MATHEMATICAL ANALYSIS 1

Irena Penev

Computer Science Institute
of Charles University (IÚUK)

INFORMATICKÝ ÚSTAV
UNIVERZITY KARLOVY
Matematicko–fyzikální fakulta
Univerzita Karlova

Contents

Chapter 1

Number systems: rational and real numbers

Notation: In what follows, we will use the following notation:

- $\mathbb{N}$ is the set of all natural numbers (positive integers);
- $\mathbb{N}_0$ is the set of all non-negative integers;
- $\mathbb{Z}$ is the set of all integers;
- $\mathbb{Q}$ is the set of all rational numbers;
- $\mathbb{R}$ is the set of all real numbers;
- $\mathbb{C}$ is the set of all complex numbers.

1.1 A brief introduction to "complete ordered fields"

We begin by reviewing a couple of definitions. You have seen "fields" in Linear Algebra, and you have seen "strict total orders" in Discrete Math. The formal definitions are as follows.

Fields. A *field* is an ordered triple $(\mathbb{F}, +, \cdot)$, where $\mathbb{F}$ is a set, and $+$ and $\cdot$ are binary operations on $\mathbb{F}$ (i.e. functions from $\mathbb{F} \times \mathbb{F}$ to $\mathbb{F}$), called *addition* and *multiplication*, respectively, satisfying the following axioms:

1. [**Associativity of addition and multiplication**] addition and multiplication are associative, that is, for all $a, b, c \in \mathbb{F}$, we have that $a + (b + c) = (a + b) + c$ and $a \cdot (b \cdot c) = (a \cdot b) \cdot c$;
2. [**Commutativity of addition and multiplication**] addition and multiplication are commutative, that is, for all $a, b \in \mathbb{F}$, we have that $a + b = b + a$ and $a \cdot b = b \cdot a$;

3. [**Additive and multiplicative identity**] there exist distinct elements $0_{\mathbb{F}}, 1_{\mathbb{F}} \in \mathbb{F}$ such that for all $a \in \mathbb{F}$, $a + 0_{\mathbb{F}} = a$ and $a \cdot 1_{\mathbb{F}} = a$; $0_{\mathbb{F}}$ is called the *additive identity* of $\mathbb{F}$, and $1_{\mathbb{F}}$ is called the *multiplicative identity* of $\mathbb{F}$;
4. [**Additive inverses**] for every $a \in \mathbb{F}$, there exists an element in $\mathbb{F}$, denoted by $-a$ and called the *additive inverse* of a, such that $a + (-a) = 0_{\mathbb{F}}$;
5. [**Multiplicative inverses**] for all $a \in \mathbb{F} \setminus \{0_{\mathbb{F}}\}$, there exists an element in $\mathbb{F}$, denoted by a^{-1} and called the *multiplicative inverse* of a, such that $a \cdot a^{-1} = 1_{\mathbb{F}}$;
6. [**Distributivity of multiplication over addition**] multiplication is distributive over addition, that is, for all $a, b, c \in \mathbb{F}$, we have that $a \cdot (b+c) = (a \cdot b) + (a \cdot c)$.

Notation: Usually, we write simply "field $\mathbb{F}$" rather than "field $(\mathbb{F}, +, \cdot)$." When $\mathbb{F}$ is clear from context, we write simply 0 and 1 instead of $0_{\mathbb{F}}$ and $1_{\mathbb{F}}$, respectively.

Example 1.1.1. *All the following are fields:*

- $(\mathbb{Q}, +, \cdot)$;
- $(\mathbb{R}, +, \cdot)$;
- $(\mathbb{C}, +, \cdot)$.

*However, the following are **not** fields:*

- $(\mathbb{N}, +, \cdot)$;
- $(\mathbb{N}_0, +, \cdot)$;
- $(\mathbb{Z}, +, \cdot)$.

Strict total orders. A *strict total order* on a non-empty set A is a binary relation $<$ on A that satisfies the following two axioms:

1. [**Transitivity**] For all $a, b \in A$, if $a < b$ and $b < c$, then $a < c$.
2. [**Trichotomy**] For all $a, b \in A$, exactly one of $a < b$, $a = b$, and $b < a$ is true.

Ordered fields. An *ordered field* is a field $\mathbb{F}$ with a strict total order $<$ such that:

- if $a < b$, then $a + c < b + c$;
- if $0 < a$ and $0 < b$, then $0 < ab$.

We define $\leq$ as follows: $a \leq b$ if $a < b$ or $a = b$. Furthermore, we write $a > b$ when $b < a$, and we write $a \geq b$ when $b \leq a$.

Remark: It is not hard to show that, in an ordered field, all the usual algebraic properties of the relations $<$ and $\leq$ hold. We omit the proof.

Fact 1.1.2. *$\mathbb{Q}$ and $\mathbb{R}$ are ordered fields (under the usual $<$ relation). However, $\mathbb{C}$ is **not** an ordered field.*

Upper and lower bounds. If $\mathbb{F}$ is an ordered field, $S \subseteq \mathbb{F}$, $x \in \mathbb{F}$, then

- x is called an *upper bound* for S if for all $s \in S$, we have $s \leq x$;
- x is called the *least upper bound* (or *supremum*) for S if x is an upper bound for S, and every upper bound y for S satisfies $x \leq y$.
- x is called a *lower bound* for S if for all $s \in S$, we have $x \leq s$;
- x is called the *greatest lower bound* (or *infimum*) for S if x is a lower bound for S, and every lower bound y for S satisfies $y \leq x$.

A subset of $\mathbb{F}$ is *bounded above* if it has an upper bound, and it is *bounded below* if it has a lower bound. A subset of $\mathbb{F}$ is *bounded* if it is both bounded above and bounded below.

Remarks:

- An upper or lower bound for a set S may, but need not, belong to S. In particular, the supremum or infimum of a set S (if it exists) may, but need not, belong to S.
- A subset of an ordered field does not necessarily have an upper or a lower bound. For instance, in the ordered field $\mathbb{R}$, the set $\mathbb{R}$ itself has neither an upper nor a lower bound.
- If a set does have a supremum, then that supremum is unique, and the same goes for the infimum (see Proposition 1.1.4 below).

Example 1.1.3. *In* $\mathbb{R}$*:*

- *the set* $[-2, 3)$ *has both a supremum (namely* 3*) and an infimum (namely* -2*);*
- *the set* $(-\infty, 4]$ *has a supremum (namely* 4*), but no infimum;*
- *the set* $(2, \infty)$ *has no supremum, but does have an infimum (namely* 2*);*
- *the set* $(2, 3) \cup (5, 7)$ *has both a supremum (namely* 7*) and an infimum (namely* 2*).*

Proposition 1.1.4. *Let* $\mathbb{F}$ *be an ordered field, and let* $S \subseteq \mathbb{F}$*. Then* S *has at most one supremum and at most one infimum.*[1]

Proof. Suppose that x and y are suprema of S; we must show that $x = y$. Since x is an supremum of S and y is an upper bound of S, we have that $x \leq y$. Similarly, since y is a supremum of S and x is an upper bound of S, we have that $y \leq x$. Since both $x \leq y$ and $y \leq x$ hold, we see that $x = y$. This proves that S does indeed have at most one supremum. Analogously, S has at most one infimum. □

---

[1] Once again, supremum of S, if it exists, may or may not belong to S, and the same applies to the infimum.

Notation: If $\mathbb{F}$ is an ordered field and the set $S \subseteq \mathbb{F}$ has a supremum, then by Proposition 1.1.4, that supremum is unique, and we denote it by $\sup(S)$. Similarly, if S has an infimum, then by Proposition 1.1.4, that infimum is unique, and we denote it by $\inf(S)$.

Complete ordered fields. An ordered field $\mathbb{F}$ is *complete* if every non-empty subset of $\mathbb{F}$ that is bounded above has the least upper bound (i.e. supremum).

Remark: It is not hard to show that in a complete ordered field, every non-empty set that is bounded below has the greatest lower bound (i.e. infimum).

Fact 1.1.5. *$\mathbb{R}$ is a complete ordered field. In fact, up to "isomorphism" (essentially, a renaming of elements), $\mathbb{R}$ is the* ***only*** *complete ordered field.*

Remark: In particular, $\mathbb{Q}$ is **not** a complete ordered field.

We omit the proof of Fact 1.1.5. To give a formal proof, we would need to fully formalize the field of real numbers (e.g. using "Dedekind cuts" or equivalence classes of "Cauchy sequences" of rational numbers), which is beyond the scope of this course. The point is that when we speak of a "complete ordered field," we are in fact referring to $\mathbb{R}$ (equipped with the usual addition, multiplication, and less-than relation).

Remark: The fact that $\mathbb{R}$ is a complete ordered field is essential for formally defining functions such as $\sin, \cos, \exp$ (and many others).

Example 1.1.6. *Note that the set $S = \{x \in \mathbb{Q} \mid x^2 \leq 2\}$ is bounded in $\mathbb{Q}$ (and therefore in $\mathbb{R}$ as well). For example, 2 is an upper bound of S (both in $\mathbb{Q}$ and in $\mathbb{R}$). In $\mathbb{R}$, $\sqrt{2}$ is the least upper bound (i.e. supremum) of S. In $\mathbb{Q}$, S does not have the least upper bound. Here, we are using the well-known fact that $\sqrt{2}$ is an irrational number.*

Actually, let us formally prove that $\sqrt{2}$ is irrational!

Theorem 1.1.7. *$\sqrt{2}$ is irrational.*

Proof. Suppose otherwise, and set $\sqrt{2} = \frac{p}{q}$, where $p \in \mathbb{Z}$, $q \in \mathbb{N}$, and p, q are relatively prime (i.e. they have no common divisor greater than 1). Then $p^2 = 2q^2$, which implies that $2 \mid p^2$. Since 2 is prime, it follows that $2 \mid p$. Thus, there exists some $r \in \mathbb{Z}$ such that $p = 2r$. Now $4r^2 = 2q^2$, and it follows that $2r^2 = q^2$. Now $2 \mid q^2$, and so since 2 is prime, we see that $2 \mid q$. But now 2 is a common divisor of p and q, contrary to the fact that p and q are relatively prime. $\square$

1.2 Cardinality: comparing infinities

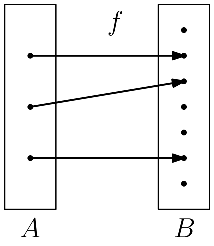
We know that

$$\mathbb{N} \subsetneqq \mathbb{Z} \subsetneqq \mathbb{Q} \subsetneqq \mathbb{R} \subsetneqq \mathbb{C}.$$

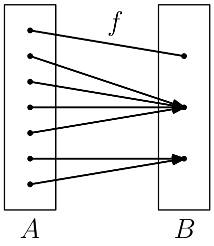
So, in a sense, there are "more" integers than there are natural numbers, "more" rational numbers than integers, "more" real numbers than real numbers, and "more" complex numbers than real numbers. However, there is another way of comparing the sizes of two sets: using bijections. Let us first recall the definition of a bijection.

Injections, surjections, and bijections. A function $f : A \to B$ (where A and B are some sets) is:

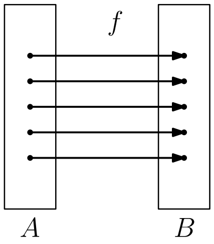
- *one-to-one* (or *injective*, or an *injection*) if for all $a_1, a_2 \in A$ such that $a_1 \neq a_2$, we have that $f(a_1) \neq f(a_2)$;[2]



- *onto* (or *surjective*, or a *surjection*) if for all $b \in B$, there exists some $a \in A$ such that $f(a) = b$;



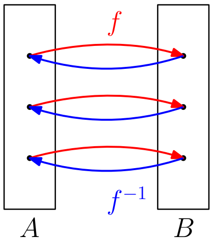
- *bijective* (or a *bijection*) if it is both one-to-one and onto.



Remark: Recall that if $f : A \to B$ is a bijection, then it has an *inverse function* $f^{-1} : B \to A$ such that for all $b \in B$, $f^{-1}(b)$ is the unique element $a \in A$ such that $f(a) = b$. In other words, if $f : A \to B$ is a bijection, then for all $a \in A$ and $b \in B$, we have that $f(a) = b$ if and only if $f^{-1}(b) = a$. Clearly, the inverse of a bijection is also a bijection.

---

[2] Equivalently: if for all $a_1, a_2 \in A$ such that $f(a_1) = f(a_2)$, we have that $a_1 = a_2$.



Cardinality. Sets A and B are said to have *the same cardinality* if there exists a bijection $f : A \to B$, and in this case, we write $|A| = |B|$.

Remark: Thus, we can think of two sets as being of "the same size" (or having "the same number of elements") if we can set up a bijection between them.

Proposition 1.2.1. *All the following hold:*

(a) for all sets A, $|A| = |A|$;

(b) for all sets A and B, if $|A| = |B|$, then $|B| = |A|$;

(c) for all sets A, B, and C, if $|A| = |B|$ and $|B| = |C|$, then $|A| = |C|$.

Proof. Fix sets A, B, and C.

For (a), we simply observe that the identity function on A is a bijection.[3]

For (b), we observe that if $f : A \to B$ is a bijection, then so is $f^{-1} : B \to A$.

For (c), we observe that if $f_1 : A \to B$ and $f_2 : B \to C$ are bijections, then $f_2 \circ f_1 : A \to C$ is also a bijection. ☐

Proposition 1.2.2. $|\mathbb{N}| = |\mathbb{Z}|$.

Proof. We define the function $f : \mathbb{N} \to \mathbb{Z}$ by setting

$$f(n) \quad = \quad \begin{cases} (n-1)/2 & \text{if } n \text{ is odd} \\ -n/2 & \text{if } n \text{ is even} \end{cases}$$

for all $n \in \mathbb{N}$. Thus, we have the following:

- $f(2) = -1$;
- $f(4) = -2$;
- $f(6) = -3$;
- ...

- $f(1) = 0$;
- $f(3) = 1$;
- $f(4) = 2$;
- $f(7) = 3$;
- ...

---

[3] The *identity function* on A is the function $\mathrm{Id}_A : A \to A$ given by $f(a) = a$ for all $a \in A$.

and it is not difficult to formally check that f is a bijection.[4] This proves that $|\mathbb{N}| = |\mathbb{Z}|$. $\square$

Given sets A and B, we say that the cardinality of A is *no greater than* the cardinality of B, and we write $|A| \leq |B|$ if there exists a one-to-one function $f : A \to B$.

Notation: For sets A and B:

- we write $|A| \geq |B|$ if $|B| \leq |A|$;
- we write $|A| < |B|$ if $|A| \leq |B|$ and $|A| \neq |B|$ (i.e. there exists a one-to-one function from A to B, but there is no bijection between A and B);
- we write $|A| > |B|$ if $|B| < |A|$.

Remark: Note that if sets A, B, and C satisfy $|A| \leq |B|$ and $|B| \leq |C|$, then they also satisfy $|A| \leq |C|$. Indeed, if $f_1 : A \to B$ and $f_2 : B \to C$ are one-to-one functions, then $f_2 \circ f_1 : A \to C$ is also a one-to-one function.

Cantor–Schröder–Bernstein theorem. *If sets A and B satisfy $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$.*

Proof. Omitted. $\square$

Remark: The Cantor–Schröder–Bernstein theorem may seem obvious, but it is in fact not! Fully spelled out, it states the following:

> For all sets A and B, if there exist one-to-one functions $f_1 : A \to B$ and $f_2 : B \to A$, then there exists a bijection $f : A \to B$.

The statement is indeed true, but the proof is beyond the scope of this course.

Remark: It is possible to prove that for all sets A and B, either $|A| \leq |B|$ or $|B| \leq |A|$. However, the proof (which we omit) is not easy and uses the so called "Axiom of Choice."

Proposition 1.2.3. *For all sets A and B, if $A \subseteq B$, then $|A| \leq |B|$.*

Proof. Let A and B be sets such that $A \subseteq B$. Then the function $f : A \to B$ given by $f(a) = a$ for all $a \in A$ is one-to-one, and consequently, $|A| \leq |B|$. $\square$

Remark: Since

$$\mathbb{N} \subsetneqq \mathbb{Z} \subsetneqq \mathbb{Q} \subsetneqq \mathbb{R} \subsetneqq \mathbb{C},$$

Proposition 1.2.3 guarantees that

$$|\mathbb{N}| \leq |\mathbb{Z}| \leq |\mathbb{Q}| \leq |\mathbb{R}| \leq |\mathbb{C}|,$$

---

[4] Check this!

and moreover, by Proposition 1.2.2, we further have that $|\mathbb{N}| = |\mathbb{Z}|$.

Countable and uncountable sets. A set S is *countable* if one of the following holds:

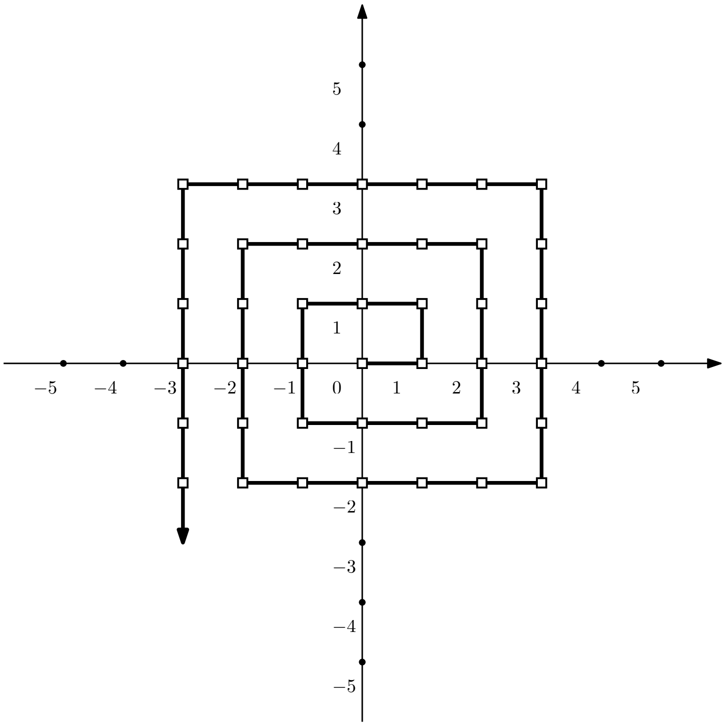
- S is finite;
- there exists a bijection $f : \mathbb{N} \to S$.

A set is *denumerable* (or *countably infinite*) if it is both countable and infinite. So, if S is infinite, then it is countable if and only if its members can be enumerated as $s_1, s_2, s_3, \dots$.[5] A set is *uncountable* if it is not countable.

Remark: By Proposition 1.2.2, $\mathbb{Z}$ is denumerable (i.e. countably infinite).

Theorem 1.2.4. *$\mathbb{Z} \times \mathbb{Z}$ and $\mathbb{Q}$ are both denumerable.*

Proof (slightly informal). The picture below shows that $\mathbb{Z} \times \mathbb{Z}$ can be enumerated as $(p_1, q_1), (p_2, q_2), (p_3, q_3), \dots$. So, $\mathbb{Z} \times \mathbb{Z}$ is countable.



Rational numbers correspond to $\mathbb{Z} \times \mathbb{Z}$ in a natural way. Indeed, a fraction $\frac{p}{q}$ corresponds to the ordered pair (p, q). Except that there's a problem: each rational corresponds to more than one ordered pair of integers (because e.g. $\frac{p}{q} = \frac{2p}{2q}$).

---

[5] Here, the elements, $s_1, s_2, s_3, \dots$ are required to be pairwise distinct, and all the elements of S must appear on the list $s_1, s_2, s_3, \dots$.

Furthermore, the denominator of a fraction can never be zero. We deal with this as follows. First, we take our list of ordered pairs of integers, and we delete from it all pairs (p, q) where $q \leq 0$. Then, we remove all ordered pairs (p, q) where p and q are not relatively prime. Now each rational corresponds to exactly one ordered pair on the remaining list. This proves that $\mathbb{Q}$ is countable. $\square$

Theorem 1.2.5. $\mathbb{R}$ *is uncountable.*

Cantor's diagonal proof. Suppose otherwise. Then in particular, the interval $[0, 1]$ is countable. We now enumerate the members of $[0, 1]$ as follows:[6]

$$\begin{aligned} x_1 &= 0.x_{1,1}x_{1,2}x_{1,3}x_{1,4}\dots \\ x_2 &= 0.x_{2,1}x_{2,2}x_{2,3}x_{2,4}\dots \\ x_3 &= 0.x_{3,1}x_{3,2}x_{3,3}x_{3,4}\dots \\ x_4 &= 0.x_{4,1}x_{4,2}x_{4,3}x_{4,4}\dots \\ &\vdots \end{aligned}$$

We now create a number $a = 0.a_1a_2a_3a_4\dots$ as follows. For each $i \in \mathbb{N}$, we set

$$a_i = \begin{cases} 5 & \text{if} \quad x_{i,i} \neq 5 \\ 6 & \text{if} \quad x_{i,i} = 5 \end{cases}$$

But now $a \in [0, 1]$, and it does not appear on our list (sequence) $x_1, x_2, x_3, \dots$, a contradiction. So, $\mathbb{R}$ is uncountable. $\square$

Remark: It can be shown (but we omit the proof) that $|\mathbb{R}| = |\mathbb{C}|$. This, combined with Propositions 1.2.3 and 1.2.2 and Theorems 1.2.4 and 1.2.5, yield the following:

$$|\mathbb{N}| \;=\; |\mathbb{Z}| \;=\; |\mathbb{Q}| \;<\; |\mathbb{R}| \;=\; |\mathbb{C}|.$$

Density of rationals in the reals. Theorems 1.2.4 and 1.2.5 essentially imply that there are "many more" reals than rationals. We do, however, have the following.

Fact 1.2.6. $\mathbb{Q}$ *is* dense *in* $\mathbb{R}$*, i.e. for all* $a, b \in \mathbb{R}$ *such that* $a < b$*, there exists some* $c \in \mathbb{Q}$ *such that* $a < c < b$.

Remark: We omit a formal proof of Fact 1.2.6, but we note that this fact is the reason why we are able to approximate real numbers by rationals (with an arbitrarily small error).

---

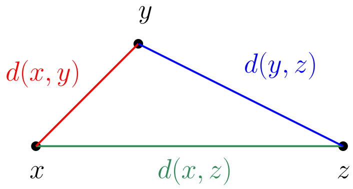
[6] Note that $1 = 0.9999999\dots$.

1.3 A (very) brief introduction to metric spaces

A *metric space* is an ordered pair (M, d), where M is a non-empty set, and $d : M \times M \to \mathbb{R}$ is a *metric* on M, i.e. a function satisfying the following properties:

- for all $x, y \in M$, $d(x, y) \geq 0$,
- for all $x, y \in M$, $d(x, y) = 0$ if and only if $x = y$;
- for all $x, y \in M$, $d(x, y) = d(y, x)$;
- for all $x, y, z \in M$, $d(x, z) \leq d(x, y) + d(y, z)$.

The inequality from the third bullet point is referred to as the *triangle inequality*.



The discrete metric. For any non-empty set M, the function $d : M \times M$ given by

$$d(x, y) \quad = \quad \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}$$

for all $x, y \in M$ is a metric, called the *discrete metric*.

The Euclidean metric. Each of $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ can be turned into a metric space simply by setting $d(x, y) = |x - y|$ for all x and y.

As a matter of fact, $\mathbb{R}^d$ can also be turned into a metric space by setting $d(x, y) = |x - y|$, where for a vector

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_d \end{bmatrix},$$

we have $|\mathbf{v}| = \sqrt{v_1^2 + \cdots + v_d^2}$. This metric is called the *Euclidean metric*.

We will soon start studying sequences. We will develop the theory of sequences for $\mathbb{R}$. However, many (most) of the properties of real sequences are fully generalizable to sequences in general metric spaces.

For future reference, we give the following inequalities.

The Triangle Inequality. *All $x, y \in \mathbb{R}$ satisfy $|x + y| \leq |x| + |y|$.*

Corollary 1.3.1. *All $x, y \in \mathbb{R}$ satisfy $|x - y| \geq |x| - |y|$.*

Proof. Fix $x, y \in \mathbb{R}$. The triangle inequality applied to $x - y$ and y implies that $\underbrace{|(x - y) + y|}_{=|x|} \leq |x - y| + |y|$, and it follows that $|x - y| \geq |x| - |y|$. □

1.4 Bernoulli's inequality

Bernoulli's inequality. *For all integers* $n \geq 0$ *and all real numbers* $x \geq -1$*, we have* $(1+x)^n \geq 1+nx$.

Proof. We fix a real number $x \geq -1$, and we proceed by induction on n.

Base case: For $n = 0$, we have

$$(1+x)^0 = 1 = 1 + 0 \cdot x.$$

Induction step: Fix a non-negative integer n, and assume inductively that

$$(1+x)^n \geq 1+nx.$$

We must show that

$$(1+x)^{n+1} \geq 1+(n+1)x.$$

We now compute:

$$\begin{array}{rcll}
(1+x)^{n+1} & = & (1+x)^n(1+x) & \\
 & \geq & (1+nx)(1+x) & \text{by the induction hypothesis} \\
 & & & \text{and the fact that } 1+x \geq 0 \\
 & = & 1+(n+1)x+nx^2 & \\
 & \geq & 1+(n+1)x. &
\end{array}$$

This completes the induction. □

Chapter 2

Sequences

2.1 Limits of sequences: definition and examples

A *sequence* of real numbers is any function $a : \mathbb{N} \to \mathbb{R}$. By convention, we write a_n instead of $a(n)$. We denote sequences by $\{a_n\}_{n=1}^{\infty}$, by $\{a_n\}_{n \in \mathbb{N}}$, or simply by $a_1, a_2, a_3, \dots$.

Remark: One can also speak of sequences of rational numbers, complex numbers, vectors in $\mathbb{R}^d$, etc. However, we will work almost exclusively with sequences of real numbers.

We now define the "limit" of a sequence.[1] Intuitively, L is the limit of a sequence $\{a_n\}_{n=1}^{\infty}$ if, when n is very large, a_n is very close to L. Let us now formalize this.

The limit of a sequence. We say that a sequence $\{a_n\}_{n=1}^{\infty}$ of real numbers *converges* to a real number L provided that the following holds:

> For all real numbers $\varepsilon > 0$, there exists some $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$: if $n \geq N$, then $|a_n - L| < \varepsilon$.

Under such circumstances, we say that L is the *limit* of the sequence $\{a_n\}_{n=1}^{\infty}$, and we write

$$L = \lim_{n \to \infty} a_n,$$

or

$$a_n \to L \quad \text{as} \quad n \to \infty.$$

A sequence is *convergent* (or *converges*) if it has a limit. Otherwise, it is *divergent* (or *diverges*).

Remark: Note that

$$\begin{aligned} |a_n - L| < \varepsilon \qquad &\iff \qquad L - \varepsilon < a_n < L + \varepsilon \\ &\iff \qquad a_n \in (L - \varepsilon, L + \varepsilon). \end{aligned}$$

---

[1] **Warning:** Not all sequences have limits!

Thus, informally, "$L = \lim_{n\to\infty} a_n$" means that no matter how small we choose our real number $\varepsilon > 0$, at some point, the a_n's all start landing in the open interval $(L-\varepsilon, L+\varepsilon)$. Or, more formally, no matter how small we choose our $\varepsilon > 0$, we can find some positive integer N, so that, with the possible exception of $a_1, \dots, a_N$ (the first N terms of our sequence), all the a_n's belong to the interval open interval $(L-\varepsilon, L+\varepsilon)$.

Remark: It turns out that if a sequence converges, then its limit is unique. We will prove this (see Theorem 2.2.1), but first, let us take a look at a couple of examples.

Example 2.1.1. *Show that* $\lim_{n\to\infty} \frac{1}{n} = 0$.

Solution. Fix $\varepsilon > 0$. Let $N \in \mathbb{N}$ be such that $N > \frac{1}{\varepsilon}$. (Thus, $\frac{1}{N} < \varepsilon$.) Fix $n \in \mathbb{N}$ such that $n \geq N$. Then

$$\left|\tfrac{1}{n} - 0\right| \;=\; \tfrac{1}{n} \;\leq\; \tfrac{1}{N} \;<\; \varepsilon.$$

This proves that $\lim_{n\to\infty} \frac{1}{n} = 0$. □

Example 2.1.2. *Show that* $\lim_{n\to\infty} (1 + \frac{1}{n^2}) = 1$.

Solution. Fix $\varepsilon > 0$. Let $N \in \mathbb{N}$ be such that $N > \frac{1}{\sqrt{\varepsilon}}$.[2] Note that this implies that $\frac{1}{N^2} < \varepsilon$. Now, fix $n \in \mathbb{N}$ such that $n \geq N$. We now have the following:

$$\begin{aligned} |(1+\tfrac{1}{n^2}) - 1| &= \tfrac{1}{n^2} && \text{because } \tfrac{1}{n^2} > 0 \\ &\leq \tfrac{1}{N^2} && \text{because } n \geq N > 0 \\ &< \varepsilon. \end{aligned}$$

This proves that $\lim_{n\to\infty} (1 + \frac{1}{n^2}) = 1$. □

Example 2.1.3. *Show that* $\lim_{n\to\infty} (1 + (-\frac{1}{2})^n) = 1$.

Solution. Fix $\varepsilon > 0$. Let $N \in \mathbb{N}$ be such that $N > \log_2(\frac{1}{\varepsilon})$. (Thus, $\frac{1}{2^N} < \varepsilon$.) Fix $n \in \mathbb{N}$ such that $n \geq N$. Then

$$\left|\Big(1 + (-\tfrac{1}{2})^n\Big) - 1\right| = \tfrac{1}{2^n} \leq \tfrac{1}{2^N} < \varepsilon.$$

This proves that $\lim_{n\to\infty} (1 + (-\frac{1}{2})^n) = 1$. □

---

[2] This footnote is **not** a formal part of the proof (think of it as being on "scratch paper"), but here's how we "guessed" the correct value of N. Our goal is to get $|(1 + \frac{1}{n^2}) - 1| < \varepsilon$. This is equivalent to $\frac{1}{n^2} < \varepsilon$, and that, in turn is equivalent to $n > \frac{1}{\sqrt{\varepsilon}}$. Thus, we should choose $N \in \mathbb{N}$ such that $N > \frac{1}{\sqrt{\varepsilon}}$.

We now take a look at a couple of examples of divergent sequences, i.e. sequences that do **not** have a limit. By definition, a sequence $\{a_n\}_{n=1}^{\infty}$ of real number diverges if the following is satisfied:

> For all $L \in \mathbb{R}$, there exists some real number $\varepsilon > 0$ such that for all $N \in \mathbb{N}$, there exists some $n \in \mathbb{N}$ such that $n \geq N$ and $|a_n - L| \geq \varepsilon$.

In other words (informally), $\{a_n\}_{n=1}^{\infty}$ diverges if no matter which L we choose, we can find some small enough $\varepsilon > 0$ such that infinitely many a_n's land outside of the interval $(L - \varepsilon, L + \varepsilon)$.

Example 2.1.4. *Prove that the sequence* $\{\sqrt{n}\}_{n=1}^{\infty}$ *diverges.*

Remark: Clearly, $\sqrt{n}$ gets very large (arbitrarily large) as n gets very large. So, the sequence $\{\sqrt{n}\}_{n=1}^{\infty}$ "diverges to infinity." However, the goal is to prove formally that a limit does not exist.

Solution. We must prove the following statement:

> For all $L \in \mathbb{R}$, there exists some $\varepsilon > 0$ such that for all $N \in \mathbb{N}$, there exists some $n \in \mathbb{N}$ such that $n \geq N$ and $|\sqrt{n} - L| \geq \varepsilon$.

Fix an arbitrary $L \in \mathbb{R}$. Next, fix some $\varepsilon > 0$.[3] Fix an arbitrary $N \in \mathbb{N}$. Now, choose $n \in \mathbb{N}$ so that $n \geq \max\{N, (L+\varepsilon)^2\}$. Clearly, $n \geq N$. Our goal is to show that $|\sqrt{n} - L| \geq \varepsilon$.

By construction, $n \geq (L+\varepsilon)^2$, and so (since $n \geq 0$) we have that $\sqrt{n} \geq |L + \varepsilon| \geq L + \varepsilon$, and consequently, $\sqrt{n} - L \geq \varepsilon$, which in turn implies that

$$|\sqrt{n} - L| \geq \sqrt{n} - L \geq \varepsilon.$$

This proves that $\{\sqrt{n}\}_{n=1}^{\infty}$ diverges, which is what we needed to show. □

Example 2.1.5. *Prove that the sequence* $\{(-1)^n\}_{n=1}^{\infty}$ *diverges.*

Remark: Clearly, the sequence $\{(-1)^n\}_{n=1}^{\infty}$ "jumps" between -1 and 1, and so it does not have a limit. However, we need to give a formal proof of the the non-existence of a limit.

Solution. We must prove the following statement:

> For all $L \in \mathbb{R}$, there exists some $\varepsilon > 0$ such that for all $N \in \mathbb{N}$, there exists some $n \in \mathbb{N}$ such that $n \geq N$ and $|(-1)^n - L| \geq \varepsilon$.

Fix an arbitrary $L \in \mathbb{R}$. First, note that

---

[3] In this particular example, any choice of $\varepsilon > 0$ will do. We will later see examples where ε must be chosen more carefully.

$$\begin{aligned} |(-1)-L|+|1-L| &= |L+1|+|1-L| \\ &\geq |(L+1)+(1-L)| \qquad \text{by the Triangle Inequality} \\ &= 2, \end{aligned}$$

and so either $|(-1)-L| \geq 1$ or $|1-L| \geq 1$.

We now set $\varepsilon = 1$. Fix an arbitrary $N \in \mathbb{N}$. Now, we consider two cases: when $|(-1)-L| \geq 1$, and when $|1-L| \geq 1$.

Case 1: $|(-1)-L| \geq 1$. In this case, we fix an odd $n \in \mathbb{N}$ such that $n \geq N$, so that $(-1)^n = -1$. We now have that

$$|(-1)^n - L| \;=\; |(-1)-L| \;\geq\; 1 \;=\; \varepsilon,$$

which is what we needed to show.

Case 2: $|1-L| \geq 1$. In this case, we fix an even $n \in \mathbb{N}$ such that $n \geq N$, so that $(-1)^n = 1$. We now have that

$$|(-1)^n - L| \;=\; |1-L| \;\geq\; 1 \;=\; \varepsilon$$

which is what we needed to show. □

2.2 Properties of limits

The theorems that we prove in this section are stated (and proven) for sequences of real numbers. However, they remain true for sequences in $\mathbb{C}$ or $\mathbb{R}^d$, with very minor (if any) modifications of the proofs.

Theorem 2.2.1. *A sequence of real numbers can have at most one limit. So, every convergent sequence of real numbers has a unique limit.*

Proof. Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers, and suppose that this sequence has two distinct limits, call them L_1 and L_2. Set $\varepsilon := \frac{1}{2}|L_1 - L_2|$; then $|L_1 - L_2| = 2\varepsilon$.

Using the fact that $\{a_n\}_{n=1}^{\infty}$ converges to L_1, we fix $N_1 \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N_1$, then $|a_n - L_1| < \varepsilon$.

Using the fact that $\{a_n\}_{n=1}^{\infty}$ converges to L_2, we fix $N_2 \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N_2$, then $|a_n - L_2| < \varepsilon$.

Set $N = \max\{N_1, N_2\}$, and fix $n \in \mathbb{N}$ such that $n \geq N$. Then $|a_n - L_1| < \varepsilon$ and $|a_n - L_2| < \varepsilon$. We now have the following:

$$\begin{aligned} |L_1 - L_2| &= |(L_1 - a_n) + (a_n - L_2)| \\ &\leq |L_1 - a_n| + |a_n - L_2| \qquad \text{by the Triangle Inequality} \end{aligned}$$

$$\begin{aligned} &= |a_n - L_1| + |a_n - L_2| \\ &< \varepsilon + \varepsilon \\ &= 2\varepsilon. \end{aligned}$$

But this contradicts our choice of ε, i.e. the fact that $|L_1 - L_2| = 2\varepsilon$. $\square$

Remark: Note that whether a sequence of real numbers converges or diverges is determined by what happens in the sequence "eventually," i.e. the first few terms do not count. The same applies to the value of the limit, if it exists. Formally, we have the following lemma.

Lemma 2.2.2. *Let* $\{a_n\}_{n=1}^{\infty}$ *and* $\{b_n\}_{n=1}^{\infty}$ *be sequences of real numbers, and assume that there exits some* $N \in \mathbb{N}$ *such that for all* $n \in \mathbb{N}$*, if* $n \geq N$*, then* $a_n = b_n$*.*[4] *Then:*

(a) $\{a_n\}_{n=1}^{\infty}$ *converges if and only if* $\{b_n\}_{n=1}^{\infty}$ *converges;*

(b) if $\{a_n\}_{n=1}^{\infty}$ *and* $\{b_n\}_{n=1}^{\infty}$ *both converge, then* $\lim\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} b_n$.

Proof. Exercise. $\square$

In chapter 1, we defined upper and lower bounds of sets of real numbers, as well as their suprema and infima. We can analogously (and straightforwardly) define these concepts for sequences of real numbers, as follows.

Bounded sequences. For a sequence $\{a_n\}_{n=1}^{\infty}$ of real numbers:

- an *upper bound* of $\{a_n\}_{n=1}^{\infty}$ is any real number x such that for all $n \in \mathbb{N}$, we have $a_n \leq x$.
- the *least upper bound* (or *supremum*) of $\{a_n\}_{n=1}^{\infty}$ is an upper bound x of $\{a_n\}_{n=1}^{\infty}$ such that for every upper bound y of $\{a_n\}_{n=1}^{\infty}$, we have that $x \leq y$;
- $\{a_n\}_{n=1}^{\infty}$ is *bounded above* if it has an upper bound;
- a *lower bound* of $\{a_n\}_{n=1}^{\infty}$ is any real number x such that for all $n \in \mathbb{N}$, we have $x \leq a_n$.
- the *greatest lower bound* (or *infimum*) of $\{a_n\}_{n=1}^{\infty}$ is a lower bound x of $\{a_n\}_{n=1}^{\infty}$ such that for every lower bound y of $\{a_n\}_{n=1}^{\infty}$, we have that $y \leq x$;
- $\{a_n\}_{n=1}^{\infty}$ is *bounded below* if it has a lower bound;

---

[4] In other words, the two sequences may possibly differ in the first N terms, but after that, they coincide.

- $\{a_n\}_{n=1}^{\infty}$ is *bounded* if it is both bounded above and bounded below.

Remark: Since the ordered field $\mathbb{R}$ is complete, any sequence of real numbers that is bounded above has a supremum, and any sequence of real numbers that is bounded below has an infimum.

Remark: Note that if $\{a_n\}_{n=1}^{\infty}$ is a bounded sequence of real numbers, then there exists a real number M such that for all $n \in \mathbb{N}$, we have that $|a_n| \leq M$, and in particular, the sequence $\{|a_n|\}_{n=1}^{\infty}$ is also bounded. Indeed, suppose that x is an upper bound of $\{a_n\}_{n=1}^{\infty}$, and that y is a lower bound of $\{a_n\}_{n=1}^{\infty}$. Set $M := \max\{|x|, |y|\}$. Then for all $n \in \mathbb{N}$, we have that $-M \leq y \leq a_n \leq x \leq M$, and consequently, $|a_n| \leq M$.

Lemma 2.2.3. *Let $\{a_n\}_{n=1}^{\infty}$ be a convergent sequence of real numbers. Then we have the following:*

(a) $\{a_n\}_{n=1}^{\infty}$ is bounded;

(b) for all $m, M \in \mathbb{R}$ such that m is a lower bound and M an upper bound of $\{a_n\}_{n=1}^{\infty}$, we have that $m \leq \lim\limits_{n\to\infty} a_n \leq M$.

Proof. (a) Fix an arbitrary real number $\varepsilon > 0$,[5] and using the fact that $L = \lim\limits_{n\to\infty} a_n$, choose $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $|a_n - L| < \varepsilon$. Now, set

$$m := \min\{a_1, \dots, a_N, L-\varepsilon\} \quad \text{and} \quad M := \max\{a_1, \dots, a_N, L+\varepsilon\}.$$

We claim that m and M are a lower and upper bound, respectively, of the sequence $\{a_n\}_{n=1}^{\infty}$. Fix $n \in \mathbb{N}$; we must show that $m \leq a_n \leq M$.

If $n \leq N$, then by construction, we have that

$$m \quad \leq \quad \min\{a_1, \dots, a_N\} \quad \leq \quad a_n \quad \leq \quad \max\{a_1, \dots, a_N\} \quad \leq \quad M,$$

and we are done.

From now on, we assume that $n > N$. Then $|a_n - L| < \varepsilon$, i.e. $a_n \in (L-\varepsilon, L+\varepsilon)$. But now

$$m \quad \leq \quad L-\varepsilon \quad \leq \quad a_n \quad \leq \quad L+\varepsilon \quad \leq \quad M,$$

and again we are done.

(b) Let $m, M \in \mathbb{R}$ be such that m is a lower bound and M an upper bound of $\{a_n\}_{n=1}^{\infty}$. Set $L = \lim\limits_{n\to\infty} a_n$. We must show that $m \leq L \leq M$.

We first show that $L \leq M$. Suppose toward a contradiction that $L > M$. Set $\varepsilon_M = L - M$. Fix $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $|a_n - L| < \varepsilon_M$. Now, fix $n \in \mathbb{N}$ such that $n \geq N$. Then

$$\varepsilon_M \quad > \quad |a_n - L| \quad \geq \quad L - a_n \quad = \quad (\underbrace{L-M}_{=\varepsilon_M}) + (\underbrace{M-a_n}_{\geq 0}) \quad \geq \quad \varepsilon_M.$$

---

[5] For the proof of (a), any $\varepsilon > 0$ will do. We could, for example, choose $\varepsilon = 1$.

We have now proven that $\varepsilon_M > \varepsilon_M$, a contradiction. So, $L \leq M$.

The proof that $m \leq L$ is similar. Indeed, suppose that $L < m$, and set $\varepsilon_m = m - L$. Fix $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $|a_n - L| < \varepsilon_m$. Now, fix $n \in \mathbb{N}$ such that $n \geq N$. Then

$$\varepsilon_m \quad > \quad |a_n - L| \quad \geq \quad a_n - L \quad = \quad \underbrace{(a_n - m)}_{\geq 0} + \underbrace{(m - L)}_{\varepsilon_m} \quad \geq \quad \varepsilon_m.$$

We have now proven that $\varepsilon_m > \varepsilon_m$, a contradiction. So, $m \leq L$. ☐

Lemma 2.2.4. *For all $c \in \mathbb{R}$, the constant sequence $c, c, c, \dots$ converges, and $\lim\limits_{n\to\infty} c = c$.*

Proof. This is "obvious," but here's a formal proof. Fix $\varepsilon > 0$, and set $N = 1$. Fix $n \in \mathbb{N}$ such that $n \geq N$. Then $|c - c| = 0 < \varepsilon$. So, $\lim\limits_{n\to\infty} c = c$. ☐

Theorem 2.2.5. *Let $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ be convergent sequences of real numbers, and let $c \in \mathbb{R}$. Then all the following hold:*

(a) $\lim\limits_{n\to\infty}(ca_n) = c \lim\limits_{n\to\infty}(a_n)$;

(b) $\lim\limits_{n\to\infty}(a_n + b_n) = \big(\lim\limits_{n\to\infty} a_n\big) + \big(\lim\limits_{n\to\infty} b_n\big)$;

(c) $\lim\limits_{n\to\infty}(a_n - b_n) = \big(\lim\limits_{n\to\infty} a_n\big) - \big(\lim\limits_{n\to\infty} b_n\big)$;

(d) $\lim\limits_{n\to\infty}(a_n b_n) = \big(\lim\limits_{n\to\infty} a_n\big)\big(\lim\limits_{n\to\infty} b_n\big)$;

(e) if $b_n \neq 0$ for all $n \in \mathbb{N}$ and $\lim\limits_{n\to\infty} b_n \neq 0$, then $\lim\limits_{n\to\infty}\big(\frac{a_n}{b_n}\big) = \frac{\lim_{n\to\infty} a_n}{\lim_{n\to\infty} b_n}$.

Proof. For notational convenience, set

$$a := \lim_{n\to\infty} a_n \qquad \text{and} \qquad b := \lim_{n\to\infty} b_n.$$

We must prove:

(a) $\lim\limits_{n\to\infty}(ca_n) = ca$;

(b) $\lim\limits_{n\to\infty}(a_n + b_n) = a + b$;

(c) $\lim\limits_{n\to\infty}(a_n - b_n) = a - b$;

(d) $\lim\limits_{n\to\infty}(a_n b_n) = ab$;

(e) if $b_n \neq 0$ for all $n \in \mathbb{N}$ and $b \neq 0$, then $\lim\limits_{n\to\infty}\big(\frac{a_n}{b_n}\big) = \frac{a}{b}$.

(a) If $c = 0$, then the result follows from Lemma 2.2.4. So assume that $c \neq 0$. Fix $\varepsilon > 0$. Using the fact that $\lim\limits_{n\to\infty} a_n = a$, we fix $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $|a_n - a| < \frac{\varepsilon}{|c|}$. Now, fix $n \in \mathbb{N}$ such that $n \geq N$. Then

$$|ca_n - ca| \quad = \quad |c||a_n - a| \quad < \quad |c|\tfrac{\varepsilon}{|c|} \quad = \quad \varepsilon.$$

This proves that $\lim\limits_{n\to\infty}(ca_n) = ca$.

(b) Fix $\varepsilon > 0$. Using the fact that $\lim\limits_{n\to\infty} a_n = a$, we choose $N_1 \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N_1$, then $|a_n - a| < \frac{\varepsilon}{2}$. Similarly, using the fact that $\lim\limits_{n\to\infty} b_n = b$, we choose $N_2 \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N_2$, then $|b_n - b| < \frac{\varepsilon}{2}$. Set $N := \max\{N_1, N_2\}$, and fix $n \in \mathbb{N}$ such that $n \geq N$. Then $n \geq N_1$ and $n \geq N_2$; the former implies that $|a_n - a| < \frac{\varepsilon}{2}$, whereas the later implies that $|b_n - b| < \frac{\varepsilon}{2}$. We now compute:

$$\begin{array}{rcll} |(a_n + b_n) - (a + b)| & = & |(a_n - a) + (b_n - b)| & \\ & \leq & |a_n - a| + |b_n - b| & \text{by the Triangle Inequality} \\ & < & \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \quad = \quad \varepsilon, & \end{array}$$

and so $\lim\limits_{n\to\infty}(a_n + b_n) = a + b$.

(c) By (a), $\lim\limits_{n\to\infty}(-b_n) = -b$, and so the result follows from (b).[6]

(d) By Lemma 2.2.3(a), since $\{a_n\}_{n=1}^{\infty}$ is convergent, it is bounded. Fix a real number $A > 0$ such that for all $n \in \mathbb{N}$, we have that $|a_n| \leq A$.[7] Now, fix $\varepsilon > 0$. Using the fact that $\lim\limits_{n\to\infty} a_n = a$, we fix $N_1 \in \mathbb{N}$ such that for all $n \in \mathbb{N}$ satisfying $n \geq N_1$, we have that $|a_n - a| < \frac{\varepsilon}{2(|b|+1)}$. Further, using the fact that $\lim\limits_{n\to\infty} b_n = b$, we fix $N_2 \in \mathbb{N}$ such that for all $n \in \mathbb{N}$ satisfying $n \geq N_2$, we have that $|b_n - b| < \frac{\varepsilon}{2A}$. Set $N := \max\{N_1, N_2\}$, and fix $n \in \mathbb{N}$ such that $n \geq N$. Note that this means that $n \geq N_1$ and $n \geq N_2$; the former implies that $|a_n - a| < \frac{\varepsilon}{2(|b|+1)}$, whereas the latter implies that $|b_n - b| < \frac{\varepsilon}{2A}$. We then have the following:

$$\begin{array}{rcll} |a_nb_n - ab| & = & |(a_nb_n - a_nb) + (a_nb - ab)| & \\ & \leq & |a_nb_n - a_nb| + |a_nb - ab| & \text{by the Triangle Inequality} \end{array}$$

---

[6] Indeed, we simply apply (b) to the sequences $\{a_n\}_{n=1}^{\infty}$ and $\{-b_n\}_{n=1}^{\infty}$.

[7] To see that such an A exists, fix a lower bound m and an upper bound M of the sequence $\{a_n\}_{n=1}^{\infty}$, and let $A = \max\{|m|, |M|, 1\}$. (Note: we cannot take $A = \max\{|m|, |M|\}$ because it is possible that $m = M = 0$, and we need that $A > 0$.)

$$\begin{array}{rl} = & |a_n||b_n - b| + |a_n - a||b| \\ \leq & A|b_n - b| + |a_n - a||b| \\ < & A\frac{\varepsilon}{2A} + \frac{\varepsilon}{2(|b|+1)}|b| \quad \leq \quad \varepsilon. \end{array}$$

Thus, $\lim\limits_{n\to\infty}(a_nb_n) = ab$.

(e) Suppose that $b_n \neq 0$ for all $n \in \mathbb{N}$ and $b \neq 0$. Let us first show that $\lim\limits_{n\to\infty}\left(\frac{1}{b_n}\right) = \frac{1}{b}$. We first prove the following.

> ***Claim.*** *There exist some $N_1 \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N_1$, then $|b_n| > \frac{|b|}{2}$.*

Proof of Claim. Set $\varepsilon = \frac{|b|}{2}$. Let $N_1 \in \mathbb{N}$ be such that for all $n \in \mathbb{N}$, if $n \geq N_1$, then $|b_n - b| < \varepsilon$ (i.e. $|b_n - b| < \frac{|b|}{2}$). Now, fix $n \in \mathbb{N}$ such that $n \geq N_1$. We claim that $|b_n| > \frac{|b|}{2}$. Suppose otherwise, i.e. suppose that $|b_n| \leq \frac{|b|}{2}$. Then $\frac{|b|}{2} \leq |b| - |b_n|$. But now we have the following:

$$\frac{|b|}{2} \quad \leq \quad |b| - |b_n| \quad \overset{(*)}{\leq} \quad |b - b_n| \quad = \quad |b_n - b| \quad < \quad \frac{|b|}{2},$$

where (*) follows from Corollary 1.3.1. Thus, $\frac{|b|}{2} < \frac{|b|}{2}$, a contradiction. This proves the Claim. ◆

Let N_1 be as in the Claim. Now, fix $\varepsilon > 0$, and choose $N_2 \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N_2$, then $|b_n - b| < \frac{b^2}{2}\varepsilon$. Set $N = \max\{N_1, N_2\}$. Then for all $n \in \mathbb{N}$ such that $n \geq N$, we have the following:

$$\begin{array}{rcll} |\frac{1}{b_n} - \frac{1}{b}| & = & |\frac{b-b_n}{b_nb}| & \\ & = & \frac{|b_n-b|}{|b||b_n|} & \\ & \leq & \frac{|b_n-b|}{|b|\cdot\frac{|b|}{2}} & \text{because } n \geq N_1\text{, and so } |b_n| > \frac{|b|}{2} \\ & = & \frac{2}{b^2}|b_n - b| & \\ & < & \frac{2}{b^2}\cdot\frac{b^2}{2}\varepsilon & \text{because } n \geq N_2\text{, and so } |b_n - b| < \frac{b^2}{2}\varepsilon \\ & = & \varepsilon. & \end{array}$$

Thus, $\lim\limits_{n\to\infty}\left(\frac{1}{b_n}\right) = \frac{1}{b}$, which together with (d) implies that $\lim\limits_{n\to\infty}\left(\frac{a_n}{b_n}\right) = \frac{a}{b}$.[8] □

Remark: Recall that by Example 2.1.1, we have that $\lim\limits_{n\to\infty}\left(\frac{1}{n}\right) = 0$. Combined with Theorem 2.2.5, this readily yields the following proposition.

Proposition 2.2.6. *For all real numbers a and positive integers p, we have that*

$$\lim_{n\to\infty}\left(\tfrac{a}{n^p}\right) \quad = \quad 0.$$

Proof. We compute:

$$\begin{array}{rcll}
\lim\limits_{n\to\infty}\left(\frac{a}{n^p}\right) & = & \lim\limits_{n\to\infty}\left(a(\frac{1}{n})^p\right) & \\
& = & a\Big(\lim\limits_{n\to\infty}(\frac{1}{n})^p\Big) & \text{by Theorem 2.2.5(a)} \\
& = & a\Big(\lim\limits_{n\to\infty}\frac{1}{n}\Big)^p & \begin{array}{l}\text{by repeated applicaton}\\ \text{of Theorem 2.2.5(d)}\end{array} \\
& = & a0^p & \text{by Example 2.1.1} \\
& = & 0. &
\end{array}$$

□

Example 2.2.7. *Compute* $\lim\limits_{n\to\infty}\frac{4n^3+2n^2-n+7}{3n^3-3n^2-12}$.

Solution. Here, we have a rational expression, where both the numerator and the denominator have the same degree. The trick is to divide both the numerator and the denominator by n^3 (the highest degree term of both), and then compute. A fully formal proof using Theorem 2.2.5 and Proposition 2.2.6 would look like this:

$$\begin{array}{rcl}
\lim\limits_{n\to\infty}\frac{4n^3+2n^2-n+7}{3n^3-3n^2-12} & \overset{(*)}{=} & \lim\limits_{n\to\infty}\frac{4+\frac{2}{n}-\frac{1}{n^2}+\frac{7}{n^3}}{3-\frac{3}{n}-\frac{12}{n^3}} \\
& = & \frac{\lim\limits_{n\to\infty}\left(4+\frac{2}{n}-\frac{1}{n^2}+\frac{7}{n^3}\right)}{\lim\limits_{n\to\infty}\left(3-\frac{3}{n}-\frac{12}{n^3}\right)} \\
& = & \frac{\lim\limits_{n\to\infty}(4)+\lim\limits_{n\to\infty}(\frac{2}{n})-\lim\limits_{n\to\infty}(\frac{1}{n^2})+\lim\limits_{n\to\infty}(\frac{7}{n^3})}{\lim\limits_{n\to\infty}(3)-\lim\limits_{n\to\infty}(\frac{3}{n})-\lim\limits_{n\to\infty}(\frac{12}{n^3})} \\
& = & \frac{4+0-0+0}{3-0-0} \\
& = & \frac{4}{3},
\end{array}$$

---

[8] Indeed, we simply apply (d) to the sequences $\{a_n\}_{n=1}^{\infty}$ and $\{\frac{1}{b_n}\}_{n=1}^{\infty}$.

where (*) was obtained by diving both the numerator and the denominator by n^3. However, we do not normally include this level of detail in our computation! Typically, we would evaluate the limit this way:

$$\lim_{n\to\infty}\frac{4n^3+2n^2-n+7}{3n^3-3n^2-12} \overset{(*)}{=} \lim_{n\to\infty}\frac{4+\overset{\nearrow 0}{\boxed{\frac{2}{n}}}-\overset{\nearrow 0}{\boxed{\frac{1}{n^2}}}+\overset{\nearrow 0}{\boxed{\frac{7}{n^3}}}}{3-\underset{\searrow 0}{\boxed{\frac{3}{n}}}-\underset{\searrow 0}{\boxed{\frac{12}{n^3}}}} = \frac{4}{3},$$

where (*) was obtained by dividing both the numerator and the denominator by n^3. (Here, arrows indicate what limit the circled expressions converge to as $n \to \infty$.) $\square$

2.3 The Monotone Sequence Theorem

A sequence $\{a_n\}_{n=1}^{\infty}$ of real numbers is

- *non-decreasing* if for all $n \in \mathbb{N}$, we have that $a_n \leq a_{n+1}$;
- *strictly increasing* (or simply *increasing*) if for all $n \in \mathbb{N}$, we have that $a_n < a_{n+1}$;
- *non-increasing* if for all $n \in \mathbb{N}$, we have that $a_n \geq a_{n+1}$;
- *strictly decreasing* (or simply *decreasing*) if for all $n \in \mathbb{N}$, we have that $a_n > a_{n+1}$;
- *monotone* if it is either non-decreasing or non-increasing.

Lemma 2.3.1. *Let $\{a_n\}_{n=1}^{\infty}$ be a non-decreasing sequence of real numbers bounded above. Then $\{a_n\}_{n=1}^{\infty}$ converges, and its limit is precisely the supremum of the sequence.*[9]

Proof. Let S be the supremum of the sequence $\{a_n\}_{n=1}^{\infty}$. We must show that $\lim\limits_{n\to\infty} a_n = S$. Fix an arbitrary $\varepsilon > 0$, and fix some $N \in \mathbb{N}$ such that $a_N > S - \varepsilon$; such an N exists because otherwise, $S - \varepsilon$ would be an upper bound of $\{a_n\}_{n=1}^{\infty}$, contrary to the fact that S is the supremum of the sequence. Now, fix $n \in \mathbb{N}$ such that $n \geq N$. Then

$$S - \varepsilon \;<\; a_N \;\overset{(*)}{\leq}\; a_n \;\overset{(**)}{\leq}\; S,$$

where (*) follows from the fact that our sequence is non-decreasing, wheres (**) follows from the fact that S is the supremum of the sequence. Thus, $-\varepsilon < a_n - S \leq 0$, and consequently, $|a_n - S| < \varepsilon$. This completes the argument. $\square$

---

[9] Note that we are using the completeness of the field $\mathbb{R}$, and in particular, the fact that every non-empty subset of $\mathbb{R}$ (and therefore, every sequence in $\mathbb{R}$) bounded above has a (unique) supremum, i.e. the least upper bound. This theorem fails for e.g. $\mathbb{Q}$, since $\mathbb{Q}$ is not complete.

Lemma 2.3.2. *Let $\{a_n\}_{n=1}^{\infty}$ be a non-increasing sequence of real numbers bounded below. Then $\{a_n\}_{n=1}^{\infty}$ converges, and its limit is precisely the infimum of the sequence.*[10]

Proof. Exercise (similar to the proof of Lemma 2.3.1). □

The Monotone Sequence Theorem. *Every monotone and bounded sequence of real numbers is convergent.*

Proof. This follows immediately from Lemmas 2.3.1 and 2.3.2. □

Remark: The Monotone Sequence Theorem is often used to prove that a recursively defined sequence converges, and we give an example below (see Example 2.3.4). First, we need one more limit law, which we state without proof.

Lemma 2.3.3. *Let $\{a_n\}_{n=1}^{\infty}$ be a convergent sequence, and let $p \in \mathbb{R}$. Then*

$$\lim_{n\to\infty}(a_n^p) \quad = \quad \big(\lim_{n\to\infty} a_n\big)^p,$$

as long as both the sequence $\{a_n^p\}_{n=1}^{\infty}$ and the number $\big(\lim\limits_{n\to\infty} a_n\big)^p$ are both defined.

Remark: If the real number p is **not** a non-negative integer, then a_n^p may possibly be undefined for some value(s) of $n \in \mathbb{N}$, in which case, the sequence $\{a_n^p\}_{n=1}^{\infty}$ is undefined. On the other hand, since the sequence $\{a_n\}_{n=1}^{\infty}$ is convergent, we do know that $\lim\limits_{n\to\infty} a_n$ is defined; however, depending on the value of $\lim\limits_{n\to\infty} a_n$ and p, it is possible that $\big(\lim\limits_{n\to\infty} a_n\big)^p$ is undefined. For instance, if $\lim\limits_{n\to\infty} a_n = 0$ and $p < 0$, then $\big(\lim\limits_{n\to\infty} a_n\big)^p$ is undefined.

Proof. Omitted. □

Example 2.3.4. *Let $\{a_n\}_{n=1}^{\infty}$ be the sequence defined recursively as follows:*

- $a_1 = \sqrt{2}$;
- $a_{n+1} = \sqrt{2a_n}$ *for all* $n \in \mathbb{N}$.

Show that $\{a_n\}_{n=1}^{\infty}$ converges, and find its limit.

Proof. We first use the Monotone Sequence Theorem to prove that the sequence $\{a_n\}_{n=1}^{\infty}$ converges (i.e. has a limit), and then we compute that limit.

Proving convergence. We first prove the sequence $\{a_n\}_{n=1}^{\infty}$ is monotone and bounded. More precisely, we will show that for all $n \in \mathbb{N}$, we have that $\sqrt{2} \leq a_n \leq 2$ and $a_n \leq a_{n+1}$. We proceed by induction on n.

---

[10] Once again, we are using the completeness of $\mathbb{R}$, and in particular, the fact that every non-empty subset of $\mathbb{R}$ (and therefore, every sequence in $\mathbb{R}$) bounded below has a (unique) infimum, i.e. the greatest lower bound. This theorem fails for e.g. $\mathbb{Q}$, since $\mathbb{Q}$ is not complete.

For $n = 1$, we simply observe that $\sqrt{2} = a_1 \leq 2$ and that $a_1 = \sqrt{2} \leq \sqrt{2\sqrt{2}} = \sqrt{2a_1} = a_2$.

Now, fix $n \in \mathbb{N}$, and suppose inductively that $\sqrt{2} \leq a_n \leq 2$ and $a_n \leq a_{n+1}$. We must show that $\sqrt{2} \leq a_{n+1} \leq 2$ and $a_{n+1} \leq a_{n+2}$. First, since $\sqrt{2} \leq a_n \leq 2$, we have that $2\sqrt{2} \leq 2a_n \leq 4$, and so $\sqrt{2\sqrt{2}} \leq \sqrt{2a_n} \leq 2$. Since $\sqrt{2} \leq \sqrt{2\sqrt{2}}$ and $a_{n+1} = \sqrt{2a_n}$, we deduce that $\sqrt{2} \leq a_{n+1} \leq 2$. On the other hand, since $0 \leq a_{n+1} \leq 2$, we have that $a_{n+1} = \sqrt{a_{n+1}^2} \leq \sqrt{2a_{n+1}} = a_{n+2}$. This completes the induction.

We have now shown that $\{a_n\}_{n=1}^{\infty}$ is monotone and bounded, and so by the Monotone Sequence Theorem, the sequence converges.

Computing the limit. Set $a = \lim\limits_{n\to\infty} a_n$. (The existence of the limit follows from the convergence of the sequence $\{a_n\}_{n=1}^{\infty}$, proven above.) Then

$$a \;=\; \lim_{n\to\infty} a_{n+1} \;=\; \lim_{n\to\infty} \sqrt{2a_n} \;=\; \sqrt{2\lim_{n\to\infty} a_n} \;=\; \sqrt{2a}.$$

So, $a^2 = 2a$, which implies that $a = 0$ or $a = 2$. However, since $\{a_n\}_{n=1}^{\infty}$ is bounded below by $\sqrt{2}$, we know that $\sqrt{2} \leq a$,[11] and in particular, $a \neq 0$. So, $a = 2$.

Conclusion. We have shown that $\{a_n\}_{n=1}^{\infty}$ converges and satisfies

$$\lim_{n\to\infty} a_n \;=\; 2,$$

and we are done. □

Remark: When solving problems similar to Example 2.3.4, **it is imperative that you prove that the limit actually exists**, i.e. that the sequence in question does converge. It is **not** enough to simply compute the limit. Let us explain why. Suppose we are given a sequence $\{a_n\}_{n=1}^{\infty}$, defined recursively as follows:

- $a_1 = -1$;
- $a_{n+1} = -a_n$ for all $n \in \mathbb{N}$.

Suppose we simply set $a := \lim\limits_{n\to\infty} a_n$ and then try to evaluate a. We get:

$$a \;=\; \lim_{n\to\infty} a_{n+1} \;=\; \lim_{n\to\infty} (-a_n) \;=\; -\lim_{n\to\infty} a_n \;=\; -a.$$

Thus, $a = -a$, and consequently, $2a = 0$, i.e. $a = 0$. But this doesn't work! Indeed, our sequence satisfies $a_n = (-1)^n$ for all $n \in \mathbb{N}$, and therefore does not converge (by Example 2.1.5). What our computation actually showed is that **if** $\lim\limits_{n\to\infty} a_n$ exists, **then** $\lim\limits_{n\to\infty} a_n = 0$. It did **not** show that $\lim\limits_{n\to\infty} a_n$ does in fact exist. As a matter of fact, the limit does **not** exist.

---

[11] Here, we are using Lemma 2.2.3(b).

2.3.1 Euler's number

You may recall from high school that *Euler's number* e is an (irrational) real number that satisfies $e \approx 2.71828$. Formally, Euler's number is defined as follows:

$$e \quad := \quad \lim_{n\to\infty} \big(1+\tfrac{1}{n}\big)^n.$$

This immediately raises the following question: how do we even know that this limit exists? To prove that it does, we will use the Monotone Sequence Theorem and Bernoulli's inequality (see section 1.4), as follows.

First, we consider the auxiliary sequence $\{a_n\}_{n=1}^{\infty}$ given by

$$a_n \quad := \quad \big(1+\tfrac{1}{n}\big)^{n+1}$$

for all $n \in \mathbb{N}$. We will show that $\{a_n\}_{n=1}^{\infty}$ is non-increasing and bounded below, and is consequently convergent (by the Monotone Sequence Theorem).[12]

It is clear that $a_n > 0$ for all $n \in \mathbb{N}$, and in particular, the sequence $\{a_n\}_{n=1}^{\infty}$ is bounded below. Next, for all $n \in \mathbb{N}$, we have that

$$\begin{aligned}
\tfrac{a_n}{a_{n+1}} \quad &= \quad \tfrac{(1+\frac{1}{n})^{n+1}}{(1+\frac{1}{n+1})^{n+2}} \\
&= \quad \Big(\tfrac{1+\frac{1}{n}}{1+\frac{1}{n+1}}\Big)^{n+2} \cdot \tfrac{1}{1+\frac{1}{n}} \\
&= \quad \Big(\tfrac{n^2+2n+1}{n^2+2n}\Big)^{n+2} \cdot \tfrac{1}{1+\frac{1}{n}} \\
&= \quad \Big(1+\tfrac{1}{n^2+2n}\Big)^{n+2} \cdot \tfrac{1}{1+\frac{1}{n}} \\
&\geq \quad \big(1+(n+2)\cdot\tfrac{1}{n^2+2n}\big) \cdot \tfrac{1}{1+\frac{1}{n}} \qquad \text{by Bernoulli's inequality} \\
&= \quad \big(1+\tfrac{1}{n}\big) \cdot \tfrac{1}{1+\frac{1}{n}} \\
&= \quad 1,
\end{aligned}$$

and consequently (since $a_{n+1} > 0$), that $a_n \geq a_{n+1}$. So, the sequence $\{a_n\}_{n=1}^{\infty}$ is non-increasing. We have now shown that $\{a_n\}_{n=1}^{\infty}$ is monotone and bounded, and consequently, it is convergent (by the Monotone Sequence Theorem). In other words, $\lim\limits_{n\to\infty} a_n$ exists.

---

[12] Technically, we are using Lemma 2.3.2. However, note also that any non-increasing sequence is automatically bounded above (for example, by the first term of the sequence). So, any sequence that is non-increasing and bounded below, is in particular monotone and bounded, and consequently convergent (by the Monotone Sequence Theorem).

Finally, we compute:

$$\lim_{n\to\infty}\left(1+\tfrac{1}{n}\right)^n \;=\; \lim_{n\to\infty}\frac{a_n}{1+\frac{1}{n}} \;=\; \frac{\lim\limits_{n\to\infty} a_n}{\lim\limits_{n\to\infty}(1+\frac{1}{n})} \;=\; \frac{\lim\limits_{n\to\infty} a_n}{1} \;=\; \lim_{n\to\infty} a_n.$$

This proves that the sequence $\left\{(1+\frac{1}{n})^n\right\}_{n=1}^{\infty}$ is in fact convergent, and that Euler's number

$$e \;:=\; \lim_{n\to\infty}\left(1+\tfrac{1}{n}\right)^n$$

is indeed well defined.

Remark: Note that for all $n \in \mathbb{N}$, we have that

$$(1+\tfrac{1}{n})^n \;\overset{(*)}{\geq}\; 1+n\tfrac{1}{n} \;=\; 2,$$

where (*) follows from Bernoulli's inequality. So, 2 is a lower bound of the sequence $\left\{(1+\frac{1}{n})^n\right\}_{n=1}^{\infty}$. So, by Lemma 2.2.3(b), we have that $e = \lim\limits_{n\to\infty}\left(1+\frac{1}{n}\right)^n \geq 2$. To obtain an upper bound for e, we recall that $e = \lim\limits_{n\to\infty} a_n$, and that the sequence $\{a_n\}_{n=1}^{\infty}$ is non-increasing and is therefore bounded above by a_1. Therefore,

$$e \;=\; \lim_{n\to\infty} a_n \;\overset{(*)}{\leq}\; a_1 \;=\; 4,$$

where (*) follows from Lemma 2.2.3(b). Thus, we have obtained the following estimate for Euler's number e:

$$2 \;\leq\; e \;\leq\; 4.$$

By using numerical methods, it is possible to obtain an ever more precise estimate of e. For example, it is known that $e \approx 2.71828$.

2.4 The Squeeze Theorem for sequences

The Squeeze Theorem for sequences. *Let $\{a_n\}_{n=1}^{\infty}$, $\{b_n\}_{n=1}^{\infty}$, and $\{c_n\}_{n=1}^{\infty}$ be sequences of real numbers such that $a_n \leq b_n \leq c_n$ for all $n \in \mathbb{N}$. Assume that $\{a_n\}_{n=1}^{\infty}$ and $\{c_n\}_{n=1}^{\infty}$ converge to the same limit L. Then $\{b_n\}_{n=1}^{\infty}$ also converges to L.*

Remark: Schematically (and informally), we can represent the Squeeze Theorem like this:

$$\begin{tikzcd} a_n \arrow[dr] & \leq & b_n & \leq & c_n \arrow[dr] & \forall n \in \mathbb{N} & \Longrightarrow & \lim\limits_{n\to\infty} b_n = L \\ & L & & & & L & & \end{tikzcd}$$

Proof. Fix $\varepsilon > 0$. Using the fact that $\lim\limits_{n\to\infty} a_n = L$, we fix some $N_1 \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N_1$, then $|a_n - L| < \varepsilon$. Similarly, using the fact that $\lim\limits_{n\to\infty} c_n = L$, we fix some $N_2 \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N_2$, then $|c_n - L| < \varepsilon$. Let $N = \max\{N_1, N_2\}$. Fix $n \in \mathbb{N}$ such that $n \geq N$. Then $n \geq N_1$, and it follows that $|a_n - L| < \varepsilon$, and so $-\varepsilon < a_n - L < \varepsilon$. Similarly, $n \geq N_2$, and it follows that $|c_n - L| < \varepsilon$, and so $-\varepsilon < c_n - L < \varepsilon$. Since $a_n \leq b_n \leq c_n$, we have that

$$-\varepsilon \quad < \quad a_n - L \quad \leq \quad b_n - L \quad \leq \quad c_n - L \quad < \quad \varepsilon,$$

and we deduce that $|b_n - L| < \varepsilon$. This proves that $\lim\limits_{n\to\infty} b_n = L$, as we had claimed. □

Example 2.4.1. *Compute* $\lim\limits_{n\to\infty} \big((-1)^n \frac{1}{n}\big)$, *or prove that the limit does not exist.*

Solution. Clearly, we have that

$$-\tfrac{1}{n} \quad \leq \quad (-1)^n \tfrac{1}{n} \quad \leq \quad \tfrac{1}{n}$$

for all $n \in \mathbb{N}$. Since

- $\lim\limits_{n\to\infty} \big(\frac{1}{n}\big) = 0$,
- $\lim\limits_{n\to\infty} \big(-\frac{1}{n}\big) = -\lim\limits_{n\to\infty} \big(\frac{1}{n}\big) = -0 = 0$,

the Squeeze theorem implies that

$$\lim_{n\to} \big((-1)^n \tfrac{1}{n}\big) \quad = \quad 0.$$

Remark: The argument above was fully formal. In practice, though, it is enough to write something like this:

$$\underset{\searrow 0}{\boxed{-\tfrac{1}{n}}} \quad \leq \quad (-1)^n \tfrac{1}{n} \quad \leq \quad \underset{\searrow 0}{\boxed{\tfrac{1}{n}}} \quad \forall n \in \mathbb{N} \quad \overset{(*)}{\Longrightarrow} \quad \lim_{n\to\infty} \big((-1)^n \tfrac{1}{n}\big) \quad = \quad 0,$$

where (*) follows from the Squeeze Theorem. □

Remark: Recall that the first few terms of a sequence have no effect on the existence or value of the limit (formally, we have Lemma 2.2.2). This allows us to state a slightly stronger version of the Squeeze Theorem, as follows. (Try to give a formal proof by yourself, using the original Squeeze Theorem, plus Lemma 2.2.2.)

The Squeeze Theorem for sequences (stronger version). *Let* $\{a_n\}_{n=1}^{\infty}$, $\{b_n\}_{n=1}^{\infty}$, *and* $\{c_n\}_{n=1}^{\infty}$ *be sequences of real numbers. Assume that there exists an integer* $N \in \mathbb{N}$ *such that for all* $n \in \mathbb{N}$ *satisfying* $n \geq N$, *we have that* $a_n \leq b_n \leq c_n$. *Assume that* $\{a_n\}_{n=1}^{\infty}$ *and* $\{c_n\}_{n=1}^{\infty}$ *converge to the same limit* L. *Then* $\{b_n\}_{n=1}^{\infty}$ *also converges to* L.

Remark: A schematic (and informal) representation of the theorem above looks like this:

$$\begin{tikzcd}[column sep=small] \enclose{circle}{a_n} \arrow[dr] & \leq & b_n & \leq & \enclose{circle}{c_n} \arrow[dr] & \forall n \geq N & \Longrightarrow & \lim\limits_{n\to\infty} b_n = L \\ & L & & & & L & & \end{tikzcd}$$

2.4.1 Some limits with roots

Proposition 2.4.2. $\lim\limits_{n\to\infty} \sqrt[n]{n} = 1$.

