

(i) *If $S \subseteq A$, then $S \subseteq f^{-1}(f(S))$.*

(ii) *f is injective if and only if $f^{-1}(f(S)) = S$ for all $S \subseteq A$.*

(iii) *If $T \subseteq B$, then $f(f^{-1}(T)) \subseteq T$.*

(iv) *f is surjective if and only if $f(f^{-1}(T)) = T$ for all $T \subseteq B$.*

2   Real-valued functions of a real variable

Definition 2.1. Let $f : A \to B$. If $B = \mathbb{R}$, then we say that f is a *real-valued function*. If $A \subseteq \mathbb{R}$, then we say that f is a *function of a real variable*.

In this series of lectures, we will be mainly concerned with real-valued functions of a real variable, that is, functions whose domain is a subset of $\mathbb{R}$ and whose codomain is $\mathbb{R}$.

As already mentioned, in this case the graph of the function is a subset of the plane $\mathbb{R}^2$, so to study such functions it is often convenient to examine their graphs. Many of the properties of functions discussed in Section 1.7 can be clearly described in terms of "geometric" properties of the graph. For example:

- A subset Γ of the plane $\mathbb{R} \times \mathbb{R}$ is the graph of a function if and only if every vertical line intersects Γ in at most one point.

- A real-valued function f of a real variable is injective if and only if every horizontal line intersects the graph Γ_f of f in at most one point.

Remark 2.2. It is not always so easy to sketch the graph of a function. For instance, the reader may consider how one could sketch the graph of the *Dirichlet function* $d : \mathbb{R} \to \mathbb{R}$, defined by the following rule:

$$d(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q}, \\ 0 & \text{if } x \notin \mathbb{Q}, \end{cases}$$

for all $x \in \mathbb{R}$. In the course of these lectures we will introduce several properties to rule out such "pathological examples" and consider classes of "well behaved" functions.

2.1   The Domain Convention

According to Definition 1.1, to define a function one needs to specify a domain, a codomain and a rule.

However, when working with real-valued functions of a real variable, it is common practice to define functions just by giving a formula, such as "the function $f(x) = x^2 + 3$", or "the function $x \mapsto \sqrt{1+x}$", thus expressing a rule, without explicitly specifying the domain (we always assume the codomain to be $\mathbb{R}$, consistently with Definition 2.1). In these cases, we adopt the following

---

Domain Convention: when not otherwise specified, the domain is the largest subset of $\mathbb{R}$ for which the given expression makes sense (also called the *natural domain*).

The Domain Convention is better illustrated with a few examples.

Example 2.3. According to the Domain Convention:

1. the domain of the function

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

is the set $\mathbb{R} \setminus \{2\}$, that is, the set of the $x \in \mathbb{R}$ such that $x - 2 \neq 0$ (a fraction is defined when the denominator is nonzero);

2. the domain of the function

$$g(x) = \sqrt{x + 2}$$

   is the set $[-2, \infty)$, that is, the set of the $x \in \mathbb{R}$ such that $x + 2 \geq 0$ (a square root is defined when its argument is nonnegative);

3. the domain of the function

$$h(x) = \sqrt[3]{x}$$

   is the whole $\mathbb{R}$ (the cube root of any real number is a well-defined real number);

4. more generally, if $n \in \mathbb{N}$, then the domain of the n-th root function

$$x \mapsto \sqrt[n]{x}$$

   is $[0, \infty)$ when n is even, while it is $\mathbb{R}$ when n is odd;

5. the domain of the *tangent function*

$$\tan x = \frac{\sin x}{\cos x}$$

   is the set $\mathbb{R} \setminus \{\pi/2 + k\pi : k \in \mathbb{Z}\}$, that is, the set of all $x \in \mathbb{R}$ such that $\cos x \neq 0$;

6. the domain of the *cotangent function*

$$\cot x = \frac{\cos x}{\sin x}$$

   is the set $\mathbb{R} \setminus \{k\pi : k \in \mathbb{Z}\}$, that is, the set of all $x \in \mathbb{R}$ such that $\sin x \neq 0$.

2.2　Real-valued inverse function

According to Definition 2.1, when we speak of a real-valued function, we require that its codomain is the whole real line $\mathbb{R}$. As a matter of fact, the analysis developed in these lectures could be applied more generally to functions $f : A \to B$ where $A, B \subseteq \mathbb{R}$. Indeed, given any such function $f : A \to B$, we can always construct a new function $\tilde{f} = \mathrm{id}_{\mathbb{R}} \circ f : A \to \mathbb{R}$, which has the same domain and the same graph as the original function f, but whose codomain is $\mathbb{R}$. Since we will be mainly interested in properties of functions that depend only on the graph, it is convenient not to bother with different choices of codomain and just assume that the codomain is $\mathbb{R}$.

There is one drawback in working only with functions whose codomain is $\mathbb{R}$ (rather than a subset of $\mathbb{R}$): if we use the definition of inverse function given in Section 1.9, then a real-valued function whose inverse is also a real-valued function must have domain, codomain and image all equal to $\mathbb{R}$. In order to relax this constraint, we need to slightly adjust the definition of inverse function given in Section 1.9.

Definition 2.4. Let $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$ for some $A, B \subseteq \mathbb{R}$. We say that g is the *real-valued inverse* of f if

$$f(A) = B, \qquad g(B) = A,$$

and moreover

$$f(x) = y \qquad \text{if and only if} \qquad g(y) = x$$

for all $x \in A$ and $y \in B$.

The relation between Definitions 2.4 and 1.13 is described in the following proposition, which says that "inverse" and "real-valued inverse" are essentially the same thing, if one forgets about codomains and only looks at graphs.

Proposition 2.5. *Let $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$ for some $A, B \subseteq \mathbb{R}$. Let $\hat{f} = \mathrm{id}_{f(A)} \circ f : A \to f(A)$ be the function with the same graph as f, but whose codomain is the image of f. Similarly, let $\hat{g} = \mathrm{id}_{g(B)} \circ g : B \to g(B)$ be the functions with the same graph as g, but whose codomain is the image of g. Then the following are equivalent:*

(i) g is the real-valued inverse of f (in the sense of Definition 2.4);

(ii) $\hat{g}$ is the inverse of $\hat{f}$ (in the sense of Definition 1.13).

From this result and the discussion in Section 1.9 one easily derives the following properties.

Corollary 2.6. *Let $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$ for some $A, B \subseteq \mathbb{R}$. Then the following hold.*

(i) f has a real-valued inverse if and only if f is injective.

(ii) If g is the real-valued inverse of f, then f is the real-valued inverse of g.

The proofs of Proposition 2.5 and Corollary 2.6 are left as an exercise to the interested reader.

From now on, with a slight abuse of language, when working with real-valued functions of a real variable, we will normally use the expression "inverse" in place of the more precise "real-valued inverse".

From Definition 2.4, it is not difficult to see that, if a function $f : A \to \mathbb{R}$ has an inverse $g : B \to \mathbb{R}$, then the graph of g is obtained by reflecting the graph of f along the line $\{(x, y) \in \mathbb{R}^2 : x = y\}$.

Example 2.7. Let $a \in (0, \infty)$. The *base-a exponential function* $\exp_a : \mathbb{R} \to \mathbb{R}$ is defined by

$$\exp_a(x) = a^x.$$

If $a = 1$, then $\exp_a$ is constant; indeed $\exp_1(x) = 1^x = 1$ for all $x \in \mathbb{R}$. If $a \neq 1$, instead, $\exp_a$ is injective and its image is $(0, \infty)$; in this case, its (real-valued) inverse is the *base-a logarithm function* $\log_a : (0, \infty) \to \mathbb{R}$, which satisfies

$$\log_a(a^x) = x \qquad \text{and} \qquad a^{\log_a y} = y$$

for all $x \in \mathbb{R}$ and $y \in (0, \infty)$.

2.3 Sum, difference, product and quotient of functions

We have already seen in Section 1.8 that, given two functions, under suitable conditions, we can compose them in order to produce another function. When we work with real-valued functions, we have other ways of combining existing functions to produce another function.

Definition 2.8. Let $A \subseteq \mathbb{R}$. Let $f : A \to \mathbb{R}$ and $g : A \to \mathbb{R}$. Let $\lambda \in \mathbb{R}$.

(i) The *sum* of f and g is the function $f + g : A \to \mathbb{R}$ defined by

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

for all $x \in A$.

(ii) The *difference* of f and g is the function $f - g : A \to \mathbb{R}$ defined by

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

for all $x \in A$.

(iii) The *product* of f and g is the function $f \cdot g : A \to \mathbb{R}$ defined by

$$(f \cdot g)(x) = f(x) \cdot g(x)$$

for all $x \in A$.

(iv) The *product* of λ and f is the function $\lambda f : A \to \mathbb{R}$ defined by

$$(\lambda f)(x) = \lambda f(x)$$

for all $x \in A$.

(v) Let $D = \{x \in A : g(x) \neq 0\}$. The *quotient* of f and g is the function $f/g : D \to \mathbb{R}$ defined by

$$(f/g)(x) = f(x)/g(x)$$

for all $x \in D$.

Example 2.9. Here are familiar examples of sums, products and quotients of functions.

1. A *monomial* is a function from $\mathbb{R}$ to $\mathbb{R}$ of the form

$$x \mapsto ax^n$$

for some $n \in \mathbb{N}_0$ and $a \in \mathbb{R}$. If $a \neq 0$, then n is called the degree of the monomial. The monomial $x \mapsto ax^n$ is the product of the number a and the function $x \mapsto x^n$; the latter is the n-fold product of the identity function $x \mapsto x$ with itself.

2. A *polynomial* is a sum of finitely many monomials. Namely, a polynomial is a function $f : \mathbb{R} \to \mathbb{R}$ of the form

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1} + a_n x^n$$

for some $n \in \mathbb{N}_0$ and some $a_0, a_1, \ldots, a_n \in \mathbb{R}$; if $a_n \neq 0$, then the number n is called the *degree* of the polynomial f and the real numbers $a_0, \ldots, a_n$ are called the *coefficients* of f.

3. A *rational function* is the quotient of two polynomials. In other words, a rational function is a real-valued function f of the form

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

for some $n, m \in \mathbb{N}_0$ and some $a_0, \ldots, a_n, b_0, \ldots, b_m \in \mathbb{R}$. Note that the domain of a rational function is the set of the real numbers x such that the denominator $b_0 + b_1 x + \cdots + b_m x^m$ does not vanish.

2.4   Elementary functions

In mathematics, an elementary function is a function of a single variable (real-valued function of a real variable in this module) that is defined as taking sums, products, and compositions of finitely many of the following basic functions:

- Constant functions: C;

- Power functions: x^a with $a \in \mathbb{R}$;

- Exponential functions: a^x with $a > 0$;

- Trigonometric functions: $\sin x$, $\cos x$, etc.

- Hyperbolic functions: $\sinh x$, $\cosh x$, etc.

- Inverse functions of the above (if exist), such as logarithms, inverse trigonometric functions, inverse hyperbolic functions.

Most functions we have seen are elementary functions. For example

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

Recall the *modulus function*, which is the function from $\mathbb{R}$ to $\mathbb{R}$ that associates to every $x \in \mathbb{R}$ its modulus (or absolute value) $|x|$, given by

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

The modulus function is an elementary function, since $|x| = \sqrt{x^2}$.

2.5   Absolute value, inequalities, sign analysis

Recall that in Section 1.7 we defined the *modulus* (or *absolute value*) $|x|$ of a real number x as

$$|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0. \end{cases} \tag{2.1}$$

From the definition, we immediately see that

$$|x| = |-x| \geq 0 \tag{2.2}$$

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

The next statement lists fundamental properties of the modulus function in relation to the operations of addition and multiplication on $\mathbb{R}$.

Proposition 2.10. *For all $x, y \in \mathbb{R}$,*

$$x \le |x|, \tag{2.3}$$
$$|x + y| \le |x| + |y|, \tag{2.4}$$
$$||x| - |y|| \le |x - y|, \tag{2.5}$$
$$|xy| = |x| \, |y|. \tag{2.6}$$

Proof. To prove (2.3), we distinguish two cases. If $x \ge 0$, then $|x| = x$ by definition, and we are done. If instead $x < 0$, then $|x| \ge 0 > x$ by (2.2), and again (2.3) is satisfied.

Let us now prove (2.4). We distinguish two cases. If $x + y \ge 0$, then

$$|x + y| = x + y \le |x| + |y|,$$

where we used that $x \le |x|$ and $y \le |y|$ by (2.3). If instead $x + y < 0$, then

$$|x + y| = -(x + y) = (-x) + (-y) \le |-x| + |-y| = |x| + |y|,$$

where we used that $|-x| = |x|$ and $|-y| = |y|$ by (2.2).

The proofs of (2.5) and (2.6) are done similarly, and left as an exercise to the reader. $\square$

Remark 2.11. The inequality (2.4) above is also known as the *triangle inequality*, while (2.5) is sometimes called the *reverse triangle inequality*. Their names derive by the fact that, if $a, b \in \mathbb{R}$, then the expression

$$|a - b|$$

can be geometrically interpreted as the *distance* between the points a and b on the real line, and therefore the inequalities (2.4) and (2.5) admit a geometric interpretation.

Indeed, it is an elementary geometric fact that, for any given triangle in the plane, the sum of the length of two of its sides is not smaller than the length of the remaining side. This fact remains true even for "degenerate" triangles, whose vertices all lie on the same line. In particular, if we consider the degenerate triangle on the real line with vertices the three points $a, b, c \in \mathbb{R}$, then the side lengths of this triangle are $|a - b|$, $|a - c|$ and $|b - c|$, so the aforementioned relation between side lengths reads

$$|a - c| \le |a - b| + |b - c|. \tag{2.7}$$

This inequality is actually equivalent to the triangle inequality (2.4) above: indeed (2.4) follows from (2.7) by choosing $a = x$, $b = 0$, $c = -y$; conversely, (2.7) follows from (2.4) by taking $x = a - b$ and $y = b - c$

The identity (2.6) is related to elementary properties of the multiplication of real numbers, and especially how the sign of the product is related to the sign of the factors. These properties are crucial tools when solving inequalities, as the following examples show.

Example 2.12. Assume that we want to solve the inequality

$$x^3 \ge 2x - 1, \tag{2.8}$$

that is, determine the set of all the $x \in \mathbb{R}$ that satisfy this inequality. It is convenient to move all terms to the left-hand side, thus obtaining the equivalent inequality

$$x^3 - 2x + 1 \geq 0.$$

It is easily observed that $x = 1$ is a root of the cubic polynomial in the right-hand side; this can be used to factorise it and rewrite the inequality as

$$(x - 1)(x^2 + x - 1) \geq 0.$$

After a further factorisation we obtain

$$(x - 1)(x - a)(x - b) \geq 0$$

where $a = -(1 + \sqrt{5})/2$ and $b = (\sqrt{5} - 1)/2$. Since $a < b < 1$, we can use the following diagram to represent the sign of the factors in the left-hand side, and the corresponding sign of the product.

	$x < a$	$x = a$	$a < x < b$	$x = b$	$b < x < 1$	$x = 1$	$1 < x$
$x - a$	$-$	0	$+$	$+$	$+$	$+$	$+$
$x - b$	$-$	$-$	$-$	0	$+$	$+$	$+$
$x - 1$	$-$	$-$	$-$	$-$	$-$	0	$+$
$x^3 - 2x + 1$	$-$	0	$+$	0	$-$	0	$+$

From the above diagram, it is clear that $x^3 - 2x + 1 \geq 0$ if and only if $a \leq x \leq b$ or $x \geq 1$. In other words, the set of solutions $\{x \in \mathbb{R} : x^3 \geq 2x - 1\}$ of the inequality (2.8) can be written by using the interval notation as

$$[a, b] \cup [1, \infty).$$

Similar techniques can be applied when dealing with fractions, since $1/x$ has the same sign as x; however one should be careful to remember that $1/x$ is not defined when $x = 0$, so one needs to exclude from the solution set those values for which denominators vanish.

Example 2.13. Let us solve the inequality

$$\frac{x^2 - 3x + 2}{x^3 + x} \geq 0. \tag{2.9}$$

Here numerator and denominator factorise as

$$x^2 - 2x = (x - 1)(x - 2), \qquad x^3 + x = x(x^2 + 1),$$

and we note that the quadratic polynomial $x^2 + 1$ has negative discriminant, so has no real roots and is always strictly positive (since its leading coefficient is). This means that the denominator $x^3 + x$ vanishes if and only if $x = 0$, and this value must be excluded from the solution set. The following diagram allows us to analyse the sign of the left-hand side of (2.9).

	$x < 0$	$x = 0$	$0 < x < 1$	$x = 1$	$1 < x < 2$	$x = 2$	$2 < x$
$x - 2$	$-$	$-$	$-$	$-$	$-$	0	$+$
$x - 1$	$-$	$-$	$-$	0	$+$	$+$	$+$
x	$-$	0	$+$	$+$	$+$	$+$	$+$
$x^2 + 1$	$+$	$+$	$+$	$+$	$+$	$+$	$+$
$\frac{x^2-3x+2}{x^3+x}$	$-$	n.d.	$+$	0	$-$	0	$+$

("n.d." stands for "not defined".) Based on this analysis, we conclude that the inequality (2.9) is satisfied if and only if $0 < x \leq 1$ or $x \geq 2$, that is, if and only if x belongs to the set

$$(0,1] \cup [2, \infty).$$

2.6 Parity and periodicity

In addition to the properties of functions discussed in Section 1.7, there are a number of properties that are specific of real-valued functions of a real variable.

Definition 2.14 (parity)**.** Let $f : A \to \mathbb{R}$ for some $A \subseteq \mathbb{R}$.

(i) We say that f is an *even* function if, for all $x \in A$, we have $-x \in A$ and

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

(ii) We say that f is an *odd* function if, for all $x \in A$, we have $-x \in A$ and

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

These properties can also be characterised in terms of the function's graph; indeed one can easily check that:

- a function is even if and only if its graph is symmetric with respect to the vertical coordinate axis;

- a function is odd if and only if its graph is symmetric with respect to the origin.

Example 2.15. Here are some simple examples of even and odd functions.

1. The function $q : \mathbb{R} \to \mathbb{R}$ defined by

   $$q(x) = x^2$$

   is an even function; indeed

   $$q(-x) = (-x)^2 = x^2 = q(x)$$

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

2. The function $c : \mathbb{R} \to \mathbb{R}$ defined by

   $$c(x) = x^3$$

   is an odd function, since

   $$c(-x) = (-x)^3 = -x^3 = -c(x)$$

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

3. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

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

   is neither even nor odd. Indeed $f(1) = 2$, while $f(-1) = 0$, so $f(-1) \neq f(1)$ and $f(-1) \neq -f(1)$.

Definition 2.16 (periodicity). Let $f : A \to \mathbb{R}$ for some $A \subseteq \mathbb{R}$. We say that f is *periodic* if there exists a real number $\omega > 0$ such that, for all $x \in A$, we have $x + \omega \in A$ and

$$f(x + \omega) = f(x).$$

In this case we say that ω is a *period* of f, and we say that f is ω-periodic. If f has a minimum period ω, then ω is called the *fundamental period* of f.

It is easily seen that a function is ω-periodic if and only if its graph is invariant under a horizontal translation by a distance of ω.

Example 2.17. Here are some examples of periodic and nonperiodic functions.

1. The sine and cosine functions

   $$\sin : \mathbb{R} \to \mathbb{R}, \qquad \cos : \mathbb{R} \to \mathbb{R}$$

   are both periodic with fundamental period 2π.

2. The function $h : \mathbb{R} \to \mathbb{R}$ defined by

   $$h(x) = 2\cos(37x) - 2$$

   is periodic with fundamental period $2\pi/37$.

2.7   Monotonicity: increasing and decreasing functions

Definition 2.18 (monotonicity). Let $f : A \to \mathbb{R}$ for some $A \subseteq \mathbb{R}$.

(i) We say that f is *increasing* if, for all $x, x' \in A$,

$$x \le x' \qquad \implies \qquad f(x) \le f(x').$$

(ii) We say that f is *strictly increasing* if, for all $x, x' \in A$,

$$x < x' \qquad \implies \qquad f(x) < f(x').$$

(iii) We say that f is *decreasing* if, for all $x, x' \in A$,

$$x \le x' \qquad \implies \qquad f(x) \ge f(x').$$

(iv) We say that f is *strictly decreasing* if, for all $x, x' \in A$,

$$x < x' \qquad \implies \qquad f(x) > f(x').$$

(v) We say that f is *monotone* if f is increasing or f is decreasing.

Remark 2.19. [4] Being strictly increasing $[\![$ resp. strictly decreasing $]\!]$ is clearly a stronger property than being increasing $[\![$ resp. decreasing $]\!]$. Note also that strictly increasing and strictly decreasing functions are injective.

---

[4] Here and throughout, we use the parentheses $[\![\quad]\!]$ to indicate one or more alternatives. For example, the sentence

> The number x is positive $[\![$ resp. negative, null $]\!]$ if and only if $x > 0$ $[\![$ resp. $x < 0$, $x = 0$ $]\!]$.

is to be interpreted as

> The number x is positive if and only if $x > 0$, while x is negative if and only if $x < 0$, and finally x is null if and only if $x = 0$.

Example 2.20. Here are some examples of monotone and nonmonotone functions.

1. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

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

   is strictly increasing.

2. The function $g : (0, \infty) \to \mathbb{R}$ defined by

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

   is strictly decreasing.

3. The *Heaviside step function* $h : \mathbb{R} \to \mathbb{R}$, defined by

$$h(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } x \geq 0, \end{cases}$$

   is increasing, but not strictly increasing.

4. The sine and cosine functions are neither increasing nor decreasing, so they are not monotone.

The last example shows that not every function is monotone. However a function may still have a monotone behaviour when restricted to a subset of its domain. This idea is captured by the following definition.

Definition 2.21. Let $f : A \to \mathbb{R}$ for some $A \subseteq \mathbb{R}$. Let $S \subseteq A$. The function f is said to be *increasing* on S ⟦resp. *strictly increasing* on S, *decreasing* on S, *strictly decreasing* on S⟧ if the restriction $f|_S$ is increasing ⟦resp. strictly increasing, decreasing, strictly decreasing⟧.

Example 2.22. For all $k \in \mathbb{Z}$, the sine function is strictly increasing on the interval $[-\pi/2 + 2k\pi, \pi/2 + 2k\pi]$ and strictly decreasing on the interval $[\pi/2 + 2k\pi, 3\pi/2 + 2k\pi]$.

Example 2.23. Periodic functions are certainly not injective, so they have no inverse (see Corollary 2.6). However, as shown in the previous example, it may be possible to restrict them to subsets of their domain where they are injective, and we can then consider the inverses (in the sense of Section 2.2) of such restrictions. This is what is usually done with a number of trigonometric functions.

1. The sine function $\sin : \mathbb{R} \to \mathbb{R}$ is strictly increasing on $[-\pi/2, \pi/2]$, and $\sin([-\pi/2, \pi/2]) = [-1, 1]$. The inverse of the restriction $\sin|_{[-\pi/2, \pi/2]}$ is the *arcsine function*

$$\arcsin : [-1, 1] \to \mathbb{R},$$

   which is also strictly increasing, and whose image is $[-\pi/2, \pi/2]$.

2. The cosine function $\cos : \mathbb{R} \to \mathbb{R}$ is strictly decreasing on $[0, \pi]$, and $\cos([0, \pi]) = [-1, 1]$. The inverse of the restriction $\cos|_{[0, \pi]}$ is the *arccosine function*

$$\arccos : [-1, 1] \to \mathbb{R},$$

   which is also strictly decreasing, and whose image is $[0, \pi]$.

3. Let $C = \{x \in \mathbb{R} : \cos x \neq 0\}$. The tangent function $\tan : C \to \mathbb{R}$ is strictly increasing on $(-\pi/2, \pi/2)$, and $\tan((-\pi/2, \pi/2)) = \mathbb{R}$. The inverse of the restriction $\tan|_{(-\pi/2, \pi/2)}$ is the *arctangent function*

$$\arctan : \mathbb{R} \to \mathbb{R},$$

which is also strictly increasing, and whose image is $(-\pi/2, \pi/2)$.

4. Let $S = \{x \in \mathbb{R} : \sin x \neq 0\}$. The cotangent function $\cot : S \to \mathbb{R}$ is strictly decreasing on $(0, \pi)$, and $\cot((0, \pi)) = \mathbb{R}$. The inverse of the restriction $\cot|_{(0, \pi)}$ is the *arccotangent function*

$$\operatorname{arccot} : \mathbb{R} \to \mathbb{R},$$

which is also strictly decreasing, and whose image is $(0, \pi)$.

Remark 2.24. In place of arcsin, arccos, arctan, several texts use the notation $\sin^{-1}$, $\cos^{-1}$, $\tan^{-1}$. While commonly used, the latter notation may be confusing, since the sine, cosine and tangent functions are <u>not</u> invertible: as explained above, before taking the inverse, one must restrict such functions to suitable intervals. Note, for example, that

$$\arcsin(\sin(3\pi/4)) = \pi/4, \qquad \arcsin(\sin(7\pi/4)) = -\pi/4;$$

this shows that the composition $\arcsin \circ \sin$ is not the identity function.

2.8 Boundedness: upper and lower bounds, maximum and minimum, supremum and infimum

We have already introduced bounded and unbounded intervals in the real line $\mathbb{R}$. Here we extend the definition of "bounded" and "unbounded" so to apply to any subset of $\mathbb{R}$.

Definition 2.25. Let $A \subseteq \mathbb{R}$.

(i) Any real number $b \in \mathbb{R}$ such that

$$x \leq b$$

for all $x \in A$ is called an *upper bound* of A.

(ii) The set A is said to be *bounded above* if A has an upper bound, and *unbounded above* otherwise.

(iii) Any real number $b \in \mathbb{R}$ such that

$$b \leq x$$

for all $x \in A$ is called a *lower bound* of A.

(iv) The set A is said to be *bounded below* if A has a lower bound, and *unbounded below* otherwise.

(v) The set A is said to be *bounded* if A is both bounded above and bounded below, and *unbounded* otherwise.

Example 2.26. Here some examples of bounded and unbounded sets.

1. The set $\mathbb{N}_0$ of nonnegative integers is bounded below: indeed $x \geq 0$ for all $x \in \mathbb{N}_0$, so any real number $b \leq 0$ is a lower bound of $\mathbb{N}_0$. However $\mathbb{N}_0$ is <u>not</u> bounded above, because there is no real number x which is larger than every positive integer (this is known as the *Archimedean property* of real numbers). Consequently $\mathbb{N}_0$ is unbounded.

2. The interval $(-3, 17]$ is bounded. Any real number $b \geq 17$ is an upper bound of $(-3, 17]$, and any number $b \leq -3$ is a lower bound of $(-3, 17]$.

3. The half-line $(-\infty, 3)$ is bounded above: indeed any real number $b \geq 3$ is an upper bound of $(-\infty, 3)$. However $(-\infty, 3)$ has no lower bound in $\mathbb{R}$, so $(-\infty, 3)$ is not bounded below and consequently $(-\infty, 3)$ is unbounded.

It is important not to confuse the notion of *upper bound* and *lower bound* with that of *maximum* and *minimum* of a set.

Definition 2.27. Let $A \subseteq \mathbb{R}$.

(i) An element $x \in A$ such that $y \leq x$ for all $y \in A$ is called *maximum* of A and denoted by $\max A$.

(ii) An element $x \in A$ such that $x \leq y$ for all $y \in A$ is called *minimum* of A and denoted by $\min A$.

Remark 2.28.   1. A subset $A \subseteq \mathbb{R}$ has at most one maximum and at most one minimum.

2. Not every set has a maximum or a minimum. In particular, if A is unbounded above ⟦resp. unbounded below⟧, then A cannot have a maximum ⟦resp. minimum⟧. The open interval $(-1, 5)$ is bounded, but has neither maximum nor minimum.

The last remark shows that the notion of maximum and minimum for subsets of $\mathbb{R}$ has the drawback that many "interesting" sets (e.g., open or half-open intervals) do not have a maximum or a minimum, despite being bounded above or below. For this reason, it is convenient to introduce the broader notions of supremum and infimum of a set.

Definition 2.29. Let $A \subseteq \mathbb{R}$ be nonempty.

(i) If A is bounded above, then the minimum of the set of the upper bounds of A is called *supremum* (or *least upper bound*) of A and is denoted by $\sup A$; in other words,

$$\sup A = \min\{b \in \mathbb{R} : x \leq b \text{ for all } x \in A\}.$$

(ii) If A is bounded below, then the maximum of the set of the lower bounds of A is called *infimum* (or *greatest lower bound*) of A and is denoted by $\inf A$; in other words,

$$\inf A = \max\{b \in \mathbb{R} : b \leq x \text{ for all } x \in A\}.$$

Example 2.30.   1. If a set $A \subseteq \mathbb{R}$ has a maximum, then $\sup A = \max A$; similarly, if a set A has a minimum, then $\inf A = \min A$.

2. On the other hand, a set without maximum or minimum may have a supremum and an infimum. For example, as already observed, the interval $(-1, 5)$ has no maximum or minimum. However $\sup(-1, 5) = 5$ and $\inf(-1, 5) = -1$.

3. Let $B = \{1/n : n \in \mathbb{N}\}$. Then $\inf B = 0$; indeed 0 is a lower bound of B (since $1/n > 0$ for all $n \in \mathbb{N}$) and it is actually the greatest lower bound of B, since no $b > 0$ is a lower bound of B (if $b > 0$, then we can find $n \in \mathbb{N}$ sufficiently large that $n > 1/b$, and consequently $1/n < b$). On the other hand, B has no minimum (for any element $1/n \in B$ we can find a smaller element $1/n' \in B$: just take any natural number $n' > n$), that is, $\min B$ does not exist.

The above observations about the existence of supremum and infimum can be greatly generalised, as it is shown by the next statement.

Theorem 2.31 (Completeness Axiom)**.** *Let $A \subseteq \mathbb{R}$ be nonempty.*

(i) If A is bounded above, then A has a supremum.

(ii) If A is bounded below, then A has an infimum.

The above statement expresses a fundamental property of the real line $\mathbb{R}$, called *completeness*. We omit the proof, because this would require an in-depth discussion[5] of real numbers, which goes beyond the aim of this lecture series. In any case, the following example shows that the same property expressed by Theorem 2.31 does not hold if we replace $\mathbb{R}$ by $\mathbb{Q}$.

Example 2.32. Let A be the set of the nonnegative rational numbers whose square is less than 2, that is,

$$A = \{x \in \mathbb{Q} : x \geq 0 \text{ and } x^2 < 2\}.$$

It is not difficult to show that $\sup A = \sqrt{2}$, which is not a rational number. So, if we were to only work with rational numbers and forget about more general real numbers, then there would exist sets such as A that are bounded above but without a supremum. Roughly speaking, we can think that the set $\mathbb{R}$ of real numbers is obtained by "completing" the set $\mathbb{Q}$ of rational numbers in such a way that all bounded above sets have a supremum.

As already remarked, the convenience of working with inf and sup (as compared to max and min) is that we do not have to worry about their existence. Actually, to be precise, the above Definition 2.29 only deals with subsets of $\mathbb{R}$ that are nonempty and bounded above or below; it is convenient to give a meaning to inf and sup also when those conditions are not satisfied.

Definition 2.33. Let $A \subseteq \mathbb{R}$.

1. If A is unbounded above, then we say that $\sup A = \infty$.

2. If A is unbounded below, then we say that $\inf A = -\infty$.

3. If A is empty, then we say that $\sup A = -\infty$ and $\inf A = \infty$.

---

[5]Actually, in some treatments of real numbers, the property expressed by Theorem 2.31(i) is axiomatically assumed (whence the name "Completeness Axiom") and not proved.

Remark 2.34. Based on Definitions 2.29 and 2.33 (and Theorem 2.31), for <u>all</u> subsets $A \subseteq \mathbb{R}$, the supremum $\sup A$ and the infimum $\inf A$ are well-defined elements of the *extended real line* $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. Moreover, A is bounded above if and only if $\sup A < \infty$, and A is bounded below if and only if $\inf A > -\infty$.

We now want to transfer the above discussion about boundedness of sets to the case of functions.

Definition 2.35. Let $f : A \to \mathbb{R}$ for some $A \subseteq \mathbb{R}$.

(i) The function f is said to be *bounded* ⟦resp. *unbounded, bounded above, unbounded above, bounded below, unbounded below* ⟧ if its image $f(A)$ is bounded ⟦resp. unbounded, bounded above, unbounded above, bounded below, unbounded below ⟧.

(ii) Let $S \subseteq \mathbb{R}$. The *supremum* of f on S, denoted by $\sup_S f$ or $\sup_{x \in S} f(x)$, is defined by
$$\sup_S f = \sup_{x \in S} f(x) = \sup\{f(x) : x \in S\};$$

similarly, we define the *infimum*, the *maximum* and the *minimum* of f on S by

$$\inf_S f = \inf_{x \in S} f(x) = \inf\{f(x) : x \in S\},$$

$$\max_S f = \max_{x \in S} f(x) = \max\{f(x) : x \in S\},$$

$$\min_S f = \max_{x \in S} f(x) = \min\{f(x) : x \in S\}.$$

When $S = A$ is the whole domain of f, we also write $\sup f$, $\inf f$, $\max f$, $\min f$ in place of $\sup_A f$, $\inf_A f$, $\max_A f$, $\min_A f$, and speak of the *supremum, infimum, maximum, minimum* of f.

Example 2.36. Here are some examples of bounded and unbounded functions.

1. The sine and cosine functions are bounded: indeed the image of both functions is the interval $[-1, 1]$, which is bounded. Moreover,

$$\min_{x \in \mathbb{R}} \sin x = \min_{x \in \mathbb{R}} \cos x = -1, \qquad \max_{x \in \mathbb{R}} \sin x = \max_{x \in \mathbb{R}} \cos x = 1.$$

2. The tangent and cotangent functions are unbounded, and more precisely they are neither bounded above nor bounded below: indeed the image of both functions is the whole real line $\mathbb{R}$.

3. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

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

is bounded below, but it is not bounded above: indeed its image is the half-line $[0, \infty)$. Consequently f is unbounded. Moreover,

$$\min f = \inf f = 0, \qquad \sup f = \infty,$$

while $\max f$ does not exist.

4. The arctangent function $\arctan : \mathbb{R} \to \mathbb{R}$ has image $(-\pi/2, \pi/2)$ (see Example 2.23). So arctan is bounded and

$$\sup \arctan = \pi/2, \qquad \inf \arctan = -\pi/2,$$

but $\max \arctan$ and $\min \arctan$ do not exist.

The important difference between supremum and maximum (respectively, infimum and minimum) of a function f is that the former need not be in the range of f, i.e., need not be attained by f. The maximum and the minimum of a function, instead, are attained by the function (when they exist), so it makes sense to discuss where the function attains its maximum or its minimum.

Definition 2.37. Let $f : A \to \mathbb{R}$ for some $A \subseteq \mathbb{R}$.

(i) An element $x_0 \in A$ is called a *(global) maximum point* [[resp. *(global) minimum point*]] of the function f if $f(x_0) = \max f$ [[resp. $f(x_0) = \min f$]].

(ii) An element $x_0 \in A$ is called a *local maximum point* [[resp. *local minimum point*]] of the function f if there exists $\delta > 0$ such that $f(x_0) = \max_{A \cap (x_0 - \delta, x_0 + \delta)} f$ [[resp. $f(x_0) = \min_{A \cap (x_0 - \delta, x_0 + \delta)} f$]].

Example 2.38. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

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

Then clearly $f(x) \geq 0$ for all $x \in \mathbb{R}$ (the square of a real number is always nonnegative), and moreover $f(1) = f(-1) = 0$; this shows that $\min f = 0$, and that 1 and -1 are global minimum points of f. Actually, it is not difficult to show that $f(\mathbb{R}) = [0, \infty)$ (indeed, if $y \in [0, \infty)$, then it is enough to take $x = \sqrt{\sqrt{y} + 1}$ to see that $y = f(x) \in f(\mathbb{R})$), so $\sup f = \infty$ and f has no maximum. On the other hand, if $x \in [-1, 1]$, then $-1 \leq x^2 - 1 \leq 0$, and therefore $0 \leq f(x) \leq 1$; since $f(0) = 1$, this shows that $f(0) = \max_{(-1,1)} f$, and therefore 0 is a local maximum point of f.

3 Limits

3.1 Definitions

Let f be a real-valued function of a real variable. The purpose of this section is to give a meaning to the expression

$$\lim_{x \to a} f(x) = \ell \qquad (3.1)$$

which reads "the limit of $f(x)$ as x tends to a is equal to ℓ". Here a can be a real number or one of the symbols ∞ ("infinity") and $-\infty$ ("minus infinity"); moreover, the value ℓ of the limit in (3.1) can also be a real number or one of the symbols ∞ and $-\infty$. In other words, both a and ℓ are elements of the *extended real line* $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.

Sometimes, instead of (3.1), we write

$$f(x) \to \ell \text{ as } x \to a, \qquad (3.2)$$

which reads "$f(x)$ tends to ℓ as x tends to a"; we stress that (3.1) and (3.2) have exactly the same meaning, so in what follows we will only discuss the definition of (3.1), with the understanding that the alternative notation (3.2) is also allowed.

The intuitive meaning of (3.1) is that the value of $f(x)$ gets closer and closer to ℓ as the variable x gets closer and closer to a; more precisely, (3.1) says that $f(x)$ can be made arbitrarily close to ℓ, as long as x is taken sufficiently close to a. Note that the symbol x in (3.1) is a "dummy variable" and can be replaced by any other symbol.

For example, we may write

$$\lim_{x \to \infty} \frac{1}{x} = 0, \quad \lim_{t \to 0} \frac{1}{t^2} = \infty, \quad \lim_{y \to 1} \frac{y^2 - 1}{y - 1} = 2, \quad \lim_{\theta \to 0} \frac{\sin \theta}{\theta} = 1. \qquad (3.3)$$

Some of the above expressions may match your intuition, some other expressions may look more mysterious. What we want to do in this section is to give them a solid mathematical meaning.

As we will see, expressing "being close to a" in a precise way takes different forms according to whether a is a real number or one of the two symbols ∞ and $-\infty$. For this reason, we will split the definition of (3.1) into a number of cases, according to whether a and ℓ are real numbers or $\pm\infty$.

As the examples in (3.3) show, the expression $f(x)$ inside the limit (3.1) need not be defined for $x = a$. However, in order to make sense of the limit (3.1), we will require that there exist x "arbitrarily close" to a such that $f(x)$ is defined. This will be made precise in the definitions below.

3.1.1 Limit of $f(x)$ as x tends to ∞ or $-\infty$

Definition 3.1 (Limit at ∞). Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Assume that Ω is not bounded above.

(i) Let $\ell \in \mathbb{R}$. We say that

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

if, for all $\epsilon > 0$, there exists $N \in \mathbb{R}$ such that, for all $x \in \Omega$,

$$x > N \qquad \implies \qquad |f(x) - \ell| < \epsilon.$$

(ii) We say that
$$\lim_{x \to \infty} f(x) = \infty$$
if, for all $M > 0$, there exists $N \in \mathbb{R}$ such that, for all $x \in \Omega$,
$$x > N \qquad \Longrightarrow \qquad f(x) > M.$$

(iii) We say that
$$\lim_{x \to \infty} f(x) = -\infty$$
if, for all $M > 0$, there exists $N \in \mathbb{R}$ such that, for all $x \in \Omega$,
$$x > N \qquad \Longrightarrow \qquad f(x) < -M.$$

Definition 3.2 (Limit at $-\infty$). Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Assume that Ω is not bounded below.

(i) Let $\ell \in \mathbb{R}$. We say that
$$\lim_{x \to -\infty} f(x) = \ell$$
if, for all $\epsilon > 0$, there exists $N \in \mathbb{R}$ such that, for all $x \in \Omega$,
$$x < N \qquad \Longrightarrow \qquad |f(x) - \ell| < \epsilon.$$

(ii) We say that
$$\lim_{x \to -\infty} f(x) = \infty$$
if, for all $M > 0$, there exists $N \in \mathbb{R}$ such that, for all $x \in \Omega$,
$$x < N \qquad \Longrightarrow \qquad f(x) > M.$$

(iii) We say that
$$\lim_{x \to -\infty} f(x) = -\infty$$
if, for all $M > 0$, there exists $N \in \mathbb{R}$ such that, for all $x \in \Omega$,
$$x < N \qquad \Longrightarrow \qquad f(x) < -M.$$

Example 3.3. Let us prove, using the definition, that

$$\lim_{x \to \infty} \frac{1}{x} = 0. \tag{3.4}$$

Here Definition 3.1(i) applies, where $f(x) = \frac{1}{x}$, $\Omega = \mathbb{R} \setminus \{0\}$ (by the domain convention), and $\ell = 0$; note that Ω is indeed unbounded above. So we need to show that, for all $\epsilon > 0$, there exists $N \in \mathbb{R}$ such that, for all $x \in \mathbb{R} \setminus \{0\}$,

$$x > N \qquad \Longrightarrow \qquad \left| \frac{1}{x} - 0 \right| < \epsilon. \tag{3.5}$$

Let an arbitrary $\epsilon > 0$ be given. Note that

$$\left| \frac{1}{x} - 0 \right| = \left| \frac{1}{x} \right| = \frac{1}{|x|};$$

in order for this quantity to be less than ϵ, it is necessary and sufficient that

$$|x| > \frac{1}{\epsilon}.$$

In particular, if we take $M = \frac{1}{\epsilon}$, then M is a real number and, for all $x \in \mathbb{R} \setminus \{0\}$, if $x > M$ then $|\frac{1}{x}| < \frac{1}{M} = \epsilon$. In other words, for every given $\epsilon > 0$, we have found an $M \in \mathbb{R}$ such that (3.5) holds, and this by Definition 3.1(i) proves (3.4).

Example 3.4. Let us prove, using the definition, that

$$\lim_{x \to -\infty} (x^2 - x) = \infty. \tag{3.6}$$

Here Definition 3.2(ii) applies, where $f(x) = x^2 - x$ and $\Omega = \mathbb{R}$ (by the domain convention); note that Ω is indeed unbounded below. So we need to show that, for all $M > 0$, there exists $N \in \mathbb{R}$ such that, for all $x \in \mathbb{R}$,

$$x < N \qquad \implies \qquad x^2 - x > M. \tag{3.7}$$

Note that, for all $N > 0$, from $x < -N$ we deduce that $x^2 > N^2$ and $-x > N$, whence

$$x^2 - x > N^2 + N \geq N.$$

So, for any arbitrarily given $M > 0$, in order to ensure that (3.7) holds it is enough to choose $N = M$, because in this way from $x < -N$ we deduce $x^2 - x \geq N = M$. By Definition 3.2(ii), this proves (3.6).

3.1.2 Limit of $f(x)$ as x tends to a real number

Definition 3.5. Let $\Omega \subseteq \mathbb{R}$. We say that a real number $a \in \mathbb{R}$ is an *accumulation point* of Ω if, for all $\delta > 0$, there exists $x \in \Omega$ such that $0 < |x - a| < \delta$.

In other words, for a point $a \in \mathbb{R}$ to be an accumulation point of a subset $\Omega \subseteq \mathbb{R}$, we must be able to find elements of Ω which are arbitrarily close to a (but different from a itself). The main properties that we will use about accumulation points are included in the next statement.

Proposition 3.6. *The following hold.*

(i) *If $A \subseteq B \subseteq \mathbb{R}$, then any accumulation point of A is also an accumulation point of B.*

(ii) *Let $\Omega \subseteq \mathbb{R}$ and $a \in \mathbb{R}$. Then a is an accumulation point of Ω if and only if*
$$a = \inf(\Omega \cap (a, \infty)) \quad or \quad a = \sup(\Omega \cap (-\infty, a)).$$

(iii) *Let $b, c \in \mathbb{R}$ with $b < c$. Every real number $a \in [b, c]$ is an accumulation points of the open interval (b, c)*

(iv) *Let $b \in \mathbb{R}$. No real number $a < b$ is an accumulation point of $[b, \infty)$. Similarly, no real number $a > b$ is an accumulation point of $(-\infty, b]$.*

Proof. (i). Assume that a is an accumulation point of A. Then, for all $\delta > 0$, we can find $x \in A$ such that $0 < |x - a| < \delta$. However, since $A \subseteq B$, such point x will also be an element of B. Since $\delta > 0$ was arbitrary, by Definition 3.5 this proves that a is an accumulation point of B.

(ii). If $a = \inf(\Omega \cap (a, \infty))$, then, for any $\delta > 0$, $c + \delta$ is not a lower bound of $\Omega \cap (c, \infty)$, hence there exists $x \in \Omega \cap (c, \infty)$ such that $x < c + \delta$; in particular, $x \in \Omega$ and $a < x < a + \delta$, so $0 < |x - a| < \delta$; as $\delta > 0$ was arbitrary, this proves that a is an accumumation point of Ω. In a similar way one proves that, if $a = \sup(\Omega \cap (-\infty, a))$, then a is an accumulation point of Ω.

Finally, assume instead that a is neither the supremum of $\Omega \cap (-\infty, a)$ nor the infimum of $\Omega \cap (a, \infty)$. Note that a is a lower bound of $\Omega \cap (a, \infty)$; as a is not the greatest lower bound of $\Omega \cap (a, \infty)$, there exists a real number $b > a$ which is a lower bound of $\Omega \cap (a, \infty)$, and therefore $\Omega \cap (a, \infty) \subseteq [b, \infty)$. In a similar way, as a is not the least upper bound of $\Omega \cap (-\infty, a)$, one proves that there exists a real number $b' < a$ such that $\Omega \cap (-\infty, a) \subseteq (-\infty, b']$. Let now $\delta = \min\{b - a, a - b'\}$. Then there is no point $x \in \Omega$ such that $0 < |x - a| < \delta$: indeed, such point x would either satisfy $x > a$ or $x < a$; in the case $x > a$, we would have $x \in \Omega \cap (c, \infty) \subseteq [b, \infty)$, therefore $x \geq b$, and $|x - a| = x - a \geq b - a \geq \delta$, which contradicts the inequality $|x - a| < \delta$; a similar contradiction could be derived in the case $x < a$, as one would have $|x - a| = a - x \geq a - b' \geq \delta$. As $\delta > 0$, this proves that a is not an accumulation point of Ω.

(iii). Let $\Omega = (b, c)$. If $a \in [b, c]$, then $a > b$ or $a < c$; in the first case, $\sup(\Omega \cap (-\infty, a)) = \sup(b, a) = a$, while in the second $\inf(\Omega \cap (a, \infty)) = \inf(a, c) = a$, and in either case part (ii) gives that a is an accumulation point of Ω.

(iv). We first show that any $a < b$ is not an accumulation point of $[b, \infty)$. Let $\Omega = [b, \infty)$. Then $\Omega \cap (-\infty, a) = \emptyset$, so a is not its supremum. Moreover, $\inf(\Omega \cap (a, \infty)) = \inf[b, \infty) = b > a$, so a is not the infimum of $\Omega \cap (a, \infty)$. Hence part (ii) gives that a is not an accumulation point of Ω. The proof that any $a > b$ is not an accumulation point of $(-\infty, b]$ is done analogously. $\square$

Corollary 3.7. *Let Ω be any of the intervals (b, c), $[b, c]$, $(b, c]$, $[b, c)$, where $b, c \in \mathbb{R}$ and $b < c$. Then the accumulation points of Ω are exactly the elements of the closed interval $[b, c] = \Omega \cup \{b, c\}$.*

Proof. Since $\Omega \supseteq (b, c)$, the fact that every element of $[b, c]$ is an accumulation point of Ω follows from parts (iii) and (i) of Proposition 3.6. Conversely, since $\Omega \subseteq [b, \infty)$ and $\Omega \subseteq (-\infty, c]$, from parts (iv) and (i) of Proposition 3.6 it follows that no point $a < b$ and no point $a > c$ is an accumulation point of Ω. $\square$

Remark 3.8. In this series of lectures we will be mainly be concerned with subsets of $\mathbb{R}$ that are intervals or finite unions thereof, and Proposition 3.6 allows one to determine the accumulation points of such sets. However the definition of accumulation point applies to more complicated subsets of $\mathbb{R}$ too; for example:

1. No real number is an accumulation point of $\mathbb{Z}$.

2. Every real number is an accumulation point of $\mathbb{Q}$.

The proof of these statements is left as an exercise to the interested reader.

We can now introduce the definition of limit of $f(x)$ as x tends to a real number.

Definition 3.9 (Limit at a real number)**.** Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \mathbb{R}$ be an accumulation point of Ω.

(i) Let $\ell \in \mathbb{R}$. We say that
$$\lim_{x \to a} f(x) = \ell$$
if, for all $\epsilon > 0$, there exists $\delta > 0$ such that, for all $x \in \Omega$,
$$0 < |x - a| < \delta \qquad \Longrightarrow \qquad |f(x) - \ell| < \epsilon.$$

(ii) We say that
$$\lim_{x \to a} f(x) = \infty$$
if, for all $M > 0$, there exists $\delta > 0$ such that, for all $x \in \Omega$,
$$0 < |x - a| < \delta \qquad \Longrightarrow \qquad f(x) > M.$$

(iii) We say that
$$\lim_{x \to a} f(x) = -\infty$$
if, for all $M > 0$, there exists $\delta > 0$ such that, for all $x \in \Omega$,
$$0 < |x - a| < \delta \qquad \Longrightarrow \qquad f(x) < -M.$$

Example 3.10. Let us prove, using the definition, that
$$\lim_{x \to 1} \frac{x^2 - 1}{x - 1} = 2. \tag{3.8}$$

Here Definition 3.9(i) applies, where $f(x) = \frac{x^2-1}{x-1}$, $\Omega = \mathbb{R} \setminus \{1\}$ (by the domain convention), $\ell = 2$ and $a = 1$; note that indeed 1 is an accumulation point of Ω. So we need to show that, for all $\epsilon > 0$, there exists $\delta > 0$ such that, for all $x \in \mathbb{R} \setminus \{1\}$,
$$0 < |x - 1| < \delta \qquad \Longrightarrow \qquad \left| \frac{x^2 - 1}{x - 1} - 2 \right| < \epsilon. \tag{3.9}$$

Note that, for all $x \in \mathbb{R} \setminus \{1\}$,
$$\frac{x^2 - 1}{x - 1} - 2 = \frac{(x-1)(x+1)}{x - 1} - 2 = x + 1 - 2 = x - 1,$$

whence the condition $\left| \frac{x^2-1}{x-1} - 2 \right| < \epsilon$ is the same as $|x - 1| < \epsilon$. For any arbitrarily given $\epsilon > 0$, in order to ensure that the implication (3.9) holds, it is then enough to choose $\delta = \epsilon$. By Definition 3.9(i), this proves (3.8).

Example 3.11. Let us prove, using the definition, that
$$\lim_{x \to 0} \frac{1}{x^2} = \infty. \tag{3.10}$$

Here Definition 3.9(ii) applies, where $f(x) = \frac{1}{x^2}$, $\Omega = \mathbb{R} \setminus \{0\}$ (by the domain convention), and $a = 0$; note that indeed 0 is an accumulation point of Ω. So we need to show that, for all $M > 0$, there exists $\delta > 0$ such that, for all $x \in \mathbb{R} \setminus \{0\}$,
$$0 < |x - 0| < \delta \qquad \Longrightarrow \qquad \frac{1}{x^2} > M. \tag{3.11}$$

Note that, for all $x \in \mathbb{R} \setminus \{0\}$, $|x - 0| = |x|$. Moreover, for any given $M > 0$, the condition $\frac{1}{x^2} > M$ is satisfied if and only if $x^2 < \frac{1}{M}$, that is, $|x| < \sqrt{\frac{1}{M}}$. In order to ensure that the implication (3.11) holds, it is then enough to choose $\delta = \sqrt{\frac{1}{M}}$. By Definition 3.9(ii), this proves (3.10).