Proof. Clearly, for all $n \in \mathbb{N}$, we have that $\sqrt[n]{n} > 1$, and consequently,

$$r_n \quad := \quad \sqrt[n]{n} - 1 \quad > \quad 0.$$

Our goal is to show that $\lim\limits_{n\to\infty} r_n = 0$. This is enough, for it will imply that

$$\lim_{n\to\infty} \sqrt[n]{n} \quad = \quad \lim_{n\to\infty} (1 + r_n) \quad = \quad 1 + \lim_{n\to\infty} r_n \quad = \quad 1 + 0 \quad = \quad 1,$$

which is what we need.

Claim. $0 < r_n < \sqrt{\frac{2}{n-1}}$ for all integers $n \geq 2$.

Proof of the Claim. Fix an integer $n \geq 2$. We already saw that $r_n > 0$. For the other inequality, we observe that

$$n \quad \overset{(*)}{=} \quad (r_n + 1)^n \quad \overset{(**)}{=} \quad \sum_{k=0}^{n} \tbinom{n}{k} r_n^k \quad \overset{(***)}{>} \quad \tbinom{n}{2} r_n^2 \quad = \quad \tfrac{n(n-1)}{2} r_n^2,$$

and consequently $r_n < \sqrt{\frac{2}{n-1}}$, where

- (*) follows from the definition of r_n;
- (**) follows from the Binomial Theorem;
- (***) follows from the fact that each summand in the sum $\sum\limits_{k=0}^{n} \binom{n}{k} r_n^k$ is strictly positive (because $r_n > 0$), and so each summand is strictly smaller than the whole sum (because there is more than one summand).[13]

This proves the Claim. ◆

---

[13] Note that there are $n + 1 \geq 3$ summands, and $\binom{n}{2} r_n^2$ is the third summand (the one we get for $k = 2$.

Clearly, $\lim\limits_{n\to\infty} \sqrt{\frac{2}{n-1}} = 0$. So, by the Claim and the Squeeze Theorem, we have that $\lim\limits_{n\to\infty} r_n = 0$, and we are done.[14] □

Proposition 2.4.3. *For all real numbers $a > 0$, we have that $\lim\limits_{n\to\infty} \sqrt[n]{a} = 1$.*

Remark: We could prove Proposition 2.4.3 similarly to the way we proved Proposition 2.4.2.[15] However, since we have already proven Proposition 2.4.2, we can simply use it, together with the Squeeze Theorem, to prove Proposition 2.4.3.

Proof. Fix a real number $a > 0$. We consider two cases: when $a \geq 1$, and when $0 < a < 1$.

Case 1: $a \geq 1$. Fix $N \in \mathbb{N}$ such that $a \leq N$. Then for all $n \in \mathbb{N}$ such that $n \geq N$, we have that

$$1 \quad \leq \quad \sqrt[n]{a} \quad \leq \quad \sqrt[n]{n}.$$

Since $\lim\limits_{n\to\infty} \sqrt[n]{n} = 1$ (by Proposition 2.4.2), the Squeeze Theorem guarantees that $\lim\limits_{n\to\infty} \sqrt[n]{a} = 1$.[16]

Case 2: $0 < a < 1$. Then $\frac{1}{a} > 1$, and so by Case 1, we have that $\lim\limits_{n\to\infty} \sqrt[n]{\frac{1}{a}} = 1$. We now compute:

$$\lim_{n\to\infty} \sqrt[n]{a} \quad = \quad \lim_{n\to\infty} \frac{1}{\sqrt[n]{\frac{1}{a}}} \quad = \quad \frac{1}{\lim\limits_{n\to\infty} \sqrt[n]{\frac{1}{a}}} \quad = \quad \frac{1}{1} \quad = \quad 1,$$

and we are done. □

2.5 Subsequences and the Bolzano-Weierstrass Theorem

A *subsequence* of a sequence $\{a_n\}_{n=1}^{\infty}$ is a sequence of the form $\{a_{n_j}\}_{j=1}^{\infty}$, where $\{n_j\}_{j=1}^{\infty}$ is a strictly increasing sequence of positive integers.

Informally, a subsequence of a sequence $\{a_n\}_{n=1}^{\infty}$ is any sequence that can be obtained from $\{a_n\}_{n=1}^{\infty}$ by possibly deleting some terms, but so that infinitely many terms still remain. (In particular, every sequence is a subsequence of itself.)

---

[14] Indeed, the Claim guarantees that for all integers $n \geq 2$, we have that

$$0 \quad < \quad r_n \quad < \quad \underset{\searrow 0}{\boxed{\sqrt{\frac{2}{n-1}}}} \quad ,$$

and so by the Squeeze Theorem, $\lim\limits_{n\to\infty} r_n = 0$. (Here, we are using our "stronger version" of the Squeeze Theorem, for $N = 2$.)

[15] Try it!

[16] Note that we are using our "stronger version" of the Squeeze Theorem.

Proposition 2.5.1. *For all sequences* $\{a_n\}_{n=1}^{\infty}$ *of real numbers, the following hold:*

(a) $\{a_n\}_{n=1}^{\infty}$ *converges if and only if all subsequences of* $\{a_n\}_{n=1}^{\infty}$ *converge;*

(b) if $\{a_n\}_{n=1}^{\infty}$ *converges, then all subsequences of* $\{a_n\}_{n=1}^{\infty}$ *converge to* $\lim\limits_{n\to\infty} a_n$.

Proof. By definition, $\{a_n\}_{n=1}^{\infty}$ is a subsequence of itself. Consequently, if all subsequences of $\{a_n\}_{n=1}^{\infty}$ converge, then in particular, $\{a_n\}_{n=1}^{\infty}$ also converges.[17]

Suppose now that $\{a_n\}_{n=1}^{\infty}$ converges, and set $L := \lim\limits_{n\to\infty} a_n$. We will show that all subsequences of $\{a_n\}_{n=1}^{\infty}$ converge to L.[18] Fix any subsequence $\{a_{n_j}\}_{j=1}^{\infty}$ of $\{a_n\}_{n=1}^{\infty}$, where $\{n_j\}_{j=1}^{\infty}$ is a strictly increasing sequence of positive integers. We must show that $\lim\limits_{j\to\infty} a_{n_j} = L$. Fix any $\varepsilon > 0$, and using the fact that $L := \lim\limits_{n\to\infty} a_n$, fix $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $|a_n - L| < \varepsilon$. Since $\{n_j\}_{j=1}^{\infty}$ is a strictly increasing sequence of positive integers, we see that for all $j \in \mathbb{N}$ such that $j \geq N$, we have that $n_j \geq n_N \geq N$, and consequently, $|a_{n_j} - L| < \varepsilon$. This proves that $L = \lim\limits_{j\to\infty} a_{n_j}$. □

Lemma 2.5.2. *Every sequence* $\{a_n\}_{n=1}^{\infty}$ *of real numbers has a monotone subsequence.*

Proof. Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers. Let us say that a positive integer m is a *peak* of the sequence $\{a_n\}_{n=1}^{\infty}$ if for all $n \in \mathbb{N}$ such that $n > m$, we have that $a_m > a_n$ (i.e. a_m is strictly greater than every subsequent term of the sequence).

Suppose first that $\{a_n\}_{n=1}^{\infty}$ has infinitely many peaks, and let $\{n_j\}_{j=1}^{\infty}$ be the sequence of all the peaks, arranged in increasing order (i.e. $n_1 < n_2 < n_3 < \dots$). But now $a_{n_1} > a_{n_2} > a_{n_3} > \dots$, i.e. $\{a_{n_j}\}_{j=1}^{\infty}$ is a strictly decreasing (and therefore monotone) subsequence of $\{a_n\}_{n=1}^{\infty}$.

From now on, we assume that $\{a_n\}_{n=1}^{\infty}$ has only finitely many (if any) peaks. Fix some $N \in \mathbb{N}$ such that all the peaks of $\{a_n\}_{n=1}^{\infty}$ are strictly smaller than N.[19] Note that this means that for all $m \in \mathbb{N}$ such that $m \geq N$, the integer m is **not** a peak, i.e. there exists some $n \in \mathbb{N}$ such that $m < n$ and $a_m \leq a_n$.

Our goal is now to (recursively) form an increasing sequence $\{n_j\}_{j=1}^{\infty}$ of positive integers such that for all $j \in \mathbb{N}$, we have that $a_{n_j} \leq a_{n_{j+1}}$. Note that once we have formed such a sequence $\{n_j\}_{j=1}^{\infty}$, we will have that $\{a_{n_j}\}_{j=1}^{\infty}$ is a non-decreasing (and therefore monotone) subsequence of $\{a_n\}_{n=1}^{\infty}$, which is what we need. First, fix $n_1 = N$. Next, fix $j \in \mathbb{N}$, and suppose we have constructed positive integers $n_1, \dots, n_j$ so that $n_1 < \dots < n_j$ and $a_{n_1} \leq \dots \leq a_{n_j}$. Since n_j is not a peak (because

---

[17] Note that this proves the "$\Longleftarrow$" part of (a).

[18] This will prove the "$\Longrightarrow$" part of (a), as well as all of part (b).

[19] Let us check that such an N exists. If $\{a_n\}_{n=1}^{\infty}$ has no peaks, then we can choose N to be any positive integer (for example, $N = 1$). If $\{a_n\}_{n=1}^{\infty}$ has at least one peak, then (using the fact that $\{a_n\}_{n=1}^{\infty}$ has only finitely many peaks) we let $m_1, \dots, m_k$ be all the peaks of $\{a_n\}_{n=1}^{\infty}$, and we let N be any integer greater than $\max\{m_1, \dots, m_k\}$ (for example, $N = \max\{m_1, \dots, m_k\} + 1$).

$n_j \geq N$), there exists some $n \in \mathbb{N}$ such that $n_j < n$ and $a_{n_j} \leq a_n$; choose n_{j+1} to be the smallest such n.[20]

By construction, $\{n_j\}_{j=1}^{\infty}$ is a strictly increasing sequence of positive integers, and $\{a_{n_j}\}_{j=1}^{\infty}$ is a non-decreasing (and therefore monotone) subsequence of $\{a_n\}_{n=1}^{\infty}$. □

The Bolzano-Weierstrass Theorem. *Every bounded sequence of real numbers has a convergent subsequence.*

Proof. Let $\{a_n\}_{n=1}^{\infty}$ be a bounded sequence of real numbers. By Lemma 2.5.2, $\{a_n\}_{n=1}^{\infty}$ has a monotone subsequence, say $\{a_{n_j}\}_{j=1}^{\infty}$, where $\{n_j\}_{j=1}^{\infty}$ is an increasing sequence of positive integers. Since $\{a_n\}_{n=1}^{\infty}$ is bounded, so is its subsequence $\{a_{n_j}\}_{j=1}^{\infty}$.[21] Now $\{a_{n_j}\}_{j=1}^{\infty}$ is a monotone and bounded sequence of real numbers, and so by the Monotone Sequence Theorem, it converges. □

2.6 Accumulation points

An *accumulation point* of a sequence $\{a_n\}_{n=1}^{\infty}$ of real numbers is a real number A such that for all real numbers $\varepsilon > 0$ and all $N \in \mathbb{N}$, there exists some $n \in \mathbb{N}$ such that $n \geq N$ and $|a_n - A| < \varepsilon$.

Proposition 2.6.1. *Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers, and let $A \in \mathbb{R}$. Then the following are equivalent:*

(i) A is an accumulation point of $\{a_n\}_{n=1}^{\infty}$;

(ii) for all real numbers $\varepsilon > 0$, the interval $(A - \varepsilon, A + \varepsilon)$ contains infinitely many terms of the sequence $\{a_n\}_{n=1}^{\infty}$;

(iii) some subsequence of $\{a_n\}_{n=1}^{\infty}$ converges to A.

Terminology: For a real number $\varepsilon > 0$, the open interval $(A - \varepsilon, A + \varepsilon)$ is called the *ε-neighborhood of A.*

Proof. It suffices to prove the following sequence of implications: "(i) $\Longrightarrow$ (iii) $\Longrightarrow$ (ii) $\Longrightarrow$ (i)."

We first assume (i) and prove (iii). We define a strictly increasing sequence $\{n_j\}_{j=1}^{\infty}$ recursively as follows. First, using the fact that A is an accumulation point of $\{a_n\}_{n=1}^{\infty}$, we fix $n_1 \in \mathbb{N}$ such that $|a_{n_1} - A| < 1$.[22] Next, assume that for some

---

[20] Actually, we could have chosen n_{j+1} to be **any** such n, and the argument would still go through. It would, however, rely on the (once controversial) Axiom of Choice, which we will not discuss in any detail in this course, but which (very roughly) states that we are allowed to make infinitely many arbitrary choices. By choosing n_{j+1} to be the **minimal** n having the desired property, our choice of n_{j+1} stops being arbitrary, and so the Axiom of Choice is not needed.

[21] Indeed, any upper bound of $\{a_n\}_{n=1}^{\infty}$ is automatically an upper bound of the subsequence $\{a_{n_j}\}_{j=1}^{\infty}$, and similarly, any lower bound of $\{a_n\}_{n=1}^{\infty}$ is a lower bound of $\{a_{n_j}\}_{j=1}^{\infty}$.

[22] Here, we are using the definition of an accumulation point for $\varepsilon = 1$ and $N = 1$.

$j \in \mathbb{N}$, we have defined the positive integer n_j; we then define $n_{j+1} \in \mathbb{N}$ as follows. Since A is an accumulation point of $\{a_n\}_{n=1}^{\infty}$, we let n_{j+1} be the smallest positive integer satisfying $n_{j+1} \geq n_j + 1$ and $|A - a_{n_{j+1}}| < \frac{1}{j+1}$.[23] We have now constructed a strictly increasing sequence $\{n_j\}_{j=1}^{\infty}$ of positive integers such that for all $j \in \mathbb{N}$, we have that $|a_{n_j} - A| < \frac{1}{j}$.

It is now easy to verify that $\lim\limits_{j\to\infty} a_{n_j} = A$. Indeed, fix $\varepsilon > 0$. Let $J \in \mathbb{N}$ be such that $J > \frac{1}{\varepsilon}$. Fix $j \in \mathbb{N} \in \mathbb{N}$ such that $j \geq J$. Then

$$\begin{array}{rcll} |a_{n_j} - A| & < & \frac{1}{j} & \text{by the construction of } n_j \\ & \leq & \frac{1}{J} & \text{because } j \geq J \\ & < & \varepsilon & \text{because } J > \frac{1}{\varepsilon}. \end{array}$$

This proves that $\lim\limits_{j\to\infty} a_{n_j} = A$, i.e. (iii) holds.

Next, we assume (iii) and prove (ii). Using (iii), we fix a strictly increasing sequence $\{n_j\}_{n=1}^{\infty}$ of positive integers such that $\lim\limits_{j\to\infty} a_{n_j} = A$. Now fix $\varepsilon > 0$. Using the definition of a limit, we now fix $J \in \mathbb{N}$ such that for all $j \in \mathbb{N}$, if $j \geq J$, then $|a_{n_j} - A| < \varepsilon$, i.e. $a_{n_j} \in (A - \varepsilon, A + \varepsilon)$. But now $a_{n_J}, a_{n_{J+1}}, a_{n_{J+2}}, \dots$ all belong to $(A - \varepsilon, A + \varepsilon)$. This proves (ii).

Finally, we assume (ii) and prove (i). Fix $\varepsilon > 0$ and $N \in \mathbb{N}$. Now, using (ii), we know that there exist infinitely many positive integers n such that $a_n \in (A-\varepsilon, A+\varepsilon)$, i.e. $|a_n - A| < \varepsilon$. Since there are infinitely many such n's, one of them (in fact, infinitely many of them) must satisfy $n \geq N$. This proves (i). □

Theorem 2.6.2. *Let* $\{a_n\}_{n=1}^{\infty}$ *be a convergent sequence of real numbers. Then* $\{a_n\}_{n=1}^{\infty}$ *has exactly one accumulation point, namely* $L := \lim\limits_{n\to\infty} a_n$.

Proof. First of all, $\{a_n\}_{n=1}^{\infty}$ is a subsequence of itself, and it converges to L; so, by Proposition 2.6.1, L is indeed an accumulation point of $\{a_n\}_{n=1}^{\infty}$. On the other hand, by Proposition 2.5.1, all subsequences of $\{a_n\}_{n=1}^{\infty}$ converge to L, and so by Proposition 2.6.1, $\{a_n\}_{n=1}^{\infty}$ has no accumulation points other than L. □

Remark: By Theorem 2.6.2, if a sequence has more than one accumulation point, then it diverges. For instance, the sequence $\{(-1)^n\}_{n=1}^{\infty}$ (which we saw in Example 2.1.5) has two accumulation points, namely 1 and -1, and so it diverges. However, the converse of Theorem 2.6.2 is false in general, i.e. some sequences that only have one accumulation point nevertheless diverge. One example is the sequence $\{a_n\}_{n=1}^{\infty}$

---

[23] Here, we are using the definition of an accumulation point for $\varepsilon = \frac{1}{j+1}$ and $N = n_j + 1$.

given by

$$a_n \quad := \quad \begin{cases} \frac{1}{n} & \text{if} \quad n \text{ is even} \\ n & \text{if} \quad n \text{ is odd} \end{cases}$$

for all $n \in \mathbb{N}$.[24] Clearly, the only accumulation point of the sequence $\{a_n\}_{n=1}^{\infty}$ is 0, and yet the sequence diverges.

2.7 Divergence to infinity

Divergence to (positive) infinity. A sequence $\{a_n\}_{n=1}^{\infty}$ *diverges to infinity*, and we write

$$\lim_{n\to\infty} a_n = \infty \qquad \text{or} \qquad \lim_{n\to\infty} a_n = +\infty,$$

or alternatively,

$$a_n \to \infty \quad \text{as} \quad n \to \infty \qquad\qquad \text{or} \qquad\qquad a_n \to +\infty \quad \text{as} \quad n \to \infty,$$

if for all $M \in \mathbb{R}$, there exists some $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $a_n > M$.

Divergence to negative infinity. A sequence $\{a_n\}_{n=1}^{\infty}$ *diverges to negative infinity*, and we write

$$\lim_{n\to\infty} a_n = -\infty,$$

or alternatively,

$$a_n \to -\infty \quad \text{as} \quad n \to \infty,$$

if for all $M \in \mathbb{R}$, there exists some $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $a_n < M$.

Example 2.7.1. *Using the definition, show that* $\lim\limits_{n\to\infty} n^2 = +\infty$.

Solution. Fix $M \in \mathbb{R}$. Let $N \in \mathbb{N}$ be such that $N > M$. Then for all $n \in \mathbb{N}$ such that $n \geq N$, we have that $n^2 \geq N^2 \geq N > M$. This proves that $\lim\limits_{n\to\infty} n^2 = +\infty$. □

Example 2.7.2. *Using the definition, show that* $\lim\limits_{n\to\infty} \big(-\sqrt{n}\big) = -\infty$.

Proof. Fix $M \in \mathbb{R}$. Fix $N \in \mathbb{N}$ such that $N > M^2$. Then $\sqrt{N} > |M| \geq -M$, and consequently, $-\sqrt{N} < M$. It follows that for all $n \in \mathbb{N}$ such that $n \geq N$, we have that $-\sqrt{n} \leq -\sqrt{N} < M$. This proves that $\lim\limits_{n\to\infty} \big(-\sqrt{n}\big) = -\infty$. □

Proposition 2.7.3. *Let* $\{a_n\}_{n=1}^{\infty}$ *be a sequence of real numbers. Then:*

---

[24] This is the sequence $1, \frac{1}{2}, 3, \frac{1}{4}, 5, \frac{1}{6}, 7, \frac{1}{8}, \dots$.

(a) if $\lim\limits_{n\to\infty} a_n = +\infty$, *then* $\{a_n\}_{n=1}^{\infty}$ *is bounded below, but is not bounded above;*

(b) if $\lim\limits_{n\to\infty} a_n = -\infty$, *then* $\{a_n\}_{n=1}^{\infty}$ *is bounded above, but is not bounded below.*

Proof. Exercise. $\square$

Remark: Properties of limits from Theorem 2.2.5 readily generalize to divergence to (positive or negative) infinity. We do not state an analogous theorem formally. Instead, we focus on the "danger zones." In particular, the following forms are "indeterminate," i.e. they can in principle be anything:

$$\tfrac{0}{0} \qquad \tfrac{\infty}{\infty} \qquad 0\cdot\infty \qquad 1^{\infty} \qquad \infty-\infty \qquad 0^0 \qquad \infty^0$$

We now consider a few examples involving these indeterminate forms.

Example 2.7.4. *Consider the behavior of the following "$\frac{\infty}{\infty}$" forms:*

(a) $$\lim_{n\to\infty} \frac{\overset{\nearrow +\infty}{\boxed{2n}}}{\underset{\searrow +\infty}{\boxed{7n}}} = \lim_{n\to\infty} \tfrac{2}{7} = \tfrac{2}{7};$$

(b) $$\lim_{n\to\infty} \frac{\overset{\nearrow -\infty}{\boxed{-2\sqrt{n}}}}{\underset{\searrow -\infty}{\boxed{-3\sqrt{n}}}} = \lim_{n\to\infty} \tfrac{2}{3} = \tfrac{2}{3}$$

(c) $$\lim_{n\to\infty} \frac{\overset{\nearrow +\infty}{\boxed{3n^2}}}{\underset{\searrow +\infty}{\boxed{2n}}} = \lim_{n\to\infty} \tfrac{3n}{2} = +\infty;$$

(d) $$\lim_{n\to\infty} \frac{\overset{\nearrow -\infty}{\boxed{-2n^2}}}{\underset{\searrow +\infty}{\boxed{5n^3}}} = \lim_{n\to\infty} \left(-\tfrac{2}{5n}\right) = 0.$$

Example 2.7.5. *Compute the following limits:*

(a) $\lim\limits_{n\to\infty}\frac{2n^2-3n+7}{-n^2+2}$;

(b) $\lim\limits_{n\to\infty}\frac{-n^3+5n-1}{2n-1}$;

(c) $\lim\limits_{n\to\infty}\frac{n^2-1}{n^5+n^4+n^2}$.

(Note that each of the above is of the form "$\frac{\infty}{\infty}$.")

Solution. In each part, we start by factoring out the largest degree term from both the numerator and the denominator, and then we evaluate.

(a) $$\lim_{n\to\infty}\frac{2n^2-3n+7}{-n^2+2} = \lim_{n\to\infty}\frac{\cancel{n^2}\left(2-\overset{\nearrow 0}{\boxed{\frac{3}{n}}}+\overset{\nearrow 0}{\boxed{\frac{7}{n^2}}}\right)}{\cancel{n^2}\left(-1+\underset{\searrow 0}{\boxed{\frac{2}{n^2}}}\right)} = -2$$

(b)

$$\begin{aligned}\lim_{n\to\infty}\frac{-n^3+5n-1}{2n-1} &= \lim_{n\to\infty}\frac{n^3\left(-1+\frac{5}{n^2}-\frac{1}{n^3}\right)}{n\left(2-\frac{1}{n}\right)}\\ &= \lim_{n\to\infty}\left(\overset{\nearrow +\infty}{\boxed{n^2}}\cdot\overset{\nearrow -\frac{1}{2}}{\boxed{\frac{-1+\frac{5}{n^2}-\frac{1}{n^3}}{2-\frac{1}{n}}}}\right)\\ &= -\infty\end{aligned}$$

(c)

$$\begin{aligned}\lim_{n\to\infty}\frac{n^2-1}{n^5+n^4+n^2} &= \lim_{n\to\infty}\frac{n^2\left(1-\frac{1}{n^2}\right)}{n^5\left(1+\frac{1}{n}+\frac{1}{n^3}\right)}\\ &= \lim_{n\to\infty}\left(\overset{\nearrow 0}{\boxed{\frac{1}{n^3}}}\cdot\overset{\nearrow 1}{\boxed{\frac{1-\frac{1}{n^2}}{1+\frac{1}{n}+\frac{1}{n^3}}}}\right)\end{aligned}$$

$$= \quad 0$$

□

Example 2.7.6. *Compute the following limits:*

(a) $\lim\limits_{n\to\infty}\left(\sqrt{7n^2+n}-\sqrt{2n^2-2n}\right)$*;*

(b) $\lim\limits_{n\to\infty}\left(\sqrt{n^2+n}-\sqrt{n^3-1}\right)$*;*

(c) $\lim\limits_{n\to\infty}\left(\sqrt{n^2+n}-\sqrt{n^2-2n}\right)$*.*

(Note that each of the above is of the form "$\infty-\infty$.")

Solution. Informally, the idea is as follows:

(a) $\sqrt{7n^2+n}-\sqrt{2n^2-2n}\approx\sqrt{7n^2}-\sqrt{2n^2}=n(\sqrt{7}-\sqrt{2})\to\infty$;

(b) $\sqrt{n^2+n}-\sqrt{n^3-1}\approx\sqrt{n^2}-\sqrt{n^3}=n-n\sqrt{n}\to-\infty$ (because $n\sqrt{n}$ increases much faster than n);

(c) $\sqrt{n^2+n}-\sqrt{n^2-2n}\approx\sqrt{n^2}-\sqrt{n^2}$ (pure "$\infty-\infty$").

- Here, we do indeed have $\sqrt{n^2+n}\approx\sqrt{n^2}$ and $\sqrt{n^2-2n}\approx\sqrt{n^2}$, but the problem is that the error is only small relative to $\sqrt{n^2}$, and **cannot** necessarily be made smaller than an arbitrarily small $\varepsilon>0$. Thus, we **cannot** deduce that $\lim\limits_{n\to\infty}\left(\sqrt{n^2+n}-\sqrt{n^2-2n}\right)=0$, and we need to compute more intelligently in order to figure out what the limit is.

However, the above only gives the intuition, and it does **not** count as a proper proof! Let us try to formalize this, i.e. give a proper solution.

(a)

$$\begin{aligned}\lim_{n\to\infty}\left(\sqrt{7n^2+n}-\sqrt{2n^2-2n}\right) &= \lim_{n\to\infty}\Big(\textcircled{n}\overset{+\infty}{\nearrow}\quad\underbrace{\Big(\sqrt{7+\frac{1}{n}}-\sqrt{2-\frac{2}{n}}\Big)}_{\longrightarrow\sqrt{7}-\sqrt{2}}\\ &= +\infty\end{aligned}$$

(b)

$$\begin{aligned}\lim_{n\to\infty}\left(\sqrt{n^2+n}-\sqrt{n^3-1}\right) &= \lim_{n\to\infty}\Big(\Big(\overset{\nearrow +\infty}{\boxed{n\sqrt{n}}}\quad\Big(\underbrace{\sqrt{\frac{1}{n}+\frac{1}{n^2}}}_{\longrightarrow 0}-\underbrace{\sqrt{1-\frac{1}{n^3}}}_{\longrightarrow 1}\Big)\\ &= -\infty\end{aligned}$$

(c) Here, the trick is to multiply and divide by $\sqrt{n^2+n}+\sqrt{n^2-2n}$, and then make use of the familiar formula $(x-y)(x+y)=x^2-y^2$. This way, we will eliminate square roots in the numerator, while obtaining the sum (rather than difference) of square roots in the denominator. Formally, we have the following.

$$\begin{aligned}\lim_{n\to\infty}\left(\sqrt{n^2+n}-\sqrt{n^2-2n}\right) &= \lim_{n\to\infty}\tfrac{\left(\sqrt{n^2+n}-\sqrt{n^2-2n}\right)\left(\sqrt{n^2+n}+\sqrt{n^2-2n}\right)}{\sqrt{n^2+n}+\sqrt{n^2-2n}}\\ &= \lim_{n\to\infty}\tfrac{(n^2+n)-(n^2-2n)}{\sqrt{n^2+n}+\sqrt{n^2-2n}}\\ &= \lim_{n\to\infty}\tfrac{3n}{n\left(\sqrt{1+\frac{1}{n}}+\sqrt{1-\frac{2}{n}}\right)}\\ &= \lim_{n\to\infty}\tfrac{3}{\sqrt{1+\frac{1}{n}}+\sqrt{1-\frac{2}{n}}}\\ &= \tfrac{3}{2}\end{aligned}$$

□

Theorem 2.7.7. *Let $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ be sequences of real numbers such that $a_n \leq b_n$ for all $n \in \mathbb{N}$. Then*

(a) if $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ both converge, then $\lim\limits_{n\to\infty} a_n \leq \lim\limits_{n\to\infty} b_n$;

(b) if $\lim\limits_{n\to\infty} a_n = +\infty$, then $\lim\limits_{n\to\infty} b_n = +\infty$;

(c) if $\lim\limits_{n\to\infty} b_n = -\infty$, then $\lim\limits_{n\to\infty} a_n = -\infty$.

Proof. We prove (a) and (b). The proof of (c) is similar to the proof of (b).

(a) Assume that $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ both converge, and set $a := \lim\limits_{n\to\infty} a_n$ and $b := \lim\limits_{n\to\infty} b_n$. We must show that $a \leq b$. Suppose otherwise, so that $b < a$. Set $\varepsilon := \frac{a-b}{2}$, and note that this implies that $a-\varepsilon = b+\varepsilon$. Using the fact that $a = \lim\limits_{n\to\infty} a_n$, we fix some $N_1 \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N_1$, then $|a_n - a| < \varepsilon$, i.e. $a-\varepsilon < a_N < a+\varepsilon$. Similarly, using the fact that $b = \lim\limits_{n\to\infty} b_n$, we fix some

$N_2 \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N_2$, then $|b_n - b| < \varepsilon$, i.e. $b - \varepsilon < b_N < b + \varepsilon$. Set $N := \max\{N_1, N_2\}$. Then $N \geq N_1$ and $N \geq N_2$, and consequently,

$$b_N \quad < \quad b + \varepsilon \quad = \quad a - \varepsilon \quad < \quad a_N,$$

contrary to the fact that $a_N \leq b_N$.

(b) Assume that $\lim\limits_{n\to\infty} a_n = +\infty$. We must show that $\lim\limits_{n\to\infty} b_n = +\infty$. Fix $M \in \mathbb{R}$. Using the fact that $\lim\limits_{n\to\infty} a_n = +\infty$, we fix some $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $a_n > M$. But now for all $n \in \mathbb{N}$ such that $n \geq N$, we have that $b_n \geq a_n > M$. This proves that $\lim\limits_{n\to\infty} b_n = +\infty$. □

Example 2.7.8. *Compute* $\lim\limits_{n\to\infty} \big(n + (-1)^n\sqrt{n}\big)$.

Solution. Clearly, we have that $n - \sqrt{n} \leq n + (-1)^n\sqrt{n}$ for all $n \in \mathbb{N}$. Since

$$\lim_{n\to\infty} \big(n - \sqrt{n}\big) \quad = \quad \lim_{n\to\infty} \Big(\overset{\nearrow +\infty}{\textcircled{n}} \quad \big(1 - \overset{\nearrow 0}{\bigcirc\!\!\!\!\!\!\frac{1}{\sqrt{n}}}\big)\Big) \quad = \quad +\infty,$$

Theorem 2.7.7(b) implies that $\lim\limits_{n\to\infty} \big(n + (-1)^n\sqrt{n}\big) = +\infty$, and we are done. □

Proposition 2.7.9. *Let* $q \in \mathbb{R}$. *Then:*

(a) if $|q| < 1$, *then* $\lim\limits_{n\to\infty} q^n = 0$;

(b) if $q = 1$, $\lim\limits_{n\to\infty} q^n = 1$;

(c) if $q > 1$, *then* $\lim\limits_{n\to\infty} q^n = +\infty$;

(d) if $q \leq -1$, *then* $\lim\limits_{n\to\infty} q^n$ *does not exist, i.e. the sequence* $\{q^n\}_{n=1}^{\infty}$ *diverges, but neither to* $+\infty$ *nor to* $-\infty$.

Proof. Part (b) is obvious. Let us prove (c). Assume that $q > 1$. Set $x := q - 1$. Then $x > 0$, and so by Bernoulli's inequality, we have that

$$q^n \quad = \quad (1 + x)^n \quad \geq \quad 1 + nx,$$

and clearly, $\lim\limits_{n\to\infty} (1 + nx) = +\infty$.[25] Therefore, by Theorem 2.7.7(b), we have that $\lim\limits_{n\to\infty} q^n = +\infty$. This proves (c).

---

[25] Here is a formal proof. Fix any $M \in \mathbb{R}$, and let $N \in \mathbb{N}$ be such that $N > \frac{M-1}{x}$. Fix any $n \in \mathbb{N}$ such that $n \geq N$. Since $x > 0$, we deduce that $1 + nx \geq 1 + Nx > 1 + (M - 1) = M$. So, $\lim\limits_{n\to\infty} (1 + nx) = +\infty$.

Let us now prove (a). So, we assume that $|q| < 1$, and we prove that $\lim\limits_{n\to\infty} q^n = 0$. If $q = 0$, then this is obviously true. Suppose now that $0 < q < 1$. Then $\frac{1}{q} > 1$, and so by (c), we have that $\lim\limits_{n\to\infty} \left(\frac{1}{q}\right)^n = +\infty$, and consequently,

$$\lim_{n\to\infty} q^n \quad = \quad \lim_{n\to\infty} \frac{1}{\underbrace{\left(\frac{1}{q}\right)^n}_{+\infty}} \quad = \quad 0.$$

Suppose now that $-1 < q < 0$. Then $0 < -q < 1$, and so by what we just showed, we have that $\lim\limits_{n\to\infty} (-q)^n = 0$. Since

$$-(-q)^n \quad \leq \quad q^n \quad \leq \quad (-q)^n$$

for all $n \in \mathbb{N}$, the Squeeze Theorem now implies that $\lim\limits_{n\to\infty} q^n = 0$. This proves (a).

It remains to prove (d). First, if $q = -1$, then $\{q^n\}_{n=1}^{\infty}$ is simply the sequence $\{(-1)^n\}_{n=1}^{\infty}$, and $\lim\limits_{n\to\infty} q^n$ does not exist.[26]

Suppose now that $q < -1$. Then by (c), the subsequence $\{q^{2n}\}_{n=1}^{\infty}$ diverges to $-\infty$,[27] whereas the subsequence $\{q^{2n+1}\}_{n=1}^{\infty}$ diverges to $-\infty$.[28] Thus, the subsequence $\{q^{2n}\}_{n=1}^{\infty}$ is unbounded above, whereas the subsequene $\{q^{2n+1}\}_{n=1}^{\infty}$ is unbounded below. Consequently, the sequence $\{q^n\}_{n=1}^{\infty}$ is bounded neither above nor below, and consequently, by Lemma 2.2.3(a) and Proposition 2.7.3, $\lim\limits_{n\to\infty} q^n$ does not exist.[29] □

For a sequence $\{a_n\}_{n=1}^{\infty}$ of real numbers:

- we define $\sup\limits_{n\geq 1} a_n$ to be the supremum of the subsequence $\{a_n\}_{n=1}^{\infty}$ if this subsequence is bounded above,[30] and we set $\sup\limits_{n\geq 1} a_n := +\infty$ otherwise;

---

[26] The fact that the sequence $\{(-1)^n\}_{n=1}^{\infty}$ diverges was formally proven in Example 2.1.5. Moreover, this sequence is bounded above (by 1), and therefore does not diverge to $+\infty$; similarly, the sequence is bounded below (by -1), and therefore does not diverge to $-\infty$. Thus, $\lim\limits_{n\to\infty} (-1)^n$ does not exit.

[27] Indeed, the subsequence $\{q^{2n}\}_{n=1}^{\infty}$ is precisely equal to the sequence $\{|q|^{2n}\}_{n=1}^{\infty}$, and so since $|q| > 1$, part (c) implies that $\lim\limits_{n\to\infty} |q|^{2n} = +\infty$.

[28] Indeed, the subsequence $\{q^{2n+1}\}_{n=1}^{\infty}$ is precisely the subsequence $\{-|q|^{2n+1}\}_{n=1}^{\infty}$. Since $|q| > 1$, part (c) implies that $\lim\limits_{n\to} |q|^{2n+1} = +\infty$, and consequently, $\lim\limits_{n\to} \big(-|q|^{2n+1}\big) = -\infty$.

[29] Indeed, since $\{q^n\}_{n=1}^{\infty}$ is not bounded, Lemma 2.2.3(a) guarantees that it diverges. But since $\{q^n\}_{n=1}^{\infty}$ is bounded neither above nor below, Proposition 2.7.3 guarantees that this sequence diverges neither to $+\infty$ nor to $-\infty$. Thus, $\lim\limits_{n\to\infty} q^n$ does not exist.

[30] Since the ordered field $\mathbb{R}$ is complete, we know that if the subsequence $\{a_n\}_{n=1}^{\infty}$ is bounded above, then it does indeed have the least upper bound (supremum).

- we define $\inf\limits_{n \geq 1} a_n$ to be the infimum of the subsequence $\{a_n\}_{n=1}^{\infty}$ if this subsequence is bounded below,[31] and we set $\inf\limits_{n \geq 1} a_n := -\infty$ otherwise.

Proposition 2.7.10. *Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers. Then:*

(a) if $\{a_n\}_{n=1}^{\infty}$ is non-decreasing, then $\lim\limits_{n \to \infty} a_n = \sup\limits_{n \geq 1} a_n$;

(b) if $\{a_n\}_{n=1}^{\infty}$ is non-increasing, then $\lim\limits_{n \to \infty} a_n = \inf\limits_{n \geq 1} a_n$.

Remark: In other words:

- part (a) states that if a non-decreasing sequence is bounded above, then it converges to its supremum, and otherwise it diverges to $+\infty$;
- part (b) states that if a non-increasing sequence is bounded below, then it diverges to its infimum, and otherwise it diverges to $-\infty$.

Proof. We prove (a); the proof of (b) is similar. So, assume that $\{a_n\}_{n=1}^{\infty}$ is non-decreasing. If $\{a_n\}_{n=1}^{\infty}$ is bounded above, then Lemma 2.3.1 guarantees that it converges to its supremum. Suppose now that $\{a_n\}_{n=1}^{\infty}$ is not bounded above, so that $\sup\limits_{n \geq 1} a_n = +\infty$; we must show that $\lim\limits_{n \to \infty} a_n = +\infty$. Fix $M \in \mathbb{R}$. Since $\{a_n\}_{n=1}^{\infty}$ is not bounded above, we know that there exists some $N \in \mathbb{N}$ such that $a_N > M$. Since $\{a_n\}_{n=1}^{\infty}$ is also non-decreasing, it follows that for all $n \in \mathbb{N}$ such that $n \geq N$, we have that $a_n \geq a_N > M$. This proves that $\sup\limits_{n \geq 1} a_n = +\infty$. □

2.8 Limit superior and limit inferior

Proposition 2.8.1. *Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers. Then:*

(a) the following are equivalent:

- *(a.1) $\{a_n\}_{n=1}^{\infty}$ is bounded above,*
- *(a.2) all subsequences of $\{a_n\}_{n=1}^{\infty}$ are bounded above,*
- *(a.3) for all $k \in \mathbb{N}$, the subsequence $\{a_n\}_{n=k}^{\infty}$ is bounded above,*
- *(a.4) there exists some $k \in \mathbb{N}$ such that the subsequence $\{a_n\}_{n=k}^{\infty}$ is bounded above;*

(b) the following are equivalent:

- *(b.1) $\{a_n\}_{n=1}^{\infty}$ is bounded below,*
- *(b.2) all subsequences of $\{a_n\}_{n=1}^{\infty}$ are bounded below,*

---

[31] Since the ordered field $\mathbb{R}$ is complete, we know that if the subsequence $\{a_n\}_{n=1}^{\infty}$ is bounded below, then it does indeed have the greatest lower bound (infimum).

(b.3) for all $k \in \mathbb{N}$, the subsequence $\{a_n\}_{n=k}^{\infty}$ is bounded below,

(b.4) there exists some $k \in \mathbb{N}$ such that the subsequence $\{a_n\}_{n=k}^{\infty}$ is bounded below.

Proof. We prove (a); the proof of (b) is analogous. It suffices to prove the following implications: "(a.1) $\Longrightarrow$ (a.2) $\Longrightarrow$ (a.3) $\Longrightarrow$ (a.4) $\Longrightarrow$ (a.1)."

Obviously, if some $M \in \mathbb{R}$ is an upper bound of the sequence $\{a_n\}_{n=1}^{\infty}$, then M is also an upper bound of all the subsequences of $\{a_n\}_{n=1}^{\infty}$; this proves the implication "(a.1) $\Longrightarrow$ (a.2)."

The implications "(a.2) $\Longrightarrow$ (a.3)" and "(a.3) $\Longrightarrow$ (a.4)" are obvious.

It remains to prove the implication "(a.4) $\Longrightarrow$ (a.1)." So, suppose that (a.4) holds, and fix some $k \in \mathbb{N}$, the subsequence $\{a_n\}_{n=k}^{\infty}$ is bounded above, say, by some $M_k \in \mathbb{R}$. But then $\{a_n\}_{n=1}^{\infty}$ is clearly bounded above by $M := \max\{a_1, \dots, a_{k-1}, M_k\}$, and in particular, (a.1) holds. This proves the implication "(a.4) $\Longrightarrow$ (a.1)," which completes the proof of (a). $\square$

For a sequence $\{a_n\}_{n=1}^{\infty}$ of real numbers and a positive integer k:

- we define $\sup\limits_{n \geq k} a_n$ to be the supremum of the subsequence $\{a_n\}_{n=k}^{\infty}$ if this subsequence is bounded above,[32] and we set $\sup\limits_{n \geq k} a_n := +\infty$ otherwise;
- we define $\inf\limits_{n \geq k} a_n$ to be the infimum of the subsequence $\{a_n\}_{n=k}^{\infty}$ if this subsequence is bounded below,[33] and we set $\inf\limits_{n \geq k} a_n := -\infty$ otherwise.

Example 2.8.2. *Consider the sequence $\left\{\frac{(-1)^n}{n}\right\}_{n=1}^{\infty}$, i.e. the sequence*

$$-1, \tfrac{1}{2}, -\tfrac{1}{3}, \tfrac{1}{4}, -\tfrac{1}{5}, \tfrac{1}{6}, \dots.$$

Then:

- *the sequence $\left\{\sup\limits_{n \geq k}\left(\frac{(-1)^n}{n}\right)\right\}_{k=1}^{\infty}$ is the (non-increasing) sequence*

$$\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{6}, \tfrac{1}{6} \dots;$$

- *the sequence $\left\{\inf\limits_{n \geq k}\left(\frac{(-1)^n}{n}\right)\right\}_{k=1}^{\infty}$ is the (non-decreasing) sequence*

$$-1, -\tfrac{1}{3}, -\tfrac{1}{3}, -\tfrac{1}{5}, -\tfrac{1}{5}, \dots.$$

---

[32] By Proposition 2.8.1(a), we know that $\{a_n\}_{n=k}^{\infty}$ is bounded above if and only if the sequence $\{a_n\}_{n=1}^{\infty}$ is bounded above. Moreover, since the ordered field $\mathbb{R}$ is complete, we know that if the subsequence $\{a_n\}_{n=k}^{\infty}$ is bounded above, then it does indeed have the least upper bound (supremum).

[33] By Proposition 2.8.1(b), we know that $\{a_n\}_{n=k}^{\infty}$ is bounded below if and only if the sequence $\{a_n\}_{n=1}^{\infty}$ is bounded below. Moreover, since the ordered field $\mathbb{R}$ is complete, we know that if the subsequence $\{a_n\}_{n=k}^{\infty}$ is bounded below, then it does indeed have the greatest lower bound (infimum).

Proposition 2.8.3. *Let* $\{a_n\}_{n=1}^{\infty}$ *be a sequence of real numbers. Then all the following hold:*

(a) for all $k \in \mathbb{N}$*, we have that* $\inf\limits_{n \geq k} a_n \leq a_k \leq \sup\limits_{n \geq k} a_n$*;*

(b) the sequence $\{\sup\limits_{n \geq k} a_n\}_{k=1}^{\infty}$ *is non-increasing;*

(c) the sequence $\{\inf\limits_{n \geq k} a_n\}_{k=1}^{\infty}$ *is non-decreasing.*

Proof. We first prove (a). Fix $k \in \mathbb{N}$. Then $\inf\limits_{n \geq k} a_n$ is a lower bound of the subsequence $\{a_n\}_{n=k}^{\infty}$, and in particular, it satisfies $\inf\limits_{n \geq k} a_n \leq a_k$. Similarly, $\sup\limits_{n \geq k} a_n$ is an upper bound of the subsequence $\{a_n\}_{n=k}^{\infty}$, and in particular, it satisfies $a_k \leq \sup\limits_{n \geq k} a_n$. This proves (a).

For (b), we simply observe that for all $k \in \mathbb{N}$, we have that

$$\sup_{n \geq k} a_n \quad = \quad \max\Big\{a_k, \sup_{n \geq k+1} a_n\Big\} \quad \geq \quad \sup_{n \geq k+1} a_n.$$

Similarly, for (c), we observe that for all $k \in \mathbb{N}$, we have that

$$\inf_{n \geq k} a_n \quad = \quad \min\Big\{a_k, \inf_{n \geq k+1} a_n\Big\} \quad \leq \quad \inf_{n \geq k+1} a_n.$$

This completes the argument. □

Notation: As a convention, for the constant sequence $+\infty, +\infty, +\infty, \dots$, we define $\lim\limits_{n \to \infty}(+\infty) = +\infty$, as well as $\sup\limits_{n \geq k}(+\infty) = +\infty$ and $\inf\limits_{n \geq k}(+\infty) = +\infty$ (for each $k \in \mathbb{N}$). Likewise, for the constant sequence $-\infty, -\infty, -\infty, \dots$, we define $\lim\limits_{n \to \infty}(-\infty) = -\infty$, as well as $\sup\limits_{n \geq k}(-\infty) = -\infty$ and $\inf\limits_{n \geq k}(-\infty) = -\infty$ (for each $k \in \mathbb{N}$).

The limit superior and the limit inferior. The *limit superior* of a sequence $\{a_n\}_{n=1}^{\infty}$ of real numbers is defined to be

$$\limsup_{n \to \infty} a_n \quad := \quad \lim_{k \to \infty} \big(\sup_{n \geq k} a_n\big) \quad = \quad \inf_{k \geq 1} \big(\sup_{n \geq k} a_n\big),$$

whereas the *limit inferior* of $\{a_n\}_{n=1}^{\infty}$ is defined to be

$$\liminf_{n \to \infty} a_n \quad := \quad \lim_{k \to \infty} \big(\inf_{n \geq k} a_n\big) \quad = \quad \sup_{k \geq 1} \big(\inf_{n \geq k} a_n\big).$$

Remark: As Proposition 2.8.4 (below) shows, the limit superior and limit inferior are indeed well defined for any sequence $\{a_n\}_{n=1}^{\infty}$ of real numbers. We emphasize that each of $\limsup\limits_{n \to \infty} a_n$ and $\liminf\limits_{n \to \infty} a_n$ may possibly be a real number, $+\infty$, or $-\infty$. This is in contrast to the fact that $\lim\limits_{n \to \infty} a_n$ need not exist.

Proposition 2.8.4. *Let* $\{a_n\}_{n=1}^{\infty}$ *be a sequence of real numbers. Then* $\{a_n\}_{n=1}^{\infty}$ *has both the limit superior and the limit inferior, that is, both* $\limsup\limits_{n\to\infty} a_n$ *and* $\liminf\limits_{n\to\infty} a_n$ *are well defined. Moreover, the following hold:*

- $\limsup\limits_{n\to\infty} a_n = +\infty$ *if and only if* $\{a_n\}_{n=1}^{\infty}$ *is not bounded above;*
- $\liminf\limits_{n\to\infty} a_n = -\infty$ *if and only if* $\{a_n\}_{n=1}^{\infty}$ *is not bounded below.*

Proof. We first deal with the limit superior. First of all, if $\{a_n\}_{n=1}^{\infty}$ is not bounded above, then Proposition 2.8.1(a) guarantees that for all $k \in \mathbb{N}$, the subsequence $\{a_n\}_{n=k}^{\infty}$ is not bounded above, either, and consequently, $\sup\limits_{n\geq k} a_n = +\infty$. Thus, if $\{a_n\}_{n=1}^{\infty}$ is not bounded above, then we simply have that

$$\limsup_{n\to\infty} a_n \quad = \quad \lim_{k\to\infty}\big(\sup_{n\geq k} a_n\big) \quad = \quad \inf_{k\geq 1}\big(\sup_{n\geq k} a_n\big) \quad = \quad +\infty.$$

Suppose now that $\{a_n\}_{n=1}^{\infty}$ is bounded above. Then for all $k \in \mathbb{N}$, the subsequence $\{a_n\}_{n=k}^{\infty}$ is also bounded above,[34] and so since the ordered field $\mathbb{R}$ is complete, $\sup\limits_{n\geq k} a_n$ is defined and is a real number. By Proposition 2.8.3(b), the sequence $\{\sup\limits_{n\geq k} a_n\}_{k=1}^{\infty}$ is non-increasing, and therefore, by Proposition 2.7.10(b), we have that

$$\lim_{k\to\infty}\big(\sup_{n\geq k} a_n\big) \quad = \quad \inf_{k\geq 1}\big(\sup_{n\geq k} a_n\big),$$

which in particular means that

$$\limsup_{n\to\infty} a_n \quad = \quad \lim_{k\to\infty}\big(\sup_{n\geq k} a_n\big) \quad = \quad \inf_{k\geq 1}\big(\sup_{n\geq k} a_n\big),$$

is well defined and is either a real number or $-\infty$.

It remains to deal with the limit inferior. First of all, if $\{a_n\}_{n=1}^{\infty}$ is not bonded below, then Proposition 2.8.1(b) guarantees that for all $k \in \mathbb{N}$, the subsequence $\{a_n\}_{n=k}^{\infty}$ is not bounded below, either, and consequently, $\inf\limits_{n\geq k} a_n = -\infty$. Thus, if $\{a_n\}_{n=1}^{\infty}$ is not bounded below, then we simply have that

$$\liminf_{n\to\infty} a_n \quad = \quad \lim_{k\to\infty}\big(\inf_{n\geq k} a_n\big) \quad = \quad \sup_{k\geq 1}\big(\inf_{n\geq k} a_n\big) \quad = \quad -\infty.$$

Suppose now that $\{a_n\}_{n=1}^{\infty}$ is bounded below. Then for all $k \in \mathbb{N}$, the subsequence $\{a_n\}_{n=k}^{\infty}$ is also bounded below,[35] and so since the ordered field $\mathbb{R}$ is complete, $\inf\limits_{n\geq k} a_n$

---

[34] This is obvious, but it also follows from Proposition 2.8.1(a).

[35] This is obvious, but it also follows from Proposition 2.8.1(b).

is defined and is a real number. By Proposition 2.8.3(c), the sequence $\{\sup\limits_{n\geq k} a_n\}_{k=1}^{\infty}$ is non-decreasing, and therefore, by Proposition 2.7.10(a), we have that

$$\lim_{k\to\infty}\big(\inf_{n\geq k} a_n\big) \quad = \quad \sup_{k\geq 1}\big(\inf_{n\geq k} a_n\big),$$

which in particular means that

$$\liminf_{n\to\infty} a_n \quad = \quad \lim_{k\to\infty}\big(\inf_{n\geq k} a_n\big) \quad = \quad \sup_{k\geq 1}\big(\inf_{n\geq k} a_n\big).$$

is well defined, and is either a real number or $-\infty$. □

Remark: As we shall see (see Theorem 2.8.8), for a sequence $\{a_n\}_{n=1}^{\infty}$ of real numbers, $\lim\limits_{n\to\infty} a_n$ exists (as a real number, as $+\infty$, or as $-\infty$) if and only if the limit superior and the limit inferior of the sequence are equal, and in this case, $\lim\limits_{n\to\infty} a_n$ is equal to both the limit superior and the limit inferior. For now, we have the following example.

Example 2.8.5. *Consider the sequence* $\big\{\frac{(-1)^n}{n}\big\}_{n=1}^{\infty}$, *i.e. the sequence*

$$-1, \tfrac{1}{2}, -\tfrac{1}{3}, \tfrac{1}{4}, -\tfrac{1}{5}, \tfrac{1}{6}, \dots.$$

Then:

- *the sequence* $\big\{\sup\limits_{n\geq k}\big(\frac{(-1)^n}{n}\big)\big\}_{k=1}^{\infty}$ *is the sequence* $\frac{1}{2}, \frac{1}{2}, \frac{1}{4}, \frac{1}{4}, \frac{1}{6}, \frac{1}{6} \dots$, *and we see that* $\limsup\limits_{n\to\infty} a_n = \lim\limits_{k\to\infty}\big(\sup\limits_{n\geq k}\big(\frac{(-1)^n}{n}\big)\big) = 0$;
- *the sequence* $\big\{\sup\limits_{n\geq k}\big(\frac{(-1)^n}{n}\big)\big\}_{k=1}^{\infty}$ *is the sequence* $-1, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{5}, -\frac{1}{5}, \dots$, *and we see that* $\liminf\limits_{n\to\infty} a_n = \lim\limits_{k\to\infty}\big(\inf\limits_{n\geq k}\big(\frac{(-1)^n}{n}\big)\big) = 0$.

Note also that $\lim\limits_{n\to\infty}\big(\frac{(-1)^n}{n}\big) = 0$, *so that the limit, the limit superior, and the limit inferior of the sequence* $\big\{\frac{(-1)^n}{n}\big\}_{n=1}^{\infty}$ *are all equal.*

Proposition 2.8.6. *Let* $\{a_n\}_{n=1}^{\infty}$ *be a bounded sequence of real numbers. Then both* $\limsup\limits_{n\to\infty} a_n$ *and* $\liminf\limits_{n\to\infty} a_n$ *are both real numbers, and moreover, they are both accumulation points of the sequence* $\{a_n\}_{n=1}^{\infty}$.

Proof. We prove the statement for the limit superior; the proof of the limit inferior is analogous.

Let $m \in \mathbb{R}$ be a lower bound of the sequence $\{a_n\}_{n=1}^{\infty}$. Further, to simplify notation, set $S := \limsup\limits_{n\to\infty} a_n$, and for all $k \in \mathbb{N}$, set $S_k := \sup\limits_{n\geq k} a_n$. Since $\{a_n\}_{n=1}^{\infty}$ is bounded, we know that S_k is a real number for all $k \in \mathbb{N}$, and moreover, that

$$S \quad = \quad \limsup_{n\to\infty} a_n \quad = \quad \lim_{k\to\infty}\big(\sup_{n\geq k} a_n\big) \quad = \quad \lim_{k\to\infty} S_k.$$

Let us first explain why S is a real number. Clearly, for all $k \in \mathbb{N}$, we have that $m \leq a_k \leq \sup\limits_{n\geq k} a_n = S_k$. Thus, the sequence $\{S_k\}_{k=1}^{\infty}$ is bounded below (by m). Since the sequence $\{S_k\}_{k=1}^{\infty}$ is also non-increasing (by Proposition 2.8.3(b)), Lemma 2.3.2 guarantees that it converges (to its infimum). Thus, S is indeed a real number.

It remains to show that S is an accumulation point of the sequence $\{a_n\}_{n=1}^{\infty}$. Fix $\varepsilon > 0$ and $N \in \mathbb{N}$. We must show that there exists some $n \in \mathbb{N}$ such that $n \geq N$ and $|a_n - S| < \varepsilon$.

Using the fact that $S = \lim\limits_{k\to\infty} S_k$, we fix some $K \in \mathbb{N}$ such that for all $k \in \mathbb{N}$, if $k \geq K$, then $|S_k - S| < \frac{\varepsilon}{2}$. Now, set $k := \max\{N, K\}$. Then $k \geq K$, and consequently, $|S_k - S| < \frac{\varepsilon}{2}$. Further, by definition, we have that $S_k = \sup\limits_{n\geq k} a_n$, and consequently, there exists some $n \in \mathbb{N}$ such that $n \geq k$ and $S_k - \frac{\varepsilon}{2} < a_n \leq S_k$. (The existence of such an n follows from the fact that S_k is an upper bound of the subsequence $\{a_n\}_{n=k}^{\infty}$, whereas $S_k - \frac{\varepsilon}{2}$ is not.) Note that this implies that $|a_n - S_k| < \frac{\varepsilon}{2}$. We now have that $n \geq N$, and that

$$\begin{array}{rcll} |a_n - S| & = & |(a_n - S_k) + (S_k - S)| & \\ & \leq & |a_n - S_k| + |S_k - S| & \text{by the Triangle Inequality} \\ & < & \frac{\varepsilon}{2} + \frac{\varepsilon}{2} & \\ & = & \varepsilon, & \end{array}$$

which proves that S is indeed an accumulation point of $\{a_n\}_{n=1}^{\infty}$. □

Proposition 2.8.7. *Let* $\{a_n\}_{n=1}^{\infty}$ *be a sequence of real numbers. Then* $\liminf\limits_{n\to\infty} a_n \leq \limsup\limits_{n\to\infty} a_n$.

Proof. If $\liminf\limits_{n\to\infty} a_n = -\infty$ or $\limsup\limits_{n\to\infty} a_n = +\infty$, then the result is immediate. We may therefore assume that $\liminf\limits_{n\to\infty} a_n > -\infty$ and $\limsup\limits_{n\to\infty} a_n < +\infty$. So, by Proposition 2.8.4, the sequence $\{a_n\}_{n=1}^{\infty}$ is bounded. Consequently, $\sup\limits_{n\geq k} a_k$ and $\inf\limits_{n\geq k} a_k$ are real numbers for all $k \in \mathbb{N}$, and moreover, by Proposition 2.8.6, both

$\liminf\limits_{n\to\infty} a_n = \lim\limits_{k\to\infty} \big(\inf\limits_{n\geq k} a_n \big)$ and $\limsup\limits_{n\to\infty} a_n = \lim\limits_{k\to\infty} \big(\sup\limits_{n\geq k} a_n \big)$ are real numbers, i.e. sequences $\big\{ \inf\limits_{n\geq k} a_n \big\}_{k=1}^{\infty}$ and $\big\{ \sup\limits_{n\geq k} a_n \big\}_{k=1}^{\infty}$ both converge. But note that for all $k \in \mathbb{N}$, we have that

$$\inf_{n\geq k} a_n \quad\leq\quad a_k \quad\leq\quad \sup_{n\geq k} a_n.$$

Therefore, by Theorem 2.7.7, we have that

$$\underbrace{\lim_{k\to\infty} \big(\inf_{n\geq k} a_n \big)}_{=\liminf\limits_{n\to\infty} a_n} \quad\leq\quad \underbrace{\lim_{k\to\infty} \big(\sup_{n\geq k} a_n \big)}_{=\liminf\limits_{n\to\infty} a_n},$$

and we are done. □

Theorem 2.8.8. *Let* $\{a_n\}_{n=1}^{\infty}$ *be a sequence of real numbers. Then* $\lim\limits_{n\to\infty} a_n$ *exists (as a real number, as* $+\infty$*, or as* $-\infty$*) if and only if* $\liminf\limits_{n\to\infty} a_n = \limsup\limits_{n\to\infty} a_n$*, and in this case, we have that* $\lim\limits_{n\to\infty} a_n = \liminf\limits_{n\to\infty} a_n = \limsup\limits_{n\to\infty} a_n$*.*

Proof. To simplify notation, set $S := \limsup\limits_{n\to\infty} a_n$ and $I := \liminf\limits_{n\to\infty} a_n$. Then at least one of the following holds: the sequence $\{a_n\}_{n=1}^{\infty}$ is bounded; the sequence $\{a_n\}_{n=1}^{\infty}$ is unbounded above; the sequence $\{a_n\}_{n=1}^{\infty}$ is unbounded below.[36] We consider the first two cases; the third case is similar to the second.

Case 1: The sequence $\{a_n\}_{n=1}^{\infty}$ is bounded. Then clearly, the subsequence $\{a_n\}_{n=k}^{\infty}$ is bounded for all $k \in \mathbb{N}$, and consequently, $\inf\limits_{n\geq k} a_n$ and $\sup\limits_{n\geq k} a_n$ are real numbers for all $k \in \mathbb{N}$. Moreover, by Proposition 2.8.6, both I and S are real numbers and accumulation points of $\{a_n\}_{n=1}^{\infty}$.

Suppose first that $L := \lim\limits_{n\to\infty} a_n$ exists. Since $\{a_n\}_{n=1}^{\infty}$ is bounded, Proposition 2.7.3 guarantees that $L \neq \pm\infty$; consequently, L is a real number, i.e. the sequence $\{a_n\}_{n=1}^{\infty}$ converges to L. Consequently, by Theorem 2.6.2, L is the only accumulation point of the sequence $\{a_n\}_{n=1}^{\infty}$. Since S and I are accumulation points of $\{a_n\}_{n=1}^{\infty}$, it follows that $L = S = I$.

Suppose now that $S = I$. Now, note that

$$\inf_{n\geq k} a_n \quad\leq\quad a_k \quad\leq\quad \sup_{n\geq k} a_n$$

for all $k \in \mathbb{N}$. Since $\lim\limits_{k\to\infty} \inf\limits_{n\geq k} a_n = I = S = \lim\limits_{k\to\infty} \sup\limits_{n\geq k} a_n$, the Squeeze Theorem guarantees that $\lim\limits_{k\to\infty} a_k = I = S$.

Case 2: The sequence $\{a_n\}_{n=1}^{\infty}$ is unboudned above. Then by Proposition 2.8.4, we have that $\limsup\limits_{n\to\infty} a_n = +\infty$.

---

[36] Clearly, it is possible that $\{a_n\}_{n=1}^{\infty}$ is unbounded above and also unbounded below, i.e. there is overlap between the last two cases.

Claim. If $\lim\limits_{n\to\infty} a_n$ exists, then $\lim\limits_{n\to\infty} a_n = +\infty$.

Proof of the Claim. Assume that $\lim\limits_{n\to\infty} a_n$ exists. Since $\{a_n\}_{n=1}^{\infty}$ is unbounded above, Lemma 2.2.3 guarantees that $\{a_n\}_{n=1}^{\infty}$ diverges, i.e. $\lim\limits_{n\to\infty} a_n$ is not a real number. On the other hand, since $\{a_n\}_{n=1}^{\infty}$ is unbounded above, Proposition 2.7.3(b) guarantees that $\lim\limits_{n\to\infty} a_n \neq -\infty$. Thus, $\lim\limits_{n\to\infty} a_n = +\infty$. ♦

In view of the Claim, it is now enough to show that $\lim\limits_{n\to\infty} a_n = +\infty$ if and only if $\liminf\limits_{n\to\infty} a_n = +\infty$,[37] that is, that $\lim\limits_{n\to\infty} a_n = +\infty$ if and only if $\lim\limits_{k\to\infty} \big(\inf\limits_{n\geq k} a_n\big) = +\infty$. Before proceeding with the proof, we observe that since $\{a_n\}_{n=1}^{\infty}$ is bounded below, Proposition 2.8.1(b) guarantees that for all $k \in \mathbb{N}$, the subsequence $\{a_n\}_{n=k}^{\infty}$ is also bounded below, and consequently, $\inf\limits_{n\geq k} a_n$ is a real number.

Suppose first that $\lim\limits_{n\to\infty} a_n = +\infty$. We must show that $\lim\limits_{k\to\infty} \big(\inf\limits_{n\geq k} a_n\big) = +\infty$. Fix any $M \in \mathbb{R}$. Using the fact that $\lim\limits_{n\to\infty} a_n = +\infty$, we fix some $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $a_n > M + 1$. Now, fix any $k \in \mathbb{N}$ such that $k \geq N$; we must show that $\inf\limits_{n\geq k} a_k > M$. First, for all for all $n \in \mathbb{N}$ such that $n \geq k$, we have that $n \geq N$, and consequently, $a_k > M + 1$. Therefore, $M + 1$ is a lower bound of the subsequence $\{a_n\}_{n=k}^{\infty}$, and it follows that $M < M + 1 \leq \inf\limits_{n\geq k} a_k$. This proves that $\lim\limits_{k\to\infty} \big(\inf\limits_{n\geq k} a_n\big) = +\infty$.

Conversely, suppose that $\lim\limits_{k\to\infty} \big(\inf\limits_{n\geq k} a_n\big) = +\infty$. Now note that $\inf\limits_{n\geq k} a_n \leq a_k$ for all $k \in \mathbb{N}$. By Theorem 2.7.7(b), it follows that $\lim\limits_{n\to\infty} a_n = +\infty$. □

2.9 Series

Suppose we are given a sequence $\{a_n\}_{n=1}^{\infty}$ of real numbers. We can then form the following infinite sequence of *partial sums*, as follows:

- $\sum\limits_{n=1}^{1} a_n = a_1$;
- $\sum\limits_{n=1}^{2} a_n = a_1 + a_2$;

---

[37] Indeed, suppose that we have shown that $\lim\limits_{n\to\infty} a_n = +\infty$ if and only if $\liminf\limits_{n\to\infty} a_n = +\infty$. Let us verify that this is enough to prove the theorem for our Case 2. Suppose first that $\lim\limits_{n\to\infty} a_n$ exists. Then by the Claim, we have that $\lim\limits_{n\to\infty} a_n = +\infty$, and consequently (by our supposition), $\liminf\limits_{n\to\infty} a_n = +\infty$. We already know that $\limsup\limits_{n\to\infty} a_n = +\infty$, and it follows that $\lim\limits_{n\to\infty} a_n = \liminf\limits_{n\to\infty} a_n = \limsup\limits_{n\to\infty} a_n$. Suppose now that $\liminf\limits_{n\to\infty} a_n = \limsup\limits_{n\to\infty} a_n$. Since $\limsup\limits_{n\to\infty} a_n = +\infty$, it follows that $\liminf\limits_{n\to\infty} a_n = +\infty$. Consequently, by our supposition, we have that $\lim\limits_{n\to\infty} a_n = +\infty$. Thus, $\lim\limits_{n\to\infty} a_n$ exists, and moreover, $\lim\limits_{n\to\infty} a_n = \liminf\limits_{n\to\infty} a_n = \limsup\limits_{n\to\infty} a_n$.

- $\sum\limits_{n=1}^{3} a_n = a_1 + a_2 + a_3$;
- $\sum\limits_{n=1}^{4} a_n = a_1 + a_2 + a_3 + a_4$;
- $\sum\limits_{n=1}^{5} a_n = a_1 + a_2 + a_3 + a_4 + a_5$;
- ...

We can also consider the associated *infinite series* (or simply *series*) $\sum\limits_{n=1}^{\infty} a_n$ (also denoted $a_1 + a_2 + a_3 + \dots$). We say that the series $\sum\limits_{n=1}^{\infty} a_n$ *converges* (or is *convergent*) if the sequence $\{\sum\limits_{n=1}^{k} a_n\}_{k=1}^{\infty}$ of partial sums converges; otherwise, the series $\sum\limits_{n=1}^{\infty} a_n$ *diverges* (or is *divergent*). If the series $\sum\limits_{n=1}^{\infty} a_n$ converges, then its *sum* is the limit $s := \lim\limits_{k\to\infty} \big(\sum\limits_{n=1}^{k} a_n\big)$, and we write $\sum\limits_{n=1}^{\infty} a_n = s$. If $\lim\limits_{k\to\infty} \big(\sum\limits_{n=1}^{k} a_n\big) = +\infty$, then we say that the series $\sum\limits_{n=1}^{\infty} a_n$ diverges to $+\infty$, and we write $\sum\limits_{n=1}^{\infty} a_n = +\infty$. Similarly, if $\lim\limits_{k\to\infty} \big(\sum\limits_{n=1}^{k} a_n\big) = -\infty$, then we say that the series $\sum\limits_{n=1}^{\infty} a_n$ diverges to $-\infty$, and we write $\sum\limits_{n=1}^{\infty} a_n = -\infty$.

Remark: Sometimes, it is convenient to start our series at an index other than $n = 1$. So, we may get series of the form $\sum\limits_{n=0}^{\infty} a_n$, or $\sum\limits_{n=5}^{\infty} a_n$, or even $\sum\limits_{n=-10}^{\infty} a_n$. Convergence and divergence of such series, as well as their sums, are defined in a natural way.

Proposition 2.9.1. *Let $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ be sequences of real numbers, and let $c \in \mathbb{R}$. Then:*

(a) if $\sum\limits_{n=1}^{\infty} a_n$ converges, then so does $\sum\limits_{n=1}^{\infty} (ca_n)$, and in that case, we have that $\sum\limits_{n=1}^{\infty} (ca_n) = c\big(\sum\limits_{n=1}^{\infty} a_n\big)$;

(b) if $\sum\limits_{n=1}^{\infty} a_n$ and $\sum\limits_{n=1}^{\infty} b_n$ both converge, then so does $\sum\limits_{n=1}^{\infty} (a_n + b_n)$, and in that case, we have that $\sum\limits_{n=1}^{\infty} (a_n + b_n) = \big(\sum\limits_{n=1}^{\infty} a_n\big) + \big(\sum\limits_{n=1}^{\infty} b_n\big)$;

(c) if $\sum\limits_{n=1}^{\infty} a_n$ *and* $\sum\limits_{n=1}^{\infty} b_n$ *both converge, then so does* $\sum\limits_{n=1}^{\infty}(a_n - b_n)$*, and in that case, we have that* $\sum\limits_{n=1}^{\infty}(a_n - b_n) = \big(\sum\limits_{n=1}^{\infty} a_n\big) - \big(\sum\limits_{n=1}^{\infty} b_n\big)$*;*

Proof. This follows from Theorem 2.2.5(a-c), applied to the relevant partial sums. □

The geometric series. A *geometric series* is a series of the form

$$\sum_{n=0}^{\infty} bq^n \quad = \quad b + bq + bq^2 + bq^3 + bq^4 + \dots,$$

where $b, q \in \mathbb{R}$, with $b \neq 0$. Let us check that this series converges if and only if $|q| < 1$, and that in this case (i.e. if $|q| < 1$), we have that

$$\sum_{n=0}^{\infty} bq^n \quad = \quad \tfrac{b}{1-q},$$

First, if $q = 1$, then our series is simply of the form $b + b + b + b + \dots$, and since $b \neq 0$, it obviously diverges (to $+\infty$ if $b > 0$, and to $-\infty$ if $b < 0$). So, let us assume that $q \neq 1$. Then for all integers $k \geq 0$, we observe that

$$\begin{aligned}(1-q)\big(\sum_{n=0}^{k} q^n\big) \quad &= \quad (1-q) + (q-q^2) + (q^2-q^3) + \dots + (q^k - q^{k+1}) \\ &= \quad 1 - q^{k+1},\end{aligned}$$

and consequently (since $q \neq 1$),

$$\sum_{n=0}^{k} q^n \quad = \quad \tfrac{1-q^{k+1}}{1-q}$$

and therefore

$$\sum_{n=0}^{k} bq^n \quad = \quad \tfrac{b-bq^{k+1}}{1-q}$$

Now, if $|q| < 1$, then by Proposition 2.7.9(a), we have that $\lim\limits_{n\to\infty} q^n = 0$, and consequently,

$$\sum_{n=0}^{\infty} bq^n \quad = \quad \lim_{k\to\infty}\big(\sum_{n=0}^{k} q^n\big) \quad = \quad \lim_{k\to\infty} \tfrac{b-bq^{k+1}}{1-q} \quad = \quad \tfrac{b}{1-q},$$

and in particular, the series $\sum\limits_{n=0}^{\infty} bq^n$ converges. On the other hand, if $|q| \geq 1$, then by Proposition 2.7.9, we know that $\{q^n\}_{n=1}^{\infty}$ diverges,[38] and consequently, the series

---

[38] Note that we are also using the fact that $q \neq 1$.

$\sum\limits_{n=0}^{\infty} bq^n$ diverges as well.[39]

Theorem 2.9.2. *Let* $\{a_n\}_{n=1}^{\infty}$ *be a sequence of real numbers. If the series* $\sum\limits_{n=1}^{\infty} a_n$ *converges, then* $\lim\limits_{n\to\infty} a_n = 0$*.*

Proof. Assume that the series $\sum\limits_{n=1}^{\infty} a_n$ converges, and set $s := \sum\limits_{n=1}^{\infty} a_n$. For each $k \in \mathbb{N}$, set $s_k := \sum\limits_{n=1}^{k} a_n$, so that $s = \lim\limits_{k\to\infty} s_k$. For convenience, we may set $s_0 := 0$. We then have the following:

$$0 \quad = \quad \underbrace{\big(\lim_{k\to\infty} s_k\big)}_{=s} - \underbrace{\big(\lim_{k\to\infty} s_{k-1}\big)}_{=s} \quad = \quad \lim_{k\to\infty} \big(s_k - s_{k-1}\big) \quad = \quad \lim_{k\to\infty} a_k,$$

and we are done. $\square$

Remark: Theorem 2.9.2 can sometimes be used to show that a series **diverges**: if a sequence $\{a_n\}_{n=1}^{\infty}$ does not converge to 0, then the series $\sum\limits_{n=1}^{\infty} a_n$ diverges. However, the converse of Theorem 2.9.2 is **false** in general. For example, we shall see in subsection 2.9.1 that the so called “harmonic series”

$$\sum_{n=1}^{\infty} \tfrac{1}{n} \quad = \quad 1 + \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{4} + \tfrac{1}{5} + \tfrac{1}{6} \dots$$

diverges, even though $\lim\limits_{n\to\infty} \frac{1}{n} = 0$.

2.9.1 The Comparison Test and the harmonic series

The Comparison Test. *Let* $\{a_n\}_{n=1}^{\infty}$ *and* $\{b_n\}_{n=1}^{\infty}$ *be sequences of real numbers such that* $0 \le a_n \le b_n$ *for all* $n \in \mathbb{N}$*. Then:*

---

[39] Let us explain this in a bit more detail. We have assumed that $b \neq 0$, $q \neq 1$, and $|q| \geq 1$. If the series $\sum\limits_{n=0}^{\infty} bq^n$ converged, then we would have that

$$s \quad := \quad \sum_{n=0}^{\infty} bq^n \quad = \quad \lim_{k\to\infty} \big(\sum_{n=0}^{k} bq^n\big) \quad = \quad \lim_{k\to\infty} \big(\tfrac{b-bq^{k+1}}{1-q}\big) \quad = \quad \frac{b-b\big(\lim\limits_{k\to\infty} q^{k+1}\big)}{1-q},$$

and consequently

$$\lim_{k\to\infty} q^{k+1} \quad = \quad \tfrac{b-(1-q)s}{b},$$

contrary to the fact that $\{q^{k+1}\}_{k=0}^{\infty}$ diverges (because $\{q^n\}_{n=1}^{\infty}$ does). So, the series $\sum\limits_{n=0}^{\infty} bq^n$ indeed diverges.

(a) if the series $\sum\limits_{n=1}^{\infty} b_n$ *converges, then so does the series* $\sum\limits_{n=1}^{\infty} a_n$*, and moreover,* $\sum\limits_{n=1}^{\infty} a_n \leq \sum\limits_{n=1}^{\infty} b_n$*;*

(b) if the series $\sum\limits_{n=1}^{\infty} a_n$ *diverges, then so does the series* $\sum\limits_{n=1}^{\infty} b_n$*.*

Proof. Note that (b) is simply the contrapositive of (a). So, it is enough to prove (a). Let us therefore assume that the series $\sum\limits_{n=1}^{\infty} b_n$ converges. For all $k \in \mathbb{N}$, consider the partial sums

$$A_k := \sum_{n=1}^{k} a_n \qquad \text{and} \qquad B_k := \sum_{n=1}^{k} b_n$$

Since all terms of the sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ are non-negative, we see that the sequences $\{A_k\}_{k=1}^{\infty}$ and $\{B_k\}_{k=1}^{\infty}$ are both non-decreasing, and moreover, we have that

$$0 \quad \leq \quad A_k \quad \leq \quad B_k \qquad \forall k \in \mathbb{N}.$$

Moreover, since $\{B_k\}_{k=1}^{\infty}$ converges (by assumption), Lemma 2.2.3 guarantees that it is bounded. Consequently, $\{A_k\}_{k=1}^{\infty}$ is also bounded.[40] Now the sequence $\{A_k\}_{k=1}^{\infty}$ is monotone and bounded, and so by the Monotone Sequence Theorem, it converges, that is, the series $\sum\limits_{n=1}^{\infty} a_n$ converges. Moreover, by Theorem 2.7.7(a), we have that $\lim\limits_{k\to\infty} A_k \leq \lim\limits_{k\to\infty} B_k$, that is, $\sum\limits_{n=1}^{\infty} a_n \leq \sum\limits_{n=1}^{\infty} b_n$. □

The harmonic series. The *harmonic series* is the series

$$\sum_{n=1}^{\infty} \tfrac{1}{n} \quad = \quad 1 + \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{4} + \tfrac{1}{5} + \tfrac{1}{6} \dots$$

We will use the Comparison Test to show that this sequence **diverges**. Define the sequence $\{a_n\}_{n=1}^{\infty}$ by setting

$$a_{2^k+\ell} \quad = \quad \tfrac{1}{2^{k+1}}$$

for all $k, \ell \in \mathbb{N}_0$ such that $0 \leq \ell \leq 2^k - 1$.[41] Note that the sequence $\{a_n\}_{n=1}^{\infty}$ is precisely the sequence

$$\underbrace{\frac{1}{2}}_{1}, \underbrace{\frac{1}{4}, \frac{1}{4}}_{2}, \underbrace{\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}}_{4}, \underbrace{\frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}}_{8}, \dots$$

---

[40] Indeed, $\{A_k\}_{k=1}^{\infty}$ is bounded below by 0, and it is bounded above by any upper bound of $\{B_k\}_{k=1}^{\infty}$.

[41] Note that every $n \in \mathbb{N}$ can be expressed (uniquely) in the form $n = 2^k + \ell$ for some $k, \ell \in \mathbb{N}_0$ such that $0 \leq \ell \leq 2^k - 1$. Indeed, fix any $n \in \mathbb{N}_0$, and let $k \in \mathbb{N}_0$ be such that $2^k \leq n \leq 2^{k+1} - 1$; clearly, such a k exists and is unique. Now set $\ell := n - 2^k$. Then $n = 2^k + \ell$, and moreover, we have that $0 \leq \ell = n - 2^k \leq (2^{k+1} - 1) - 2^k = 2^k - 1$.

Clearly, for all $n \in \mathbb{N}$, we have that $0 < a_n < \frac{1}{n}$.[42] Thus, by the Comparison Test, we need only show that $\sum\limits_{n=1}^{\infty} a_n$ diverges. But note that

$$\begin{array}{rcl} \sum\limits_{n=1}^{\infty} a_n & = & \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \dots \\ & = & \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \dots \\ & = & +\infty, \end{array}$$

and we are done.

Remark: The harmonic series and the Comparison Test can often be used to show that a series diverges. For example, note that for all $n \in \mathbb{N}$, we have that $0 < \frac{1}{n} \leq \frac{1}{\sqrt{n}}$. So, since the harmonic series $\sum\limits_{n=1}^{\infty} \frac{1}{n}$ diverges, so does the series $\sum\limits_{n=1}^{\infty} \frac{1}{\sqrt{n}}$.

2.9.2 The Alternating Series Test and the alternating harmonic sequence

The Alternating Series Test. *Let $\{a_n\}_{n=1}^{\infty}$ be the a non-increasing sequence of non-negative real numbers such that $\lim\limits_{n\to\infty} a_n = 0$. Then the series*

$$\sum_{n=1}^{\infty} (-1)^{n-1} a_n \quad = \quad a_1 - a_2 + a_3 - a_4 + a_5 - a_6 + \dots$$

converges.

Remark: Here, we are assuming that the following are satisfied:

- $a_n \geq 0$ for all $n \in \mathbb{N}$;
- $a_1 \geq a_2 \geq a_3 \geq a_4 \geq \dots$;
- $\lim\limits_{n\to\infty} a_n = 0$.

The Alternating Series Test states that, under these circumstances, the "alternating series" $\sum\limits_{n=1}^{\infty} (-1)^{n-1} a_n = a_1 - a_2 + a_3 - a_4 + a_5 - a_6 + \dots$ converges.

---

[42] Let us verify this. Fix any $n \in \mathbb{N}$, and let $k \in \mathbb{N}_0$ be such that $2^k \leq n \leq 2^{k+1} - 1$ (clearly, there is exactly one such k). Next, set $\ell := n - 2^k$, and observe that $0 \leq \ell \leq 2^k - 1$ and $2^k + \ell < 2^{k+1}$. Then $a_n = \frac{1}{2^{k+1}} > 0$. On the other hand,

$$a_n \quad = \quad \tfrac{1}{2^{k+1}} \quad < \quad \tfrac{1}{2^k + \ell} \quad = \quad \tfrac{1}{n},$$

as required.

Proof. For each $k \in \mathbb{N}$, consider the partial sum $s_k := \sum\limits_{n=1}^{k} (-1)^{n-1} a_n$. We will show that $\lim\limits_{k\to\infty} s_{2k}$ and $\lim\limits_{k\to\infty} s_{2k+1}$ converge to the same limit; this will imply that $\lim\limits_{k\to\infty} s_k$ also converges to that same limit.[43]

First, we note that for all $k \in \mathbb{N}$, we have that

$$s_{2k+2} \quad = \quad s_{2k} + a_{2k+1} - a_{2k+2} \quad \overset{(*)}{\geq} \quad s_{2k},$$

where (*) follows from the fact that the sequence $\{a_n\}_{n=1}^{\infty}$ is non-increasing, and consequently, $a_{2k+1} - a_{2k+2} \geq 0$. Thus, the subsequence $\{s_{2k}\}_{k=1}^{\infty}$ is non-decreasing. On the other hand, we note that for all $k \in \mathbb{N}$, we have that

$$s_{2k} \quad = \quad a_1 - \big(\underbrace{a_2 - a_3}_{\overset{(*)}{\geq} 0}\big) - \big(\underbrace{a_4 - a_5}_{\overset{(*)}{\geq} 0}\big) - \dots - \big(\underbrace{a_{2k-2} - a_{2k-1}}_{\overset{(*)}{\geq} 0}\big) - \underbrace{a_{2k}}_{\overset{(**)}{\geq} 0} \quad \leq \quad a_1,$$

where each instance of (*) follows from the fact that the sequence $\{a_n\}_{n=1}^{\infty}$ is non-increasing, and where (**) follows from the fact that all terms of the sequence $\{a_n\}_{n=1}^{\infty}$ are non-negative. Thus, the subsequence $\{s_{2k}\}_{k=1}^{\infty}$ is bounded above by a_1.[44] The Monotone Sequence Theorem now implies that $\{s_{2k}\}_{k=1}^{\infty}$ converges.

Now, set $s := \lim\limits_{k\to\infty} s_k$. Then

$$\begin{aligned} \lim_{k\to\infty} s_{2k+1} \quad &= \quad \lim_{k\to\infty} (s_{2k} + a_{2k+1}) \\ &= \quad \big(\lim_{k\to\infty} s_{2k}\big) + \big(\lim_{k\to\infty} a_{2k+1}\big) \\ &= \quad s + 0 \\ &= \quad s. \end{aligned}$$

This completes the argument. □

The alternating harmonic series. By the Alternating Series Test, the *alternating harmonic series*

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \quad = \quad 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{5} - \tfrac{1}{6} + \dots$$

converges.

---

[43] Prove this! That is, prove that if $\lim\limits_{k\to\infty} s_{2k}$ and $\lim\limits_{k\to\infty} s_{2k+1}$ converge to the same limit, say s, then $\lim\limits_{k\to\infty} s_k = s$.

[44] Since the sequence $\{a_{2k}\}_{k=1}^{\infty}$ is non-decreasing, it is obviously bounded below by its first term, that is, by the term $s_2 = a_1 - a_2$.

2.9.3 The p-series

A *p-series* is a series of the form

$$\sum_{n=1}^{\infty} \frac{1}{n^p},$$

where p is some (fixed) real number. Note that for $p = 1$, we simply obtain the harmonic series $\sum\limits_{n=1}^{\infty} \frac{1}{n}$ (which diverges, as we saw in subsection 2.9.1). As Theorem 2.9.3 (below) shows, the p-series converges if and only if $p > 1$.

Theorem 2.9.3. *Let* $p \in \mathbb{R}$.

(a) If $p > 1$*, then the p-series* $\sum\limits_{n=1}^{\infty} \frac{1}{n^p}$ *converges;*

(b) If $p \leq 1$*, then the p-series* $\sum\limits_{n=1}^{\infty} \frac{1}{n^p}$ *diverges.*

Proof. We first prove (b). Assume that $p \leq 1$. Then for all $n \in \mathbb{N}$, we have that $0 < \frac{1}{n} \leq \frac{1}{n^p}$. Since the harmonic series $\sum\limits_{n=1}^{\infty} \frac{1}{n}$ diverges, the Comparison Test now guarantees that the p-series $\sum\limits_{n=1}^{\infty} \frac{1}{n^p}$ also diverges. This proves (b).

It remains to prove (a). Assume that $p > 1$. We must show that $\sum\limits_{n=1}^{\infty} \frac{1}{n^p}$ converges. For all $n \in \mathbb{N}$, set $a_n := 2^{-pk}$, where k is the (unique) non-negative integer such that $2^k \leq n \leq 2^{k+1} - 1$. Our goal is to show that $0 < \frac{1}{n^p} \leq a_n$ for all $n \in \mathbb{N}$, and that the series $\sum\limits_{n=1}^{\infty} a_n$ converges; the Comparison Test will then guarantee that the p-series $\sum\limits_{n=1}^{\infty} \frac{1}{n^p}$ also converges, which is what we need.

Claim 1. For all $n \in \mathbb{N}$, we have that $0 < \frac{1}{n^p} \leq a_n$.

Proof of Claim 1. Fix $n \in \mathbb{N}$. Obviously, we have $0 < \frac{1}{n^p}$, and we just need to show that $\frac{1}{n^p} \leq a_n$. Let k be the unique non-negative integer such that $2^k \leq n \leq 2^{k+1} - 1$, so that $a_n = 2^{-pk}$. We now have the following sequence of equivalent statements:

$$\begin{array}{lcll} \frac{1}{n^p} \leq a_n & \iff & \frac{1}{n^p} \leq 2^{-pk} & \text{because } a_n = 2^{-pk} \\ & \iff & (\frac{1}{n})^p \leq (\frac{1}{2^k})^p & \\ & \iff & \frac{1}{n} \leq \frac{1}{2^k} & \text{because } p > 0 \\ & \iff & 2^k \leq n. & \end{array}$$

The last statement ("$2^k \leq n$") is true by the choice of k. So, the first statement ("$\frac{1}{n^p} \leq a_n$") is true as well. ◆

Claim 2. The series $\sum\limits_{n=1}^{\infty} a_n$ converges.

Proof of Claim 2. We compute:

$$\begin{aligned}
\sum_{n=1}^{\infty} a_n &= \sum_{k=0}^{\infty} \Big(\sum_{n=2^k}^{2^{k+1}-1} a_n \Big) \\
&= \sum_{k=0}^{\infty} \Big(\sum_{n=2^k}^{2^{k+1}-1} 2^{-pk} \Big) \\
&= \sum_{k=0}^{\infty} \Big((2^{k+1} - 2^k) 2^{-pk} \Big) \\
&= \sum_{k=0}^{\infty} \big(2^k \cdot 2^{-pk} \big) \\
&= \sum_{k=0}^{\infty} 2^{k-pk} \\
&= \sum_{k=0}^{\infty} (2^{1-p})^k.
\end{aligned}$$

But note that $\sum\limits_{k=0}^{\infty} (2^{1-p})^k$ is a geometric series.[45] Since $p > 1$, we have that $1 - p < 0$, and we deduce that $0 < 2^{1-p} < 1$.[46] Therefore, the geometric series $\sum\limits_{k=0}^{\infty} (2^{1-p})^k$ converges, and consequently, the series $\sum\limits_{n=1}^{\infty} a_n$ converges as well. ◆

Claims 1 and 2 and the Comparison Test now guarantee that the p-series $\sum\limits_{n=1}^{\infty} \frac{1}{n^p}$ converges. This proves (a). □

2.9.4 Absolute convergence

For a sequence $\{a_n\}_{n=1}^{\infty}$ of real numbers, the series $\sum\limits_{n=1}^{\infty} a_n$ is said to *converge absolutely* (or to be *absolutely convergent*) if the series $\sum\limits_{n=1}^{\infty} |a_n|$ converges. A series *converges*

---

[45] To see this, simply set $b := 1$ and $q := 2^{1-p}$.

[46] Thus, for our $q := 2^{1-p}$, we have that $|q| < 1$.

conditionally (or is *conditionally convergent*) if it is convergent, but not absolutely convergent.

Theorem 2.9.4. *Every absolutely convergent series is convergent. In other words, for every sequence* $\{a_n\}_{n=1}^{\infty}$ *of real numbers, if the series* $\sum\limits_{n=1}^{\infty} |a_n|$ *converges, then so does the series* $\sum\limits_{n=1}^{\infty} a_n$*.*

Proof. Fix a sequence $\{a_n\}_{n=1}^{\infty}$ of real numbers, and assume that the series $\sum\limits_{n=1}^{\infty} |a_n|$ converges. Consequently (by Proposition 2.9.1(a)), the series $\sum\limits_{n=1}^{\infty} 2|a_n|$ also converges. Now, note that for all $n \in \mathbb{N}$, we have that either $a_n = |a_n|$ or $a_n = -|a_n|$, and consequently,

$$0 \leq a_n + |a_n| \leq 2|a_n|.$$

Since the series $\sum\limits_{n=1}^{\infty} 2|a_n|$ converges, the Comparison Test guarantees that the series $\sum\limits_{n=1}^{\infty} \big(a_n + |a_n|\big)$ also converges. But now

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} \Big(\big(a_n + |a_n|\big) - |a_n|\Big) = \underbrace{\Big(\sum_{n=1}^{\infty} \big(a_n + |a_n|\big)\Big)}_{\text{converges}} - \underbrace{\Big(\sum_{n=1}^{\infty} |a_n|\Big)}_{\text{converges}},$$

and it follows that $\sum\limits_{n=1}^{\infty} a_n$ converges.[47] □

Remark: Obviously, if $\{a_n\}_{n=1}^{\infty}$ is a sequence of **non-negative** real numbers, so that $|a_n| = a_n$ for all $n \in \mathbb{N}$, then the series $\sum\limits_{n=1}^{\infty} a_n$ cannot be conditionally convergent, and instead, it is convergent if and only if it is absolutely convergent.

Remark: Not all convergent series are absolutely convergent. For example, as we saw in subsection 2.9.2, the alternating harmonic series

$$\sum_{n=1}^{\infty} \tfrac{(-1)^{n-1}}{n} = 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{5} - \tfrac{1}{6} + \dots$$

converges, but the series

$$\sum_{n=1}^{\infty} |\tfrac{(-1)^{n-1}}{n}| = \sum_{n=1}^{\infty} \tfrac{1}{n} = 1 + \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{4} + \tfrac{1}{5} + \tfrac{1}{6} + \dots$$

is simply the harmonic series, which diverges (as we saw in subsection 2.9.1). This means that the alternating harmonic series is conditionally convergent.

---

[47] Note that we were implicitly using Proposition 2.9.1(c).

2.9.5 Rearranging the terms of a series

We note that rearranging the terms of a convergent series may alter the sum of the series, or even produce a divergent series. Let us take a look at an example. Recall from subsection 2.9.2 that the alternating harmonic series

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \quad = \quad 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{5} - \tfrac{1}{6} + \tfrac{1}{7} - \tfrac{1}{8} + \dots$$

converges; in fact, as we saw in subsection 2.9.4, it converges conditionally. Let us show that simply rearranging the terms of this series can produce a divergent series. Set

$$s \quad := \quad \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \quad = \quad 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{5} - \tfrac{1}{6} + \tfrac{1}{7} - \tfrac{1}{8} + \dots.$$

Note that $1 > \frac{1}{2}$, $\frac{1}{3} > \frac{1}{4}$, $\frac{1}{5} > \frac{1}{6}$, and so on; therefore, $s > 0$.[48] Now, by multiplying both sides of the equality

$$s \quad = \quad 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{5} - \tfrac{1}{6} + \tfrac{1}{7} - \tfrac{1}{8} + \dots.$$

by $\frac{1}{2}$, we obtain

$$\tfrac{s}{2} \quad = \quad \tfrac{1}{2} - \tfrac{1}{4} + \tfrac{1}{6} - \tfrac{1}{8} + \tfrac{1}{10} - \tfrac{1}{12} + \tfrac{1}{14} - \tfrac{1}{16} + \dots.$$

By inserting some zeros, we obtain

$$\tfrac{s}{2} \quad = \quad 0 + \tfrac{1}{2} + 0 - \tfrac{1}{4} + 0 + \tfrac{1}{6} + 0 - \tfrac{1}{8} + 0 + \tfrac{1}{10} + 0 - \tfrac{1}{12} + 0 + \tfrac{1}{14} + 0 - \tfrac{1}{16} + \dots$$

Now we sum up the following two equalities:

$$s \quad = \quad 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \tfrac{1}{5} - \tfrac{1}{6} + \tfrac{1}{7} - \tfrac{1}{8} + \tfrac{1}{9} - \tfrac{1}{10} + \tfrac{1}{11} - \tfrac{1}{12} + \tfrac{1}{13} - \tfrac{1}{14} + \tfrac{1}{15} - \tfrac{1}{16} + \dots$$

$$\tfrac{s}{2} \quad = \quad 0 + \tfrac{1}{2} + 0 - \tfrac{1}{4} + 0 + \tfrac{1}{6} + 0 - \tfrac{1}{8} + 0 + \tfrac{1}{10} + 0 - \tfrac{1}{12} + 0 + \tfrac{1}{14} + 0 - \tfrac{1}{16} + \dots$$

to obtain

$$\begin{aligned} \tfrac{3s}{2} \quad &= \quad 1 + 0 + \tfrac{1}{3} - \tfrac{1}{2} + \tfrac{1}{5} + 0 + \tfrac{1}{7} - \tfrac{1}{4} + \tfrac{1}{9} + 0 + \tfrac{1}{11} - \tfrac{1}{6} + \tfrac{1}{13} + 0 + \tfrac{1}{15} - \tfrac{1}{8} + \dots \\ &= \quad 1 + \tfrac{1}{3} - \tfrac{1}{2} + \tfrac{1}{5} + \tfrac{1}{7} - \tfrac{1}{4} + \tfrac{1}{9} + \tfrac{1}{11} - \tfrac{1}{6} + \tfrac{1}{13} + \tfrac{1}{15} - \tfrac{1}{8} + \dots \end{aligned}$$

Note that the last series that we obtained has exactly the same terms as the alternating harmonic series, but in a different order. However, the sum of this last

---

[48] Here is a formal proof. Note that

$$s \quad = \quad \sum_{n=1}^{\infty} \tfrac{(-1)^{n-1}}{n} \quad = \quad \sum_{k=1}^{\infty} \left(\tfrac{1}{2k-1} - \tfrac{1}{2k}\right) \quad = \quad \sum_{k=1}^{\infty} \tfrac{1}{(2k-1)(2k)}.$$

Since $\frac{1}{(2k-1)(2k)} > 0$ for all $k \in \mathbb{N}$, it follows that $s > 0$. In fact, it can be shown that $s = \ln 2$, i.e. the sum of the alternating harmonic series is $\ln 2$, but we will not use this fact. We will only use the fact that $s \neq 0$.

series is $\frac{3s}{2}$, rather than s. (Since $s \neq 0$, we see that $\frac{3s}{2} \neq s$, i.e. the sum has indeed changed.)

The phenomenon that we observed above may only happen with conditionally convergent series. In the case of absolutely convergent series, rearranging the terms will never alter the sum of the series (see Theorem 2.9.6 below); we remark that this is one of the reasons why we care about absolute convergence.

Recall that a *permutation* of a set X is simply a bijection from X to X. For a sequence $\{a_n\}_{n=1}^{\infty}$ of real numbers, a *rearrangement* of the series $\sum\limits_{n=1}^{\infty} a_n$ is any series of the form

$$\sum_{n=1}^{\infty} a_{\sigma(n)} \quad = \quad a_{\sigma(1)} + a_{\sigma(2)} + a_{\sigma(3)} + a_{\sigma(4)} + \dots,$$

where σ is a permutation of $\mathbb{N}$. For example, if $\sigma : \mathbb{N} \to \mathbb{N}$ is given by

$$\sigma(n) \quad = \quad \begin{cases} n+1 & \text{if} \quad n \text{ is odd} \\ n-1 & \text{if} \quad n \text{ is even} \end{cases}$$

for all $n \in \mathbb{N}$,[49] then we have that

$$\begin{aligned} \sum_{n=1}^{\infty} a_{\sigma(n)} \quad &= \quad a_{\sigma(1)} + a_{\sigma(2)} + a_{\sigma(3)} + a_{\sigma(4)} + a_{\sigma(5)} + a_{\sigma(6)} + \dots \\ &= \quad a_2 + a_1 + a_4 + a_3 + a_6 + a_5 + \dots \end{aligned}$$

Proposition 2.9.5. *Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of non-negative real numbers, and let σ be a permutation of $\mathbb{N}$. Then the series $\sum\limits_{n=1}^{\infty} a_n$ converges if and only if the series $\sum\limits_{n=1}^{\infty} a_{\sigma(n)}$ converges, and in this case, $\sum\limits_{n=1}^{\infty} a_n = \sum\limits_{n=1}^{\infty} a_{\sigma(n)}$.*

Proof. For all $n \in \mathbb{N}$, set $b_n := a_{\sigma(n)}$. We must show that $\sum\limits_{n=1}^{\infty} a_n$ converges if and only if $\sum\limits_{n=1}^{\infty} b_n$ converges, and that in this case, $\sum\limits_{n=1}^{\infty} a_n = \sum\limits_{n=1}^{\infty} b_n$. Clearly, it suffices to prove the following two statements:

(i) if $\sum\limits_{n=1}^{\infty} a_n$ converges, then so does $\sum\limits_{n=1}^{\infty} b_n$, and in this case, $\sum\limits_{n=1}^{\infty} b_n \leq \sum\limits_{n=1}^{\infty} a_n$;

(ii) if $\sum\limits_{n=1}^{\infty} b_n$ converges, then so does $\sum\limits_{n=1}^{\infty} a_n$, and in this case, $\sum\limits_{n=1}^{\infty} a_n \leq \sum\limits_{n=1}^{\infty} b_n$.

But note that since σ is a permutation of $\mathbb{N}$, so is σ^{-1}, and moreover, we have that $a_n = b_{\sigma^{-1}(n)}$ for all $n \in \mathbb{N}$.[50] Thus, it is enough to prove (i), for the proof of (ii) will be analogous.

---

[49] Note that this permutation swaps 1 and 2, swaps 3 and 4, swaps 5 and 6, and so on.

[50] Indeed, for all $n \in \mathbb{N}$, we have that $b_{\sigma^{-1}(n)} = a_{\sigma(\sigma^{-1}(n))} = a_n$.

So, let us assume that $\sum\limits_{n=1}^{\infty} a_n$ converges, and set $s := \sum\limits_{n=1}^{\infty} a_n$. We must show that $\sum\limits_{n=1}^{\infty} b_n$ converges and satisfies $\sum\limits_{n=1}^{\infty} b_n \leq s$.

First, for all $k \in \mathbb{N}$, consider the partial sums $s_k := \sum\limits_{n=1}^{k} a_n$ and $t_k := \sum\limits_{n=1}^{k} b_n = \sum\limits_{n=1}^{k} a_{\sigma(n)}$. Since $a_n \geq 0$ for all $n \in \mathbb{N}$, both $\{s_k\}_{k=1}^{\infty}$ and $\{t_k\}_{k=1}^{\infty}$ are non-decreasing sequences. Moreover, by definition, we have that $s := \sum\limits_{n=1}^{\infty} a_n = \lim\limits_{k \to \infty} s_k$. So, by Lemma 2.3.1, s is the least upper bound (supremum) of the sequence $\{s_k\}_{k=1}^{\infty}$.[51] It now suffices to show that s is an upper bound of the non-decreasing sequence $\{t_k\}_{k=1}^{\infty}$, for Lemma 2.3.1 will then imply that $\lim\limits_{k \to \infty} t_k$ converges to its least upper bound (supremum), and in particular, that $\lim\limits_{k \to \infty} t_k \leq s$, and consequently, $\sum\limits_{n=1}^{\infty} b_n \leq s$, which is what we need to show.

Fix any $k \in \mathbb{N}$. Since $\sigma(1), \dots, \sigma(k)$ is a finite list of (pairwise distinct) positive integers, we know that it is bounded above, say, by some positive integer N. But then $\{\sigma(1), \dots, \sigma(k)\} \subseteq \{1, \dots, N\}$, and so since $a_1, \dots, a_N$ are all non-negative, we see that $\sum\limits_{n=1}^{k} a_{\sigma(n)} \leq \sum\limits_{n=1}^{N} a_n$. Consequently,

$$t_k \quad = \quad \sum_{n=1}^{k} b_n \quad = \quad \sum_{n=1}^{k} a_{\sigma(n)} \quad \leq \quad \sum_{n=1}^{N} a_n \quad = \quad s_N \quad \leq \quad s.$$

This proves that s is indeed an upper bound of $\{t_k\}_{k=1}^{\infty}$, and we are done. □

Theorem 2.9.6. *Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers such that the series $\sum\limits_{n=1}^{\infty} a_n$ converges **absolutely**. Then for all permutations σ of $\mathbb{N}$, the series $\sum\limits_{n=1}^{\infty} a_{\sigma(n)}$ converges absolutely, and moreover, $\sum\limits_{n=1}^{\infty} a_{\sigma(n)} = \sum\limits_{n=1}^{\infty} a_n$.*

Proof. Fix a permutation σ of $\mathbb{N}$. By assumption, $\sum\limits_{n=1}^{\infty} |a_n|$ converges; since all terms of the sequence $\{|a_n|\}_{n=1}^{\infty}$ are non-negative, Proposition 2.9.5 guarantees that $\sum\limits_{n=1}^{\infty} |a_{\sigma(n)}|$ converges, i.e. that the series $\sum\limits_{n=1}^{\infty} a_{\sigma(n)}$ converges absolutely.

---

[51] Let us explain this in a bit more detail. Since the sequence $\{s_k\}_{k=1}^{\infty}$ converges, Lemma 2.2.3(a) guarantees that it is bounded. Since $\{s_k\}_{k=1}^{\infty}$ is non-decreasing and bounded above, Lemma 2.3.1 guarantees that it converges to its least upper bound (supremum). Since $s = \lim\limits_{k \to \infty} s_k$, it follows that s is the least upper bound of $\{s_k\}_{k=1}^{\infty}$.

We define sequences $\{a_n^+\}_{n=1}^\infty$ and $\{a_n^+\}_{n=1}^\infty$ by setting

$$a_n^+ \quad := \quad \begin{cases} |a_n| & \text{if} \quad a_n \geq 0 \\ 0 & \text{if} \quad a_n < 0 \end{cases} \qquad \text{and} \qquad a_n^- \quad := \quad \begin{cases} 0 & \text{if} \quad a_n \geq 0 \\ |a_n| & \text{if} \quad a_n < 0 \end{cases}$$

for all $n \in \mathbb{N}$. We observe that

$$a_n \quad = \quad a_n^+ - a_n^-$$

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

Obviously, $0 \leq a_n^+, a_n^- \leq |a_n|$ for all $n \in \mathbb{N}$; since $\sum\limits_{n=1}^{\infty} |a_n|$ converges (by assumption), the Comparison Test guarantees that $\sum\limits_{n=1}^{\infty} a_n^+$ and $\sum\limits_{n=1}^{\infty} a_n^-$ both converge. Since all terms of the sequences $\{a_n^+\}_{n=1}^\infty$ and $\{a_n^-\}_{n=1}^\infty$ are non-negative, Proposition 2.9.5 guarantees that

$$\sum_{n=1}^{\infty} a_{\sigma(n)}^+ \quad = \quad \sum_{n=1}^{\infty} a_n^+ \qquad \text{and} \qquad \sum_{n=1}^{\infty} a_{\sigma(n)}^- \quad = \quad \sum_{n=1}^{\infty} a_n^-.$$

It now follows that:

$$\begin{aligned} \sum_{n=1}^{\infty} a_{\sigma(n)} \quad &= \quad \sum_{n=1}^{\infty} \left(a_{\sigma(n)}^+ - a_{\sigma(n)}^-\right) \\ &= \quad \Big(\sum_{n=1}^{\infty} a_{\sigma(n)}^+\Big) - \Big(\sum_{n=1}^{\infty} a_{\sigma(n)}^-\Big) \\ &= \quad \Big(\sum_{n=1}^{\infty} a_n^+\Big) - \Big(\sum_{n=1}^{\infty} a_n^-\Big) \\ &= \quad \sum_{n=1}^{\infty} \left(a_n^+ - a_n^-\right) \\ &= \quad \sum_{n=1}^{\infty} a_n, \end{aligned}$$

and we are done. □

In the case of conditionally convergent series, we have the following theorem (whose proof we omit).

The Riemann Rearrangement Theorem. *Let $\{a_n\}_{n=1}^\infty$ be a sequence of real numbers such that the series $\sum\limits_{n=1}^{\infty} a_n$ converges **conditionally**. Then all the following hold:*

(a) for all $s \in \mathbb{R}$*, there exists a permutation* σ *of* $\mathbb{N}$ *such that* $\sum\limits_{n=1}^{\infty} a_{\sigma(n)} = s$*;*

(b) there exists a permutation σ *of* $\mathbb{N}$ *such that* $\sum\limits_{n=1}^{\infty} a_{\sigma(n)} = +\infty$*;*

(c) there exists a permutation σ *of* $\mathbb{N}$ *such that* $\sum\limits_{n=1}^{\infty} a_{\sigma(n)} = -\infty$*;*

(d) there exists a permutation σ *of* $\mathbb{N}$ *such that* $\sum\limits_{n=1}^{\infty} a_{\sigma(n)}$ *diverges (but neither to* $+\infty$ *nor to* $-\infty$*).*

While we do not give the full details of the proof of the Riemann Rearrangement Theorem, let us try to give a bit of an intuition. Suppose that $\{a_n\}_{n=1}^{\infty}$ is a sequence of real numbers such that the series $\sum\limits_{n=1}^{\infty} a_n$ converges conditionally. As in the proof of Theorem 2.9.6, we can define the sequences $\{a_n^+\}_{n=1}^{\infty}$ and $\{a_n^-\}_{n=1}^{\infty}$ by setting

$$a_n^+ \;:=\; \begin{cases} |a_n| & \text{if} \quad a_n \geq 0 \\ 0 & \text{if} \quad a_n < 0 \end{cases} \qquad \text{and} \qquad a_n^- \;:=\; \begin{cases} 0 & \text{if} \quad a_n \geq 0 \\ |a_n| & \text{if} \quad a_n < 0 \end{cases}$$

for all $n \in \mathbb{N}$. Clearly,

$$a_n \;=\; a_n^+ - a_n^-$$

for all $n \in \mathbb{N}$. However, unlike in the proof of Theorem 2.9.6, the series $\sum\limits_{n=1}^{\infty} a_n^+$ and $\sum\limits_{n=1}^{\infty} a_n^-$ diverge, for otherwise, it would be easy to show that the series $\sum\limits_{n=1}^{\infty} a_n$ converges absolutely, a contradiction. So, if we tried to write

$$\sum_{n=1}^{\infty} a_n \;=\; \sum_{n=1}^{\infty} \big(a_n^+ - a_n^-\big) \;=\; \Big(\sum_{n=1}^{\infty} a_n^+\Big) - \Big(\sum_{n=1}^{\infty} a_n^-\Big),$$

we would get an expression of the form "$\infty - \infty$," which can be anything. And indeed, by carefully rearranging the terms of the series, the "gap" between the two infinities can be arranged to be anything we like: a real number of our choice, $+\infty$, or $-\infty$. Moreover, it is possible to rearrange the terms of the series $\sum\limits_{n=1}^{\infty} a_n$ so that the partial sums of the resulting series behave "erratically," i.e. they neither converge to a real number, nor diverge to $+\infty$ or $-\infty$; in this case, the resulting series simply diverges (but neither to $+\infty$ nor to $-\infty$).

2.9.6 The Ratio Test

The Ratio Test. *Let* $\{a_n\}_{n=1}^{\infty}$ *be a sequence of real numbers, and assume that* $L := \lim\limits_{n\to\infty} |\frac{a_{n+1}}{a_n}|$ *exists (either as a real number or as* $+\infty$*).*

(a) If $L < 1$, then the series $\sum\limits_{n=1}^{\infty} a_n$ *converges absolutely;*

(b) If $L > 1$,[52] *then the series* $\sum\limits_{n=1}^{\infty} a_n$ *diverges.*

Remark: If $L = 1$, then the Ratio Test gives us no information about the convergence/divergence of the series $\sum\limits_{n=1}^{\infty} a_n$.

Proof. The idea of the proof is as follows. For (a), we will use a suitably chosen convergent geometric series and the Comparison Test. For (b), we will show that the sequence $\{a_n\}_{n=1}^{\infty}$ does not converge to 0, and consequently, $\sum\limits_{n=1}^{\infty} a_n$ diverges (by Theorem 2.9.2). Let us now give the formal details.

We first prove (a). So, suppose that $L < 1$. Set $\varepsilon := \frac{1-L}{2}$; clearly, $\varepsilon > 0$. Now, using the fact that $\lim\limits_{n\to\infty} |\frac{a_{n+1}}{a_n}| = L$, fix $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $\big||\frac{a_{n+1}}{a_n}| - L\big| < \varepsilon$, and consequently, $0 \leq |\frac{a_{n+1}}{a_n}| < L + \varepsilon = \frac{1+L}{2} =: q$;[53] Since $L < 1$, we deduce that $0 \leq q < 1$, and in particular, $|q| \leq 1$. Therefore, the geometric series

$$\sum_{n=0}^{\infty} |a_N| q^n$$

converges.

Claim 1. For all $n \in \mathbb{N}_0$, we have that $0 \leq |a_{N+n}| \leq |a_N| q^n$.

Proof of Claim 1. This is "obvious," but here is a formal inductive proof. The statement is obviously true for $n = 0$. Now fix $n \in \mathbb{N}_0$, and assume inductively that $0 \leq |a_{N+n}| \leq |a_N| q^n$. Clearly, $N + n \geq N$, and so $|\frac{a_{N+n+1}}{a_{N+n}}| < q$. Then $0 \leq |a_{N+n+1}| < q|a_{N+n}| \overset{(*)}{\leq} q\big(|a_N| q^n\big) = |a_N| q^{n+1}$, where (*) follows from the induction hypothesis. This completes the induction. ♦

Since the geometric series $\sum\limits_{n=0}^{\infty} |a_N| q^n$ converges, Claim 1 and the Comparison Test guarantee that the series $\sum\limits_{n=0}^{\infty} |a_{N+n}| = \sum\limits_{n=N}^{\infty} |a_n|$ converges. If $N = 1$, then we are done. So, let us suppose that $N \geq 2$. Then

$$\sum_{n=1}^{\infty} |a_n| \;\; = \;\; \underbrace{\Big(\sum_{n=1}^{N-1} |a_n|\Big)}_{\text{finite sum}} + \underbrace{\Big(\sum_{n=N}^{\infty} |a_n|\Big)}_{\text{convergent}},$$

---

[52] Possibly, $\lim\limits_{n\to\infty} |\frac{a_{n+1}}{a_n}| = +\infty$.

[53] Technically, the fact that $\big||\frac{a_{n+1}}{a_n}| - L\big| < \varepsilon$ implies that $L - \varepsilon < |\frac{a_{n+1}}{a_n}| < L + \varepsilon$. However, because of the absolute value, it is obvious that $|\frac{a_{n+1}}{a_n}| \geq 0$.

and so $\sum\limits_{n=1}^{\infty} |a_n|$ converges. This proves (a).

It remains to prove (b). So, assume that $L > 1$. Our goal is to show that $\{a_n\}_{n=1}^{\infty}$ does not converge to 0; in view of Theorem 2.9.2, this will imply that $\sum\limits_{n=1}^{\infty} a_n$ diverges, which is what we need.

> **Claim 2.** There exists some $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $|\frac{a_{n+1}}{a_n}| > 1$.

Proof of Claim 2. If $L = +\infty$, i.e. $\lim\limits_{n\to\infty} |\frac{a_{n+1}}{a_n}| = +\infty$, then this follows immediately from the definition of divergence to infinity.[54]

So, let us suppose first that L is a real number. Set $\varepsilon := L - 1$; since $L > 1$, we know that $\varepsilon > 0$. Using the fact that $\lim\limits_{n\to\infty} |\frac{a_{n+1}}{a_n}| = L$, we fix some $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $\big||\frac{a_{n+1}}{a_n}| - L\big| < \varepsilon$. Consequently, for all $n \in \mathbb{N}$ such that $n \geq N$, we have that $|\frac{a_{n+1}}{a_n}| > L - \varepsilon = 1$. ◆

> **Claim 3.** There exists some $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $|a_n| > |a_N| > 0$.

Proof of Claim 3. Using Claim 2, fix $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $|\frac{a_{n+1}}{a_n}| > 1$, and consequently, $|a_{n+1}| > |a_n|$. Then the sequence

$$|a_N|, |a_{N+1}|, |a_{N+2}|, |a_{N+3}|, \dots$$

is strictly increasing. Moreover, since $a_N \neq 0$ (for otherwise, the ratio $\frac{a_{N+1}}{a_N}$ would be undefined), we have that $|a_N| > 0$. This proves Claim 3. ◆

> **Claim 4.** The sequence $\{a_n\}_{n=1}^{\infty}$ does not converge to 0.

Proof of Claim 4. Suppose otherwise, that is, suppose that $\lim\limits_{n\to\infty} a_n = 0$. First, let $N \in \mathbb{N}$ be as in Claim 3, so that for all $n \in \mathbb{N}$ such that $n \geq N$, we have that $|a_n| > |a_N| > 0$. Set $\varepsilon := \frac{|a_N|}{2}$; clearly, $\varepsilon > 0$.

Using the fact that $\lim\limits_{n\to\infty} a_n = 0$, we fix $N' \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N'$, then $|a_n - 0| < \varepsilon$, i.e. $|a_n| < \frac{|a_N|}{2}$. But now for $n := \max\{N, N'\}$, we have that

$$|a_N| \overset{(*)}{<} |a_n| \overset{(**)}{<} \frac{|a_N|}{2},$$

where (*) follows from the fact that $n \geq N$, and (**) follows from the fact that $n \geq N'$. We have now obtained $|a_N| < \frac{|a_N|}{2}$, a contradiction. ◆

Claim 4 and Theorem 2.9.2 together imply that the series $\sum\limits_{n=1}^{\infty} a_n$ diverges, which completes the proof of (b). □

---

[54] We simply take $M := 1$ and apply the definition of divergence to infinity.

Example 2.9.7. *Consider the series* $\sum\limits_{n=1}^{\infty}(-1)^n\frac{n^7}{7^n}$. *Note that*

$$\lim_{n\to\infty}\left|\frac{(-1)^{n+1}\frac{(n+1)^7}{7^{n+1}}}{(-1)^n\frac{n^7}{7^n}}\right| = \lim_{n\to\infty}\frac{(n+1)^7}{7n^7} = \lim_{n\to\infty}\frac{(1+\frac{1}{n})^7}{7} = \frac{1}{7} < 1,$$

and so by the Ratio Test, the series $\sum\limits_{n=1}^{\infty}(-1)^n\frac{n^7}{7^n}$ *converges absolutely.*

Example 2.9.8. *Consider the series* $\sum\limits_{n=1}^{\infty}\frac{n^n}{n!}$. *Note that*

$$\lim_{n\to\infty}\left|\frac{\frac{(n+1)^{n+1}}{(n+1)!}}{\frac{n^n}{n!}}\right| = \lim_{n\to\infty}\frac{(n+1)^n}{n^n} = \lim_{n\to\infty}(1+\tfrac{1}{n})^n = e > 1,$$

and so by the Ratio Test, the series $\sum\limits_{n=1}^{\infty}\frac{n^n}{n!}$ *diverges.*

Remark: As pointed out above, if $\{a_n\}_{n=1}^{\infty}$ is a sequence of real numbers such that $\lim\limits_{n\to\infty}|\frac{a_{n+1}}{a_n}| = 1$, then the Ratio Test gives us no information about the convergence/divergence of the series $\sum\limits_{n=1}^{\infty}a_n$, and in principle, it is possible that the series converges absolutely, converges conditionally, or diverges. For example, consider the following three series:

1. the harmonic series $\sum\limits_{n=1}^{\infty}\frac{1}{n}$ diverges, and we have that

$$\lim_{n\to\infty}\left|\frac{\frac{1}{n+1}}{\frac{1}{n}}\right| = \lim_{n\to\infty}\frac{n}{n+1} = \lim_{n\to\infty}\frac{1}{1+\frac{1}{n}} = 1;$$

2. the alternating harmonic series $\sum\limits_{n=1}^{\infty}\frac{(-1)^{n-1}}{n}$ converges conditionally, and we have that

$$\lim_{n\to\infty}\left|\frac{\frac{(-1)^n}{n+1}}{\frac{(-1)^{n-1}}{n}}\right| = \lim_{n\to\infty}\frac{n}{n+1} = \lim_{n\to\infty}\frac{1}{1+\frac{1}{n}} = 1;$$

3. the p-series $\sum\limits_{n=1}^{\infty}\frac{1}{n^p}$ converges (absolutely) or diverges depending on the value of the real number p;[55] however, for all real numbers p, we have that

$$\lim_{n\to\infty}\left|\frac{\frac{1}{(n+1)^p}}{\frac{1}{n^p}}\right| = \lim_{n\to\infty}\frac{n^p}{(n+1)^p} = \lim_{n\to\infty}\frac{1}{(1+\frac{1}{n})^p} = 1.$$

---

[55] Since all terms of a p-series are positive, the series cannot converge conditionally.

2.9.7 The Root Test

The Root Test. *Let* $\{a_n\}_{n=1}^{\infty}$ *be a sequence of real numbers, and assume that* $L := \lim\limits_{n\to\infty} \sqrt[n]{|a_n|}$ *exists (either as a real number or as* $+\infty$*).*

(a) If $L < 1$*, then the series* $\sum\limits_{n=1}^{\infty} a_n$ *converges absolutely.*

(b) If $L > 1$*,*[56] *then the series* $\sum\limits_{n=1}^{\infty} a_n$ *diverges.*

Remark: If $L = 1$, then the Root Test gives us no information about the convergence/divergence of the series $\sum\limits_{n=1}^{\infty} a_n$.

Proof. The proof is similar to that of the Ratio Test. Once again, we will prove (a) by using a suitably chosen convergent geometric series and the Comparison Test. We will prove (b) by showing that $\{a_n\}_{n=1}^{\infty}$ does not converge to 0, and consequently, the series $\sum\limits_{n=1}^{\infty} a_n$ diverges (by Theorem 2.9.2). We now give the formal details.

We first prove (a). So, suppose that $L < 1$. First, since all terms of the sequence $\{\sqrt[n]{|a_n|}\}_{n=1}^{\infty}$ are non-negative, we see that $L \geq 0$.[57]

Set $\varepsilon := \frac{1-L}{2}$; since $0 \leq L < 1$, we see that $0 < \varepsilon \leq \frac{1}{2}$. Now, using the fact that $\lim\limits_{n\to\infty} \sqrt[n]{|a_n|} = L$, we fix $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $|\sqrt[n]{|a_n|} - L| < \varepsilon$, and consequently, $\sqrt[n]{|a_n|} < L + \varepsilon = \frac{1+L}{2} =: q$. It follows that for all $n \in \mathbb{N}$ such that $n \geq N$, we have that $0 \leq |a_n| \leq q^n$. On the other hand, since $0 \leq L < 1$, we have that $\frac{1}{2} \leq q < 1$, and in particular, $|q| < 1$. So, the geometric series

$$\sum_{n=N}^{\infty} q^n$$

converges.[58] Thus, by the Comparison Test, the series

$$\sum_{n=N}^{\infty} |a_n|$$

---

[56] Possibly, $\lim\limits_{n\to\infty} \sqrt[n]{|a_n|} = +\infty$.

[57] Formally, $\{\sqrt[n]{|a_n|}\}_{n=1}^{\infty}$ is bounded below by 0, and so by Lemma 2.2.3(b), we have that $L = \lim\limits_{n\to\infty} \sqrt[n]{|a_n|} \geq 0$.

[58] Let us explain this in a bit more detail. We set $b := q^N$, and we note that $\sum\limits_{n=N}^{\infty} q^n = \sum\limits_{n=N}^{\infty} bq^{n-N} = \sum\limits_{n=0}^{\infty} bq^n$; since $|q| < 1$, this geometric series really does converge.

converges. If $N = 1$, then we are done. So, let us assume that $N \geq 2$. In this case, we simply observe that

$$\sum_{n=1}^{\infty} |a_n| \;=\; \underbrace{\Big(\sum_{n=1}^{N-1} |a_n|\Big)}_{\text{finite sum}} + \underbrace{\Big(\sum_{n=N}^{\infty} |a_n|\Big)}_{\text{convergent}},$$

and so $\sum\limits_{n=1}^{\infty} |a_n|$ converges. This proves (a).

It remains to prove (b). So, let us assume that $L > 1$.

> **Claim.** There exists some $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $|a_n| > 1$.

Proof of the Claim. Suppose first that $L = +\infty$, i.e. $\lim\limits_{n\to\infty} \sqrt[n]{|a_n|} = +\infty$. In this case, we know that there exists some $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $\sqrt[n]{|a_n|} > 1$,[59] and consequently, $|a_n| > 1$.

Let us now suppose that L is a real number. Set $\varepsilon := L - 1$; since $L > 1$, we see that $\varepsilon > 0$. Now, using the fact that $\lim\limits_{n\to\infty} \sqrt[n]{|a_n|} = L$, we fix some $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $|\sqrt[n]{|a_n|} - L| < \varepsilon$, and consequently, $L - \varepsilon < \sqrt[n]{|a_n|} < L + \varepsilon$. But note that $L - \varepsilon = 1$. Thus, for all $n \in \mathbb{N}$ such that $n \geq N$, we have that $\sqrt[n]{|a_n|} > L - \varepsilon = 1$, and consequently, $|a_n| > 1^n = 1$. ♦

We can now show that the sequence $\{a_n\}_{n=1}^{\infty}$ does not converge to 0. Suppose otherwise, that is suppose that $\lim\limits_{n\to\infty} a_n = 0$. Fix $N' \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N'$, then $|a_n| < 1$.[60] Now, let $N \in \mathbb{N}$ be as in the Claim, so that for all $n \in \mathbb{N}$ such that $n \geq N$, we have that $|a_n| > 1$. Then for $n := \max\{N, N'\}$, we have that

$$1 \overset{(*)}{<} |a_n| \overset{(**)}{<} 1,$$

where (*) follows from the fact that $n \geq N$, whereas (**) follows from the fact that $n \geq N'$. We have now obtained $1 < 1$, a contradiction. This proves that $\{a_n\}_{n=1}^{\infty}$ does not converge to 0. So, by Theorem 2.9.2, the series $\sum\limits_{n=1}^{\infty} a_n$ diverges. ☐

Example 2.9.9. *Consider the series* $\sum\limits_{n=1}^{\infty} \frac{(-1)^n n}{5^n}$. *Recall that* $\lim\limits_{n\to\infty} \sqrt[n]{n} = 1$ *(by Proposition 2.4.2), and note that*

$$\lim_{n\to\infty} \sqrt[n]{\left|\tfrac{(-1)^n n}{5^n}\right|} \;=\; \lim_{n\to\infty} \frac{\sqrt[n]{n}}{5} \;=\; \frac{1}{5} \;<\; 1.$$

---

[59] Indeed, set $M := 1$, and apply the definition of divergence to $+\infty$.

[60] Let us explain this in a bit more detail. Set $\varepsilon := 1$. Then, since $\lim\limits_{n\to\infty} a_n = 0$, there exists some $N' \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N'$, then $|a_n - 0| < \varepsilon$, that is, $|a_n| < 1$.

So, by the Root Test, the series $\sum\limits_{n=1}^{\infty} \frac{(-1)^n n}{5^n}$ *converges absolutely.*

Example 2.9.10. *Consider the series* $\sum\limits_{n=1}^{\infty} \left(\frac{16n+3}{-7n+6}\right)^n$*. Note that*

$$\begin{array}{rcl} \lim\limits_{n\to\infty} \sqrt[n]{\left|\left(\frac{16n+3}{-7n+6}\right)^n\right|} & = & \lim\limits_{n\to\infty} \left|\frac{16n+3}{-7n+6}\right| \\ & = & \lim\limits_{n\to\infty} \frac{16n+3}{7n-6} \\ & = & \lim\limits_{n\to\infty} \frac{16+\frac{3}{n}}{7-\frac{6}{n}} \\ & = & \frac{16}{7} \\ & > & 1, \end{array}$$

and so by the Root Test, the series $\sum\limits_{n=1}^{\infty} \left(\frac{-7n+6}{16n+3}\right)^n$ *diverges.*

Remark: As pointed out above, if $\{a_n\}_{n=1}^{\infty}$ is a sequence of real numbers such that $\lim\limits_{n\to\infty} \sqrt[n]{|a_n|} = 1$, then the Root Test gives us no information about the convergence/divergence of the series $\sum\limits_{n=1}^{\infty} a_n$, and in principle, it is possible that the series converges absolutely, converges conditionally, or diverges. Recall that $\lim\limits_{n\to\infty} \sqrt[n]{n} = 1$ (by Proposition 2.4.2), and consider, for example, the following three series:

1. the harmonic series $\sum\limits_{n=1}^{\infty} \frac{1}{n}$ diverges, and we have that
$$\lim_{n\to\infty} \sqrt[n]{\left|\tfrac{1}{n}\right|} \;=\; \lim_{n\to\infty} \frac{1}{\sqrt[n]{n}} \;=\; 1;$$
2. the alternating harmonic series $\sum\limits_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}$ converges conditionally, and we have that
$$\lim_{n\to\infty} \sqrt[n]{\left|\tfrac{(-1)^{n-1}}{n}\right|} \;=\; \lim_{n\to\infty} \frac{1}{\sqrt[n]{n}} \;=\; 1;$$
3. the p-series $\sum\limits_{n=1}^{\infty} \frac{1}{n^p}$ converges (absolutely) or diverges depending on the value of the real number p;[61] however, for all real numbers p, we have that
$$\lim_{n\to\infty} \sqrt[n]{\left|\tfrac{1}{n^p}\right|} \;=\; \lim_{n\to\infty} \frac{1}{(\sqrt[n]{n})^p} \;=\; \frac{1}{(\lim\limits_{n\to\infty} \sqrt[n]{n})^p} \;=\; \frac{1}{1^p} \;=\; 1.$$

---

[61] Since all terms of a p-series are positive, the series cannot converge conditionally.

2.9.8 Defining functions via series: the exponential function

Some important functions can be defined as series. For instance, we may define the function $\exp : \mathbb{R} \to \mathbb{R}$ by setting

$$\exp(x) \quad := \quad \sum_{n=0}^{\infty} \tfrac{x^n}{n!} \quad = \quad 1 + \tfrac{x}{1!} + \tfrac{x^2}{2!} + \tfrac{x^3}{3!} + \tfrac{x^4}{4!} + \tfrac{x^5}{5!} + \dots$$

for all $x \in \mathbb{R}$. Let us explain why this series converges for all $x \in \mathbb{R}$ (and so the function exp is well defined). Obviously, the series above converges for $x = 0$, and we have that $\exp(0) = 1$. On the other hand, for fixed $x \in \mathbb{R} \setminus \{0\}$, we have that

$$\lim_{n\to\infty} \left| \frac{\frac{x^{n+1}}{(n+1)!}}{\frac{x^n}{n!}} \right| \quad = \quad \lim_{n\to\infty} \tfrac{|x|}{n+1} \quad \overset{(*)}{=} \quad |x| \lim_{n\to\infty} \tfrac{1}{n+1} \quad = \quad |x| \cdot 0 \quad = \quad 0 \quad < \quad 1,$$

where (*) follows from the fact that x is fixed (and may therefore be treated as a constant). So, by the Ratio Test, the series $\exp(x)$ converges for all $x \in \mathbb{R} \setminus \{0\}$. This proves that the function exp is well defined.

Remark: It can be shown that $\exp(x) = e^x$ for all $x \in \mathbb{R}$; in other words, for all $x \in \mathbb{R}$, we have that

$$e^x \quad = \quad \sum_{n=0}^{\infty} \tfrac{x^n}{n!},$$

but we omit the proof. We do note that this implies, in particular, that

$$e \quad = \quad \sum_{n=0}^{\infty} \tfrac{1}{n!} \quad = \quad 1 + \tfrac{1}{1!} + \tfrac{1}{2!} + \tfrac{1}{3!} + \tfrac{1}{4!} + \tfrac{1}{5!} + \dots.$$

Chapter 3

Limits of functions

Notation: When we write "$f : A \subseteq \mathbb{R} \to \mathbb{R}$," we mean that A is a subset of $\mathbb{R}$, and f is a function from A to $\mathbb{R}$. Similarly, "$f : A \subsetneqq \mathbb{R} \to \mathbb{R}$" means that A is a proper subset of $\mathbb{R}$ (i.e. A is a subset of $\mathbb{R}$, but is not equal to $\mathbb{R}$), and f is a function from A to $\mathbb{R}$.

The domain of a function. In what follows, when we define a function f via a formula (without specifying the domain and the codomain), we will assume that the domain of f is the subset of all real numbers x for which the expression $f(x)$ makes sense and is a real number. For example:

- the domain of $f_1(x) = \frac{x-1}{x+3}$ is $\mathbb{R} \setminus \{-3\}$, whereas the codomain is $\mathbb{R}$;
- the domain of $f_2(x) = \sqrt{x}$ is $[0, +\infty)$, whereas the codomain is $\mathbb{R}$;
- both the domain and the codomain of $f_3(x) = \sqrt[3]{x}$ are $\mathbb{R}$.

Intervals. As usual, an *interval* in $\mathbb{R}$ is any set of the following form:

- $[a, b]$, where $a, b \in \mathbb{R}$ are such that $a < b$;
- $[a, b)$, where $a, b \in \mathbb{R}$ are such that $a < b$;
- $(a, b]$, where $a, b \in \mathbb{R}$ are such that $a < b$;
- (a, b), where $a, b \in \mathbb{R}$ are such that $a < b$;
- $(a, +\infty)$, where $a \in \mathbb{R}$;
- $(-\infty, b)$, where $b \in \mathbb{R}$;
- $(-\infty, +\infty)$.

A *closed interval* is an interval of the form $[a, b]$, whereas an *open interval* is an interval of the form (a, b), $(a, +\infty)$, $(-\infty, b)$, or $(-\infty, +\infty)$.

3.1 The limit of a function at a point: an informal introduction

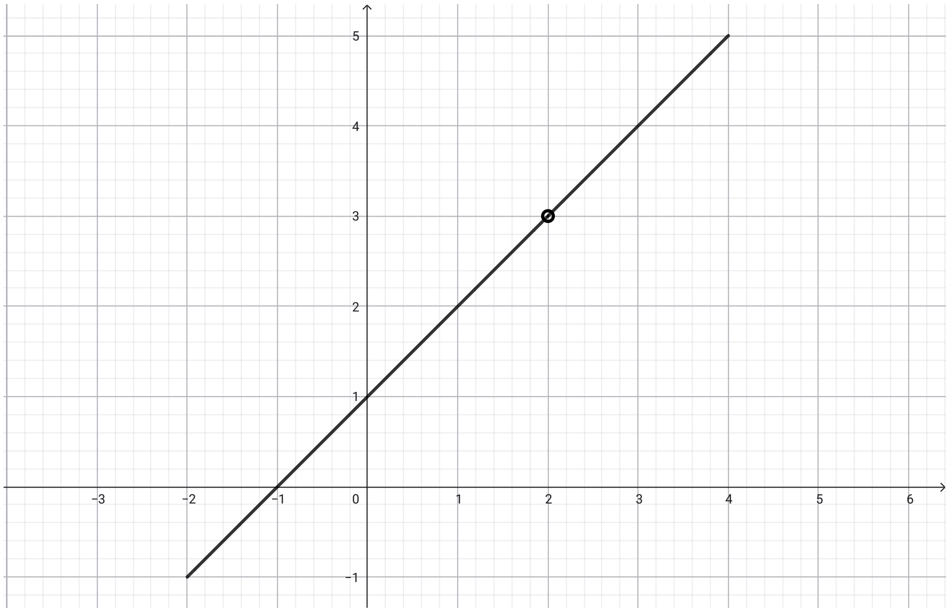
Suppose that $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is a function, and that $a \in \mathbb{R}$ (f may or may not be defined at a, i.e. both $a \in A$ and $a \notin A$ are possible). We would like to say that $\lim\limits_{x\to a} f(x) = L$ (which we read: "L is the limit of $f(x)$ as x approaches a") provided that whenever $x \approx a$, we have that $f(x) \approx L$. To make this formal, we need to explain what exactly "$\approx$" means, which we do in section 3.2. Here, we take a look at an (informal) example.

Example 3.1.1. *Consider the function* $f(x) = \frac{x^2-x-2}{x-2}$. *The domain of this function is* $\mathbb{R} \setminus \{2\}$ *(since we cannot divide by zero). Now, for* $x \in \mathbb{R} \setminus \{2\}$, *we have that*

$$f(x) \quad = \quad \tfrac{x^2-x-2}{x-2} \quad = \quad \tfrac{(x+1)(x-2)}{x-2} \quad = \quad x+1.$$

As pointed out above, $f(2)$ *is undefined. However (informally), for* $x \approx 2$, *we obviously have that* $f(x) \approx 2+1 = 3$, *and so* $\lim\limits_{x\to 2} f(x) = 3$.

Remark: Below is the graph of the function f from Example 3.1.1. The function is undefined at $x = 2$. However, for values of x close to 2, the value of $f(x)$ is close to 3, which is why $\lim\limits_{x\to 2} f(x) = 3$.



3.2 The Limit of a function at a point: the formal definition

In this section, we define accumulation points of subsets of $\mathbb{R}$, and we give a formal definition of a limit of a function at a point. Moreover, we prove the uniqueness of limits (provided they exist; see Theorem 3.2.3 below), and we take a look at a few examples.

Accumulation points of sets. An *accumulation point* of a set $A \subseteq \mathbb{R}$ is a point $a \in \mathbb{R}$ (note that a may or may not belong to A) such that for all real numbers $\varepsilon > 0$, there exists some $a' \in A$ such that $0 < |a' - a| < \varepsilon$.[1]

Remark: By definition, for a set $A \subseteq \mathbb{R}$ and a real number a (which may or may not belong to A), we have that a is **not** an accumulation point if and only if there exists some $\varepsilon > 0$ such that the set A contain **no points** from interval $(a - \varepsilon, a + \varepsilon)$ other than possibly a itself. So, a is an accumulation point of A if no such $\varepsilon > 0$ exists.

Example 3.2.1. *Every real number is an accumulation point of* $\mathbb{Q}$ *in* $\mathbb{R}$*. Indeed, consider any* $a \in \mathbb{R}$*, and fix any real number* $\varepsilon > 0$*. Since* $\mathbb{Q}$ *is dense in* $\mathbb{R}$*, we know that* $(a - \varepsilon, a + \varepsilon) \setminus \{a\} = (a - \varepsilon, a) \cup (a, a + \varepsilon)$ *contains a rational number* q*.*[2] *Then* $q \in \mathbb{Q} \setminus \{a\}$*, and moreover,* $0 < |q - a| < \varepsilon$*. So, all real numbers are indeed accumulation points of* $\mathbb{R}$*.*

Lemma 3.2.2. *Let* $A \subseteq \mathbb{R}$ *and* $a \in \mathbb{R}$*. If there exist* $p, q \in \mathbb{R}$ *such that* $p < a < q$ *and* $(p, q) \setminus \{a\} \subseteq A$*,*[3] *then* a *is an accumulation point of* A*.*

Proof. Assume that there exist some $p, q \in \mathbb{R}$ such that $p < a < q$ and $(p, q) \setminus \{a\} \subseteq A$. Fix $\varepsilon > 0$. Set $p' = \max\{p, a - \varepsilon\}$ and $q' = \min\{q, a + \varepsilon\}$. Then $p' < a < q'$ and $(p', q') \setminus \{a\} \subseteq A$. Now fix any $a' \in (p', q') \setminus \{a\}$. Then $a' \in A$ and $0 < |a' - a| < \varepsilon$. This proves that a is an accumulation point of A. □

The limit of a function at a point. Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, let $a \in \mathbb{R}$ be an accumulation point of A,[4] and let $L \in \mathbb{R}$. We say that L *is the limit of* $f(x)$ *as* x *approaches* a, or that $f(x)$ *tends to* L *as* x *approaches* a, provided that the following holds:

---

[1] So, a is an accumulation point of A if and only one can find a points of A arbitrarily close to a (but distinct from a).

[2] Indeed, by the density of $\mathbb{Q}$ in $\mathbb{R}$ we know that each of the intervals $(a - \varepsilon, a)$ and $(a, a + \varepsilon)$ contains a rational number.

[3] Equivalently: $(p, a) \cup (a, q) \subseteq A$.

[4] In practice, we will most often encounter the situation from the statement of Lemma 3.2.2: the function f will be defined on some (possibly very small) open interval containing a, except possibly at a itself. Occasionally, we will see slightly more complicated situations.

for every $\varepsilon > 0$, there exists some $\delta > 0$, such that for all $x \in A$, if $0 < |x - a| < \delta$, then $|f(x) - L| < \varepsilon$.

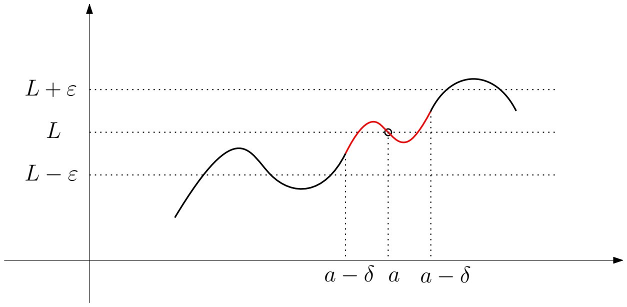
Under such circumstances, we write

$$L = \lim_{x \to a} f(x) \qquad \text{or} \qquad f(x) \to L \quad \text{as} \quad x \to a.$$

Remark: Suppose $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is a function, $a \in \mathbb{R}$ is an accumulation point of A, and L is a real number, as in the definition above. Then $L = \lim_{x \to a} f(x)$ means that for every $\varepsilon > 0$, we can choose some $\delta > 0$, such that

$$\big\{f(x) \mid x \in A \setminus \{a\}, x \in (a - \delta, a + \delta)\big\} \quad \subseteq \quad (L - \varepsilon, L + \varepsilon),$$

as in the picture below. In other words, all elements of $A \setminus \{a\}$ inside the interval $(a - \delta, a + \delta)$ get mapped by f to the interval $(L - \varepsilon, L + \varepsilon)$. The existence and value of $f(a)$ is irrelevant for the existence and value of $L = \lim_{x \to a} f(x)$.



Remark: We emphasize that the existence and value of $\lim_{x \to a} f(x)$ is determined by the behavior of $f(x)$ for values of x **close to** a, but is unaffected by the existence and value $f(a)$ itself. Indeed, each of the following in possible:

- $f(a)$ and $\lim_{x \to a} f(x)$ are both defined, and $f(a) = \lim_{x \to a} f(x)$;
- $f(a)$ and $\lim_{x \to a} f(x)$ are both defined, but $f(a) \neq \lim_{x \to a} f(x)$;
- $f(a)$ is defined, but $\lim_{x \to a} f(x)$ is undefined;
- $f(a)$ is undefined, but $\lim_{x \to a} f(x)$ is defined;
- $f(a)$ and $\lim_{x \to a} f(x)$ are both undefined.

Remark: In what follows, we will occasionally write something like "$\lim_{x \to a} f(x)$ exists (as a real number)." Of course, so far, we have defined $\lim_{x \to a} f(x)$ to be a real number (if it exists). However, we will later introduce the notion of infinite limits, and will introduce the notation "$\lim\limits_{x \to a} f(x) = +\infty$" and "$\lim\limits_{x \to a} f(x) = -\infty$." Hence, we sometimes emphasize that we wish $\lim\limits_{x \to a} f(x)$ to be a real number.

Theorem 3.2.3. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, and let a be an accumulation point of A. Then $f(x)$ has at most one limit as x approaches a.*

Proof. Suppose otherwise, and let L_1 and L_2 be two distinct limits of $f(x)$ as x approaches a. Set $\varepsilon := \frac{|L_1 - L_2|}{2}$. (Obviously, $\varepsilon > 0$.)

Using the fact that L_1 is a limit of $f(x)$ as x approaches a, we fix $\delta_1 > 0$ such that for all $x \in A$, if $0 < |x - a| < \delta_1$, then $|f(x) - L_1| < \varepsilon$.

Using the fact that L_2 is a limit of $f(x)$ as x approaches a, we fix $\delta_2 > 0$ such that for all $x \in A$, if $0 < |x - a| < \delta_2$, then $|f(x) - L_2| < \varepsilon$.

Set $\delta := \min\{\delta_1, \delta_2\}$. (Obviously, $\delta > 0$.) Fix $x \in A$ such that $0 < |x - a| < \delta$. Then $0 < |x - a| < \delta_1$ and $0 < |x - a| < \delta_2$; the former implies that $|f(x) - L_1| < \varepsilon$, whereas the latter implies that $|f(x) - L_2| < \varepsilon$. We now have the following:

$$\begin{array}{rcll} |L_1 - L_2| & = & \big|\big(L_1 - f(x)\big) + \big(f(x) - L_2\big)\big| & \\ & \leq & |L_1 - f(x)| + |f(x) - L_2| & \text{by the Triangle Inequality} \\ & = & |f(x) - L_1| + |f(x) - L_2| & \\ & < & \varepsilon + \varepsilon & \\ & = & 2\varepsilon & \\ & = & |L_1 - L_2| & \text{by the choice of } \varepsilon, \end{array}$$

a contradiction. ☐

Example 3.2.4. *Let $f(x) = \frac{x^2 - x - 2}{x - 2}$. Compute $\lim\limits_{x \to 2} f(x)$ (and give an "ε-δ proof" of the correctness of your answer).*

Remark: Here, we are revisiting Example 3.1.1, but this time, we give a formal proof of the fact that $\lim\limits_{x \to 2} f(x) = 3$.

Solution. Clearly, f is defined on $A := \mathbb{R} \setminus \{2\}$. Our goal is to show that $\lim\limits_{x \to 2} f(x) = 3$. Fix $\varepsilon > 0$. Let $\delta := \varepsilon$. Fix $x \in A$ such that $0 < |x - 2| < \delta$. (In particular, $x \neq 2$.) Then

$$|f(x) - 3| \;=\; \left|\tfrac{x^2 - x - 2}{x - 2} - 3\right| \;=\; |(x + 1) - 3| \;=\; |x - 2| \;<\; \delta \;=\; \varepsilon.$$

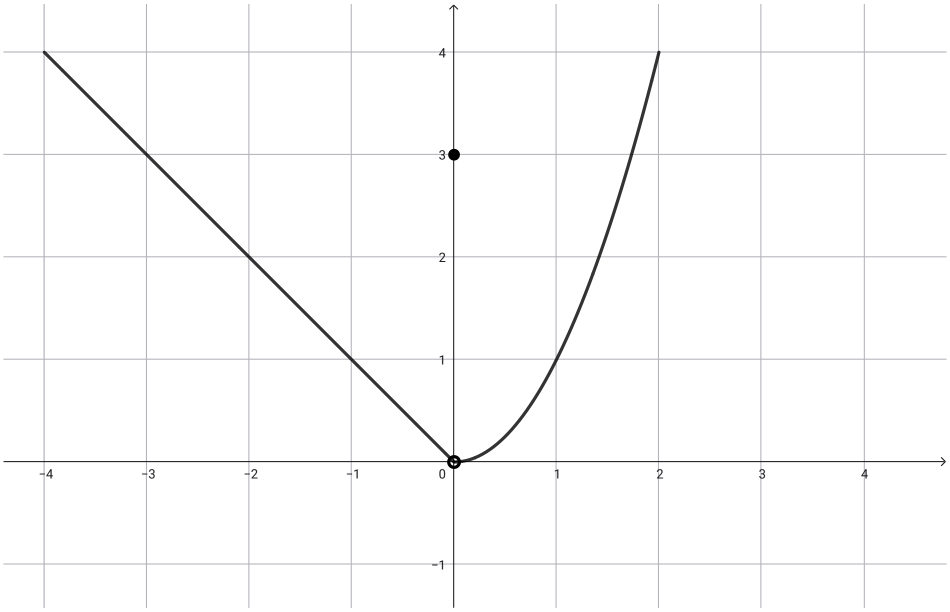
This proves that $\lim\limits_{x \to 2} f(x) = 3$, as we had claimed. ☐

Example 3.2.5. *Let* $f : \mathbb{R} \to \mathbb{R}$ *be given by*

$$f(x) \quad = \quad \begin{cases} -x & \textit{if} \quad x < 0 \\ 3 & \textit{if} \quad x = 0 \\ x^2 & \textit{if} \quad x > 0 \end{cases}$$

for all $x \in \mathbb{R}$*. Either find* $\lim\limits_{x \to 0} f(x)$ *(and give an "*ε*-*δ *proof" of the correctness of your answer), or explain why the limit does not exist (as a real number).*

Remark: The graph of this function is shown below. By definition, $f(0) = 3$. However, when x is very close to 0 (but not actually equal to 0), we have that $f(x)$ is very close to 0. So, $\lim\limits_{x \to 0} f(x) = 0$. Our solution (below) gives a formal proof of this fact.



Solution. Our goal is to show that $\lim\limits_{x \to 0} f(x) = 0$. Fix $\varepsilon > 0$. Let $\delta := \min\{\varepsilon, \sqrt{\varepsilon}\}$. Fix $x \in \mathbb{R}$ such that $0 < |x - 0| < \delta$. We must show that $|f(x) - 0| < \varepsilon$.

Since $0 < |x - 0| < \delta$, i.e. $0 < |x| < \delta$, we have that either $-\delta < x < 0$ or $0 < x < \delta$. We consider both cases.

Case 1: $-\delta < x < 0$. In this case, we have that $f(x) = -x$. We now compute:

$$|f(x) - 0| \quad = \quad |-x - 0| \quad = \quad |x| \quad < \quad \delta \quad = \quad \min\{\varepsilon, \sqrt{\varepsilon}\} \quad \leq \quad \varepsilon,$$

which is what we needed.

Case 2: $0 < x < \delta$. In this case, we have that $f(x) = x^2$ and $x^2 < \delta^2$. We now compute:

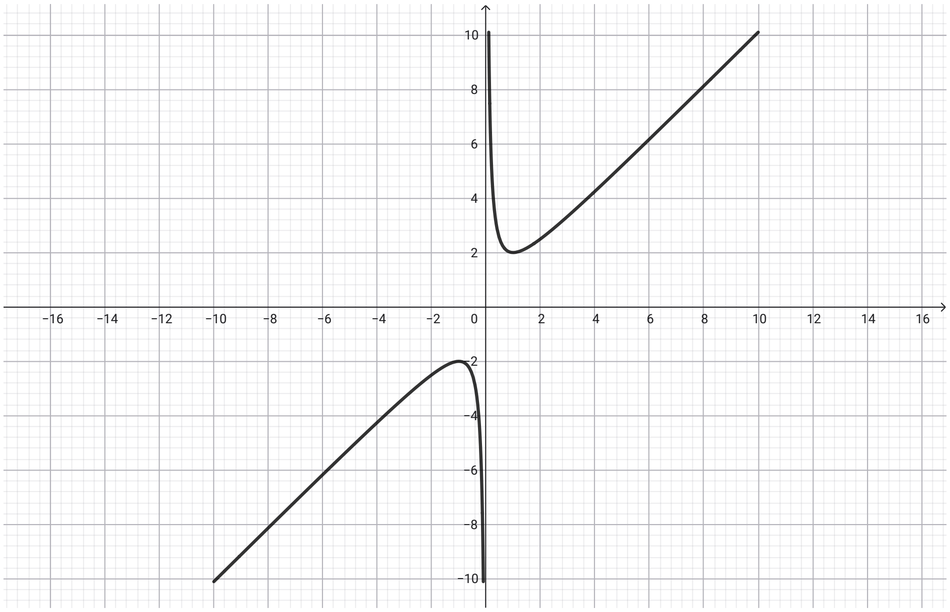
$$|f(x)-0| \quad = \quad |x^2-0| \quad = \quad x^2 \quad < \quad \delta^2 \quad = \quad \min\{\varepsilon^2,\varepsilon\} \quad \leq \quad \varepsilon,$$

which is what we needed.