3.2   Existence and uniqueness of limits

With the above definitions, it may appear that we have given a clear meaning to the expression $\lim_{x \to a} f(x)$ for any function $f : \Omega \to \mathbb{R}$ and $a \in \overline{\mathbb{R}}$ (provided Ω contains values "arbitrarily close" to a). Actually, two problems remain open.

- The *existence* of limits: can we be sure that there exists an $\ell \in \overline{\mathbb{R}}$ such that $\lim_{x \to a} f(x) = \ell$? If not, the expression $\lim_{x \to a} f(x)$ remains undefined.

- The *uniqueness* of limits: can we be sure that at most one $\ell \in \overline{\mathbb{R}}$ satisfies $\lim_{x \to a} f(x) = \ell$? If more than one ℓ satisfies such condition, then the expression $\lim_{x \to a} f(x)$ is ill-defined.

As for the existence problem, the answer is in general negative, that is,

---

LIMITS NEED NOT EXIST.

Let us illustrate this with a couple of examples.

Example 3.12. Let us show that the limit

$$\lim_{x \to 0} \frac{x}{|x|}$$

does not exist. In other words, there is no $\ell \in \overline{\mathbb{R}}$ such that

$$\lim_{x \to 0} \frac{x}{|x|} = \ell. \tag{3.12}$$

Assume, for a contradiction, that (3.12) holds for some $\ell \in \overline{\mathbb{R}}$. Note first that

$$\frac{x}{|x|} = \begin{cases} 1 & \text{if } x > 0, \\ -1 & \text{if } x < 0, \end{cases}$$

and in particular, for all $x \in \mathbb{R} \setminus \{0\}$,

$$-1 \le \frac{x}{|x|} \le 1. \tag{3.13}$$

So (3.12) cannot hold for $\ell = \infty$: indeed Definition 3.9(ii) would require that, for any given $M > 0$, there exist $x \in \mathbb{R} \setminus \{0\}$ such that $\frac{x}{|x|} > M$, but due to (3.13) this is impossible if we take, say, $M = 1$. Similarly we rule out that (3.12) holds for $\ell = -\infty$.

It remains to consider the case where $\ell \in \mathbb{R}$. In this case, if (3.12) holds, by Definition 3.9(i) we deduce that $\left| \frac{x}{|x|} - \ell \right|$ can be made arbitrarily small (that

is, smaller than any given $\epsilon > 0$), provided $|x|$ is small enough (that is, smaller than a suitable $\delta > 0$). However,

$$\left| \frac{x}{|x|} - \ell \right| = \begin{cases} |1 - \ell| & \text{if } x > 0, \\ |-1 - \ell| & \text{if } x < 0. \end{cases}$$

Note also that, for all $\delta > 0$, the condition $0 < |x| < \delta$ is always satisfied by some $x > 0$ and also by some $x < 0$; in other words, irrespective of how small we take δ, the expression $\left| \frac{x}{|x|} - \ell \right|$ takes both the values $|1 - \ell|$ and $|-1 - \ell|$ on the range $0 < |x| < \delta$. In conclusion, in order to satisfy Definition 3.9(i), we would have to find $\ell \in \mathbb{R}$ so that both $|1 - \ell|$ and $|-1 - \ell|$ are smaller than any $\epsilon > 0$, but this would only be possible if $|1 - \ell| = 0$ and $|-1 - \ell| = 0$, that is $\ell = 1$ and simultaneously $\ell = -1$, which is a contradiction.

Example 3.13. Let us show that the limit

$$\lim_{x \to \infty} \sin x$$

does not exist. In other words, there is no $\ell \in \overline{\mathbb{R}}$ such that

$$\lim_{x \to \infty} \sin x = \ell. \tag{3.14}$$

We note first that

$$-1 \le \sin x \le 1 \tag{3.15}$$

for all $x \in \mathbb{R}$. Similarly as in Example 3.12, this allows us to rule out that (3.14) may hold for $\ell = \pm\infty$, and it remains to consider the case $\ell \in \mathbb{R}$.

Let then $\ell \in \mathbb{R}$ be fixed. We note now that there must exist an $x_0 \in \mathbb{R}$ such that $\sin x_0 \ne \ell$: indeed the image of the function $\sin$ is the whole interval $[-1, 1]$, which is infinite, so it is enough to choose x_0 such that $\sin x_0 \in [-1, 1] \setminus \{\ell\}$.

Take $\epsilon = |\sin x_0 - \ell|/2$, so that $\epsilon > 0$. If (3.14) held, by Definition 3.9(i) there would exist an $N \in \mathbb{R}$ such that, for all $x \in \mathbb{R}$,

$$x > N \qquad \Longrightarrow \qquad |\sin x - \ell| < \epsilon. \tag{3.16}$$

On the other hand, since $\sin$ is periodic of period 2π, for all $k \in \mathbb{N}$,

$$|\sin(x_0 + 2k\pi) - \ell| = |\sin(x_0) - \ell| = 2\epsilon > \epsilon,$$

and clearly we can choose $k \in \mathbb{N}$ sufficiently large that $x_0 + 2k\pi > N$ (just take $k \in \mathbb{N}$ such that $k > \frac{N - x_0}{2\pi}$); hence, by choosing $x = x_0 + 2k\pi$, we contradict (3.16). This proves that (3.14) cannot hold for any ℓ.

The above examples show that we must be careful in writing expressions such as $\lim_{x \to a} f(x)$ and treating them as they had a defined meaning, because in general they may well be undefined.

Luckily enough, the uniqueness problem instead has a positive answer.

Proposition 3.14 (Uniqueness of limits). *Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a, \ell_1, \ell_2 \in \overline{\mathbb{R}}$. If*

$$\lim_{x \to a} f(x) = \ell_1 \qquad and \qquad \lim_{x \to a} f(x) = \ell_2, \tag{3.17}$$

then

$$\ell_1 = \ell_2.$$

Proof. We only give the proof in the case $a, \ell_1, \ell_2 \in \mathbb{R}$. The remaining cases are left as an exercise to the interested reader.

We note first that, from the existence of the limits in (3.17) and Definition 3.9, it follows that a is an accumulation point of Ω.

Assume for a contradiction that $\ell_1 \neq \ell_2$. Then $|\ell_1 - \ell_2| > 0$, so we can find $\epsilon > 0$ such that

$$\epsilon < |\ell_1 - \ell_2|/2. \tag{3.18}$$

Since $\lim_{x \to a} f(x) = \ell_1$, by Definition 3.9 there exists $\delta_1 > 0$ such that, for all $x \in \Omega$,

$$0 < |x - a| < \delta_1 \qquad \Longrightarrow \qquad |f(x) - \ell_1| < \epsilon. \tag{3.19}$$

Similarly, since $\lim_{x \to a} f(x) = \ell_2$, there exists $\delta_2 > 0$ such that, for all $x \in \Omega$,

$$0 < |x - a| < \delta_2 \qquad \Longrightarrow \qquad |f(x) - \ell_2| < \epsilon. \tag{3.20}$$

Let $\delta = \min\{\delta_1, \delta_2\}$, so $\delta > 0$. Since a is an accumulation point of Ω, there exists $x_0 \in \Omega$ such that $0 < |x_0 - a| < \delta$. If we apply (3.19) and (3.20) with $x = x_0$, then we deduce that

$$|f(x_0) - \ell_1| < \epsilon \qquad \text{and, at the same time,} \qquad |f(x_0) - \ell_2| < \epsilon. \tag{3.21}$$

Then

$$\begin{aligned}
|\ell_1 - \ell_2| &= |\ell_1 - f(x_0) + f(x_0) - \ell_2| \\
&\leq |\ell_1 - f(x_0)| + |f(x_0) - \ell_2| \qquad \text{by the triangle inequality} \\
&< \epsilon + \epsilon \qquad\qquad\qquad\qquad\qquad \text{by (3.21)} \\
&= 2\epsilon \\
&< |\ell_1 - \ell_2| \qquad\qquad\qquad\qquad \text{by (3.18),}
\end{aligned}$$

which is a contradiction. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

3.3   One-sided limits of $f(x)$ as x tends to a real number

By examining Example 3.12, one notes that the obstruction to the existence of the limit in that case is the fact that the expression $\frac{x}{|x|}$ takes two different values, 1 and -1, for x arbitrarily close to 0, so there is no single value in $\mathbb{R}$ that the function approaches as x tends to 0. Things would change if we were allowed to decide "how" x tends to 0, that is, whether x tends to 0 from the left or from the right (that is, restricting our discussion to $x < 0$ or $x > 0$), because in that case a specific value would exist. This motivates the following definitions.

Definition 3.15 (Right and left limits)**.** Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \mathbb{R}$ and $\ell \in \overline{\mathbb{R}}$.

(i) We say that "the limit of $f(x)$ as x tends to a from above (or from the right) is equal to ℓ", and write

$$\lim_{x \to a^+} f(x) = \ell,$$

whenever

$$\lim_{x \to a} f_{a,+}(x) = \ell,$$

where $f_{a,+}$ is the restriction $f|_{\Omega \cap (a, \infty)}$ of f to $\Omega \cap (a, \infty)$.

(ii) We say that "the limit of $f(x)$ as x tends to a from below (or from the left) is equal to ℓ", and write

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

whenever

$$\lim_{x \to a} f_{a,-}(x) = \ell,$$

where $f_{a,-}$ is the restriction $f|_{\Omega \cap (-\infty, a)}$ of f to $\Omega \cap (-\infty, a)$.

Remark 3.16. By comparing Definitions 3.15 and 3.9, it is clear that, in order for the right limit $\lim_{x \to a^+} f(x)$ to be defined, it is necessary that a is an accumulation point of $\Omega \cap (a, \infty)$; similarly, in order for the left limit $\lim_{x \to a^-} f(x)$ to be defined, it is necessary that a is an accumulation point of $\Omega \cap (-\infty, a)$.

Example 3.17. As a continuation to the discussion of Example 3.12, we now show that

$$\lim_{x \to 0^+} \frac{x}{|x|} = 1, \qquad \lim_{x \to 0^-} \frac{x}{|x|} = -1.$$

To prove the first identity, we apply Definition 3.15(i), with $f(x) = \frac{x}{|x|}$, $\Omega = \mathbb{R} \setminus \{0\}$, $a = 0$, and $\ell = 1$. Since $\Omega \cap (0, \infty) = (0, \infty)$ in this case, we must show that $\lim_{x \to 0} f|_{(0,\infty)} = 1$. According to Definition 3.9(i), we are then required to show that, for all $\epsilon > 0$, there exists $\delta > 0$ such that, for all $x \in (0, \infty)$,

$$0 < |x| < \delta \qquad \Longrightarrow \qquad \left| \frac{x}{|x|} - 1 \right| < \epsilon. \tag{3.22}$$

On the other hand, if $x \in (0, \infty)$, then $x/|x| = 1$, so (3.22) trivially holds whatever choice of $\delta > 0$ we make. This proves that $\lim_{x \to 0^+} \frac{x}{|x|} = 1$; the other identity $\lim_{x \to 0^-} \frac{x}{|x|} = -1$ is proved in a similar way.

The previous Examples 3.12 and 3.17 indicate that the obstruction to the existence of the "two-sided limit" $\lim_{x \to 0} \frac{x}{|x|}$ is the fact that the two one-sided limits $\lim_{x \to 0^+} \frac{x}{|x|}$ and $\lim_{x \to 0^-} \frac{x}{|x|}$ have different values. This reflects a general fact, stated in the following proposition; its proof is omitted and left as an exercise to the reader.

Proposition 3.18. *Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \mathbb{R}$ be an accumulation point of both $\Omega \cap (-\infty, a)$ and $\Omega \cap (a, \infty)$. Let $\ell \in \bar{\mathbb{R}}$. Then*

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

if and only if

$$\lim_{x \to a^+} f(x) = \ell \qquad and \qquad \lim_{x \to a^-} f(x) = \ell.$$

Remark 3.19. One should notice that the mismatch of the values of the one-sided limits is not the only possible reason for the nonexistence of the two-sided limit. Indeed, either of the one-sided limits might not exist as well. As a matter of fact, the discussion about existence and uniqueness of limits of Section 3.2 applies to the one-sided limits as well, and in particular the analogue of Proposition 3.14 holds for one-sided limits $\lim_{x \to a^\pm} f(x)$ in place of $\lim_{x \to a} f(x)$.

3.4   Locality of limits

An important property of limits, as defined in the previous sections, is that the limit $\lim_{x \to a} f(x)$ only depends on the behaviour of $f(x)$ when x is "close to a" (except possibly for the value of $f(x)$ at $x = a$). In other words, the concept of limit has a *local* nature. This is made precise by the following statement, telling us that, in order to compute the limit $\lim_{x \to a} f(x)$, it is enough to consider the limit of the *restriction* of f to a set containing only the points that are "close to a" in a suitable sense (except for a itself).

Proposition 3.20 (Locality of limits). *Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \overline{\mathbb{R}}$ and $\Omega' \subseteq \Omega$. Assume that one of the following conditions holds.*

(a) *$a \in \mathbb{R}$ and $\Omega' = \Omega \cap (a - r, a + r) \setminus \{a\}$ for some $r > 0$.*

(b) *$a = \infty$ and $\Omega' = \Omega \cap (N, \infty)$ for some $N \in \mathbb{R}$.*

(c) *$a = -\infty$ and $\Omega' = \Omega \cap (-\infty, N)$ for some $N \in \mathbb{R}$.*

Then the limit

$$\lim_{x \to a} f(x) \tag{3.23}$$

exists if and only if the limit

$$\lim_{x \to a} f|_{\Omega'}(x) \tag{3.24}$$

exists, and the two limits are equal.

Proof. We only prove this in the case (a); the proofs in the other cases are similar and left as an exercise to the reader.

We first prove that, in this case,

a is an accumulation point of $\Omega' \iff a$ is an accumulation point of Ω.

Since $\Omega' \subseteq \Omega$, the implication $\Rightarrow$ follows immediately from Proposition 3.6(i); so we only need to prove the reverse implication $\Leftarrow$. Assume therefore that a is an accumulation point of Ω. Now, given any $\delta > 0$, let $\delta' = \min\{\delta, r\}$. Since $\delta' > 0$ and a is an accumulation point of Ω, there exists $x \in \Omega$ such that

$$0 < |x - a| < \delta'.$$

Consequently $0 < |x - a| < r$, that is, $x \in (a - r, a + r) \setminus \{a\}$, and therefore

$$x \in \Omega \cap (a - r, a + r) \setminus \{a\} = \Omega'; \tag{3.25}$$

moreover, from the above inequality we also deduce that $0 < |x - a| < \delta$. In other words, we have found a point $x \in \Omega'$ such that $0 < |x - a| < \delta$. Since we can do this for any $\delta > 0$, this proves that a is an accumulation point of Ω'.

Now, assume that the limit (3.23) exists and is equal to $\ell \in \overline{\mathbb{R}}$. If $\ell \in \mathbb{R}$, by Definition 3.9 we deduce that for all $\epsilon > 0$ there exists $\delta > 0$ such that, for all $x \in \Omega$,

$$0 < |x - a| < \delta \qquad \Longrightarrow \qquad |f(x) - \ell| < \epsilon.$$

Since $\Omega' \subseteq \Omega$, clearly the above implication holds in particular for all $x \in \Omega'$. On the other hand, by definition, $f(x) = f|_{\Omega'}(x)$ for all $x \in \Omega'$. Hence we have proved that for all $\epsilon > 0$ there exists $\delta > 0$ such that, for all $x \in \Omega'$,

$$0 < |x - a| < \delta \qquad \Longrightarrow \qquad |f|_{\Omega'}(x) - \ell| < \epsilon.$$

By Definition 3.9, this tells us that the limit (3.24) exists too, and is equal to ℓ. The cases where $\ell = \infty$ and $\ell = -\infty$ can be discussed analogously, by using the appropriate cases of Definition 3.9. In conclusion, we have proved that, if the limit (3.23) exists, then the limit (3.24) exists and they are equal.

We now prove the reverse implication. Assume that the limit (3.24) exists and is equal to $\ell \in \overline{\mathbb{R}}$. Again, we only discuss the case where $\ell \in \mathbb{R}$, the other cases being analogous. Let $\epsilon > 0$ be given. Then from Definition 3.9 applied to the limit (3.24) we deduce that there exists $\delta > 0$ such that, for all $x \in \Omega'$,

$$0 < |x - a| < \delta \qquad \Longrightarrow \qquad |f|_{\Omega'}(x) - \ell| < \epsilon. \tag{3.26}$$

Take now $\delta' = \min\{\delta, r\}$. Then, for all $x \in \Omega$, if $0 < |x - a| < \delta'$, by arguing as in the proof of (3.25) above we deduce that $x \in \Omega'$ (so $f(x) = f|_{\Omega'}(x)$) and $0 < |x - a| < \delta$; therefore we can apply the implication (3.26) to x and deduce that $|f(x) - \ell| < \epsilon$. To sum up, we have proved that, for all $x \in \Omega$,

$$0 < |x - a| < \delta' \qquad \Longrightarrow \qquad |f(x) - \ell| < \epsilon.$$

Since we can find such $\delta' > 0$ for any given $\epsilon > 0$, this proves by Definition 3.9 that the limit (3.23) exists and is equal to ℓ. □

Example 3.21. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 0 & \text{if } |x| < 1, \\ \cos(x^2 + \arctan x) & \text{otherwise.} \end{cases}$$

If we want to determine the limit

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

then, in light of Proposition 3.20, it is enough to study the restriction of f to any set of the form $\Omega' = \mathbb{R} \cap (-r, r) \setminus \{0\}$ for some $r > 0$. If we choose $r = 1$, then $\Omega' = (-1, 1) \setminus \{0\}$, and

$$f|_{\Omega}(x) = f(x) = 0$$

for all $x \in \Omega'$. In other words, our initial problem reduces to determining the limit as $x \to 0$ of the constant function 0, which is easily seen to be equal to 0 from the definition of limit.

3.5   The Algebra of Limits

By using the definition of limit, it is straightforward to verify the following statements about the limits of the *identity* function $x \mapsto x$, namely

$$\lim_{x \to a} x = a \qquad \forall a \in \overline{\mathbb{R}}, \tag{3.27}$$

and also those of any *constant* function $x \mapsto c$, where $c \in \mathbb{R}$, namely

$$\lim_{x \to a} c = c \qquad \forall a \in \overline{\mathbb{R}}. \tag{3.28}$$

However, determining the limit of more complicated expressions by directly applying the definition of limit can be challenging. In order to make the computation of limits easier and more systematic, we now introduce a number of rules that allow us to compute the limit of more complicated expressions starting from the knowledge of the limit of simpler ones, such as the "basic limits" in (3.27) and (3.28) above.

3.5.1   Limits and sums

We partially extend the sum operation of $\mathbb{R}$ to the extended real line $\overline{\mathbb{R}}$. Namely, we define

$$\infty + \infty = \infty,$$
$$-\infty + (-\infty) = -\infty$$

and, for all $\ell \in \mathbb{R}$,

$$\infty + \ell = \ell + \infty = \infty,$$
$$-\infty + \ell = \ell + (-\infty) = -\infty.$$

Note that

> the expressions $\infty + (-\infty)$ and $-\infty + \infty$ remain <u>undefined</u>.

Proposition 3.22. *Let $f, g : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \overline{\mathbb{R}}$. If*

$$\lim_{x \to a} f(x) = \ell_1 \qquad and \qquad \lim_{x \to a} g(x) = \ell_2$$

for some $\ell_1, \ell_2 \in \overline{\mathbb{R}}$, then

$$\lim_{x \to a} (f(x) + g(x)) = \ell_1 + \ell_2,$$

except for the cases where one of ℓ_1, ℓ_2 is ∞ and the other is $-\infty$.

Proof. We only discuss the case where $a, \ell_1, \ell_2 \in \mathbb{R}$. The proof of the other cases is left as an exercise to the reader.

Let $\epsilon > 0$ be given, and note that $\epsilon/2 > 0$ as well. Since $\lim_{x \to a} f(x) = \ell_1$ and $\lim_{x \to a} g(x) = \ell_2$, by Definition 3.9 we can find $\delta_1, \delta_2 > 0$ such that, for all $x \in \Omega$,

$$0 < |x - a| < \delta_1 \qquad \Longrightarrow \qquad |f(x) - \ell_1| < \epsilon/2$$

and

$$0 < |x - a| < \delta_2 \qquad \Longrightarrow \qquad |g(x) - \ell_2| < \epsilon/2.$$

Consequently, if we take $\delta = \min\{\delta_1, \delta_2\}$, then $\delta > 0$, and moreover, for all $x \in \Omega$, if $0 < |x - a| < \delta$, then

$$\begin{aligned}
|(f(x) + g(x)) - (\ell_1 + \ell_2)| &= |(f(x) - \ell_1) + (g(x) - \ell_2)| \\
&\leq |f(x) - \ell_1| + |g(x) - \ell_2| \\
&< \epsilon/2 + \epsilon/2 \\
&= \epsilon.
\end{aligned}$$

Since $\epsilon > 0$ was arbitrary, this shows, by Definition 3.9, that $\lim_{x \to a} (f(x) + g(x)) = \ell_1 + \ell_2$. $\qquad\square$

The above result allows us to deduce, in many cases, the limit of the sum $f + g$ of two functions f and g by knowing the limits of the two functions separately. An analogous statement holds for the limit of the difference $f - g$ of two functions.

As a matter of fact, the statement about the difference can be derived from Proposition 3.22 by noticing that $f - g = f + (-g)$; so, in order to treat differences, we only need to relate the limit of $-g$ to the limit of g. Since $-g = (-1) \cdot g$, this can be done via Proposition 3.23 below.

3.5.2   Limits and products

We partially extend the product operation of $\mathbb{R}$ to the extended real line $\overline{\mathbb{R}}$. Namely, we define

$$(-\infty) \cdot (-\infty) = (-\infty) \cdot (-\infty) = \infty,$$
$$(-\infty) \cdot \infty = \infty \cdot (-\infty) = -\infty.$$

Moreover, for all $\ell \in \mathbb{R}$, if $\ell > 0$, then we define

$$\infty \cdot \ell = \ell \cdot \infty = \infty,$$
$$(-\infty) \cdot \ell = \ell \cdot (-\infty) = -\infty,$$

while, if $\ell < 0$, then we define

$$\infty \cdot \ell = \ell \cdot \infty = -\infty,$$
$$(-\infty) \cdot \ell = \ell \cdot (-\infty) = \infty.$$

Note that

> the expressions $0 \cdot \infty$, $0 \cdot (-\infty)$, $\infty \cdot 0$ and $(-\infty) \cdot 0$ remain <u>undefined</u>.

Proposition 3.23. *Let $f, g : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \overline{\mathbb{R}}$. If*

$$\lim_{x \to a} f(x) = \ell_1 \qquad and \qquad \lim_{x \to a} g(x) = \ell_2$$

for some $\ell_1, \ell_2 \in \overline{\mathbb{R}}$, then

$$\lim_{x \to a} (f(x) \cdot g(x)) = \ell_1 \cdot \ell_2,$$

except for the case where one of ℓ_1, ℓ_2 is 0 and the other is $\pm\infty$.

Proof. We only give the proof in the case where $a, \ell_1, \ell_2 \in \mathbb{R}$. The proof of the other cases is left to the interested reader.

Let $\epsilon > 0$ be given. Take $\epsilon' = \min\{1, \epsilon/(1 + |\ell_1| + |\ell_2|)\}$, and note that

$$\epsilon' > 0, \quad \epsilon' \leq 1, \quad \epsilon'(1 + |\ell_1| + |\ell_2|) \leq \epsilon.$$

Since $\epsilon' > 0$, by our assumption and Definition 3.9 we deduce that there exist $\delta_1, \delta_2 > 0$ such that, for all $x \in \Omega$,

$$0 < |x - a| < \delta_1 \qquad \Longrightarrow \qquad |f(x) - \ell_1| < \epsilon'$$

and

$$0 < |x - a| < \delta_2 \qquad \Longrightarrow \qquad |g(x) - \ell_2| < \epsilon'.$$

Let $\delta = \min\{\delta_1, \delta_2\}$. Then $\delta > 0$. Moreover, for all $x \in \Omega$, if $0 < |x - a| < \delta$, then

$$\begin{aligned}
|f(x) \cdot g(x) - \ell_1 \cdot \ell_2| &= |f(x) \cdot g(x) - f(x) \cdot \ell_2 + f(x) \cdot \ell_2 - \ell_1 \cdot \ell_2| \\
&\leq |f(x) \cdot g(x) - f(x) \cdot \ell_2| + |f(x) \cdot \ell_2 - \ell_1 \cdot \ell_2| \\
&= |f(x)||g(x) - \ell_2| + |f(x) - \ell_1||\ell_2| \\
&\leq (|f(x) - \ell_1| + |\ell_1|)|g(x) - \ell_2| + |f(x) - \ell_1||\ell_2| \\
&< (\epsilon' + |\ell_1|)\epsilon' + \epsilon'|\ell_2| \\
&\leq (1 + |\ell_1| + |\ell_2|)\epsilon' \\
&\leq \epsilon.
\end{aligned}$$

Since $\epsilon > 0$ was arbitrary, by Definition 3.9 this shows that $\lim_{x \to a}(f(x) \cdot g(x)) = \ell_1 \cdot \ell_2$. $\qquad\square$

3.5.3 Limits and quotients

Proposition 3.23 allows us to determine, under certain assumptions, the limit of the product $f \cdot g$ of two functions from the limits of the two functions f and g separately. Similarly, we may be interested in the limit of the quotient f/g. Since $f/g = f \cdot (1/g)$, it is actually enough to discuss the relation between the limits of $1/g$ and g. This is done in the following statements.

Proposition 3.24. *Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \overline{\mathbb{R}}$. Then the following hold.*

(i) If

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

for some $\ell \in \mathbb{R} \setminus \{0\}$, then

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

(ii) If

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

then

$$\lim_{x \to a} \frac{1}{f(x)} = 0.$$

Proof. We only give the proof of part (i) in the case $a \in \mathbb{R}$. The proof of the remaining cases is left as an exercise.

We first notice that, since $\lim_{x \to a} f(x) = \ell \in \mathbb{R} \setminus \{0\}$, then, by applying Definition 3.9 with $\epsilon = |\ell|/2$, we deduce the existence of $r > 0$ such that, for all $x \in \Omega$ such that $0 < |x - a| < r$, we have

$$|f(x) - \ell| < \ell/2,$$

and in particular

$$|f(x)| \geq |\ell| - |f(x) - \ell| > \ell/2 > 0.$$

This shows that the domain of the function $1/f$, that is, the set $\tilde{\Omega} = \{x \in \Omega : f(x) \neq 0\}$ (see Definition 2.8), contains all the points $x \in \Omega$ such that $0 < |x - a| < r$, so a is an accumulation point of $\tilde{\Omega}$ too.

Let $\epsilon > 0$ be given. Define $\epsilon' = 2\epsilon/|\ell|^2$, and note that $\epsilon' > 0$ too. By our assumption and Definition 3.9, we deduce that there exists $\delta > 0$ such that, for all $x \in \Omega$,

$$0 < |x - a| < \delta \qquad \implies \qquad |f(x) - \ell| < \epsilon'.$$

Let $\delta' = \min\{\delta, r\}$. Then $\delta' > 0$. Moreover, for all $x \in \tilde{\Omega}$, if $0 < |x - a| < \delta$, then

$$\left| \frac{1}{f(x)} - \frac{1}{\ell} \right| = \left| \frac{\ell - f(x)}{\ell f(x)} \right| \leq \frac{|f(x) - \ell|}{|\ell||f(x)|} < \frac{\epsilon'}{|\ell| \cdot |\ell|/2} = \frac{2\epsilon'}{|\ell|^2} = \epsilon.$$

Since $\epsilon > 0$ was arbitrary, this shows, by Definition 3.9, that $\lim_{x \to a}(1/f(x)) = 1/\ell$. $\qquad\square$

Note that Proposition 3.24 does not cover the case where $\lim_{x \to a} f(x) = 0$. This is because such information is not enough to determine the limit $\lim_{x \to \infty} 1/f(x)$; as a matter of fact, in general such limit need not exist. The following statement allows us to treat the case where $\lim_{x \to a} f(x) = 0$ under an additional assumption on the sign of f. The proof is omitted and left as an exercise to the interested reader.

Proposition 3.25. *Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \overline{\mathbb{R}}$, and assume that*

$$f(x) > 0 \qquad [\![\ resp. \ f(x) < 0 \ ]\!] \tag{3.29}$$

for all $x \in \Omega \setminus \{a\}$. If

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

then

$$\lim_{x \to a} \frac{1}{f(x)} = \infty \qquad \left[\!\left[\ resp. \ \lim_{x \to a} \frac{1}{f(x)} = -\infty \ \right]\!\right]. \tag{3.30}$$

Remark 3.26. In light of Proposition 3.20, in order to apply the previous result, it is enough to check the sign condition (3.29) just for those points $x \in \Omega \setminus \{a\}$ that are "sufficiently close to a" (in the sense made precise by Proposition 3.20). For example, if $a \in \mathbb{R}$, then it will be enough to check that (3.29) holds for all $x \in \Omega \cap (a - r, a + r) \setminus \{a\}$, where $r > 0$ can be chosen arbitrarily.

3.6 Finite limits and absolute values. Null limits

An immediate consequence of the definition of limit (see Definitions 3.1, 3.2, 3.9) is the following criterion, that tells us that $f(x)$ tends to $\ell \in \mathbb{R}$ if and only if the distance $|f(x) - \ell|$ between $f(x)$ and ℓ tends to zero.[6] We omit the proof, which is left as an exercise to the interested reader.

Proposition 3.27. *Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \overline{\mathbb{R}}$ and $\ell \in \mathbb{R}$. Then*

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

if and only if

$$\lim_{x \to a} |f(x) - \ell| = 0.$$

In the case $\ell = 0$, the above criterion turns into a particularly useful tool to prove that a limit is zero, so we state it as a corollary.

Corollary 3.28 (Absolute Value Rule for Null Limits)**.** *Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \overline{\mathbb{R}}$. Then*

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

if and only if

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

---

[6]This idea can actually be used to *define* the limit of functions that are not real-valued, but take values in some other set where a notion of "distance" is defined (think, e.g., of a function having as codomain the set $\mathbb{R}^2$ of points of the plane). We will not pursue this in the course of these lectures.

3.7   Limits and inequalities. The Sandwich Theorem

As previously mentioned (see Section 1.5), we extend the order on the real line $\mathbb{R}$ to the extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ in the natural way; namely, we stipulate that

$$-\infty < x < \infty$$

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

Proposition 3.29. *Let $f, g : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \overline{\mathbb{R}}$. Assume that*

$$f(x) \leq g(x) \tag{3.31}$$

for all $x \in \Omega \setminus \{a\}$. If

$$\lim_{x \to a} f(x) = \ell_1 \qquad and \qquad \lim_{x \to a} g(x) = \ell_2 \tag{3.32}$$

for some $\ell_1, \ell_2 \in \overline{\mathbb{R}}$, then

$$\ell_1 \leq \ell_2.$$

Proof. We only give the proof in the case $a, \ell_1, \ell_2 \in \mathbb{R}$. The proof of the remaining cases is similar and is left as an exercise.

Let $\epsilon > 0$ be arbitrary. From our assumption (3.32), we deduce the existence of $\delta_1, \delta_2 > 0$ such that, for all $x \in \Omega$,

$$0 < |x - a| < \delta_1 \qquad \Longrightarrow \qquad |f(x) - \ell_1| < \epsilon \tag{3.33}$$

and

$$0 < |x - a| < \delta_2 \qquad \Longrightarrow \qquad |g(x) - \ell_2| < \epsilon. \tag{3.34}$$

Moreover, by our assumptions, there exists $r > 0$ such that (3.31) holds for all $x \in \Omega$ with $0 < |x - a| < r$.

Since $\delta_1, \delta_2, r > 0$ and a is an accumulation point of Ω, we can find $x \in \Omega$ such that $0 < |x - a| < \min\{\delta_1, \delta_2, r\}$. Consequently

$$
\begin{aligned}
\ell_1 - \epsilon &< f(x) & &\text{by (3.33)}, \\
&\leq g(x) & &\text{by (3.31)}, \\
&< \ell_2 + \epsilon & &\text{by (3.34)},
\end{aligned}
$$

whence

$$\ell_1 - \ell_2 < 2\epsilon. \tag{3.35}$$

Since $\epsilon > 0$ was arbitrary, from (3.35) we deduce that

$$\ell_1 - \ell_2 \leq 0,$$

that is, $\ell_1 \leq \ell_2$, as desired. □

It is important to notice that, in Proposition 3.29, we <u>assume</u> a priori the existence of both limits in (3.32). The next results show that, under stronger assumptions, it is possible to deduce the existence of the limit of one function from the existence of the limit of other functions.

Proposition 3.30 (Sandwich Theorem). *Let $f, g, h : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \overline{\mathbb{R}}$. Assume that*

$$f(x) \leq g(x) \leq h(x) \tag{3.36}$$

for all $x \in \Omega \setminus \{a\}$. If

$$\lim_{x \to a} f(x) = \ell = \lim_{x \to a} h(x) \tag{3.37}$$

for some $\ell \in \mathbb{R}$, then

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

Proof. Let $\epsilon > 0$ be arbitrary. By Definition 3.9, from our assumption (3.37) we deduce that there exist $\delta_1 > 0$ and $\delta_2 > 0$ such that, for all $x \in \Omega$,

$$0 < |x - a| < \delta_1 \qquad \implies \qquad |f(x) - \ell| < \epsilon \tag{3.38}$$

and

$$0 < |x - a| < \delta_2 \qquad \implies \qquad |h(x) - \ell| < \epsilon. \tag{3.39}$$

Let $r > 0$ be such that (3.36) is satisfied for all $x \in \Omega$ with $0 < |x - a| < r$.

Since $\delta_1, \delta_2, r > 0$, we have that $\delta = \min\{\delta_1, \delta_2, r\} > 0$ as well. Consequently, for all $x \in \Omega$ such that $0 < |x - a| < \delta$, from (3.36) and (3.38) we deduce that

$$g(x) \geq f(x) > \ell - \epsilon,$$

while from (3.36) and (3.38) we deduce that

$$g(x) \leq h(x) < \ell + \epsilon.$$

In conclusion, we have proved that, for all $x \in \Omega$,

$$0 < |x - a| < \delta \qquad \implies \qquad |g(x) - \ell| < \epsilon.$$

By Definition 3.9, this shows that $\lim_{x \to a} g(x) = \ell$. $\qquad\qquad\square$

Example 3.31. Let us prove, using the Sandwich Theorem, that

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

From (3.15) we deduce that, for all $x \in \mathbb{R} \setminus \{0\}$,

$$-|x| \leq x \sin(2^{1/x}) \leq |x|.$$

By the Sandwich Theorem, it is enough to prove that

$$\lim_{x \to 0} (-|x|) = 0 \qquad \text{and} \qquad \lim_{x \to 0} |x| = 0.$$

The validity of the latter limits is however straightforward from the definition: indeed, for all $\epsilon > 0$, if we take $\delta = \epsilon$, then, for all $x \in \mathbb{R}$, if $0 < |x| < \delta$, then $|-|x| - 0| = |x - 0| = |x| < \epsilon$ as well.

The Sandwich Theorem holds in the case $\ell = \pm\infty$ as well. In this case, only one bounding function is actually needed; this is stated precisely in the next proposition (whose proof is omitted and left as an exercise to the reader).

Proposition 3.32 (Sandwich Theorem for Infinite Limits)**.** *Let* $f, g : \Omega \to \mathbb{R}$ *for some* $\Omega \subseteq \mathbb{R}$. *Let* $a \in \overline{\mathbb{R}}$.

$$f(x) \leq g(x) \tag{3.40}$$

for all $x \in \Omega \setminus \{a\}$. *Then the following hold.*

(i) If

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

then

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

(ii) If

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

then

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

Example 3.33. Let us prove, using Proposition 3.32, that

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

Note that arctan is odd and increasing, and moreover $\arctan 1 = \pi/4$; hence, for all $x \in (-1, 1)$,

$$-\pi/4 \leq \arctan x \leq \pi/4.$$

In particular, for all $x \in \mathbb{R}$ such that $0 < |x| < 1$,

$$\frac{1 + \arctan x}{x^2} \geq \frac{1 - \pi/4}{x^2};$$

by Proposition 3.32, we are then reduced to showing that

$$\lim_{x \to 0} \frac{1 - \pi/4}{x^2} = \infty.$$

Since $1 - \pi/4 > 0$, the latter limit is easily proved directly from the definition (the proof is almost verbatim the one given for (3.10)).

Remark 3.34. In light of Proposition 3.20, in order to apply Propositions 3.29, 3.30 and 3.32, it is enough to check the respective conditions (3.31), (3.36) and (3.40) just for those $x \in \Omega \setminus \{a\}$ that are "sufficiently close to a".

4   Continuity

In this section, we define a class of particulary "well-behaved" functions, known as *continuous functions*. For these functions, computing limits (at any point of the domain) turns out to be particularly easy.

4.1   Definition

Definition 4.1. Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$.

  (i) Let $a \in \Omega$. We say that f is *continuous at the point a* if, for all $\epsilon > 0$, there exists $\delta > 0$ such that, for all $x \in \Omega$,

$$|x - a| < \delta \qquad \Longrightarrow \qquad |f(x) - f(a)| < \epsilon.$$

  (ii) If f is continuous at each point $a \in \Omega$, then we say that f is *continuous*.

  (iii) More generally, if $A \subseteq \Omega$, then we say that f is *continuous on A* if f is continuous at each point $a \in A$.

Example 4.2. Here are some examples of continuous functions.

1. Constant functions are continuous. Indeed, if $f : \mathbb{R} \to \mathbb{R}$ is constant, then

$$|f(x) - f(a)| = 0 < \epsilon$$

for all $\epsilon > 0$ and all $x, a \in \mathbb{R}$; so any choice of $\delta > 0$ trivially satisfies the definition.

2. The identity function $f : \mathbb{R} \to \mathbb{R}$, given by $f(x) = x$, is continuous. Indeed, for all $x, a \in \mathbb{R}$,
$$|f(x) - f(a)| = |x - a|;$$
therefore, for all $\epsilon > 0$, we can choose $\delta = \epsilon$ to satisfy the definition.

3. The modulus function $g : \mathbb{R} \to \mathbb{R}$, given by $g(x) = |x|$, is continuous. Indeed, for all $x, a \in \mathbb{R}$,

$$|g(x) - g(a)| = ||x| - |a|| \le |x - a|$$

by the triangle inequality; therefore, for all $\epsilon > 0$, we can choose $\delta = \epsilon$ to satisfy the definition.

4. More generally, one can prove that the following functions are continuous (on their respective domains):

  (a) polynomials and rational functions (see Example 2.9);

  (b) the modulus function $x \mapsto |x|$;

  (c) the nth root function $x \mapsto \sqrt[n]{x}$ (for all $n \in \mathbb{N}$, see Example 2.3);

  (d) trigonometric and inverse trigonometric functions:

$$\sin, \cos, \tan, \cot, \arcsin, \arccos, \arctan, \operatorname{arccot}$$

(see Examples 2.3 and 2.23);

(e) exponential and logarithm:

$$\exp_a, \ \log_a$$

(for all $a \in (0, \infty) \setminus \{1\}$, see Example 2.7).

We omit the proofs. However we note that the fact that polynomials and rational functions are continuous can be deduced from the first two examples and the Algebra of Continuous Functions (see Section 4.3 below).

Example 4.3. The sign function $s : \mathbb{R} \to \mathbb{R}$ (see (1.2)) is continuous at every point of $\mathbb{R} \setminus \{0\}$, but is not continuous at 0. Indeed, let $a \in \mathbb{R}$. If $a \neq 0$, then, if we take $\delta = |a|$, for all $x \in \mathbb{R}$ such that $|x - a| < \delta$, we have $s(x) = s(a)$ (why?) and therefore $|s(x) - s(a)| = 0 < \epsilon$ for any $\epsilon > 0$. If instead $a = 0$, then $|s(x) - s(a)| = 1$ for all $x \neq 0$ (why?); therefore, if we choose $\epsilon = 1/2$, it is not possible to find $\delta > 0$ so that $|s(x) - s(a)| < 1/2$ for all $x \in \mathbb{R}$ such that $|x| < \delta$ (indeed, the interval $(-\delta, \delta)$ contains nonzero numbers for any $\delta > 0$).

Example 4.4. The Dirichlet function $d : \mathbb{R} \to \mathbb{R}$ (see Remark 2.2) is not continuous at any $a \in \mathbb{R}$. Indeed, if $a \in \mathbb{Q}$, then $d(a) = 1$; however, if we take $\epsilon = 1/2$, then the inequality

$$|d(x) - d(a)| < 1/2$$

is only satisfied when $x \in \mathbb{Q}$ (if $x \notin \mathbb{Q}$, then $d(x) = 0$ and $|d(x) - d(a)| = 1$), and therefore there is no $\delta > 0$ such that the inequality is satisfied for all $x \in \mathbb{R}$ such that $|x - a| < \delta$ (indeed, in any interval of the form $(a - \delta, a + \delta)$ we can find irrational numbers). A similar argument shows that d is not continuous at any $a \in \mathbb{R} \setminus \mathbb{Q}$.

4.2 Continuity and limits

By comparing the definitions of continuity and limit (Definitions 3.9 and 4.1) one can see that the two notions are closely related; this is made esplicit in Proposition 4.6 below. However, one should note that, in order for the function $f : \Omega \to A$ to be continuous at $a \in \mathbb{R}$, the point a must be an element of the domain Ω of f (but a need not be an accumulation point of Ω); instead, in order for $\lim_{x \to a} f(x)$ to be defined, the point a must be an accumulation point of Ω (but a need not be an element of Ω). For this reason, the continuity of f in the case where $a \in \Omega$ but a is not an accumulation point of Ω must be treated separately; this is done in the following statement.

Proposition 4.5. *Lef $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \Omega$. If a is not an accumulation point of Ω, then f is continuous at a.*

Proof. Let $\epsilon > 0$ be given. Since a is not an accumulation point of Ω, there exists $\delta > 0$ such that no element $x \in \Omega$ satisfies $0 < |x - a| < \delta$. Consequently, for all $x \in \Omega$, if $|x - a| < \delta$, then it must be $|x - a| = 0$, that is, $x = a$, and therefore $|f(x) - f(a)| = |f(a) - f(a)| = 0 < \epsilon$. Since $\epsilon > 0$ was arbitrary, this proves that f is continuous at a. $\qquad\square$

The following result shows in particular that computing the limit of a continuous function at any point of its domain is particularly easy (that is, it is the same as evaluating the function at that point).

Proposition 4.6. *Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \Omega$ be an accumulation point of Ω. The following are equivalent:*

(i) f is continuous at a;

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

Proof. (i) $\Rightarrow$ (ii). If f is continuous at a, then, by Definition 4.1, for all $\epsilon > 0$ there exists $\delta > 0$ such that, for all $x \in \Omega$,

$$|x - a| < \delta \qquad \Longrightarrow \qquad |f(x) - f(a)| < \epsilon.$$

By Definition 3.9, this clearly implies that $\lim_{x \to a} f(x) = f(a)$.

(ii) $\Rightarrow$ (i). If $\lim_{x \to a} f(x) = f(a)$, then by Definition 3.9 we deduce that for all $\epsilon > 0$ there exists $\delta > 0$ such that, for all $x \in \Omega$,

$$0 < |x - a| < \delta \qquad \Longrightarrow \qquad |f(x) - f(a)| < \epsilon. \tag{4.1}$$

Note now that, if $|x - a| = 0$, then $x = a$ and therefore trivially $|f(x) - f(a)| = 0 < \epsilon$ for any $\epsilon > 0$. In other words, from (4.1) we actually deduce the formally stronger statement

$$|x - a| < \delta \qquad \Longrightarrow \qquad |f(x) - f(a)| < \epsilon.$$

By Definition 4.1, this implies that f is continuous at a. $\qquad\square$

Example 4.7. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x \sin \frac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

is continuous at 0. This can be seen by applying Proposition 4.6: indeed,

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

where the fact that $\lim_{x \to 0} x \sin \frac{1}{x} = 0$ can be justified by means of the Sandwich Theorem (compare Example 3.31).

4.3   The Algebra of Continuous Functions

Given the relation between limits and continuity (see Proposition 4.6), one may expect that the "Algebra of Limits" discussed in Section 3.5 has consequences for continuous functions. This is made precise in the following statement.

Proposition 4.8. *Let $f, g : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \Omega$ and assume that f and g are both continuous at a. Then the following hold.*

(i) The sum $f + g$ is continuous at a.

(ii) The product $f \cdot g$ is continuous at a.

(iii) If $f(a) \neq 0$, then $1/f$ is continuous at a.

Proof. If a is not an accumulation point of Ω, then by Proposition 4.5 the functions $f + g$, $f \cdot g$ and $1/f$ are trivially continuous at a.

Assume instead that a is an accumulation point of Ω. Then, by Proposition 4.6,

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

whence, by the Algebra of Limits,

$$\lim_{x \to a} (f(x) + g(x)) = f(a) + g(a), \qquad \lim_{x \to a} (f(x) \cdot g(x)) = f(a) \cdot g(a),$$

and (in the case $f(a) \neq 0$)

$$\lim_{x \to a} \frac{1}{f(x)} = \frac{1}{f(a)};$$

by Proposition 4.6 again, this implies that $f + g$, $f \cdot g$ and $1/f$ are continuous at a. $\qquad\square$

Example 4.9. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = \sin x + \cos^2 x$ is continuous. Indeed, the functions sin and cos are continuous (see Example 4.2), therefore f is continuous by the Algebra of Continuous Functions: more precisely, $x \mapsto \cos^2 x$ is continuous (it is the product of cos with itself) and finally f is continuous (it is the sum of sin and $x \mapsto \cos^2 x$).

4.4   Limits, continuity and composition

Proposition 4.10. *Let $f : \Omega \to \mathbb{R}$ and $g : A \to \mathbb{R}$ for some $\Omega, A \subseteq \mathbb{R}$ such that $f(\Omega) \subseteq A$. Let $a \in \overline{\mathbb{R}}$. Assume that*

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

for some $\ell \in \mathbb{R}$. Further, assume that $\ell \in A$ and g is continuous at ℓ. Then

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

Proof. We consider only the case $a \in \mathbb{R}$. The proof in the remaining cases is similar and is left as an exercise.

Let $\epsilon > 0$ be given. Since g is continuous at ℓ, by Definition 4.1 we can find $\delta > 0$ such that, for all $y \in A$,

$$|y - \ell| < \delta \quad \implies \quad |g(y) - g(\ell)| < \epsilon. \tag{4.2}$$

Now, since $\delta > 0$ and $\lim_{x \to a} f(x) = \ell$, by Definition 3.9 there exists $\sigma > 0$ such that, for all $x \in \Omega$,

$$0 < |x - a| < \sigma \quad \implies \quad |f(x) - \ell| < \delta. \tag{4.3}$$

In conclusion, for all $x \in \Omega$, if $0 < |x - a| < \sigma$, then by (4.3) $|f(x) - \ell| < \delta$, and therefore (by (4.2) applied with $y = f(x)$) $|g(f(x)) - g(\ell)| < \epsilon$. Since $\epsilon > 0$ was arbitrary, by Definition 3.9 this proves that $\lim_{x \to a} g(f(x)) = g(\ell)$. $\qquad\square$

As a consequence of the previous result, we may show that the composition of continuous functions is continuous.

Proposition 4.11. *Let $f : \Omega \to \mathbb{R}$ and $g : A \to \mathbb{R}$ for some $\Omega, A \subseteq \mathbb{R}$ such that $f(\Omega) \subseteq A$.*

(i) *Let $a \in \Omega$. If f is continuous at a and g is continuous at $f(a)$, then $g \circ f$ is continuous at a.*

(ii) *If f and g are continuous, then $g \circ f$ is continuous.*

Proof. It is enough to prove part (i); indeed part (ii) follows by applying part (i) to every $a \in \Omega$.

Let $a \in \Omega$, and assume that f is continuous at a and g is continuous at $f(a)$. If a is not an accumulation point of Ω, then $g \circ f : \Omega \to \mathbb{R}$ is trivially continuous at a by Propsition 4.5. Assume instead that a is an accumulation point of Ω. Then, by Proposition 4.6, since f is continuous at a,

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

therefore, since g is continuous at $f(a)$, by Proposition 4.10 applied with $\ell = f(a)$,

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

By Proposition 4.6 again, we conclude that $g \circ f$ is continuous at a. $\qquad\square$

Example 4.12. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = \arctan(\log(1 + x^2))$ is continuous. Indeed the polynomial $x \mapsto 1 + x^2$, the logarithm function $\log$ and the arctangent function $\arctan$ are all continuous functions (see Example 4.2), therefore their composition is continuous (by applying Proposition 4.11 twice).

Example 4.13. The function $f : \mathbb{R} \to \mathbb{R}$ defined in Example 4.7 is continuous (at all points of its domain $\mathbb{R}$). Indeed, from Example 4.7 we already know that f is continuous at 0. On the other hand, for $x \neq 0$,

$$f(x) = x \sin \frac{1}{x},$$

and the function $x \mapsto x \sin \frac{1}{x}$ is continuous (at every point of its domain $\mathbb{R} \setminus \{0\}$) by the Algebra of Continuous Functions and Proposition 4.11 (indeed $x \mapsto x \sin \frac{1}{x}$ is the product of the functions $x \mapsto x$ and $x \mapsto \sin \frac{1}{x}$, and the latter is the composition of $x \mapsto \sin x$ and $x \mapsto \frac{1}{x}$); so, for all $a \in \mathbb{R} \setminus \{0\}$,

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

which shows by Proposition 4.6 that f is continuous at a.

Another very useful consequence of Proposition 4.10 is the following "change of variable" result for limits, that shows that the implication

$$\left[\quad \lim_{x \to a} f(x) = b \quad \text{and} \quad \lim_{x \to b} g(x) = \ell \quad \right] \quad \Longrightarrow \quad \lim_{x \to a} g(f(x)) = \ell$$

holds under suitable conditions on the functions f and g.

Proposition 4.14 ("change of variable in limits")**.** *Let $f : \Omega \to \mathbb{R}$ and $g : A \to \mathbb{R}$ for some $\Omega, A \subseteq \mathbb{R}$ such that $f(\Omega) \subseteq A$. Let $a \in \overline{\mathbb{R}}$. Assume that*

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

for some $b \in \overline{\mathbb{R}}$, and that

$$f(x) \neq b \tag{4.4}$$

for all $x \in \Omega \setminus \{a\}$. If

$$\lim_{x \to b} g(x) = \ell$$

for some $\ell \in \overline{\mathbb{R}}$, then

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

Proof. We only give the proof in the case where $a, \ell \in \mathbb{R}$. Let $\tilde{g} : A \cup \{b\} \to \mathbb{R}$ be defined by

$$\tilde{g}(x) = \begin{cases} g(x) & \text{if } x \neq b, \\ \ell & \text{if } x = b. \end{cases}$$

Then, since the limit as $x \to b$ of a function does not depend on the value of the function at b (see Definition 3.9),

$$\lim_{x \to b} \tilde{g}(x) = \lim_{x \to b} g(x) = \ell = \tilde{g}(b),$$

and therefore, by Proposition 4.6, $\tilde{g}$ is continuous at b. In addition,

$$\tilde{g}(f(x)) = g(f(x))$$

for all $x \in \Omega \setminus \{a\}$, because $f(x) \neq b$; therefore,

$$\lim_{x \to a} g(f(x)) = \lim_{x \to a} \tilde{g}(f(x)) = \tilde{g}(b) = \ell,$$

where Proposition 4.10 was applied. $\square$

Remark 4.15. In the case $b = \pm\infty$, the assumption (4.4) is trivially satisfied, because f is real-valued. In the case $b \in \mathbb{R}$, however, the assumption (4.4) is crucial and cannot generally be omitted, unless the function g is continuous at b (in which case Proposition 4.10 applies).