So, in either case (i.e. both in Case 1 and in Case 2), we get that $|f(x)-0| < \varepsilon$. This proves that $\lim\limits_{x\to 0} f(x) = 0$. $\square$

Example 3.2.6. *Let $f : \mathbb{R}\setminus\{0\} \to \mathbb{R}$ be a given by $f(x) = x + \frac{1}{x}$. Either compute $\lim\limits_{x\to 0} f(x)$, or prove that the limit does not exist (as a real number).*

Remark: The graph of the function f is shown below. As we can see, $\lim\limits_{x\to 0} f(x)$ does not exist, which we will prove formally in our solution.



Solution. We will prove that $\lim\limits_{x\to 0} f(x)$ does not exist (as a real number). For this, we must prove the following:

> For all $L \in \mathbb{R}$, there exists some $\varepsilon > 0$, such that for all $\delta > 0$, there exists some $x \in \mathbb{R}\setminus\{0\}$, such that $0 < |x-0| < \delta$ and $|f(x)-L| \geq \varepsilon$.

Fix $L \in \mathbb{R}$, and fix an arbitrary $\varepsilon > 0$.[5] Fix $\delta > 0$. We need to exhibit some $x \in \mathbb{R}\setminus\{0\}$ such that $0 < |x-0| < \delta$ and $|f(x)-L| \geq \varepsilon$.

Claim. There exists some $x \in \mathbb{R}\setminus\{0\}$ such that $0 < x < \delta$ and $\frac{1}{x} > L+\varepsilon$.

---

[5] In this particular case, any $\varepsilon > 0$ will do. In other examples, ε may need to be carefully chosen.

Proof of the Claim. If $L + \varepsilon \leq 0$, then we simply choose any $x \in (0, \delta)$, and we observe that $0 < x < \delta$ and $\frac{1}{x} > 0 \geq L + \varepsilon$.

Suppose now that $L + \varepsilon > 0$. We then choose an arbitrary $x \in (0, \min\{\delta, \frac{1}{L+\varepsilon}\})$. Clearly, $0 < x < \delta$. On the other hand, we have that $0 < x < \frac{1}{L+\varepsilon}$, and consequently, $\frac{1}{x} > L + \varepsilon$. ♦

Now, fix x as in the Claim above. Since $0 < x < \delta$, we have that $0 < |x - 0| < \delta$. Further, we compute:

$$|f(x) - L| \;=\; |x + \tfrac{1}{x} - L| \;\geq\; x + \tfrac{1}{x} - L \;\overset{(*)}{>}\; 0 + (L + \varepsilon) - L \;=\; \varepsilon,$$

where (*) follows from the fact that $x > 0$ and $\frac{1}{x} > L + \varepsilon$. This completes the argument. □

3.3 A relationship between limits of functions and limits of sequences

Theorem 3.3.1. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$, let $a \in \mathbb{R}$ be an accumulation point of the set A,*[6] *and let $L \in \mathbb{R}$. Then the following are equivalent:*

(i) $\lim\limits_{x \to a} f(x) = L$;

(ii) for all sequences $\{a_n\}_{n=1}^{\infty}$ of real numbers that all belong to the set $A \setminus \{a\}$, if $\lim\limits_{n\to\infty} a_n = a$, then $\lim\limits_{n\to\infty} f(a_n) = L$.

Proof. First, we assume (i) and prove (ii). Fix a sequence $\{a_n\}_{n=1}^{\infty}$ of real numbers that all belong to the set $A \setminus \{a\}$, and assume that $\lim\limits_{n\to\infty} a_n = a$. We must show that $\lim\limits_{n\to\infty} f(a_n) = L$. Fix $\varepsilon > 0$. By (i), we have that $\lim\limits_{x\to a} f(x) = L$. So, fix $\delta > 0$ such that for all $x \in A$, if $0 < |x - a| < \delta$, then $|f(x) - L| < \varepsilon$. Now, using the fact that $\lim\limits_{n\to\infty} a_n = a$, we fix $N \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, if $n \geq N$, then $|a_n - a| < \delta$. But recall that $a_n \in A \setminus \{a\}$ (and in particular, $a_n \neq a$) for all $n \in \mathbb{N}$. We conclude that for all $n \in \mathbb{N}$, if $n \geq N$, then $0 < |a_n - a| < \delta$ (with $a_n \in A$), and therefore, by the choice of δ, we have that $|f(a_n) - L| < \varepsilon$. Thus, $\lim\limits_{n\to\infty} f(a_n) = L$, which proves (ii).

We now assume that (i) is false, and we show that (ii) is false. Using the fact that (i) is false, we fix $\varepsilon > 0$ such that for all $\delta > 0$, there exists some $x \in A$ such that $0 < |x - a| < \delta$ and $|f(x) - L| \geq \varepsilon$. Now, for all $n \in \mathbb{N}$, we choose some $a_n \in A$ such that $0 < |a_n - a| < \frac{1}{n}$ and $|f(a_n) - L| \geq \varepsilon$.[7] Obviously, $\{a_n\}_{n=1}^{\infty}$ is a sequence

---

[6] As usual, a may, but need not, belong to the set A.

[7] Let us explain why such an a_n exists. Here, we are setting $\delta := \frac{1}{n}$. By our assumption, we know there exists some $x \in A$ such that $0 < |x - a| < \delta$ and $|f(x) - L| \geq \varepsilon$; we let a_n be any such x.

Remark: Note that we made infinitely many arbitrary choices: for each $n \in \mathbb{N}$, there may have been many suitable "candidates" for a_n and we chose one arbitrarily. We can do this because of the so called "Axiom of Choice," which you will study if you take Set Theory.

of real numbers that all belong to $A \setminus \{a\}$. Our goal is to show that $\lim\limits_{n\to\infty} a_n = a$, but that the sequence $\{f(a_n)\}_{n=1}^{\infty}$ fails to converge to L; this will prove that (ii) is false.

We first show that $\lim\limits_{n\to\infty} a_n = a$. Fix $\varepsilon' > 0$. Let $N \in \mathbb{N}$ be such that $N > \frac{1}{\varepsilon'}$. Fix any $n \in \mathbb{N}$ such that $n \geq N$. Then

$$\begin{array}{rcll} |a_n - a| & < & \frac{1}{n} & \text{by the choice of } a_n \\ & \leq & \frac{1}{N} & \text{because } n \geq N \\ & < & \varepsilon' & \text{because } N > \frac{1}{\varepsilon'}. \end{array}$$

This proves that $\lim\limits_{n\to\infty} a_n = a$.

It remains to show that the sequence $\{f(a_n)\}_{n=1}^{\infty}$ does not converge to L. By the construction of our sequence $\{a_n\}_{n=1}^{\infty}$, we know that for all $n \in \mathbb{N}$, we have that $|f(a_n) - L| \geq \varepsilon$, that is, $f(a_n) \notin (L - \varepsilon, L + \varepsilon)$. By Proposition 2.6.1, this implies that L is not an accumulation point of the sequence $\{f(a_n)\}_{n=1}^{\infty}$.[8] Therefore, by Theorem 2.6.2, the sequence $\{f(a_n)\}_{n=1}^{\infty}$ does not converge to L. This completes the argument. □

3.4 Properties of limits

Lemma 3.4.1. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$, and let a be an accumulation point of A. Let $c \in \mathbb{R}$ be such that $f(x) = c$ for all $x \in A \setminus \{a\}$.*[9] *Then $\lim\limits_{x\to a} f(x) = c$.*

Proof. Fix $\varepsilon > 0$. Fix an arbitrary $\delta > 0$.[10] Fix $x \in A$ such that $0 < |x - a| < \delta$. In particular, $x \neq a$, and we see that $x \in A \setminus \{a\}$. So, $f(x) = c$, and we deduce that $|f(x) - c| = 0 < \varepsilon$. □

Lemma 3.4.2. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$, and let $a \in \mathbb{R}$ be an accumulation point of A. If $\lim\limits_{x\to a} f(x)$ exists (as a real number), then there exist a real number $\delta > 0$ such that the set $\{f(x) \mid x \in A, |x - a| < \delta\}$ is bounded.*

Remark: Informally, this lemma states that if $\lim\limits_{x\to a} f(x)$ exists (as a real number), then $f(x)$ is bounded for $x \approx a$.

Proof. Assume that $L := \lim\limits_{x\to a} f(x)$ exists (and is a real number). Fix $\varepsilon > 0$. Using the fact that $L = \lim\limits_{x\to a} f(x)$, we fix $\delta > 0$ such that for all $x \in A$, if $0 < |x - a| < \delta$,

---

[8] Note that we are using the "(i) $\iff$ (ii)" part of Proposition 2.6.1.

[9] So, $f(a)$ may or may not be defined; if $f(a)$ is defined, it may or may not be equal to c. However, the function f is assumed to be constant on $A \setminus \{a\}$, always taking the value c.

[10] In this particular case, any $\delta > 0$ will do. In other proofs, δ may need to be picked more carefully.

then $|f(x) - L| < \varepsilon$ (i.e. $L - \varepsilon < f(x) < L + \varepsilon$). If $a \in A$ (i.e. $f(a)$ is defined), then let $p = \min\{L - \varepsilon, f(a)\}$ and $q = \max\{L + \varepsilon, f(a)\}$; if $f(a)$ is not defined, then we let $p = L - \varepsilon$ and $q = L + \varepsilon$. Then by construction, we have that for all $x \in A$, if $|x - a| < \delta$, then $p \leq f(x) \leq q$. This proves that the set $\{f(x) \mid x \in A, |x - a| < \delta\}$ is bounded (indeed, p is a lower bound, and q is an upper bound). $\square$

Lemma 3.4.3. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$, and let $a \in \mathbb{R}$ be an accumulation point of A. Assume that $L := \lim\limits_{x \to a} f(x)$ exists (as a real number), and let $L' \in \mathbb{R} \setminus \{L\}$. Then a is an accumulation point of the set $A' := \{x \in A \mid f(x) \neq L'\}$. Moreover, there exists a real number $\delta > 0$ such that $(A \setminus \{a\}) \cap (a - \delta, a + \delta) \subseteq A'$.*

Proof. We start with a technical claim.

> **Claim.** There exists a real number $\delta > 0$ such that for all $x \in A$, if $0 < |x - a| < \delta$, then $f(x) \neq L'$.

Proof of the Claim. Set $\varepsilon := \frac{|L'-L|}{2}$; since $L' \neq L$, we know that $\varepsilon > 0$. Using the fact that $L = \lim\limits_{x \to a} f(x)$, we fix some $\delta > 0$ such that for all $x \in A$, if $0 < |x - a| < \delta$, then $|f(x) - L| < \varepsilon$, i.e. $f(x) \in (L - \varepsilon, L + \varepsilon)$. It is now enough to show that $L' \notin (L - \varepsilon, L + \varepsilon)$, for the Claim will then follow immediately. We consider two cases: when $L' < L$, and when $L' > L$.

Suppose first that $L' < L$. Then $\varepsilon = \frac{|L'-L|}{2} = \frac{L-L'}{2}$, and we deduce that $L - \varepsilon = L - \frac{L-L'}{2} = \frac{L+L'}{2} > L'$. Consequently, $L' \notin (L - \varepsilon, L + \varepsilon)$.

Suppose now that $L' > L$. Then $\varepsilon = \frac{|L'-L|}{2} = \frac{L'-L}{2}$, and we deduce that $L + \varepsilon = L + \frac{L'-L}{2} = \frac{L+L'}{2} < L'$, and it follows that $L' \notin (L - \varepsilon, L + \varepsilon)$. This completes the proof of the Claim. ◆

Let δ be as in the Claim. Then clearly, $(A \setminus \{a\}) \cap (a - \delta, a + \delta) \subseteq A'$.[11]

It remains to show that a is an accumulation point of A'. Fix an arbitrary $\varepsilon > 0$. Set $\varepsilon' := \min\{\delta, \varepsilon\}$. Using the fact that a is an accumulation point of A, we fix $a' \in A$ such that $0 < |a' - a| < \varepsilon'$. Since $a' \in A$ and $0 < |a' - a| < \varepsilon' \leq \delta$, our choice of δ guarantees that $f(a') \neq L'$, and it follows that $a' \in A'$. On the other hand, we obviously have that $0 < |a' - a| < \varepsilon' \leq \varepsilon$. This proves that a is indeed an accumulation point of A'. $\square$

Theorem 3.4.4. *Let $f_1 : A \subseteq \mathbb{R} \to \mathbb{R}$ and $f_2 : A \subseteq \mathbb{R} \to \mathbb{R}$ be functions, and let $a \in \mathbb{R}$ be an accumulation point of A. Assume that both $\lim\limits_{x \to a} f_1(x)$ and $\lim\limits_{x \to a} f_2(x)$ exist. Then all the following hold:*

(a) for all $c \in \mathbb{R}$, $\lim\limits_{x \to a} (cf_1)(x) = c \lim\limits_{x \to a} f_1(x)$;

---

[11] Indeed, fix $a' \in (A \setminus \{a\}) \cap (a - \delta, a + \delta)$. Then $a' \in A$, $a' \neq a$, and $|a' - a| < \delta$. Consequently, $a' \in A$ and $0 < |a' - a| < \delta$. So, by the choice of δ, we have that $f(a') \neq L'$. It follows that $a' \in A'$. This proves that $(A \setminus \{a\}) \cap (a - \delta, a + \delta) \subseteq A'$.

(b) $\lim\limits_{x\to a}(f_1+f_2)(x) = (\lim\limits_{x\to a} f_1(x)) + (\lim\limits_{x\to a} f_2(x))$*;*

(c) $\lim\limits_{x\to a}(f_1-f_2)(x) = (\lim\limits_{x\to a} f_1(x)) - (\lim\limits_{x\to a} f_2(x))$*;*

(d) $\lim\limits_{x\to a}(f_1f_2)(x) = (\lim\limits_{x\to a} f_1(x))(\lim\limits_{x\to a} f_2(x))$*;*

(e) if $\lim\limits_{x\to a} f_2(x) \neq 0$*, then* a *is an accumulation point of the set* $A' := \{x \in A \mid f_2(x) \neq 0\}$*, and we have that* $\lim\limits_{x\to a}(\frac{f_1}{f_2})(x) = \frac{\lim_{x\to a} f_1(x)}{\lim_{x\to a} f_2(x)}$*, where we consider the domain of the function* $\frac{f_1}{f_2}$ *to be* A'*.*

Remark: The proof of this theorem is very similar to that of the corresponding theorem for sequences (see Theorem 2.2.5 and its proof).

Proof. Set

$$L_1 := \lim_{x\to a} f_1(x) \qquad \text{and} \qquad L_2 := \lim_{x\to a} f_2(x).$$

We must prove the following:

(a) for all $c \in \mathbb{R}$, $\lim\limits_{x\to a}(cf_1)(x) = cL_1$;

(b) $\lim\limits_{x\to a}(f_1+f_2)(x) = L_1 + L_2$;

(c) $\lim\limits_{x\to a}(f_1-f_2)(x) = L_1 - L_2$;

(d) $\lim\limits_{x\to a}\big(f_1f_2\big)(x) = L_1L_2$;

(e) if $L_2 \neq 0$, then a is an accumulation point of the set $A' := \{x \in A \mid f_2(x) \neq 0\}$, and we have that $\lim\limits_{x\to a}(\frac{f_1}{f_2})(x) = \frac{L_1}{L_2}$, where we consider the domain of the function $\frac{f_1}{f_2}$ to be A'.

(a) Fix $c \in \mathbb{R}$. If $c = 0$, then the result follows from Lemma 3.4.1. So assume that $c \neq 0$. Fix $\varepsilon > 0$. Using the fact that $\lim\limits_{x\to a} f_1(x) = L_1$, we fix $\delta > 0$ such that for all $x \in A$, if $0 < |x-a| < \delta$, then $|f_1(x) - L_1| < \frac{\varepsilon}{|c|}$. We now fix $x \in A$ such that $0 < |x-a| < \delta$, and we observe that

$$|cf_1(x) - cL_1| \quad = \quad |c||f_1(x) - L_1| \quad < \quad |c|\tfrac{\varepsilon}{|c|} \quad = \quad \varepsilon.$$

This proves that $\lim\limits_{x\to a}(cf_1) = cL_1$.

(b) Fix $\varepsilon > 0$. Using the fact that $\lim\limits_{x\to a} f_1(x) = L_1$, we choose $\delta_1 > 0$ such that for all $x \in A$, if $0 < |x-a| < \delta_1$, then $|f_1(x) - L_1| < \frac{\varepsilon}{2}$. Similarly, using the fact that $\lim\limits_{x\to a} f_2(x) = L_2$, we choose $\delta_2 > 0$ such that for all $x \in A$, if $0 < |x-a| < \delta_2$, then $|f_2(x) - L_2| < \frac{\varepsilon}{2}$. Now, set $\delta := \min\{\delta_1, \delta_2\}$. Fix $x \in A$ such that $0 < |x-a| < \delta$.

Then both $0 < |x-a| < \delta_1$ and $0 < |x-a| < \delta_2$ hold; the former implies that $|f_1(x) - L_1| < \frac{\varepsilon}{2}$, whereas the latter implies that $|f_2(x) - L_2| < \frac{\varepsilon}{2}$. We now compute

$$\begin{aligned} \big|(f_1+f_2)(x) - (L_1+L_2)\big| &= \big|\big(f_1(x) - L_1\big) + \big(f_2(x) - L_2\big)\big| \\ &\overset{(*)}{\leq} |f_1(x) - L_1| + |f_2(x) - L_2| \\ &< \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} \;=\; \varepsilon, \end{aligned}$$

where (*) follows from the Triangle Inequality. This proves that $\lim\limits_{x\to a}(f_1+f_2)(x) = L_1 + L_2$.

(c) This follows from (a) and (b). Indeed,

$$\begin{aligned} \lim_{x\to a}(f_1 - f_2)(x) &= \lim_{x\to a}(f_1 + (-1)f_2)(x) \\ &\overset{\text{(a)}}{=} \big(\lim_{x\to a} f_1(x)\big) + \big(\lim_{x\to a}(-1)f_2(x)\big) \\ &\overset{\text{(b)}}{=} \big(\lim_{x\to a} f_1(x)\big) + (-1)\big(\lim_{x\to a} f_2(x)\big) \\ &= L_1 - L_2, \end{aligned}$$

and we are done.

(d) Using Lemma 3.4.2,[12] we fix $\delta_0 > 0$ such that the set $\{f_1(x) \mid x \in A, |x-a| < \delta_0\}$ is bounded. Fix a real number $M > 0$ such that for all $x \in A$ that satisfy $|x-a| < \delta_0$, we have that $|f_1(x)| < M$.[13]

Now, fix $\varepsilon > 0$. We must show that there exists some $\delta > 0$ such that for all $x \in A$, if $0 < |x-a| < \delta$, then $|(f_1f_2)(x) - L_1L_2| < \varepsilon$.

Using the fact that $\lim\limits_{x\to a} f_1(x) = L_1$, we fix $\delta_1 > 0$ such that for all $x \in A$, if $0 < |x-a| < \delta_1$, then $|f_1(x) - L_1| < \frac{\varepsilon}{2(|L_2|+1)}$.

Using the fact that $\lim\limits_{x\to a} f_2(x) = L_2$, we fix $\delta_2 > 0$ such that for all $x \in A$, if $0 < |x-a| < \delta_2$, then $|f_2(x) - L_2| < \frac{\varepsilon}{2M}$.

Now, set $\delta := \min\{\delta_0, \delta_1, \delta_2\}$. Now, fix $x \in A$ such that $0 < |x-a| < \delta$. Note that:

- $|f_1(x)| < M$ (because $x \in A$ and $|x-a| < \delta_0$);

---

[12] We are using the fact that $\lim\limits_{x\to a} f_1(x)$ exists, and so Lemma 3.4.2 applies.

[13] Let us explain why such an M exists. Let m_- be a lower bound of the set $\{f_1(x) \mid x \in A, |x-a| < \delta_0\}$, and let m^+ be an upper bound of the set. We can now choose M to be any real number that satisfies $M > \max\{|m_-|, |m_+|\}$.

- $|f_1(x) - L_1| < \frac{\varepsilon}{2(|L_2|+1)}$ (because $x \in A$ and $0 < |x-a| < \delta_1$);
- $|f_2(x) - L_2| < \frac{\varepsilon}{2M}$ (because $x \in A$ and $0 < |x-a| < \delta_2$).

We now compute:

$$\begin{array}{rcl} \big|(f_1f_2)(x) - L_1L_2\big| & = & \big|\big((f_1f_2)(x) - f_1(x)L_2\big) + \big(f_1(x)L_2 - L_1L_2\big)\big| \\ & \overset{(*)}{\leq} & \big|(f_1f_2)(x) - f_1(x)L_2| + |f_1(x)L_2 - L_1L_2\big| \\ & = & \big|f_1(x)\big(f_2(x) - L_2\big)\big| + \big|\big(f_1(x) - L_1\big)L_2| \\ & = & |f_1(x)||f_2(x) - L_2| + |f_1(x) - L_1||L_2| \\ & < & M\frac{\varepsilon}{2M} + \frac{\varepsilon}{2(|L_2|+1)}|L_2| \quad < \quad \varepsilon, \end{array}$$

where (*) follows from the Triangle Inequality. This proves that $\lim\limits_{x \to a}(f_1f_2)(x) = L_1L_2$.

(e) Assume that $L_2 \neq 0$. By Lemma 3.4.3, we have that a is an accumulation point of the set $A' = \{x \in A \mid f(x) \neq 0\}$.[14] It remains to show that $\lim\limits_{x \to a} \frac{1}{f_2(x)} = \frac{1}{L_2}$ (where the domain of the function $\frac{1}{f_2}$ is assumed to be A'), for then we will have that

$$\lim_{x \to a}\big(\tfrac{f_1}{f_2}\big)(x) \quad = \quad \overset{(d)}{=} \quad \big(\lim_{x \to a} f_1(x)\big)\big(\lim_{x \to a} \tfrac{1}{f_2(x)}\big) \quad = \quad L_1 \cdot \tfrac{1}{L_2} \quad = \quad \tfrac{L_1}{L_2},$$

which is what we need.

> **Claim.** There exists some $\delta > 0$ such that for all $x \in A$, if $0 < |x-a| < \delta$, then $|f_2(x)| > \frac{|L_2|}{2}$.

Proof of the Claim. Set $\varepsilon = \frac{|L_2|}{2}$. (Since $L_2 \neq 0$, we see that $\varepsilon > 0$.) Using the fact that $\lim\limits_{x \to a} f_2(x) = L_2$, we fix some $\delta > 0$ such that for all $x \in A$, if $0 < |x-a| < \delta$, then $|f_2(x) - L_2| < \varepsilon$, that is $f_2(x) \in (L_2 - \varepsilon, L_2 + \varepsilon)$. Now, consider such an x, i.e. $x \in A$ and $0 < |x-a| < \delta$, and consequently, $f_2(x) \in (L_2 - \varepsilon, L_2 + \varepsilon)$. We must show that $|f_2(x)| > \frac{|L_2|}{2}$.

Suppose first that $L_2 > 0$, so that $L_2 = |L_2|$. Then $f(x) > L_2 - \varepsilon = L_2 - \frac{|L_2|}{2} = \frac{|L_2|}{2}$, and we are done.

Suppose now that $L_2 < 0$, so that $L_2 = -|L_2|$. Then $f(x) < L_2 + \varepsilon = -|L_2| + \frac{|L_2|}{2} = -\frac{|L_2|}{2}$. So, $f(x) < -\frac{|L_2|}{2} < 0$, which implies that $|f_2(x)| = -f(x) > \frac{|L_2|}{2}$, and once again we are done. ♦

---

[14] We are simply applying Lemma 3.4.3, with L_2 and 0 playing the roles of L and L', respectively.

Let $\delta > 0$ be as in the Claim. Now, fix $\varepsilon > 0$. Using the fact that $\lim\limits_{x \to a} f(x) = L_2$, we fix $\delta' > 0$ such that for all $x \in A$, if $0 < |x - a| < \delta'$, then $|f_2(x) - L_2| < \frac{L_2^2}{2}\varepsilon$. (Here, we are using the fact that $L_2 \neq 0$ and $\varepsilon > 0$, and so $\frac{L^2}{2}\varepsilon > 0$.) Next, set $\delta'' := \min\{\delta, \delta'\}$. Fix an arbitrary $x \in A'$ such that $0 < |x - a| < \delta''$. Then

- $|f_2(x) - L_2| < \frac{L_2^2}{2}\varepsilon$ (because $x \in A$ and $0 < |x - a| < \delta'$);
- $|f_2(x)| > \frac{|L_2|}{2}$ (because $x \in A$ and $0 < |x - a| < \delta$).

We now compute:

$$|\tfrac{1}{f_2(x)} - \tfrac{1}{L_2}| \;=\; |\tfrac{L_2 - f_2(x)}{f_2(x)L_2}| \;=\; \tfrac{|L_2 - f_2(x)|}{|f_2(x)||L_2|} \;<\; \frac{\frac{L_2^2}{2}\varepsilon}{\frac{|L_2|}{2}|L_2|} \;=\; \varepsilon.$$

This completes the argument. □

3.5 Continuity

Continuity of a function at a point. A function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is *continuous* at a point $a \in A$ if the following holds:

> for all $\varepsilon > 0$, there exists some $\delta > 0$, such that for all $x \in A$, if $|x - a| < \delta$, then $|f(x) - f(a)| < \varepsilon$.

Remark: Suppose that $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is a function, and $a \in A$, as in the definition above. Note that if a is an accumulation point of A, then f is continuous at a if and only if $\lim\limits_{x \to a} f(x) = f(a)$ (and in particular, $\lim\limits_{x \to a} f(x)$ exists). If a is **not** an accumulation point of A, then f is automatically continuous at a.

Continuous functions. If a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is continuous at all points $a \in A$, then we simply say that f is *continuous*. If $I \subseteq A$ is an interval, then we say that f is *continuous* on I provided that $f \restriction I$ is continuous at all points in I.[15]

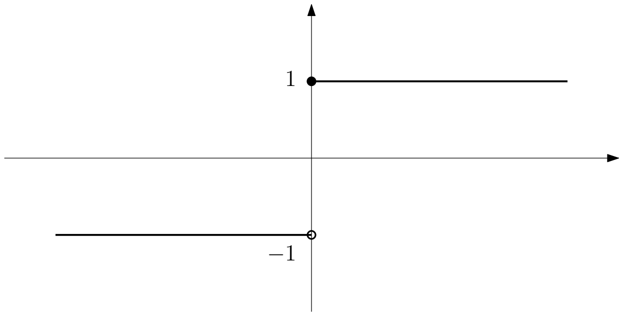
Remark: Suppose that $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is a function and that $I \subseteq A$ is an interval. If f is continuous on all points in I, then f is indeed continuous on I, as we would expect. Somewhat surprisingly, the converse is false in general! For instance, consider the function $f : \mathbb{R} \to \infty$ given by

$$f(x) \quad = \quad \begin{cases} -1 & \text{if} \quad x < 0 \\ 1 & \text{if} \quad x \geq 0 \end{cases}$$

for all $x \in \mathbb{R}$. (The graph of this function is given below.) Then f is **not** continuous at the point $x = 0$, but it **is** continuous at on interval $[0, +\infty)$. This is because whether or not a function is continuous on an interval is determined solely by the behavior of the function on the interval in question, and not by its behavior elsewhere.

---

[15] Here, $f \restriction I$ is the restriction of f to I. In other words, $(f \restriction I) : I \to \mathbb{R}$ is the function given by $(f \restriction I)(x) = f(x)$ for all $x \in I$.



Remark: The intuition behind the concept of continuity is that a function is supposed to be continuous on an interval if its graph (restricted to the interval in question) can be drawn on a piece of paper without lifting the pen. Of course, pens and paper are not mathematical objects! The definition above attempts to formalize our intuition, although some functions that are continuous according to our formal definition are somewhat strange, and drawing their graphs (with our without lifting the pen) would be quite difficult. We do, however, have the following fact, which we state without proof.

Fact 3.5.1. *The following functions are all continuous:*

- *polynomial functions,*
- *rational functions,*
- *root functions,*
- *exponential and logarithmic functions,*
- *trigonometric and inverse trigonometric functions.*

Theorem 3.5.2. *Let* $f_1 : A \subseteq \mathbb{R} \to \mathbb{R}$ *and* $f_2 : A \subseteq \mathbb{R} \to \mathbb{R}$ *be functions that are continuous at a point* $a \in A$*. Then all the following hold:*

(a) for all $c \in \mathbb{R}$*, the function* cf_1 *is continuous at* a*;*

(b) functions $f_1 + f_2$*,* $f_1 - f_2$*, and* $f_1 f_2$ *are continuous at* a*;*

(c) if $f_2(a) \neq 0$*, then the function* $\frac{f_1}{f_2}$ *is continuous at* a*.*

Proof. If a is an accumulation point of A, then this readily follows from Theorem 3.4.4 and from the definition of continuity. On the other hand, if a is **not** accumulation point of A, then any real-valued function whose domain is A (or any subset of A that contains the point a) is continuous at a. $\square$

Theorem 3.5.3. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ and $g : B \subseteq \mathbb{R} \to \mathbb{R}$ (with $f[A] \subseteq B$, so that the function $g \circ f : A \subseteq \mathbb{R} \to \mathbb{R}$ is defined). Let $a \in \mathbb{R}$, and assume that g is continuous at $b := \lim\limits_{x \to a} f(x)$ (in particular, this limit exists and belongs to B). Then:*

(a) $\lim\limits_{x \to a} (g \circ f)(x) = g\big(\lim\limits_{x \to a} f(x)\big) = g(b)$;

(b) if f is continuous at a, then $g \circ f$ is continuous at a.

Proof. We begin by observing that a is an accumulation point of A, since $\lim\limits_{x \to a} f(x)$ exists.

We first prove (a). Fix $\varepsilon > 0$. Using the fact that g is continuous at b, we fix some $\varepsilon' > 0$ such that for all $x \in B$, if $|x - b| < \varepsilon'$, then $|g(x) - g(b)| < \varepsilon$. Next, using the fact that $b = \lim\limits_{x \to a} f(a)$, we fix some $\delta > 0$ such that for all $x \in A$, if $0 < |x - a| < \delta$, then $|f(x) - b| < \varepsilon'$. Now, fix $x \in A$ such that $0 < |x - a| < \delta$. Then $|f(x) - b| < \varepsilon'$. Since $f[A] \subseteq B$, we see that $f(x) \in B$. Now by the choice of ε', we have that $|g(f(x)) - g(b)| < \varepsilon$, i.e. $|(g \circ f)(x) - g(b)| < \varepsilon$. This proves (a).

For (b), we simply observe that, if f is continuous at a, then

$$\begin{aligned} \lim_{x \to a} (g \circ f)(x) &= g\big(\lim_{x \to a} f(x)\big) && \text{by (a)} \\ &= g(f(a)) && \text{because } f \text{ is continuous at } a \\ &= (g \circ f)(a), \end{aligned}$$

which by definition means that $g \circ f$ is continuous at a. □

Example 3.5.4. *Compute $\lim\limits_{x \to 2} e^{\frac{x^2 - 3x + 2}{x - 2}}$.*

Solution. Since the function e^x is continuous, we may compute:

$$\lim_{x \to 2} e^{\frac{x^2-3x+2}{x-2}} = e^{\lim\limits_{x \to 2} \frac{x^2-3x+2}{x-2}} = e^{\lim\limits_{x \to 2} \frac{(x-2)(x-1)}{x-2}} = e^{\lim\limits_{x \to 2} (x-1)} = e,$$

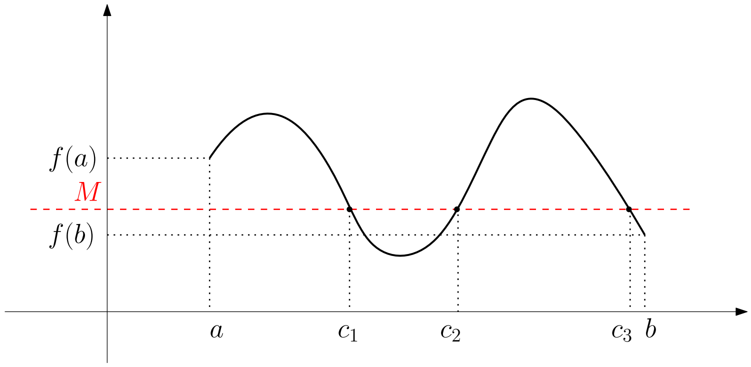
and we are done. □

3.5.1 The Intermediate Value Theorem

The Intermediate Value Theorem. *Let a and b be real numbers such that $a < b$, let $f : [a, b] \to \mathbb{R}$ be a continuous function, and let $M \in \mathbb{R}$ be such that* $\min\{f(a), f(b)\} < M < \max\{f(a), f(b)\}$.[16] *Then there exists some $c \in (a, b)$ such that $f(c) = M$.*

---

[16] So, we are assuming that $f(a) \neq f(b)$, and that either $f(a) < M < f(b)$ or $f(b) < M < f(a)$.



Remark: In practice, we may have a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ that is continuous on some closed interval $[a, b] \subseteq A$, though not necessarily on all of A. We can still apply the Intermediate Value Theorem to f and $[a, b]$. Technically, we are applying the theorem to the function $f \restriction [a, b]$, i.e. the restriction of f to the interval $[a, b]$.

Proof. By hypothesis, we have that either $f(a) < M < f(b)$ or $f(b) < M < f(a)$. We will prove the theorem for the case when $f(a) < M < f(b)$; the other case is similar. Let

$$S \quad := \quad \{x \in [a, b] \mid f(x) < M\}.$$

Then $a \in S$ (because $f(a) < M$), and in particular, $S \neq \emptyset$. On the other hand, the set S is bounded above (by b). So, by the completeness of the ordered field $\mathbb{R}$, S has a supremum, call it c. Our goal is to show that $c \in (a, b)$ and $f(c) = M$.

Claim 1. $c \in [a, b]$, and in particular, $f(c)$ is defined.

Proof of Claim 1. Since c is an upper bound of S, and $a \in S$, we have that $a \leq c$. On the other hand, since b is an upper bound of S, and c is the supremum (i.e. the least upper bound) of S, we have that $c \leq b$. So, $c \in [a, b]$. This proves Claim 1. ◆

Claim 2. $c < b$.

Proof of Claim 2. By Claim 1, we have that $c \in [a, b]$, and consequently, $c \leq b$. So, we need only show that $c \neq b$. Suppose otherwise, i.e. suppose that $c = b$, so that b is the supremum of S.

Recall that $M < f(b)$, and set $\varepsilon := \frac{f(b)-M}{2}$; then $\varepsilon > 0$. Using the continuity of f at b, we fix some $\delta > 0$ such that for all $x \in [a, b]$, if $|x-b| < \delta$, then $|f(x)-f(b)| < \varepsilon$.

Since b is the supremum of S, we see that there exists some $x \in S$ such that $b - \delta < x \leq b$.[17] We now have that $x \in [a, b]$ and $|x - b| < \delta$,[18] and consequently $|f(x) - f(b)| < \varepsilon$. Therefore:

---

[17] Indeed, if no such x existed, then $b - \delta$ would be an upper bound of S, contrary to the fact that b is the least upper bound of S.

[18] Indeed, since $x \in S \subseteq [a, b]$, we obviously have that $x \in [a, b]$. On the other hand, since $b - \delta < x \leq b$, we have that $|x - b| < \delta$.

$$\begin{array}{rcll} f(x) & > & f(b)-\varepsilon & \text{because } |f(x)-f(b)|<\varepsilon \\ & = & f(b)-\frac{f(b)-M}{2} & \text{because } \varepsilon=\frac{f(b)-M}{2} \\ & = & \frac{f(b)+M}{2} & \\ & > & M & \text{because } M<f(b). \end{array}$$

We have now shown that $f(x) > M$, contrary to the fact that $x \in S$. This proves Claim 2. ♦

Claim 3. $f(c) = M$.

Proof of Claim 3. Our goal is to show that, for all $\varepsilon > 0$, we have that $M - \varepsilon < f(c) < M + \varepsilon$. Indeed, if this holds for every single $\varepsilon > 0$, then it is clear that $f(c) = M$, which is what we need to show.

So, fix $\varepsilon > 0$. Using the continuity of f and the fact that $c \in [a, b]$ (by Claim 1), we fix some $\delta > 0$ such that for all $x \in [a, b]$, if $|x - c| < \delta$, then $|f(x) - f(c)| < \varepsilon$.

We first show that $f(c) < M + \varepsilon$. Since c is the supremum of S, we see that there exists some $a^* \in S$ such that $c - \delta < a^* \le c$ (indeed, if no such a^* existed, then $c - \delta$ would be an upper bound of S, contrary to the fact that c is the least upper bound of S). By the definition of S, we have that $f(a^*) \le M$. Now, since $a^* \in (c - \delta, c]$, we see that $|a^* - c| < \delta$. Also, $a^* \in [a, b]$ (because $a^* \in S \subseteq [a, b]$). So, by the choice of δ, we have that $|f(a^*) - f(c)| < \varepsilon$. It follows that $f(c) < f(a^*) + \varepsilon \le f(c) < M + \varepsilon$.

It remains to show that $M - \varepsilon < f(c)$. By Claim 2, we have that $c < b$. So, fix some $b^* \in (c, \min\{c + \delta, b\})$. Then $b^* \in [a, b]$ and $|b^* - c| < \delta$, and it follows that $|f(b^*) - f(c)| < \varepsilon$. Consequently, $f(b^*) - \varepsilon < f(c)$. On the other hand, since $c < b^*$, and c is the supremum (and an in particular, an upper bound) of S, we deduce that $b^* \notin S$. Since $b^* \in [a, b]$ and $b^* \notin S$, we see that $M \le f(b^*)$. So, $M - \varepsilon \le f(b^*) - \varepsilon < f(c)$. This proves Claim 3. ♦

Now, by Claims 1 and 2, we have that $c \in [a, b)$, and by Claim 3, we have that $f(c) = M$. It remains to show that $c \neq a$. But this follows from the fact that $f(a) < M = f(c)$, and in particular, $f(a) \neq f(c)$. □

Example 3.5.5. *Prove that the equation* $\sin x = \ln x$ *has at least one solution.*

Solution. Let $a = e^{-2}$ and $b = e^2$, and let $f : [a, b] \to \mathbb{R}$ be defined by $f(x) = \sin x - \ln x$. Clearly, f is continuous on $[a, b]$. Furthermore, we have that:

- $f(a) = \sin(e^{-2}) - \ln(e^{-2}) = \sin(e^{-2}) + 2 \ge 1$;
- $f(b) = \sin(e^2) - \ln(e^2) = \sin(e^2) - 2 \le -1$.

It follows that $f(b) < 0 < f(a)$, and so the Intermediate Value Theorem implies that there exists some $x \in (a, b)$ such that $f(x) = 0$, i.e. such that $\sin x = \ln x$. □

3.5.2 The Extreme Value Theorem

Bounded functions. A function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is

- *bounded above* if there exists some $U \in \mathbb{R}$, called an *upper bound* for f, such that for all $a \in A$, we have that $f(a) \leq U$;
- *bounded below* if there exists some $L \in \mathbb{R}$, called a *lower bound* for f, such that for all $a \in A$, we have that $L \leq f(a)$;
- *bounded* on A if it is bounded above and bounded below.

For $B \subseteq A$, we say that

- f is *bounded above* on B if $f \restriction B$ is bounded above;[19]
- f is *bounded below* on B if $f \restriction B$ is bounded below;
- f is *bounded* on B if $f \restriction B$ is bounded.

Notation: For $a \in \mathbb{R}$, we set $[a, a] = \{a\}$. So, we introduce a slight generalization of closed intervals, where the set consisting of a single real number is a closed interval.

Theorem 3.5.6. *Let $a, b \in \mathbb{R}$ be such that $a < b$, and let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then f is bounded on $[a, b]$.*

Remark: Note that Theorem 3.5.6 fails for open intervals. For instance, $f(x) = \tan x$ is continuous on the open interval $(-\frac{\pi}{2}, \frac{\pi}{2})$, but it is bounded neither above nor below on this interval.

Proof. Let

$$B \quad := \quad \big\{x \in [a, b] \mid f \text{ is bounded on } [a, x]\big\}.$$

Our goal is to show that $b \in B$; this will immediately imply that f is bounded on $[a, b]$, which is what we need to show.

Clearly, $a \in B$,[20] and in particular, $B \neq \emptyset$. Furthermore, B is bounded above by b. The completeness of the ordered field $\mathbb{R}$ now guarantees that B has a supremum, call it s. Clearly, $a \leq s \leq b$,[21] that is, $s \in [a, b]$.

Claim 1. *$s \in B$.*

---

[19] $f \restriction B$ is the restriction of f to B

[20] Indeed, f is bounded both above and below on $[a, a] = \{a\}$ by $f(a)$.

[21] Indeed, since $a \in B$ and s is an upper bound of B, we have that $a \leq s$. On the other hand, since s is the least upper bound of B, and b is some upper bound of B, we have that $s \leq b$.

Proof of Claim 1. If $s = a$, then this follows from the fact that $a \in B$. So, assume that $s > a$. We must show that f is bounded on $[a, s]$. (Since $s \in [a, b]$, this will immediately imply that $s \in B$, which is what we need.)

Fix $\varepsilon > 0$. Using the continuity of f, we fix some $\delta > 0$ such that for all $x \in [a, b]$, if $|x - s| < \delta$, then $|f(x) - f(s)| < \varepsilon$.

Now, using the fact that s is the least upper bound of B, we fix some $b' \in B$ such that $s - \delta < b' \leq s$.[22] Since $b' \in B$, we have that $b' \in [a, b]$ and that f is bounded on $[a, b']$; fix $L', U' \in \mathbb{R}$ such that for all $x \in [a, b']$, we have that $L' \leq f(x) \leq U'$.[23]

Set $L := \min\{L', f(s) - \varepsilon\}$ and $U := \max\{U', f(s) + \varepsilon\}$. We claim that for all $x \in [a, s]$, we have that $L \leq f(x) \leq U$. Note that this will immediately imply that f is bounded on $[a, s]$, which is what we need to show.

Fix any $x \in [a, s]$. Suppose first that $x > s - \delta$. Then $|s - x| < \delta$, and consequently, $|f(x) - f(s)| < \varepsilon$, which implies that $L \leq f(s) - \varepsilon < f(x) < f(s) + \varepsilon \leq U$.

Suppose now that $x \leq s - \delta$. Then $a \leq x \leq s - \delta < b'$, and in particular, $x \in [a, b']$. But now $L \leq L' \leq f(x) \leq U' \leq U$. This proves Claim 1. ◆

Claim 2. $s = b$.

Proof of Claim 2. By Claim 1, we have that $s \in B$, and in particular, $s \in [a, b]$. So, we may assume that $a \leq s < b$, for otherwise we are done.

Fix $\varepsilon > 0$. Using the continuity of f on $[a, b]$, we fix some $\delta > 0$ such that for all $x \in [a, b]$, if $|x - s| < \delta$, then $|f(x) - f(s)| < \varepsilon$.

Set $s' := \min\left\{b, s + \frac{\delta}{2}\right\}$. Then for all $x \in [s, s']$, we have that $x \in [a, b]$ and $|x - s| < \delta$, and consequently, $|f(x) - f(s)| < \varepsilon$, that is, $f(s) - \varepsilon < f(x) < f(s) + \varepsilon$. Thus, f is bounded on $[s, s']$.

By Claim 1, f is bounded on $[a, s]$, and by what we just showed, f is bounded on $[s, s']$. Therefore f is bounded on $[a, s']$,[24] and consequently, $s' \in B$. But this is impossible since s is an upper bound of B, and $s < s'$. This proves Claim 2. ◆

By Claim 2, we have that $s = b$, and so Claim 1 implies that $b \in B$. It then follows from the definition of B that f is bounded on $[a, b]$. ☐

Maxima and minima. Given a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ and $a \in A$, we say that

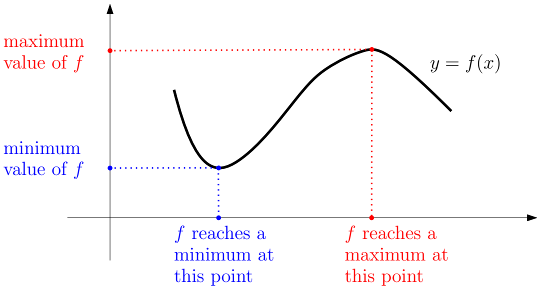
- f reaches a *global maximum* (or simply *maximum*) at a, and that $f(a)$ is the *maximum value* of f, if for all $x \in A$, we have that $f(x) \leq f(a)$;
- f reaches a *global minimum* (or simply *minimum*) at a, and that $f(a)$ is the *minimum value* of f, if for all $x \in A$, we have that $f(x) \geq f(a)$.

---

[22] If no such b' existed, then $s - \delta$ would be an upper bound of B, contrary to the fact that s is the least upper bound of B.

[23] In other words, L' and U' are, respectively, a lower and upper bound of $f \restriction [a, b']$.

[24] Indeed, suppose that L_1 and U_1 are, respectively, a lower and upper bound of $f \restriction [a, s]$, and that L_2 and U_2 are, respectively, a lower and upper bound of $f \restriction [s, s']$. Then $L := \min\{L_1, L_2\}$ and $U := \max\{U_1, U_2\}$ are, respectively, a lower and upper bound of $f \restriction [a, s']$.



Lemma 3.5.7. *Let $a, b \in \mathbb{R}$ be such that $a < b$, and let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then f reaches a maximum on the interval $[a, b]$.*

Proof. By Theorem 3.5.6, f is bounded on $[a, b]$, i.e. the set

$$Y \quad := \quad \big\{f(x) \mid x \in [a, b]\big\}$$

is bounded. Clearly, $Y \neq \emptyset$, and so by the completeness of $\mathbb{R}$, Y has a supremum, call it s. We need to show that there exists some $x \in [a, b]$ such that $f(x) = s$.[25] Suppose otherwise. Then for all $x \in [a, b]$, we have that $f(x) < s$, and in particular, $s - f(x) \neq 0$. Now, define $g : [a, b] \to \mathbb{R}$ by setting

$$g(x) \quad := \quad \tfrac{1}{s-f(x)} \qquad \forall x \in [a, b].$$

Since f is continuous on $[a, b]$, so is g (by Theorem 3.5.2). So, by Theorem 3.5.6, g is bounded. Let U be an upper bound for g. We will derive a contradiction by exhibiting an $x \in [a, b]$ such that $g(x) > U$.

Note that $g(x) > 0$ for all $x \in [a, b]$.[26] Consequently, $U > 0$. Since s is the supremum of Y, we know that $s - \frac{1}{U}$ is not an upper bound of Y, and consequently, there exist some $x \in [a, b]$ such that $f(x) > s - \frac{1}{U}$. But then $\frac{1}{s-f(x)} > U$, i.e. $g(x) > U$, contrary to the fact that U is an upper bound for g. ☐

Lemma 3.5.8. *Let $a, b \in \mathbb{R}$ be such that $a < b$, and let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then f reaches a minimum on the interval $[a, b]$.*

Proof. Since f is continuous on $[a, b]$, so is $-f$. So, by Lemma 3.5.7, $-f$ reaches a maximum at some $x \in [a, b]$. But then f reaches a minimum at x. ☐

The Extreme Value Theorem. *Let $a, b \in \mathbb{R}$ be such that $a < b$, and let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then f reaches both a maximum and a minimum on $[a, b]$.*

Proof. This follows immediately from Lemmas 3.5.7 and 3.5.8. ☐

---

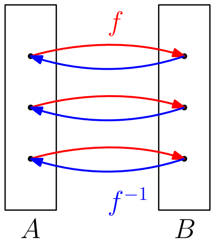
[25] Indeed, if such an x exists, then f reaches a maximum at x, and we are done.

[26] This is because $f(x) < s$ for all $x \in [a, b]$. So, $g(x) = \frac{1}{s-f(x)} > 0$ for all $x \in [a, b]$.

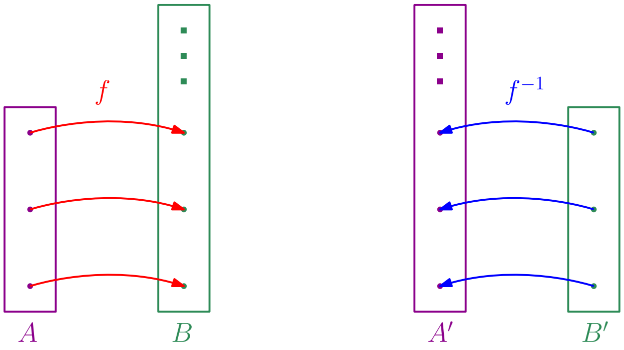
Remark: Just like Theorem 3.5.6, the Extreme Value Theorem fails for open intervals. For instance, $\tan x$ is continuous on the open interval $(-\frac{\pi}{2}, \frac{\pi}{2})$, but it is bounded neither above nor below, and consequently, it reaches neither a maximum nor a minimum on $(-\frac{\pi}{2}, \frac{\pi}{2})$.

3.5.3 The continuity of inverse functions

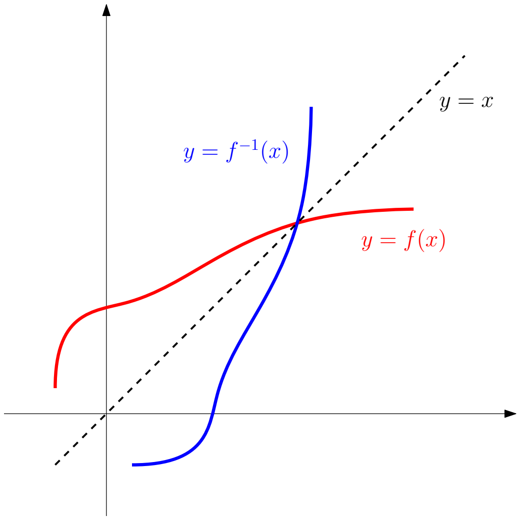
We know that if $f : A \to B$ is a bijection, then it has an *inverse function* $f^{-1} : B \to A$ that satisfies the following: for all $b \in B$, we have that $f^{-1}(b) = a$, where a is the unique member of A such that $f(a) = b$.



Let us try to generalize this. Suppose that $f : A \to B$ is a one-to-one function, and let $B' := \text{Im}(f) = B' := \{f(x) \mid x \in A\}$. If we restrict the codomain of f to B', then $f : A \to B'$ becomes a bijection, and it therefore has an inverse $f^{-1} : B' \to A$. Sometimes, it is convenient to then extend the codomain of f^{-1} to a set $A' \supseteq A$, so that we get $f^{-1} : B' \to A'$.



Note that if $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is a one-to-one function, and $B := \text{Im}(f)$, then the graph of $f^{-1} : B \to \mathbb{R}$ is obtained by reflecting the graph of f about the $y = x$ line, as in the picture below.



If a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is one-to-one and continuous, its inverse need not be continuous (see Example 3.5.14). However, if the domain A happens to be an interval (open, closed, or half-open), then f^{-1} will, in fact, be continuous; this is proven below (see Theorem 3.5.12 and Corollary 3.5.13).

Before giving a formal proof, let us give some intuition. So, let us suppose that $f : I \subseteq \mathbb{R}$ is a continuous one-to-one function. Now, our intuition is that the the graph of the continuous function f on the interval I can be drawn without lifting a pen. So, the graph obtained by reflecting about the $y = x$ can also be drawn without lifting a pen, and this is precisely the graph of f^{-1}. So, intuitively, f^{-1} should be continuous (and the formal proof is given below).

Monotone functions. Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function.

- We say that f is *non-decreasing* if for all $a, a' \in A$ such that $a \leq a'$, we have that $f(a) \leq f(a')$.
- We say that f is *strictly increasing* (or simply *increasing*) if for all $a, a' \in A$ such that $a < a'$, we have that $f(a) < f(a')$.
- We say that f is *non-increasing* if for all $a, a' \in A$ such that $a \leq a'$, we have that $f(a) \geq f(a')$.
- We say that f is *strictly decreasing* (or simply *decreasing*) if for all $a, a' \in A$ such that $a < a'$, we have that $f(a) > f(a')$.
- We say that f is *monotone* if it is non-decreasing or non-increasing.

For $A' \subseteq A$, we say that f is *non-decreasing* (resp. *increasing*, *non-increasing*, *decreasing*, *monotone*) on A' if the restriction $f \restriction A'$ is non-decreasing (resp. increasing, non-increasing, decreasing, monotone).

Remark: Suppose that $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is a continuous function, and that $a, b \in A$ are such that $a < b$, $[a, b] \subseteq A$, and $f(a) \neq f(b)$. When we say that "we apply the Intermediate Value Theorem to the function f and the interval $[a, b]$," we technically mean that we are applying the Intermediate Value Theorem to $f \upharpoonright [a, b]$, i.e. to the restriction of f to the interval $[a, b]$. We may do this because the continuity of the function f immediately implies the continuity of $f \upharpoonright [a, b]$.

Proposition 3.5.9. *Let $a, b \in \mathbb{R}$ be such that $a < b$, and let $f : [a, b] \to \mathbb{R}$ be a continuous one-to-one function. Then the following hold:*

(a) if $f(a) < f(b)$, then f is strictly increasing, and $Im(f) = [f(a), f(b)]$;

(b) if $f(a) > f(b)$, then f is strictly decreasing, and $Im(f) = [f(b), f(a)]$.

Proof. We prove (a); the proof of (b) is similar and is left as an exercise. So, let us assume that $f(a) < f(b)$.

Claim. For all $x \in (a, b)$, we have that $f(a) < f(x) < f(b)$.

Proof of the Claim. Fix $x \in (a, b)$. We will prove that $f(a) < f(x)$; the proof of the fact that $f(x) < f(b)$ is similar.

We suppose toward a contradiction that $f(x) \leq f(a)$. Since f is one-to-one and $x \neq a$, we have that $f(x) \neq f(a)$; therefore, $f(x) < f(a)$. Set $M := \frac{f(x)+f(a)}{2}$; then $f(x) < M < f(a)$. By applying the Intermediate Value Theorem to the continuous function f and the closed interval $[a, x]$, we obtain a point $c \in (a, x)$ such that $f(c) = M$. Next, since $f(a) < f(b)$, we see that $f(x) < M < f(b)$, and so by applying the Intermediate Value Theorem to the continuous function f and the closed interval $[x, b]$, we obtain a point $d \in (x, b)$ such that $f(d) = M$. But now $c \neq d$ and $f(c) = f(d)$, contrary to the fact that f is one-to-one. ♦

We now show that f is strictly increasing. Fix any $c, d \in [a, b]$ such that $c < d$; we must show that $f(c) < f(d)$. If $c = a$, then this follows immediately from the Claim that we just proved. We may therefore assume that $c \neq a$, and therefore, $a < c < d \leq b$. Now, let us assume toward a contradiction that $f(c) \geq f(d)$. Since f is one-to-one and $c \neq d$, we have that $f(c) \neq f(d)$; therefore, $f(c) > f(d)$. Set $M := \frac{f(c)+f(d)}{2}$. Then $f(d) < M < f(c)$. By applying the Intermediate Value Theorem to the continuous function f and the interval $[c, d]$, we obtain a point $x_1 \in (c, d)$ such that $f(x_1) = M$. Now, in view of the Claim above, we have that $f(a) < f(d)$, and it follows that $f(a) < M < f(c)$. So, by applying the Intermediate Value Theorem to the continuous function f and the interval $[a, c]$, we obtain a point $x_2 \in (a, c)$ such that $f(x_2) = M$. But now $x_1 \neq x_2$ and $f(x_1) = f(x_2)$, contrary to the fact that f is one-to-one.

It remains to show that $\mathrm{Im}(f) = [f(a), f(b)]$. The fact that $\mathrm{Im}(f) \subseteq [f(a), f(b)]$ follows immediately from the Claim. It remains to prove the reverse inclusion. So, fix some $M \in [f(a), f(b)]$; we must show that $M \in \text{Im}(f)$, i.e. that there exists some $x \in [a, b]$ such that $f(x) = M$. If $M = f(a)$, then we may choose $x := a$; similarly, if $M = f(b)$, then we may choose $x := b$. We may therefore assume that $M \in \big(f(a), f(b)\big)$, i.e. $f(a) < M < f(b)$. But then by the Intermediate Value Theorem applied to the continuous function f and the interval $[a, b]$, there exists some $x \in (a, b)$ such that $f(x) = M$, and we are done. $\square$

Proposition 3.5.10. *Let I be an interval (open, closed, or half-open) in $\mathbb{R}$, and let $f : I \to \mathbb{R}$ be a continuous one-to-one function. Then f is either strictly increasing or strictly decreasing.*

Proof. We may assume that f is not strictly decreasing, for otherwise we are done. So, there exist some $a, b \in I$ such that $a < b$ and $f(a) \leq f(b)$. Since $a \neq b$ and f is one-to-one, we have that $f(a) \neq f(b)$; consequently, $f(a) < f(b)$.

Now, our goal is to show that f is strictly increasing. Fix any $c, d \in I$ such that $c < d$; we must show that $f(c) < f(d)$. Set $a_0 := \min\{a, c\}$ and $b_0 := \max\{d, b\}$. Then $a, b, c, d \in [a_0, b_0] \subseteq I$. By Proposition 3.5.9 applied to the continuous function f and the interval $[a_0, b_0]$,[27] we see that f is either strictly increasing or strictly decreasing on the interval $[a_0, b_0]$. Since $a, b \in [a_0, b_0]$, and since $a < b$ and $f(a) < f(b)$, we see that f is not strictly decreasing on $[a_0, b_0]$. Therefore, f is strictly increasing on $[a_0, b_0]$. Since $c, d \in [a_0, b_0]$ and $c < d$, it follows that $f(c) < f(d)$. $\square$

Proposition 3.5.11. *Let $f : I \subseteq \mathbb{R} \to \mathbb{R}$ be a one-to-one continuous function, where I is an interval. The Im(f) is an interval of the same kind (open, closed, or half-open) as the interval I.*

Proof (outline). If I is a closed interval, then this follows immediately from Proposition 3.5.9. Here we give an outline of the proof for the case when I is an open interval. The case when I is a half-open interval is left as an exercise.

So, let us suppose that I is an open interval. By Proposition 3.5.10, f is either strictly increasing or strictly decreasing. We consider the case when f is strictly increasing; the other case is similar. It is then not hard to show that there exist sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ of real numbers such that $\dots < a_3 < a_2 < a_1 < b_1 < b_2 < b_3 < \dots$ and $I = \bigcup_{n=1}^{\infty} [a_n, b_n]$.[28] Then

$$\text{Im}(f) \;=\; \textstyle\bigcup_{n=1}^{\infty} \text{Im}\big(f \restriction [a_n, b_n]\big) \;\overset{(*)}{=}\; \textstyle\bigcup_{n=1}^{\infty} [f(a_n), f(b_n)],$$

---

[27] Technically, we are applying Proposition 3.5.9 to the function $f \restriction [a_0, b_0]$, i.e. the restriction of f to $[a_0, b_0]$.

[28] For instance, if $I = (a, b)$ for some $a, b \in \mathbb{R}$, then for $c := \frac{a+b}{2}$, we may define $a_n := \frac{(n-1)a+c}{n}$ and $a_n := \frac{(n-1)b+c}{n}$ for all $n \in \mathbb{N}$. If $I = (-\infty, +\infty)$, then we may simply set $a_n := -n$ and $b_n := n$ for all $n \in \mathbb{N}$. If $I = (a, +\infty)$ for some $a \in \mathbb{R}$, then we may set $a_n := a + \frac{1}{n}$ and $b_n := a + n + 2$ for all $n \in \mathbb{R}$. Similarly, if $I = (-\infty, b)$ for some $b \in \mathbb{R}$, then we may set $a_n := b - n - 2$ and $b_n := b - \frac{1}{n}$ for all $n \in \mathbb{R}$.

where (*) follows from Proposition 3.5.9. Since f is strictly increasing, we see that $\dots < f(a_3) < f(a_2) < f(a_1) < f(b_1) < f(b_2) < f(b_3) < \dots$, and then it is not difficult to show that $\bigcup_{n=1}^{\infty}[f(a_n), f(b_n)]$ is an open interval. So, $\text{Im}(f)$ is an open interval, as we had claimed. ☐

Theorem 3.5.12. *Let I be an open interval, and let $f : I \to \mathbb{R}$ be a continuous one-to-one function. Then f^{-1} is also continuous.*

Proof. By Proposition 3.5.10, we know that f is either strictly increasing or strictly decreasing. We will prove the theorem for the case when f is strictly increasing; the other case is similar.

Set $J := \text{Im}(f)$, so that $f^{-1} : J \to \mathbb{R}$.[29] We fix an arbitrary $y_0 \in J$, and we prove that f^{-1} is continuous at y_0. Fix $\varepsilon > 0$. We must show that there exists some $\delta > 0$ such that for all $y \in J$, if $|y - y_0| < \delta$, then $|f^{-1}(y) - f^{-1}(y_0)| < \varepsilon$.

Set $x_0 := f^{-1}(y_0)$, so that $f(x_0) = y_0$. Now, we know that $x_0 \in I$ and that I is an open interval. So, both $I \cap (x_0 - \varepsilon, x_0)$ and $I \cap (x_0, x_0 + \varepsilon)$ are non-empty. Fix $a \in I \cap (x_0 - \varepsilon, x_0)$ and $b \in (x_0 - \varepsilon, x_0)$. Then $x_0 \in (a, b)$ and $[a, b] \subseteq I \cap (x_0 - \varepsilon, x_0 + \varepsilon)$. Now, since f is strictly increasing, Proposition 3.5.9 guarantees that $\text{Im}\big(f \restriction [a, b]\big) = [f(a), f(b)]$. Moreover, since $x_0 \in (a, b)$ and f is one-to-one, it follows that $y_0 = f(x_0) \in (f(a), (b))$.[30] Now, choose any $\delta > 0$ such that $(y_0 - \delta, y_0 + \delta) \subseteq (f(a), f(b))$.

Fix $y \in J$ such that $|y - y_0| < \delta$. So, $y \in (y_0 - \delta, y_0 + \delta) \subseteq (f(a), f(b))$. Since $\text{Im}(f \restriction [a, b]) = [f(a), f(b)]$, it follows that there exists some $x \in [a, b]$ such that $f(x) = y$. So, $x = f^{-1}(y)$. But now $x \in [a, b] \subseteq (x_0 - \varepsilon, x_0 + \varepsilon)$, and consequently, $|f^{-1}(y) - f^{-1}(y_0)| = |x - x_0| < \varepsilon$. ☐

Corollary 3.5.13. *Let I be an interval (open, closed, or half-open) in $\mathbb{R}$, and let $f : I \to \mathbb{R}$ be a continuous one-to-one function. Then f^{-1} is also continuous.*

Proof. By Proposition 3.5.10, f is either strictly increasing or strictly decreasing. We consider the case when f is strictly increasing; the other case is similar. Now, if I is an open interval, then the result follows immediately from Theorem 3.5.12. We will prove the corollary for the case when I is a closed interval; the case when I is a half-open interval is similar.

So, let us assume that $I = [a, b]$ for some $a, b \in \mathbb{R}$ such that $a < b$. Since f is strictly increasing, Proposition 3.5.9 guarantees that $\text{Im}(f) = [f(a), f(b)]$. Now, define the function $g : \mathbb{R} \to \mathbb{R}$ by setting

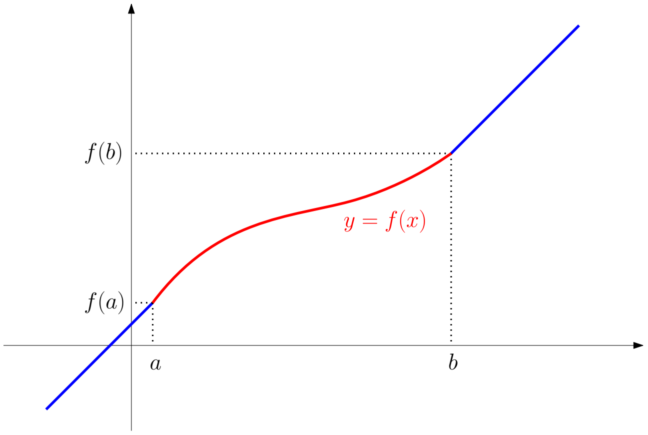
$$g(x) \;:=\; \begin{cases} x + f(a) - a & \text{if} \quad x < a \\ f(x) & \text{if} \quad a \le x \le b \\ x + f(b) - b & \text{if} \quad x > b \end{cases}$$

---

[29] By Proposition 3.5.11, J is an open interval. However, we will not actually use this fact.

[30] Indeed, since $x \in [a, b]$, it is clear that $f(x_0) \in \text{Im}(f \restriction [a, b])$, and so $f(x_0) \in [f(a), f(b)]$. But since f is one-to-one, and since $a \neq x_0 \neq b$, it follows that $f(a) \neq f(x_0) \neq f(b)$, and so $f(x_0) \in (f(a), f(b))$.

for all $x \in \mathbb{R}$. A schematic representation of the graph of the function g is given below (the graph of the original function f is in red, whereas the parts of the graph of g for $x < a$ and $x > b$ are in blue). Now, g is a continuous one-to-one function defined whose domain is the open interval $(-\infty, +\infty)$. Therefore, by Theorem 3.5.12, g^{-1} is continuous. Since f^{-1} is simply the restriction of g^{-1} to the interval $[f(a), f(b)]$, it follows that f^{-1} is also continuous.



$\square$

Remark: If $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is a one-to-one continuous function, but the domain A of f is **not** an interval, then f^{-1} need not be continuous (as the example below shows).

Example 3.5.14. *Consider the function* $f : [0,1) \cup (1,2)$ *given by*

$$f(x) \quad := \quad \begin{cases} x & \textit{if} \quad x \in [0,1) \\ x-2 & \textit{if} \quad x \in (1,2) \end{cases}$$

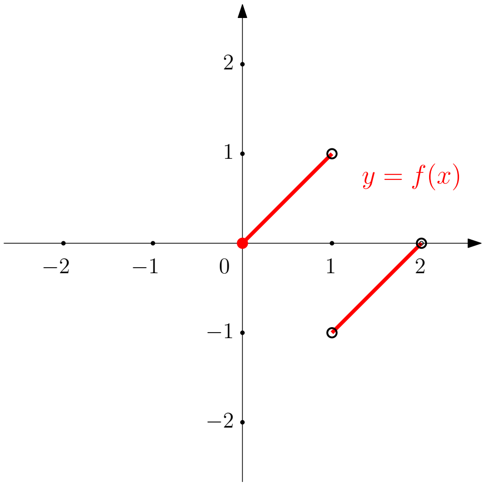
for all $x \in [0,1) \cup (1,2)$. *Clearly, the function* f *is continuous and one-to-one.*[31] *But note that* $Im(f) = (-1,1)$, *and that* $f^{-1} : (-1,1) \to \mathbb{R}$ *is given by*

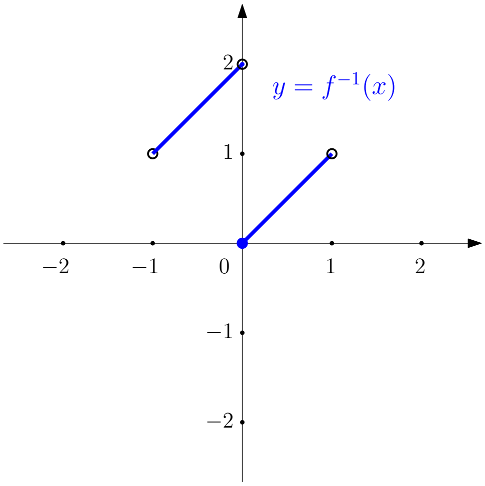
$$f^{-1}(x) \quad = \quad \begin{cases} x & \textit{if} \quad x \in [0,1) \\ x+2 & \textit{if} \quad x \in (-1,0) \end{cases}$$

for all $x \in (-1,1)$. *Then* f^{-1} *is not continuous (because it is not continuous at* 0*). The graphs of* f *and* f^{-1} *are represented below.*

---

[31] The function f is continuous because it is continuous at all points in its domain. The non-existence of $\lim\limits_{x\to 1} f(x)$ is no obstacle to the continuity of f because 1 is not in the domain of the function f.





3.6 Variations on the theme: infinite limits, one-sided limits, and limits at infinity. Asymptotes

3.6.1 Infinite limits

Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, and let $a \in \mathbb{R}$ be an accumulation point of A.

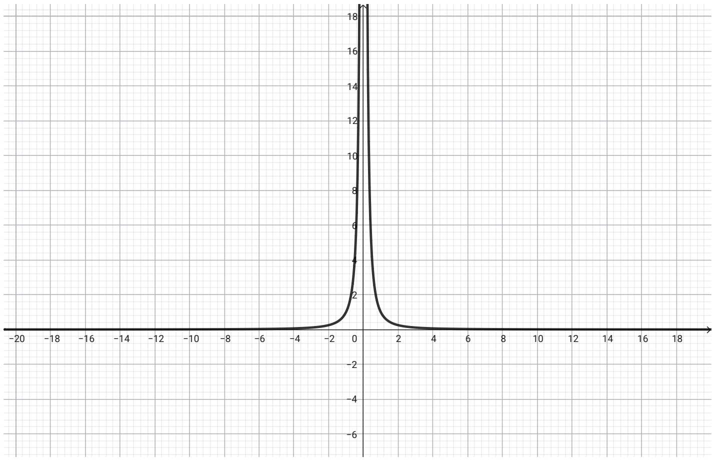
- We write
  $$\lim_{x \to a} f(x) = +\infty$$
  if the following holds:

  > for all $M \in \mathbb{R}$, there exists some $\delta > 0$, such that for all $x \in A$, if $0 < |x - a| < \delta$, then $f(x) > M$.

- We write
  $$\lim_{x \to a} f(x) = -\infty$$
  if the following holds:

  > for all $M \in \mathbb{R}$, there exists some $\delta > 0$, such that for all $x \in A$, if $0 < |x - a| < \delta$, then $f(x) < M$.

For example, we have that $\lim\limits_{x \to 0} \frac{1}{x^2} = +\infty$ (the graph of this function is shown below).



3.6.2 One-sided limits and vertical asymptotes

Sometimes, it is useful to consider "one-sided limits." Intuitively, we write $\lim\limits_{x \to a^+} f(x) = L$ if, whenever x is very slightly greater than a, we have that $f(x) \approx L$. Similarly, we have that $\lim\limits_{x \to a^-} f(x) = L$ if, whenever x is very slightly smaller than a, we have that $f(x) \approx L$. We formalize this below.

Remark: Note that for any set $A \subseteq \mathbb{R}$, a point $a \in \mathbb{R}$ is an accumulation point of A if and only if it is an accumulation point of at least one of $A \cap (-\infty, a)$ and $A \cap (a, +\infty)$.[32]

Left-hand limits. Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, and let $a \in \mathbb{R}$ be an accumulation point of $A \cap (-\infty, a)$.[33]

- For $L \in \mathbb{R}$, we write
$$\lim_{x \to a^-} f(x) = L$$
if the following holds:

    for every $\varepsilon > 0$, there exists some $\delta > 0$, such that for all $x \in A$, if $a - \delta < x < a$, then $|f(x) - L| < \varepsilon$.

---

[32] This is "obvious," but here is a proof. Fix a set $A \subseteq \mathbb{R}$ and a point $a \in \mathbb{R}$. Since $A \cap (-\infty, a)$ and $A \cap (a, +\infty)$ are both subsets of A, it is clear that if a is an accumulation point of $A \cap (-\infty, a)$ or $A \cap (a, +\infty)$, then it is also an accumulation point of A. Suppose now that a is an accumulation point neither of $A \cap (-\infty, a)$ nor of $A \cap (a, +\infty)$. Then there exist some $\varepsilon^-, \varepsilon^+$ such that $\big(A \cap (-\infty, a)\big) \cap \big((a - \varepsilon^-, a) \cup (a, a + \varepsilon^-)\big)$ and $\big(A \cap (a, +\infty)\big) \cap \big((a - \varepsilon^+, a) \cup (a, a + \varepsilon^+)\big)$ are both empty. Set $\varepsilon := \min\{\varepsilon^-, \varepsilon^+\}$. Then $A \cap \big((a - \varepsilon, a) \cup (a, a + \varepsilon)\big)$ is empty, and it follows that a is not an accumulation point of A.

[33] So, we are assuming that there are points of A to the left of a, and arbitrarily close to a. The typical situation is that the domain of f includes some open interval of the form $(a - \delta, a)$ for some small $\delta > 0$.

- We write
  $$\lim_{x\to a^-} f(x) = +\infty$$
  if the following holds:

  > for all $M \in \mathbb{R}$, there exists some $\delta > 0$ such that for all $x \in A$, if $a - \delta < x < a$, then $f(x) > M$.

- We write
  $$\lim_{x\to a^-} f(x) = -\infty$$
  if the following holds:

  > for all $M \in \mathbb{R}$, there exists some $\delta > 0$ such that for all $x \in A$, if $a - \delta < x < a$, then $f(x) < M$.