Example 4.16. Let us determine

$$\lim_{x \to \infty} (\sqrt{x+4} - \sqrt{x+1}). \tag{4.5}$$

We first observe that both

$$\lim_{x \to \infty} (x+4) = \infty \quad \text{and} \quad \lim_{x \to \infty} (x+1) = \infty$$

by the Algebra of Continuous Functions; since

$$\lim_{y \to \infty} \sqrt{y} = \infty \tag{4.6}$$

(this can be easily checked by the definition of limit, see Definition 3.1 and the following examples), from Proposition 4.14 we conclude that

$$\lim_{x \to \infty} \sqrt{x+4} = \infty \quad \text{and} \quad \lim_{x \to \infty} \sqrt{x+1} = \infty.$$

Unfortunately this is not enough to determine (4.5) directly by means of the Algebra of Limits, because the case "∞ minus ∞" is excluded by the assumptions of Proposition 3.22. However we can rewrite the given expression as follows:

$$\sqrt{x+4} - \sqrt{x+1} = \frac{(x+4)-(x+1)}{\sqrt{x+4}+\sqrt{x+1}} = \frac{3}{\sqrt{x+4}+\sqrt{x+1}}$$

(in the first step we are multiplying and dividing the expression by $\sqrt{x+4} + \sqrt{x+1}$). From the Algebra of Limits, it is now immediate to deduce that

$$\lim_{x\to\infty} (\sqrt{x+4}+\sqrt{x+1}) = \infty,$$

and therefore

$$\lim_{x\to\infty} (\sqrt{x+4}-\sqrt{x+1}) = \lim_{x\to\infty} \frac{3}{\sqrt{x+4}+\sqrt{x+1}} = 0.$$

Example 4.17. Let us determine

$$\lim_{x\to\infty} (\sqrt{2x-5} - \sqrt{4x+3}). \tag{4.7}$$

Note that, for all sufficiently large $x \in \mathbb{R}$,

$$\sqrt{2x-5} - \sqrt{4x+3} = \sqrt{x}\left(\sqrt{2-5/x} - \sqrt{4+3/x}\right).$$

By the Algebra of Limits, we immediately have that

$$\lim_{x\to\infty} (2-5/x) = 2 \quad \text{and} \quad \lim_{x\to\infty} (4+3/x) = 4,$$

whence, by Proposition 4.10,

$$\lim_{x\to\infty} \sqrt{2-5/x} = \sqrt{2} \quad \text{and} \quad \lim_{x\to\infty} \sqrt{4+3/x} = 2,$$

because the function $y \mapsto \sqrt{y}$ is continuous; consequently, by the Algebra of Limits,

$$\lim_{x\to\infty} \left(\sqrt{2-5/x} - \sqrt{4+3/x}\right) = \sqrt{2} - 2 < 0.$$

Hence, by (4.6) and the Algebra of Limits, we conclude that
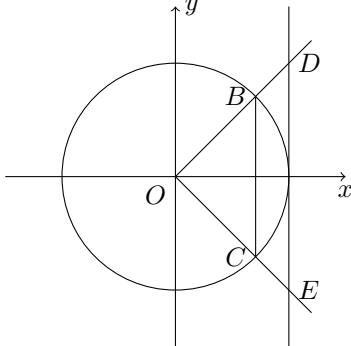
$$\lim_{x\to\infty} (\sqrt{2x-5} - \sqrt{4x+3}) = \lim_{x\to\infty} \sqrt{x}\left(\sqrt{2-5/x} - \sqrt{4+3/x}\right) = -\infty.$$

Remark 4.18. In the previous examples, the two similarly looking limits (4.5) and (4.7) are tackled through somewhat different strategies. The reader may want to check whether the strategy used in one case can be applied to the other case as well.

Theorem 4.19. *Elementary functions are continuous (on their domain).*

Proof. Recall the definition of elementary functions in Subsection 2.4. Then, the proof is obtained by collecting Example 4.2, Proposition 4.8, and Proposition 4.11. □

Figure 1: Construction described in the proof of Lemma 4.21.

4.5   Some notable limits

Proposition 4.6 and the results in Sections 4.3 and 4.4 are particulary useful tools to compute a wide variety of limits, when dealing with continuous functions. There are however certain limits that cannot be directly dealt with by using those tools. Some of these limits are of fundamental importance in what follows and are discussed in this section.

Proposition 4.20. *The following equalities hold.*

$$\lim_{x \to 0} \frac{\sin x}{x} = 1. \tag{4.8}$$

$$\lim_{x \to 0} \frac{1 - \cos x}{x^2} = \frac{1}{2}. \tag{4.9}$$

To prove Proposition 4.20, we need the following inequality.

Lemma 4.21. *For all* $\theta \in [0, \pi/2)$,

$$\sin \theta \leq \theta \leq \tan \theta.$$

Proof. Consider the unit circle centred at the origin $O = (0,0)$ of the xy-plane. The two half-lines emanating from O and making angles of $\pm\theta$ with the x-axis intersect the circle at the points B and C, while they intersect the line $x = 1$ at the points D and E (see Figure 1). From this construction we can see that the length of the straight line segment joining B and C is $2\sin\theta$, while the length of the arc of the circle between B and C is 2θ; since the straight line segment is the shortest line joining the two points, we deduce that $2\sin\theta \leq 2\theta$, that is,

$$\sin \theta \leq \theta.$$

To prove the remaining inequality, we note that the circular sector enclosed by the two radii OB and OC has area θ (its central angle is 2θ, so its area is $\frac{2\theta}{2\pi}$ times the area π of the unit circle), while the triangle ODE has area $\tan\theta$ (its base DE has length $2\tan\theta$ and the corresponding height is 1). Since the circular sector is contained in the triangle, the area of the former is not greater than the area of the latter, that is,

$$\theta \leq \tan\theta,$$

and we are done.   □

Proof of Proposition 4.20. From Lemma 4.21 we deduce the inequalities

$$\sin x \leq x \leq \tan x$$

for all $x \in [0, \pi/2)$. These inequalities imply that

$$\frac{\sin x}{x} \leq 1 \leq \frac{\tan x}{x} = \frac{1}{\cos x} \frac{\sin x}{x},$$

for all $x \in (0, \pi/2)$, that is,

$$\cos x \leq \frac{\sin x}{x} \leq 1. \tag{4.10}$$

The inequalities (4.10) actually hold for all $x \in (-\pi/2, \pi/2) \setminus \{0\}$; indeed, if $-\pi/2 < x < 0$, it is enough to apply (4.10) to $-x$ in place of x, and use the fact that cos is even, while sin is odd.

Since cos is continuous (see Example 4.2),

$$\lim_{x \to 0} \cos x = \cos 0 = 1.$$

From (4.10) and the Sandwich Theorem (Proposition 3.30), we deduce that

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

that is, (4.8).

As for (4.9), by standard trigonometric identities we deduce that

$$1 - \cos x = 2 \sin^2(x/2),$$

so

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

By the Algebra of Limits, (4.9) holds provided we can prove that

$$\lim_{x \to 0} \frac{\sin(x/2)}{x/2} = 1. \tag{4.11}$$

This however follows from (4.8) by a "change of variables" (Proposition 4.14), since

$$\lim_{x \to 0} (x/2) = 0$$

and $x/2 \neq 0$ for all $x \neq 0$. $\square$

In what follows, the letter e denotes Euler's constant (also known as "the base of natural logarithms"). We omit the proof of the following result[7].

Proposition 4.22. *The following equalities hold.*

$$\lim_{x \to \infty} \left(1 + \frac{1}{x} \right)^x = e. \tag{4.12}$$

$$\lim_{x \to -\infty} \left(1 + \frac{1}{x} \right)^x = e. \tag{4.13}$$

$$\lim_{x \to 0} (1 + x)^{1/x} = e. \tag{4.14}$$

[7]Proposition 4.22 extends a similar result discussed within "Sequences and Series".

In what follows, the base-e exponential and logarithm functions $\exp_e$ and $\log_e$ will be simply denoted by exp and log (and referred to as the "natural exponential" and "natural logarithm" functions)[8].

Proposition 4.23. *The following equalities hold.*

$$\lim_{x \to 0} \frac{\log(1+x)}{x} = 1. \tag{4.15}$$

$$\lim_{x \to 0} \frac{e^x - 1}{x} = 1. \tag{4.16}$$

Proof. Note that

$$\frac{\log(1+x)}{x} = \log((1+x)^{1/x}).$$

By Proposition 4.22, we know that

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

Since log is continuous, by Proposition 4.10 we deduce that

$$\lim_{x \to 0} \frac{\log(1+x)}{x} = \lim_{x \to 0} \log((1+x)^{1/x}) = \log e = 1.$$

As for (4.16), let $f : (-1, \infty) \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} \frac{\log(1+x)}{x} & \text{if } x \neq 0, \\ 1 & \text{if } x = 0. \end{cases}$$

In view of (4.15) and Proposition 4.6, we deduce that f is continuous. Moreover, for all $x \neq 0$,

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

in other words, proving (4.16) reduces to proving that

$$\lim_{x \to 0} \frac{1}{f(e^x - 1)} = 1,$$

which, by the Algebra of Limits, is equivalent to proving that

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

On the other hand, since $x \mapsto e^x - 1$ is continuous (by the Algebra of Continuous Functions), from Proposition 4.11 we deduce that $x \mapsto f(e^x - 1)$ is continuous as well, and therefore

$$\lim_{x \to 0} f(e^x - 1) = f(e^0 - 1) = f(0) = 1,$$

as desired.   □

---

[8]Please note that in some texts the symbol ln is used (in place of log) to denote the natural logarithm, while log is used to denote the base-10 logarithm $\log_{10}$.

We finally state a few important results about limits involving powers, exponential and logarithm functions at the endpoints of their domains. Indeed, by using the definition of limit it is not difficult to show that, for all $b \in (0, \infty)$,

$$\lim_{x \to \infty} x^b = \infty, \qquad\qquad \lim_{x \to 0^+} x^b = 0, \qquad\qquad (4.17)$$

$$\lim_{x \to \infty} e^x = \infty, \qquad\qquad \lim_{x \to -\infty} e^x = 0, \qquad\qquad (4.18)$$

$$\lim_{x \to \infty} \log x = \infty, \qquad\qquad \lim_{x \to 0^+} \log x = -\infty. \qquad\qquad (4.19)$$

However it is also important to be able to compare the "growth rates" of these functions to those of power laws. The proof of the following result (extending a result from "Sequences and Series") is omitted and left as an exercise.

Proposition 4.24. *For all $b > 0$, the following equalities hold.*

$$\lim_{x \to \infty} \frac{e^x}{x^b} = \infty. \qquad\qquad (4.20)$$

$$\lim_{x \to -\infty} |x|^b e^x = 0. \qquad\qquad (4.21)$$

$$\lim_{x \to \infty} \frac{\log x}{x^b} = 0. \qquad\qquad (4.22)$$

$$\lim_{x \to 0^+} x^b \log x = 0. \qquad\qquad (4.23)$$

Remark 4.25. Recall that, for all real numbers A, B such that $A > 0$,

$$A^B = \exp(B \log A). \qquad\qquad (4.24)$$

This identity may be useful in dealing with limits involving exponentiation.

Example 4.26. Let us determine

$$\lim_{x \to \infty} x^{-1/x}.$$

Note that, by (4.24), for all $x > 0$,

$$x^{-1/x} = \exp(-(\log x)/x).$$

By Proposition 4.24 we know that

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

hence, by the Algebra of Limits,

$$\lim_{x \to \infty} \left(-\frac{\log x}{x} \right) = 0$$

too; since exp is a continuous function, by Proposition 4.6 we deduce that

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

4.6   Consequences of continuity

We discuss here two fundamental properties of continuous functions.

The first statement corresponds to the naive idea that "the graph of a continuous function (whose domain is an interval) can be drawn without having to raise the pen from the paper". While this intuitive idea is not completely correct[9], the following statement is.

Theorem 4.27 (Intermediate Value Theorem). *Let $f : [a,b] \to \mathbb{R}$ be continuous, where $a, b \in \mathbb{R}$ and $a \le b$. If $f(a) \le f(b)$ ⟦ resp. $f(b) \le f(a)$ ⟧ then, for all $y \in [f(a), f(b)]$ ⟦ resp. $y \in [f(b), f(a)]$ ⟧, there exists $x \in [a,b]$ such that $f(x) = y$.*

Proof. Up to replacing f with $-f$ (which is also continuous by the Algebra of Continuous Functions), we may assume that $f(a) \le f(b)$. In addition, up to replacing f with $f - y$ (which is also continuous by the Algebra of Continuous Functions) we may assume that $y = 0$. In other words, we have a continuous function $f : [a,b] \to \mathbb{R}$ such that $f(a) \le 0 \le f(b)$, and we want to prove that there exists $x \in [a,b]$ such that $f(x) = 0$.

Clearly, if $f(a) = 0$ we can take $x = a$, and if $f(b) = 0$ we can take $x = b$, so it remains to treat the case where $f(a) < 0 < f(b)$.

Let $A = \{x \in [a,b] : f(x) < 0\}$. Note that $A \subseteq [a,b]$, so A is bounded; moreover $a \in A$, so A is nonempty. Therefore, by Theorem 2.31, the supremum $\bar{x} = \sup A$ is a real number. Moreover $a \le \bar{x} \le b$, because $a \in A$ and b is an upper bound of A, so $\bar{x} \in [a,b]$.

We now claim that $f(\bar{x}) = 0$ (that is, $\bar{x}$ is the element of $[a,b]$ that we are looking for). For a contradiction, assume instead that $f(\bar{x}) \ne 0$. Then it is either $f(\bar{x}) < 0$ or $f(\bar{x}) > 0$.

Assume first that $f(\bar{x}) > 0$ (so $\bar{x} > a$, because $f(a) < 0$), and take $\epsilon = f(\bar{x})$. Since f is continuous at $\bar{x}$, we can find $\delta > 0$ such that $|f(x) - f(\bar{x})| < \epsilon$ for all $x \in [a,b]$ such that $|x - \bar{x}| < \delta$; in particular, for all $x \in (\bar{x} - \delta, \bar{x}) \cap [a,b]$, we have $f(x) > f(\bar{x}) - \epsilon = 0$, and in particular no point of $(\bar{x} - \delta, \bar{x}) \cap [a,b]$ belongs to A. This implies that any element of $(\bar{x} - \delta, \bar{x}) \cap [a,b]$ is an upper bound of A; since $(\bar{x} - \delta, \bar{x}) \cap [a,b]$ is nonempty (because $a < \bar{x} \le b$), this contradicts the fact that $\bar{x}$ is the supremum (that is, the least upper bound) of A.

Assume instead that $f(\bar{x}) < 0$ (so $\bar{x} < b$, because $f(b) > 0$). By using the continuity of f at $\bar{x}$ and arguing as above, we can find δ such that $f(x) < 0$ for all $x \in [a,b]$ such that $|x - \bar{x}| < \delta$. In particular, the points of $(\bar{x}, \bar{x} + \delta) \cap [a,b]$ belong to A. Since $(\bar{x}, \bar{x} + \delta) \cap [a,b]$ is not empty (because $a \le \bar{x} < b$), this contradicts the fact that $\bar{x}$ is an upper bound of A.

Since assuming that $f(\bar{x}) \ne 0$ leads to a contradiction, we finally conclude that $f(\bar{x}) = 0$. □

The second statement is also known as the "Extreme Value Theorem", and should be compared with the discussion in Section 2.8.

---

[9]The interested reader may want, e.g., to try and plot the graph of the continuous function $f : [-1,1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x \sin \frac{1}{x} & \text{if } x \ne 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Theorem 4.28 (Boundedness Theorem). *Let $f : [a,b] \to \mathbb{R}$ be continuous, where $a, b \in \mathbb{R}$ and $a \le b$. Then f is bounded and attains its bounds. In other words, there exist points x_m and x_M in $[a,b]$ such that, for all $x \in [a,b]$,*

$$f(x_m) \le f(x) \le f(x_M).$$

The proof of this result, which we present below, is based in turn on a fundamental result about sequences of real numbers. Recall that a (real-valued) *sequence* is a function from $\mathbb{N}$ to $\mathbb{R}$, and that we also use the notation $(a_n)_n$ or $(a_n)_{n \in \mathbb{N}}$ to denote the sequence $n \mapsto a_n$; moreover, we say that the sequence $(a_n)_n$ is *convergent* if $\lim_{n \to \infty} a_n = \ell$ for some $\ell \in \mathbb{R}$. Finally, a *subsequence* of a sequence $(a_n)_n$ is any sequence of the form $(a_{n_k})_k$, where $k \mapsto n_k$ is a strictly increasing function from $\mathbb{N}$ to $\mathbb{N}$.

Theorem 4.29 (Bolzano–Weierstrass Theorem). *Every bounded sequence has a convergent subsequence.*

We omit the proof of the Bolzano–Weierstrass Theorem.[10] Based on this result, we can present a proof of the Boundedness Theorem.

Proof of Theorem 4.28. We want to show that f has a maximum and a minimum. We will show that f has a maximum; the proof for the minimum is similar and is omitted.

Let $\ell = \sup f$. At this stage ℓ could be either a real number or ∞ (we will see at the end that actually $\ell \in \mathbb{R}$). We now claim that there exists a sequence $(x_n)_{n \in \mathbb{N}}$ taking values in $[a,b]$ such that $\lim_{n \to \infty} f(x_n) = \ell$ (this is called an "extremising sequence" for f).

To construct the sequence, we distinguish between the cases where $\ell < \infty$ and $\ell = \infty$.

In the case where ℓ is a real number, for all $n \in \mathbb{N}$, we note that $\ell - 1/n < \ell$; so $\ell - 1/n$ cannot be an upper bound of the range $f([a,b])$ (recall that ℓ is the least upper bound of the range), and therefore there exists $x_n \in [a,b]$ such that $f(x_n) > \ell - 1/n$. Since $\ell = \sup f$, we also have $f(x_n) \le \ell$. Noting that $\lim_{n \to \infty}(\ell - 1/n) = \lim_{n \to \infty} \ell = \ell$, by the Sandwich Theorem (Proposition 3.30) we conclude that $\lim_{n \to \infty} f(x_n) = \ell$ too. So the sequence $(x_n)_{n \in \mathbb{N}}$ constructed in this way satisfies the required property.

If $\ell = \infty$, instead, for all $n \in \mathbb{N}$, the number n cannot be an upper bound of $f([a,b])$ (recall that $\sup f = \infty$ means that $f([a,b])$ is unbounded above), so there exists $x_n \in [a,b]$ such that $f(x_n) > n$. Again, by the Sandwich Theorem for Infinite Limits (Proposition 3.32), since $\lim_{n \to \infty} n = \infty$, we conclude that $\lim_{n \to \infty} f(x_n) = \infty = \ell$. So again the sequence $(x_n)_{n \in \mathbb{N}}$ constructed in this way satisfies the required property.

Since $x_n \in [a,b]$ for all $n \in \mathbb{N}$, the sequence $(x_n)_n$ is bounded; hence, by the Bolzano–Weierstrass Theorem (Theorem 4.29), $(x_n)_n$ has a convergent subsequence. Up to replacing $(x_n)_n$ with its subsequence, we may assume that the sequence $(x_n)_n$ itself converges. Let $\bar{x} = \lim_{n \to \infty} x_n$; by Proposition 3.29, since $a \le x_n \le b$, we also deduce that $a \le \bar{x} \le b$, that is, $\bar{x} \in [a,b]$.

Since $\bar{x} \in [a,b]$, the function f is continuous at $\bar{x}$, and moreover $\lim_{n \to \infty} x_n = \bar{x}$; therefore, by Proposition 4.10,

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

---

[10]Students taking the module "Sequences and Series" will discuss this result in greater depth.

On the other hand, by construction,

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

This implies (by the uniqueness of limits, see Proposition 3.14) that $f(\bar{x}) = \ell$. In particular, ℓ is indeed a real number and belongs to the range $f([a, b])$; since $\ell = \sup f$, we conclude that $\ell = \max f$ and indeed $f(x) \leq \ell = f(\bar{x})$ for all $x \in [a, b]$. $\qquad \square$

5 Differentiability

5.1 Definition

Definition 5.1. Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$.

(i) Let $a \in \Omega$. If

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \ell \tag{5.1}$$

for some $\ell \in \mathbb{R}$, then we say that f is *differentiable at the point a*; moreover, the value ℓ of the limit is called *derivative* of f at a, and is denoted by $f'(a)$.

(ii) If f is differentiable at each point $a \in \Omega$, then we just say that f is *differentiable*.

(iii) More generally, if $A \subseteq \Omega$, we say that f is *differentiable on A* if f is differentiable at every point $a \in A$.

(iv) Let $\tilde{\Omega}$ be the set of the points $a \in \Omega$ such that f is differentiable at a. The function which associates to each $a \in \tilde{\Omega}$ the derivative $f'(a)$ of f at a is called the *derivative* of f, and is denoted by f'.

Remark 5.2. The ratio

$$\frac{f(x) - f(a)}{x - a} \tag{5.2}$$

that appears in the definition of the derivative of f is called *Newton quotient* or *difference quotient* (at the point a). Using basic Cartesian geometry it is not difficult to check that this ratio is the slope of the straight line passing through the points $(a, f(a))$ and $(x, f(x))$ of the graph of f. Correspondingly, its limit as $x \to a$ (if it exists) is the slope of the straight line that is tangent to the graph of f at the point $(a, f(a))$. This gives us a *geometric interpretation* of differentiability: the differentiability of a function f at a point a is related to the existence of a <u>nonvertical</u> tangent line to the graph of f at the point $(a, f(a))$.

Remark 5.3. There is also a *physical interpretation* of the derivative. If the function f represents the position (as a function of time) of a point mass moving on a straight line, then the Newton quotient (5.2) is the *average velocity* of the mass between the time instants a and x (that is, the ratio between the space displacement $f(x) - f(a)$ and the time interval $x - a$). Correspondingly, its limit as $x \to 0$, that is, the derivative $f'(a)$, is the *instantaneous velocity* of the mass at time a.

Remark 5.4. Note that, if we set $h = x - a$ (so $x = a + h$), then the Newton quotient (5.2) can be rewritten as

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

the latter is called the Newton quotient of f at the point a with increment h. Consequently, using a change of variables (Proposition 4.14), we can equivalently rewrite the above definition of derivative as

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

Remark 5.5. There are many different ways in which the derivative f' of a function f can be denoted, such as

$$Df, \tag{5.3}$$

$$\frac{df}{dx}, \tag{5.4}$$

$$\frac{d}{dx}f, \tag{5.5}$$

where x is the name given to the variable of the function. The notation (5.4) (known as *Leibniz's notation*), which is formally written as a quotient, is reminiscent of the fact that the derivative is the limit of the Newton quotient

$$\frac{\Delta f}{\Delta x} = \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

as the increment Δx tends to 0. When using the notation (5.4), the derivative $f'(a)$ of f at a point a can be also written as

$$\left(\frac{df}{dx}\right)_{x=a} \qquad \text{or} \qquad \left.\frac{df}{dx}\right|_{x=a}.$$

Notation such as (5.4) or (5.5), where the variable x is explicitly named, is particularly convenient when the function f is directly expressed as a formula involving the variable x (see the discussion in Example 2.3). For example, if we consider the function $f(x) = x^2 + 2$, then we can write its derivative f' as

$$\frac{d(x^2 + 2)}{dx}, \qquad \text{or} \qquad \frac{d}{dx}(x^2 + 2).$$

The notation (5.4) is also convenient in contexts such as physics, where functions and variables are used to represent various physical quantities; so, for example, one can write that the velocity (denoted by v) is the derivative of the position (denoted by s) with respect to time (denoted by t) by writing

$$v = \frac{ds}{dt}.$$

Here are a few examples of differentiable and nondifferentiable functions.

Example 5.6. Polynomials, rational functions, trigonometric functions, exponential and logarithm functions are all differentiable functions (on their respective domains). We discuss these examples in detail in Section 5.5 below.

Example 5.7. The modulus function $x \mapsto |x|$ is not differentiable at 0. Indeed the Newton quotient at 0 is given by

$$\frac{|x| - |0|}{x - 0} = \frac{|x|}{x},$$

and we already know that the limit of this quotient as $x \to 0$ does not exist (see Example 3.12). On the other hand, it is not difficult to check that the function $x \mapsto |x|$ is differentiable at every point $a \in \mathbb{R} \setminus \{0\}$. Geometrically, this corresponds to the fact that the graph of the modulus function $x \mapsto |x|$ has a corner at $x = 0$: this corner prevents the existence of a tangent line to the graph at $(0,0)$ and is the geometric counterpart to the lack of differentiability at 0.

Example 5.8. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x \sin \frac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

is not differentiable at 0. Indeed, the Newton quotient at 0 is given by

$$\frac{f(x) - f(0)}{x - 0} = \sin \frac{1}{x},$$

and the limit of this expression as $x \to 0$ does not exist (compare Example 3.13).

Example 5.9. Let $f : [0, \infty) \to \mathbb{R}$ be defined by $f(x) = \sqrt{x}$. Then f is not differentiable at 0. Indeed, the Newton quotient at 0 is given by

$$\frac{f(x) - f(0)}{x - 0} = \frac{\sqrt{x}}{x} = \frac{1}{\sqrt{x}}$$

and

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

however, according to Definition 5.1, the limit of the Newton quotient must be finite for the function to be differentiable. Note that, in this case, the graph of the function f has a vertical tangent line at the point $(0, 0)$.