Right-hand limits. Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, and let $a \in \mathbb{R}$ be an accumulation point of $A \cap (a, +\infty)$.[34]

- For $L \in \mathbb{R}$, we write
  $$\lim_{x\to a^+} f(x) = L$$
  if the following holds:

  > for every $\varepsilon > 0$, there exists some $\delta > 0$, such that for all $x \in A$, if $a < x < a + \delta$, then $|f(x) - L| < \varepsilon$.

- We write
  $$\lim_{x\to a^+} f(x) = +\infty$$
  if the following holds:

  > for all $M \in \mathbb{R}$, there exists some $\delta > 0$ such that for all $x \in A$, if $a < x < a + \delta$, then $f(x) > M$.

- We write
  $$\lim_{x\to a^+} f(x) = -\infty$$
  if the following holds:

  > for all $M \in \mathbb{R}$, there exists some $\delta > 0$ such that for all $x \in A$, if $a < x < a + \delta$, then $f(x) < M$.

---

[34] So, we are assuming that there are points of A to the right of a, and arbitrarily close to a. The typical situation is that the domain of f includes some open interval of the form $(a, a + \delta)$ for some small $\delta > 0$.

Remark: Rules for one-sided limits (i.e. "$\lim\limits_{x \to a^-} f(x)$" and "$\lim\limits_{x \to a^+} f(x)$") are very similar to those for ordinary limits, i.e. Theorem 3.4.4 can be adapted to one-sided limits in a completely straightforward way. The details are left as an exercise.

Lemma 3.6.1. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, and let $a \in \mathbb{R}$ be an accumulation point of A. Then a is an accumulation point of at least one of $A \cap (-\infty, a)$ and $A \cap (a, +\infty)$. Moreover:*

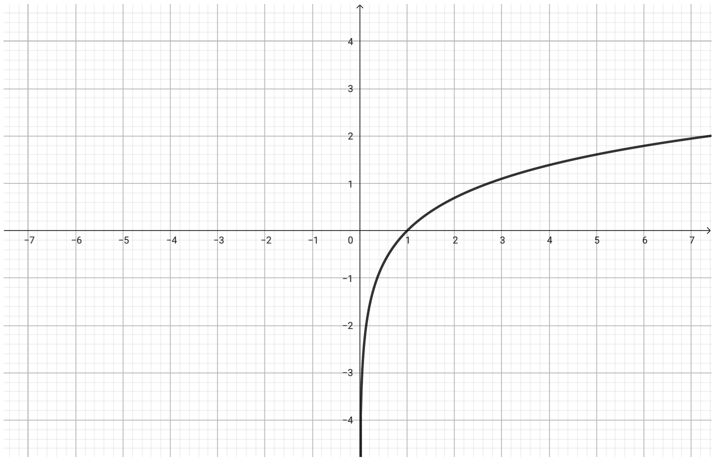
(a) if $\lim\limits_{x \to a} f(x)$ exists (as a real number, or $+\infty$, or $-\infty$) and a is an accumulation point of $A \cap (-\infty, a)$, then $\lim\limits_{x \to a^-} f(x) = \lim\limits_{x \to a} f(x)$;

(b) if $\lim\limits_{x \to a} f(x)$ exists (as a real number, or $+\infty$, or $-\infty$) and a is an accumulation point of $A \cap (a, +\infty)$, then $\lim\limits_{x \to a^+} f(x) = \lim\limits_{x \to a} f(x)$.

Proof. This readily follows from the appropriate definitions. The details are left as an exercise. □

Remark: Note that Lemma 3.6.1 in particular implies that if $\lim\limits_{x \to a} f(x)$ exists, then at least one of $\lim\limits_{x \to a^-} f(x)$ and $\lim\limits_{x \to a^+} f(x)$ exists.

Example 3.6.2. *The domain of the function $\ln x$ (whose graph is represented below) is $A := (0, +\infty)$. Clearly, 0 is an accumulation point of A. Moreover, 0 is an accumulation point of $A \cap (0, +\infty) = A$, but not of $A \cap (-\infty, 0) = \emptyset$. We have that $\lim\limits_{x \to 0} \ln x = \lim\limits_{x \to 0^+} \ln x = -\infty$, whereas $\lim\limits_{x \to 0^-} \ln x$ does not exist. Finally, $x = 0$ is a vertical asymptote of $\ln x$.*



Remark: In Example 3.6.2, a crucial point is that 0 is a not an accumulation point of $A \cap (-\infty, 0)$, and that is the reason why $\lim\limits_{x \to 0^-} \ln x$ does not exist, and yet this is no obstacle to the existence of $\lim\limits_{x \to 0} \ln x$. However, if $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is a function and $a \in \mathbb{R}$, then:

- if a is an accumulation point of $A \cap (-\infty, a)$ and yet $\lim\limits_{x \to a^-} f(x)$ does not exist, then $\lim\limits_{x \to a} f(x)$ does not exist either;
- if a is an accumulation point of $A \cap (a, +\infty)$ and yet $\lim\limits_{x \to a^+} f(x)$ does not exist, then $\lim\limits_{x \to a} f(x)$ does not exist either.

Lemma 3.6.3. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$, let $a \in \mathbb{R}$ be an accumulation point both of $A \cap (-\infty, a)$ and of $A \cap (a, +\infty)$.*[35] *Then the following are equivalent:*

(i) $\lim\limits_{x \to a} f(x)$ *exists (as a real number, or $+\infty$, or $-\infty$);*

(ii) $\lim\limits_{x \to a^-} f(x)$ *and* $\lim\limits_{x \to a^+} f(x)$ *both exist (as real numbers, or $+\infty$, or $-\infty$) and are equal.*

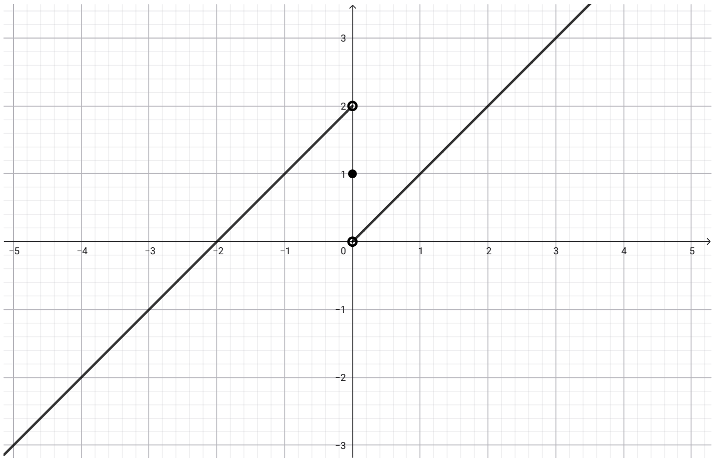
Moreover, in this case, we have that $\lim\limits_{x \to a} f(x) = \lim\limits_{x \to a^-} f(x) = \lim\limits_{x \to a^+} f(x)$.

Proof. This readily follows from the appropriate definitions. The details are left as an exercise. □

Example 3.6.4. *Consider the function $f : \mathbb{R} \to \mathbb{R}$ be given by*

$$f(x) = \begin{cases} x + 2 & \text{if} \quad x < 0 \\ 1 & \text{if} \quad x = 0 \\ x & \text{if} \quad x > 0 \end{cases}$$

for all $x \in \mathbb{R}$. (The graph of this function is shown below.)



Then

---

[35] The typical situation is that for some $\delta > 0$, we have that $(a - \delta, a) \cup (a, a + \delta) \subseteq A$, i.e. f is defined on some open interval containing a, except possibly at a itself.

- $\lim\limits_{x\to 0^-} f(x) = 2$;
- $\lim\limits_{x\to 0^+} f(x) = 0$;
- $\lim\limits_{x\to 0} f(x)$ *does not exist (because* $\lim\limits_{x\to 0^-} f(x) \neq \lim\limits_{x\to 0^+} f(x)$*).*

Example 3.6.5. *Consider the function* $f : \mathbb{R} \to \mathbb{R}$ *be given by*

$$f(x) = \begin{cases} x^2 & \text{if} \quad x \neq \pi \\ 100 & \text{if} \quad x = \pi \end{cases}$$

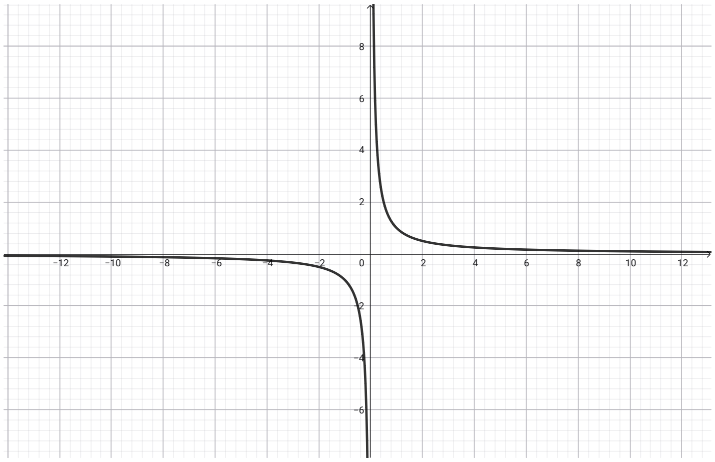
for all $x \in \mathbb{R}$. *Then*

- $\lim\limits_{x\to \pi^-} f(x) = \pi^2$;
- $\lim\limits_{x\to \pi^+} f(x) = \pi^2$;
- $\lim\limits_{x\to \pi} f(x) = \pi^2$.

Vertical asymptotes. Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function and let $a \in \mathbb{R}$. We say that $x = a$ is a *vertical asymptote* of f if at least one of $\lim\limits_{x\to a^-} f(x)$ and $\lim\limits_{x\to a^+} f(x)$ exists and is equal to $+\infty$ or $-\infty$.

Example 3.6.6. *Consider the function* $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ *given by* $f(x) = \frac{1}{x}$ *for all* $x \in \mathbb{R} \setminus \{0\}$. *(The graph of this function is shown below.) Then*

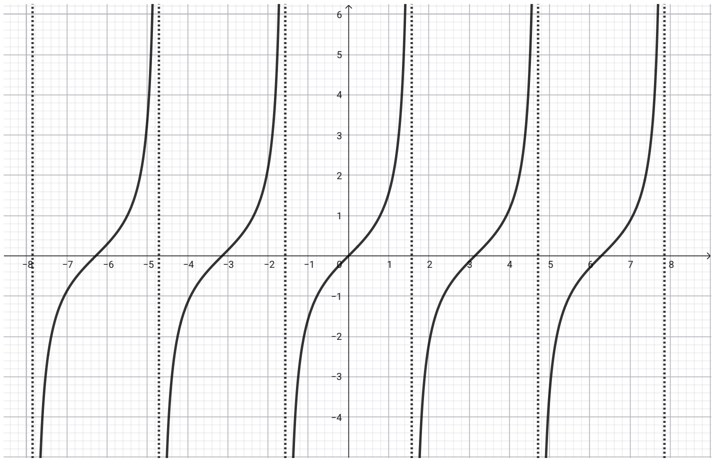
- $\lim\limits_{x\to 0^-} f(x) = -\infty$;
- $\lim\limits_{x\to 0^+} f(x) = +\infty$;
- $\lim\limits_{x\to 0} f(x)$ *does not exist (because* $\lim\limits_{x\to 0^-} f(x) \neq \lim\limits_{x\to 0^+} f(x)$*);*
- $x = 0$ *is a vertical asymptote of* $f(x)$.



Example 3.6.7. *For any integer k, the following hold:*

- $\lim\limits_{x \to (\frac{2k+1}{2}\pi)^-} \tan x = +\infty$*;*
- $\lim\limits_{x \to (\frac{2k+1}{2}\pi)^+} \tan x = -\infty$*;*
- $\lim\limits_{x \to \frac{2k+1}{2}\pi} \tan x$ *does not exist;*
- $x = \frac{2k+1}{2}\pi$ *is a vertical asymptote of* $\tan x$*.*

Remark: The function $\tan x$ is represented in the graph below. The dotted vertical lines are the vertical asymptotes, and they are all of the form $x = \frac{2k+1}{2}\pi$ for $k \in \mathbb{Z}$.

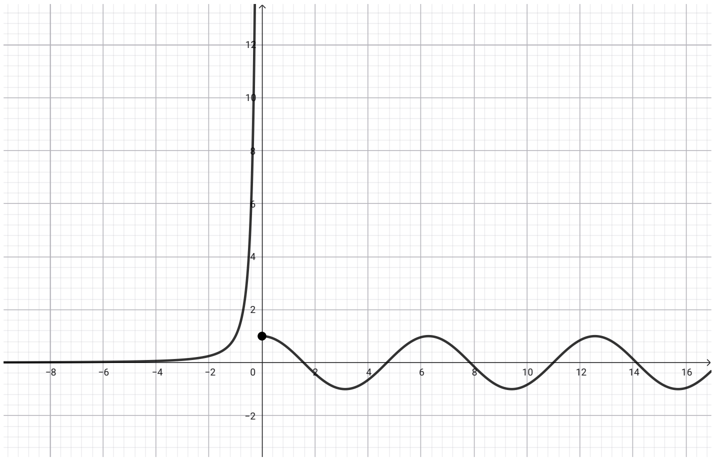


Example 3.6.8. *Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by*

$$f(x) \quad := \quad \begin{cases} x^{-2} & \text{if} \quad x < 0 \\ \cos x & \text{if} \quad x \geq 0 \end{cases}$$

for all $x \in \mathbb{R}$. (The graph of this function is shown below.) Then

- $\lim\limits_{x \to 0^-} f(x) = +\infty$*;*
- $\lim\limits_{x \to 0^+} f(x) = 1$*;*
- $\lim\limits_{x \to 0} f(x)$ *does not exist (because* $\lim\limits_{x \to 0^-} f(x) \neq \lim\limits_{x \to 0^+} f(x)$*);*
- $x = 0$ *is a vertical asymptote of* $f(x)$*.*



Example 3.6.9. *Consider the function* $f : \mathbb{R} \to \mathbb{R}$ *given by*

$$f(x) \quad := \quad \begin{cases} x & \textit{if} \quad x < 0 \\ 0 & \textit{if} \quad x \geq 0 \textit{ and } x \in \mathbb{Q} \\ 1 & \textit{if} \quad x \geq 0 \textit{ and } x \notin \mathbb{Q} \end{cases}$$

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

- $\lim\limits_{x \to 0^-} f(x) = 0$;
- $\lim\limits_{x \to 0^+} f(x)$ *does not exist;*
- $\lim\limits_{x \to 0} f(x)$ *does not exist (because* $\lim\limits_{x \to 0^+} f(x)$ *does not exist, even though* 0 *is an accumulation point of* $\mathbb{R} \cap (0, +\infty)$*, where* $\mathbb{R}$ *is the domain of our function* f*).*

One-sided continuity. Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, and let $a \in A$.

- We say that f is *continuous at* a *from the left* if the following holds:

  for all $\varepsilon > 0$, there exists some $\delta > 0$, such that for all $x \in A$, if $a - \delta < x \leq a$, then $|f(x) - f(a)| < \varepsilon$.

- We say that f is *continuous at* a *from the right* if the following holds:

  for all $\varepsilon > 0$, there exists some $\delta > 0$, such that for all $x \in A$, if $a \leq x < a + \delta$, then $|f(x) - f(a)| < \varepsilon$.

Remark: Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, and let $a \in A$, as in the definition above. Then the following hold:

- if a is an accumulation point of $A \cap (-\infty, a)$, then f is continuous at a from the left if and only if $\lim\limits_{x \to a^-} f(x) = f(a)$;

- if a is an accumulation point of $A \cap (a, +\infty)$, then f is continuous at a from the right if and only if $\lim\limits_{x \to a^+} f(x) = f(a)$;
- if a is an accumulation point both of $A \cap (-\infty, a)$ and of $A \cap (a, +\infty)$, then f is continuous at a if and only if it is continuous at a both from the left and from the right.

Example 3.6.10. *Consider the function* $f : \mathbb{R} \to \mathbb{R}$ *given by*

$$f(x) \quad := \quad \begin{cases} -x & \textit{if} \quad x < 0 \\ 0 & \textit{if} \quad x = 0 \\ x + 7 & \textit{if} \quad x > 0 \end{cases}$$

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

- $\lim\limits_{x \to 0^-} f(x) = 0 = f(0)$*, and so* f *is continuous at* 0 *from the left;*
- $\lim\limits_{x \to 0^+} f(x) = 7 \neq f(0)$*, and so* f *is* ***not*** *continuous at* 0 *from the right;*
- f *is* ***not*** *continuous at* 0 *(because it is not continuous at* 0 *from the right).*

3.6.3 Limits at infinity and horizontal asymptotes

For a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$, we would like to define $\lim\limits_{x \to +\infty} f(x)$ and $\lim\limits_{x \to -\infty} f(x)$. The idea is as follows:

- $\lim\limits_{x \to +\infty} f(x)$ should be equal to a real number L if when $x \in A$ is very, very large, $f(x)$ is very, very close to L;
- $\lim\limits_{x \to +\infty} f(x) = +\infty$ if when $x \in A$ is very, very large, $f(x)$ is very, very large;
- $\lim\limits_{x \to +\infty} f(x) = -\infty$ if when $x \in A$ is very, very large, $f(x)$ is very, very large in the negative direction (i.e. $f(x)$ is a negative number whose absolute value is very large).

The intuition for $\lim\limits_{x \to -\infty} f(x)$ is similar. Let us now formalize this.

Limit at $+\infty$. Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$, where the domain A of f is not bounded above.[36] Then:

- for $L \in \mathbb{R}$, we write $\lim\limits_{x \to +\infty} f(x) = L$ if the following hold:

  for all real numbers $\varepsilon > 0$, there exists some $N \in \mathbb{R}$ such that for all $x \in A$, if $x > N$, then $|f(x) - L| < \varepsilon$;

---

[36] So, we are assuming that for all $N \in \mathbb{R}$, there exists some $x \in A$ such that $x > N$.

- we write $\lim\limits_{x\to+\infty} f(x) = +\infty$ if the following hold:

  for all $M \in \mathbb{R}$, there exists some $N \in \mathbb{R}$ such that for all $x \in A$, if $x > N$, then $f(x) > M$;

- we write $\lim\limits_{x\to+\infty} f(x) = -\infty$ if the following hold:

  for all $M \in \mathbb{R}$, there exists some $N \in \mathbb{R}$ such that for all $x \in A$, if $x > N$, then $f(x) < M$.

Limit at $-\infty$. Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$, where the domain A of f is not bounded below.[37] Then:

- for $L \in \mathbb{R}$, we write $\lim\limits_{x\to-\infty} f(x) = L$ if the following hold:

  for all real numbers $\varepsilon > 0$, there exists some $N \in \mathbb{R}$ such that for all $x \in A$, if $x < N$, then $|f(x) - L| < \varepsilon$;

- we write $\lim\limits_{x\to-\infty} f(x) = +\infty$ if the following hold:

  for all $M \in \mathbb{R}$, there exists some $N \in \mathbb{R}$ such that for all $x \in A$, if $x < N$, then $f(x) > M$;

- we write $\lim\limits_{x\to-\infty} f(x) = -\infty$ if the following hold:

  for all $M \in \mathbb{R}$, there exists some $N \in \mathbb{R}$ such that for all $x \in A$, if $x < N$, then $f(x) < M$.

Horizontal asymptotes. For a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ and a real number $b \in \mathbb{R}$, we say that:

- $y = b$ is a *horizontal asymptote* of f as $x \to +\infty$ if $\lim\limits_{x\to+\infty} f(x) = b$;
- $y = b$ is a *horizontal asymptote* of f as $x \to -\infty$ if $\lim\limits_{x\to-\infty} f(x) = b$;
- y is a *horizontal asymptote* of f if at least one of $\lim\limits_{x\to+\infty} f(x)$ and $\lim\limits_{x\to-\infty} f(x)$ exists and is equal to b.

Remark: The techniques and tricks that we use to compute limits of functions at $+\infty$ and $-\infty$ are similar to those that we used for computing the limits of sequences. Let us take a look at a few examples with rational functions.

Example 3.6.11. *Consider the function* $f(x) = \frac{3x^2-2x+5}{7x^2-1}$*. (Note that the roots of* $7x^2 - 1 = 0$ *are* $\pm\frac{1}{\sqrt{7}}$*, and so the domain of* f *is* $\mathbb{R} \setminus \left\{\frac{1}{\sqrt{7}}, -\frac{1}{\sqrt{7}}\right\}$*.) We now compute:*

---

[37] So, we are assuming that for all $N \in \mathbb{R}$, there exists some $x \in A$ such that $x < N$.

$$\begin{aligned}
\lim_{x\to+\infty} f(x) &= \lim_{x\to+\infty} \tfrac{3x^2-2x+5}{7x^2-1} \\
&= \lim_{x\to+\infty} \tfrac{x^2\left(3-\frac{2}{x}+\frac{5}{x^2}\right)}{x^2\left(7-\frac{1}{x^2}\right)} \\
&= \lim_{x\to+\infty} \frac{3-\overset{\nearrow 0}{\boxed{\frac{2}{x}}}+\overset{\nearrow 0}{\boxed{\frac{5}{x^2}}}}{7-\underset{\searrow 0}{\boxed{\frac{1}{x^2}}}} = \tfrac{3}{7}.
\end{aligned}$$

Similarly, we have the following:

$$\begin{aligned}
\lim_{x\to-\infty} f(x) &= \lim_{x\to-\infty} \tfrac{3x^2-2x+5}{7x^2-1} \\
&= \lim_{x\to-\infty} \tfrac{x^2\left(3-\frac{2}{x}+\frac{5}{x^2}\right)}{x^2\left(7-\frac{1}{x^2}\right)} \\
&= \lim_{x\to-\infty} \frac{3-\overset{\nearrow 0}{\boxed{\frac{2}{x}}}+\overset{\nearrow 0}{\boxed{\frac{5}{x^2}}}}{7-\underset{\searrow 0}{\boxed{\frac{1}{x^2}}}} = \tfrac{3}{7}.
\end{aligned}$$

Note that this means that $y = \frac{3}{7}$ *is a horizontal asymptote of* f *(both as* $x \to +\infty$ *and as* $x \to -\infty$*).*

Example 3.6.12. *Consider the function* $f(x) = \frac{-x^3+x^2-x+1}{3x^2+1}$. *(Note that* $3x^2+1 > 0$ *for all* $x \in \mathbb{R}$, *and so the domain of* f *is* $\mathbb{R}$.*) We compute:*

$$\begin{aligned}
\lim_{x\to+\infty} f(x) &= \lim_{x\to+\infty} \tfrac{-x^3+x^2-x+1}{3x^2+1} \\
&= \lim_{x\to+\infty} \tfrac{x^3\left(-1+\frac{1}{x}-\frac{1}{x^2}+\frac{1}{x^3}\right)}{x^2\left(3+\frac{1}{x^2}\right)} \\
&= \lim_{x\to+\infty} \frac{\overset{\nearrow +\infty}{\boxed{x}}\left(-1+\overset{\nearrow 0}{\boxed{\frac{1}{x}}}-\overset{\nearrow 0}{\boxed{\frac{1}{x^2}}}+\overset{\nearrow 0}{\boxed{\frac{1}{x^3}}}\right)}{3+\underset{\searrow 0}{\boxed{\frac{1}{x^2}}}}
\end{aligned}$$

$$= \quad -\infty.$$

Similarly, we have the following:

$$\begin{aligned}
\lim_{x\to-\infty} f(x) &= \lim_{x\to-\infty} \tfrac{-x^3+x^2-x+1}{3x^2+1} \\
&= \lim_{x\to-\infty} \tfrac{x^3\left(-1+\frac{1}{x}-\frac{1}{x^2}+\frac{1}{x^3}\right)}{x^2\left(3+\frac{1}{x^2}\right)} \\
&= \lim_{x\to-\infty} \frac{\overset{\nearrow -\infty}{\boxed{x}}\left(-1+\overset{\nearrow 0}{\boxed{\frac{1}{x}}}-\overset{\nearrow 0}{\boxed{\frac{1}{x^2}}}+\overset{\nearrow 0}{\boxed{\frac{1}{x^3}}}\right)}{3+\underset{\searrow 0}{\boxed{\frac{1}{x^2}}}} \\
&= +\infty.
\end{aligned}$$

So, f has no horizontal asymptotes.

Example 3.6.13. *Consider the function $f(x) = \frac{x^4+x^3+1}{x^6-3x^4+3x-7}$. (It is not so simply to determine the domain of f, since there is no formula for computing the roots of polynomials of degree at least five. However, it is clear that $f(x)$ is defined when x is a real number, positive or negative, of very large absolute value.) We compute:*

$$\begin{aligned}
\lim_{x\to+\infty} f(x) &= \lim_{x\to+\infty} \tfrac{x^4+x^3+1}{x^6-3x^4+3x-7} \\
&= \lim_{x\to+\infty} \tfrac{x^4\left(1+\frac{1}{x}+\frac{1}{x^4}\right)}{x^6\left(1-\frac{3}{x^2}+\frac{3}{x^5}-\frac{7}{x^6}\right)} \\
&= \lim_{x\to+\infty} \frac{1+\overset{\nearrow 0}{\boxed{\frac{1}{x}}}+\overset{\nearrow 0}{\boxed{\frac{1}{x^4}}}}{\underset{\searrow +\infty}{\boxed{x^2}}\left(1-\underset{\searrow 0}{\boxed{\frac{3}{x^2}}}+\underset{\searrow 0}{\boxed{\frac{3}{x^5}}}-\underset{\searrow 0}{\boxed{\frac{7}{x^6}}}\right)} \quad = \quad 0.
\end{aligned}$$

Similarly, we have the following:

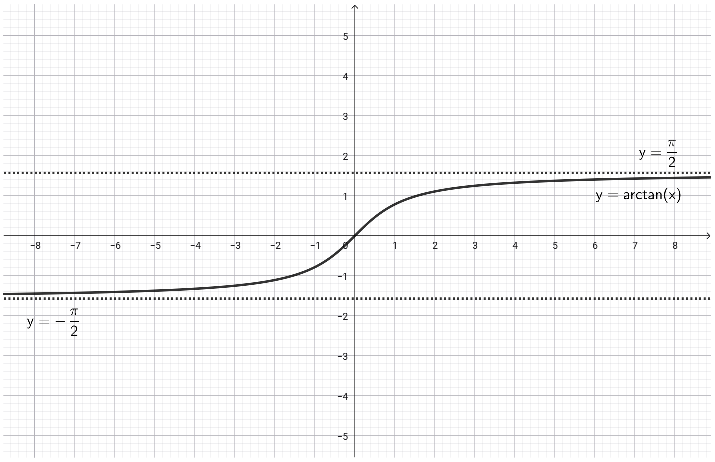
$$\lim_{x\to-\infty} f(x) \quad = \quad \lim_{x\to-\infty} \tfrac{x^4+x^3+1}{x^6-3x^4+3x-7}$$

$$= \lim_{x\to-\infty} \frac{x^4\left(1+\frac{1}{x}+\frac{1}{x^4}\right)}{x^6\left(1-\frac{3}{x^2}+\frac{3}{x^5}-\frac{7}{x^6}\right)}$$

$$= \lim_{x\to-\infty} \frac{1+\overset{\nearrow 0}{\left(\frac{1}{x}\right)}+\overset{\nearrow 0}{\left(\frac{1}{x^4}\right)}}{\underset{\searrow +\infty}{\left(x^2\right)}\left(1-\underset{\searrow 0}{\left(\frac{3}{x^2}\right)}+\underset{\searrow 0}{\left(\frac{3}{x^5}\right)}-\underset{\searrow 0}{\left(\frac{7}{x^6}\right)}\right)} = 0.$$

So, $y = 0$ is a horizontal asymptote of f (both as $x \to +\infty$ and as $x \to -\infty$).

Example 3.6.14. *Note that $\lim\limits_{x\to\infty} \arctan x = \frac{\pi}{2}$ and $\lim\limits_{x\to\infty} \arctan x = -\frac{\pi}{2}$ (see the picture below). So, $y = \frac{\pi}{2}$ is a horizontal asymptote of* $\arctan x$ *as $x \to +\infty$, and $y = -\frac{\pi}{2}$ is a horizontal asymptote of* $\arctan x$ *as $x \to -\infty$.*



3.6.4 Slant asymptotes

For a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ and real numbers $a \neq 0$ and b, we say that:

- $y = ax+b$ is a *slant asymptote* of f as $x \to +\infty$ if $\lim\limits_{x\to+\infty} \big(f(x)-(ax+b)\big) = 0$;[38]
- $y = ax+b$ is a *slant asymptote* of f as $x \to -\infty$ if $\lim\limits_{x\to-\infty} \big(f(x)-(ax+b)\big) = 0$;[39]
- $y = ax+b$ is a *slant asymptote* of f as $x \to -\infty$ if either $\lim\limits_{x\to+\infty} \big(f(x)-(ax+b)\big) = 0$ or $\lim\limits_{x\to-\infty} \big(f(x)-(ax+b)\big) = 0$.

---

[38] So, for very large x, we have that $f(x) \approx ax + b$.

[39] So, for a negative real number x of very large absolute value, we have that $f(x) \approx ax + b$.

Proposition 3.6.15. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, and let $a \neq 0$ and b be real numbers. Then:*

(a) $y = ax + b$ is a slant asymptote of f as $x \to +\infty$ if and only if the following hold:

$$\lim_{x \to +\infty} \tfrac{f(x)}{x} = a \qquad \textit{and} \qquad \lim_{x \to +\infty} \big(f(x) - ax\big) = b;$$

(b) $y = ax + b$ is a slant asymptote of f as $x \to -\infty$ if and only if the following hold:

$$\lim_{x \to -\infty} \tfrac{f(x)}{x} = a \qquad \textit{and} \qquad \lim_{x \to -\infty} \big(f(x) - ax\big) = b.$$

Proof. We prove (a); the proof of (b) is similar.

Suppose first that $y = ax + b$ is a slant asymptote of f as $x \to +\infty$, so that $\lim\limits_{x \to +\infty} \big(f(x) - (ax + b)\big) = 0$. Then

$$\begin{aligned} \lim_{x \to +\infty} \big(f(x) - ax\big) &= \lim_{x \to +\infty} \big(b + f(x) - (ax + b)\big) \\ &= b + \lim_{x \to +\infty} \big(f(x) - (ax + b)\big) \\ &= b + 0 \;=\; b. \end{aligned}$$

This obviously implies that

$$\lim_{x \to +\infty} \tfrac{f(x) - ax}{x} \;=\; 0,$$

and therefore,

$$\lim_{x \to +\infty} \tfrac{f(x)}{x} \;=\; \lim_{x \to +\infty} \big(a + \tfrac{f(x) - ax}{x}\big) \;=\; a + \lim_{x \to +\infty} \tfrac{f(x) - ax}{x} \;=\; a.$$

We have now shown that $\lim\limits_{x \to +\infty} \frac{f(x)}{x} = a$ and $\lim\limits_{x \to \infty} \big(f(x) - ax\big) = b$, as required.

Suppose conversely that $\lim\limits_{x \to +\infty} \frac{f(x)}{x} = a$ and $\lim\limits_{x \to \infty} \big(f(x) - ax\big) = b$. Then

$$\lim_{x \to \infty} \big(f(x) - (ax + b)\big) \;=\; \Big(\lim_{x \to \infty} \big(f(x) - ax\big)\Big) - b \;=\; b - b \;=\; 0,$$

which proves that $y = ax + b$ is a slant asymptote of f.[40] □

Example 3.6.16. *Determine whether the function $f(x) = \frac{x^2 + x + 1}{x - 2}$ has any slant asymptotes, and if so, compute them.*

---

[40] Note that for this direction, we only used the fact that $\lim\limits_{x \to \infty} \big(f(x) - ax\big) = b$, and we never actually made use of the fact that $\lim\limits_{x \to +\infty} \frac{f(x)}{x} = a$.

Solution. First, we note that

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

and

$$\begin{aligned} b_+ \quad &= \quad \lim_{x\to+\infty} \Big(f(x)-a_+x\Big) \\ &= \quad \lim_{x\to+\infty} \Big(\tfrac{x^2+x+1}{x-2}-x\Big) \\ &= \quad \lim_{x\to+\infty} \tfrac{3x+1}{x-2} \\ &= \quad \lim_{x\to+\infty} \tfrac{3+\frac{1}{x}}{1-\frac{2}{x}} \\ &= \quad 3. \end{aligned}$$

So, by Proposition 3.6.15, $y = x+3$ is a slant asymptote of f as $x \to +\infty$.

Similarly, we can compute

$$a_- = \lim_{x\to-\infty} \tfrac{f(x)}{x} = 1 \quad \text{and} \quad b_- = \lim_{x\to-\infty} \Big(f(x)-a_-x\Big) = 3,$$

and so Proposition 3.6.15 guarantees that $y = x+3$ is a slant asymptote of f as $x \to -\infty$. □

3.7 The Squeeze Theorem for functions

The Squeeze Theorem for functions. *Let $f, g, h : A \subseteq \mathbb{R} \to \mathbb{R}$ be functions such that*

$$f(x) \quad \leq \quad g(x) \quad \leq \quad h(x) \qquad \forall x \in A,$$

let $a \in \mathbb{R}$ be an accumulation point of A, and let $L \in \mathbb{R}$. If $\lim\limits_{x\to a} f(x) = \lim\limits_{x\to a} h(x) = L$, then $\lim\limits_{x\to a} g(x) = L$.

Proof. Assume that $\lim\limits_{x\to a} f(x) = \lim\limits_{x\to a} h(x) = L$, and fix $\varepsilon > 0$. Using the fact that $\lim\limits_{x\to a} f(x) = L$, we fix $\delta_1 > 0$ such that for all $x \in A$, if $0 < |x-a| < \delta_1$, then $|f(x)-L| < \varepsilon$, i.e. $L-\varepsilon < f(x) < L+\varepsilon$. Similarly, using the fact that $\lim\limits_{x\to a} h(x) = L$, we fix $\delta_2 > 0$ such that for all $x \in A$, if $0 < |x-a| < \delta_2$, then $|h(x)-L| < \varepsilon$, i.e. $L-\varepsilon < h(x) < L+\varepsilon$. Set $\delta = \min\{\delta_1, \delta_2\}$. Then for all $x \in A$ such that $0 < |x-a| < \delta$, we have that

$$L-\varepsilon \quad < \quad f(x) \quad \leq \quad g(x) \quad \leq \quad h(x) \quad < \quad L+\varepsilon,$$

and it follows that $|g(x)-L| < \varepsilon$. This proves that $\lim\limits_{x\to a} g(x) = L$, as we had claimed. □

Example 3.7.1. *Compute* $\lim\limits_{x \to 0} \left(x \sin(\frac{1}{x})\right)$.

Solution. Clearly,

$$-|x| \quad \leq \quad x \sin(\tfrac{1}{x}) \quad \leq \quad |x| \qquad \forall x \in \mathbb{R}.$$

Obviously, $\lim\limits_{x \to 0}(-|x|) = \lim\limits_{x \to 0} |x| = 0$. So, by the Squeeze Theorem, we have that $\lim\limits_{x \to 0} \left(x \sin(\frac{1}{x})\right) = 0$. □

Remark: To apply the Squeeze Theorem, we do not really need the inequality $f(x) \leq g(x) \leq h(x)$ to hold for all x in the domain. It is enough to have the inequality for values of x close to a (but not necessarily at a itself), as follows. (The proof is very similar to that of the Squeeze Theorem above, and it is left as an exercise.)

The Squeeze Theorem for functions (stronger version). *Let $f, g, h : A \subseteq \mathbb{R} \to \mathbb{R}$ be functions, let $a \in \mathbb{R}$ be an accumulation point of A, and let $L \in \mathbb{R}$. Assume that for some $\delta > 0$, the following holds:*

$$f(x) \quad \leq \quad g(x) \quad \leq \quad h(x) \qquad \forall x \in A \cap \Big((a-\delta, a) \cup (a, a+\delta)\Big)$$

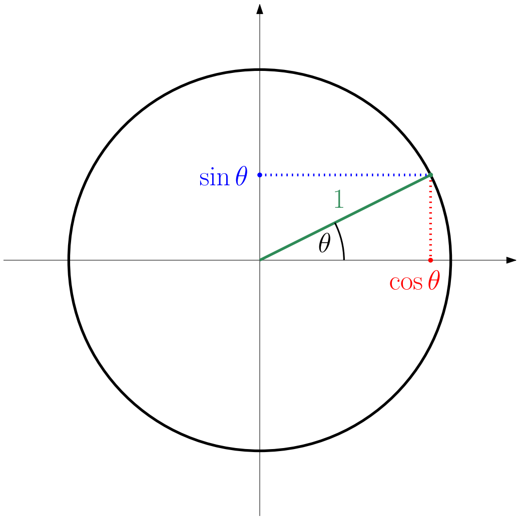
If $\lim\limits_{x \to a} f(x) = \lim\limits_{x \to a} h(x) = L$, then $\lim\limits_{x \to a} g(x) = L$.

Proof. Exercise. □

Remark: Similar "Squeeze Theorems" can obviously be formulated for one-sided limits ("$x \to a^-$" or "$x \to a^+$"), as well as for limits at infinity ("$x \to +\infty$" or "$x \to -\infty$"). The exact formulations and the proofs of these theorems are left as an exercise for the reader.

Theorem 3.7.2 (below) establishes that $\lim\limits_{x \to 0} \frac{\sin x}{x} = 1$. We will prove the theorem using some basic trigonometry and geometry, combined with the Squeeze Theorem. To this end, let us us recall some basic trigonometric and geometric facts.

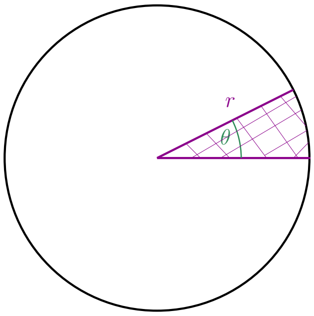
Trigonometric functions. Given a real number θ (which we consider to be an angle in radians), we obtain $\sin\theta$ and $\cos\theta$ as follows. First, we consider the unit circle (i.e. the circle in the coordinate plane, centered at the origin and of radius 1). We consider the point on the unit circle determined by the angle θ (starting at the x-axis, we move θ radians counterclockwise if $\theta \geq 0$, and we move $-\theta$ radians clockwise if $\theta < 0$). The coordinates of this point are $(\cos\theta, \sin\theta)$, as in the picture below.



If θ is not an odd integer multiple of $\frac{\pi}{2}$ (and so $\cos\theta \neq 0$), then we define $\tan\theta := \frac{\sin\theta}{\cos\theta}$. Similarly, if θ is not an integer multiple of π (and so $\sin\theta \neq 0$), then we define $\cot\theta := \frac{\cos\theta}{\sin\theta}$.

Remark: Unless explicitly specified otherwise, when we write $\sin x$, $\cot x$, $\tan x$, or $\cot x$ (in analysis), we always assume that x is an angle in **radians** (not degrees).

The area of a sector. Suppose we are given real numbers $r > 0$ and $\theta \in [0, 2\pi]$, where we think of θ as an angle in radians. The area of a disk with radius r is $r^2\pi$. Meanwhile, the area of the sector of the disk determined by the angle θ (see the picture below) is $\frac{\theta}{2\pi}r^2\pi = \frac{1}{2}r^2\theta$.



Theorem 3.7.2. $\lim\limits_{x\to 0} \frac{\sin x}{x} = 1$.

Proof. Our goal is to show that $\lim\limits_{x\to 0^+} \frac{\sin x}{x} = 1$. This is enough, because once we have shown this, it will follow that

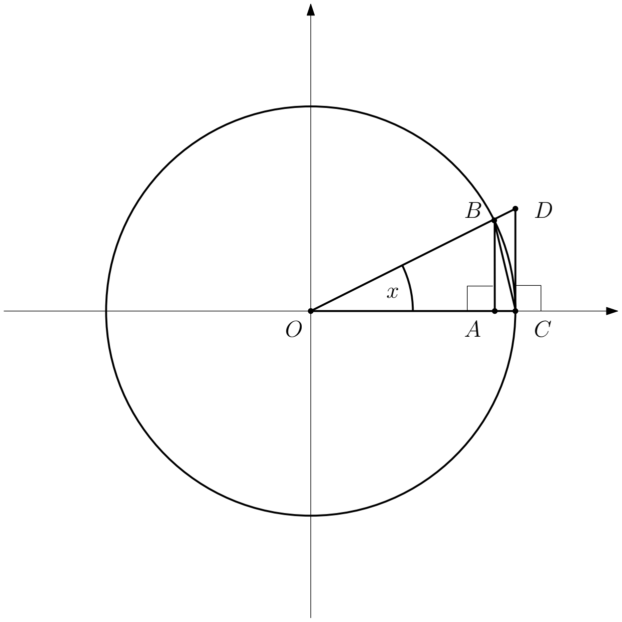
$$\lim_{x\to 0^-} \frac{\sin x}{x} \quad = \quad \lim_{x\to 0^-} \frac{-\sin x}{-x} \quad = \quad \lim_{x\to 0^-} \frac{\sin(-x)}{-x} \quad = \quad \lim_{x\to 0^+} \frac{\sin x}{x} \quad = \quad 1,$$

and consequently, $\lim\limits_{x\to 0^+} \frac{\sin x}{x} = \lim\limits_{x\to 0^-} \frac{\sin x}{x} = 1$, which in turn implies that $\lim\limits_{x\to 0} \frac{\sin x}{x} = 1$ (by Lemma 3.6.3). We will prove that $\lim\limits_{x\to 0^+} \frac{\sin x}{x} = 1$ by showing that

$$\cos x \quad < \quad \frac{\sin x}{x} \quad < \quad 1 \qquad \forall x \in \big(0, \tfrac{\pi}{2}\big).$$

Since $\lim\limits_{x\to 0^+} \cos x = 1$, the Squeeze Theorem will immediately imply that $\lim\limits_{x\to 0^+} \frac{\sin x}{x} = 1$, which is what we need to show.

Now, fix $x \in (0, \frac{\pi}{2})$, and consider the diagram below (where the circle is the unit circle, i.e. its radius is 1).



As usual, for points X and Y, we will denote by $|XY|$ the length of the line segment XY. Note that:

- $|OB| = |OC| = 1$;
- $|AB| = \sin x$;
- $|OA| = \cos x$;
- $|CD| = \tan x$.[41]

We now observe the following:

---

[41] You may know this fact from trigonometry, but in case you do not, here is a way to derive it. The triangle ΔOCD is similar to the triangle ΔOAB, and so $\frac{|CD|}{|AB|} = \frac{|OC|}{|OA|}$. Therefore, $|CD| = \frac{|OC|}{|OA|}|AB| = \frac{1}{\cos x}\sin x = \tan x$.

- the area of the triangle ΔOCB is $\frac{1}{2}|OC||AB| = \frac{1}{2}\sin x$;
- the area of the sector OCB is $\frac{1}{2}x$;
- the area of the triangle ΔOCD is $\frac{1}{2}|OC||CD| = \frac{1}{2}\tan x$.

Obviously,

$$\text{area}\big(\Delta OCB\big) \quad < \quad \text{area}\big(\text{sector } OCB\big) \quad < \quad \text{area}\big(\Delta OCD\big),$$

and consequently,

$$\tfrac{1}{2}\sin x \quad < \quad \tfrac{1}{2}x \quad < \quad \tfrac{1}{2}\tan x.$$

Since $0 < x < \frac{\pi}{2}$, we know that $\sin x > 0$; therefore, we may multiply the inequality above by $\frac{2}{\sin x}$ to obtain

$$1 \quad < \quad \tfrac{x}{\sin x} \quad < \quad \tfrac{1}{\cos x}.$$

Since all values in the inequality above are positive (because $0 < x < \frac{\pi}{2}$), we may take reciprocals to obtain

$$1 \quad > \quad \tfrac{\sin x}{x} \quad > \quad \cos x,$$

and we are done. □

Analogues of the Squeeze Theorem for infinite limits. Theorems 3.7.3 and 3.7.4 (see below) can be thought of as analogues of the Squeeze Theorem for infinite limits.

Theorem 3.7.3. *Let $f, g : A \subseteq \mathbb{R} \to \mathbb{R}$ be functions such that*

$$f(x) \quad \leq \quad g(x) \qquad \forall x \in A.$$

Let $a \in \mathbb{R}$ be an accumulation point of A. Then the following hold:

(a) if $\lim\limits_{x \to a} f(x) = +\infty$, then $\lim\limits_{x \to a} g(x) = +\infty$;

(b) if $\lim\limits_{x \to a} g(x) = -\infty$, then $\lim\limits_{x \to a} f(x) = -\infty$.

Proof. We prove (a); the proof of (b) is similar. Fix $M \in \mathbb{R}$. Using the fact that $\lim\limits_{x \to a} f(x) = +\infty$, we fix some $\delta > 0$ such that for all $x \in A$, if $0 < |x - a| < \delta$, then $f(x) > M$. But now for all $x \in A$ such that $0 < |x - a| < \delta$, we have that

$$g(x) \quad \geq \quad f(x) \quad > \quad M,$$

which proves that $\lim\limits_{x \to a} f(x) = +\infty$. □

Remark: As in the case of the Squeeze Theorem, Theorem 3.7.3 can be be strengthened, as follows: we do not need to assume that $f(x) \leq g(x)$ for all points x in the domain; it is enough that this inequality should hold for values of x close to a. More precisely, we have the theorem below.

Theorem 3.7.4. *Let* $f, g : A \subseteq \mathbb{R} \to \mathbb{R}$ *be functions, and let* $a \in \mathbb{R}$ *be an accumulation point of* A*. Assume that for some* $\delta > 0$*, we following holds:*

$$f(x) \quad \leq \quad g(x) \qquad\qquad \forall x \in A \cap \Big((a-\delta, a) \cup (a, a+\delta)\Big).$$

Then the following hold:

(a) if $\lim\limits_{x \to a} f(x) = +\infty$*, then* $\lim\limits_{x \to a} g(x) = +\infty$*;*

(b) if $\lim\limits_{x \to a} g(x) = -\infty$*, then* $\lim\limits_{x \to a} f(x) = -\infty$*.*

Proof. Exercise. □

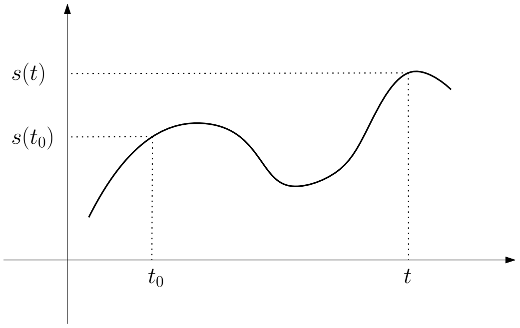
Remark: Theorems 3.7.3 and 3.7.4 could easily be formulated and proven for one-sided limits as well. The details are left as an exercise for the reader.

Chapter 4

Differentiation

4.1 Motivating the derivative: instantaneous velocity and tangent to the curve

Instantaneous velocity. Suppose a particle is traveling along a straight line (possibly changing speed and direction over time). Suppose we then plot the position of the particle against time. The horizontal axis (the "t-axis" in the picture below) denotes time, the vertical axis (the "s-axis") denotes position, and $s(t)$ denotes the position of the particle at time t.



Now, consider two moments in time: t_0 and t_1 (for convenience, let us assume that $t_0 < t_1$). The average velocity of the particle during the time interval $[t_0, t_1]$ is

$$v_{[t_0,t_1]} \quad := \quad \tfrac{s(t_1)-s(t_0)}{t_1-t_0}.$$

If $t_1 \approx t_0$, then the velocity above is something like the instantaneous velocity of the particle. More formally, we can take the limit! So, the instantaneous velocity of the particle at time t_0 is

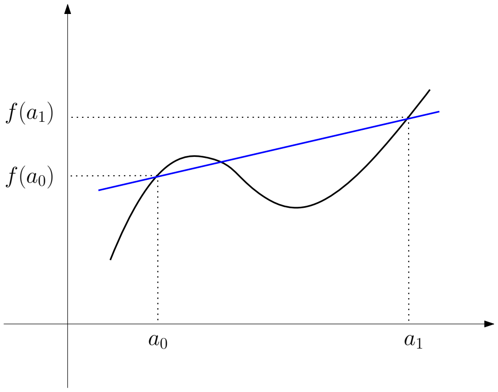
$$v_{t_0} \quad := \quad \lim_{t \to t_0} \tfrac{s(t)-s(t_0)}{t-t_0},$$

if the limit exists.

The tangent to the curve. Suppose we are given a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ and distinct points $a_0, a_1 \in A$. Then the equation of the secant line to the graph of the function f through the points $\big(a_0, f(a_0)\big)$ and $\big(a_1, f(a_1)\big)$ is

$$y \quad = \quad \tfrac{f(a_1)-f(a_0)}{a_1-a_0}(x-a_0)+f(a_0),$$

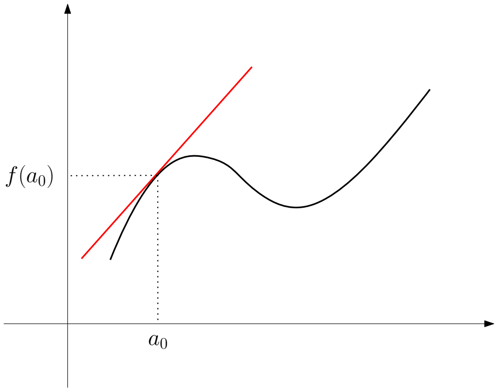
The slope of this line (i.e. the coefficient in front of x) is $\frac{f(a_1)-f(a_0)}{a_1-a_0}$. (See the picture below. The secant line in question is in blue.)



Now, suppose we take $a_1 \approx a_0$. Then the secant line that we obtain gets very close to a tangent line. The actual equation of the tangent to the graph of f through the point $\big(a_0, f(a_0)\big)$ is obtained by taking the limit, as follows:

$$y \quad := \quad \big(\lim_{a \to a_0} \tfrac{f(a)-f(a_0)}{a-a_0}\big)(x-a_0)+f(a_0),$$

provided the limit exists. The slope of this tangent is $\lim\limits_{a \to a_0} \frac{f(a)-f(a_0)}{a-a_0}$. (See the picture below. The tangent in question is in red.)



4.2 The definition of the derivative

The derivative. Suppose that $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is a function and $a_0 \in A$ is an accumulation point of A. Then the *derivative* of f at a_0 is defined to be

$$f'(a_0) \quad := \quad \lim_{a \to a_0} \tfrac{f(a)-f(a_0)}{a-a_0},$$

provided the limit exists (and is a real number). If the limit above exists, then we say that f is *differentiable* at a_0. We say that the function f is *differentiable* provided that f is differentiable at all points in its domain.

Differentiability on an open interval. Suppose that $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is a function, and that $I \subseteq A$ is an open interval. (In particular, we are assuming that f is **defined** at all points in I.) We say that f is *differentiable on* I if f is differentiable at all points in I.

Remark: In practice, we will most commonly deal with the situation where we have a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$, and a point $a_0 \in I \subseteq A$, where I is some open interval. In this case, f is defined on some open neighborhood (open interval) containing a_0, and in particular, a_0 belongs to A and is an accumulation point of A. With this set-up, the definition above applies, i.e. the derivative of f at a_0 is defined to be

$$f'(a_0) \quad := \quad \lim_{a \to a_0} \tfrac{f(a)-f(a_0)}{a-a_0},$$

provided the limit exists (as a real number). Sometimes, it is convenient to rewrite the above limit as

$$f'(a_0) \quad := \quad \lim_{h \to 0} \tfrac{f(a_0+h)-f(a_0)}{h},$$

provided the limit exists (as a real number).

Example 4.2.1. *Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ for all $x \in \mathbb{R}$. Consider any $a_0 \in \mathbb{R}$. Then*

$$\begin{aligned} f'(a_0) &= \lim_{h \to 0} \tfrac{f(a_0+h)-f(a_0)}{h} \\ &= \lim_{h \to 0} \tfrac{(a_0+h)^2-a_0^2}{h} \\ &= \lim_{h \to 0} \tfrac{(a_0^2+2a_0h+h^2)-a_0^2}{h} \\ &= \lim_{h \to 0} \tfrac{2a_0h+h^2}{h} \\ &= \lim_{h \to 0} (2a_0+h) \\ &= 2a_0. \end{aligned}$$

The derivative as a function. Suppose that $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is a function, and A' is the set of all points at which f is differentiable. Then $f' : A' \to \mathbb{R}$ is a function in its own right.

Example 4.2.2. *Consider the function* $f : \mathbb{R} \to \mathbb{R}$ *given by* $f(x) = x^2$ *for all* $x \in \mathbb{R}$. *It then follows from Example 4.2.1 that the function* f *is differentiable, and the function* $f' : \mathbb{R} \to \mathbb{R}$ *is given by* $f'(x) = 2x$ *for all* $x \in \mathbb{R}$.

The Leibniz notation for the derivative. The notation $f'(x)$ that we introduced above is due to Lagrange. However, we sometimes use the following notation due to Leibniz instead:

$$\tfrac{df(x)}{dx} \qquad \text{or} \qquad \tfrac{df}{dx}(x) \qquad \text{or} \qquad \tfrac{d}{dx}f(x).$$

The three expressions above have exactly the same meaning as $f'(x)$.

The Leibniz notation may be convenient when a function f is given by a formula, so that we may write $\frac{d}{dx}(x^4 - 3x^2 + 1)$ rather than $(x^4 - 3x^2 + 1)'$. The Leibniz notation is even more convenient when we need to differentiate an expression that involves more than one "letter," one of which is the variable with respect to which we are differentiating, whereas the others are treated as constants. For instance, $\frac{d}{dx}(x^\alpha)$ tells us that we are differentiating with respect to the variable x, while treating α as a constant. Similarly, $\frac{d}{dt}(ct^2)$ tells us that we are differentiating with respect to t, while treating c as a constant.

We will also occasionally see equations of the form $y = f(x)$, in which case we may write $\frac{dy}{dx}$ instead of $\frac{df}{dx}(x)$ or $f'(x)$. If we mean to write "the derivative of f evaluated at a_0," then we may write

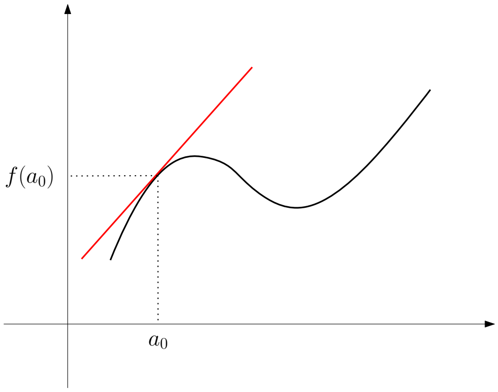
$$\tfrac{dy}{dx}\Big|_{x=a_0}$$

instead of $f'(a_0)$.

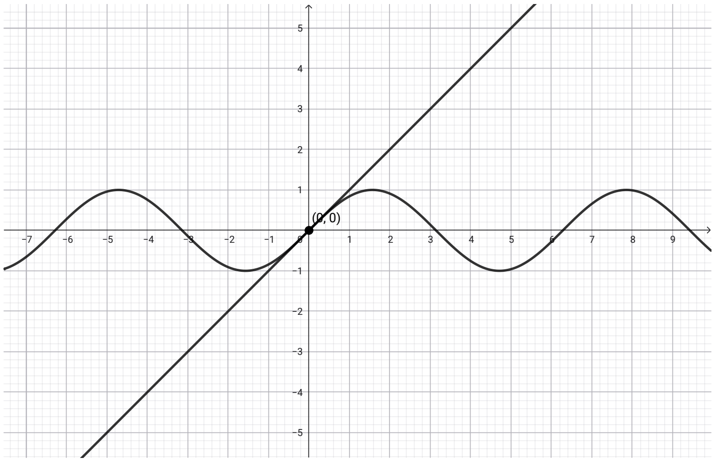
The derivative as the rate of change. Suppose that $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is a function and $a_0 \in A$ is an accumulation point of A, as in the definition of the derivative. If f is differentiable at a_0, then we may think of $f'(a_0)$ as the "rate of change" at the point a_0 of the function f. Note that this is consistent with our discussion of "instantaneous velocity" in section 4.1.

Tangent lines. Suppose $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is a function, differentiable at a point $a_0 \in A$ (in particular, a_0 is an accumulation point of A). Consistently with our discussion in section 4.1, we define the *tangent line* (or simply *tangent*) to the graph of f though the point $\big(a_0, f(a_0)\big)$ to be the line though that point, and with slope $f'(a_0)$. Note that the equation of this line (if it is exists) is

$$y \;=\; f'(a_0)(x - a_0) + f(a_0).$$



Remark: Note that, according to our definition, the tangent to the graph of a function through a given point may possibly "pierce" the graph of the function (rather than touching it). For instance, it can be shown that $\frac{d}{dx}(\sin x) = \cos x$ (see Proposition 4.4.9), and it readily follows that the equation of the tangent to $\sin x$ through the point $(0, \sin 0) = (0, 0)$ is $y = x$.[1] This is represented in the graph below; note that the tangent "pierces" the graph of $\sin x$ in the point $(0, 0)$.



Example 4.2.3. *Compute the equations of the tangents to the graph of the function* $f(x) = x^2$ *through each of the following points:* $(0, 0)$, $(-1, 1)$, $(2, 4)$.

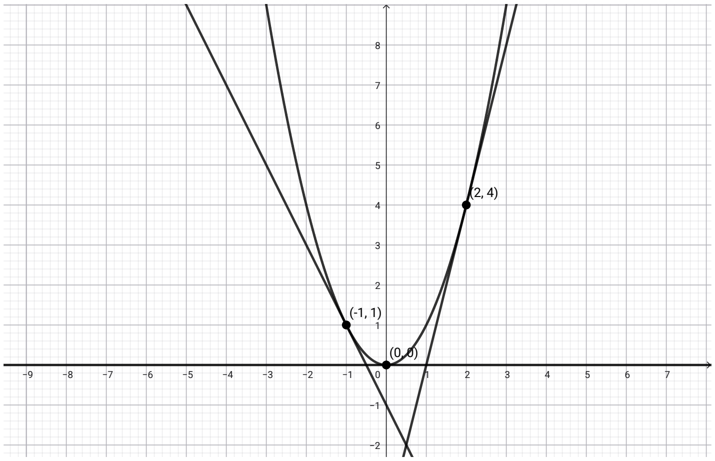
Proof. By Example 4.2.2, we have that $f'(x) = 2x$ for all $x \in \mathbb{R}$. Now, by definition, for any point $a \in \mathbb{R}$, the equation of the tangent to the graph of f through the point $(a, f(a)) = (a, a^2)$ is $y = f'(a)(x - a) + f(a)$, that is, $y = 2a(x - a) + a^2$, i.e. $y = 2ax - a^2$. Therefore:

---

[1] Indeed, the equation of the tangent to $\sin x$ through $(0, \sin 0)$ is $y = \sin'(0)(x - 0) + \sin 0$. Since $\sin'(0) = \cos 0 = 1$ and $\sin 0 = 0$, it follows that the equation of our tangent is $y = x$, as we had claimed.

- the equation of the tangent to the graph of f through the point $(0,0)$ is $y=0$ (i.e. the x-axis);
- the equation of the tangent to the graph of f through the point $(-1,1)$ is $y=-2x-1$;
- the equation of the tangent to the graph of f through the point $(2,4)$ is $y=4x-4$.

The graph of f and the three tangent lines are represented in the graph below.



$\square$

4.3 Differentiability and continuity

Theorem 4.3.1. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, and let $a_0 \in A$ be an accumulation point of A. If f is differentiable at a_0, then f is continuous at a_0.*

Proof. We assume that f is differentiable at a_0, so that

$$f'(a_0) \quad = \quad \lim_{a\to a_0} \frac{f(a)-f(a_0)}{a-a_0}$$

exists (as a real number). To prove that f is continuous at a_0, it suffices to show that $\lim\limits_{a\to a_0} f(a) = f(a_0)$. For this, we compute:

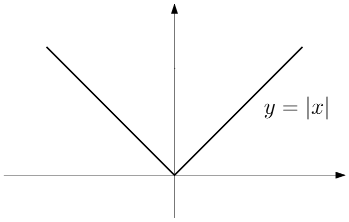
$$\begin{aligned}
\lim_{a\to a_0} f(a) \quad &= \quad f(a_0) + \lim_{a\to a_0} \big(f(a)-f(a_0)\big) \\
&= \quad f(a_0) + \lim_{a\to a_0} \Big({\color{red}\tfrac{f(a)-f(a_0)}{a-a_0}}{\color{blue}(a-a_0)}\Big) \\
&= \quad f(a_0) + \Big({\color{red}\lim_{a\to a_0} \tfrac{f(a)-f(a_0)}{a-a_0}}\Big)\Big({\color{blue}\lim_{a\to a_0}(a-a_0)}\Big)
\end{aligned}$$

$$= \quad f(a_0) + f'(a_0) \cdot 0 \quad = \quad f(a_0).$$

This proves that f is indeed continuous at a_0, as we had claimed. □

Remark: The converse of Theorem 4.3.1 is false. Indeed, as the example below shows, there are continuous functions that are not differentiable.

Example 4.3.2. *Consider the function* $f : \mathbb{R} \to \mathbb{R}$ *given by* $f(x) = |x|$ *for all* $x \in \mathbb{R}$. *Show that* f *is continuous at* 0, *but is not differentiable at* 0.



Solution. By definition, we have that

$$f(x) \quad = \quad \begin{cases} x & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x = 0 \\ -x & \text{if} \quad x < 0 \end{cases}$$

for all $x \in \mathbb{R}$. We then have that

- $\lim\limits_{x \to 0^+} f(x) = \lim\limits_{x \to 0^+} x = 0$;
- $\lim\limits_{x \to 0^-} f(x) = \lim\limits_{x \to 0^-} (-x) = 0$.

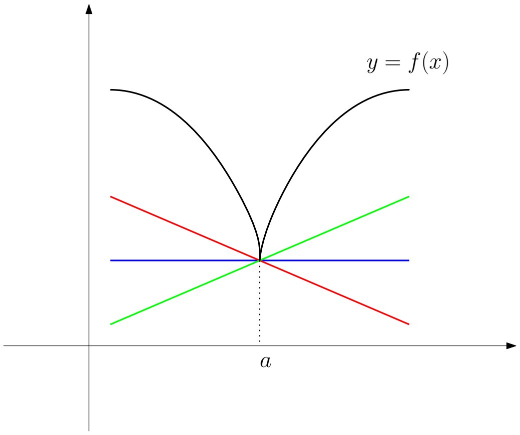
Since $\lim\limits_{x \to 0^+} f(x) = \lim\limits_{x \to 0^-} f(x) = 0$, it follows that $\lim\limits_{x \to 0} f(x) = 0$, i.e. $\lim\limits_{x \to 0} f(x) = f(0)$. So, f is continuous at 0.

It remains to show that f is **not** differentiable at 0. Note that:

- $\lim\limits_{x \to 0^+} \frac{f(x)-f(0)}{x-0} = \lim\limits_{x \to 0^+} \frac{x-0}{x-0} = 1$;
- $\lim\limits_{x \to 0^-} \frac{f(x)-f(0)}{x-0} = \lim\limits_{x \to 0^-} \frac{-x-0}{x-0} = -1$.

Thus, $\lim\limits_{x \to 0^+} \frac{f(x)-f(0)}{x-0} \neq \lim\limits_{x \to 0^-} \frac{f(x)-f(0)}{x-0}$. Consequently, $\lim\limits_{x \to 0} \frac{f(x)-f(0)}{x-0}$ does not exist, that is, $f'(0)$ does not exist. So, f is **not** differentiable at 0. □

Remark: Intuitively, a function that is continuous at a point a, but has a "sharp corner" at a, is not differentiable at a. Note that the presence of "sharp corners" prevents the existence of a tangent.



4.3.1 Sums and scalar multiples of differentiable functions

Proposition 4.3.3. *Let $f, g : A \subseteq \mathbb{R} \to \mathbb{R}$ be functions, and let $a_0 \in A$ be an accumulation point of A. If f and g are both differentiable at a_0, then so is the function $f+g$, and moreover, we have that $(f+g)'(a_0) = f'(a_0) + g'(a_0)$.*

Proof. Assume that f and g are both differentiable at a_0. Then

$$\begin{aligned}
(f+g)'(a_0) &= \lim_{a \to a_0} \tfrac{(f+g)(a)-(f+g)(a_0)}{a-a_0} \\
&= \lim_{a \to a_0} \tfrac{f(a)+g(a)-f(a_0)-g(a_0)}{a-a_0} \\
&= \lim_{a \to a_0} \tfrac{\big(f(a)-f(a_0)\big)+\big(g(a)-g(a_0)\big)}{a-a_0} \\
&= \Big(\lim_{a \to a_0} \tfrac{f(a)-f(a_0)}{a-a_0}\Big) + \Big(\lim_{a \to a_0} \tfrac{g(a)-g(a_0)}{a-a_0}\Big) \\
&= f'(a_0) + g'(a_0).
\end{aligned}$$

In particular, $f+g$ is indeed differentiable at a_0. □

Proposition 4.3.4. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, let $a_0 \in A$ be an accumulation point of A, and let $c \in \mathbb{R}$ be a constant. If f is differentiable at a_0, then so is the function cf, and moreover, we have that $(cf)'(a_0) = c\big(f'(a_0)\big)$.*

Proof. Assume that f is differentiable at a_0. Then

$$(cf)'(a_0) = \lim_{a \to a_0} \tfrac{(cf)(a)-(cf)(a_0)}{a-a_0}$$

$$
\begin{array}{rl}
= & \lim\limits_{a\to a_0} \frac{c\big(f(a)-f(a_0)\big)}{a-a_0} \\[2ex]
= & c \lim\limits_{a\to a_0} \frac{f(a)-f(a_0)}{a-a_0} \\[2ex]
= & c\big(f'(a_0)\big).
\end{array}
$$

In particular, the function cf is differentiable at a_0. □

4.4 The derivatives of some important functions

4.4.1 Constant functions

A function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is *constant* if there exists some $c \in \mathbb{R}$ such that for all $x \in A$, we have that $f(x) = c$.

Proposition 4.4.1. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a constant function, and let $a_0 \in A$ be an accumulation point of A. Then $f'(a_0) = 0$.*

Proof. We compute:

$$
f'(a_0) \quad = \quad \lim_{a\to a_0} \frac{f(a)-f(a_0)}{a-a_0} \quad \overset{(*)}{=} \quad \lim_{a\to a_0} \frac{0}{a-a_0} \quad = \quad 0,
$$

where (*) follows form the fact that f is constant, and so $f(a) = f(a_0)$ for all $a \in A$. □

4.4.2 Polynomial functions

Proposition 4.4.2. *Let n be a positive integer. Then $\frac{d}{dx}(x^n) = nx^{n-1}$ for $x \in \mathbb{R}$.*

Proof. For $x \in \mathbb{R}$, we compute:

$$
\begin{array}{rcl}
\frac{d}{dx}(x^n) & = & \lim\limits_{h\to 0} \frac{(x+h)^n - x^n}{h} \\[2ex]
& = & \lim\limits_{h\to 0} \frac{\Big(\sum\limits_{i=0}^{n} \binom{n}{i} x^{n-i}h^i\Big) - x^n}{h} \\[2ex]
& = & \lim\limits_{h\to 0} \frac{\Big(x^n h^0 + \sum\limits_{i=1}^{n} \binom{n}{i} x^{n-i}h^i\Big) - x^n}{h} \\[2ex]
& = & \lim\limits_{h\to 0} \frac{\sum\limits_{i=1}^{n} \binom{n}{i} x^{n-i}h^i}{h}
\end{array}
$$

$$\begin{aligned}
&= \lim_{h \to 0} \Big(\sum_{i=1}^{n} \tbinom{n}{i} x^{n-i} h^{i-1} \Big) \\
&= \sum_{i=1}^{n} \tbinom{n}{i} x^{n-i} 0^{i-1} \\
&\overset{(*)}{=} \tbinom{n}{1} x^{n-1} \\
&= n x^{n-1},
\end{aligned}$$

where (*) follows from the fact that $0^0 = 1$, whereas $0^k = 0$ for all $k \in \mathbb{N}$. $\square$

Example 4.4.3. *Consider the polynomial function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 4x^3 + 7x^2 - 5x + 3$ for all $x \in \mathbb{R}$. Then*

$$\begin{aligned}
f'(x) &= \tfrac{d}{dx}\big(4x^3 + 7x^2 - 5x + 3\big) \\
&= 4\tfrac{d}{dx}(x^3) + 7\tfrac{d}{dx}(x^2) - 5\tfrac{d}{dx}(x) + \tfrac{d}{dx}(3) && \textit{by Propositions 4.3.3 and 4.3.4} \\
&= 4(3x^2) + 7(2x) - 5(1) + (0) && \textit{by Propositions 4.4.1 and 4.4.2} \\
&= 12x^2 + 14x - 5.
\end{aligned}$$

Remark: Once we develop more theory, we will see that the formula from Proposition 4.4.2 generalizes to **real** exponents α (in place of the positive integer exponent n), though we may possibly require $x > 0$ to ensure that all terms are defined. More precisely, for a fixed real number α, we will get the formula

$$\tfrac{d}{dx}(x^\alpha) = \alpha x^{\alpha - 1} \qquad \text{for } x \in (0, +\infty).$$

We cannot yet derive this formula in its full generality. We do, however, have the following proposition. (Recall that for an integer $n \geq 2$, we have that $x^{1/n} = \sqrt[n]{x}$.)

Proposition 4.4.4. *Let $n \geq 2$ be an integer. Then:*

(a) if n is odd, then $\frac{d}{dx}(x^{1/n}) = \frac{1}{n}x^{(1/n)-1} = \frac{1}{n}x^{-(n-1)/n}$ for $x \in \mathbb{R} \setminus \{0\}$;

(b) if n is even, then $\frac{d}{dx}(x^{1/n}) = \frac{1}{n}x^{(1/n)-1} = \frac{1}{n}x^{-(n-1)/n}$ for $x \in (0, +\infty)$.

Remark: Under our assumption that $n \geq 2$, the expression $0^{-(n-1)/n}$ is undefined (regardless of the parity of n), an in fact, the function $x^{1/n}$ is **not** differentiable at 0.

Proof. We prove both parts simultaneously. If n is odd, then we allow x to be any non-zero integer in the computation below, and if n is even, then x is assumed to be strictly positive. We now compute:

$$\begin{array}{rcl}
\frac{d}{dx}(x^{1/n}) & = & \frac{d}{dx}(\sqrt[n]{x}) \;=\; \lim\limits_{h\to 0} \frac{\sqrt[n]{x+h}-\sqrt[n]{x}}{h} \\[2ex]
& = & \lim\limits_{h\to 0} \frac{\left(\sqrt[n]{x+h}-\sqrt[n]{x}\right)\left(\sum\limits_{i=0}^{n-1}(\sqrt[n]{x+h})^i(\sqrt[n]{x})^{n-i-1}\right)}{h\left(\sum\limits_{i=0}^{n-1}(\sqrt[n]{x+h})^i(\sqrt[n]{x})^{n-i-1}\right)} \\[2ex]
& = & \lim\limits_{h\to 0} \frac{(x+h)-x}{h\left(\sum\limits_{i=0}^{n-1}(x+h)^{i/n}\; x^{(n-i-1)/n}\right)} \\[2ex]
& = & \lim\limits_{h\to 0} \frac{1}{\sum\limits_{i=0}^{n-1}(x+h)^{i/n}\; x^{(n-i-1)/n}} \\[2ex]
& = & \frac{1}{\sum\limits_{i=0}^{n-1}(x+0)^{i/n}\; x^{(n-i-1)/n}} \\[2ex]
& = & \frac{1}{\sum\limits_{i=0}^{n-1} x^{(n-1)/n}} \\[2ex]
& = & \frac{1}{nx^{(n-1)/n}} \\[2ex]
& = & \frac{1}{n}x^{-(n-1)/n} \\[2ex]
& = & \frac{1}{n}x^{(1/n)-1},
\end{array}$$

and we are done. □

4.4.3 The exponential and logarithmic functions

As we saw in subsection 2.3.1, the sequence $\{(1+\frac{1}{n})^n\}_{n=1}^{\infty}$ converges, and we defined Euler's number e to be the limit of this sequence, that is,

$$e \;:=\; \lim_{n\to\infty}(1+\tfrac{1}{n})^n.$$

It can be shown (but it is not simple, and we omit the details) that we also have the following **function** limit:

$$e \;=\; \lim_{x\to\infty}(1+\tfrac{1}{x})^x.$$

In the above, x may take real values (not just integer values). In what follows, we will use the above fact without proof.

Proposition 4.4.5. *All the following hold:*

(a) $\lim\limits_{x\to 0}(1+x)^{1/x}=e$*;*

(b) $\lim\limits_{x\to 0}\frac{\ln(1+x)}{x}=1$*;*

(c) $\lim\limits_{x\to 0}\frac{e^x-1}{x}=1$*.*

Proof. (a) First, we compute:

$$\lim_{x\to 0^+}(1+x)^{1/x} \overset{(*)}{=} \lim_{y\to+\infty}(1+\tfrac{1}{y})^y = e,$$

where for (*), we used the substitution $y=\frac{1}{x}$. Next, we compute:

$$\begin{aligned}
\lim_{x\to 0^-}(1+x)^{1/x} &= \lim_{y\to+\infty}(1-\tfrac{1}{y})^{-y} && \text{via substitution } y=-1/x \\
&= \lim_{y\to+\infty}(\tfrac{y}{y-1})^y \\
&= \lim_{z\to+\infty}(\tfrac{z+1}{z})^{z+1} && \text{via substitution } z=y-1 \\
&= \lim_{z\to+\infty}(1+\tfrac{1}{z})^{z+1} \\
&= \big(\lim_{z\to+\infty}(1+\tfrac{1}{z})^z\big)\big(\lim_{z\to+\infty}(1+\tfrac{1}{z})\big) \\
&= e\cdot 1 = e.
\end{aligned}$$

We have now shown that $\lim\limits_{x\to 0^+}(1+x)^{1/x}=\lim\limits_{x\to 0^-}(1+x)^{1/x}=e$, and it follows that $\lim\limits_{x\to 0}(1+x)^{1/x}=e$.

(b) We compute:

$$\lim_{x\to 0}\frac{\ln(1+x)}{x} = \lim_{x\to 0}\ln\big((1+x)^{1/x}\big) \overset{(*)}{=} \ln\big(\lim_{x\to 0}(1+x)^{1/x}\big) \overset{(a)}{=} \ln e = 1,$$

where (*) follows from the continuity of the logarithmic functions.

(c) We compute:

$$\lim_{x\to 0}\frac{e^x-1}{x} \overset{(*)}{=} \lim_{y\to 0}\frac{y}{\ln(1+y)} = \lim_{y\to 0}\frac{1}{\frac{\ln(1+y)}{y}} = \frac{1}{\lim\limits_{y\to 0}\frac{\ln(1+y)}{y}} \overset{(b)}{=} \frac{1}{1} = 1,$$

where (*) was obtained via substitution $y=e^x-1$. □

Proposition 4.4.6. *$\frac{d}{dx}(e^x) = e^x$ for $x \in \mathbb{R}$.*

Proof. For $x \in \mathbb{R}$, we compute:

$$\frac{d}{dx}(e^x) \quad = \quad \lim_{h \to 0} \frac{e^{x+h}-e^x}{h} \quad = \quad \lim_{h \to 0} \frac{e^x(e^h-1)}{h} \quad = \quad e^x\big(\lim_{h \to 0} \frac{e^h-1}{h}\big) \quad \overset{(*)}{=} \quad e^x,$$

where (*) follows from Proposition 4.4.5(c). ☐

Proposition 4.4.7. *$\frac{d}{dx}(\ln x) = \frac{1}{x}$ for $x \in (0, +\infty)$.*

Proof. For $x \in (0, +\infty)$, we compute:

$$\begin{aligned} \frac{d}{dx}(\ln x) \quad &= \quad \lim_{h \to 0} \frac{\ln(x+h)-\ln x}{h} \\ &= \quad \lim_{h \to 0} \frac{\ln\left(\frac{x+h}{x}\right)}{h} \\ &= \quad \frac{1}{x} \lim_{h \to 0} \frac{\ln\left(1+\frac{h}{x}\right)}{\frac{h}{x}} \\ &\overset{(*)}{=} \quad \frac{1}{x} \lim_{y \to 0} \frac{\ln(1+y)}{y} \\ &\overset{(**)}{=} \quad \frac{1}{x} \cdot 1 \quad = \quad \frac{1}{x}. \end{aligned}$$

where (*) is obtained via substitution $y = \frac{h}{x}$ (with x treated as a constant and h as a variable inside the limit), whereas (**) follows from Proposition 4.4.5(b). ☐

4.4.4 The sine and cosine functions

Recall from trigonometry that

$$\sin^2 x + \cos^2 x \quad = \quad 1 \qquad \forall x \in \mathbb{R}.$$

Furthermore, we have the following addition formulas for sine and cosine:

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

for all $x, y \in \mathbb{R}$. Recall also that by Theorem 3.7.2, we have that

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

Proposition 4.4.8. $\lim\limits_{x \to 0} \frac{1-\cos x}{x} = 0.$

Proof. We compute:

$$\begin{aligned}
\lim_{x\to 0}\tfrac{1-\cos x}{x} &= \lim_{x\to 0}\tfrac{1-\cos(\frac{x}{2}+\frac{x}{2})}{x} \\
&= \lim_{x\to 0}\tfrac{1-\left(\cos^2(\frac{x}{2})-\sin^2(\frac{x}{2})\right)}{x} \\
&= \lim_{x\to 0}\tfrac{1-\cos^2(\frac{x}{2})+\sin^2(\frac{x}{2})}{x} \\
&= \lim_{x\to 0}\tfrac{2\sin^2(\frac{x}{2})}{x} \\
&= \lim_{x\to 0}\tfrac{\sin^2(\frac{x}{2})}{\frac{x}{2}} \\
&= \lim_{y\to 0}\tfrac{\sin^2 y}{y} && \text{via substitution } y=\tfrac{x}{2} \\
&= \big(\lim_{y\to 0}\sin y\big)\big(\lim_{y\to 0}\tfrac{\sin y}{y}\big) \\
&= (\sin 0)\cdot 1 && \text{by the continuity of the sine function and by Theorem 3.7.2} \\
&= 0,
\end{aligned}$$

and we are done. □

Proposition 4.4.9. *$\frac{d}{dx}(\sin x)=\cos x$ and $\frac{d}{dx}(\cos x)=-\sin x$ for $x\in\mathbb{R}$.*

Proof. **The derivative of sine.** For $x\in\mathbb{R}$, we compute:

$$\begin{aligned}
\tfrac{d}{dx}(\sin x) &= \lim_{h\to 0}\tfrac{\sin(x+h)-\sin x}{h} \\
&= \lim_{h\to 0}\tfrac{\sin x\cos h+\sin h\cos x-\sin x}{h} \\
&= \lim_{h\to 0}\tfrac{\sin h\cos x-\sin x(1-\cos h)}{h} \\
&= \lim_{h\to 0}\big(\cos x\tfrac{\sin h}{h}-\sin x\tfrac{1-\cos h}{h}\big) \\
&= (\cos x)\big(\lim_{h\to 0}\tfrac{\sin h}{h}\big)-(\sin x)\big(\lim_{h\to 0}\tfrac{1-\cos h}{h}\big) \\
&\overset{(*)}{=} (\cos x)\cdot 1-(\sin x)\cdot 0 \quad = \quad \cos x,
\end{aligned}$$

where (*) follows from Theorem 3.7.2 and Proposition 4.4.8.

The derivative of cosine. For $x \in \mathbb{R}$, we compute:

$$\begin{array}{rcl}
\frac{d}{dx}(\cos x) & = & \lim\limits_{h \to 0} \frac{\cos(x+h) - \cos x}{h} \\
& = & \lim\limits_{h \to 0} \frac{\cos x \cos h - \sin x \sin h - \cos x}{h} \\
& = & -\lim\limits_{h \to 0} \frac{-\cos x \cos h + \sin x \sin h + \cos x}{h} \\
& = & -\lim\limits_{h \to 0} \frac{\sin x \sin h + \cos x (1 - \cos h)}{h} \\
& = & -\lim\limits_{h \to 0} \big(\sin x \frac{\sin h}{h} + \cos x \frac{1 - \cos h}{h} \big) \\
& = & -(\sin x) \big(\lim\limits_{h \to 0} \frac{\sin h}{h} \big) - (\cos x) \big(\lim\limits_{h \to 0} \frac{1 - \cos h}{h} \big) \\
& \overset{(*)}{=} & -(\sin x) \cdot 1 - (\cos x) \cdot 0 \quad = \quad -\sin x,
\end{array}$$

where (*) follows from Theorem 3.7.2 and Proposition 4.4.8. □

4.5 The Product and Quotient Rules for derivatives

The Product Rule. *Let $f, g : A \subseteq \mathbb{R} \to \mathbb{R}$ be functions, let $a_0 \in A$ be an accumulation point of A, and assume that both f and g are differentiable at a_0. Then fg is also differentiable at a_0, and moreover, we have the following formula:*

$$(fg)'(a_0) \quad = \quad f'(a_0)g(a_0) + f(a_0)g'(a_0).$$

Remark: If $f, g : A \subseteq \mathbb{R} \to \mathbb{R}$ are differentiable functions (i.e. if they are differentiable at all points in their domain), then the Product Rule states that fg is also differentiable, and it gives us the formula $(fg)' = f'g + fg'$.

Proof. First of all, since f and g are differentiable at a_0, Theorem 4.3.1 implies that they are continuous at a_0, and consequently,

$$\lim_{a \to a_0} f(a) = f(a_0) \qquad \text{and} \qquad \lim_{a \to a_0} g(a) = g(a_0).$$

We now compute:

$$(fg)'(a_0) \quad = \quad \lim_{a \to a_0} \frac{(fg)(a) - (fg)(a_0)}{a - a_0}$$

$$\begin{array}{rl}
= & \lim\limits_{a\to a_0} \frac{f(a)g(a)-f(a_0)g(a_0)}{a-a_0} \\[2ex]
= & \lim\limits_{a\to a_0} \frac{f(a)g(a)-f(a_0)g(a)+f(a_0)g(a)-f(a_0)g(a_0)}{a-a_0} \\[2ex]
= & \lim\limits_{a\to a_0} \frac{\big(f(a)-f(a_0)\big)g(a)+f(a_0)\big(g(a)-g(a_0)\big)}{a-a_0} \\[2ex]
= & \lim\limits_{a\to a_0} \big(\frac{f(a)-f(a_0)}{a-a_0}g(a)\big) + \lim\limits_{a\to a_0} \big(f(a_0)\frac{g(a)-g(a_0)}{a-a_0}\big) \\[2ex]
= & \big(\lim\limits_{a\to a_0} \frac{f(a)-f(a_0)}{a-a_0}\big)\big(\lim\limits_{a\to a_0} g(a)\big) + f(a_0)\lim\limits_{a\to a_0} \big(\frac{g(a)-g(a_0)}{a-a_0}\big) \\[2ex]
= & f'(a_0)\big(\lim\limits_{a\to a_0} g(a)\big) + f(a_0)g'(a_0) \\[2ex]
\overset{(*)}{=} & f'(a_0)g(a_0) + f(a_0)g'(a_0),
\end{array}$$

where (*) follows from the continuity of g at a_0. $\square$

Example 4.5.1. *Consider the function* $f : (0, +\infty) \to \mathbb{R}$ *given by* $f(x) = x^5 \ln x$ *for all* $x \in (0, +\infty)$*. Then, for* $x \in \mathbb{R}$*, we compute:*

$$f'(x) \;=\; \frac{d}{dx}(x^5 \ln x) \;\overset{(*)}{=}\; (5x^4)\ln x + x^5 \frac{1}{x} \;=\; 5x^4 \ln x + x^4,$$

where (*) was obtained via the Product Rule.

The Quotient Rule. *Let* $f, g : A \subseteq \mathbb{R} \to \mathbb{R}$ *be functions, and let* $a_0 \in A$ *be an accumulation point of* A*. Assume that both* f *and* g *are differentiable at* a_0 *and that* $g(a_0) \neq 0$*. Then* $\frac{f}{g}$ *is also differentiable at* a_0*, and moreover, we have the following formula:*

$$\big(\tfrac{f}{g}\big)'(a_0) \;=\; \frac{f'(a_0)g(a_0)-f(a_0)g'(a_0)}{(g(a_0))^2}.$$

Proof. We begin with a technical claim.

> **Claim.** The function $\frac{1}{g}$ is differentiable at a_0, and moreover, we have the following formula:
>
> $$\big(\tfrac{1}{g}\big)'(a_0) \;=\; -\frac{g'(a_0)}{(g(a_0))^2}.$$

Proof of the Claim. First, since g is differentiable at a_0, Theorem 4.3.1 guarantees that g is also continuous at a_0, and consequently,

$$\lim_{a\to a_0} g(a) \;=\; g(a_0).$$

We now compute:

$$\begin{array}{rcl}
\big(\frac{1}{g}\big)'(a_0) & = & \lim\limits_{a\to a_0} \frac{\frac{1}{g(a)}-\frac{1}{g(a_0)}}{a-a_0} \\
& = & \lim\limits_{a\to a_0} \frac{g(a_0)-g(a)}{\big(g(a)g(a_0)\big)(a-a_0)} \\
& = & \big(\lim\limits_{a\to a_0} \frac{1}{g(a)g(a_0)}\big)\big(\lim\limits_{a\to a_0} \frac{g(a_0)-g(a)}{a-a_0}\big) \\
& = & \frac{1}{(\lim\limits_{a\to a_0} g(a))g(a_0)}\big(-\lim\limits_{a\to a_0} \frac{g(a)-g(a_0)}{a-a_0}\big) \\
& = & \frac{1}{(g(a_0))^2}(-g'(a_0)) \\
& = & -\frac{g'(a_0)}{(g(a_0))^2}.
\end{array}$$

This proves the Claim. ♦

The result now readily follows from the Product Rule and the above Claim, applied to the functions f and $\frac{1}{g}$. Indeed, we compute:

$$\begin{array}{rcll}
\big(\frac{f}{g}\big)'(a_0) & = & (f\cdot\frac{1}{g})'(a_0) & \\
& = & f'(a_0)\big((\frac{1}{g})(a_0)\big)+f(a_0)\big((\frac{1}{g})'(a_0)\big) & \text{by the Product Rule} \\
& = & \frac{f'(a_0)}{g(a_0)}+f(a_0)\big(-\frac{g'(a_0)}{(g(a_0))^2}\big) & \text{by the Claim} \\
& = & \frac{f'(a_0)g(a_0)-f(a_0)g'(a_0)}{(g(a_0))^2}. &
\end{array}$$

This completes the argument. □

Proposition 4.5.2. $\frac{d}{dx}(\tan x)=\frac{1}{\cos^2 x}$ *for* $x\in\mathbb{R}\setminus\big\{\frac{2k+1}{2}\pi \mid k\in\mathbb{Z}\big\}$.

Proof. First, recall that by Proposition 4.4.9, we have that

$$\frac{d}{dx}(\sin x)=\cos x \qquad \text{and} \qquad \frac{d}{dx}(\cos x)=-\sin x$$

for $x\in\mathbb{R}$. Now, for $x\in\mathbb{R}\setminus\{\frac{2k+1}{2}\pi\mid k\in\mathbb{Z}\}$, we compute:

$$\begin{array}{rcll}
\frac{d}{dx}(\tan x) & = & \big(\frac{\sin x}{\cos x}\big)' & \\
& = & \frac{(\sin x)'\cdot(\cos x)-(\sin x)\cdot(\cos x)'}{\cos^2 x} & \text{by the Quotient Rule} \\
& = & \frac{(\cos x)\cdot(\cos x)-(\sin x)\cdot(-\sin x)}{\cos^2 x} &
\end{array}$$

$$
\begin{aligned}
&= \tfrac{\cos^2 x+\sin^2 x}{\cos^2 x} \\
&= \tfrac{1}{\cos^2 x}.
\end{aligned}
$$

□

4.6 The Chain Rule

The Chain Rule. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ and $g : B \subseteq \mathbb{R} \to \mathbb{R}$ be functions such that $Im(f) \subseteq B$ (so that $g \circ f : A \to \mathbb{R}$ is defined). Assume that f is differentiable at a point $a_0 \in A$, and that g is differentiable at the point $b_0 := f(a_0)$. (In particular, a_0 belongs to and is an accumulation point of A, whereas b_0 belongs to and is an accumulation point of B.) Then $g \circ f$ is differentiable at a_0, and moreover, we have that*

$$(g \circ f)'(a_0) \quad = \quad g'(b_0)f'(a_0).$$

Proof. First of all, by Theorem 4.3.1, the differentiability of f at a_0 implies the continuity of f at a_0, and consequently (since a_0 is an accumulation point of A), we have that

$$\lim_{a \to a_0} f(a) \quad = \quad f(a_0).$$

Now, we define an auxiliary "error" function $\eta : B \to \mathbb{R}$ by setting

$$\eta(b) \quad = \quad \begin{cases} \frac{g(b)-g(b_0)}{b-b_0} - g'(b_0) & \text{if} \quad b \neq b_0 \\ 0 & \text{if} \quad b = b_0 \end{cases}$$

for all $b \in B$.

Claim 1. $\lim\limits_{a \to a_0} \eta(f(a)) = 0$.

Proof of Claim 1. Clearly,

$$
\begin{aligned}
\lim_{b \to b_0} \eta(b) \quad &= \quad \lim_{b \to b_0} \left(\tfrac{g(b)-g(b_0)}{b-b_0} - g'(b_0)\right) \\
&= \quad \Big(\lim_{b \to b_0} \tfrac{g(b)-g(b_0)}{b-b_0}\Big) - g'(b_0) \\
&= \quad g'(b_0) - g'(b_0) \quad = \quad 0 \quad = \quad \eta(b_0),
\end{aligned}
$$

and so η is continuous at $b_0 = f(a_0)$. Now, recall that f is continuous at a_0, and that $\lim\limits_{a \to a_0} f(a) = f(a_0)$. We deduce that

$$\lim_{a \to a_0} \eta(f(a)) \quad \overset{(*)}{=} \quad \eta\Big(\lim_{a \to a_0} f(a)\Big) \quad = \quad \eta(f(a_0)) \quad = \quad \eta(b_0) \quad = \quad 0,$$

where $(*)$ follows from Theorem 3.5.3, since η is continuous at $\lim\limits_{a \to a_0} f(a) = f(a_0)$. ♦

Claim 2. For all $b \in B$, we have that

$$g(b) - g(b_0) \quad = \quad \big(g'(b_0) + \eta(b)\big)(b - b_0).$$

Proof of Claim 2. For $b \in B \setminus \{b_0\}$, the equality follows immediately from the definition of the function η, whereas for $b = b_0$, both sides of the equality are equal to zero. ♦

Claim 3. For all $a \in A \setminus \{a_0\}$, we have that:

$$\tfrac{(g \circ f)(a) - (g \circ f)(a_0)}{a - a_0} \quad = \quad \big(g'(b_0) + \eta(f(a))\big) \tfrac{f(a) - f(a_0)}{a - a_0}.$$

Proof of Claim 3. Fix $a \in A \setminus \{a_0\}$, and set $b := f(a)$. Then

$$\begin{aligned}
\tfrac{(g \circ f)(a) - (g \circ f)(a_0)}{a - a_0} &= \tfrac{g(f(a)) - g(f(a_0))}{a - a_0} \\
&= \tfrac{g(b) - g(b_0)}{a - a_0} \\
&= \tfrac{\big(g'(b_0) + \eta(b)\big)(b - b_0)}{a - a_0} && \text{by Claim 2} \\
&= \big(g'(b_0) + \eta(f(a))\big) \tfrac{f(a) - f(a_0)}{a - a_0},
\end{aligned}$$

which proves the claim. ♦

We now compute:

$$\begin{aligned}
(g \circ f)'(a_0) &= \lim_{a \to a_0} \tfrac{(g \circ f)(a) - (g \circ f)(a_0)}{a - a_0} \\
&= \lim_{a \to a_0} \Big(\big(g'(b_0) + \eta(f(a))\big) \tfrac{f(a) - f(a_0)}{a - a_0}\Big) && \text{by Claim 3} \\
&= \Big(g'(b_0) + \lim_{a \to a_0} \eta(f(a))\Big)\Big(\lim_{a \to a_0} \tfrac{f(a) - f(a_0)}{a - a_0}\Big) \\
&= \Big(g'(b_0) + 0\Big) f'(a_0) && \text{by Claim 1} \\
&= g'(b_0) f'(a_0).
\end{aligned}$$

This completes the argument. □

Example 4.6.1. *Compute* $\frac{d}{dx}\big((\sin x)^{100}\big)$.

Solution. Set $h(x) = (\sin x)^{100}$, so that $h = g \circ f$, where $f(x) = \sin x$ and $g(x) = x^{100}$. We know that $f'(x) = \cos x$ (by Proposition 4.4.9) and $g'(x) = 100x^{99}$ (by Proposition 4.4.2). Therefore,

$$\begin{aligned} \tfrac{d}{dx}\big((\sin x)^{100}\big) &= h'(x) \\ &= g'(f(x))f'(x) \\ &= 100(\sin^{99} x)(\cos x). \end{aligned}$$

Remark: The idea is that we are differentiating the function $\Big(\boxed{(\sin x)}\Big)^{100}$. For $\square = \boxed{(\sin x)}$, we first differentiate $\square^{100}$ to obtain $100\square^{99} = \sin^{99} x$, and then we differentiate $\square$ to obtain $\cos x$; we multiply the two to obtain $100(\sin^{99} x)(\cos x)$. □

Example 4.6.2. *Compute* $\frac{d}{dx}(\sin^3 x + 2)^2$.

Solution. By repeatedly applying the Chain Rule, we obtain:

$$\begin{aligned} \tfrac{d}{dx}(\sin^3 x + 2)^2 &= 2\big(\sin^3 x + 2\big) \cdot \tfrac{d}{dx}(\sin^3 x + 2) \\ &= 2\big(\sin^3 x + 2\big)\big(\tfrac{d}{dx}(\sin^3 x) + \tfrac{d}{dx}(2)\big) \\ &= 2\big(\sin^3 x + 2\big)\big(3\sin^2 x \tfrac{d}{dx}(\sin x) + 0\big) \\ &= 2(\sin^3 x + 2)\big(3(\sin^2 x)(\cos x)\big) \\ &= 6(\sin^2 x)(\cos x)(\sin^3 x + 2). \end{aligned}$$

□

Remark: When differentiating, we often use the following trick:

$$\square^{\Delta} = e^{\ln\left(\square^{\Delta}\right)} = e^{\Delta \ln \square}.$$

This is useful for differentiating via the Chain Rule, further using the fact that $\frac{d}{dx}e^x = e^x$ (by Proposition 4.4.6).

Proposition 4.6.3. *Let* $a \in (0,1) \cup (1,+\infty)$ *be a fixed real number. Then:*

(a) $\frac{d}{dx}a^x = a^x \ln a$ *for* $x \in \mathbb{R}$*;*

(b) $\frac{d}{dx}\log_a(x) = \frac{1}{x \ln a}$ *for* $x \in (0,+\infty)$.

Proof. First, recall that $\frac{d}{dx}e^x = e^x$ for $x \in \mathbb{R}$ (by Proposition 4.4.6) and $\frac{d}{dx}\ln x = \frac{1}{x}$ for $x \in (0, +\infty)$ (by Proposition 4.4.7).

For (a), we compute (for $x \in \mathbb{R}$):

$$\begin{array}{rcll} \frac{d}{dx}a^x & = & \frac{d}{dx}\big(e^{\ln(a^x)}\big) & \\ & = & \frac{d}{dx}\big(e^{x\ln a}\big) & \\ & = & e^{x\ln a}\frac{d}{dx}(x\ln a) & \text{by the Chain Rule} \\ & = & e^{\ln(a^x)}\ln a & \\ & = & a^x \ln a. & \end{array}$$

For (b), we compute (for $x \in (0, +\infty)$):

$$\frac{d}{dx}\log_a(x) \;=\; \frac{d}{dx}\Big(\frac{\ln x}{\ln a}\Big) \;=\; \frac{1}{\ln a}\frac{d}{dx}(\ln x) \;=\; \frac{1}{\ln a}\cdot\frac{1}{x} \;=\; \frac{1}{x\ln a}.$$

□

Proposition 4.6.4. *Let $\alpha \in \mathbb{R}$ be a fixed constant. Then $\frac{d}{dx}(x^\alpha) = \alpha x^{\alpha-1}$ for $x \in (0, +\infty)$.*

Proof. For $x \in (0, +\infty)$, we compute:

$$\begin{array}{rcll} \frac{d}{dx}(x^\alpha) & = & \frac{d}{dx}\big(e^{\ln(x^\alpha)}\big) & \\ & = & \frac{d}{dx}\big(e^{\alpha\ln x}\big) & \\ & = & e^{\alpha\ln x}\frac{d}{dx}(\alpha\ln x) & \text{by the Chain Rule} \\ & = & e^{\ln(x^\alpha)}\cdot\frac{\alpha}{x} & \\ & = & x^\alpha\cdot\frac{\alpha}{x} & \\ & = & \alpha x^{\alpha-1}, & \end{array}$$

and we are done. □

Example 4.6.5. *Compute $\frac{d}{dx}\big(x^x\big)$ for $x \in (0, +\infty)$.*

Proof. For $x \in (0, +\infty)$, we compute:

$$\begin{array}{rcll} \frac{d}{dx}\big(x^x\big) & = & \frac{d}{dx}\big(e^{\ln(x^x)}\big) & \\ & = & \frac{d}{dx}\big(e^{x \ln x}\big) & \\ & = & e^{x \ln x} \cdot \frac{d}{dx}(x \ln x) & \text{by the Chain Rule} \\ & = & x^x \cdot \frac{d}{dx}(x \ln x) & \\ & = & x^x(\ln x + x \cdot \frac{1}{x}) & \text{by the Product Rule} \\ & = & x^x(1 + \ln x). & \end{array}$$

□

4.7 The Inverse Function Theorem

Suppose that $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is one-to-one and differentiable at a point $a_0 \in A$, and that f^{-1} is differentiable at the point $b_0 := f(a_0)$. (In particular, a_0 belongs to and is an accumulation point of A, whereas b_0 belongs to and is an accumulation point of $\text{Im}(f)$.) Then by the Chain Rule, we have that

$$(f^{-1} \circ f)'(a_0) \quad = \quad (f^{-1})'\big(f(a_0)\big)f'(a_0) \quad = \quad (f^{-1})'(b_0)f'(a_0)$$

But note that $f^{-1} \circ f = \text{Id}_A$, that is, $(f^{-1} \circ f) : A \to A$ is given by $(f^{-1} \circ f)(x) = x$ for all $x \in A$. Therefore, $(f^{-1} \circ f)'(x) = 1$ for all $x \in A$ such that x is an accumulation point of A. So, the computation above yields $1 = (f^{-1})'(b_0)f'(a_0)$. Consequently, both $f'(a_0)$ and $(f^{-1})'(b_0)$ are non-zero, and moreover, we have that

$$(f^{-1})'(b_0) \quad = \quad \tfrac{1}{f'(a_0)}.$$

The Inverse Function Theorem (see below) weakens the assumption for f^{-1} from the above in the sense that it does not assume that f^{-1} is differentiable at $b_0 := f(a_0)$. Instead, it strengthens the assumption for f in a suitable way, and it concludes (rather than assuming) that f^{-1} is differentiable at b_0, with the same formula for $(f^{-1})'(b_0)$ as the one that we obtained above.

The Inverse Function Theorem. *Let I be an open interval in $\mathbb{R}$, and let $f : I \to \mathbb{R}$ be a one-to-one function that is differentiable on I. Let $a_0 \in I$, set $b_0 := f(a_0)$, and assume that $f'(a_0) \neq 0$. Then f^{-1} is differentiable at b_0, and moreover,*

$$(f^{-1})'(b_0) \quad = \quad \tfrac{1}{f'(a_0)}.$$

Proof. First of all, since f is one-to-one, we know that for any $a \in I \setminus \{a_0\}$, we have that $f(a) \neq f(a_0)$, and consequently, $f(a) - f(a_0) \neq 0$. We will implicitly use this fact in our computation below (ensuring that we are not dividing by zero, so that our terms are defined).

Since f is differentiable on I, Theorem 4.3.1 guarantees that f is continuous on I. Moreover, Proposition 3.5.11 guarantees that $J := \text{Im}(f)$ is an open interval (because the domain I of f is an open interval), and Theorem 3.5.12 guarantees that $f^{-1} : J \to \mathbb{R}$ is continuous. We now compute:

$$\begin{aligned}
(f^{-1})'(b_0) &= \lim_{b \to b_0} \tfrac{f^{-1}(b) - f^{-1}(b_0)}{b - b_0} \\
&\overset{(*)}{=} \lim_{a \to a_0} \tfrac{a - a_0}{f(a) - f(a_0)} \\
&= \lim_{a \to a_0} \tfrac{1}{\frac{f(a) - f(a_0)}{a - a_0}} \\
&\overset{(**)}{=} \tfrac{1}{\lim\limits_{a \to a_0} \frac{f(a) - f(a_0)}{a - a_0}} \\
&= \tfrac{1}{f'(a_0)}.
\end{aligned}$$

For (*), we used the substitution $a = f^{-1}(b)$, so that $b = f(a)$. This is valid by the continuity of f^{-1} at b_0 and the continuity of f at a_0, ensuring that

$$b \to b_0 \qquad \Longleftrightarrow \qquad a = f^{-1}(b) \to f^{-1}(b_0) = a_0.$$

Meanwhile, (**) uses the fact that $\lim\limits_{a \to a_0} \frac{f(a) - f(a_0)}{a - a_0} = f'(a_0) \neq 0$, which in particular means that we are not dividing by zero. $\square$

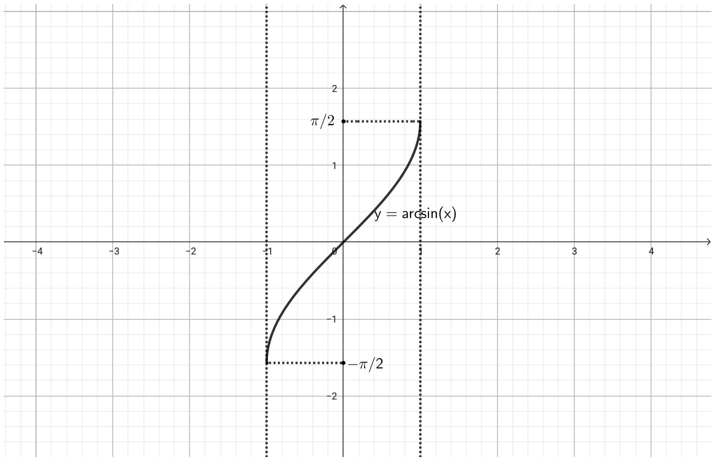
4.7.1 Inverse trigonometric functions

By Propositions 4.4.9 and 4.5.2, we have the following formulas:

- $\frac{d}{dx} \sin x = \cos x$ and $\frac{d}{dx} \cos x = -\sin x$ for $x \in \mathbb{R}$;
- $\frac{d}{dx} \tan x = \frac{1}{\cos^2 x}$ for $x \in \mathbb{R} \setminus \left\{ \frac{2k+1}{2}\pi \mid k \in \mathbb{Z} \right\}$.

Here, we use the formulas above, as well as the Inverse Function Theorem, to compute the derivatives of the inverse trigonometric functions $\arcsin x$ and $\arctan x$. (The case of $\arccos x$ is left as an exercise.)

The arcsine function. The sine function is not one-to-one, and therefore, it does not have an inverse. However, we can turn it into a one-to-one function by restricting its domain to $[-\frac{\pi}{2}, \frac{\pi}{2}]$. The arcsine function is simply the inverse function of the sine function restricted to $[-\frac{\pi}{2}, \frac{\pi}{2}]$. The domain of the arcsine function is $[-1, 1]$, its image is $[-\frac{\pi}{2}, \frac{\pi}{2}]$, and its graph is represented below.



Proposition 4.7.1. *For all* $x \in [-1,1]$*, we have that* $\sin(\arcsin x) = x$ *and* $\cos(\arcsin x) = \sqrt{1-x^2}$.

Proof. Fix $x \in [-1,1]$. The fact that $\sin(\arcsin x) = x$ follows immediately from the definition of the arcsine function. Since $\sin^2 y + \cos^2 y = 1$ for all $y \in \mathbb{R}$, it follows that

$$\cos^2(\arcsin x) \quad = \quad 1 - \sin^2(\arcsin x) \quad = \quad 1 - x^2.$$

But note that $\arcsin x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, and consequently, $\cos(\arcsin x) \geq 0$. Therefore,

$$\cos(\arcsin x) \quad = \quad \sqrt{1-x^2},$$

and we are done. □

Remark: For $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, we do in fact have that $\arcsin(\sin x) = x$. However, this equality does **not** hold for $x \in \mathbb{R} \setminus [-\frac{\pi}{2}, \frac{\pi}{2}]$. For example, note that $\arcsin(\sin \pi) = \arcsin(0) = 0 \neq \pi$.

Proposition 4.7.2. $\frac{d}{dx} \arcsin x = \frac{1}{\sqrt{1-x^2}}$ *for* $x \in (-1,1)$.

Proof. Recall that $\frac{d}{dx} \sin x = \cos x$ for $x \in \mathbb{R}$ (by Proposition 4.4.9). Note the function $\sin \upharpoonright (-\frac{\pi}{2}, \frac{\pi}{2}) : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ is one-to-one and differentiable on the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$. Now, fix any $y_0 \in (-1,1)$, and set $x_0 := \arcsin y_0$, so that $y_0 = \sin x_0$. Then $x_0 \in (-\frac{\pi}{2}, \frac{\pi}{2})$, and so $\sin'(x_0) = \cos x_0 \neq 0$. We now compute:

$$\arcsin'(y_0) \quad \overset{(*)}{=} \quad \frac{1}{\sin'(x_0)} \quad = \quad \frac{1}{\cos(x_0)} \quad = \quad \frac{1}{\cos(\arcsin y_0)} \quad \overset{(**)}{=} \quad \frac{1}{\sqrt{1-y_0^2}},$$

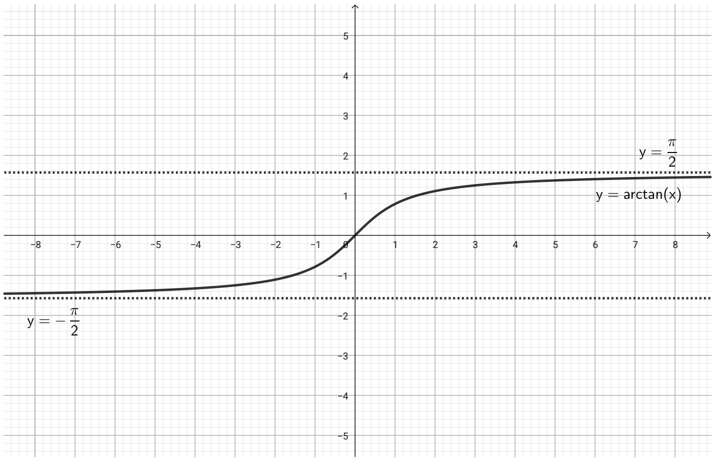
where (*) follows from the Inverse Function Theorem, whereas (**) follows from Proposition 4.7.1. □

Remark: Note that $\arcsin x$ is **not** differentiable at $x = \pm 1$. Indeed, for any $x_0 \in [-1, 1]$ for which the arcsine function is differentiable, we have that

$$\begin{array}{rcll} 1 & = & \frac{d}{dx} x\big|_{x=x_0} & \\ & = & (\sin \circ \arcsin)'(x_0) & \\ & = & \big(\cos(\arcsin x_0)\big)(\arcsin' x_0) & \text{by the Chain Rule} \\ & = & \sqrt{1-x_0^2}\,(\arcsin' x_0) & \text{by Proposition 4.7.1,} \end{array}$$

which implies that $\sqrt{1-x_0^2} \neq 0$, and consequently, $x_0 \neq \pm 1$. (Note that the computation above essentially reproduces the argument from the beginning of the section.)

The arctangent function. The tangent function is not one-to-one, and therefore, it does not have an inverse. However, we can turn it into a one-to-one function by restricting its domain to $(-\frac{\pi}{2}, \frac{\pi}{2})$. The arctangent function is simply the inverse function of the tangent function restricted to $(-\frac{\pi}{2}, \frac{\pi}{2})$. The domain of the arctagent function is $\mathbb{R}$, its range is $(-\frac{\pi}{2}, \frac{\pi}{2})$, and its graph is represented below.



Remark: By the definition of the arctangent function, we have that $\tan(\arctan x) = x$ for all $x \in \mathbb{R}$. Meanwhile, for $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$, we have that $\arctan(\tan x) = x$. However, this last equality fails for values of $x \notin (-\frac{\pi}{2}, \frac{\pi}{2})$. For example, $\arctan(\tan \pi) = \arctan 0 = 0 \neq \pi$.

Proposition 4.7.3. $\frac{d}{dx} \arctan x = \frac{1}{1+x^2}$ *for* $x \in \mathbb{R}$.

Proof. First, we know that the function $\tan \restriction (-\frac{\pi}{2}, \frac{\pi}{2}) : (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ is one-to-one and differentiable on the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$. Now, fix $y_0 \in \mathbb{R}$, and set $x_0 := \arctan y_0$, so that $y_0 = \tan x_0$. By Proposition 4.5.2, we have that $\tan'(x_0) = \frac{1}{\cos^2 x_0} \neq 0$. We now compute:

$$\arctan'(y_0) \overset{(*)}{=} \frac{1}{\tan' x_0} = \frac{1}{\frac{1}{\cos^2 x_0}} \overset{(**)}{=} \frac{1}{1+\tan^2 x_0} = \frac{1}{1+y_0^2},$$

where (*) follows from the Inverse Function Theorem, whereas (**) follows from the following computation:

$$1+\tan^2 x_0 = 1+\frac{\sin^2 x_0}{\cos^2 x_0} = \frac{\cos^2 x_0+\sin^2 x_0}{\cos^2 x_0} = \frac{1}{\cos^2 x_0}.$$

$\square$

4.8 A summary of important derivatives

Putting together the formulas that we obtained in the previous few sections, we obtain the table below. These formulas should be memorized.

function $f(x)$	derivative $f'(x)$	differentiable for
$c = \text{const.}$	0	$x \in \mathbb{R}$
x^α (for a fixed constant $\alpha \in \mathbb{R}$)	$\alpha x^{\alpha-1}$	$x > 0$
e^x	e^x	$x \in \mathbb{R}$
$\ln x$	$\frac{1}{x}$	$x > 0$
$\sin x$	$\cos x$	$x \in \mathbb{R}$
$\cos x$	$-\sin x$	$x \in \mathbb{R}$
$\tan x$	$\frac{1}{\cos^2 x}$	$x \neq \frac{2k+1}{2}\pi,\ k \in \mathbb{Z}$
$\arcsin x$	$\frac{1}{\sqrt{1-x^2}}$	$x \in (-1,1)$
$\arctan x$	$\frac{1}{1+x^2}$	$x \in \mathbb{R}$

Remark: As the table above shows, for a fixed constant $\alpha \in \mathbb{R}$, the formula $\frac{d}{dx}(x^\alpha) = \alpha x^{\alpha-1}$ is valid for $x \in (0,+\infty)$. However:

- for the special case when α is a positive integer, the formula $\frac{d}{dx}(x^\alpha) = \alpha x^{\alpha-1}$ is valid for $x \in \mathbb{R}$;
- for the special case when $\alpha = \frac{1}{n}$ for some **odd** $n \in \mathbb{N}$, the formula $\frac{d}{dx}(x^\alpha) = \alpha x^{\alpha-1}$ is valid for $x \in \mathbb{R} \setminus \{0\}$;
- for the special case when α is a negative integer or when $\alpha = -\frac{1}{n}$ for some **odd** $n \in \mathbb{N}$, the formula $\frac{d}{dx}(x^\alpha) = \alpha x^{\alpha-1}$ is valid for $x \in \mathbb{R} \setminus \{0\}$.

Indeed, the first two bullet points above follow from Propositions 4.4.2 and 4.4.4(a), respectively, and the third bullet point follows from an argument similar to the one given in those two propositions.

Remark: Consider $a \in (0,1) \cup (1,+\infty)$. In addition to the formulas from the table above, you should either memorize the following formals (from Proposition 4.6.3), or be able to derive them very quickly:

- $\frac{d}{dx}(a^x) = a^x \ln a$ for $x \in \mathbb{R}$;
- $\frac{d}{dx}(\log_a x) = \frac{1}{x \ln a}$ for $x \in (0,+\infty)$.

4.9 Implicit differentiation

Sometimes a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is given "implicitly," without an explicit formula for $f(x)$. For instance, a function may be given by an equation such as

$$x^2 + y^2 = 4,$$

or

$$x^3 + y^3 = 6xy,$$

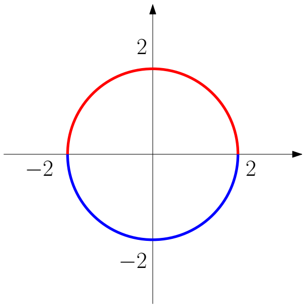
where we think of y as a function of x. Technically, an equation of the type given above might define more than one function.[2] For instance, the equation

$$x^2 + y^2 = 4$$

defines the circle of radius 2, centered at the origin, and it in fact defines two distinct functions:

- $y = \sqrt{4 - x^2}$ for $x \in [-2,2]$;
- $y = -\sqrt{4 - x^2}$ for $x \in [-2,2]$.

The graph of the first function is given by the red part of the circle below, whereas the graph of the second function is given by the blue part of the circle below.



---

[2] It is also possible that the equation does not define **any** function (with a non-empty domain). For example, this is the case for the equation $x^2 + y^2 = -1$. However, such equations are not interesting.

Sometimes, it may be very difficult (or even impossible) to find an actual formula for y in terms of x. Luckily, it is not necessary to find such a formula in order to compute the derivative of our implicitly defined function. Instead, the idea is to differentiate both sides of the equation, and then solve for $\frac{dy}{dx}$ in terms of x and y. Here, is it important to remember that y is a function of x, and so we must not forget to apply the Chain Rule. For instance, we have $\frac{d}{dx}(y^7) = 7y^6\frac{dy}{dx}$ and $\frac{d}{dx}(\sin y) = (\cos y)\frac{dy}{dx}$.

Example 4.9.1. *Compute* $\frac{dy}{dx}$ *for* $x^2 + y^2 = 4$. *Then, compute the equation of the tangent to the curve defined by this equation passing through the point* $(1, \sqrt{3})$.

Solution. We differentiate both sides of the equation to obtain

$$\tfrac{d}{dx}(x^2 + y^2) \quad = \quad \tfrac{d}{dx}(4),$$

which yields

$$2x + 2y\tfrac{dy}{dx} \quad = \quad 0.$$

Solving for $\frac{dy}{dx}$, we obtain

$$\tfrac{dy}{dx} \quad = \quad -\tfrac{x}{y}.$$

For $(x_0, y_0) = (1, \sqrt{3})$, we get

$$\tfrac{dy}{dx}\Big|_{(x,y)=(1,\sqrt{3})} \quad = \quad -\tfrac{1}{\sqrt{3}}.$$

So, the tangent to our curve through the point $(x_0, y_0) = (1, \sqrt{3})$ has slope $-\sqrt{3}$, and therefore, the equation of this tangent is

$$y \quad = \quad -\tfrac{1}{\sqrt{3}}(x - 1) + \sqrt{3},$$

or equivalently,

$$y \quad = \quad -\tfrac{x}{\sqrt{3}} + \sqrt{3} + \tfrac{1}{\sqrt{3}}.$$

□

Example 4.9.2. *Compute* $\frac{dy}{dx}$ *for* $x^3 + y^3 = 6xy$. *Then, compute the tangent to this curve through the point* $(x_0, y_0) = (3, 3)$.

Solution. We differentiate both sides of the equation:

$$\tfrac{d}{dx}(x^3 + y^3) \quad = \quad \tfrac{d}{dx}(6xy),$$

which yields

$$3x^2 + 3y^2\tfrac{dy}{dx} \quad = \quad 6y + 6x\tfrac{dy}{dx}$$

Dividing both sides by 3, we obtain

$$x^2 + y^2\tfrac{dy}{dx} \quad = \quad 2y + 2x\tfrac{dy}{dx}.$$

This implies that

$$y^2\frac{dy}{dx} - 2x\frac{dy}{dx} = 2y - x^2,$$

which yields

$$\frac{dy}{dx} = \frac{2y-x^2}{y^2-2x}.$$

For $(x_0, y_0) = (3,3)$, we obtain

$$\frac{dy}{dx}\Big|_{(x,y)=(3,3)} = -1.$$

Therefore, the equation of the tangent to our curve through the point $(x_0, y_0) = (3,3)$ is

$$y = -(x-3)+3,$$

that is,

$$y = -x+6.$$

□

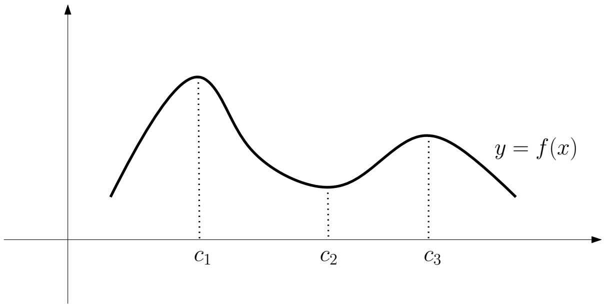
4.10 Rolle's Theorem and the Mean Value Theorems

4.10.1 Local extrema

Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ and $c \in A$.

- We say that f reaches a *local maximum* at c if there exists some $\delta > 0$ such that for all $x \in A$, if $|x-c| < \delta$, then $f(c) \geq f(x)$.
- We say that f reaches a *local minimum* at c if there exists some $\delta > 0$ such that for all $x \in A$, if $|x-c| < \delta$, then $f(c) \leq f(x)$.
- We say that f reaches a *local extremum* at c if f reaches a local maximum or a local minimum at c.

For instance, in the picture below, the function f reaches local maxima at c_1 and c_3, and it reaches a local minimum at c_2.



Example 4.10.1. *The function $f(x) = \sin x$ reaches a local maximum at each point of the form $x = \frac{\pi}{2} + 2k\pi$ with $k \in \mathbb{Z}$ (for such x, we have that $f(x) = 1$), and it reaches a local minimum at each point of the form $x = -\frac{\pi}{2} + 2k\pi$ with $k \in \mathbb{Z}$ (for such x, we have that $f(x) = -1$).*

Theorem 4.10.2. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, and assume that f reaches a local extremum at a point $c_0 \in I \subseteq A$, where I is an open interval. If f is differentiable at c_0, then $f'(c_0) = 0$.*

Proof. We prove the theorem for the case when f reaches a local maximum at c_0; the proof for the case when f reaches a local minimum at c_0 is analogous.

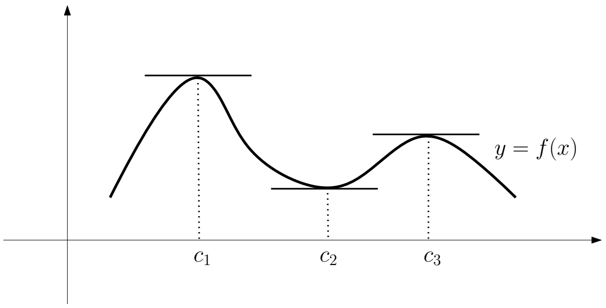
So, let us assume f reaches a local maximum at c_0, and that f is differentiable at c_0. Fix $\delta > 0$ such that for all $c \in A$, if $|c - c_0| < \delta$, then $f(c_0) \geq f(c)$. We may assume that $(c_0 - \delta, c_0 + \delta) \subseteq I \subseteq A$, for otherwise, we simply choose a smaller δ. Therefore, we in fact have that for all $c \in \mathbb{R}$, if $|c - c_0| < \delta$, then $f(c_0) \geq f(c)$.

Now, by definition, $f'(c_0) = \lim\limits_{c \to c_0} \frac{f(c)-f(c_0)}{c-c_0}$. Since f is defined on an open interval containing c_0, we deduce that

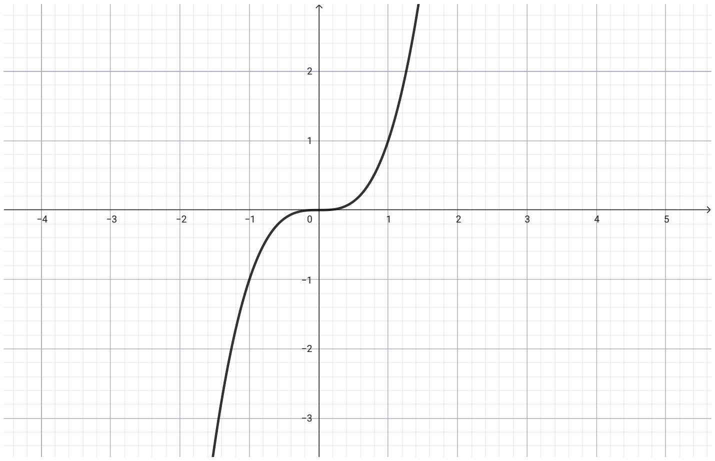
$$f'(c_0) = \lim_{c \to c_0^-} \tfrac{f(c)-f(c_0)}{c-c_0} \qquad \text{and} \qquad f'(c_0) = \lim_{c \to c_0^+} \tfrac{f(c)-f(c_0)}{c-c_0}.$$

Now, for all $c \in (c_0 - \delta, c_0)$, we have that $c < c_0$ and $f(c) \leq f(c_0)$, which implies that $\frac{f(c)-f(c_0)}{c-c_0} \geq 0$, and consequently, $f'(c_0) = \lim\limits_{c \to c_0^-} \frac{f(c)-f(c_0)}{c-c_0} \geq 0$. On the other hand, for all $c \in (c_0, c_0 + \delta)$, we have that $c > c_0$ and $f(c) \leq f(c_0)$, which implies that $\frac{f(c)-f(c_0)}{c-c_0} \leq 0$, and consequently, $f'(c_0) = \lim\limits_{c \to c_0^-} \frac{f(c)-f(c_0)}{c-c_0} \leq 0$. We have now shown that $f'(c_0) \geq 0$ and $f'(c_0) \leq 0$, and it follows that $f'(c_0) = 0$. □

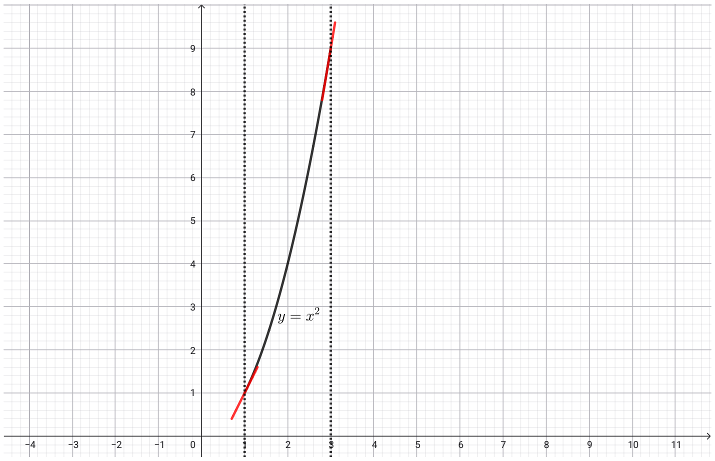
Remark: Geometrically, Theorem 4.10.2 can be interpreted as follows: if a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ reaches a local extremum at a point $c \in I \subseteq A$, where I is some open interval, and f is also differentiable at c, then the tangent to the curve $y = f(x)$ through the point $(c, f(c))$ is a horizontal line.



Remark: The converse of Theorem 4.10.2 is false. Indeed, it is possible for the function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ to be differentiable at a point $c \in I \subseteq A$, where I is some open interval, and that $f'(c) = 0$, but that f does **not** reach a local extremum at c. For instance, the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^3$ for all $x \in \mathbb{R}$ is differentiable, and it satisfies $f'(0) = 0$, but f does **not** reach a local extremum at $c = 0$.



Remark: Theorem 4.10.2 does **not** apply when a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ reaches a local extremum at a point $c \in A$ that does **not** belong to any open interval $I \subseteq A$. For instance, consider the function $f : [1, 3] \to \mathbb{R}$ given by $f(x) = x^2$ for all $x \in [1, 3]$. Then f reaches a global (and therefore also local) minimum at $c_1 = 1$ and a global (and therefore also local) maximum at $c_2 = 3$. However, $f'(c_1) = 2c_1 = 2$ and $f'(c_2) = 2c_2 = 6$. Note that the tangents to the graph of f through the points $(1, f(1)) = (1, 1)$ and $(3, f(3)) = (3, 9)$ are **not** horizontal (see the picture below; the tangents in question are in red).

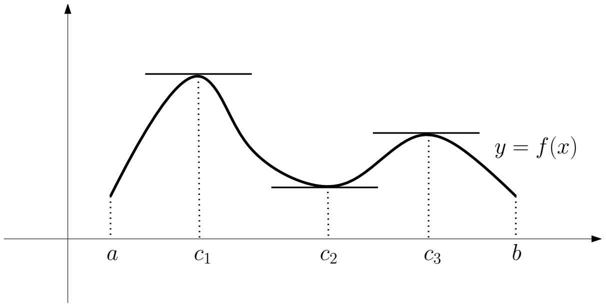


4.10.2 Rolle's Theorem

Rolle's Theorem. *Let $a, b \in \mathbb{R}$ be such that $a < b$, and let $f : [a, b] \to \mathbb{R}$ be a function that satisfies all the following:*

- *f is continuous on $[a, b]$;*
- *f is differentiable on (a, b);*
- *$f(a) = f(b)$.*

Then there exists some $c \in (a, b)$ such that $f'(c) = 0$.



Proof. Since f is continuous on $[a, b]$, the Extreme Value Theorem (see section 3.5.2) guarantees that f reaches both a global minimum and a global maximum on $[a, b]$. Let $m, M \in [a, b]$ be such that f reaches a global minimum at m and a global maximum at M. Clearly, $f(m) \leq f(M)$. If $f(m) = f(M)$, then f is a constant function, and so by Proposition 4.4.1, we have that $f'(c) = 0$ for all $c \in (a, b)$.

It remains to consider the case when $f(m) < f(M)$. Since $f(a) = f(b)$, we deduce that at least one of m, M belongs to (a, b).[3] But then f reaches a local extremum at some point $c \in (a, b)$,[4] and so by Theorem 4.10.2, we have that $f'(c) = 0$. $\square$

4.10.3 Lagrange's Mean Value Theorem

Lagrange's Mean Value Theorem. *Let $a, b \in \mathbb{R}$ be such that $a < b$, and let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). Then there exists some $c \in (a, b)$ such that $f'(c) = \frac{f(b)-f(a)}{b-a}$.*

Remark: Note that Rolle's Theorem is a special case of Lagrange's Mean Value Theorem for $f(a) = f(b)$.

---

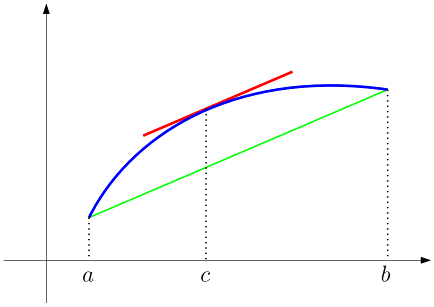
[3] Indeed, since $f(m) < f(M)$ and $f(a) = f(b)$, we see that at most one of $f(m)$ and $f(M)$ may be equal to $f(a) = f(b)$. Thus, at most one of m and M belongs to $\{a, b\}$, and consequently, the other one belongs to the open interval (a, b).

[4] Indeed, if $m \in (a, b)$, then we set $c = m$. Otherwise, we have that $M \in (a, b)$, and we set $c = M$. In either case, we have that $c \in (a, b)$, and f reaches a local extremum at c.

Remark: Note that the line through the points $\big(a, f(a)\big)$ and $\big(b, f(b)\big)$ is given by the equation

$$y = \tfrac{f(b)-f(a)}{b-a}(x-a) + f(a),$$

and in particular, the slope of this line is $\frac{f(b)-f(a)}{b-a}$. So, the outcome of Lagrange's Mean Value Theorem is that there exists a point $c \in (a, b)$ such that the tangent to the graph of f through the point $\big(c, f(c)\big)$ has the same slope as the line above, i.e. the line through $\big(a, f(a)\big)$ and $\big(b, f(b)\big)$.



Proof of Lagrange's Mean Value Theorem. Define the function $F : [a, b] \to \mathbb{R}$ by setting

$$F(x) = f(x) - \tfrac{f(b)-f(a)}{b-a}(x-a)$$

for all $x \in [a, b]$. Then F is continuous on $[a, b]$ and differentiable on (a, b). Furthermore, we have that $F(a) = f(a) = F(b)$. Thus, F satisfies the hypotheses of Rolle's Theorem, and so there exists some $c \in (a, b)$ such that $F'(c) = 0$. But by the construction of F, we have that

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

for all $x \in (a, b)$. Setting $x = c$, we get that $f'(c) = \frac{f(b)-f(a)}{b-a}$, which is what we needed. □

Terminology: Lagrange's Mean Value Theorem is often referred to simply as the "Mean Value Theorem."

Corollary 4.10.3. *Let I be an open interval in $\mathbb{R}$, and let $f : I \to \mathbb{R}$ be a differentiable function. Then all the following hold:*

(a) if $f'(x) = 0$ for all $x \in I$, then f constant on I;

(b) if $f'(x) \geq 0$ for all $x \in I$, then f is non-decreasing on I;

(c) if $f'(x) > 0$ for all $x \in I$, then f is strictly increasing on I;

(d) if $f'(x) \leq 0$ for all $x \in I$, then f is non-increasing on I;

(e) if $f'(x) < 0$ for all $x \in I$, then f is strictly decreasing on I.

Proof. We prove (a) and (b). The proofs of (c), (d), and (e) are similar to the proof of (b).

We first prove (a). Assume that $f'(x) = 0$ for all $x \in I$. Fix $a, b \in I$; we must show that $f(a) = f(b)$. We may assume that $a \neq b$, for otherwise we are done. By symmetry, we may further assume that $a < b$. Then f restricted to $[a, b]$ satisfies the hypotheses of the Mean Value Theorem, and we deduce that there exists some $c \in (a, b)$ such that $f'(c) = \frac{f(b)-f(a)}{b-a}$. Since $f'(c) = 0$, it follows that $\frac{f(b)-f(a)}{b-a} = 0$, and consequently, $f(a) = f(b)$. This proves (a).

We now prove (b). Assume that $f'(x) \geq 0$ for all $x \in I$. Fix $a, b \in I$ such that $a \leq b$; we must show that $f(a) \leq f(b)$. We may assume that $a \neq b$ (and therefore $a < b$), for otherwise we are done. Clearly, f restricted to $[a, b]$ satisfies the hypotheses of the Mean Value Theorem, and so there exists some $c \in (a, b)$ such that $f'(c) = \frac{f(b)-f(a)}{b-a}$. But $f'(c) \geq 0$ and $b - a > 0$; consequently, $f(b) - f(a) \geq 0$, and we deduce that $f(a) \leq f(b)$. This proves (b). □

We have the following partial converse of Corollary 4.10.3.

Proposition 4.10.4. *Let I be an open interval in $\mathbb{R}$, and let $f : I \to \mathbb{R}$ be a differentiable function. Then all the following hold:*

(a) if f is constant on I, then $f'(x) = 0$ for all $x \in I$;

(b) if f is non-decreasing on I, then $f'(x) \geq 0$ for all $x \in I$;

(c) if f is non-increasing on I, then $f'(x) \leq 0$ for all $x \in I$.

Proof. Part (a) follows immediately from Proposition 4.4.1. We now prove (b); the proof of (c) is similar. So, let us assume that f is non-decreasing on I, and fix an arbitrary $x_0 \in I$. Then

$$f'(x_0) \quad = \quad \lim_{x \to x_0} \tfrac{f(x)-f(x_0)}{x-x_0} \quad = \quad \lim_{x \to x_0^+} \tfrac{f(x)-f(x_0)}{x-x_0} \quad \overset{(*)}{\geq} \quad 0,$$

where (*) follows from the fact that f is non-decreasing, and so for any $x \in I$ such that $x > x_0$, we have that $f(x) \geq f(x_0)$, and consequently, $\frac{f(x)-f(x_0)}{x-x_0} \geq 0$. □

Remark: The converses of parts (c) and (e) of Corollary 4.10.3 are false. For example, the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^3$ for all $x \in \mathbb{R}$ is strictly increasing, but $f'(0) = 0$. Similarly, the function $g : \mathbb{R} \to \mathbb{R}$ given by $g(x) = -x^3$ for all $x \in \mathbb{R}$ is strictly decreasing, but $g'(0) = 0$.

Example 4.10.5. *Prove that $|\arctan x_1 - \arctan x_2| \leq |x_1 - x_2|$ for all $x_1, x_2 \in \mathbb{R}$.*

Solution. Fix $x_1, x_2 \in \mathbb{R}$. We may assume that $x_1 \neq x_2$ (for otherwise, the result is immediate), and by symmetry, we may further assume that $x_1 > x_2$. So, $|x_1 - x_2| = x_1 - x_2 > 0$. Moreover, since the arctangent function is strictly increasing,[5] we see that $\arctan x_1 > \arctan x_2$, and consequently, we have that

$$|\arctan x_1 - \arctan x_2| \;=\; \arctan x_1 - \arctan x_2 \;>\; 0.$$

Now, the function $\arctan x$ is continuous on $[x_2, x_1]$ and differentiable on (x_2, x_1), and so by the Mean Value Theorem, there exists some $y \in (x_2, x_1)$ such that

$$\arctan'(y) \;=\; \tfrac{\arctan x_1 - \arctan x_2}{x_1 - x_2}.$$

We now compute:

$$\tfrac{|\arctan x_1 - \arctan x_2|}{|x_1 - x_2|} \;=\; \tfrac{\arctan x_1 - \arctan x_2}{x_1 - x_2} \;=\; \arctan'(y) \;=\; \tfrac{1}{1+y^2} \;\leq\; 1.$$

Multiplying both sides by the positive number $|x_1 - x_2|$, we obtain

$$|\arctan x_1 - \arctan x_2| \;\leq\; |x_1 - x_2|,$$

which is what we needed to show. □

4.10.4 Cauchy's Mean Value Theorem

Cauchy's Mean Value Theorem. *Let $a, b \in \mathbb{R}$ be such that $a < b$, and let $f, g : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b), with $g'(x) \neq 0$ for all $x \in (a, b)$. Then $g(a) \neq g(b)$, and there exists some $c \in (a, b)$ such that*

$$\tfrac{f'(c)}{g'(c)} \;=\; \tfrac{f(b)-f(a)}{g(b)-g(a)}.$$

Remark: Cauchy's Mean Value Theorem generalizes Lagrange's Mean Value Theorem. Indeed, when $g(x) = x$, then Cauchy's Mean Value Theorem becomes Lagrange's Mean Value Theorem.

Proof. We have that $g(a) \neq g(b)$, because otherwise, Rolle's Theorem would imply that $g'(c) = 0$ for some $c \in (a, b)$, a contradiction. Now, define $F : [a, b] \to \mathbb{R}$ by setting

$$F(x) \;=\; f(x) - \tfrac{f(b)-f(a)}{g(b)-g(a)} g(x)$$

for all $x \in [a, b]$. Clearly, F is continuous on $[a, b]$ and differentiable on (a, b), and furthermore, $F(a) = \frac{f(a)g(b)-f(b)g(a)}{g(b)-g(a)} = F(b)$. So, by Rolle's Theorem, there exists some $c \in (a, b)$ such that $F'(c) = 0$. But note that

$$F'(x) \;=\; f'(x) - \tfrac{f(b)-f(a)}{g(b)-g(a)} g'(x)$$

---

[5] Here, we may simply rely on the familiar fact from trigonometry that $\arctan x$ is a strictly increasing function. Alternatively, we may rely on the fact that for $x \in \mathbb{R}$, we have that $\frac{d}{dx} \arctan x = \frac{1}{1+x^2} > 0$, and so by Corollary 4.10.3(c), the function $\arctan x$ is strictly increasing.

for all $x \in (a, b)$. Setting $x = c$, we get that

$$\tfrac{f'(c)}{g'(c)} \quad = \quad \tfrac{f(b)-f(a)}{g(b)-g(a)},$$

which is what we needed to show. $\square$

4.11 L'Hôpital's Rule

When evaluating limits, the following forms are "indeterminate," i.e. they can in principle be anything:

$$\tfrac{0}{0} \qquad \tfrac{\infty}{\infty} \qquad 0 \cdot \infty \qquad 1^{\infty} \qquad \infty - \infty \qquad 0^0 \qquad \infty^0$$

In this section, we cover L'Hôpital's Rule, which can (sometimes) be used to evaluate limits that look like indeterminate forms. More precisely, L'Hôpital's Rule is (sometimes) used for evaluating indeterminate forms of types "$\frac{0}{0}$" and "$\frac{\infty}{\infty}$," though as we shall see, other types of indeterminate forms can sometimes be transformed into these two in convenient ways.

In fact, there are several related results, all of which are referred to as "L'Hôpital's Rule." The proofs of these rules are somewhat complicated (here, we note that they rely on Cauchy's Mean Value Theorem). So, the remainder of the section is organized as follows. In subsection 4.11.1, we state the rules (without proof), and we give a number of relevant examples. Then, in subsection 4.11.2, we prove all versions of L'Hôpital's Rule.

4.11.1 L'Hôpital's Rule: theorem statements and examples

L'Hôpital's Rule for right-hand limits. *Let $a_0 \in \mathbb{R}$ and $\delta > 0$, and assume that functions $f, g : (a_0, a_0 + \delta) \to \mathbb{R}$ are both differentiable, and that $g'(x) \neq 0$ for all $x \in (a_0, a_0 + \delta)$. Assume moreover that one of the following holds:*

(i) $\lim\limits_{x \to a_0^+} f(x) = \lim\limits_{x \to a_0^+} g(x) = 0$;

(ii) $\lim\limits_{x \to a_0^+} f(x) = \pm\infty$ *and* $\lim\limits_{x \to a_0^+} g(x) = \pm\infty$.

Then, if $\lim\limits_{x \to a_0^+} \frac{f'(x)}{g'(x)}$ *exists (either as a real number or as $\pm\infty$), then* $\lim\limits_{x \to a_0^+} \frac{f(x)}{g(x)}$ *exists as well, and in this case, we have that*

$$\lim_{x \to a_0^+} \tfrac{f(x)}{g(x)} \quad = \quad \lim_{x \to a_0^+} \tfrac{f'(x)}{g'(x)}.$$

L'Hôpital's Rule for left-hand limits. *Let $a_0 \in \mathbb{R}$ and $\delta > 0$, and assume that functions $f, g : (a_0 - \delta, a_0) \to \mathbb{R}$ are both differentiable, and that $g'(x) \neq 0$ for all $x \in (a_0 - \delta, a_0)$. Assume moreover that one of the following holds:*

(i) $\lim\limits_{x \to a_0^-} f(x) = \lim\limits_{x \to a_0^-} g(x) = 0$*;*

(ii) $\lim\limits_{x \to a_0^-} f(x) = \pm\infty$ *and* $\lim\limits_{x \to a_0^-} g(x) = \pm\infty$*.*

Then, if $\lim\limits_{x \to a_0^-} \frac{f'(x)}{g'(x)}$ *exists (either as a real number or as* $\pm\infty$*), then* $\lim\limits_{x \to a_0^-} \frac{f(x)}{g(x)}$ *exists as well, and in this case, we have that*

$$\lim_{x \to a_0^-} \tfrac{f(x)}{g(x)} \quad = \quad \lim_{x \to a_0^-} \tfrac{f'(x)}{g'(x)}.$$

L'Hôpital's Rule for limits at a point. *Let* $a_0 \in \mathbb{R}$ *and* $\delta > 0$*, and assume that functions* $f, g : (a_0 - \delta, a_0) \cup (a_0, a_0 + \delta) \to \mathbb{R}$ *are both differentiable, and that* $g'(x) \neq 0$ *for all* $x \in (a_0 - \delta, a_0) \cup (a_0, a_0 + \delta)$*. Assume moreover that one of the following holds:*

(i) $\lim\limits_{x \to a_0} f(x) = \lim\limits_{x \to a_0} g(x) = 0$*;*

(ii) $\lim\limits_{x \to a_0} f(x) = \pm\infty$ *and* $\lim\limits_{x \to a_0} g(x) = \pm\infty$*.*

Then, if $\lim\limits_{x \to a_0} \frac{f'(x)}{g'(x)}$ *exists (either as a real number or as* $\pm\infty$*), then* $\lim\limits_{x \to a_0} \frac{f(x)}{g(x)}$ *exists as well, and in this case, we have that*

$$\lim_{x \to a_0} \tfrac{f(x)}{g(x)} \quad = \quad \lim_{x \to a_0} \tfrac{f'(x)}{g'(x)}.$$

L'Hôpital's Rule for limits at $+\infty$. *Let* $a_0 \in \mathbb{R}$*, and assume that functions* $f, g : (a_0, +\infty) \to \mathbb{R}$ *are both differentiable, and that* $g'(x) \neq 0$ *for all* $x \in (a_0, +\infty)$*. Assume moreover that one of the following holds:*

(i) $\lim\limits_{x \to +\infty} f(x) = \lim\limits_{x \to +\infty} g(x) = 0$*;*

(ii) $\lim\limits_{x \to +\infty} f(x) = \pm\infty$ *and* $\lim\limits_{x \to +\infty} g(x) = \pm\infty$*.*

Then, if $\lim\limits_{x \to +\infty} \frac{f'(x)}{g'(x)}$ *exists (either as a real number or as* $\pm\infty$*), then* $\lim\limits_{x \to +\infty} \frac{f(x)}{g(x)}$ *exists as well, and in this case, we have that*

$$\lim_{x \to +\infty} \tfrac{f(x)}{g(x)} \quad = \quad \lim_{x \to +\infty} \tfrac{f'(x)}{g'(x)}.$$

L'Hôpital's Rule for limits at $-\infty$. *Let* $a_0 \in \mathbb{R}$*, and assume that functions* $f, g : (-\infty, a_0) \to \mathbb{R}$ *are both differentiable, and that* $g'(x) \neq 0$ *for all* $x \in (-\infty, a_0)$*. Assume moreover that one of the following holds:*

(i) $\lim\limits_{x \to -\infty} f(x) = \lim\limits_{x \to -\infty} g(x) = 0$*;*

(ii) $\lim\limits_{x\to-\infty} f(x) = \pm\infty$ *and* $\lim\limits_{x\to-\infty} g(x) = \pm\infty$.

Then, if $\lim\limits_{x\to-\infty} \frac{f'(x)}{g'(x)}$ *exists (either as a real number or as* $\pm\infty$*), then* $\lim\limits_{x\to-\infty} \frac{f(x)}{g(x)}$ *exists as well, and in this case, we have that*

$$\lim_{x\to-\infty} \frac{f(x)}{g(x)} \quad = \quad \lim_{x\to-\infty} \frac{f'(x)}{g'(x)}.$$

Remark/Warning: In order to apply L'Hôpital's Rule (any version), we must make sure we really do have an indeterminate form of the form "$\frac{0}{0}$" or "$\frac{\infty}{\infty}$," for otherwise, L'Hôpital's Rule does **not** apply. For example, $\lim\limits_{x\to 0} \frac{1+x}{1+2x} = \frac{\lim\limits_{x\to 0}(1+x)}{\lim\limits_{x\to 0}(1+2x)} = 1$, but $\lim\limits_{x\to 0} \frac{(1+x)'}{(1+2x)'} = \lim\limits_{x\to 0} \frac{1}{2} = \frac{1}{2}$, and so

$$\lim_{x\to 0} \frac{1+x}{1+2x} \quad \neq \quad \lim_{x\to 0} \frac{(1+x)'}{(1+2x)'}.$$

The reason this we could not apply L'Hôpital's Rule in this case is because our original limit $\lim\limits_{x\to 0} \frac{1+x}{1+2x}$ was not of the form "$\frac{0}{0}$" or "$\frac{\infty}{\infty}$," i.e. neither (i) nor (ii) from the statement of the appropriate L'Hôpital's Rule was satisfied.

Example 4.11.1. *Compute* $\lim\limits_{x\to 0} \frac{\tan x - x}{x - \sin x}$.

Solution. Define $f, g : (-\frac{\pi}{2}, 0) \cup (0, \frac{\pi}{2})$ by setting $f(x) := \tan x - x$ and $g(x) := x - \sin x$ for all $x \in (-\frac{\pi}{2}, 0) \cup (0, \frac{\pi}{2})$. Clearly, both f and g are differentiable, and moreover, for $x \in (-\frac{\pi}{2}, 0) \cup (0, \frac{\pi}{2})$, we have that $g'(x) = \frac{d}{dx}(x - \sin x) = 1 - \cos(x) \neq 0$. Furthermore, we have that $\lim\limits_{x\to 0} f(x) = \lim\limits_{x\to 0} g(x) = 0$. Next, we now compute:

$$\begin{aligned}
\lim_{x\to 0} \frac{f'(x)}{g'(x)} \quad &= \quad \lim_{x\to 0} \frac{\frac{d}{dx}(\tan x - x)}{\frac{d}{dx}(x - \sin x)} \\
&= \quad \lim_{x\to 0} \frac{\frac{1}{\cos^2 x} - 1}{1 - \cos x} \\
&= \quad \lim_{x\to 0} \frac{1 - \cos^2 x}{\cos^2 x(1 - \cos x)} \\
&= \quad \lim_{x\to 0} \frac{(1 - \cos x)(1 + \cos x)}{\cos^2 x(1 - \cos x)} \\
&= \quad \lim_{x\to 0} \frac{1 + \cos x}{\cos^2 x} \\
&= \quad 2.
\end{aligned}$$

So, by L'Hôpital's Rule, we have that $\lim\limits_{x\to 0} \frac{f(x)}{g(x)} = 2$, that is,

$$\lim_{x\to 0} \frac{\tan x - x}{x - \sin x} \quad = \quad 2.$$

Remark: Typically, when computing limits using L'Hôpital's Rule, we do not include the level of detail that we included above. Normally, we simply write something like this:

$$\begin{array}{rcl}
\lim\limits_{x\to 0}\frac{\tan x-x}{x-\sin x} & \overset{(*)}{=} & \lim\limits_{x\to 0}\frac{\frac{d}{dx}(\tan x-x)}{\frac{d}{dx}(x-\sin x)} \\
& = & \lim\limits_{x\to 0}\frac{\frac{1}{\cos^2 x}-1}{1-\cos x} \\
& = & \lim\limits_{x\to 0}\frac{1-\cos^2 x}{\cos^2 x(1-\cos x)} \\
& = & \lim\limits_{x\to 0}\frac{(1-\cos x)(1+\cos x)}{\cos^2 x(1-\cos x)} \\
& = & \lim\limits_{x\to 0}\frac{1+\cos x}{\cos^2 x} \\
& = & 2,
\end{array}$$

where (*) follows from L'Hôpital's Rule, which we could apply because we have that $\lim\limits_{x\to 0}(\tan x - x) = 0$ and $\lim\limits_{x\to 0}(x - \sin x) = 0$. $\square$

Remark/Warning: For indeterminate forms of type "$\frac{0}{0}$" and "$\frac{\infty}{\infty}$," L'Hôpital's Rule gives us only a one-way implication, as follows:

$$\lim_{x\to a_0}\frac{f'(x)}{g'(x)} = L \quad \begin{array}{c}\Longrightarrow \\ \xcancel{\Longleftarrow}\end{array} \quad \lim_{x\to a_0}\frac{f(x)}{g(x)} = L,$$

and similar for one-sided limits and limits at ±∞. This means that if $\lim\limits_{x\to a_0}\frac{f'(x)}{g'(x)}$ does not exist (either as a real number or as $\pm\infty$), then we have no information about $\lim\limits_{x\to a_0}\frac{f(x)}{g(x)}$, that is, $\lim\limits_{x\to a_0}\frac{f(x)}{g(x)}$ may be a real number, may be $\pm\infty$, or may fail to exist. This further means that if, having verified that $\lim\limits_{x\to a_0}\frac{f(x)}{g(x)}$ an indeterminate form of type "$\frac{0}{0}$" or "$\frac{\infty}{\infty}$," we attempt to compute

$$\lim_{x\to a_0}\frac{f(x)}{g(x)} \overset{\text{L'Hôpital}}{=} \lim_{x\to a_0}\frac{f'(x)}{g'(x)} = \dots,$$

our computation is only valid if $\lim\limits_{x\to a_0}\frac{f'(x)}{g'(x)}$ does, in fact, exist. If $\lim\limits_{x\to a_0}\frac{f'(x)}{g'(x)}$ does **not** exist, then we **cannot** conclude that $\lim\limits_{x\to a_0}\frac{f(x)}{g(x)}$ does not exist, either. In this case, we need to scratch out the whole computation

$$\xcancel{\lim_{x\to a_0}\frac{f(x)}{g(x)} \overset{\text{L'Hôpital}}{=} \lim_{x\to a_0}\frac{f'(x)}{g'(x)} = \dots}$$

and start over: L'Hôpital's Rule will not help us, and we need to find some other way to compute the limit $\lim\limits_{x \to a_0} \frac{f(x)}{g(x)}$ (or to show that the limit does not exist). Similar remarks apply to the use of L'Hôpital's Rule for one-sided limits, and for limits at $\pm\infty$. A concrete example is given below.