5.2 Differentiability and continuity

Proposition 5.10. *Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \Omega$ be an accumulation point of Ω. Then the following are equivalent.*

(i) f is differentiable at a.

(ii) There exists a function $g : \Omega \to \mathbb{R}$, which is continuous at a, such that

$$f(x) = f(a) + g(x)(x - a) \tag{5.6}$$

for all $x \in \Omega$.

In this case, moreover, $f'(a) = g(a)$.

Proof. (i) $\Rightarrow$ (ii). Assume that f is differentiable at a. Define $g : \Omega \to \mathbb{R}$ by

$$g(x) = \begin{cases} \frac{f(x) - f(a)}{x - a} & \text{if } x \neq a, \\ f'(a) & \text{if } x = a. \end{cases}$$

Then clearly (by Definition 5.1) $\lim_{x \to a} g(x) = f'(a) = g(a)$, and therefore g is continuous at a by Proposition 4.6. Moreover, from the above definition of g it is clear that the equality (5.6) is satisfied for all $x \in \Omega$.

(ii) $\Rightarrow$ (i). Assume that there exists a function $g : \Omega \to \mathbb{R}$ that satisfies (5.6) and is continuous at a. Then from (5.6) we deduce that, for all $x \in \Omega \setminus \{a\}$,

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

and therefore, since g is continuous at a, by Proposition 4.6,

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \lim_{x \to a} g(x) = g(a) \in \mathbb{R}.$$

By Definition 5.1, this means that f is differentiable at a and its derivative $f'(a)$ is equal to $g(a)$. $\qquad\square$

Corollary 5.11. *Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \Omega$ be an accumulation point of Ω. If f is differentiable at a, then f is continuous at a.*

Proof. This is an immediate consequence of Proposition 5.10 and the Algebra of Continuous Functions: indeed, from the identity (5.6), the continuity of g at a and the Algebra of Continuous Functions, one deduces the continuity of f at a. $\qquad\square$

Remark 5.12. It is important to remember that the implication in Corollary 5.11 in general cannot be reversed, that is, continuity does not imply differentiability. Indeed, all the functions considered in Examples 5.7, 5.8 and 5.9 are continuous at 0, but not differentiable at 0.

5.3   One-sided derivatives

Definition 5.13. Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$.

(i) Let $a \in \Omega$ be an accumulation point of $\Omega \cap (a, \infty)$. If

$$\lim_{x \to a^+} \frac{f(x) - f(a)}{x - a} = \ell$$

for some $\ell \in \mathbb{R}$, then f is said to be *right-differentiable at the point a*; moreover, the value ℓ is called the *right derivative of f at a* and is denoted by $f'_+(a)$.

(ii) Let $a \in \Omega$ be an accumulation point of $\Omega \cap (-\infty, a)$. If

$$\lim_{x \to a^-} \frac{f(x) - f(a)}{x - a} = \ell$$

for some $\ell \in \mathbb{R}$, then f is said to be *left-differentiable at the point a*; moreover, the value ℓ is called the *left derivative of f at a* and is denoted by $f'_-(a)$.

An immediate consequence of the definitions and Proposition 3.18 is the following result.

Proposition 5.14. *Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \Omega$ be an accumulation point of both $\Omega \cap (a, \infty)$ and $\Omega \cap (-\infty, a)$. Let $\ell \in \mathbb{R}$. Then the following are equivalent.*

(i) f is differentiable at a and $f'(a) = \ell$.

(ii) f is both left-differentiable and right-differentiable at a, and moreover $f'_+(a) = f'_-(a) = \ell$.

Example 5.15. Going back to the discussion in Example 5.7, it is easily seen that, if $f(x) = |x|$ for all $x \in \mathbb{R}$, then f is left-differentiable and right-differentiable at 0, but the left and right derivatives at 0 are different, that is,

$$f'_+(0) = 1 \quad \text{and} \quad f'_-(0) = -1,$$

and therefore f is not differentiable at 0.

Remark 5.16. Recalling the discussion in Remark 3.19, one should notice that the mismatch of the values of the one-sided derivatives is not the only possible reason for the lack of differentiability of a function. Indeed, either of the one-sided derivatives might not exist. For example, the function f of Example 5.8 is neither left-differentiable nor right-differentiable at 0.

5.4   Differentiation rules

Proposition 5.17. *Let $f, g : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \Omega$ be such that f and g are both differentiable at a. Then the following hold.*

(i) *The sum $f + g$ is differentiable at a, and*

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

(ii) *(Leibniz rule) The product fg is differentiable at a, and*

$$(fg)'(a) = f(a)\,g'(a) + f'(a)\,g(a).$$

(iii) *(quotient rule) If $g(a) \neq 0$, then f/g is differentiable at a, and*

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

Proof. By Proposition 5.10 there exist $f_0, g_0 : \Omega \to \mathbb{R}$ that are continuous at a and such that

$$f(x) = f(a) + f_0(x)\,(x - a), \qquad g(x) = g(a) + g_0(x)\,(x - a)$$

for all $x \in \Omega$, and moreover $f_0(a) = f'(a)$ and $g_0(a) = g'(a)$. Consequently

$$\begin{aligned}(f + g)(x) = f(x) + g(x) &= f(a) + f_0(x)\,(x - a) + g(a) + g_0(x)\,(x - a) \\ &= (f + g)(a) + (f_0 + g_0)(x)\,(x - a)\end{aligned}$$

for all $x \in \Omega$; since $f_0 + g_0$ is continuous at a (by the Algebra of Continuous Function), the last identity proves (by Proposition 5.10) that $f + g$ is differentiable at a, and moreover $(f + g)'(a) = (f_0 + g_0)(a) = f'(a) + g'(a)$.

Similarly,

$$\begin{aligned}(fg)(x) &= f(x)\,g(x) \\ &= (f(a) + f_0(x)\,(x - a))\,(g(a) + g_0(x)\,(x - a)) \\ &= f(a)\,g(a) + (f_0(x)g(a) + f(a)g_0(x) + f_0(x)g_0(x)(x - a))\,(x - a) \\ &= (fg)(a) + h(x)\,(x - a),\end{aligned}$$

for all $x \in \Omega$, where we have set $h(x) = f_0(x)g(a)+f(a)g_0(x)+f_0(x)g_0(x)(x-a)$. By the Algebra of Continuous Functions, the function $h : \Omega \to \mathbb{R}$ is continuous at a. From Proposition 5.10 we finally deduce that fg is differentiable at a, and $(fg)'(a) = h(a) = f_0(a)g(a) + f(a)g_0(a) = f'(a)g(a) + f(a)g'(a)$.

The proof of part (iii) is postponed to the next section. $\qquad\square$

Proposition 5.18 (chain rule). *Let $f : \Omega \to \mathbb{R}$ and $g : A \to \mathbb{R}$ for some $\Omega, A \subseteq \mathbb{R}$ such that $f(\Omega) \subseteq A$. Let $a \in \Omega$ be such that f is differentiable at a and g is differentiable at $f(a)$. Then the composition $g \circ f$ is differentiable at a, and*

$$(g \circ f)'(a) = (g' \circ f)(a)\, f'(a). \tag{5.7}$$

Proof. Let $b = f(a)$. Since f is differentiable at a and g is differentiable at b, by Proposition 5.10 we deduce that there exist $f_0 : \Omega \to \mathbb{R}$ and $g_0 : A \to \mathbb{R}$, continuous at a and b respectively, such that

$$f(x) = f(a) + f_0(x)\,(x - a), \qquad g(y) = g(b) + g_0(y)\,(y - b)$$

for all $x \in \Omega$ and $y \in A$, and moreover $f_0(a) = f'(a)$ and $g_0(b) = g'(b)$.

Consequently

$$
\begin{aligned}
(g \circ f)(x) &= g(f(x)) \\
&= g(f(a) + f_0(x)\,(x - a)) \\
&= g(b) + g_0(f(a) + f_0(x)\,(x - a))\,(f(a) + f_0(x)\,(x - a) - b) \\
&= g(f(a)) + h(x)\,(x - a),
\end{aligned}
$$

for all $x \in \Omega$, where we have set $h(x) = g_0(f(a) + f_0(x)\,(x - a))\, f_0(x)$ and used that $b = f(a)$. By the Algebra of Continuous Functions, we deduce that $h : \Omega \to \mathbb{R}$ is continuous at a. Hence, by Proposition 5.10, we conclude that $g \circ f$ is differentiable at a, and $(g \circ f)'(a) = h(a) = g_0(f(a))\, f_0(a) = g'(f(a))\, f'(a)$. $\quad\square$

Remark 5.19. A commonly used way to remember the chain rule is to write it by using the notation of Remark 5.5. Indeed, if

$$y = f(x) \qquad \text{and} \qquad z = g(y),$$

then we can write $f' = \frac{dy}{dx}$ and $g' = \frac{dz}{dy}$. Moreover

$$z = g(y) = g(f(x)) = g \circ f(x),$$

hence we can write $(g \circ f)' = \frac{dz}{dx}$. With this notation, the chain rule (5.7) is often written as follows:

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}.$$

This expression (which appears to be obtained by formally "multiplying and dividing by dy") may be easier to remember than (5.7), however it must be used with great care, because it does not clearly specify <u>at what points</u> the functions must be computed; a somewhat more precise version of the above expression is

$$
\begin{aligned}
\frac{dz}{dx}(x) &= \frac{dz}{dy}(y) \cdot \frac{dy}{dx}(x) \\
&= \frac{dz}{dy}(f(x)) \cdot \frac{dy}{dx}(x),
\end{aligned}
$$

which is more clearly resembling (5.7).

The Chain Rule can also be used to work out the formula for the derivative of the inverse of a differentiable function. Indeed, if $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$ are differentiable functions with $f(A) \subseteq B$ and

$$g(f(x)) = x$$

for all $x \in A$, then by differentiating both sides of the above identity and applying the Chain Rule we obtain that

$$g'(f(x))f'(x) = 1,$$

which implies that $f'(x) \neq 0$ and

$$g'(y) = \frac{1}{f'(x)},$$

where $y = f(x)$. Note that the above discussion requires us to assume that both f and g are differentiable. The following proposition gives a more precise result, in that we can *deduce* that the inverse g is differentiable from the differentiability of f (and the continuity of g), provided we know that the derivative f' does not vanish.

Proposition 5.20 (derivative of the inverse). *Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \Omega$. Assume that f has a real-valued inverse $g : A \to \mathbb{R}$, which is continuous at $f(a)$. Assume moreover that f is differentiable at a and $f'(a) \neq 0$. Then g is differentiable at $f(a)$ and*

$$g'(f(a)) = \frac{1}{f'(a)}.$$

Proof. Since f is differentiable at a, by Proposition 5.10 we deduce that there exist $f_0 : \Omega \to \mathbb{R}$, continuous at a, such that

$$f(x) = f(a) + f_0(x)\,(x - a) \tag{5.8}$$

for all $x \in \Omega$, and moreover $f_0(a) = f'(a)$. For all $y \in A$, by applying (5.8) to $x = g(y)$, we deduce that

$$y = f(g(y)) = f(a) + f_0(g(y))\,(g(y) - a) = b + f_0(g(y))\,(g(y) - g(b)),$$

where we have set $b = f(a)$.

The last identity can be rewritten as

$$y - b = f_0(g(y))\,(g(y) - g(b)),$$

which shows that, if $y \neq b$, then $f_0(g(y)) \neq 0$. On the other hand, $f_0(g(b)) = f_0(a) = f'(a) \neq 0$. So we can define the function $h : A \to \mathbb{R}$ by $h(y) = 1/f_0(g(y))$, which is continuous at b by the Algebra of Continuous Functions, and moreover

$$g(y) - g(b) = h(y)\,(y - b)$$

for all $y \in A$. By Proposition 5.10 we conclude that g is differentiable at b and $g'(b) = h(b) = 1/f_0(g(b)) = 1/f'(a)$. $\qquad\square$

5.5   Some notable derivatives

Proposition 5.21. *Let $n \in \mathbb{N}$. The function $x \mapsto x^n$ is differentiable, and its derivative is $x \mapsto nx^{n-1}$.*

Proof. Note that, for all $a, x \in \mathbb{R}$,

$$x^n - a^n = (x - a)(x^{n-1} + x^{n-2}a + \cdots + xa^{n-2} + a^{n-1}),$$

and therefore, if $x \neq a$,

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

By the Algebra of Limits, each of the n summands in the right-hand side tends to a^{n-1} as $x \to a$, hence

$$\lim_{x \to a} \frac{x^n - a^n}{x - a} = na^{n-1},$$

as desired. $\qquad\square$

Remark 5.22. An alternative proof of Proposition 5.21 can be given by induction, starting with the simpler case $n = 1$, and then applying the Leibniz rule.

Proposition 5.23. *The function $x \mapsto 1/x$ is differentiable, and its derivative is $x \mapsto -1/x^2$.*

Proof. Note that, for all $x, a \in \mathbb{R} \setminus \{0\}$,

$$\frac{1}{x} - \frac{1}{a} = \frac{a - x}{ax},$$

whence, if $x \neq a$,

$$\frac{\frac{1}{x} - \frac{1}{a}}{x - a} = -\frac{1}{ax}.$$

By the Algebra of Limits, we conclude that

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

as desired. $\qquad\square$

Remark 5.24. By combining the previous results, one can extend the validity of Proposition 5.21 to the case where n is a negative integer. Indeed, if $n \in \mathbb{Z}$ and $n < 0$, then $x^n = 1/(x^m)$, where $m = -n \in \mathbb{N}$. By Propositions 5.21 and 5.23 and the chain rule (Proposition 5.18), we deduce that the function $f : x \mapsto x^n$ is differentiable (on its domain $\mathbb{R} \setminus \{0\}$) and its derivative is given by

$$f'(x) = -\frac{1}{x^{2m}} \left(mx^{m-1} \right) = -\frac{m}{x^{m+1}} = nx^{n-1}.$$

By combining the previous results, we finally prove the "quotient rule" of Proposition 5.17

Proof of part (iii) of Proposition 5.17. By Proposition 5.23 and the chain rule, the function $1/g$ is differentiable at a, and

$$\left(\frac{1}{g}\right)'(a) = -\frac{1}{(g(a))^2} \cdot g'(a).$$

Note now that

$$\frac{f}{g} = f \cdot \frac{1}{g}.$$

Consequently, by the Leibniz rule, the function f/g is differentiable at a, and

$$\left(\frac{f}{g}\right)'(a) = f(a)\left(-\frac{1}{(g(a))^2} \cdot g'(a)\right) + f'(a)g(a) = \frac{f'(a)g(a) - f(a)g'(a)}{(g(a))^2},$$

as desired. $\qquad\square$

Remark 5.25. Using the previous results and the differentiation rules, it is not difficult to prove that polynomials and rational functions are differentiable (on their respective domains).

We now discuss the differentiability of trigonometric functions.

Proposition 5.26. *The sine function* $\sin : \mathbb{R} \to \mathbb{R}$ *and the cosine function* $\cos : \mathbb{R} \to \mathbb{R}$ *are differentiable, and their derivatives are given by*

$$\sin'(x) = \cos x, \qquad \cos'(x) = -\sin x.$$

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

Proof. Note that, for all $x, h \in \mathbb{R}$,

$$\sin(x + h) = \sin x \cos h + \cos h \sin x,$$

whence, if $h \neq 0$,

$$\frac{\sin(x + h) - \sin x}{h} = \sin x \frac{\cos h - 1}{h} + \cos x \frac{\sin h}{h}.$$

Note now that, by Proposition 4.20 and the Algebra of Limits,

$$\lim_{h \to 0} \frac{\cos h - 1}{h} = -\lim_{h \to 0} h \frac{1 - \cos h}{h^2} = 0,$$

$$\lim_{h \to 0} \frac{\sin h}{h} = 1,$$

whence, again by the Algebra of Limits,

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

This proves that sin is differentiable and its derivative is cos. Since

$$\cos x = \sin(\pi/2 - x)$$

for all $x \in \mathbb{R}$, and the function $x \mapsto \pi/2 - x$ is differentiable, by Proposition 5.18 we deduce that cos is differentiable too, and

$$\cos'(x) = -\sin'(\pi/2 - x) = \cos(\pi/2 - x) = \sin x$$

for all $x \in \mathbb{R}$. $\qquad\square$

From Propositions 5.26 and the differentiation rules we easily derive the following result.

Corollary 5.27. *The tangent function* tan *and the cotangent function* cot *are differentiable, and their derivatives are given by*

$$\tan'(x) = \frac{1}{\cos^2 x} = 1 + \tan^2 x, \qquad \cot'(y) = -\frac{1}{\sin^2 y} = -(1 + \cot^2 y).$$

for all x *in the domain of* tan *and all* y *in the domain of* cot.

From the previous results and Proposition 5.20 we also obtain formulas for the derivatives of the inverse trigonometric functions.

Proposition 5.28. *The arcsine function* arcsin *and the arccosine function* arccos *are both differentiable on the interval* $(-1, 1)$, *and their derivatives are given by*

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

for all $x \in (-1, 1)$.

Proof. We only discuss the differentiability of arcsin and the formula for its derivative. The proof for the function arccos is similar and is omitted.

Note that arcsin is the real-valued inverse of the function $h = \sin|_{[-\pi/2, \pi/2]}$. From Proposition 5.26 we deduce that h is differentiable and

$$h'(x) = \cos x = \sqrt{1 - \sin^2 x} = \sqrt{1 - (h(x))^2}$$

for all $x \in [-\pi/2, \pi/2]$; indeed, note that $\cos x \geq 0$ for all $x \in [-\pi/2, \pi/2]$, and actually $\cos x > 0$ if $x \in (-\pi/2, \pi/2)$. Since arcsin is continuous, from Proposition 5.20 we deduce that, for all $x \in (-\pi/2, \pi/2)$, the function arcsin is differentiable at $h(x)$ and

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

Note now that, for all $y \in (-1, 1)$, there exists $x \in (-\pi/2, \pi/2)$ such that $y = \sin x = h(x)$; consequently from the previous formula we deduce that

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

for all $y \in (-1, 1)$, as desired. $\qquad\square$

The proof of the following proposition is similar to the proof of the previous one (using Corollary 5.27 in place of Proposition 5.26) and is omitted.

Proposition 5.29. *The functions* $\arctan : \mathbb{R} \to \mathbb{R}$ *and* $\mathrm{arccot} : \mathbb{R} \to \mathbb{R}$ *are differentiable, and their derivatives are given, for all* $x \in \mathbb{R}$, *by*

$$\arctan'(x) = \frac{1}{1 + x^2}, \qquad \mathrm{arccot}'(x) = -\frac{1}{1 + x^2}.$$

Finally we treat the differentiability of exponential and logarithm functions.

Proposition 5.30. *The exponential function* $\exp : \mathbb{R} \to \mathbb{R}$ *and the logarithm function* $\log : (0, \infty) \to \mathbb{R}$ *are differentiable, and their derivatives are given by*

$$\exp'(x) = \exp(x), \qquad \log'(y) = \frac{1}{y}$$

for all $x \in \mathbb{R}$ *and* $y \in (0, \infty)$.

Proof. Note that, for all $x, h \in \mathbb{R}$,

$$\exp(x + h) - \exp(x) = e^{x+h} - e^x = e^x(e^h - 1),$$

whence, if $h \neq 0$,

$$\frac{\exp(x + h) - \exp(x)}{h} = \exp(x)\frac{e^h - 1}{h}.$$

By Proposition 4.23 and the Algebra of Limits,

$$\lim_{h \to 0} \frac{\exp(x + h) - \exp(x)}{h} = \exp(x) \lim_{h \to 0} \frac{e^h - 1}{h} = \exp(x);$$

this proves that exp is differentiable and its derivative is exp itself.

Similarly, note that, for all $y \in (0, \infty)$ and $h \in \mathbb{R}$,

$$\log(y+h) - \log y = \log(y(1+h/y)) = \log y + \log(1+h/y) - \log y = \log(1+h/y),$$

whence, if $h \neq 0$,

$$\frac{\log(y + h) - \log y}{h} = \frac{\log(1 + h/y)}{h} = \frac{1}{y}\frac{\log(1 + h/y)}{h/y}.$$

From Proposition 4.23 and a change of variables (see Proposition 4.14) we deduce that

$$\lim_{h \to 0} \frac{\log(1 + h/y)}{h/y} = \lim_{x \to 0} \frac{\log(1 + x)}{x} = 1$$

(indeed, note that $\lim_{h \to 0}(h/y) = 0$ and that $h/y \neq 0$ if $h \neq 0$), so, by the Algebra of Limits,

$$\lim_{h \to 0} \frac{\log(y + h) - \log y}{h} = \frac{1}{y} \lim_{h \to 0} \frac{\log(1 + h/y)}{h/y} = \frac{1}{y}.$$

This proves that log is differentiable and that $\log'(y) = 1/y$. $\square$

Remark 5.31. Proposition 5.30 only treats the case of the natural (base-e) exponential and logarithm. Note however that, for all $a \in (0, \infty) \setminus \{1\}$,

$$\exp_a(x) = a^x = \exp(x \log a), \qquad \log_a x = \frac{\log x}{\log a},$$

hence by the differentiation rules it is not difficult to check that $\exp_a$ and $\log_a$ are differentiable as well, and to deduce formulas for the respective derivatives.

Remark 5.32. By using the previous results, we can further extend Proposition 5.21 to the case where the exponent is a real number. Namely, let $f : (0, \infty) \to \mathbb{R}$ be given by $f(x) = x^{\alpha}$, where $\alpha \in \mathbb{R}$. Then we can write

$$f(x) = \exp(\alpha \log x)$$

for all $x \in (0, \infty)$. From Proposition 5.30 and the differentiation rules, we deduce that f is differentiable and

$$f'(x) = \exp(\alpha \log x)\frac{\alpha}{x} = \alpha\frac{x^{\alpha}}{x} = \alpha x^{\alpha-1},$$

as desired.

Remark 5.33. In both Remarks 5.31 and 5.32 the identity (4.24) was used to reduce exponentiation to a composition of functions involving exp and log. This can be generalised as follows: if $f, g : \Omega \to \mathbb{R}$ are functions such that $f(\Omega) \subseteq (0, \infty)$, then we can write

$$f(x)^{g(x)} = \exp(g(x) \log f(x));$$

in particular, if f and g are both differentiable, then the function $x \mapsto f(x)^{g(x)}$ is differentiable too, and its derivative can be expressed in terms of the functions f, g, f', g' by using Proposition 5.30 and the differentiation rules.

Example 5.34. Recall that the *hyperbolic functions* are defined as follows:

- the *hyperbolic sine* $\sinh : \mathbb{R} \to \mathbb{R}$ is defined by

$$\sinh x = \frac{e^x - e^{-x}}{2};$$

- the *hyperbolic cosine* $\cosh : \mathbb{R} \to \mathbb{R}$ is defined by

$$\cosh x = \frac{e^x + e^{-x}}{2};$$

- the *hyperbolic tangent* $\tanh : \mathbb{R} \to \mathbb{R}$ is defined by

$$\tanh x = \frac{\sinh x}{\cosh x};$$

- the *hyperbolic cotangent* $\coth : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ is defined by

$$\coth x = \frac{\cosh x}{\sinh x}.$$

Using Proposition 5.30 and the differentiation rules, it is not difficult to check that sinh, cosh, tanh and coth are differentiable and

$$\sinh'(x) = \cosh x, \qquad \cosh'(x) = \sinh x, \qquad \tanh'(x) = \frac{1}{\cosh^2 x} = 1 - \tanh^2 x$$

for all $x \in \mathbb{R}$, while

$$\coth'(x) = -\frac{1}{\sinh^2 x} = 1 - \coth^2 x$$

for all $x \in \mathbb{R} \setminus \{0\}$.

5.6   Higher-order derivatives

Given a real-valued function f of a real variable, according to Definition 5.1 we can construct another function f' (the derivative of f), whose domain is a subset[11] of the domain of f (namely, the set of the points at which f is differentiable).

We can now iterate this construction, and consider the derivative f'' of f', the derivative f''' of f'', and so on. These functions are known as *higher-order derivatives* of the function f; specifically, f'' is called the *second derivative* of f, f''' is called the *third derivative* of f, and so on. In this context, the derivative f' of f is also called the *first derivative* of f, and the function f itself can be thought of as the *zeroth derivative* of f.

Clearly the notation f', f'', f''', etc., for the subsequent derivatives of a function f becomes a bit cumbersome if one wants to consider derivatives of very high order; in this case the alternative notation $f^{(n)}$ for the nth derivative can be used instead (so $f^{(0)} = f$, $f^{(1)} = f'$, $f^{(2)} = f''$, and so on). Using this notation, a precise definition of the nth derivative of a function can be given recursively as follows.

Definition 5.35. Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. For all $n \in \mathbb{N}_0$, the *nth derivative* $f^{(n)}$ of f is defined by

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

where $(f^{(n-1)})'$ denotes the derivative of $f^{(n-1)}$.

Remark 5.36. The alternative notation for higher-order derivatives that corresponds to the notation (5.4)-(5.5) described in Remark 5.5 is as follows: for all $n \in \mathbb{N}$, the nth derivative $f^{(n)}$ of f can also be denoted by

$$\frac{d^n f}{dx^n} \qquad \text{or} \qquad \frac{d^n}{dx^n} f.$$

Again, this notation is particularly expressive in contexts such physics. For example, when discussing the motion of a point mass on a line, the acceleration (denoted by a) is the derivative of the velocity (denoted by v) with respect to time (denoted by t), hence it is the second derivative of the position (denoted by s) with respect to time:

$$a = \frac{dv}{dt} = \frac{d^2 s}{dt^2}.$$

Note that, according to Definition 5.1, a function f is differentiable at a if and only if a is an element of the domain of the derivative f'. Using this idea, we can introduce the notion of "higher-order differentiability" as follows.