Example 4.11.2. *Compute* $\lim\limits_{x \to 0} \frac{x^2 \sin(\frac{1}{x})}{x}$. *What happens if you try to apply L'Hôpital's Rule?*

Solution. Clearly,

$$-|x| \quad \leq \quad \frac{x^2 \sin(\frac{1}{x})}{x} \quad \leq \quad |x| \qquad \forall x \in \mathbb{R} \setminus \{0\}.$$

Since $\lim\limits_{x \to 0}(-|x|) = 0$ and $\lim\limits_{x \to 0} |x| = 0$, the Squeeze Theorem guarantees that

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

and in particular, our limit exists.

Now, consider the functions $f, g : (-\infty, 0) \cup (0, +\infty) \to \mathbb{R}$ given by $f(x) = x^2 \sin(\frac{1}{x})$ and $g(x) = x$ for all $x \in (-\infty, 0) \cup (0, +\infty)$. Then f and g are both differentiable, and moreover, we have that $g'(x) = 1 \neq 0$ for all $x \in (-\infty, 0) \cup (0, +\infty)$. We further have that $\lim\limits_{x \to 0} f(x) = 0$ and $\lim\limits_{x \to 0} x = 0$.[6] However,

$$\begin{aligned}
\lim_{x \to 0} \frac{f'(x)}{g'(x)} \quad &= \quad \lim_{x \to 0} \frac{\frac{d}{dx}(x^2 \sin(\frac{1}{x}))}{\frac{d}{dx}(x)} \\
&= \quad \lim_{x \to 0} \frac{2x \sin(\frac{1}{x}) + x^2 \cos(\frac{1}{x})(-\frac{1}{x^2})}{1} \\
&= \quad \lim_{x \to 0} \big(2x \sin(\tfrac{1}{x}) - \cos(\tfrac{1}{x})\big) \\
&= \quad 2\big(\lim_{x \to 0} x \sin(\tfrac{1}{x})\big) - \big(\lim_{x \to 0} \cos(\tfrac{1}{x})\big) \\
&\overset{(*)}{=} \quad -\lim_{x \to 0} \cos(\tfrac{1}{x}),
\end{aligned}$$

where (*) follows from the fact $\lim\limits_{x \to 0} x \sin(\frac{1}{x}) = 0$; indeed, since $-|x| \leq x \sin(\frac{1}{x}) \leq |x|$ for all $x \in \mathbb{R} \setminus \{0\}$, and since $\lim\limits_{x \to 0}(-|x|) = \lim\limits_{x \to 0} |x| = 0$, the Squeeze Theorem guarantees that $\lim\limits_{x \to 0} x \sin(\frac{1}{x}) = 0$. However, note that $\lim\limits_{x \to 0} \cos(\frac{1}{x})$ does not exist,

---

[6] For the former limit, note that $-x^2 \leq f(x) \leq x^2$ for all $x \in (-\infty, 0) \cup (0, \infty)$. Since $\lim\limits_{x \to 0}(-x^2) = 0$ and $\lim\limits_{x \to 0} x^2 = 0$, the Squeeze Theorem implies that $\lim\limits_{x \to 0} f(x) = 0$.

because for $x \approx 0$, $\cos(\frac{1}{x})$ keeps oscillating.[7] So, $\lim\limits_{x\to 0}\frac{f'(x)}{g'(x)}$ does not exist, either. Nevertheless, as we saw above, $\lim\limits_{x\to 0}\frac{f(x)}{g(x)} = 0$. □

Remark: In the remainder of this subsection, we consider several more routine applications of L'Hôpital's Rule.

Example 4.11.3. *Compute* $\lim\limits_{x\to+\infty}\frac{2^x}{x}$.

Remark: Note that this is an indeterminate form of type "$\frac{\infty}{\infty}$."

Proof. We compute:

$$\begin{aligned}
\lim_{x\to+\infty}\frac{\overset{\nearrow\, +\infty}{\textcircled{2^x}}}{\underset{\searrow\, +\infty}{\textcircled{x}}} &= \lim_{x\to+\infty}\frac{\frac{d}{dx}(2^x)}{\frac{d}{dx}(x)} \qquad \text{by L'Hôpital's Rule}\\
&= \lim_{x\to+\infty}\frac{2^x\ln 2}{1}\\
&= \lim_{x\to+\infty}(2^x\ln 2)\\
&= +\infty.
\end{aligned}$$

□

Remark: Occasionally, we need to apply l'Hôpital's Rule more than once in order to compute a limit, as our next example shows.

Example 4.11.4. *Compute* $\lim\limits_{x\to+\infty}\frac{x^2}{e^x}$.

Solution. We compute:

---

[7] Here is a formal proof. Consider the sequences $\{a_n\}_{n=0}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ of non-zero real numbers defined by setting $a_n := \frac{1}{2n\pi}$ and $b_nb_n := \frac{1}{(2n+1)\pi}$ for all $n \in \mathbb{N}$. Then $\lim\limits_{n\to\infty} a_n = 0$ and $\lim\limits_{n\to\infty} b_n = 0$. However:

- $\lim\limits_{n\to\infty}\cos(\frac{1}{a_n}) = \lim\limits_{n\to\infty}\cos(2n\pi) = \lim\limits_{n\to\infty} 1 = 1$;
- $\lim\limits_{n\to\infty}\cos(\frac{1}{b_n}) = \lim\limits_{n\to\infty}\cos((2n+1)\pi) = \lim\limits_{n\to\infty}(-1) = -1$.

So, by Theorem 3.3.1, $\lim\limits_{x\to 0}\cos(\frac{1}{x})$ does not exist.

$$\begin{aligned}
\lim_{x\to+\infty}\frac{\overset{\nearrow +\infty}{\boxed{x^2}}}{\underset{\searrow +\infty}{\boxed{e^x}}} &= \lim_{x\to+\infty}\frac{\frac{d}{dx}x^2}{\frac{d}{dx}e^x} && \text{by L'Hôpital's Rule}\\
&= \lim_{x\to+\infty}\frac{\overset{\nearrow +\infty}{\boxed{2x}}}{\underset{\searrow +\infty}{\boxed{e^x}}}\\
&= \lim_{x\to+\infty}\frac{\frac{d}{dx}(2x)}{\frac{d}{dx}e^x} && \text{by L'Hôpital's Rule}\\
&= \lim_{x\to+\infty}\frac{2}{e^x}\\
&= 0.
\end{aligned}$$

□

Remark: Sometimes, after a bit of algebraic manipulation, L'Hôpital's Rule can be used to compute limits in cases we obtain other indeterminate forms, i.e. indeterminate forms other than "$\frac{0}{0}$" or "$\frac{\infty}{\infty}$." Here are a couple of examples.

Example 4.11.5. *Compute* $\lim\limits_{x\to 0^+}(x\ln x)$.

Remark: Note that $\lim\limits_{x\to 0+} x = 0$ and $\lim\limits_{x\to 0^+}\ln x = -\infty$, and so we are dealing with an indeterminate form of type "$0\cdot\infty$."

Solution. We compute:

$$\begin{aligned}
\lim_{x\to 0^+}(x\ln x) &= \lim_{x\to 0^+}\frac{\overset{\nearrow -\infty}{\boxed{\ln x}}}{\underset{\searrow +\infty}{\boxed{\frac{1}{x}}}}\\
&= \lim_{x\to 0^+}\frac{\frac{d}{dx}(\ln x)}{\frac{d}{dx}(\frac{1}{x})} && \text{by L'Hôpital's Rule}
\end{aligned}$$

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

□

Example 4.11.6. *Compute* $\lim\limits_{x\to+\infty} x^{1/x}$.

Remark: Note that this is an indeterminate form of type "∞^0."

Solution. We compute:

$$
\begin{aligned}
\lim_{x\to+\infty} x^{1/x} &= \lim_{x\to+\infty} e^{\frac{1}{x}\ln x} \\
&= e^{\lim\limits_{x\to+\infty} \frac{\ln x}{x}} \\
&\overset{(*)}{=} e^{\lim\limits_{x\to+\infty} \frac{\frac{d}{dx}(\ln x)}{\frac{d}{dx}(x)}} \\
&= e^{\lim\limits_{x\to+\infty} \frac{\frac{1}{x}}{1}} \\
&= e^{\lim\limits_{x\to+\infty} \frac{1}{x}} \\
&= e^0 \quad = \quad 1,
\end{aligned}
$$

where in (*), we applied l'Hôpital's Rule, which was applicable because $\lim\limits_{x\to+\infty} \ln x = +\infty$ and $\lim\limits_{x\to+\infty} x = +\infty$. □

Example 4.11.7. *Compute* $\lim\limits_{x\to+\infty} \big(x\ln(x+1) - x\ln x\big)$.

Remark: Note that this is an indeterminate form of type "$\infty - \infty$."

Proof. We compute:

$$
\begin{aligned}
\lim_{x\to+\infty} \big(x\ln(x+1) - x\ln x\big) &= \lim_{x\to+\infty} x\big(\ln(x+1) - \ln x\big) \\
&= \lim_{x\to+\infty} x\ln\left(\tfrac{x+1}{x}\right) \\
&= \lim_{x\to+\infty} \frac{\ln\left(1+\frac{1}{x}\right)}{\frac{1}{x}}
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(*)}{=} \lim_{x\to+\infty} \frac{\frac{d}{dx}\ln\left(1+\frac{1}{x}\right)}{\frac{d}{dx}\left(\frac{1}{x}\right)}\\
&= \lim_{x\to+\infty} \frac{\frac{1}{1+\frac{1}{x}}(-\frac{1}{x^2})}{-\frac{1}{x^2}}\\
&= \lim_{x\to+\infty} \left(\frac{1}{1+\frac{1}{x}}\right)\\
&= \frac{1}{1+\lim\limits_{x\to+\infty}\left(\frac{1}{x}\right)}\\
&= \frac{1}{1+0} = 1,
\end{aligned}
$$

where (*) was obtained via L'Hôpital's Rule, which applies since $\lim\limits_{x\to+\infty} \ln(1+\frac{1}{x}) = \ln 1 = 0$ and $\lim\limits_{x\to+\infty}\left(\frac{1}{x}\right) = 0$. □

4.11.2 L'Hôpital's Rule: the proof

In this subsection, we prove all versions of L'Hôpital's Rule, restated below for the reader's convenience. As we shall see, the main difficulty is in proving the L'Hôpital's Rule for right-hand limits; we note that the proof of this rule relies on Cauchy's Mean Value Theorem. The rule for left-hand limits can be derived from the one for right-hand limits. Meanwhile, the rule for limits at a point follows immediately from the rules for one-sided limits and from Lemma 3.6.3. L'Hôpital's Rules for limits at $\pm\infty$ can be derived from those for one-sided limits. We begin with a simple technical lemma.

Lemma 4.11.8. *Let $f : A \subseteq \mathbb{R}$ be a function, differentiable on some open interval $I \subseteq A$ of $\mathbb{R}$. Assume that $f'(x) \neq 0$ for all $x \in I$. Then $f \restriction I$ is one-to-one, and in particular, there exists at most one point $x \in I$ such that $f(x) = 0$.*

Notation: As usual, $f \restriction I$ denotes the restriction of f to I.

Proof. Suppose otherwise, and fix distinct points $a, b \in I$ such that $f(a) = f(b)$. By symmetry, we may assume that $a < b$. Since f is differentiable on I, it is also continuous on I (by Theorem 4.3.1). So, f is continuous on $[a, b]$ and differentiable on (a, b). Since $f(a) = f(b)$, Rolle's theorem guarantees that there exists some $c \in (a, b) \subseteq I$ such that $f'(c) = 0$, a contradiction. □

L'Hôpital's Rule for right-hand limits. *Let $a_0 \in \mathbb{R}$ and $\delta > 0$, and assume that functions $f, g : (a_0, a_0 + \delta) \to \mathbb{R}$ are both differentiable, and that $g'(x) \neq 0$ for all $x \in (a_0, a_0 + \delta)$. Assume moreover that one of the following holds:*

(i) $\lim\limits_{x\to a_0^+} f(x) = \lim\limits_{x\to a_0^+} g(x) = 0$*;*

(ii) $\lim\limits_{x \to a_0^+} f(x) = \pm\infty$ *and* $\lim\limits_{x \to a_0^+} g(x) = \pm\infty$.

Then, if $\lim\limits_{x \to a_0^+} \frac{f'(x)}{g'(x)}$ *exists (either as a real number or as* $\pm\infty$*), then* $\lim\limits_{x \to a_0^+} \frac{f(x)}{g(x)}$ *exists as well, and in this case, we have that*

$$\lim_{x \to a_0^+} \tfrac{f(x)}{g(x)} \quad = \quad \lim_{x \to a_0^+} \tfrac{f'(x)}{g'(x)}.$$

Proof. First of all, in view of Lemma 4.11.8, there exists at most one point $x \in (a_0, a_0 + \delta)$ such that $g(x) = 0$. But then we may simply assume that $g(x) \neq 0$ for all $x \in (a_0, a_0 + \delta)$, for otherwise, we simply choose a suitable smaller δ. Indeed, suppose that for some $x_0 \in (a_0, a_0 + \delta)$, we have that $g(x_0) = 0$. Then we set $\delta' := x_0 - a_0$, we consider the restrictions of f and g to the interval $(a_0, a_0 + \delta') = (a_0, x_0) \subsetneqq (a_0, a + \delta)$, and we consider δ' instead of δ. We may do this because the value of the limit of a function as $x \to a_0^+$ is determined only by the values of x very slightly larger than a_0. With our additional assumption, we see that $\frac{f(x)}{g(x)}$ is defined for all $x \in (a_0, a_0 + \delta)$.

Now, we assume that $L := \lim\limits_{x \to a_0^+} \frac{f'(x)}{g'(x)}$ exists (as a real number or as $\pm\infty$), and we show that $\lim\limits_{x \to a_0^+} \frac{f(x)}{g(x)} = L$. We consider two cases: when (i) holds, and when (ii) holds.

Case 1: (i) holds, i.e. $\lim\limits_{x \to a_0^+} f(x) = \lim\limits_{x \to a_0^+} g(x) = 0$.

We extend the domain of f and g to a_0 by setting $f(a_0) := 0$ and $g(a_0) := 0$.[8] Then $\lim\limits_{x \to a_0^+} f(x) = 0 = f(a_0)$ and $\lim\limits_{x \to a_0^+} g(x) = 0 = g(a_0)$, and so f and g are both continuous at a_0 from the right. Meanwhile, f and g are differentiable, and therefore continuous (by Theorem 4.3.1) on the interval $(a_0, a_0 + \delta)$. Thus, f and g are continuous on $[a_0, a_0 + \delta)$ and differentiable on $(a_0, a_0 + \delta)$. Note that this implies that for all $b \in (a_0, a_0 + \delta)$, we may apply Cauchy's Mean Value Theorem to the functions f and g and the interval $[a_0, b]$.

Now, there are three cases to consider: when $L \in \mathbb{R}$, when $L = +\infty$, and when $L = -\infty$. We consider the first two cases; the third case is similar to the second.

Case 1.1: $L \in \mathbb{R}$.

We fix an arbitrary $\varepsilon > 0$, and using the fact that $\lim\limits_{x \to a_0^+} \frac{f'(x)}{g'(x)} = L$, we fix $\delta_0 \in (0, \delta)$ such that for all $x \in (a_0, a_0 + \delta_0)$, we have that $|\frac{f'(x)}{g'(x)} - L| < \varepsilon$. Now, fix an arbitrary

---

[8] Technically, we are defining auxiliary functions $f_1, g_1 : [a_0, a_0 + \delta) \to \mathbb{R}$ by setting

$$f_1(x) \quad = \quad \begin{cases} f(x) & \text{if} \quad x \neq a_0 \\ 0 & \text{if} \quad x = a_0 \end{cases} \qquad \text{and} \qquad g_1(x) \quad = \quad \begin{cases} g(x) & \text{if} \quad x \neq a_0 \\ 0 & \text{if} \quad x = a_0 \end{cases}$$

for all $x \in [a_0, a_0 + \delta)$.

$b_0 \in (a_0, a_0+\delta_0)$; we must show that $|\frac{f(b_0)}{g(b_0)} - L| < \varepsilon$. Note that this will immediately imply that $\lim\limits_{x \to a_0^+} \frac{f(x)}{g(x)} = L$.

First, we apply Cauchy's Mean Value Theorem to the functions f and g and the interval $[a_0, b_0]$ to obtain a point $c_0 \in (a_0, b_0)$ such that

$$\frac{f'(c_0)}{g'(c_0)} \quad = \quad \frac{f(b_0)-f(a_0)}{g(b_0)-g(a_0)} \quad \overset{(*)}{=} \quad \frac{f(b_0)}{g(b_0)},$$

where (*) follows from the fact that $f(a_0) = g(a_0) = 0$. But now

$$|\frac{f(b_0)}{g(b_0)} - L| \quad = \quad |\frac{f'(c_0)}{g'(c_0)} - L| \quad \overset{(*)}{<} \quad \varepsilon,$$

where (*) follows from the fact that $c_0 \in (a_0, b_0) \subseteq (a_0, a_0+\delta_0)$. This proves that $\lim\limits_{x \to a_0^+} \frac{f(x)}{g(x)} = L$, which is what we needed to show.

Case 1.2: $L = +\infty$.

We now have that $\lim\limits_{x \to a_0^+} \frac{f'(x)}{g'(x)} = +\infty$, and we must show that $\lim\limits_{x \to a_0^+} \frac{f(x)}{g(x)} = +\infty$.

Fix an arbitrary $M \in \mathbb{R}$. Using the fact that $\lim\limits_{x \to a_0^+} \frac{f'(x)}{g'(x)} = +\infty$, we fix $\delta_0 \in (0, \delta)$ such that for all $x \in (a_0, a_0+\delta_0)$, we have that $\frac{f'(x)}{g'(x)} > M$. Fix any $b_0 \in (a_0, a_0+\delta_0)$. We must show that $\frac{f(b_0)}{g(b_0)} > M$; note that this will immediately imply that $\lim\limits_{x \to a_0^+} \frac{f(x)}{g(x)} = +\infty$, which is what we need to show.

First, we apply Cauchy's Mean Value Theorem to the functions f and g and the interval $[a_0, b_0]$, and we obtain a point $c_0 \in (a_0, b_0)$ such that

$$\frac{f'(c_0)}{g'(c_0)} \quad = \quad \frac{f(b_0)-f(a_0)}{g(b_0)-g(a_0)} \quad \overset{(*)}{=} \quad \frac{f(b_0)}{g(b_0)},$$

where (*) follows from the fact that $f(a_0) = g(a_0) = 0$. But now

$$\frac{f(b_0)}{g(b_0)} \quad = \quad \frac{f'(c_0)}{g'(c_0)} \quad \overset{(*)}{>} \quad M,$$

where (*) follows from the fact that $c_0 \in (a_0, b_0) \subseteq (a_0, a_0+\delta_0)$. This proves that $\lim\limits_{x \to a_0^+} \frac{f(x)}{g(x)} = +\infty$, which is what we needed to show.

Case 2: (ii) holds, i.e. $\lim\limits_{x \to a_0^+} f(x) = \pm\infty$ and $\lim\limits_{x \to a_0^+} g(x) = \pm\infty$.

There are three cases to consider: when $L \in \mathbb{R}$, when $L = +\infty$, and when $L = -\infty$. We consider the first two cases; the third case is similar to the second.

Case 2.1: $L \in \mathbb{R}$.

Fix $\varepsilon > 0$. Using the fact that $\lim\limits_{x \to a_0^+} \frac{f'(x)}{g'(x)} = L$, we fix $\delta_0 \in (0, \delta)$ such that for all $x \in (a_0, a_0+\delta_0)$, we have that $|\frac{f'(x)}{g'(x)} - L| < \frac{\varepsilon}{2}$, that is, $L - \frac{\varepsilon}{2} < \frac{f'(x)}{g'(x)} < L + \frac{\varepsilon}{2}$.

Set $b_0 := a_0 + \delta_0 < a_0 + \delta$. Now, for each $x \in (a_0, b_0)$, we apply Cauchy's Mean Value Theorem to the functions f and g and to the interval $[x, b_0]$ to obtain a point $c_x \in (x, b_0)$ such that

$$\tfrac{f'(c_x)}{g'(c_x)} \quad = \quad \tfrac{f(b_0)-f(x)}{g(b_0)-g(x)} \quad = \quad \tfrac{f(x)-f(b_0)}{g(x)-g(b_0)} \quad = \quad \tfrac{\frac{f(x)}{g(x)}-\frac{f(b_0)}{g(x)}}{1-\frac{g(b_0)}{g(x)}},$$

which is equivalent to

$$\tfrac{f'(c_x)}{g'(c_x)}\Big(1-\tfrac{g(b_0)}{g(x)}\Big) \quad = \quad \tfrac{f(x)}{g(x)}-\tfrac{f(b_0)}{g(x)},$$

and therefore to

$$\tfrac{f(x)}{g(x)} \quad = \quad \tfrac{f'(c_x)}{g'(c_x)}+\tfrac{f(b_0)-\frac{f'(c_x)}{g'(c_x)}\cdot g(b_0)}{g(x)}$$

as well. We now define an auxiliary function $r : (a_0, b_0) \to \mathbb{R}$ by setting

$$r(x) \quad := \quad \tfrac{f(b_0)-\frac{f'(c_x)}{g'(c_x)}\cdot g(b_0)}{g(x)}$$

for all $x \in (a_0, b_0)$. Note that this means that

$$\tfrac{f(x)}{g(x)} \quad = \quad \tfrac{f'(c_x)}{g'(c_x)}+r(x)$$

for all $x \in (a_0, b_0)$.

Now, we claim that $\lim\limits_{x\to a_0^+} r(x) = 0$. First of all, note that for all $x \in (a_0, b_0)$, we have that $c_x \in (x, b_0) \subseteq (a_0, a_0+\delta_0)$, and consequently, $L-\frac{\varepsilon}{2} < \frac{f'(c_x)}{g'(c_x)} < L+\frac{\varepsilon}{2}$; in particular, $\frac{f'(c_x)}{g'(c_x)}$ is bounded for $x \in (a_0, b_0)$. Since $f(b_0)$ and $g(b_0)$ are fixed constants, we see that $f(b_0)-\frac{f'(c_x)}{g'(c_x)}\cdot g(b_0)$ (i.e. the numerator of $r(x)$) is bounded for $x \in (a_0, b_0)$. Since $\lim\limits_{x\to a_0^+} g(x) = \pm\infty$, it follows that $\lim\limits_{x\to a_0^+} r(x) = 0$, as we had claimed.

Using the fact that $\lim\limits_{x\to a_0^+} r(x) = 0$, we fix $\delta_1 \in (0, \delta_0)$ such that for all $x \in (a_0, a_0+\delta_1)$, we have that $|r(x)-0| < \frac{\varepsilon}{2}$, i.e. $-\frac{\varepsilon}{2} < r(x) < \frac{\varepsilon}{2}$. Fix any $x \in (a_0, \delta_1)$. We claim that $|\frac{f(x)}{g(x)}-L| < \varepsilon$. First, since $x \in (a_0, \delta_1)$, we have that $|r(x)| < \frac{\varepsilon}{2}$. On the other hand, since $x \in (a_0, \delta_1) \subseteq (a_0, \delta_0)$, we have that $c_x \in (x, b_0) \subseteq (a_0, a_0+\delta_0)$, and consequently, $|\frac{f'(c_x)}{g'(c_x)}-L| < \frac{\varepsilon}{2}$. We now compute

$$\begin{aligned} |\tfrac{f(x)}{g(x)}-L| \quad &= \quad |(\tfrac{f'(c_x)}{g'(c_x)}+r(x))-L| \\ &= \quad |(\tfrac{f'(c_x)}{g'(c_x)}-L)+r(x)| \end{aligned}$$

$$\begin{aligned} &\leq \quad |\tfrac{f'(c_x)}{g'(c_x)} - L| + |r(x)| && \text{by the Triangle Inequality} \\ &< \quad \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} \\ &= \quad \varepsilon. \end{aligned}$$

This proves that $\lim\limits_{x \to a_0} \frac{f(x)}{g(x)} = L$, which is what we needed to show.

Case 2.2: $L = +\infty$.

We now have that $\lim\limits_{x \to a_0^+} \frac{f'(x)}{g'(x)} = +\infty$, and we must show that $\lim\limits_{x \to a_0^+} \frac{f(x)}{g(x)} = +\infty$.

We fix an arbitrary $M \in \mathbb{R}$. We must show that there exists some $\hat{\delta} \in (0, \delta)$ such that for all $x \in (a_0, a_0 + \hat{\delta})$, we have that $\frac{f(x)}{g(x)} > M$. Clearly, we may assume that $M > 0$, for otherwise, we consider any positive real number instead of M.

Using the fact that $\lim\limits_{x \to a_0^+} \frac{f'(x)}{g'(x)} = +\infty$, we fix $\delta_0 \in (0, \delta)$ such that for all $x \in (a_0, a_0 + \delta_0)$, we have that $\frac{f'(x)}{g'(x)} > 2M + \frac{1}{2}$.

We now proceed similarly as in Case 2.1. Set $b_0 := a_0 + \delta_0 < a_0 + \delta$. For each $x \in (a_0, b_0)$, we apply Cauchy's Mean Value Theorem to the interval $[x, b_0]$ to obtain a point $c_x \in (x, b_0)$ such that

$$\frac{f'(c_x)}{g'(c_x)} \quad = \quad \frac{f(b_0)-f(x)}{g(b_0)-g(x)} \quad = \quad \frac{f(x)-f(b_0)}{g(x)-g(b_0)} \quad = \quad \frac{\frac{f(x)}{g(x)} - \frac{f(b_0)}{g(x)}}{1 - \frac{g(b_0)}{g(x)}},$$

which is equivalent to

$$\frac{f'(c_x)}{g'(c_x)}\Big(1 - \frac{g(b_0)}{g(x)}\Big) \quad = \quad \frac{f(x)}{g(x)} - \frac{f(b_0)}{g(x)},$$

and therefore to

$$\frac{f(x)}{g(x)} \quad = \quad \frac{f'(c_x)}{g'(c_x)}\Big(1 - \frac{g(b_0)}{g(x)}\Big) + \frac{f(b_0)}{g(x)}$$

as well. Since $\lim\limits_{x \to a_0^+} g(x) = \pm\infty$, whereas $g(b_0)$ and $f(b_0)$ are fixed constants, it is clear that

$$\lim_{x \to a_0^-} \Big(1 - \frac{g(b_0)}{g(x)}\Big) = 1 \qquad \text{and} \qquad \lim_{x \to a_0^-} \frac{f(b_0)}{g(x)} = 0.$$

Setting $\varepsilon_1 := \frac{1}{2}$, we now fix $\delta_1 \in (0, \delta_0)$ such that for all $x \in (a_0, a_0 + \delta_1)$, we have that $|(1 - \frac{g(b_0)}{g(x)}) - 1| < \varepsilon_1$, that is, $|\frac{g(b_0)}{g(x)}| < \frac{1}{2}$, which in turn implies that $1 - \frac{g(b_0)}{g(x)} > \frac{1}{2}$.[9] Similarly, fix $\delta_2 \in (0, \delta_0)$ such that for all $x \in (a_0, a_0 + \delta_2)$, we have that $|\frac{f(b_0)}{g(x)} - 0| < \varepsilon_1$, that is, $|\frac{f(b_0)}{g(x)}| < \frac{1}{2}$, so that $\frac{f(b_0)}{g(x)} > -\frac{1}{2}$.

---

[9] Indeed, since $|\frac{g(b_0)}{g(x)}| < \frac{1}{2}$, we have that $-\frac{1}{2} < \frac{g(b_0)}{g(x)} < \frac{1}{2}$, and consequently, $-\frac{1}{2} < -\frac{g(b_0)}{g(x)} < \frac{1}{2}$. It follows that $\frac{1}{2} < 1 - \frac{g(b_0)}{g(x)} < \frac{3}{2}$.

Now, set $\hat{\delta} := \min\{\delta_1, \delta_2\} < \delta_0$, and fix any $x \in (a_0, a_0 + \hat{\delta})$. Note that this means, in particular, that $x \in (a_0, a_0 + \delta_0)$, and consequently, $\frac{f'(c_x)}{f'(c_x)} > 2M + \frac{1}{2} > 0$. Moreover, we have that $1 - \frac{g(b_0)}{g(x)} > \frac{1}{2}$ and $\frac{f(b_0)}{g(x)} > -\frac{1}{2}$. We now compute:

$$\frac{f(x)}{g(x)} \quad = \quad \frac{f'(c_x)}{g'(c_x)}\Big(1 - \frac{g(b_0)}{g(x)}\Big) + \frac{f(b_0)}{g(x)} \quad > \quad (2M+1) \cdot \frac{1}{2} - \frac{1}{2} \quad = \quad M.$$

This proves that $\lim\limits_{x \to a_0^+} \frac{f(x)}{g(x)} = +\infty$, and we are done. ☐

L'Hôpital's Rule for left-hand limits. *Let $a_0 \in \mathbb{R}$ and $\delta > 0$, and assume that functions $f, g : (a_0 - \delta, a_0) \to \mathbb{R}$ are both differentiable, and that $g'(x) \neq 0$ for all $x \in (a_0 - \delta, a_0)$. Assume moreover that one of the following holds:*

(i) $\lim\limits_{x \to a_0^-} f(x) = \lim\limits_{x \to a_0^-} g(x) = 0$*;*

(ii) $\lim\limits_{x \to a_0^-} f(x) = \pm\infty$ *and* $\lim\limits_{x \to a_0^-} g(x) = \pm\infty$.

Then, if $\lim\limits_{x \to a_0^-} \frac{f'(x)}{g'(x)}$ *exists (either as a real number or as* $\pm\infty$*), then* $\lim\limits_{x \to a_0^-} \frac{f(x)}{g(x)}$ *exists as well, and in this case, we have that*

$$\lim_{x \to a_0^-} \frac{f(x)}{g(x)} \quad = \quad \lim_{x \to a_0^-} \frac{f'(x)}{g'(x)}.$$

Proof. Obviously, this could be proven analogously to L'Hôpital's Rule for right-hand limits. However, we can also simply use L'Hôpital's Rule for right-hand limits (which we have already proven), as follows. First of all, assume that $L := \lim\limits_{x \to a_0^-} \frac{f'(x)}{g'(x)}$ exists (either as a real number or as $\pm\infty$). Now, note that for all $x \in \mathbb{R}$, we have that $x \in (a_0, a_0 + \delta)$ if and only if $2a_0 - x \in (a_0 - \delta, a_0)$. Define auxiliary functions $f_1, g_1 : (a_0, a_0 + \delta) \to \mathbb{R}$ by setting $f_1(x) = f(2a_0 - x)$ and $g_1(x) = g(2a_0 - x)$ for all $x \in (a_0, a_0 + \delta)$. First, since f and g are differentiable, so are f_1 and g_1.[10] Next, note that for $x \in (a_0, a_0 + \delta)$, we have that

$$\begin{array}{rcll} g_1'(x) & = & \frac{d}{dx} g(2a_0 - x) & \\ & = & g'(2a_0 - x) \frac{d}{dx}(2a_0 - x) & \text{by the Chain Rule} \\ & = & g'(2a_0 - x)(-1) & \\ & = & -g'(2a_0 - x) \quad \neq \quad 0. & \end{array}$$

---

[10] Indeed consider the function $h : (a_0, a_0 + \delta) \to \mathbb{R}$ given by $h(x) = 2a_0 - x$. Clearly, h is differentiable, and we have that $f_1 = f \circ h$ and $g_1 = g \circ h$. Since f, g, and h are all differentiable, so are f_1 and g_1.

Further, we have that

$$\begin{array}{rcccl} \lim\limits_{x\to a_0^+} f_1(x) & = & \lim\limits_{x\to a_0^+} f(2a_0-x) & \overset{(*)}{=} & \lim\limits_{y\to a_0^-} f(y), \\ \\ \lim\limits_{x\to a_0^+} g_1(x) & = & \lim\limits_{x\to a_0^+} g(2a_0-x) & \overset{(*)}{=} & \lim\limits_{y\to a_0^-} g(y), \end{array}$$

where both instances of (*) were obtained via the substitution $y = 2a_0 - x$. Since (i) or (ii) holds, we see that one of the following hold:

(i') $\lim\limits_{x\to a_0^+} f_1(x) = \lim\limits_{x\to a_0^+} g_1(x) = 0$;

(ii') $\lim\limits_{x\to a_0^+} f_1(x) = \pm\infty$ and $\lim\limits_{x\to a_0^+} g_1(x) = \pm\infty$.

Moreover, we have the following:

$$\begin{array}{rcll} \lim\limits_{x\to a_0^+} \frac{f_1'(x)}{g_1'(x)} & = & \lim\limits_{x\to a_0^+} \frac{\frac{d}{dx}f(2a_0-x)}{\frac{d}{dx}g(2a_0-x)} & \\ \\ & = & \lim\limits_{x\to a_0^+} \frac{f'(2a_0-x)\cdot\frac{d}{dx}(2a_0-x)}{g'(2a_0-x)\frac{d}{dx}(2a_0-x)} & \text{by the Chain Rule} \\ \\ & = & \lim\limits_{x\to a_0^+} \frac{f'(2a_0-x)\cdot(-1)}{g'(2a_0-x)\cdot(-1)} & \\ \\ & = & \lim\limits_{x\to a_0^+} \frac{f'(2a_0-x)}{g'(2a_0-x)} & \\ \\ & = & \lim\limits_{y\to a_0^-} \frac{f'(y)}{g'(y)} & \text{via substitution } y = 2a_0 - x \\ \\ & = & L. & \end{array}$$

So, by L'Hôpital's Rule for right-hand limits, we have that $\lim\limits_{x\to a_0^+} \frac{f_1(x)}{g_1(x)} = L$. It now follows that

$$\lim_{x\to a_0^-} \frac{f(x)}{g(x)} \overset{(*)}{=} \lim_{y\to a_0^+} \frac{f(2a_0-y)}{g(2a_0-y)} = \lim_{y\to a_0^+} \frac{f_1(y)}{g_1(y)} = L,$$

where (*) was obtained via substitution $y = 2a_0 - x$. This completes the argument. □

L'Hôpital's Rule for limits at a point. *Let $a_0 \in \mathbb{R}$ and $\delta > 0$, and assume that functions $f, g : (a_0-\delta, a_0) \cup (a_0, a_0+\delta) \to \mathbb{R}$ are both differentiable, and that*

$g'(x) \neq 0$ for all $x \in (a_0-\delta, a_0) \cup (a_0, a_0+\delta)$. Assume moreover that one of the following holds:

(i) $\lim\limits_{x \to a_0} f(x) = \lim\limits_{x \to a_0} g(x) = 0$*;*

(ii) $\lim\limits_{x \to a_0} f(x) = \pm\infty$ *and* $\lim\limits_{x \to a_0} g(x) = \pm\infty$*.*

Then, if $\lim\limits_{x \to a_0} \frac{f'(x)}{g'(x)}$ *exists (either as a real number or as* $\pm\infty$*), then* $\lim\limits_{x \to a_0} \frac{f(x)}{g(x)}$ *exists as well, and in this case, we have that*

$$\lim_{x \to a_0} \tfrac{f(x)}{g(x)} \quad = \quad \lim_{x \to a_0} \tfrac{f'(x)}{g'(x)}.$$

Proof. This follows immediately from L'Hopital's Rules for right-hand and left-hand limits, together with Lemma 3.6.3. □

L'Hôpital's Rule for limits at $+\infty$. *Let $a_0 \in \mathbb{R}$, and assume that functions $f, g : (a_0, +\infty) \to \mathbb{R}$ are both differentiable, and that $g'(x) \neq 0$ for all $x \in (a_0, +\infty)$. Assume moreover that one of the following holds:*

(i) $\lim\limits_{x \to +\infty} f(x) = \lim\limits_{x \to +\infty} g(x) = 0$*;*

(ii) $\lim\limits_{x \to +\infty} f(x) = \pm\infty$ *and* $\lim\limits_{x \to +\infty} g(x) = \pm\infty$*.*

Then, if $\lim\limits_{x \to +\infty} \frac{f'(x)}{g'(x)}$ *exists (either as a real number or as* $\pm\infty$*), then* $\lim\limits_{x \to +\infty} \frac{f(x)}{g(x)}$ *exists as well, and in this case, we have that*

$$\lim_{x \to +\infty} \tfrac{f(x)}{g(x)} \quad = \quad \lim_{x \to +\infty} \tfrac{f'(x)}{g'(x)}.$$

Proof. Clearly, we may assume that $a_0 > 0$ (if not, we simply restrict the domain of f and g to, for example, the open interval $(1, +\infty)$, and we consider 1 instead of a_0).

We assume that $L := \lim\limits_{x \to +\infty} \frac{f(x)}{g(x)}$ exists (either as a real number or as $\pm\infty$). Define $\hat{f}, \hat{g} : (0, \frac{1}{a_0}) \to \mathbb{R}$ by setting

$$\hat{f}(x) = f(\tfrac{1}{x}) \qquad \text{and} \qquad \hat{g}(x) = g(\tfrac{1}{x})$$

for all $x \in (0, \frac{1}{a_0})$. First, since f and g are differentiable, so are $\hat{f}$ and $\hat{g}$.[11] Next, note that for $x \in (0, \frac{1}{a_0})$, we have that

---

[11] Indeed, consider the function $h : (0, \frac{1}{a_0}) \to \mathbb{R}$ given by $h(x) = \frac{1}{x}$ for all $x \in (0, \frac{1}{a_0})$. Clearly, h is differentiable. Moreover, we have that $\hat{f} = f \circ h$ and $\hat{g} = g \circ h$. Since f, g, and h are all differentiable, so are $\hat{f}$ and $\hat{g}$.

$$\begin{array}{rcll} \hat{g}'(x) & = & \frac{d}{dx}g(\frac{1}{x}) & \\ & = & g'(\frac{1}{x})\frac{d}{dx}(\frac{1}{x}) & \text{by the Chain Rule} \\ & = & g'(\frac{1}{x})(-\frac{1}{x^2}) & \\ & = & -\frac{g'(\frac{1}{x})}{x^2} \quad \neq \quad 0. & \end{array}$$

Further, we have that

$$\begin{array}{rcccl} \lim\limits_{x\to 0^+} \hat{f}(x) & = & \lim\limits_{x\to 0^+} f(\frac{1}{x}) & \overset{(*)}{=} & \lim\limits_{y\to +\infty} f(y), \\ \lim\limits_{x\to 0^+} \hat{g}(x) & = & \lim\limits_{x\to 0^+} g(\frac{1}{x}) & \overset{(*)}{=} & \lim\limits_{y\to +\infty} g(y), \end{array}$$

where both instances of (*) were obtained via the substitution $y = \frac{1}{x}$. Since (i) or (ii) holds, we now have that at least one of the following holds:

(i') $\lim\limits_{x\to 0^+} \hat{f}(x) = \lim\limits_{x\to 0^+} \hat{g}(x) = 0$;

(ii') $\lim\limits_{x\to 0^+} \hat{f}(x) = \pm\infty$ and $\lim\limits_{x\to 0^+} \hat{g}(x) = \pm\infty$.

Moreover, we have the following:

$$\begin{array}{rcll} \lim\limits_{x\to 0^+} \frac{\frac{d}{dx}\hat{f}(x)}{\frac{d}{dx}\hat{g}(x)} & = & \lim\limits_{x\to 0^+} \frac{\frac{d}{dx}f(\frac{1}{x})}{\frac{d}{dx}g(\frac{1}{x})} & \\ & = & \lim\limits_{x\to 0^+} \frac{f'(\frac{1}{x})\cdot(-x^{-2})}{g'(\frac{1}{x})\cdot(-x^{-2})} & \text{by the Chain Rule} \\ & = & \lim\limits_{x\to 0^+} \frac{f'(\frac{1}{x})}{g'(\frac{1}{x})} & \\ & = & \lim\limits_{y\to +\infty} \frac{f'(y)}{g'(y)} & \text{via substitution } y = \frac{1}{x} \\ & = & L. & \end{array}$$

But now by L'Hôpital's Rule for right-hand limits, we have that

$$\lim_{x\to 0^+} \frac{\hat{f}(x)}{\hat{g}(x)} \quad = \quad L,$$

and we deduce that

$$\lim_{x\to +\infty} \frac{f(x)}{g(x)} \quad = \quad \lim_{y\to 0^+} \frac{f(\frac{1}{y})}{g(\frac{1}{y})} \quad = \quad \lim_{y\to 0^+} \frac{\hat{f}(y)}{\hat{g}(y)} \quad = \quad L.$$

This completes the argument. ☐

L'Hôpital's Rule for limits at $-\infty$. *Let $a_0 \in \mathbb{R}$, and assume that functions $f, g : (-\infty, a_0) \to \mathbb{R}$ are both differentiable, and that $g'(x) \neq 0$ for all $x \in (-\infty, a_0)$. Assume moreover that one of the following holds:*

(i) $\lim\limits_{x \to -\infty} f(x) = \lim\limits_{x \to -\infty} g(x) = 0$;

(ii) $\lim\limits_{x \to -\infty} f(x) = \pm\infty$ *and* $\lim\limits_{x \to -\infty} g(x) = \pm\infty$.

Then, if $\lim\limits_{x \to -\infty} \frac{f'(x)}{g'(x)}$ *exists (either as a real number or as $\pm\infty$), then* $\lim\limits_{x \to -\infty} \frac{f(x)}{g(x)}$ *exists as well, and in this case, we have that*

$$\lim_{x \to -\infty} \tfrac{f(x)}{g(x)} \quad = \quad \lim_{x \to -\infty} \tfrac{f'(x)}{g'(x)}.$$

Proof. The proof is analogous to that of L'Hôpital's Rule at $+\infty$. Alternatively, L'Hôpital's Rule for limits at $-\infty$ can be obtained using l'Hôpital's Rule for limits at $+\infty$, just we obtained l'Hôpital's Rule for left-hand limits from the one for right-hand limits. Either way, the details are left as an exercise. ☐

4.12 Higher order derivatives

Suppose that $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is a function, and that A' is the set of all points in A that f is differentiable at. Then we can treat $f' : A' \to \mathbb{R}$ as a function in its own right.[12] If A'' is the set of all points in A' that f' is differentiable at, then we can treat $f'' : A'' \to \mathbb{R}$ as a function in its own right, and we can repeat the process. Here, $f'' = (f')'$. The function f'' is called the *second derivative* of f, and likewise $f''' := (f'')'$ is the *third derivative* of f (the domain of f''' is the set of all points at which f'' is differentiable). Meanwhile, f' is simply the *first derivative* of f.

Example 4.12.1. *Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = \sin x + \cos x$ for all $x \in \mathbb{R}$. Then for $x \in \mathbb{R}$, we have:*

- $f'(x) = \cos x - \sin x$;
- $f''(x) = -\sin x - \cos x$;
- $f'''(x) = -\cos x + \sin x$.

Beyond the third derivative, it is unusual to keep adding primes (visually, it is difficult to distinguish between f'''' and f'''''). So, we use the notation $f^{(n)}$ instead. Let us be more formal and give a proper recursive definition. Suppose that $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function. We define $f^{(0)} := f$. Next, for $n \in \mathbb{N}_0$, if B is the

---

[12] It is possible that $A' = \emptyset$. In this case, f' is simply the "empty function."

set of all points at which $f^{(n)}$ is differentiable, then we define $f^{(n+1)} : B \to \mathbb{R}$ to be $f^{(n+1)} = (f^{(n)})'$.

The Leibniz notation for higher-order derivatives. Sometimes, instead of $f^{(n)}(x)$, we write $\frac{d^n}{x^n} f(x)$ or $\frac{d^n f}{x^n}(x)$. If we are given $y = f(x)$, then we may write $\frac{d^n y}{dx^x}$ instead of $f^{(n)}(x)$. Notation $\frac{d^n}{x^n} f(x)$, $\frac{d^n f}{x^n}(x)$, and $\frac{d^n y}{dx^x}$ is referred to as the *Leibniz notation.*

Higher order differentiability. Suppose that $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is a function. For $n \in \mathbb{N}$, we say that f is *n times differentiable* at a point $a \in A$ if $f^{(n)}(a)$ exists. For $n \in \mathbb{N}$, f is *n times differentiable* if $f^{(n)}(a)$ exists for all $a \in A$. The function f is *infinitely differentiable* (or *smooth*) if f is n times differentiable for all $n \in \mathbb{N}$. Now, suppose that I is an open interval. For $n \in \mathbb{N}$, we say that f is *n times differentiable on I* if $f^{(n)}(a)$ exists for all $a \in I$. The function f is *infinitely differentiable on I* (or *smooth on I*) if f is n times differentiable on I for all $n \in \mathbb{N}$.

Continuous differentiability. A function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is *continuously differentiable* if it is differentiable,[13] and in addition, $f' : A \to \mathbb{R}$ is continuous. For $n \in \mathbb{N}$, f is *n times continuously differentiable* if f is n times differentiable, and moreover, $f^{(n)} : A \to \mathbb{R}$ is continuous.

Remark: By Theorem 4.3.1, all differentiable functions are continuous. Thus, if a function f is $(n + 1)$ times differentiable (for some $n \in \mathbb{N}$), then f must be n times continuously differentiable; however, the converse is false in general.

Example 4.12.2. *Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by*

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

for all $x \in \mathbb{R}$. Show that f is differentiable, but not continuously differentiable (and consequently, f is not twice differentiable).

Solution. For $x \in \mathbb{R} \setminus \{0\}$, we have that

$$\begin{aligned} f'(x) &= \tfrac{d}{dx}\Big(x^2 \sin\left(\tfrac{1}{x}\right)\Big) \\ &= 2x \sin\left(\tfrac{1}{x}\right) + x^2 \cos\left(\tfrac{1}{x}\right) \cdot \left(-\tfrac{1}{x^2}\right) \\ &= 2x \sin\left(\tfrac{1}{x}\right) - \cos\left(\tfrac{1}{x}\right), \end{aligned}$$

whereas

$$f'(0) \;=\; \lim_{x \to 0} \tfrac{f(x) - f(0)}{x - 0} \;=\; \lim_{x \to 0} \tfrac{x^2 \sin\left(\frac{1}{x}\right) - 0}{x} \;=\; \lim_{x \to 0} x \sin\left(\tfrac{1}{x}\right) \;\overset{(*)}{=}\; 0,$$

---

[13] So, f is differentiable at all points in its domain A.

where (*) follows via an easy application of the Squeeze Theorem. Thus, f is differentiable. However, note that $\lim\limits_{x \to 0} f'(x)$ does not exist. Indeed, $\lim\limits_{x \to 0} 2x \sin\left(\frac{1}{x}\right) = 0$ (via an easy application of the Squeeze Theorem), but $\cos\left(\frac{1}{x}\right)$ keeps oscillating as x approaches 0; so, $\lim\limits_{x \to 0} f'(x)$ does not exist.[14] So, f' is not continuous at 0, and consequently, f' is not continuously differentiable. Since differentiable functions are continuous (by Theorem 4.3.1), it follows that f' is not twice differentiable. □

4.13 Local extrema: the First and Second Derivative Tests

4.13.1 Derivatives and monotonicity

The *interior* of a set $A \subseteq \mathbb{R}$, denoted by A° or $\text{int}(A)$, is the set of all points $a \in A$ for which there exists some $\delta > 0$ such that $(a - \delta, a + \delta) \subseteq A$. In other words, a point $a \in A$ belongs to the interior of A if some open interval containing a is included in A. Any subset of $\mathbb{R}$ has a (possibly empty) interior. Here, we are interested in intervals. So, for $a, b \in \mathbb{R}$ such that $a < b$, we have the following:

- the interior of (a, b) is (a, b);
- the interior of $[a, b)$ is (a, b);
- the interior of $(a, b]$ is (a, b);
- the interior of $[a, b]$ is (a, b);
- the interior of $(-\infty, b)$ is $(-\infty, b)$;
- the interior of $(-\infty, b]$ is $(-\infty, b)$;
- the interior of $(a, +\infty)$ is $(a, +\infty)$;
- the interior of $[a, +\infty)$ is $(a, +\infty)$;
- the interior of $(-\infty, +\infty)$ is $(-\infty, +\infty)$.

So, informally, the interior of an interval is the open interval that we obtain by removing the "edges" (if any) of the original interval.

Corollary 4.10.3 can be strengthened as follows (with a nearly identical proof).

Theorem 4.13.1. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, and let $I \subseteq A$ be an interval. Assume that f is continuous on I and differentiable on I°. Then all the following hold:*

---

[14] Try to give a formal proof of the fact that $\lim\limits_{x \to 0} f'(x)$ does not exist!

(a) if $f'(x) = 0$ *for all* $x \in I^\circ$*, then* f *constant on* I*;*

(b) if $f'(x) \geq 0$ *for all* $x \in I^\circ$*, then* f *is non-decreasing on* I*;*

(c) if $f'(x) > 0$ *for all* $x \in I^\circ$*, then* f *is strictly increasing on* I*;*

(d) if $f'(x) \leq 0$ *for all* $x \in I^\circ$*, then* f *is non-increasing on* I*;*

(e) if $f'(x) < 0$ *for all* $x \in I^\circ$*, then* f *is strictly decreasing on* I*.*

Proof. We prove (a) and (b). The proofs of (c), (d), and (e) are similar to the proof of (b).

We first prove (a). Assume that $f'(x) = 0$ for all $x \in I^\circ$. Fix $a, b \in I$; we must show that $f(a) = f(b)$. We may assume that $a \neq b$, for otherwise we are done. By symmetry, we may further assume that $a < b$. Then $[a, b] \subseteq I$ and $(a, b) \subseteq I^0$, and consequently, f is continuous on $[a, b]$ and differentiable on (a, b). We now apply Lagrange's Mean Value Theorem to f and the interval $[a, b]$ to obtain a point $c \in (a, b)$ such that $f'(c) = \frac{f(b)-f(a)}{b-a}$. But since $c \in (a, b) \subseteq I^\circ$, we have that $f'(c) = 0$. So, $\frac{f(b)-f(a)}{b-a} = 0$, and consequently, $f(a) = f(b)$. This proves (a).

We now prove (b). Assume that $f'(x) \geq 0$ for all $x \in I^\circ$. Fix $a, b \in I$ such that $a \leq b$; we must show that $f(a) \leq f(b)$. We may assume that $a \neq b$ (and therefore $a < b$), for otherwise we are done. Then $[a, b] \subseteq I$ and $(a, b) \subseteq I^0$, and consequently, f is continuous on $[a, b]$ and differentiable on (a, b). We now apply Lagrange's Mean Value Theorem to f and the interval $[a, b]$ to obtain a point $c \in (a, b)$ such that $f'(c) = \frac{f(b)-f(a)}{b-a}$. But since $c \in (a, b) \subseteq I^\circ$, we have that $f'(c) > 0$. So, $\frac{f(b)-f(a)}{b-a} > 0$, and consequently, $f(b) - f(a) > 0$, i.e. $f(a) < f(b)$. This proves (b). ☐

4.13.2 The First Derivative Test

The First Derivative Test. *Let* $f : A \subseteq \mathbb{R} \to \mathbb{R}$ *be a function, and let* $a, b, c \in \mathbb{R}$ *be such that* $a < c < b$ *and* $(a, b) \subseteq A$*. Assume that* f *is continuous on the open interval* (a, b) *and differentiable on the open intervals* (a, c) *and* (c, b)*.*[15]

(a) If both the following hold:

- $f'(x) \leq 0$ *for all* $x \in (a, c)$*,*
- $f'(x) \geq 0$ *for all* $x \in (c, b)$*,*

then f *reaches a local minimum at* c*.*

(b) If both the following hold:

- $f'(x) \geq 0$ *for all* $x \in (a, c)$*,*

---

[15] The function f may or may not be differentiable on the point c itself. However, by hypothesis, f is defined and continuous at c.

- *$f'(x) \leq 0$ for all $x \in (c, b)$,*

then f reaches a local maximum at c.

Proof. We first prove (a). Suppose that $f'(x) \leq 0$ for all $x \in (a, c)$, and that $f'(x) \geq 0$ for all $x \in (c, b)$. Then Theorem 4.13.1 guarantees that f is non-increasing on $(a, c]$ and non-decreasing on $[c, b)$, and it follows that f reaches a local minimum at c.

The proof of (b) is similar. Indeed, suppose that $f'(x) \geq 0$ for all $x \in (a, c)$, and that $f'(x) \leq 0$ for all $x \in (c, b)$. Then Theorem 4.13.1 guarantees that f is non-decreasing on $(a, c]$ and non-increasing on $[c, b)$, and it follows that f reaches a local maximum at c. ☐

Remark: Schematically, the First Derivative Test (and its proof) can be represented by the two tables below.

a c b

x	(a, c)	(c, b)
$f'(x)$	$-$	$+$
$f(x)$	↘ min	↗

a c b

x	(a, c)	(c, b)
$f'(x)$	$+$	$-$
$f(x)$	↗ max	↘

Technically, we would use this sort of table representation for the case when the inequalities from the First Derivative Test are strict, i.e. when we have $f'(x) > 0$ and $f'(x) < 0$, rather than merely $f'(x) \geq 0$ and $f'(x) \leq 0$.

Proposition 4.13.2. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, and let $a, b, c \in R$ be such that $a < c < b$ and $(a, b) \subseteq A$. Assume that f is continuous on the open interval (a, b) and differentiable on the open intervals (a, c) and (c, b).*

(a) If $f'(x) > 0$ for all $x \in (a, c) \cup (c, b)$, then f is strictly increasing on (a, b), and in particular, f does not reach a local extremum at c.

(b) If $f'(x) < 0$ for all $x \in (a, c) \cup (c, b)$, then f is strictly decreasing on (a, b), and in particular, f does not reach a local extremum at c.

Remark: The proposition applies even if f is not differentiable at c. However, for the proposition to hold, it is indeed necessary for f to be continuous on the whole interval (a, b), and in particular, to be defined at and continuous at the point c.

Proof. For (a), we observe that if $f'(x) > 0$ for all $x \in (a, c) \cup (c, b)$, then Theorem 4.13.1 guarantees that f is strictly increasing on the intervals $(a, c]$ and $[c, b)$, and consequently, it is strictly increasing on the interval (a, b) as well.

The proof of (b) is analogous. Indeed, if $f'(x) < 0$ for all $x \in (a, c) \cup (c, b)$, then Theorem 4.13.1 guarantees that f is strictly decreasing on the intervals $(a, c]$ and $[c, b)$, and consequently, it is strictly decreasing on the interval (a, b) as well. ☐

Remark: Schematically, Proposition 4.13.2 (and its proof) can be represented by the table below.

	a		c		b
x		(a,c)		(c,b)	
$f'(x)$		$+$		$+$	
$f(x)$		$\nearrow$		$\nearrow$	

	a		c		b
x		(a,c)		(c,b)	
$f'(x)$		$-$		$-$	
$f(x)$		$\searrow$		$\searrow$	

For a concrete example of this sort of behavior, consider the functions $f, g : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^3$ and $g(x) = -x^3$ for all $x \in \mathbb{R}$. Then $f'(x) = 2x^2$ and $g'(x) = -2x^2$. So, for $c = 0$, we have that $f'(x) = g'(c) = 0$. However, for $x \in \mathbb{R} \setminus \{0\}$, we have $f'(x) > 0$ and $g'(x) < 0$. Consistently, with Proposition 4.13.2, the function f is strictly increasing, whereas the function g is strictly decreasing.

Example 4.13.3. *Consider the function $f : \mathbb{R} \to \mathbb{R}$, given by $f(x) = 3x^5 - 5x^3 + 7$ for all $x \in \mathbb{R}$. Find all points x at which f reaches a local minimum or a local maximum, and compute $f(x)$ for those values of x. Does f reach a global maximum or minimum? What are the minimum and maximum values of f (if they exist)?*

Solution. First of all, we observe that the function f is continuous and differentiable on $\mathbb{R}$, and that

$$f'(x) \quad = \quad 15x^4 - 15x^2 \quad = \quad 15x^2(x-1)(x+1)$$

for all $x \in \mathbb{R}$. Thus, $f'(x) = 0$ for $x \in \{-1, 0, 1\}$ (and for no other value of x). Thus, by Theorem 4.10.2, the only values of x at which f may possibly reach a local extremum are $x = -1$, $x = 0$, and $x = 1$ (and we must check at which, if any, of those values of x the function f indeed reaches a local minimum or a local maximum). Now, consider the table below.

		-1		0		1	
x	$(-\infty,-1)$		$(-1,0)$		$(0,1)$		$(1,+\infty)$
x^2	$+$		$+$	0	$+$		$+$
$x-1$	$-$		$-$		$-$	0	$+$
$x+1$	$-$	0	$+$		$+$		$+$
$f'(x)$	$+$	0	$-$	0	$-$	0	$+$
$f(x)$	$\nearrow$	max	$\searrow$		$\searrow$	min	$\nearrow$

Remark: The local maximum and the local minimum in the table above were identified via the First Derivative Test. Meanwhile, by Proposition 4.13.2, f does **not** reach a local extremum at $x = 0$.

We can now summarize:

- f reaches a local maximum at $x = -1$, and we have that $f(-1) = 9$;
- f reaches a local minimum at $x = 1$, and we have that $f(1) = 5$;
- f has no other local extrema.

Meanwhile, we have the following:

- $\lim\limits_{x \to -\infty} f(x) = -\infty$;
- $\lim\limits_{x \to +\infty} f(x) = +\infty$.

Therefore, f reaches neither a global minimum nor a global maximum, and consequently, f has neither a minimum nor a maximum value.

Remark: Using computer software, we can graph our function f (see below). Note that this graph is indeed consistent with our solution.

□

Example 4.13.4. *Consider the function* $f : \mathbb{R} \to \mathbb{R}$*, given by* $f(x) = \sqrt[3]{x^2}$*. Find all points* x *at which* f *reaches a local minimum or a local maximum, and compute* $f(x)$ *for those values of* x*. Does* f *reach a global minimum or a global maximum? What are the minimum and maximum values of* f *(if they exist)?*

Solution. The function f is defined and continuous on $\mathbb{R}$, and it is differentiable on $\mathbb{R} \setminus \{0\} = (-\infty, 0) \cup (0, +\infty)$. For $x \in \mathbb{R} \setminus \{0\}$, we have that

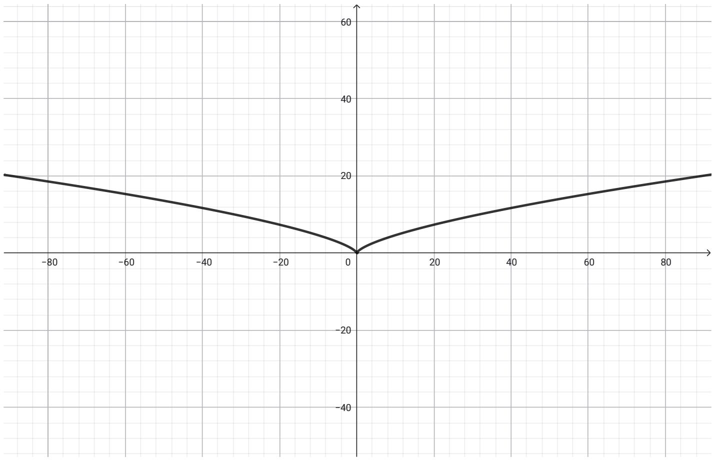
$$f'(x) = \tfrac{2}{3\sqrt[3]{x}}$$

for all $x \in \mathbb{R}$. So, we have the table below.

x	$(-\infty, 0)$	0	$(0, +\infty)$
$f'(x)$	$-$	undef	$+$
$f(x)$	↘	min	↗

So, f reaches a local minimum at $x = 0$, and we have that $f(0) = 0$. The function f has no other local extrema. Moreover, the function f is strictly decreasing on $(-\infty, 0)$ and strictly increasing on $(0, +\infty)$. So, f reaches a global minimum at $x = 0$, and the minimum value of f is $f(0) = 0$. The function f does not reach a global maximum, and consequently, f does not have a maximum value.

Remark: Using computer software, we can graph our function f (see below). Note that this graph is indeed consistent with our solution.



$\square$

4.13.3 The Second Derivative Test

The Second Derivative Test. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, and let $a, b, c \in \mathbb{R}$ be such that $a < c < b$ and $(a, b) \subseteq A$. Assume that f is differentiable on the interval (a, b), that $f'(c) = 0$, and that $f''(c)$ exists.*

(a) If $f''(c) > 0$, then f reaches a local minimum at c.

(b) If $f''(c) < 0$, then f reaches a local maximum at c.

Proof. We prove (a); the proof of (b) is similar. So, let us assume that $f''(c) > 0$. We will exhibit some $\delta > 0$ such that $(c - \delta, c + \delta) \subseteq (a, b)$, and such that the following hold:

- $f'(x) < 0$ for all $x \in (c - \delta, c)$;

- $f'(x) > 0$ for all $x \in (c, c+\delta)$.

This is enough, for the First Derivative Test will then imply that f reaches a local minimum at c, which is what we need to show.

First, note that

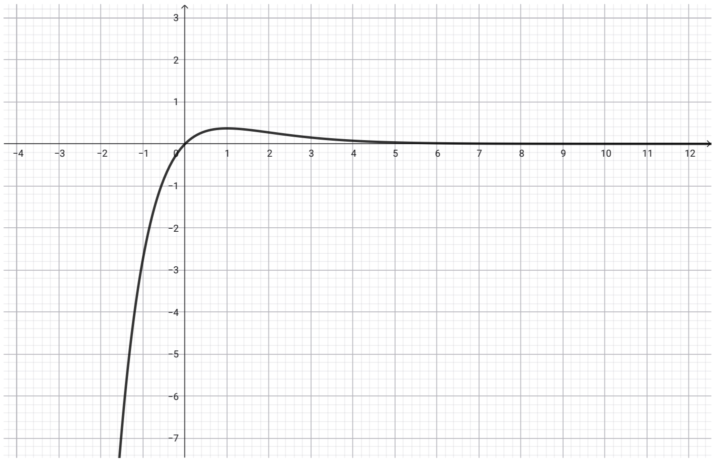
$$f''(c) \overset{(*)}{=} \lim_{x \to c} \tfrac{f'(x)-f'(c)}{x-c} \overset{(**)}{=} \lim_{x \to c} \tfrac{f'(x)}{x-c},$$

where (*) follows from the definition of the derivative, whereas (**) follows from the fact that $f'(c) = 0$. Now, set $\varepsilon := \frac{f''(c)}{2}$; since $f''(c) > 0$, we have that $\varepsilon > 0$. Fix $\delta > 0$ such that for all $x \in (a,b)$, if $0 < |x-c| < \delta$, then $|\frac{f'(x)}{x-c} - f''(c)| < \varepsilon$. Since $c \in (a,b)$, we may assume that $(c-\delta, c+\delta) \subseteq (a,b)$, for otherwise, we simply choose a smaller δ. So, for all $x \in (c-\delta, c) \cup (c, c+\delta)$, we have that $|\frac{f'(x)}{x-c} - f''(c)| < \varepsilon$, which is equivalent to $\frac{f''(c)}{2} = f''(c) - \varepsilon < \frac{f'(x)}{x-c} < f''(c) + \varepsilon = \frac{3f''(c)}{2}$, which in turn implies that $\frac{f'(x)}{x-c} > \frac{f''(c)}{2} > 0$. Therefore, for any $x \in (c-\delta, c)$, we have that $f'(x) < 0$ (because $x - c < 0$), whereas for any $x \in (c, c+\delta)$, we have that $f'(x) > 0$ (because $x - c > 0$). □

Example 4.13.5. *Find all the local minima and local maxima (if any) of the function* $f : \mathbb{R} \to \mathbb{R}$ *given by* $f(x) = xe^{-x}$ *for all* $x \in \mathbb{R}$.

Solution. Note that $f'(x) = (1-x)e^{-x}$ and $f''(x) = (x-2)e^{-x}$ for all $x \in \mathbb{R}$. Since $f'(x)$ is defined for all $x \in \mathbb{R}$, and since we have $f'(x) = 0$ only for $x = 1$, we see that the only "candidate" for a local extremum is $x = 1$. Since $f''(1) = -e^{-1} < 0$, the Second Derivative Test guarantees that f reaches a local maximum at $c = -1$ (with $f(-1) = -e^{-1}$).

Remark: Using computer software, we can graph our function f (see below). Note that this graph is indeed consistent with our solution.



□

Remark: In practice, the First Derivative Test is used much more frequently than the Second Derivative Test.

4.13.4 Finding maxima and minima of functions on closed intervals

Suppose that $f : [a, b] \to \mathbb{R}$ is a continuous function, where $a, b \in \mathbb{R}$ are such that $a < b$. Then by the Extreme Value Theorem, f reaches both a global maximum and a global minimum on $[a, b]$. While there is no algorithm for finding the points at which f reaches its global maximum and minimum, there is a strategy that will work if f is "reasonably well behaved." The strategy is as follows.

1. We find all the points $c \in (a, b)$ such that either $f'(c) = 0$ or $f'(c)$ does not exist. These points c are the "candidates" for points at which f might reach a local minimum or a local maximum on the interval (a, b).[16]
2. We compute $f(c)$ for all values of c that we found in the previous step, and we compute $f(a)$ and $f(b)$.
3. By comparing all the values that we computed in the previous step, we find the maximum and minimum values of f on the interval $[a, b]$.

Example 4.13.6. *Find the maximum and the minimum value of the function* $f(x) = \sin^3 x + \cos^3 x$ *on the interval* $[-\frac{\pi}{4}, \frac{3\pi}{4}]$.

Solution. Note that $f'(x) = 3 \sin x \cos x(\sin x - \cos x)$ for all $x \in \mathbb{R}$.[17] Note that, for $x \in (-\frac{\pi}{4}, \frac{3\pi}{4})$, we have that $f'(x) = 0$ if and only if $x \in \{0, \frac{\pi}{4}, \frac{\pi}{2}\}$. So, our "candidates" for points at which f might reach its global maximum or its global minimum are the following points: $-\frac{\pi}{4}, 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$. We now compute:

- $f(-\frac{\pi}{4}) = \sin^3(-\frac{\pi}{4}) + \cos^3(-\frac{\pi}{4}) = (-\frac{1}{\sqrt{2}})^3 + (\frac{1}{\sqrt{2}})^3 = 0;$
- $f(0) = \sin^3(0) + \cos^3(0) = 0^3 + 1^3 = 1;$
- $f(\frac{\pi}{4}) = \sin^3(\frac{\pi}{4}) + \cos^3(\frac{\pi}{4}) = (\frac{1}{\sqrt{2}})^3 + (\frac{1}{\sqrt{2}})^3 = \frac{1}{\sqrt{2}};$
- $f(\frac{\pi}{2}) = \sin^3(\frac{\pi}{2}) + \cos^3(\frac{\pi}{2}) = 1^0 + 0^3 = 1;$
- $f(\frac{3\pi}{4}) = \sin^3(\frac{3\pi}{4}) + \cos^3(\frac{3\pi}{4}) = (\frac{1}{\sqrt{2}})^3 + (-\frac{1}{\sqrt{2}})^3 = 0.$

So, f reaches a global minimum on the interval $[-\frac{\pi}{4}, \frac{3\pi}{4}]$ at the points $m_1 = -\frac{\pi}{4}$ and $m_2 = \frac{3\pi}{4}$, and the minimum value of f on this interval is $f(m_1) = f(m_2) = 0$. On the other hand, f reaches a global maximum on the interval $[-\frac{\pi}{4}, \frac{3\pi}{4}]$ at the points $M_1 = 0$ and $M_2 = \frac{\pi}{2}$, and the maximum value of f on this interval is $f(M_1) = f(M_2) = 1$.

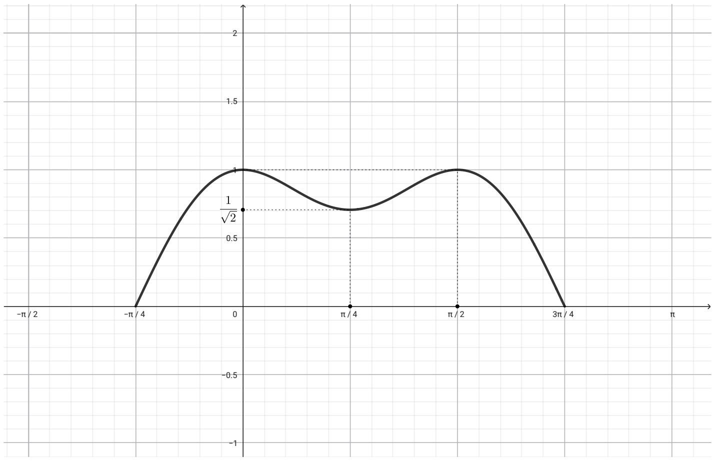
---

[16] This is the step where our strategy might fall apart if f is "poorly behaved." The implicit assumption is that there are only finitely many such points c, and that we know how to find them.

[17] In particular, f is differentiable everywhere on $\mathbb{R}$.

The minimum value of f on the interval $[-\frac{\pi}{4}, \frac{3\pi}{4}]$ is 0, and the maximum value of f on this interval is 1.

Remark: Using computer software, we get the graph below. This is consistent with our calculation above.



□

4.14 Convexity and concavity

4.14.1 Closed intervals

Recall from algebra that for $x_1, x_2 \in \mathbb{R}$ such that $x_1 < x_2$, the closed interval $[x_1, x_2]$ is given by

$$[x_1, x_2] \quad = \quad \big\{\lambda_1 x_1 + \lambda_2 x_2 \mid \lambda_1, \lambda_2 \in \mathbb{R},\ \lambda_1, \lambda_2 \geq 0,\ \lambda_1 + \lambda_2 = 1\big\}.$$

In fact, for $x \in [x_1, x_2]$, we have that

$$x \quad = \quad \tfrac{x_2 - x}{x_2 - x_1} x_1 + \tfrac{x - x_1}{x_2 - x_1} x_2,$$

and so if we set

$$\lambda_1 := \tfrac{x_2 - x}{x_2 - x_1} \qquad \text{and} \qquad \lambda_2 := \tfrac{x - x_1}{x_2 - x_1},$$

then we get $\lambda_1, \lambda_2 \geq 0$, $\lambda_1 + \lambda_2 = 1$, and $x = \lambda_1 x_1 + \lambda_2 x_2$.

4.14.2 Convex and concave functions

For a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ and an interval $I \subseteq A$, we say that:

- f is *convex* on the interval I provided that for all $x_1, x_2 \in I$, and all real numbers $\lambda_1, \lambda_2 \geq 0$ such that $\lambda_1 + \lambda_2 = 1$, we have that

$$f\big(\lambda_1 x_1 + \lambda_2 x_2\big) \quad \leq \quad \lambda_1 f(x_1) + \lambda_2 f(x_2);$$

- f is *strictly convex* on the interval I provided that for all distinct $x_1, x_2 \in I$, and all real numbers $\lambda_1, \lambda_2 > 0$ such that $\lambda_1 + \lambda_2 = 1$, we have that

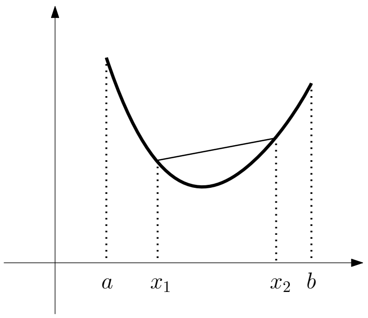
$$f\big(\lambda_1 x_1 + \lambda_2 x_2\big) \quad < \quad \lambda_1 f(x_1) + \lambda_2 f(x_2);$$

- f is *concave* on the interval I provided that for all $x_1, x_2 \in I$, and all real numbers $\lambda_1, \lambda_2 \geq 0$ such that $\lambda_1 + \lambda_2 = 1$, we have that

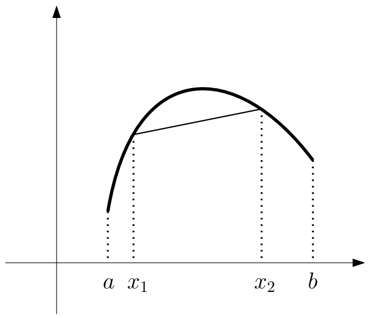
$$f\big(\lambda_1 x_1 + \lambda_2 x_2\big) \quad \geq \quad \lambda_1 f(x_1) + \lambda_2 f(x_2);$$

- f is *strictly concave* on the interval I provided that for all distinct $x_1, x_2 \in I$, and all real numbers $\lambda_1, \lambda_2 > 0$ such that $\lambda_1 + \lambda_2 = 1$, we have that

$$f\big(\lambda_1 x_1 + \lambda_2 x_2\big) \quad > \quad \lambda_1 f(x_1) + \lambda_2 f(x_2).$$



convex



concave

4.14.3 Affine functions

A function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is *affine* on an interval $I \subseteq A$ if there exist real numbers $m, b \in \mathbb{R}$ such that

$$f(x) \quad := \quad mx + b \qquad \forall x \in I.$$

Proposition 4.14.1. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, and let $I \subseteq A$ be an interval. Then the following are equivalent:*

- *f is both convex and concave on I;*
- *f is affine on I.*

Moreover, in this case, f is neither strictly convex nor strictly concave.