Definition 5.37. Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $n \in \mathbb{N}_0$.

(i) Let $a \in \Omega$. We say that f *is n times differentiable at a* if a is an element of the domain of the nth derivative $f^{(n)}$ of f.

---

[11]If f is a differentiable function, then the domain of f' is the same as the domain of f. However, for an arbitrary function f, the domain of the derivative f' could be smaller than the domain of f (it could even be empty).

(ii) If f is n times differentiable at every point of Ω, then we just say that f is *n times differentiable*.

(iii) If f is n times differentiable and $f^{(n)}$ is continuous, then we say that f is *n times continuously differentiable*, or that f is *of class C^n*.

Many of the functions that are commonly used in Calculus can be differentiated arbitrarily many times, that is, they are n times differentiable for arbitrarily large n. The following definition applies to these functions.

Definition 5.38. Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. The function f is said to be *infinitely differentiable* (or *of class C^∞*) if f is n times differentiable for all $n \in \mathbb{N}_0$.

Example 5.39. Many of the functions discussed in Section 5.5 (including polynomials, rational functions, trigonometric functions, exponential and logarithm functions, hyperbolic functions) are infinitely differentiable. This is easily deduced by the formulas in Section 5.5 and the differentiation rules.

Remark 5.40. From Corollary 5.11 we deduce that, if a function f is n times differentiable, then its $(n-1)$th derivative $f^{(n-1)}$ is continuous. In particular, if f is infinitely differentiable, then all the derivatives $f^{(n)}$ of f are continuous.

One should be mindful that not all differentiable functions are infinitely differentiable: indeed, a differentiable function may not even be continuously differentiable (that is, its derivative may not be continuous).

Example 5.41. Here we present an example of a differentiable function f, which is not twice differentiable, and whose derivative f' is not continuous. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Note that f is differentiable at 0. Indeed, the Newton quotient at 0 is given by

$$\frac{f(x) - f(0)}{x - 0} = x \sin \frac{1}{x},$$

whence

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

On the other hand, by applying the results of Section 5.5 and the differentiation rules to the restriction $f|_{\mathbb{R} \setminus \{0\}}$, it is not difficult to see that f is differentiable at every point $x \neq 0$, and

$$f'(x) = 2x \sin \frac{1}{x} - \cos \frac{1}{x}$$

for all $x \neq 0$. In conclusion, f is differentiable (at every point of its domain) and

$$f'(x) = \begin{cases} 2x \sin \frac{1}{x} - \cos \frac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

On the other hand, the derivative f' is not continuous at 0. Indeed, note that, if we define $g : \mathbb{R} \to \mathbb{R}$ by

$$g(x) = \begin{cases} x \sin \frac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

then g is continuous at 0 (see Example 4.13). So, if f' were continuous at 0, then, by the Algebra of Continuous Functions, the function $2g - f'$ would be continuous at 0 too; however,

$$2g(x) - f'(x) = \begin{cases} \cos \frac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

and one can prove that $\lim_{x \to 0} \cos \frac{1}{x}$ does not exist (cf. Example 3.13), so g is not continuous at 0 and therefore f' is also not continuous at 0 (in particular, f is differentiable, but it is not continuously differentiable). Moreover, by Corollary 5.11, we deduce that f' is not differentiable at 0, and therefore f is not twice differentiable at 0.

By adapting the previous example, one can show that, for all $n \in \mathbb{N}$, there exists a function which is n times differentiable, but not $n+1$ times differentiable. The order of differentiability of a function f (that is, the maximum $n \in \mathbb{N}$ such that f is n times differentiable, or ∞ in the case f is infinitely differentiable) can be used as a way of measuring the "regularity" (or "smoothness") of f.

6 Applications of differentiation

6.1 Maxima, minima, and stationary points

In Section 2.8 we gave the definition of local/global maximum and minimum points of a real-valued function f of a real variable (see Definition 2.37). In the case the function f is differentiable, the analysis of its derivative f' yields information about such points.

Definition 6.1. Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. A point $a \in \Omega$ is said to be a *stationary point* (or *critical point*) of f if f is differentiable at a and $f'(a) = 0$.

The next result, also known as Fermat's Theorem, shows that there is a relation between local maximum and minimum points and stationary points of a differentiable function, provided the points are "interior points" of the domain, in the sense of the following definition.

Definition 6.2. Let $\Omega \subseteq \mathbb{R}$. A point $a \in \mathbb{R}$ is said to be an *interior point* of Ω if there exists $r > 0$ such that $(a - r, a + r) \subseteq \Omega$.

Remark 6.3. We list here the main properties of interior points that will be of use for us. The proofs of the following statements are similar to those of Proposition 3.6 and Corollary 3.7, and are left as an exercise to the interested reader.

(a) If Ω is any of the intervals $[b, c]$, (b, c), $[b, c)$ and $(b, c]$, where $b, c \in \mathbb{R}$ and $b < c$, then the interior points of Ω are exactly the elements of $(b, c) = \Omega \setminus \{b, c\}$, that is, those points of Ω that are <u>not endpoints</u> of the interval. Similar considerations apply to unbounded intervals too (that is, when one or both of b and c are $\pm\infty$).

(b) If $\Omega \subseteq \Omega' \subseteq \mathbb{R}$, then any interior point of Ω is also an interior point of Ω'.

Proposition 6.4 (Fermat's Theorem)**.** *Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let a be an interior point of Ω. If a is a local maximum point or a local minimum point of f, and f is differentiable at a, then a is a stationary point of f.*

Proof. Without loss of generality, we may assume that a is a local minimum point of f. Hence we can find $r > 0$ such that $(a - r, a + r) \subseteq \Omega$ and moreover

$$f(x) \geq f(a) \tag{6.1}$$

for all $x \in (a - r, a + r)$.

Consequently, if $x \in (a, a+r)$, then $x > a$ and therefore from (6.1) we deduce that

$$\frac{f(x) - f(a)}{x - a} \geq 0.$$

Since f is differentiable at a, by Propositions 5.14 and 3.29 we deduce that

$$f'(a) = f'_+(a) = \lim_{x \to a^+} \frac{f(x) - f(a)}{x - a} \geq 0.$$

On the other hand, if $x \in (a - r, a)$, then $x < a$ and from (6.1) we deduce that

$$\frac{f(x) - f(a)}{x - a} \leq 0.$$

Since f is differentiable at a, again by Propositions 5.14 and 3.29 we obtain that

$$f'(a) = f'_-(a) = \lim_{x \to a^-} \frac{f(x) - f(a)}{x - a} \leq 0.$$

In conclusion $f'(a) = 0$, as desired. $\qquad \square$

Remark 6.5. The assumption that a is an interior point of Ω cannot be left out in general. For example, if we take $\Omega = [0, 1]$ and $f : [0, 1] \to \mathbb{R}$ given by $f(x) = x$, then 0 is clearly a global minimum point of f, and 1 a global maximum point of f. On the other hand, f is differentiable and $f'(x) = 1$ for all $x \in [0, 1]$; so in particular $f'(0) = f'(1) = 1 \neq 0$, and therefore 0 and 1 are not stationary points of f.

Remark 6.6. The implication given by Proposition 6.4 cannot be reversed in general. For example, let $f : \mathbb{R} \to \mathbb{R}$ be given by

$$f(x) = x^3.$$

Then $f'(x) = 3x^2$, and in particular $f'(0) = 0$, so 0 is a stationary point of f. On the other hand, f is strictly increasing, so 0 cannot be a (local or global) maximum or minimum point of f.

Remark 6.7. As shown in the previous remark, for an interior point of the domain of a differentiable function, being a stationary point is only a necessary condition, but not a sufficient condition for being a local maximum/minimum point. As we will see in Corollaries 6.18 and 6.21 below, the analysis of the sign of the derivative near the point, or the analysis of the sign of the second derivative at the point, can be enough to determine whether a stationary point is actually a local maximum/minimum point.

6.2   Rolle's Theorem and Mean Value Theorem

Proposition 6.8 (Rolle's Theorem)**.** *Let $f : [a, b] \to \mathbb{R}$ for some $a, b \in \mathbb{R}$ with $a < b$. Assume that f is continuous on $[a, b]$ and differentiable on (a, b). If $f(a) = f(b)$, then there exists a point $c \in (a, b)$ such that $f'(c) = 0$.*

Proof. By the Boundedness Theorem (Theorem 4.28), there exists a global maximum point $x_M \in [a, b]$ and a global minimum point $x_m \in [a, b]$ of f. If both x_M and x_m are endpoints of the interval $[a, b]$, then from our assumption $f(a) = f(b)$ we deduce that $f(x_m) = f(x_M)$, and therefore

$$f(x_m) \leq f(x) \leq f(x_M) = f(x_m)$$

for all $x \in [a, b]$, whence $f(x) = f(x_M) = f(x_m)$ for all $x \in [a, b]$. In other words, in this case f is constant and then trivially $f'(c) = 0$ for all $c \in (a, b)$.

Assume instead that one of the points x_m and x_M is not an endpoint of the interval $[a, b]$. In other words, one of the points x_m and x_M belongs to (a, b). Let c denote such point. Then c is either a local maximum point or a local minimum point for f, and moreover f is differentiable at c, so by Proposition 6.4 we deduce that $f'(c) = 0$. $\qquad \square$

Proposition 6.9 ((Lagrange's) Mean Value Theorem). *Let $f : [a, b] \to \mathbb{R}$ for some $a, b \in \mathbb{R}$ with $a < b$. Assume that f is continuous on $[a, b]$ and differentiable on (a, b). Then there exists a point $c \in (a, b)$ such that*

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

Proof. Let $g : [a, b] \to \mathbb{R}$ be defined by

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

Then, by the Algebra of Continuous Functions and the differentiation rules, g is also continuous on $[a, b]$ and differentiable on (a, b). Moreover

$$g(a) = f(a) - \frac{f(b) - f(a)}{b - a} \cdot 0 = f(a)$$

and

$$g(b) = f(b) - \frac{f(b) - f(a)}{b - a}(b - a) = f(b) - (f(b) - f(a)) = f(a),$$

whence $g(a) = g(b)$. If we apply Proposition 6.8 to the function g, we deduce the existence of a point $c \in (a, b)$ such that $g'(c) = 0$. On the other hand, by the differentiation rules, for all $x \in (a, b)$,

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

and in particular

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

whence

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

as desired. $\qquad\square$

An immediate consequence of the Mean Value Theorem is the following result, which is crucial in the development of integration theory.

Corollary 6.10. *Let $a, b \in \mathbb{R}$ with $a < b$.*

(i) *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b), and assume that*

$$f'(x) = 0$$

for all $x \in (a, b)$. Then f is constant.

(ii) *Let $f, g : [a, b] \to \mathbb{R}$ be both continuous on $[a, b]$ and differentiable on (a, b), and assume that*

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

for all $x \in (a, b)$. Then there exists $c \in \mathbb{R}$ such that

$$f(x) = g(x) + c$$

for all $x \in [a, b]$.

Proof. (i). For all $x, y \in [a, b]$ with $x < y$, we can apply Proposition 6.9 to the restriction $f|_{[x,y]}$, thus obtaining that there exists $c \in (x, y)$ such that

$$\frac{f(y) - f(x)}{y - x} = f'(c) = 0;$$

the last equality is due to our assumption on f. Consequently $f(y) - f(x) = 0$, that is, $f(x) = f(y)$. Since x, y are arbitrary points of $[a, b]$, this proves that f is constant.

(ii). Apply part (i) to the function $f - g$. □

Remark 6.11. The assumption in Corollary 6.10 that the domain of the considered functions is an interval cannot be dropped in general. Indeed, if $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ is defined by $f(x) = x/|x|$, then $f(x) = 1$ for all $x > 0$ and $f(x) = -1$ for all $x < 0$; in particular $f'(x) = 0$ for all $x \in \mathbb{R} \setminus \{0\}$, but f is not constant.

Proposition 6.12 ((Cauchy's) Generalised Mean Value Theorem)**.** *Let $f, g : [a, b] \to \mathbb{R}$ for some $a, b \in \mathbb{R}$ such that $a < b$. Assume that both f and g are continuous on $[a, b]$ and differentiable on (a, b). Assume further that $g'(x) \neq 0$ for all $x \in (a, b)$. Then $g(a) \neq g(b)$, and there exists $c \in (a, b)$ such that*

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

Proof. The proof is similar to that of Proposition 6.9, and is obtained by applying Proposition 6.8 to the function $h : [a, b] \to \mathbb{R}$ defined by

$$h(x) = (f(b) - f(a))(g(x) - g(a)) - (g(b) - g(a))(f(x) - f(a)).$$

We leave the details to the interested reader. □

6.3 Monotonicity and first derivative

In Section 2.7 we discussed monotonicity properties of real-valued functions of a real variable, such as being increasing, decreasing, strictly increasing or strictly decreasing. Here we show that, when a function is differentiable, such monotonicity properties can be related to the sign of the derivative of the function.

The following preliminary result shows that monotonicity can be characterised in terms of the sign of the Newton quotient.

Proposition 6.13. *Let $f : \Omega \to \mathbb{R}$, where $\Omega \subseteq \mathbb{R}$.*

(i) *f is increasing if and only if $\frac{f(y) - f(x)}{y - x} \geq 0$ for all $x, y \in \Omega$ with $x \neq y$.*

(ii) *f is strictly increasing if and only if $\frac{f(y) - f(x)}{y - x} > 0$ for all $x, y \in \Omega$ with $x \neq y$.*

(iii) *f is decreasing if and only if $\frac{f(y) - f(x)}{y - x} \leq 0$ for all $x, y \in \Omega$ with $x \neq y$.*

(iv) *f is decreasing if and only if $\frac{f(y) - f(x)}{y - x} < 0$ for all $x, y \in \Omega$ with $x \neq y$.*

Proof. (i). Assume that f is increasing. Take any $x, y \in \Omega$ such that $x \neq y$. If $x < y$, then $f(x) \leq f(y)$ (by definition of increasing function); from this we deduce that $f(y) - f(x) \geq 0$ and $y - x > 0$, and taking the quotient we obtain that $\frac{f(y)-f(x)}{y-x} \geq 0$, as desired. If instead $x > y$, the same argument yields $\frac{f(x)-f(y)}{x-y} \geq 0$, but this gives again the desired conclusion because $\frac{f(y)-f(x)}{y-x} = \frac{f(x)-f(y)}{x-y}$.

Conversely, assume that $\frac{f(y)-f(x)}{y-x} \geq 0$ for all $x, y \in \Omega$ with $x \neq y$. To prove that f is increasing, take any $x, y \in \Omega$ with $x \leq y$; we must prove that $f(x) \leq f(y)$. If $x = y$, then $f(x) = f(y)$ and we are done. If instead $x < y$, then we can apply our assumption and deduce that $\frac{f(y)-f(x)}{y-x} \geq 0$. Since $y - x > 0$ (because $x < y$), multiplying both sides of the previous inequality by the denominator $y - x$ finally yields $f(y) - f(x) \geq 0$, that is, $f(x) \leq f(y)$, as desired.

(ii). The proof follows the lines of that of part (i) (one just need to replace any $\leq$ by $<$ and any $\geq$ by $>$).

(iii) and (iv). The proofs for these parts are analogous to the ones for parts (i) and (ii), with small modifications. Alternatively, one can apply parts (iii) and (iv) to the function $-f$. $\qquad\square$

The previous proposition does not require the function f to be differentiable. If we assume differentiability, then the previous proposition, combined with a result about preservation of inequalities under taking limits (see Section 3.7), allows us to deduce information on the sign of the derivative from the monotonicity properties of the function.

Proposition 6.14. *Let $f : \Omega \to \mathbb{R}$, where $\Omega \subseteq \mathbb{R}$, be an increasing $[\![$ resp. decreasing $]\!]$ function. Then $f'(x) \geq 0$ $[\![$ resp. $f'(x) \leq 0$ $]\!]$ for all $x \in \Omega$ such that f is differentiable at x.*

Proof. We only consider the case where f is increasing; the proof in the other case is analogous.

Let $x \in \Omega$. Then, for all $t \in \Omega \setminus \{x\}$, from Proposition 6.13 we deduce that

$$\frac{f(t) - f(x)}{t - x} \geq 0. \tag{6.2}$$

If f is differentiable at x, then, from Proposition 3.29 we conclude that

$$f'(x) = \lim_{t \to x} \frac{f(t) - f(x)}{t - x} \geq 0,$$

as desired. $\qquad\square$

By using the Mean Value Theorem (Proposition 6.9), we can prove a sort of converse to the previous proposition, in the case where the domain is an <u>interval</u> (see Section 1.5 for a discussion of intervals in the real line; see also Remark 6.3 about interior points). The following result allows us to deduce information on strict monotonicity too, provided the information on the sign of the derivative is also given by a strict inequality.

Proposition 6.15. *Let $f : I \to \mathbb{R}$ for some interval $I \subseteq \mathbb{R}$. Let I_0 be the set of the interior points of I (that is, the points of the interval I that are not endpoints). Assume that f is continuous on I and differentiable on I_0.*

(i) If $f'(x) \geq 0$ for all $x \in I_0$, then f is increasing on I.

(ii) If $f'(x) > 0$ for all $x \in I_0$, then f is strictly increasing on I.

(iii) If $f'(x) \leq 0$ for all $x \in I_0$, then f is decreasing on I.

(iv) If $f'(x) < 0$ for all $x \in I_0$, then f is strictly decreasing on I.

Proof. We only prove parts (i) and (ii). The other parts (iii) and (iv) can be proved analogously, or alternatively can be deduced by applying parts (i) and (ii) to the function $-f$.

(i). In light of Proposition 6.13, it is enough to prove that

$$\frac{f(y) - f(x)}{y - x} \geq 0$$

for all $x, y \in \Omega$ with $x \neq y$. Since switching x and y does not change the Newton quotient in the left-hand side, in proving the inequality we may actually assume that $x < y$. We now Note now that $[x, y] \subseteq I$ and $(x, y) \subseteq I_0$, so the restriction $f|_{[x,y]}$ is continuous on $[x, y]$ and differentiable on (x, y). Hence, by the Mean Value Theorem (Proposition 6.9), there exists $c \in (x, y)$ such that

$$\frac{f(y) - f(x)}{y - x} = f'(c).$$

Since $c \in (x, y) \subseteq I_0$, from our assumption we deduce that $f'(c) \geq 0$, hence the Newton quotient is nonnegative, as desired.

(ii). The proof of this part follows the lines of that of part (i) (just replace any $\geq$ with $>$). $\qquad\square$

Remark 6.16. The assumption in Proposition 6.15 that the domain is an interval cannot be dropped in general. For example (see Corollary 5.27), the tangent function $f(x) = \tan x$ has derivative $f'(x) = 1/\cos^2 x$, which is positive for all x in the domain $D = \{x \in \mathbb{R} : \cos x \neq 0\}$ of the tangent function. However, $\tan(\pi/4) = 1 > -1 = \tan(3\pi/4)$, while $\pi/4 < 3\pi/4$, so f is not increasing.

By combining Propositions 6.14 and 6.15 we obtain the following characterisation of differentiable monotone functions on an interval.

Corollary 6.17. *Let $f : I \to \mathbb{R}$ for some interval $I \subseteq \mathbb{R}$. Assume that f is continuous on I and differentiable on the set I_0 of the interior points of I.*

(i) f is increasing on I if and only if $f'(x) \geq 0$ for all $x \in I_0$.

(ii) f is decreasing on I if and only if $f'(x) \leq 0$ for all $x \in I_0$.

Another immediate consequence of the previous results is a criterion to locate local maximum and minimum points of f based on the changes of sign of its derivative f'.

Corollary 6.18. *Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $a \in \Omega$ and $r > 0$ be such that $(a - r, a + r) \subseteq \Omega$, and assume that f is continuous on $(a - r, a + r)$ and differentiable on $(a - r, a + r) \setminus \{a\}$.*

(i) If $f'(x) \leq 0$ for all $x \in (a - r, a)$ and $f'(x) \geq 0$ for all $x \in (a, a + r)$, then a is a local minimum point for f.

(ii) If $f'(x) \geq 0$ for all $x \in (a - r, a)$ and $f'(x) \leq 0$ for all $x \in (a, a + r)$, then a is a local maximum point for f.

Remark 6.19. A point a where the derivative f' of a function changes sign (as described in Corollary 6.18) is sometimes referred to as a *turning point* of f.

Remark 6.20. Corollary 6.18 complements Fermat's Theorem (Proposition 6.4): if $f : \Omega \to \mathbb{R}$ is differentiable and a is an interior point of Ω, then in order for a to be a local maximum/minimum point for f it is necessary that a is a stationary point (that is, the derivative f' vanishes at a), and it is sufficient that a is a turning point (that is, the derivative f' changes sign at a).

Under additional regularity conditions on the function f, the analysis of the sign of f' at all points x sufficiently close to a (as in Corollary 6.18) can be replaced by an information on the sign of the second derivative f'' just at the point a.

Corollary 6.21 ("Second Derivative Test"). *Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$ be twice continuously differentiable. Let a be an interior point of Ω.*

(i) If $f'(a) = 0$ and $f''(a) > 0$, then a is a local minimum point for f.

(ii) If $f'(a) = 0$ and $f''(a) < 0$, then a is a local maximum point for f.

Proof. We only prove part (i); the proof of the other part is analogous.

Since f'' is continuous at a and $f''(a) > 0$, by taking $\epsilon = f''(a)$ and applying Definition 4.1 we can find $\delta > 0$ such that $(a - \delta, a + \delta) \subseteq \Omega$ and

$$|f''(x) - f''(a)| < f''(a)$$

for all $x \in (a - \delta, a + \delta)$; consequently

$$f''(x) > f''(a) - f''(a) = 0$$

for all $x \in (a - \delta, a + \delta)$. From Proposition 6.15 (applied to to the function f' in place of f), we deduce that f' is strictly increasing on $(a - \delta, a + \delta)$, and therefore

$$f'(x) < f'(a) < f'(y)$$

for all $x \in (a - \delta, a)$ and $y \in (a, a + \delta)$; since $f'(a) = 0$, by Corollary 6.18 we deduce that a is a local minimum point for f. $\qquad \square$

6.4   Convexity and second derivative

Definition 6.22. Let $f : I \to \mathbb{R}$ for some interval $I \subseteq \mathbb{R}$.

(i) We say that f is *convex* (or *concave upwards*) if, for all $a, b \in I$ and $t \in [0, 1]$,
$$f((1 - t)a + tb) \leq (1 - t)f(a) + tf(b).$$

(ii) We say that f is *concave* (or *concave downwards*) if, for all $x, y \in I$ and $t \in [0, 1]$,
$$f((1 - t)a + tb) \geq (1 - t)f(a) + tf(b).$$

Remark 6.23. Let $f : I \to \mathbb{R}$ for some interval $I \subseteq \mathbb{R}$, and let $a, b \in I$ be such that $a < b$. By using basic Cartesian geometry, it is easily seen that the straight line through the points $(a, f(a))$ and $(b, f(b))$ can be parametrised as follows: every point of the line has the form

$$((1-t)a + tb, (1-t)f(a) + tf(b))$$

where $t \in \mathbb{R}$; moreover, the points corresponding to $t \in [0, 1]$ are those lying on the line segment with endpoints $(a, f(a))$ and $(b, f(b))$. Starting from this observation, it is not difficult to check that the function f is convex ⟦resp. concave⟧ if and only if, for all $a, b \in I$ with $a < b$, the graph of the function f restricted to $[a, b]$ lies below ⟦resp. above⟧ the straight line passing through the points $(a, f(a))$ and $(b, f(b))$.

Similarly to the case of monotone functions, not every function is convex or concave, however it may become convex or concave when restricted to suitable intervals.

Definition 6.24. Let $f : \Omega \to \mathbb{R}$ for some $\Omega \subseteq \mathbb{R}$. Let $I \subseteq \Omega$ be an interval. We say that f is *convex on I* ⟦resp. *concave on I*⟧ if the restriction $f|_I$ is convex ⟦resp. concave⟧.

It is possible to characterise convex functions in terms of Newton quotients.

Proposition 6.25. *Let $f : I \to \mathbb{R}$ for some interval $I \subseteq \mathbb{R}$. Then the following are equivalent:*

(i) f is convex.

(ii) For all $a, b, c \in I$ such that $a < c < b$,

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

Proof (sketch). Let $a, b \in I$ with $a < b$. Let $t \in (0, 1)$ and set $c = (1-t)a + tb$, so that $t = (c - a)/(b - a)$ and $1 - t = (b - c)/(b - a)$. Then the inequality

$$f((1-t)a + tb) \leq (1-t)f(a) + tf(b)$$

that appears in the definition of convexity can be equivalently rewritten as

$$f(c) \leq \frac{b - c}{b - a}f(a) + \frac{c - a}{b - a}f(b),$$

that is,

$$(b - a)f(c) \leq (b - c)f(a) + (c - a)f(b).$$

Since $b - a = (b - c) + (c - a)$, the previous inequality is equivalent to

$$(b - c)f(c) + (c - a)f(c) \leq (b - c)f(a) + (c - a)f(b),$$

that is,

$$(b - c)(f(c) - f(a)) \leq (c - a)(f(b) - f(c));$$

dividing both sides by $(b-c)(c-a)$, we see that the last inequality is equivalent to

$$\frac{f(c) - f(a)}{c - a} \leq \frac{f(b) - f(c)}{b - c},$$

as desired. $\square$

For a differentiable function, it is possible to chararacterise its convexity/concavity in terms of its derivative.

Proposition 6.26. *Let $f : I \to \mathbb{R}$ for some interval $I \subseteq \mathbb{R}$. Assume that f is differentiable. Then f is convex $\llbracket$ resp. concave $\rrbracket$ if and only if its derivative f' is increasing $\llbracket$ resp. decreasing $\rrbracket$.*

Proof. We only treat the case of f convex; the other case is derived from the first one by replacing f with $-f$.

Suppose first that f is convex, and take $a, b \in I$ with $a < b$. Let $h > 0$ be sufficiently small that $a + h < b - h$. By applying Proposition 6.25 twice we deduce that

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

and consequently

$$\frac{f(a+h) - f(a)}{h} \leq \frac{f(b-h) - f(b)}{-h}. \tag{6.3}$$

Note now that, since f is differentiable,

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

while

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

From the inequality (6.3) and Proposition 3.29 we then deduce that

$$f'(a) \leq f'(b).$$

Since $a, b \in I$ are arbitrary points with $a < b$, this proves that f' is increasing.

Conversely, assume that f' is increasing, and let $a, b, c \in I$ be such that $a < c < b$. Then, by applying Proposition 6.9 to the functions $f|_{[a,c]}$ and $f|_{[c,b]}$, we deduce that there exist points $\xi \in (a, c)$ and $\eta \in (c, b)$ such that

$$\frac{f(c) - f(a)}{c - a} = f'(\xi) \qquad \text{and} \qquad \frac{f(b) - f(c)}{b - c} = f'(\eta).$$

Note that $\xi < c < \eta$; since f' is increasing, we then deduce that $f'(\xi) \leq f'(\eta)$, that is,

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

Since this inequality holds for arbitrary $a, b, c \in I$ with $a < c < b$, by Proposition 6.25 we deduce that f is convex. $\qquad \qquad \square$

By combining this result with Corollary 6.17, we obtain a further characterisation of convexity/concavity in terms of the sign of the second derivative.

Corollary 6.27. *Let $f : I \to \mathbb{R}$ for some interval $I \subseteq \mathbb{R}$, and let I_0 be the set of the interior points of I. Assume that f is twice differentiable. Then f is convex $\llbracket$ resp. concave $\rrbracket$ if and only if $f''(x) \geq 0$ $\llbracket$ resp. $f''(x) \leq 0$ $\rrbracket$ for all $x \in I_0$.*

Remark 6.28. Assume that $f'' : I \to \mathbb{R}$ is twice differentiable, where $I \subseteq \mathbb{R}$ is an interval. A turning point of the derivative f' (see Remark 6.19) is sometimes referred to as an *inflexion point* of f. According to Corollary 6.27, an inflexion point is a point where f changes the direction of its concavity (i.e., downwards to upwards, or vice versa).

Example 6.29. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^3 - x$. Then $f'(x) = 3x^2 - 1$ and $f''(x) = 6x$ for all $x \in \mathbb{R}$. The sign analysis of the polynomials $3x^2 - 1$ and $6x$ (see Section 2.5) tells us that:

- $f'(x)$ is zero for $x = \pm 1/\sqrt{3}$, positive for $x \in (-\infty, -1/\sqrt{3}) \cup (1/\sqrt{3}, \infty)$ and negative for $x \in (-1/\sqrt{3}, 1/\sqrt{3})$;

- $f''(x)$ is zero for $x = 0$, positive for $x > 0$ and negative for $x < 0$.

From this analysis and Proposition 6.15 we deduce that f is strictly increasing on $(-\infty, -1\sqrt{3}]$, strictly decreasing on $[-1/\sqrt{3}, 1/\sqrt{3}]$ and again strictly increasing on $[1\sqrt{3}, \infty)$; in particular, the stationary points $-1\sqrt{3}$ and $1\sqrt{3}$ are a local maximum and a local minimum for f respectively. Moreover, from Corollary 6.27 it follows that f is concave on $(-\infty, 0]$ and convex on $[0, \infty)$, and that 0 is an inflexion point.

6.5   L'Hôpital Rule

As discussed in Section 3.5, the Algebra of Limits does not allow us to directly determine the limit of the quotient f/g of two functions f and g in the case where both functions tend to 0, or in the case where both functions tend to $\pm\infty$. In such cases, if the two functions are differentiable, it may be possible to determine the limit by considering the quotient f'/g' of the derivatives in place of the original quotient f/g. The precise statement is as follows.

Proposition 6.30 (L'Hôpital Rule)**.** *Let $a, b, c \in \overline{\mathbb{R}}$ be such that $a < b$ and $a \leq c \leq b$. Let $\Omega = (a, b) \setminus \{c\}$, and let $f, g : \Omega \to \mathbb{R}$. Suppose that one of the following conditions hold.*

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

(b) $\lim_{x \to c} f(x) = \ell_1$ *and* $\lim_{x \to c} g(x) = \ell_2$, *where* $\ell_1, \ell_2 \in \{-\infty, \infty\}$.

Assume further that f and g are differentiable and that $g'(x) \neq 0$ for all $x \in \Omega$. If the limit

$$\lim_{x \to c} \frac{f'(x)}{g'(x)} \tag{6.4}$$

exists, then the limit

$$\lim_{x \to c} \frac{f(x)}{g(x)} \tag{6.5}$$

exists too, and the two limits in (6.4) *and* (6.5) *are equal.*

Proof. We only give the proof in the case where $c = a \in \mathbb{R}$ and $\lim_{x \to a} f(x) = \lim_{x \to a} g(x) = 0$. In this case, if we define $f_0 : [a, b) \to \mathbb{R}$ and $g_0 : [a, b) \to \mathbb{R}$ by setting

$$f_0(x) = \begin{cases} f(x) & \text{if } x \in (a, b), \\ 0 & \text{if } x = a, \end{cases} \qquad g_0(x) = \begin{cases} g(x) & \text{if } x \in (a, b), \\ 0 & \text{if } x = a, \end{cases}$$

then, by Proposition 4.6, both f_0 and g_0 are continuous on $[a, b)$ (because both f and g are continuous and $\lim_{x \to a} f(x) = \lim_{x \to a} g(x) = 0$) and differentiable on (a, b) (because f and g are differentiable), and moreover $f_0'(x) = f'(x)$ and $g_0'(x) = g'(x)$ for all $x \in (a, b)$. In particular, $g_0'(x) \neq 0$ for all $x \in (a, b)$.

Let $x \in (a, b)$. By applying the Generalised Mean Value Theorem (Proposition 6.12) to the restrictions $f_0|_{[a,x]}$ and $g_0|_{[a,x]}$, we obtain that there exists $c_x \in (a, x)$ such that

$$\frac{f(x)}{g(x)} = \frac{f_0(x) - f_0(a)}{g_0(x) - g_0(a)} = \frac{f_0'(c_x)}{g_0'(c_x)} = \frac{f'(c_x)}{g'(c_x)}.$$

Note that, since $a < c_x < x$, by the Sandwich Theorem we have that $c_x \to a$ as $x \to a$, and therefore, by Proposition 4.14,

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

as desired. $\square$

Example 6.31. Let us determine the limit

$$\lim_{x \to 0} \frac{x - \sin x}{x^3}.$$

Note that

$$\lim_{x \to 0} (x - \sin x) = \lim_{x \to 0} x^3 = 0,$$

so we cannot directly apply the Algebra of Limits. However, if we define $f, g : \mathbb{R} \setminus \{0\}$ by $f(x) = x - \sin x$ and $g(x) = x^3$, then both f and g are differentiable, and moreover $g'(x) = 3x^2 \neq 0$ for all $x \neq 0$. Furthermore

$$\lim_{x \to 0} \frac{f'(x)}{g'(x)} = \lim_{x \to 0} \frac{1 - \cos x}{3x^2} = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6},$$

by Proposition 4.20 and the Algebra of Limits. Therefore we can apply L'Hôpital rule to conclude that

$$\lim_{x \to 0} \frac{x - \sin x}{x^3} = \lim_{x \to 0} \frac{f(x)}{g(x)} = \lim_{x \to 0} \frac{f'(x)}{g'(x)} = \frac{1}{6}.$$

Remark 6.32. The notable limits discussed in Proposition 4.20 and 4.23 can formally be subsumed under L'Hôpital Rule; for example, since the derivative of sin is cos, and the derivative of the identity function $x \mapsto x$ is the constant 1, from Proposition 6.30 we deduce that

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

However, please note that, in order to apply L'Hôpital Rule to this case, we need to know that the derivative of sin is cos, and this was proved in Proposition 5.26 by using the notable limits of Proposition 4.20. So, from a logical point of view, we cannot prove Proposition 4.20 using L'Hôpital Rule.

Remark 6.33. It is important to observe that, on the basis of L'Hôpital's Rule, we can deduce the equality

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)} \tag{6.6}$$

only <u>provided the limit in the right-hand side exists</u>. Indeed, there exist functions f and g (satisfying the other assumptions of L'Hôpital's Rule) for which the limit in the left-hand side of (6.6) exists, but the limit in the right-hand side does not exist. One example is given by taking $a = 0$ and $f, g : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ defined by

$$f(x) = x^2 \exp(\cos(1/x)), \qquad g(x) = x.$$

Indeed, in this case f and g are infinitely differentiable and it is easily checked that $\lim_{x \to 0} f(x) = \lim_{x \to 0} g(x) = 0$. Moreover

$$\frac{f(x)}{g(x)} = x \exp(\cos(1/x)),$$

whence $|f(x)/g(x)| \le |x|e^1$ and by the Sandwich Theorem we immediately deduce that $\lim_{x \to 0} f(x)/g(x) = 0$. Instead, by using differentiation rules,

$$\frac{f'(x)}{g'(x)} = 2x \exp(\cos(1/x)) + \exp(\cos(1/x)) \sin(1/x),$$

and the limit of the above expression as $x \to 0$ does not exist. We prove this by contradiction. Indeed, assume that the limit exists. Then, by L'Hôpital's Rule it must be equal to zero too. So, by the Algebra of Limits,

$$\lim_{x \to 0} \exp(\cos(1/x)) \sin(1/x) = \lim_{x \to 0} \frac{f'(x)}{g'(x)} - 2 \lim_{x \to 0} x \exp(\cos(1/x))$$
$$= 0 - 2 \cdot 0 = 0. \tag{6.7}$$

Let us now observe that

$$|\sin(1/x)| = |\sin(1/x) \exp(\cos(1/x))| \exp(-\cos(1/x))$$
$$\le |\sin(1/x) \exp(\cos(1/x))| e^1. \tag{6.8}$$

From (6.7), Corollary 3.28 and the Algebra of Limits we deduce that the right-hand side of (6.8) tends to 0 as $x \to 0$; so, by the Sandwich Theorem, the same is true for the left-hand side of (6.8), that is

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

The change of variables $x = 1/y$ (see Proposition 4.14) then implies that $\lim_{y \to \infty} \sin y = 0$; this however contradicts the fact that the limit $\lim_{y \to \infty} \sin y$ does not exist (see Example 3.13).

6.6   Asymptotes

When analysing the behaviour of a function $f : I \to \mathbb{R}$, where $I \subseteq \mathbb{R}$ is an interval[12] of positive length, it is useful to determine (if it exists) the limit

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

---

[12]More generally, if $f : \Omega \to \mathbb{R}$, where $\Omega \subseteq \mathbb{R}$ is not an interval, but $\Omega = I_1 \cup I_2$ for some intervals I_1 and I_2, then one can apply the analysis discussed in this section to each of the restrictions $f|_{I_1}$ and $f|_{I_2}$. Similar considerations apply to unions of three or more intervals.

where $a \in \overline{\mathbb{R}}$ is either endpoint of I. In certain cases, the existence of such limit tells us that the graph of f approaches a line in the xy-plane, called an *asymptote* of the graph of f.

Vertical asymptotes   If $a \in \mathbb{R}$, and

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

for some $\ell \in \{-\infty, +\infty\}$, then we say that the line $x = a$ is a *vertical asymptote* of the graph of f.

Horizontal asymptotes   If $a = \pm\infty$, and

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

for some $\ell \in \mathbb{R}$, then we say that the line $y = \ell$ is a *horizontal asymptote* of the graph of f.

Oblique asymptotes   If $a = \pm\infty$, we say that the line $y = mx + q$ (where $m \in \mathbb{R} \setminus \{0\}$ and $q \in \mathbb{R}$) is an *oblique asymptote* of the graph of f if

$$\lim_{x \to a} (f(x) - (mx + q)) = 0. \tag{6.9}$$

Note that, by the Algebra of Limits, (6.9) is equivalent to

$$\lim_{x \to a} (f(x) - mx) = q; \tag{6.10}$$

moreover, (6.9) implies

$$\lim_{x \to a} \frac{f(x)}{x} = m, \tag{6.11}$$

and

$$\lim_{x \to a} f(x) = \pm\infty, \tag{6.12}$$

since $m \neq 0$. Consequently, in order to determine whether the graph of f has an oblique asymptote, one can follow these steps:

- First, determine whether (6.11) holds for some $m \in \mathbb{R} \setminus \{0\}$.

- If this is the case, then determine whether (6.10) holds for some $q \in \mathbb{R}$.

- If this also happens, then the line $y = mx + q$ is an oblique asymptote for the graph of f.

We remark that, in order to determine the limit in (6.11), one may try and use L'Hôpital's rule; namely, if one can prove that

$$\lim_{x \to a} f(x) = \pm\infty \qquad \text{and} \qquad \lim_{x \to a} f'(x) = m,$$

then from L'Hôpital rule it follows that (6.11) holds. We also remark that the conditions (6.11) and (6.12) are only necessary conditions for the existence of an oblique asymptote, but they are not sufficient in general.[13]

---

[13] The function $f(x) = x + \log x$ satisfies $\lim_{x \to \infty} f(x) = \infty$ and $\lim_{x \to \infty}(f(x)/x) = 1$ (see Proposition 4.24), so (6.11) is satisfied with $m = 1$; however $\lim_{x \to \infty}(f(x) - mx) = \lim_{x \to \infty} \log x = \infty$, so (6.10) does not hold for any $q \in \mathbb{R}$.

6.7 Tangent and normal lines to plane curves, implicit differentiation and polar coordinates

In this section, we discuss how differentiation can be used to determine tangent lines and normal lines to curves in the plane. Recall that the notion of tangent line was already discussed in Remark 5.2 (in connection with the definition of derivative), while the *normal line* to a curve at a point is the straight line passing through that point which is orthogonal to the tangent line.

6.7.1 Curves as graphs

Assume that the curve is given as a graph of a real-valued function of a real variable $f : \Omega \to \mathbb{R}$, that is, the curve has equation

$$y = f(x) \tag{6.13}$$

in the xy-plane. If f is differentiable at $a \in \Omega$, then, according to the discussion in Remark 5.2, the tangent line through the point $(a, f(a))$ has slope $f'(a)$, and therefore its equation is

$$y - f(a) = f'(a)(x - a).$$

Consequently, if $f'(a) \neq 0$, then the normal line has slope $-1/f'(a)$ and equation

$$y - f(a) = -\frac{1}{f'(a)}(x - a);$$

if instead $f'(a) = 0$, then the tangent line is horizontal, so the normal line is vertical and has equation

$$x = a.$$

6.7.2 Curves described via equations

More generally, curves may be described as the set of the points (x, y) of the plane that satisfy an equation of the form

$$F(x, y) = G(x, y), \tag{6.14}$$

where F and G are real-valued functions of two real variables x and y. For example, a circle of radius 1 centred at $(0, 0)$ is described as the set of points (x, y) satisfying the equation

$$x^2 + y^2 = 1. \tag{6.15}$$

Suppose that we are interested in finding the tangent and normal lines to such a curve at a given point (x_0, y_0). One possibility would be to solve the equation (6.14) of the curve for y, so to obtain another description of the curve (or, at least, part of the curve), which has the form (6.13), and then apply the previous discussion. For example, in the case of the circle, the equation (6.15) is equivalent to

$$y = \pm\sqrt{1 - x^2},$$

and one could find tangent and normal lines at (x_0, y_0) by differentiating one of the functions $x \mapsto \sqrt{1 - x^2}$ and $x \mapsto -\sqrt{1 - x^2}$ (according to whether y_0 is

positive or negative). However, for more general equations, finding an explicit solution may not be easy and therefore this method may not be viable.

There is an alternative method to determine tangent and normal lines, called *implicit differentiation*, which does not require to explicitly solve the equation (6.14) for y. The idea is roughly as follows. Assume that the curve (or, at least, a portion of the curve near the point (x_0, y_0)) can be written in the form (6.13) for some (unknown) differentiable function f. Then the equation (6.14) can be turned into an equation for the function f, namely,

$$F(x, f(x)) = G(x, f(x)).$$

Assuming that both sides are differentiable functions of x, we can differentiate them and obtain a new equation:

$$\frac{d}{dx}F(x, f(x)) = \frac{d}{dx}G(x, f(x)).$$

By using differentiation rules, it may be possible[14] to turn the last equation into an equation involving the quantities x, $f(x)$ and $f'(x)$. If we can solve such equation for $f'(x)$, then we obtain the slope of the tangent line as a function of the coordinates $(x, f(x))$ of the point.

In practice, we do not need to introduce a new symbol f for the unknown function when applying this method: it is enough to think of the variable y in the equation as a function of x (so we can use the notation $x, y, \frac{dy}{dx}$ in place of $x, f(x), f'(x)$). In the example of the circle (6.15), if we think of y as a function of x and differentiate both sides of (6.15) with respect to x, we obtain

$$\frac{d}{dx}(x^2 + y^2) = \frac{d}{dx}1,$$

that is, by the chain rule,

$$2x + 2y\frac{dy}{dx} = 0,$$

and by solving for $\frac{dy}{dx}$ we finally obtain

$$\frac{dy}{dx} = -\frac{x}{y}. \tag{6.16}$$

According to this equation, the slope of the tangent line to the circle (6.15) at the point (x_0, y_0) is given by $-x_0/y_0$ (provided $y_0 \neq 0$); consequently the tangent line at (x_0, y_0) has equation

$$y - y_0 = -\frac{x_0}{y_0}(x - x_0)$$

and (if $x_0 \neq 0$) the normal line at (x_0, y_0) has equation

$$y - y_0 = \frac{y_0}{x_0}(x - x_0).$$

---

[14]A precise discussion of this point for general functions F and G would require notions of multivariable analysis (namely, "partial derivatives" for functions of two real variables, and the corresponding "multivariable chain rule"), which are beyond the scope of these lectures. However, in many specific examples (such as the case of the circle discussed in this section), these more advanced notions are not strictly required to apply the method.

6.7.3 Curves in parametric form

A further way in which curves may be described is via a *parametrisation*. In other words, given two real-valued functions $f, g : \Omega \to \mathbb{R}$ of a real variable, we can describe a curve as the set of points $(x, y) \in \mathbb{R}$ of the form

$$\begin{cases} x = f(t) \\ y = g(t) \end{cases} \tag{6.17}$$

for some $t \in \Omega$; in this context, the variable t is also called the *parameter* of the representation (6.17). For example, the unit circle of equation (6.15) can be described in parametric form as

$$\begin{cases} x = \cos t \\ y = \sin t, \end{cases} \tag{6.18}$$

where $t \in \mathbb{R}$ (in fact, it would be enough to take $t \in [0, 2\pi)$, because sine and cosine are 2π-periodic).

In this case, if f and g are both differentiable at a point $t_0 \in \Omega$, then finding the tangent and normal lines to the curve at the point $(f(t_0), g(t_0))$ is particularly simple (provided one of $f'(t_0)$ and $g'(t_0)$ does not vanish). Namely, the tangent line is parallel[15] to the vector $(f'(t_0), g'(t_0)) \in \mathbb{R}^2$, which is sometimes called the *tangent vector* to the curve at the point $(f(t_0), g(t_0))$. In particular, if $f'(t_0) \neq 0$, then the tangent line has slope $g'(t_0)/f'(t_0)$ and equation

$$y - g(t_0) = \frac{g'(t_0)}{f'(t_0)}(x - f(t_0)), \tag{6.19}$$

while (if $g'(t_0) \neq 0$) the normal line has equation

$$y - g(t_0) = -\frac{f'(t_0)}{g'(t_0)}(x - f(t_0)). \tag{6.20}$$

In the example of the circle (6.18), we find that the tangent vector at the point $(\cos t_0, \sin t_0)$ is given by $(-\sin t_0, \cos t_0)$, so the slope of the tangent line is $-\cos t_0 / \sin t_0$. Note that this value is consistent with (6.16).

One way of thinking of the parametrisation (6.17) is as the description of the motion of a point in the plane, where x and y represent the coordinates of the point and the parameter t represents time. In this case, the tangent vector

$$\left(\left. \frac{dx}{dt} \right|_{t=t_0}, \left. \frac{dy}{dt} \right|_{t=t_0} \right)$$

---

[15]To justify this, one may observe that the line passing through the points $(f(t_0), g(t_0))$ and $(f(t_1), g(t_1))$ has slope

$$\frac{g(t_1) - g(t_0)}{f(t_1) - f(t_0)} = \frac{g(t_1) - g(t_0)}{t_1 - t_0} \frac{t_1 - t_0}{f(t_1) - f(t_0)};$$

by taking the limit as $t_1 \to t_0$, we obtain that the slope of the tangent line at $(f(t_0), g(t_0))$ is $g'(t_0)/f'(t_0)$, that is, the slope of the vector $(f'(t_0), g'(t_0))$. To be precise, this discussion only works when $f'(t_0) \neq 0$ (note that, in this case, $\frac{f(t_1)-f(t_0)}{t_1-t_0} \neq 0$ for all t_1 sufficiently close to t_0, so $f(t_1) - f(t_0) \neq 0$); in the case where $f'(t_0) = 0$ and $g'(t_0) \neq 0$, a similar discussion applies with the roles of the coordinates x and y reversed.

can be given the physical interpretation of the *velocity vector* at time t_0; according to this interpretation, the derivatives $\frac{dx}{dt}\big|_{t=t_0}$ and $\frac{dy}{dt}\big|_{t=t_0}$ are the horizontal and vertical components of the velocity (cf. Remark 5.3).

6.7.4 Polar coordinates

The discussion so far exploited the usual Cartesian coordinates (x, y) to represent points of the plane. However, other coordinate systems can be used.

For example, when using *polar coordinates*, a point P of the plane is described by two real numbers r and θ, where $r \in [0, \infty)$ is the distance of P from the origin O of the plane, while $\theta \in \mathbb{R}$ is the (oriented) angle between the positive x-axis and the vector $\overrightarrow{OP}$. It is not difficult to see that a point with polar coordinates (r, θ) has Cartesian coordinates (x, y) given by

$$\begin{cases} x = r \cos \theta \\ y = r \sin \theta. \end{cases} \tag{6.21}$$

Note that the value of r is uniquely determined by the point, since

$$r = \sqrt{x^2 + y^2}.$$

The value of θ, instead, is not uniquely determined. If the point is the origin (that is, $x = y = 0$, or, equivalently, $r = 0$), then any value of $\theta \in \mathbb{R}$ can be used. Away from the origin (that is, when $r > 0$), the coordinate θ is determined up to multiples of 2π (since sine and cosine are 2π-periodic); one could make a unique choice[16] by requiring that θ must be chosen in an interval of length 2π, such as $[-\pi, \pi)$ or $[0, 2\pi)$, however sometimes it is more convenient not to restrict the range of θ (see, e.g., Example 6.35 below).

Polar coordinates can be used to describe curves in the plane, similarly to what is done with Cartesian coordinates. For example, a curve could be given in polar coordinates (r, θ) by an equation of the form

$$r = f(\theta), \tag{6.22}$$

expressing the coordinate r as a function of the coordinate θ; here $f : \Omega \to \mathbb{R}$ is some real-valued function of a real variable. In this case, using (6.21) we can turn the equation (6.22) into the representation

$$\begin{cases} x = f(\theta) \cos \theta \\ y = f(\theta) \sin \theta, \end{cases}$$

which can be thought of as a parametric representation (in Cartesian coordinates) of the curve, where θ plays the role of the parameter; the study of the tangent and normal lines to the curve then reduces to the previous discussion of parametric curves.

---

[16]The reader should be careful when trying to derive a formula for θ in terms of x and y. For example, the equations (6.21) imply that $\tan \theta = y/x$ whenever $x \neq 0$. However the last equation is not enough to determine θ up to multiples of 2π, because tan is π-periodic; in particular, it would be wrong in general to conclude that $\theta = \arctan(y/x)$, unless we already know that $\theta \in (-\pi/2, \pi/2)$.

More generally, if the curve is parametrised in polar coordinates (r, θ) as

$$\begin{cases} r = f(t) \\ \theta = g(t), \end{cases}$$

for some functions $f, g : \Omega \to \mathbb{R}$, then using (6.21) we can again obtain a parametric representation in Cartesian coordinates (x, y), namely

$$\begin{cases} x = f(t) \cos(g(t)) \\ y = f(t) \sin(g(t)), \end{cases}$$

which can be studied as in Section 6.7.3.

Example 6.34. The circle of radius 1 centred at the origin is described, in polar coordinates (r, θ), by the equation

$$r = 1$$

(indeed, the points of the circle are exactly those points whose distance from the origin is equal to 1). By using the change-of-variable formula (6.21), the above description immediately yields the parametric representation (6.18) of the circle that we used previously.

Example 6.35. The equation

$$r = e^{\theta}$$

describes, in polar coordinates (r, θ), a curve that is called *logarithmic spiral*. To study this curve, by using the change-of-variable formula (6.21) we obtain the parametric representation

$$\begin{cases} x = e^{\theta} \cos \theta \\ y = e^{\theta} \sin \theta \end{cases}$$

in Cartesian coordinates (x, y), where $\theta \in \mathbb{R}$ now plays the role of the parameter. We note that the curve passes through the point P with Cartesian coordinates $(1, 0)$, which corresponds to $\theta = 0$. By differentiating the expressions for x and y with respect to θ,

$$\frac{dx}{d\theta} = e^{\theta}(\cos \theta - \sin \theta), \qquad \frac{dy}{d\theta} = e^{\theta}(\sin \theta + \cos \theta),$$

we obtain formulas for the components of the tangent vector at any point of the curve. In particular, by evaluating at $\theta = 0$, we obtain that the tangent vector at the point $P = (1, 0)$ is the vector $(1, 1)$; consequently (see formulas (6.19) and (6.20)), the tangent line at P has equation

$$y = x - 1,$$

while the normal line at P has equation

$$y = -(x - 1).$$