Proof. Suppose first that f is affine on I, and fix $m, b \in \mathbb{R}$ such that $f(x) = mx + b$ for all $x \in I$. Now fix any $x_1, x_2 \in I$, and fix real numbers $\lambda_1, \lambda_2 \geq 0$ such that $\lambda_1 + \lambda_2 = 1$. Then

$$
\begin{aligned}
f\big(\lambda_1 x_1 + \lambda_2 x_2\big) &= m\big(\lambda_1 x_1 + \lambda_2 x_2\big) + b \\
&= m\big(\lambda_1 x_1 + \lambda_2 x_2\big) + \underbrace{(\lambda_1 + \lambda_2)}_{=1} b \\
&= \lambda_1(m x_1 + b) + \lambda_2(m x_1 + b) \\
&= \lambda_1 f(x_1) + \lambda_2 f(x_2),
\end{aligned}
$$

and it immediately follows that f is both convex and concave. To see that f is neither strictly convex nor strictly concave, we simply choose distinct x_1, x_2, and we choose strictly positive λ_1, λ_2 for our calculation above.

Suppose conversely, that f is both convex and concave on I.

> **Claim.** For all $x_1, x_2 \in I$, and all real numbers $\lambda_1, \lambda_2 \geq 0$ such that $\lambda_1 + \lambda_2 = 1$, we have that $f\big(\lambda_1 x_1 + \lambda_2 x_2\big) = \lambda_1 f(x_1) + \lambda_2 f(x_2)$.

Proof of the Claim. This follows immediately from the fact that f is both convex and concave (we simply apply the definition of convexity and concavity). ♦

Now, fix arbitrary $x_1, x_2 \in I$ such that $x_1 < x_2$. Set $m := \frac{f(x_2) - f(x_1)}{x_2 - x_1}$ and $b := \frac{x_2 f(x_1) - x_1 f(x_2)}{x_2 - x_1}$, so that $y = mx + b$ is the line through the points $\big(x_1, f(x_1)\big)$ and $\big(x_2, f(x_2)\big)$.[18] In particular,

$$
f(x_1) = m x_1 + b \qquad \text{and} \qquad f(x_2) = m x_2 + b.
$$

We claim that $f(x) = mx + b$ for all $x \in I$; this will imply that f is affine on I, which is what we need to show.

Fix $x_0 \in I$. There are three cases to consider: when $x_1 \leq x_0 \leq x_2$; when $x_0 < x_1 < x_2$; and when $x_1 < x_2 < x_1$. We consider the first two cases; the third case is similar to the second.

Suppose first that $x_1 \leq x_0 \leq x_2$. Set $\lambda_1 := \frac{x_2 - x_0}{x_2 - x_1}$ and $\lambda_2 := \frac{x_0 - x_1}{x_2 - x_1}$, so that $\lambda_1, \lambda_2 \geq 0$, $\lambda_1 + \lambda_2 = 1$, and $x_0 = \lambda_1 x_1 + \lambda_2 x_2$. Then

$$
\begin{aligned}
f(x_0) &= f\big(\lambda_1 x_1 + \lambda_2 x_2\big) \\
&= \lambda_1 f(x_1) + \lambda_2 f(x_2) && \text{by the Claim} \\
&= \lambda_1(m x_1 + b) + \lambda_2(m x_2 + b)
\end{aligned}
$$

---

[18] Here, we simply need a line $y = mx+b$ that passes through the points $\big(x_1, f(x_1)\big)$ and $\big(x_2, f(x_2)\big)$. By solving for m and b, we get $m = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$ and $b = \frac{x_2 f(x_1) - x_1 f(x_2)}{x_2 - x_1}$. However, in what follows, we will not actually need those exact formulas for m and b. We will simply use the fact that f passes through $\big(x_1, f(x_1)\big)$ and $\big(x_2, f(x_2)\big)$.

$$\begin{aligned} &= m(\underbrace{\lambda_1 x_1 + \lambda_2 x_2}_{=x_0}) + (\underbrace{\lambda_1 + \lambda_2}_{=1})b \\ &= mx_0 + b, \end{aligned}$$

which is what we needed to show.

Suppose now that $x_0 < x_1 < x_2$. Set $\lambda_0 := \frac{x_2-x_1}{x_2-x_0}$ and $\lambda_2 := \frac{x_1-x_0}{x_2-x_0}$, so that $\lambda_0, \lambda_2 > 0$, $\lambda_0 + \lambda_2 = 1$, and $x_1 = \lambda_0 x_0 + \lambda_2 x_2$. Then

$$f(x_1) \;=\; f\big(\lambda_0 x_0 + \lambda_2 x_2\big) \;\overset{(*)}{=}\; \lambda_0 f(x_0) + \lambda_2 f(x_2),$$

where (*) follows from the Claim. Since $\lambda_0 \neq 0$, we may solve for $f(x_0)$:

$$f(x_0) \;=\; \tfrac{1}{\lambda_0}\Big(f(x_1) - \lambda_2 f(x_2)\Big).$$

We now compute:

$$\begin{aligned} f(x_0) &= \tfrac{1}{\lambda_0}\Big(f(x_1) - \lambda_2 f(x_2)\Big) \\ &= \tfrac{x_2-x_0}{x_2-x_1}\Big((mx_1+b) - \tfrac{x_1-x_0}{x_2-x_0}(mx_2+b)\Big) \\ &= \tfrac{x_2-x_0}{x_2-x_1} \cdot \tfrac{(x_2-x_0)(mx_1+b)-(x_1-x_0)(mx_2+b)}{x_2-x_0} \\ &= \tfrac{(x_2-x_0)(mx_1+b)-(x_1-x_0)(mx_2+b)}{x_2-x_1} \\ &= \tfrac{(x_2-x_0)mx_1+(x_2-x_0)b-(x_1-x_0)mx_2-(x_1-x_0)b}{x_2-x_1} \\ &= \tfrac{\big((x_2-x_0)x_1-(x_1-x_0)x_2\big)m+\big((x_2-x_0)-(x_1-x_0)\big)b}{x_2-x_1} \\ &= \tfrac{x_0(x_2-x_1)m+(x_2-x_1)b}{x_2-x_1} \\ &= mx_0 + b, \end{aligned}$$

and we are done. □

4.14.4 The first and second derivative in the context of convexity and concavity

Theorem 4.14.2. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, differentiable on an open interval $I \subseteq A$. Then all the following hold:*

(a) f' is non-decreasing on I if and only if f is convex on I;

(b) f' is strictly increasing on I if and only if f is strictly convex on I;

(c) f' is non-increasing on I if and only if f is concave on I;

(d) f' is strictly decreasing on I if and only if f is strictly concave on I.

Proof. First, since f is differentiable on I, Theorem 4.3.1 guarantees that f is continuous on I. We begin with a technical claim, which relies only on elementary algebraic manipulations (and uses neither differentiability nor continuity), and which is useful for both implications of each of the four parts of the theorem.

> **Claim 1.** Let $\star \in \{\leq, <, \geq, >\}$. Then for all $x, x_1, x_2 \in A$ such that $x_1 < x < x_2$, the following two inequalities are equivalent:
>
> - $f(x) \star \frac{x_2-x}{x_2-x_1} f(x_1) + \frac{x-x_1}{x_2-x_1} f(x_2)$;
> - $\frac{f(x)-f(x_1)}{x-x_1} \star \frac{f(x_2)-f(x)}{x_2-x}$.

Proof of Claim 1. We have the following sequence of equivalent inequalities:

$$\begin{array}{crcl} & f(x) & \star & \frac{x_2-x}{x_2-x_1} f(x_1) + \frac{x-x_1}{x_2-x_1} f(x_2) \\[2ex] \overset{(*)}{\iff} & (x_2-x_1)f(x) & \star & (x_2-x)f(x_1) + (x-x_1)f(x_2) \\[2ex] \iff & (x_2-x+x-x_1)f(x) & \star & (x_2-x)f(x_1) + (x-x_1)f(x_2) \\[2ex] \iff & (x_2-x)f(x) + (x-x_1)f(x) & \star & (x_2-x)f(x_1) + (x-x_1)f(x_2) \\[2ex] \iff & (x_2-x)\big(f(x)-f(x_1)\big) & \star & (x-x_1)\big(f(x_2)-f(x)\big) \\[2ex] \overset{(**)}{\iff} & \frac{(x_2-x)}{(x_2-x)(x-x_1)}\big(f(x)-f(x_1)\big) & \star & \frac{(x-x_1)}{(x_2-x)(x-x_1)}\big(f(x_2)-f(x)\big) \\[2ex] \iff & \frac{f(x)-f(x_1)}{x-x_1} & \star & \frac{f(x_2)-f(x)}{x_2-x} \end{array}$$

where (*) was obtained by multiplying both sides of the inequality by $x_2 - x_1 > 0$, and (**) was obtained by dividing both sides of the inequality by $(x_2-x)(x-x_1) > 0$. This proves the Claim. ♦

Now, we prove (a) and (b); the proofs of (c) and (d) are analogous.

We first prove the "if" ("$\Longleftarrow$") part of (a). Suppose that f is convex on I; we must show that f' is non-decreasing on I. Fix $x_1, x_2 \in I$ such that $x_1 \leq x_2$; we must show that $f'(x_1) \leq f'(x_2)$. We may assume that $x_1 \neq x_2$ (and therefore, $x_1 < x_2$).

Claim 2. For all $x \in (x_1, x_2)$, we have that $\frac{f(x)-f(x_1)}{x-x_1} \leq \frac{f(x_2)-f(x)}{x_2-x}$.

Proof of Claim 2. Fix $x \in (x_1, x_2)$, and set

$$\lambda_1 = \tfrac{x_2-x}{x_2-x_1} \qquad \text{and} \qquad \lambda_2 = \tfrac{x-x_1}{x_2-x_1},$$

Clearly, $\lambda_1, \lambda_2 > 0$, $\lambda_1 + \lambda_2 = 1$, and $x = \lambda_1 x_1 + \lambda_2 x_2$. Since f is convex on I, it follows that $f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2)$, i.e. that

$$f(x) \quad \leq \quad \tfrac{x_2-x}{x_2-x_1} f(x_1) + \tfrac{x-x_1}{x_2-x_1} f(x_2).$$

Claim 1 (with $\star$ replaced by $\leq$) now implies that

$$\tfrac{f(x)-f(x_1)}{x-x_1} \quad \leq \quad \tfrac{f(x_2)-f(x)}{x_2-x},$$

which is what we needed to show. ♦

We now compute:

$$\begin{aligned}
f'(x_1) \quad &= \quad \lim_{x \to x_1} \tfrac{f(x)-f(x_1)}{x-x_1} && \text{by the definition of the derivative} \\
&= \quad \lim_{x \to x_1^+} \tfrac{f(x)-f(x_1)}{x-x_1} && \\
&\leq \quad \lim_{x \to x_1^+} \tfrac{f(x_2)-f(x)}{x_2-x} && \text{by Claim 2} \\
&= \quad \tfrac{f(x_2)-f(x_1)}{x_2-x_1} && \text{because } f \text{ is continuous at } x_1 \\
&= \quad \lim_{x \to x_2} \tfrac{f(x)-f(x_1)}{x-x_1} && \text{because } f \text{ is continuous at } x_2 \\
&= \quad f'(x_2) && \text{by the definition of the derivative.}
\end{aligned}$$

So, $f'(x_1) \leq f'(x_2)$, and it follows that f' is non-decreasing on I. This proves the "if" part of (a).

Remark: If f were strictly convex, then we would get a strict inequality in Claim 2. However, in our application of Claim 2 in the third line of the computation above, we would still only get a weak inequality ("$\leq$"). So, the proof of the "if" ("$\Longleftarrow$") part of (b) is **not** analogous to the above, and we will need to deal with it in a different way, as follows.

We now prove the "if" ("$\Longleftarrow$") part of (b). So, assume that f is strictly convex on I; we must show that f' is strictly increasing on I. Suppose otherwise, and fix $x_1, x_2 \in I$ such that $x_1 < x_2$, but $f(x_1) \geq f(x_2)$. However, by (a), f' is non-decreasing on I, and it follows that $f'(x_1) = f'(x_2) =: m$, and moreover, that f' is constant on the closed interval $[x_1, x_2]$, that is, that $f'(x) = m$ for all $x \in [x_1, x_2]$. By Lagrange's Mean Value Theorem applied to the function f and the interval $[x_1, x_2]$, there exists some $c \in (x_1, x_2)$ such that $f'(c) = \frac{f(x_2)-f(x_1)}{x_2-x_1}$, and we deduce that $m = \frac{f(x_2)-f(x_1)}{x_2-x_1}$, i.e. m is the slope of the line through the points $\big(x_1, f(x_1)\big)$ and $\big(x_2, f(x_2)\big)$. We now set $b := \frac{x_2 f(x_1)-x_1 f(x_2)}{x_2-x_1}$, so that

$$f(x_1) = mx_1 + b \qquad \text{and} \qquad f(x_2) = mx_2 + b,$$

i.e. $y = mx + b$ is the equation of the line through the points $\big(x_1, f(x_1)\big)$ and $\big(x_2, f(x_2)\big)$.

Claim 3. For all $x \in [x_1, x_2]$, we have that $f(x) = mx + b$.

Proof of Claim 3. Fix $x_0 \in [x_1, x_2]$; we must show that $f(x_0) = mx_0 + b$. If $x_0 = x_1$ or $x_0 = x_2$, then we are done. So, assume that $x_0 \in (x_1, x_2)$. By Lagrange's Mean Value Theorem applied to the function f and the interval $[x_0, x_2]$, there exist some $c \in (x_0, x_2)$ such that

$$f'(c) \quad = \quad \tfrac{f(x_2)-f(x_0)}{x_2-x_0}.$$

Since $f'(c) = m$ and $f(x_2) = mx_2 + b$, we deduce that

$$m \quad = \quad \tfrac{(mx_2+b)-f(x_0)}{x_2-x_0},$$

which immediately implies that $f(x_0) = mx_0 + b$. ♦

By Claim 3, f is affine on $[x_1, x_2]$. So, by Proposition 4.14.1, f is **not** strictly convex on $[x_1, x_2]$. Since $[x_1, x_2] \subseteq I$, it follows that f is **not** strictly convex on I either, a contradiction. This proves the "if" part of (b).

We now prove the "only if" ("$\Longrightarrow$") part of (a). So, suppose that f' is non-decreasing on I; we must show that f is convex on I. Fix $x_1, x_2 \in I$, and fix $\lambda_1, \lambda_2 \geq 0$ such that $\lambda_1 + \lambda_2 = 1$; we must show that $f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2)$.

First, if $x_1 = x_2$, then the result is immediate.[19] We may therefore assume that $x_1 \neq x_2$. By symmetry, we may further assume that $x_1 \leq x_2$. Next, if $\lambda_1 = 0$ (so

---

[19] Indeed, suppose that $x_1 = x_2$. Then we have the following sequence of equivalent inequalities:

$$\begin{array}{rrcl} & f(\lambda_1 x_1 + \lambda_2 x_2) & \leq & \lambda_1 f(x_1) + \lambda_2 f(x_2) \\ \iff & f(\lambda_1 x_1 + \lambda_2 x_1) & \leq & \lambda_1 f(x_1) + \lambda_2 f(x_1) \\ \iff & f\big(\underbrace{(\lambda_1 + \lambda_2)}_{=1} x_1\big) & \leq & \underbrace{(\lambda_1 + \lambda_2)}_{=1} f(x_1) \\ \iff & f(x_1) & \leq & f(x_1) \end{array}$$

Since the last inequality is obviously true, so is the first one.

that $\lambda_2 = 1$), or if $\lambda_2 = 0$ (so that $\lambda_1 = 1$), then the result is immediate. We may therefore assume that $0 < \lambda_1, \lambda_2 < 1$.

Now, set $x := \lambda_1 x_1 + \lambda_2 x_2$, and note that we have that $x_1 < x < x_2$. Solving for λ_1 and λ_2,[20] we get that

$$\lambda_1 = \tfrac{x_2 - x}{x_2 - x_1} \qquad \text{and} \qquad \lambda_2 = \tfrac{x - x_1}{x_2 - x_1}.$$

By applying Lagrange's Mean Value Theorem to f and the interval $[x_1, x]$, we get that there exists some $c_1 \in (x_1, x)$ such that

$$f'(c_1) \quad = \quad \tfrac{f(x) - f(x_1)}{x - x_1}.$$

Similarly, by applying Lagrange's Mean Value Theorem to f and the interval $[x, x_2]$, we get that there exists some $c_2 \in (x, x_2)$ such that

$$f'(c_2) \quad = \quad \tfrac{f(x_2) - f(x)}{x_2 - x}.$$

We have that $x_1 < c_1 < x < c_2 < x_2$, and that $x_1, x_2 \in I$. Since f' is non-decreasing on the open interval I, it follows that $f'(c_1) \leq f'(c_2)$, i.e. that

$$\tfrac{f(x) - f(x_1)}{x - x_1} \quad \leq \quad \tfrac{f(x_2) - f(x)}{x_2 - x}.$$

By Claim 1, this is equivalent to

$$f(x) \quad \leq \quad \tfrac{x_2 - x}{x_2 - x_1} f(x_1) + \tfrac{x - x_1}{x_2 - x_1} f(x_2),$$

i.e. $f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2)$. This proves that f is convex on I. We have now proven the "only if" part of (a). The proof of the "only if" part of (b) is analogous. □

Corollary 4.14.3. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, differentiable on an open interval $I \subseteq A$. Then all the following hold:*

(a) f is convex on I if and only if $f''(x) \geq 0$ for all $x \in I$;

(b) if $f''(x) > 0$ for all $x \in I$, then f is strictly convex on I;

(c) f is concave on I if and only if $f''(x) \leq 0$ for all $x \in I$.

(d) if $f''(x) < 0$ for all $x \in I$, then f is strictly concave on I.

---

[20] Here, we simply solve the system

$$\begin{array}{rcl} \lambda_1 x_1 + \lambda_2 x_2 & = & x \\ \lambda_1 + \lambda_2 & = & 1 \end{array}$$

for λ_1 and λ_2.

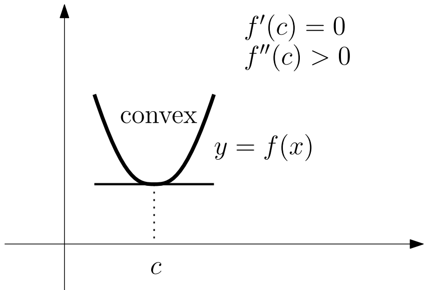
Remark: The converses of (b) and (d) are false. For instance, $f_1(x) = x^4$ is strictly convex on any open interval, but $f_1''(0) = 0$. Similarly, $f_2(x) = -x^4$ is strictly concave on any open interval, but $f_2''(0) = 0$.

Proof. We prove (a) and (b); the proofs of (c) and (d) are analogous.

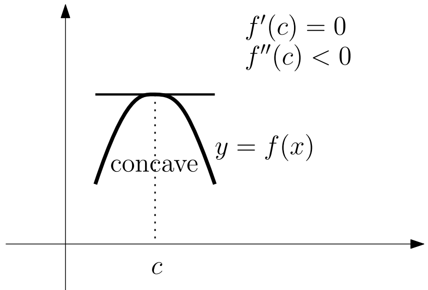
We first prove (a). By Theorem 4.14.2, f is convex on I if and only if f' is non-decreasing on I. But by Corollary 4.10.3 and Proposition 4.10.4, f' is non-decreasing on I if and only if $f''(x) \geq 0$ for all $x \in I$. This proves (a).

For (b), we observe that if $f''(x) > 0$ for all $x \in I$, then Corollary 4.10.3 implies that f' is strictly increasing, and so by Theorem 4.14.2, f is strictly convex. $\square$

Remark: Now that we have linked the second derivative to convexity and concavity, the Second Derivative Test should make intuitive sense (see the picture below).



local minimum



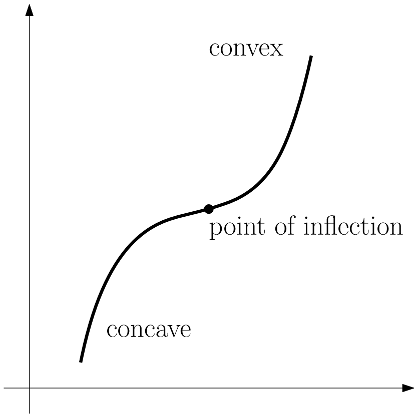
local maximum

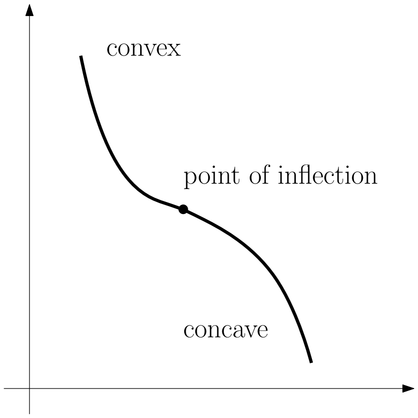
4.14.5 Points of inflection

A "point of inflection" is a point at which a function switches from being convex to being concave, or vice versa. Let us be more precise. For a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$, a point $c \in A$ is said to be a *point of inflection* (or an *inflection point*) of f if f is continuous at c, and there exists some $\delta > 0$ such that all the following hold:

- $(c - \delta, c + \delta) \subseteq A$;
- f is convex on one of the open intervals $(c - \delta, c)$ and $(c, c + \delta)$, and is concave on the other;
- for all $\delta' \in (0, \delta)$, f is affine neither on $(c - \delta', c)$ nor on $(c, c + \delta')$.

In view of Proposition 4.14.1, the third bullet point simply states that f is not simultaneously convex and concave on any small interval on one side (to the left or to the right) of c. So, points of inflection mark a "true switch" from convexity to concavity or vice versa.





Proposition 4.14.4. *Let* $f : A \subseteq \mathbb{R} \to \mathbb{R}$ *be a function, let* $c \in A$*, and let* $\delta > 0$ *be such that* $(c-\delta, c+\delta) \subseteq A$*. Assume that* f *is continuous at* c*, and is twice differentiable on* $(c-\delta, c)$ *and* $(c, c+\delta)$*. If one of the following holds:*

(i) $f''(x) > 0$ *for all* $x \in (c-\delta, c)$*, and* $f''(x) < 0$ *for all* $x \in (c, c+\delta)$*,*

(ii) $f''(x) < 0$ *for all* $x \in (c-\delta, c)$*, and* $f''(x) > 0$ *for all* $x \in (c, c+\delta)$*,*

then c *is a point of inflection of* f*.*

Proof. In view of Theorem 4.14.3, we have the following:

- if (i) holds, then f switches from being strictly convex to being strictly concave at c;
- if (ii) holds, then f switches from being strictly concave to being strictly convex at c.

In either case, c is a point of inflection of f.[21] □

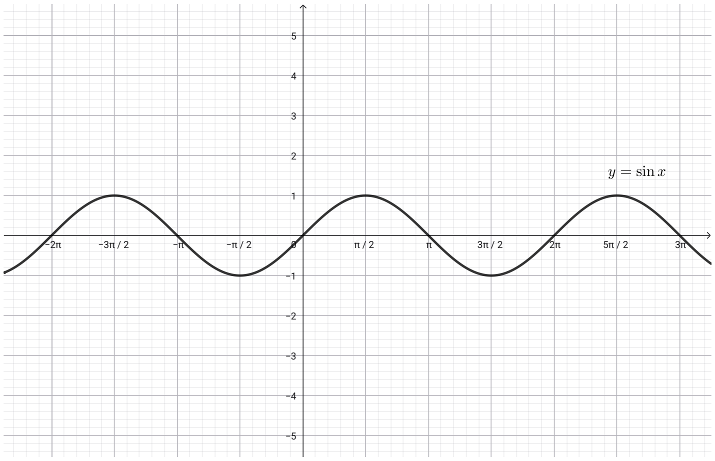
Example 4.14.5. *Consider the function* $\sin x$*. Note that* $\sin'(x) = \cos x$ *and* $\sin''(x) = -\sin x$ *for all* $x \in \mathbb{R}$*. Thus, for* $k \in \mathbb{Z}$*, the following hold:*

- $\sin''(k\pi) = 0$*;*
- $\sin''(x) > 0$ *for all* $x \in \big((2k-1)\pi, 2k\pi\big)$*;*
- $\sin''(x) < 0$ *for all* $x \in \big(2k\pi, (2k+1)\pi\big)$*.*

The points of inflection of $\sin x$ *are all the points of the form* $x = k\pi$ *(*$k \in \mathbb{Z}$*).*

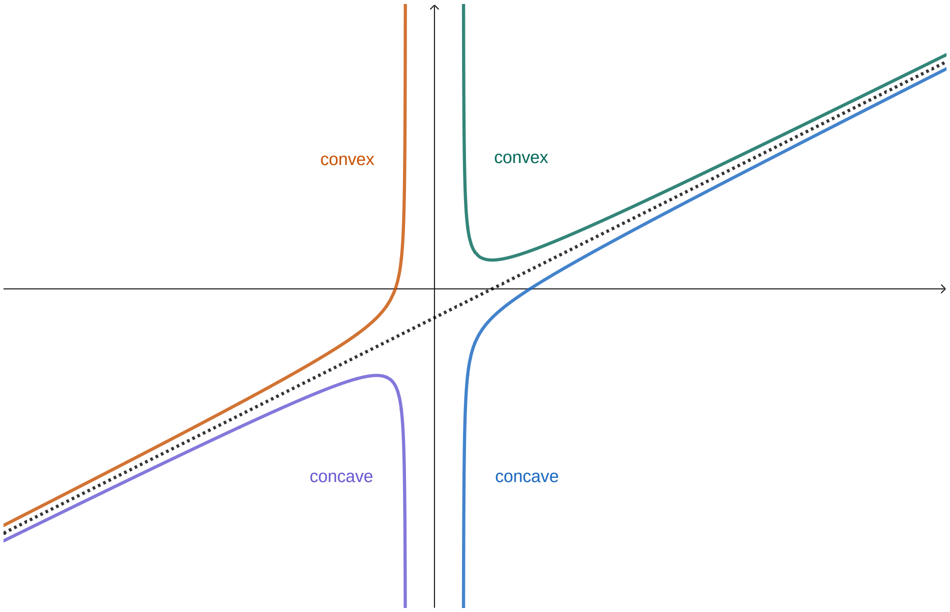
---

[21] Note that we are also using the fact that affine functions are neither strictly convex nor strictly concave. This follows immediately from the relevant definitions, or alternatively, it follows from Proposition 4.14.1.



4.14.6 Convexity and concavity in the context of asymptotes

In this subsection, we show that convex functions lie above their horizontal/slant asymptotes, whereas concave functions lie below them (see the picture below). This fact is sometimes helpful for sketching graphs of functions.



Let us outline the proof. First, Lemma 4.14.6 deals with the special case when $y = 0$ (i.e. the x-axis) is a horizontal asymptote of our function f as $x \to +\infty$. Next, Theorem 4.14.7 deals with the case when our function f has a horizontal/slant asymptote $y = ax + b$ as $x \to +\infty$.[22] To prove the theorem, we define an auxiliary

---

[22] If $a = 0$, then $y = ax + b$ (i.e. $y = b$) is a horizontal asymptote of f as $x \to +\infty$. On the other hand, if $a \neq 0$, then $y = ax + b$ is a slant asymptote of f as $x \to +\infty$.

function $g(x) := f(x) - (ax + b)$. Then $y = 0$ is a horizontal asymptote of g as $x \to +\infty$, and the convexity/concavity of g is the same as that of f; we apply Lemma 4.14.6 to g, and the result for f follows. Finally, Theorem 4.14.8 deals with the case when our function f has a horizontal/slant asymptote $y = ax + b$ as $x \to -\infty$. To prove the theorem, we define an auxiliary function $h(x) := f(-x)$; we note that the graph of h is obtained by reflecting the graph of f about the y-axis, and that $y = -ax + b$ is a horizontal/slant asymptote of h as $x \to +\infty$. We apply Theorem 4.14.7 to h, and the result for f readily follows. We now turn to the technical details.

Lemma 4.14.6. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, and let $c \in \mathbb{R}$ be such that $(c, +\infty) \subseteq A$. Assume that $y = 0$ (i.e. the x-axis) is a horizontal asymptote of f as $x \to +\infty$. Then all the following hold:*

(a) if f is convex on $(c, +\infty)$, then $f(x) \geq 0$ for all $x \in (c, +\infty)$;

(b) if f is strictly convex on $(c, +\infty)$, then $f(x) > 0$ for all $x \in (c, +\infty)$;

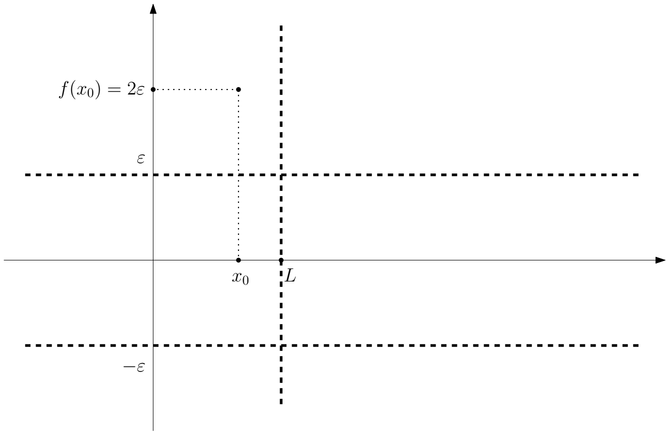
(c) if f is concave on $(c, +\infty)$, then $f(x) \leq 0$ for all $x \in (c, +\infty)$;

(d) if f is strictly concave on $(c, +\infty)$, then $f(x) < 0$ for all $x \in (c, +\infty)$.

Proof. We first prove (c). We prove the contrapositive: we assume that $f(x_0) > 0$ for some $x_0 \in (c, +\infty)$, and we prove that f is not concave on $(c, +\infty)$.

Set $\varepsilon := \frac{f(x_0)}{2}$; clearly, $\varepsilon > 0$. Using the fact that $y = 0$ is a horizontal asymptote of f as $x \to +\infty$, we fix $L \in \mathbb{R}$ such that for all $x \in A$, if $x > L$, then $|f(x) - 0| < \varepsilon$. Clearly, we may assume that $L > x_0$, for otherwise, we simply choose a larger L. So, for all $x \in (L, +\infty)$, we have that $|f(x) - 0| < \varepsilon$, that is, $-\varepsilon < f(x) < \varepsilon$.

Remark: Our set-up so far is represented in the picture below. The part of the graph of the function f to the right of the vertical dashed line (i.e. line $x = L$) lies strictly between the two horizontal dashed lines (i.e. lines $y = -\varepsilon$ and $y = \varepsilon$).



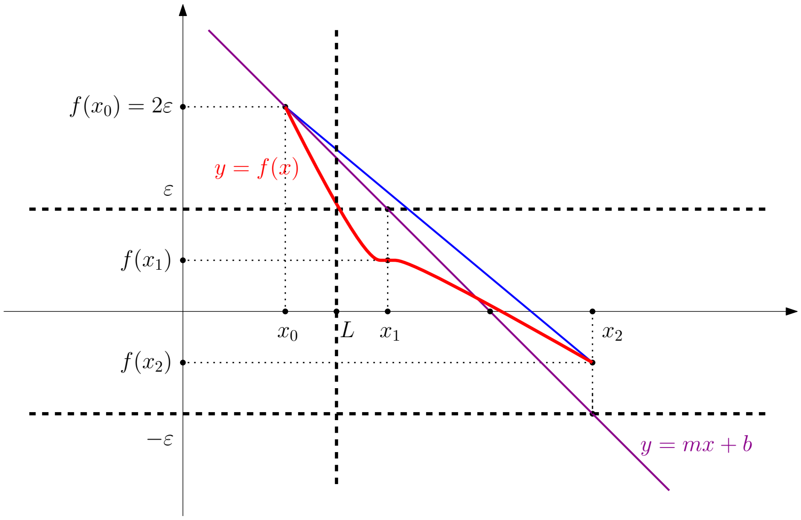
Now, fix any $x_1 \in (L, +\infty)$, and set $m = -\frac{\varepsilon}{x_1 - x_0}$ and $b = \frac{2x_1 - x_0}{x_1 - x_0}\varepsilon$, so that $y = mx + b$ is the line through the points $\big(x_0, f(x_0)\big) = (x_0, 2\varepsilon)$ and (x_1, ε).[23] Set $x_2 := 3x_1 - 2x_0$, so that the line $y = mx + b$ passes through $(x_2, -\varepsilon)$.[24] Note that

$$x_1 = \tfrac{2}{3}x_0 + \tfrac{1}{3}x_2 \qquad \text{and} \qquad c < x_0 < L < x_1 < x_2.$$

Further, since $x_1, x_2 > L$, we have that

$$-\varepsilon \quad < \quad f(x_1), f(x_2) \quad < \quad \varepsilon.$$

Remark: Our set-up is represented in the picture below. The formal argument is given below the picture, but the idea is as follows. The points $\big(x_1, f(x_1)\big)$ and $\big(x_2, f(x_2)\big)$ are both (strictly) between the two horizontal dashed lines. Therefore, the point $\big(x_1, f(x_1)\big)$ is below the line $y = mx + b$, whereas the point $\big(x_2, f(x_2)\big)$ is above it. This forces the point $\big(x_1, f(x_1)\big)$ to be below the line segment connecting points $\big(x_0, f(x_0)\big)$ and $\big(x_2, f(x_2)\big)$, which means that f is not concave.



Here are the technical details. We compute:

$$\tfrac{2}{3}f(x_0) + \tfrac{1}{3}f(x_2) \overset{(*)}{>} \tfrac{2}{3}\cdot 2\varepsilon - \tfrac{1}{3}\varepsilon \;=\; \varepsilon \;>\; f(x_1) \;=\; f\big(\tfrac{2}{3}x_0 + \tfrac{1}{3}x_2\big),$$

---

[23] We obtain these values of m and b simply by solving the system

$$\begin{array}{rcr} mx_0 + b & = & 2\varepsilon \\ mx_1 + b & = & \varepsilon \end{array}$$

for m and b.

[24] We obtained this value of x_2 simply by solving the equation $mx_2 + b = -\varepsilon$ for x_2 (and then plugging in our values for m and b).

where (*) follows from the fact that $\varepsilon = \frac{f(x_0)}{2}$ and $f(x_2) > -\varepsilon$. We have now shown that

$$f\big(\tfrac{2}{3}x_0 + \tfrac{1}{3}x_2\big) \quad < \quad \tfrac{2}{3}f(x_0) + \tfrac{1}{3}f(x_2),$$

and so f is not concave on $(c, +\infty)$. This proves (c).

We now prove (d). Suppose that f is strictly concave. By (c), we know that $f(x) \leq 0$ for all $x \in (c, +\infty)$. We must show that this inequality is strict for all $x \in (c, +\infty)$. Suppose otherwise, and fix $x_0 \in (c, +\infty)$ such that $f(x_0) = 0$. Fix any $x_1 \in (x_0, +\infty)$. Then $c < x_0 < \frac{1}{2}x_0 + \frac{1}{2}x_1 < x_1$, and we see that

$$0 \quad \overset{\text{(c)}}{\geq} \quad f\big(\tfrac{1}{2}x_0 + \tfrac{1}{2}x_1\big) \quad \overset{(*)}{>} \quad \tfrac{1}{2}f(x_0) + \tfrac{1}{2}f(x_1) \quad = \quad \tfrac{1}{2}f(x_1),$$

where (*) follows from the fact that f is strictly concave on $(c, +\infty)$. This proves that $f(x_1) < 0$. Set $\varepsilon := -f(x_1)$; clearly, $\varepsilon > 0$. Now, using the fact that $\lim\limits_{x \to +\infty} f(x) = 0$ (because $y = 0$ is a horizontal asymptote of f as $x \to +\infty$), we fix $L \in \mathbb{R}$ such that for all $x \in A$, if $x > L$, then $|f(x) - 0| < \varepsilon$, that is, $-\varepsilon < f(x) < \varepsilon$. Clearly, we may assume that $L > x_1$, for otherwise, we simply choose a larger L. Fix any $x_2 \in (L, +\infty)$, so that $c < x_0 < x_1 < L < x_2$. Note that this implies that x_0, x_2 are distinct, and that $-\varepsilon < f(x_2) < \varepsilon$.

Now, set $\lambda_0 := \frac{x_2 - x_1}{x_2 - x_0}$ and $\lambda_2 := \frac{x_1 - x_0}{x_2 - x_0}$, so that $\lambda_0, \lambda_2 > 0$, $\lambda_0 + \lambda_1 = 1$, and $x_1 = \lambda_0 x_0 + \lambda_2 x_2$. We then have that

$$\begin{array}{rcll}
-\varepsilon & = & f(x_1) & \text{because } \varepsilon = -f(x_1) \\
 & = & f\big(\lambda_0 x_0 + \lambda_2 x_2\big) & \text{because } x_1 = \lambda_0 x_0 + \lambda_2 x_2 \\
 & > & \lambda_0 f(x_0) + \lambda_2 f(x_2) & \text{because } f \text{ is strictly concave on } (c, +\infty) \\
 & = & \lambda_2 f(x_2) & \text{because } f(x_0) = 0 \\
 & > & -\lambda_2 \varepsilon & \text{because } f(x_2) > -\varepsilon \text{ and } \lambda_2 > 0 \\
 & > & -\varepsilon & \text{because } 0 < \lambda_2 < 1 \text{ and } \varepsilon > 0,
\end{array}$$

a contradiction. This proves (d).

Clearly, (a) and (b) could be proven analogously to (c) and (d). However, we can also use our result for (c) and (d) to quickly prove (a) and (b). Indeed, consider the function $-f$. It then readily follows from the appropriate definitions that the following hold:

- f is convex on $(c, +\infty)$ if and only if $-f$ is concave on $(c, +\infty)$;
- f is strictly convex on $(c, +\infty)$ if and only if $-f$ is strictly concave on $(c, +\infty)$.

Let us now prove (a). Suppose that f is convex on $(c, +\infty)$. Then $-f$ is concave on $(c, +\infty)$. By (c) applied to $-f$, we get that $-f(x) \leq 0$ for all $x \in (c, +\infty)$. So, $f(x) \geq 0$ for all $x \in (c, +\infty)$. This proves (a). The proof of (b) is analogous, except that we use (d) instead of (c). □

Theorem 4.14.7. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, and let $c \in \mathbb{R}$ be such that $(c, +\infty) \subseteq A$. Let $a, b \in \mathbb{R}$, and assume that $y = ax + b$ is a horizontal/slant asymptote of f as $x \to +\infty$.*[25] *Then all the following hold:*

(a) if f is convex on $(c, +\infty)$, then $f(x) \geq ax + b$ for all $x \in (c, +\infty)$;

(b) if f is strictly convex on $(c, +\infty)$, then $f(x) > ax + b$ for all $x \in (c, +\infty)$;

(c) if f is concave on $(c, +\infty)$, then $f(x) \leq ax + b$ for all $x \in (c, +\infty)$;

(d) if f is strictly concave on $(c, +\infty)$, then $f(x) < ax + b$ for all $x \in (c, +\infty)$.

Proof. Define $g : A \to \mathbb{R}$ by setting $g(x) = f(x) - (ax + b)$ for all $x \in \mathbb{R}$. Our goal is to apply Lemma 4.14.6 to the function g. First, we need a few technical claims.

> **Claim 1.** The line $y = 0$ (i.e. the x-axis) is a horizontal asymptote of g as $x \to +\infty$.

Proof of Claim 1. Note that:

$$\lim_{x \to +\infty} g(x) \quad = \quad \lim_{x \to +\infty} \big(f(x) - (ax + b)\big) \quad \overset{(*)}{=} \quad 0,$$

where (*) follows from the fact that $y = ax + b$ is a horizontal/slant asymptote of f. So $y = 0$ is horizontal asymptote of g as $x \to +\infty$. ♦

> **Claim 2.** Let $\star \in \{\leq, <, \geq, >\}$. Then for all $x_1, x_2 \in (c, +\infty)$ and $\lambda_1, \lambda_2 \in \mathbb{R}$ such that $\lambda_1, \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$, the following two inequalities are equivalent:
>
> - $f\big(\lambda_1 x_1 + \lambda_2 x_2\big) \star \lambda_1 f(x_1) + \lambda_2 f(x_2)$;
> - $g\big(\lambda_1 x_1 + \lambda_2 x_2\big) \star \lambda_1 g(x_1) + \lambda_2 g(x_2)$.

Proof of Claim 2. Fix $x_1, x_2 \in (c, +\infty)$ and $\lambda_1, \lambda_2 \in \mathbb{R}$ such that $\lambda_1, \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$. We then have that

$$g(\lambda_1 x_1 + \lambda_2 x_2) \quad = \quad f(\lambda_1 x_1 + \lambda_2 x_2) - \big(a(\lambda_1 x_1 + \lambda_2 x_2) + b\big),$$

and that

---

[25] If $a = 0$, then $y = ax + b$ (i.e. $y = b$) is a horizontal asymptote of f as $x \to +\infty$. On the other hand, if $a \neq 0$, then $y = ax + b$ is a slant asymptote of f as $x \to +\infty$.

$$\begin{aligned}
& \lambda_1 g(x_1) + \lambda_2 g(x_2) \\
= \quad & \lambda_1\big(f(x_1) - (ax_1 + b)\big) + \lambda_2\big(f(x_2) - (ax_2 + b)\big) \\
= \quad & \big(\lambda_1 f(x_1) + \lambda_2 f(x_2)\big) - \big(\lambda_1(ax_1 + b) + \lambda_2(ax_2 + b)\big) \\
= \quad & \big(\lambda_1 f(x_1) + \lambda_2 f(x_2)\big) - \big(a(\lambda_1 x_1 + \lambda_2 x_2) + \underbrace{(\lambda_1 + \lambda_2)}_{=1} b\big) \\
= \quad & \big(\lambda_1 f(x_1) + \lambda_2 f(x_2)\big) - \big(a(\lambda_1 x_1 + \lambda_2 x_2) + b\big).
\end{aligned}$$

Thus,

$$g(\lambda_1 x_1 + \lambda_2 x_2) - \big(\lambda_1 g(x_1) + \lambda_2 g(x_2)\big) \quad = \quad f(\lambda_1 x_1 + \lambda_2 x_2) - \big(\lambda_1 f(x_1) + \lambda_2 f(x_2)\big),$$

and the result follows. ◆

Claim 3. All the following hold:

1. f is convex on $(c, +\infty)$ if and only if g is convex on $(c, +\infty)$;
2. f is strictly convex on $(c, +\infty)$ if and only if g is strictly convex on $(c, +\infty)$;
3. f is concave on $(c, +\infty)$ if and only if g is concave on $(c, +\infty)$;
4. f is strictly concave on $(c, +\infty)$ if and only if g is strictly concave on $(c, +\infty)$.

Proof of Claim 3. For (1), we have the following sequence of equivalent statements:

$$\begin{array}{lcl}
\begin{array}{l} f \text{ is convex} \\ \text{on } (c,+\infty) \end{array} & \overset{\text{by definition}}{\iff} & \begin{array}{l} \forall x_1, x_2 \in (c,+\infty),\ \forall \lambda_1, \lambda_2 \in \mathbb{R}: \\ \text{if } \lambda_1, \lambda_2 \geq 0 \text{ and } \lambda_1 + \lambda_2 = 1, \\ \text{then } f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) \end{array} \\
 & \overset{\text{by Claim 1}}{\iff} & \begin{array}{l} \forall x_1, x_2 \in (c,+\infty),\ \forall \lambda_1, \lambda_2 \in \mathbb{R}: \\ \text{if } \lambda_1, \lambda_2 \geq 0 \text{ and } \lambda_1 + \lambda_2 = 1, \\ \text{then } g(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 g(x_1) + \lambda_2 g(x_2) \end{array} \\
 & \overset{\text{by definition}}{\iff} & g \text{ is convex on } (c,+\infty).
\end{array}$$

The proofs of (2), (3), and (4) are analogous. ◆

The result now follows immediately from Claims 1 and 3 and from Lemma 4.14.6. Indeed, for (a), we assume that f is convex on $(c, +\infty)$. Then, by Claim 3, g is convex on $(c, +\infty)$. By Claim 1, $y = 0$ is a horizontal asymptote of g as $x \to +\infty$. So, by Lemma 4.14.6, we have that $g(x) \geq 0$ for all $x \in (c, +\infty)$. It then follows from the definition of g that $f(x) \geq ax + b$ for all $x \in (c, +\infty)$. This proves (a). The proofs of (b), (c), and (d) are analogous. □

Theorem 4.14.8. *Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, and let $c \in \mathbb{R}$ be such that $(-\infty, c) \subseteq A$. Let $a, b \in \mathbb{R}$, and assume that $y = ax + b$ is a horizontal/slant asymptote of f as $x \to -\infty$.*[26] *Then all the following hold:*

(a) if f is convex on $(-\infty, c)$, then $f(x) \geq ax + b$ for all $x \in (-\infty, c)$;

(b) if f is strictly convex on $(-\infty, c)$, then $f(x) > ax + b$ for all $x \in (-\infty, c)$;

(c) if f is concave on $(-\infty, c)$, then $f(x) \leq ax + b$ for all $x \in (-\infty, c)$;

(d) if f is strictly concave on $(-\infty, c)$, then $f(x) < ax + b$ for all $x \in (-\infty, c)$.

Proof. Set $-A := \{-a \mid a \in A\}$; since $(-\infty, c) \subseteq A$, we have that $(-c, +\infty) \subseteq -A$. Define $h : -A \to \mathbb{R}$ by setting $h(x) := f(-x)$ for all $x \in -A$. Note that the graph of h is obtained by reflecting the graph of f about the y-axis. Our goal is to apply Theorem 4.14.7 to the function h in order to prove the desired result for f. First, we need a couple of technical claims.

> **Claim 1.** The line $y = -ax + b$ is a hotizontal/slant asymptote of h as $x \to +\infty$.

Proof of Claim 1. Note that:

$$\begin{array}{rcl} \lim\limits_{x \to +\infty} \big(h(x) - (-ax + b)\big) & = & \lim\limits_{x \to +\infty} \big(f(-x) - (-ax + b)\big) \\ & \overset{(*)}{=} & \lim\limits_{y \to -\infty} \big(f(y) - (ay + b)\big) \quad \overset{(*)}{=} \quad 0, \end{array}$$

where (*) was obtained via substitution $y = -x$, and (**) follows from the fact that $y = ax + b$ is a horizontal/slant asymptote of f as $x \to -\infty$. This proves that $y = -ax + b$ is a hotizontal/slant asymptote of h as $x \to +\infty$. ♦

> **Claim 2.** All the following hold:
>
> (1) f is convex on $(-\infty, c)$ if and only if h is convex on $(-c, +\infty)$;
> (2) f is strictly convex on $(-\infty, c)$ if and only if h is strictly convex on $(-c, +\infty)$;
> (3) f is concave on $(-\infty, c)$ if and only if h is concave on $(-c, +\infty)$;
> (4) f is strictly concave on $(-\infty, c)$ if and only if h is strictly concave on $(-c, +\infty)$.

Proof of Claim 2. For (1), we have the following sequence of equivalent statements:

---

[26] If $a = 0$, then $y = ax + b$ (i.e. $y = b$) is a horizontal asymptote of f as $x \to -\infty$. On the other hand, if $a \neq 0$, then $y = ax + b$ is a slant asymptote of f as $x \to -\infty$.

$$\begin{array}{lcl}
\begin{array}{l} f \text{ is convex} \\ \text{on } (-\infty, c) \end{array} & \overset{\text{by definition}}{\Longleftrightarrow} & \begin{array}{l} \forall x_1, x_2 \in (-\infty, c),\ \forall \lambda_1, \lambda_2 \in \mathbb{R}: \\ \text{if } \lambda_1, \lambda_2 \geq 0 \text{ and } \lambda_1 + \lambda_2 = 1, \text{ then} \\ f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) \end{array} \\[2ex]
 & \Longleftrightarrow & \begin{array}{l} \forall x_1, x_2 \in (-c, +\infty),\ \forall \lambda_1, \lambda_2 \in \mathbb{R}: \\ \text{if } \lambda_1, \lambda_2 \geq 0 \text{ and } \lambda_1 + \lambda_2 = 1, \text{ then} \\ f(\lambda_1(-x_1) + \lambda_2(-x_2)) \leq \lambda_1 f(-x_1) + \lambda_2 f(-x_2) \end{array} \\[2ex]
 & \Longleftrightarrow & \begin{array}{l} \forall x_1, x_2 \in (-c, +\infty),\ \forall \lambda_1, \lambda_2 \in \mathbb{R}: \\ \text{if } \lambda_1, \lambda_2 \geq 0 \text{ and } \lambda_1 + \lambda_2 = 1, \text{ then} \\ f(-(\lambda_1 x_1 + \lambda_2 x_2)) \leq \lambda_1 f(-x_1) + \lambda_2 f(-x_2) \end{array} \\[2ex]
 & \overset{\substack{\text{by the} \\ \text{definition} \\ \text{of } h}}{\Longleftrightarrow} & \begin{array}{l} \forall x_1, x_2 \in (-c, +\infty),\ \forall \lambda_1, \lambda_2 \in \mathbb{R}: \\ \text{if } \lambda_1, \lambda_2 \geq 0 \text{ and } \lambda_1 + \lambda_2 = 1, \text{ then} \\ h(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 h(x_1) + \lambda_2 h(x_2) \end{array} \\[2ex]
 & \overset{\text{by definition}}{\Longleftrightarrow} & h \text{ is convex on } (-c, +\infty).
\end{array}$$

The proofs of (2), (3), and (4) are analogous. ♦

Let us now prove (a). Suppose that f is convex on $(-\infty, c)$. Then by Claim 2, h is convex on $(-c, +\infty)$. Moreover, by Claim 1, $y = -ax + b$ is a horizontal/slant asymptote of h as $x \to +\infty$. So, by Theorem 4.14.7, we have that $h(x) \geq -ax + b$ for all $x \in (-c, +\infty)$; consequently, $f(x) = h(-x) \geq ax + b$ for all $x \in (-\infty, c)$. This proves (a). The proofs of (b), (c), and (d) are analogous. □

4.15 Sketching graphs of functions

To sketch the graph of a function $y = f(x)$, we need to do all the following:

1. Determine the domain of the function (i.e. the largest subset of $\mathbb{R}$ on which the function is defined).

2. Examine the behavior of the function near any "gaps" or "edges" of the domain,[27] at and near any points at which f is discontinuous, as well as at $+\infty$ and $-\infty$. In particular, determine whether the function has any vertical, horizontal, or slant asymptotes.

---

[27] A "gap" in the domain would be a point c at which f is undefined, whereas it is defined near c (both to the left and to the right). An "edge" of the domain would be a point c such that f defined near c to the left or to the right, but not both.

3. Determine whether the function is periodic (and if so, find its period),[28] and whether it is even and/or odd.[29]
4. Find the x- and y-intercepts of f.[30] Determine intervals at which $f(x) > 0$ and at which $f(x) < 0$.
5. Find f', and using this, determine at which intervals the function increases or decreases. Find all local maxima and minima.[31]
6. Find f'', and using this, determine the intervals at which the function is convex or concave, and find all its inflection points.

Notation: In the examples below, the symbol "$\smile$" denotes "convex," and "$\frown$" denotes "concave."

Example 4.15.1. *Sketch the graph of the function* $f(x) = \frac{(x-1)^3}{(x+1)^2}$.

Solution. The domain of the function f is $\mathbb{R} \setminus \{-1\}$, i.e. the function is defined for all real numbers $x \neq -1$. Moreover, it is clear that f is continuous (at all points in its domain). Note that

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

and so $x = -1$ is a vertical asymptote of $f(x)$. Furthermore, the graph of the function approaches the line $x = -1$ "downward" both from the left and from the right. Next, note that

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

and so f has no horizontal asymptotes. Further, note that

$$\lim_{x \to +\infty} \tfrac{f(x)}{x} \quad = \quad \lim_{x \to +\infty} \tfrac{(x-1)^3}{x(x+1)^2} \quad = \quad 1 \quad =: \quad a,$$

and that

$$\lim_{x \to +\infty} \Big(f(x) - ax \Big) \quad = \quad \lim_{x \to +\infty} \tfrac{-5x^2+2x-1}{x^2+2x+1} \quad = \quad -5 \quad =: \quad b.$$

---

[28] A function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is *periodic* if there exists a real number $p > 0$ such that for all $a \in A$ and all $k \in \mathbb{Z}$, we have that $a + kp \in A$ and $f(a + kp) = f(a)$. The smallest such p, if it exists, is called the *period* of the function.

[29] A function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is *even* if for all $x \in A$, we have that $-x \in A$ and $f(-x) = f(x)$. Meanwhile, f is *odd* if for all $x \in A$, we have that $-x \in A$ and $f(-x) = -f(x)$. Graphs of even functions are symmetric about the y-axis, and graphs of odd functions are symmetric about the origin.

[30] An *x-intercept* of f is any value of x such that $f(x) = 0$ (i.e. the value at which f intersects the x-axis). The y-intercept of f is the value $y = f(0)$ (i.e. the value at which f intersects the y-axis), if it exists.

[31] Sometimes, local maxima and minima can more easily be found using f''.

So, $y = x - 5$ is a slant asymptote of $y = f(x)$ as $x \to +\infty$. A similar calculation shows that $y = x - 5$ is a slant asymptote of $y = f(x)$ as $x \to -\infty$.

The function f is neither even nor odd (e.g. because it is defined as 1, but not at -1). The function is also not periodic because, otherwise, there would be some constant $p > 0$ (the period of the function) such that f is undefined at $-1 + kp$ for all $k \in \mathbb{Z}$.

Note that $f(0) = -1$ (so, $y = f(x)$ intersects the y axis at $y = -1$), and that $f(x) = 0$ only for $x = 1$ (so, $y = f(x)$ intersects the x-axis at $x = 1$). Further, we have the following table, showing where f is positive, negative, or zero.

x	$(-\infty,-1)$	-1	$(-1,1)$	1	$(1,+\infty)$
$x-1$	$-$		$-$	0	$+$
$x+1$	$-$	0	$+$		$+$
$f(x)$	$-$	undef	$-$	0	$+$

Further, a straightforward calculation shows that

$$f'(x) = \tfrac{(x-1)^2(x+5)}{(x+1)^3} \qquad \text{and} \qquad f''(x) = 24\tfrac{x-1}{(x+1)^4}$$

for all $x \in \mathbb{R} \setminus \{-1\}$. So, for the first derivative, we get the table below.

x	$(-\infty,-5)$	-5	$(-5,-1)$	-1	$(-1,1)$	1	$(1,+\infty)$
$x-1$	$-$		$-$		$-$	0	$+$
$x+1$	$-$		$-$	0	$+$		$+$
$x+5$	$-$	0	$+$		$+$		$+$
$f'(x)$	$+$	0	$-$	undef	$+$	0	$+$
$f(x)$	$\nearrow$	max	$\searrow$	undef	$\nearrow$		$\nearrow$

For the second derivative, we get the table below.

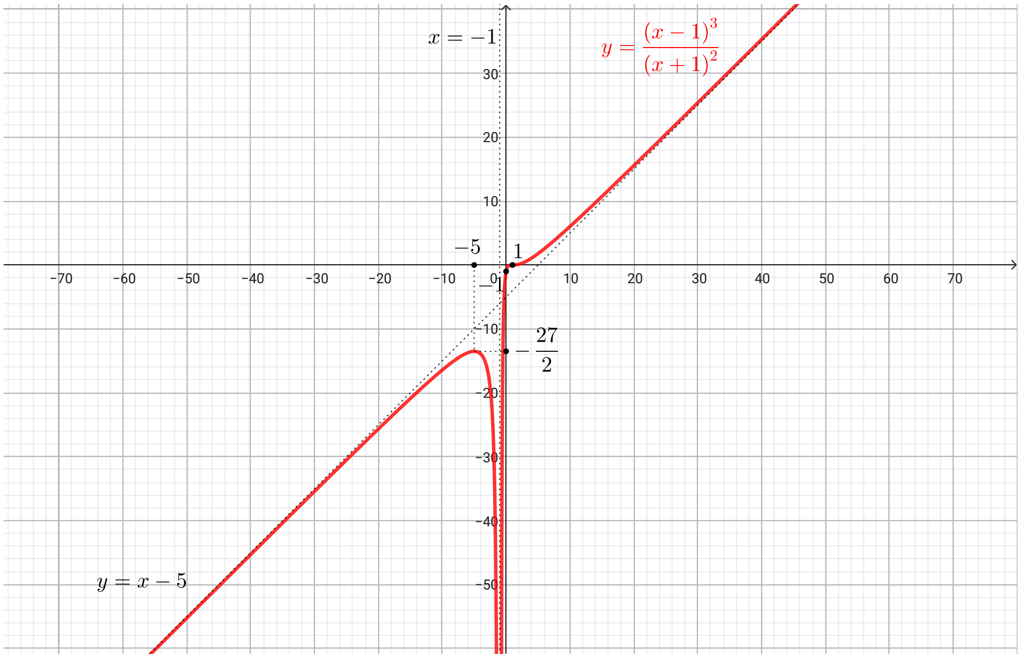
x	$(-\infty,-1)$	-1	$(-1,1)$	1	$(1,+\infty)$
$x-1$	$-$		$-$	0	$+$
$x+1$	$-$	0	$+$		$+$
$f''(x)$	$-$	undef	$-$	0	$+$
$f(x)$	$\frown$	undef	$\frown$	inflec	$\smile$

We now combine the three tables above to obtain the table below.

x	$(-\infty,-5)$	-5	$(-5,-1)$	-1	$(-1,1)$	1	$(1,+\infty)$
$f'(x)$	$+$	0	$-$	undef	$+$	0	$+$
$f''(x)$	$-$		$-$	undef	$-$	0	$+$
$f(x)$	$-$ ↗ ⌢	max	$-$ ↘ ⌢	undef	$-$ ↗ ⌢	0 inflec	$+$ ↗ ⌣

Note that f reaches a local maximum at $x = -5$, and we have that $f(-5) = -\frac{27}{2}$. (There are no other local extrema.) Furthermore, the only point of inflection is $x = 1$, and we have that $f(1) = 0$.

Putting all this together, we get the following picture.



□

Example 4.15.2. *Sketch the graph of the function* $f(x) = \frac{x-2}{\sqrt{x^2+1}}$.

Solution. The domain of the function is $\mathbb{R}$, and clearly, f is continuous. Therefore, f has no vertical asymptotes. Note that for all $x \in \mathbb{R} \setminus \{0\}$, we have that

$$f(x) \quad = \quad \frac{x-2}{\sqrt{x^2+1}} \quad = \quad \frac{x(1-\frac{2}{x})}{|x|\sqrt{1+\frac{1}{x^2}}}.$$

So,

$$\lim_{x\to+\infty} f(x) \quad = \quad \lim_{x\to+\infty} \frac{x(1-\frac{2}{x})}{|x|\sqrt{1+\frac{1}{x^2}}} \quad \overset{(*)}{=} \quad \lim_{x\to+\infty} \frac{1-\frac{2}{x}}{\sqrt{1+\frac{1}{x^2}}} \quad = \quad 1,$$

where (*) follows from the fact that $x > 0$. On the other hand,

$$\lim_{x\to-\infty} f(x) \quad = \quad \lim_{x\to-\infty} \frac{x(1-\frac{2}{x})}{|x|\sqrt{1+\frac{1}{x^2}}} \quad \overset{(*)}{=} \quad \lim_{x\to-\infty} \frac{-(1-\frac{2}{x})}{\sqrt{1+\frac{1}{x^2}}} \quad = \quad -1,$$

where (*) follows from the fact that $x < 0$. Thus, $y = 1$ is a horizontal asymptote of f as $x \to +\infty$, whereas $y = -1$ is a horizontal asymptote of f as $x \to -\infty$. Since f has a horizontal asymptote both as $x \to +\infty$ and as $x \to -\infty$, we see that f has no slant asymptotes.

Note that for $x \in \mathbb{R}$, we have that

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

and we see that f is neither even nor odd. Moreover, it is clear that f is not periodic.

Note that $f(0) = -2$ (so, $y = f(x)$ intersects the y-axis at $y = -2$), and that $f(x) = 0$ only for $x = 2$ (so, $y = f(x)$ intersects the x-axis at $x = 2$). Further, we have the following table, showing where f is positive, negative, or zero.

		2	
x	$(-\infty, 2)$		$(2, +\infty)$
$x - 2$	$-$	0	$+$
$\sqrt{x^2+1}$	$+$		$+$
$f(x)$	$-$	0	$+$

Next, a straightforward calculation shows that

$$f'(x) = \frac{2x+1}{(x^2+1)^{3/2}} \qquad \text{and} \qquad f''(x) = -\frac{4x^2+3x-2}{(x^2+1)^{5/2}} = -\frac{\left(x-\frac{-3-\sqrt{41}}{8}\right)\left(x-\frac{-3+\sqrt{41}}{8}\right)}{(x^2+1)^{5/2}}$$

for all $x \in \mathbb{R}$. So, for the first derivative, we get the table below.

		$-\frac{1}{2}$	
x	$(-\infty, -\frac{1}{2})$		$(-\frac{1}{2}, +\infty)$
$2x + 1$	$-$	0	$+$
$x^2 + 1$	$+$		$+$
$f'(x)$	$-$	0	$+$
$f(x)$	↘	min	↗

For the second derivative, we get the table below.

x	$(-\infty, \frac{-3-\sqrt{41}}{8})$	$\frac{-3-\sqrt{41}}{8}$	$(\frac{-3-\sqrt{41}}{8}, \frac{-3+\sqrt{41}}{8})$	$\frac{-3+\sqrt{41}}{8}$	$(\frac{-3+\sqrt{41}}{8}, +\infty)$
$x - \frac{-3-\sqrt{41}}{8}$	$-$	0	$+$		$+$
$x - \frac{-3+\sqrt{41}}{8}$	$-$		$-$	0	$+$
x^2+1	$+$		$+$		$+$
$f''(x)$	$-$	0	$+$	0	$-$
$f(x)$	$\frown$	inflec	$\smile$	inflec	$\frown$

We observe that $\frac{-3-\sqrt{41}}{8} < -\frac{1}{2} < 0 < \frac{-3+\sqrt{41}}{8} < 2$, and we combine the three tables above, as follows.

x	$(-\infty, \frac{-3-\sqrt{41}}{8})$	$\frac{-3-\sqrt{41}}{8}$	$(\frac{-3-\sqrt{41}}{8}, -\frac{1}{2})$	$-\frac{1}{2}$	$(-\frac{1}{2}, \frac{-3+\sqrt{41}}{8})$	$\frac{-3+\sqrt{41}}{8}$	$(\frac{-3+\sqrt{41}}{8}, 2)$	2	$(2, +\infty)$
$f'(x)$	$-$		$-$	0	$+$		$+$		$+$
$f''(x)$	$-$	0	$+$		$+$	0	$-$		$-$
$f(x)$	$-$ $\searrow$ $\frown$	inflec	$-$ $\searrow$ $\smile$	min	$-$ $\nearrow$ $\smile$	inflec	$-$ $\nearrow$ $\frown$	0	$+$ $\nearrow$ $\frown$

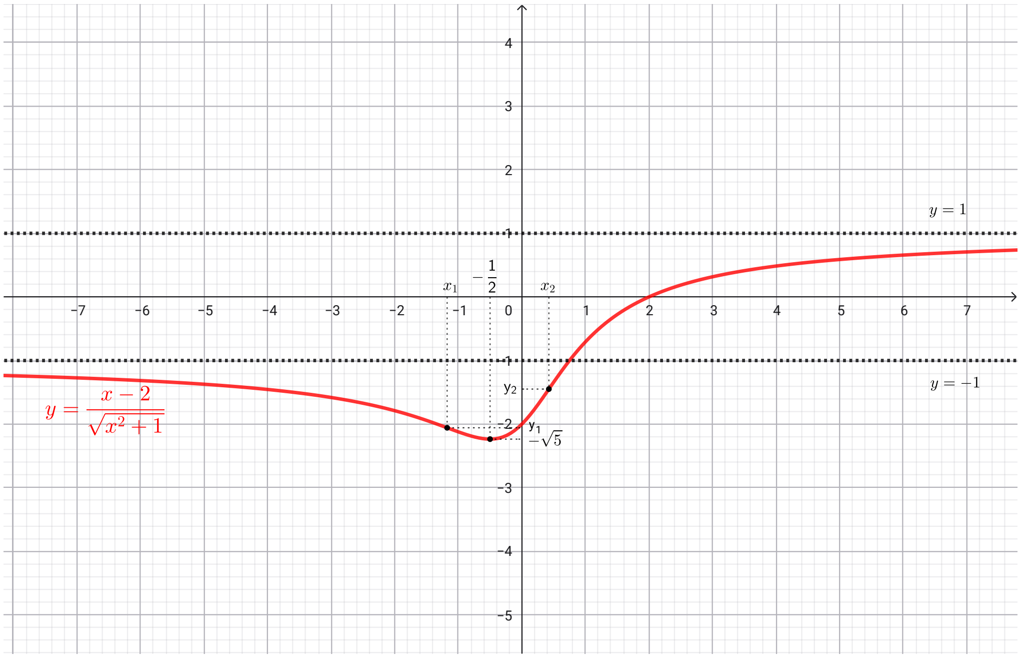
Note that f reaches a local maximum at $x = -\frac{1}{2}$, and we have that $f(-\frac{1}{2}) = -\sqrt{5}$. (There are no other local extrema.) The inflection points of f are

$$x_1 := \tfrac{-3-\sqrt{41}}{8} \qquad \text{and} \qquad x_2 := \tfrac{-3+\sqrt{41}}{8},$$

and we have that

$$y_1 := f(x_1) = \tfrac{-19-\sqrt{41}}{\sqrt{114+6\sqrt{41}}} \qquad \text{and} \qquad y_2 := f(x_2) = \tfrac{-19+\sqrt{41}}{\sqrt{114-6\sqrt{41}}}.$$

Putting all this together, we get the following picture.



□

Example 4.15.3. *Sketch the graph of the function* $f(x) = x \arctan x$*.*

Solution. The domain of the function is $\mathbb{R}$, and clearly, f is continuous. Therefore, f has no vertical asymptotes. Next, note that for all $x \in \mathbb{R}$, we have that

$$f(-x) \quad = \quad (-x)\arctan(-x) \quad \overset{(*)}{=} \quad (-x)(-\arctan x) \quad = \quad x\arctan x \quad = \quad f(x),$$

where (*) follows from the fact that the arctangent function is odd. So, f is even, which means that its graph is symmetric about the y-axis.

Now, recall that

$$\lim_{x\to+\infty} \arctan x = \tfrac{\pi}{2} \qquad \text{and} \qquad \lim_{x\to-\infty} \arctan x = -\tfrac{\pi}{2}.$$

It then readily follows that

$$\lim_{x\to+\infty} f(x) \quad = \quad \lim_{x\to+\infty} \big(x \arctan x\big) \quad = \quad +\infty$$

and

$$\lim_{x\to-\infty} f(x) \quad = \quad \lim_{x\to-\infty} \big(x \arctan x\big) \quad = \quad +\infty$$

Thus, f has no horizontal asymptotes. On the other hand, we have that

$$\lim_{x\to+\infty} \tfrac{f(x)}{x} \quad = \quad \lim_{x\to+\infty} \arctan x \quad = \quad \tfrac{\pi}{2} \quad =: \quad a,$$

and that

$$\begin{array}{rcll}
\lim\limits_{x\to+\infty}\big(f(x)-ax\big) & = & \lim\limits_{x\to+\infty}\big(x\arctan x-\tfrac{\pi}{2}x\big) & \\
& = & \lim\limits_{x\to+\infty} x(\arctan x-\tfrac{\pi}{2}) & \\
& = & \lim\limits_{x\to+\infty} \frac{\arctan x-\frac{\pi}{2}}{\frac{1}{x}} & \\
& = & \lim\limits_{x\to+\infty} \frac{\frac{d}{dx}(\arctan x-\frac{\pi}{2})}{\frac{d}{dx}(\frac{1}{x})} & \begin{array}{l}\text{via L'Hôpital's Rule, since}\\ \lim\limits_{x\to+\infty}\big(\arctan x-\frac{\pi}{2}\big)=0\\ \text{and } \lim\limits_{x\to+\infty}\big(\frac{1}{x}\big)=0\end{array} \\
& = & \lim\limits_{x\to+\infty} \frac{\frac{1}{1+x^2}}{-\frac{1}{x^2}} & \\
& = & \lim\limits_{x\to+\infty} \Big(-\frac{x^2}{x^2+1}\Big) & \\
& = & \lim\limits_{x\to+\infty} \Big(-\frac{1}{1+\frac{1}{x^2}}\Big) & \\
& = & -1 \quad =: \quad b. &
\end{array}$$

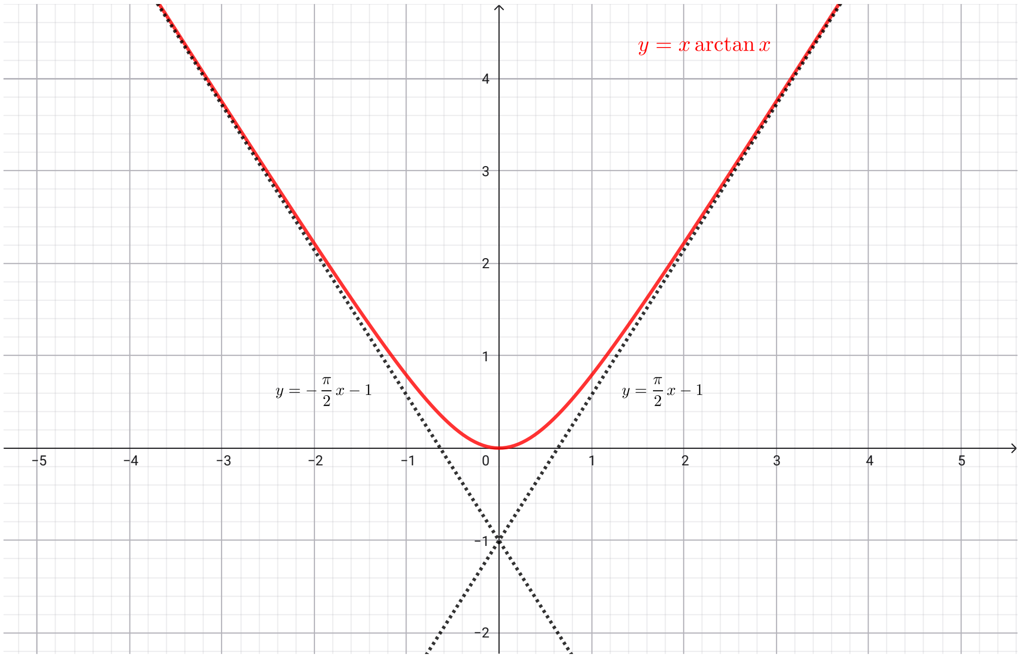
Thus, $y=\frac{\pi}{2}x-1$ is a slant asymptote of f as $x\to+\infty$. Now, the line obtained by reflecting $y=\frac{\pi}{2}x-1$ about the y-axis is $y=-\frac{\pi}{2}x-1$. Since f is even, it follows that $y=-\frac{\pi}{2}x-1$ is a slant asymptote of f as $x\to-\infty$.

Note that $f(0)=0$, and in fact, $f(x)=0$ only for $x=0$. Thus, the graph of f passes through the origin, and it does not intersect the x-axis or the y-axis anywhere else.

Next, a straightforward calculation shows that

$$f'(x)=\arctan x+\tfrac{x}{1+x^2} \qquad \text{and} \qquad f''(x)=\tfrac{2}{(1+x^2)^2}$$

Note that $f'(0)=0$, and that $f''(0)>0$. Therefore, by the Second Derivative Test, f reaches a local minimum at $x=0$, and clearly, we have that $f(0)=0$. But in fact, $f''(x)>0$ for all $x\in\mathbb{R}$, and consequently, f is strictly convex. Thus, the graph of f is as in the picture below.



Remark: Note that the fact that $f''(x) > 0$ for all $x \in \mathbb{R}$ implies that f' is strictly increasing. Since $f'(0) = 0$, it follows that $f'(x) < 0$ for all $x \in (-\infty, 0)$, whereas $f'(x) > 0$ for all $x \in (0, +\infty)$. This yields the table below, which is consistent with the graph that we obtained above. Note, however, that the behavior of f' would have been quite difficult to analyze without using f''.

x	$(-\infty, 0)$	0	$(0, +\infty)$
$f'(x)$	$-$	0	$+$
$f(x)$	$\searrow$	min	$\nearrow$

□

Chapter 5

Integration

5.1 The indefinite integral

In chapter 4, we were interested in finding the derivative of a given function. Here, we will concern ourselves with the opposite problem: determine the functions that have the given derivative.

5.1.1 The antiderivative and the indefinite integral: definition, examples, and basic properties

The antiderivative. For an interval $I \subseteq \mathbb{R}$ and a function $f : A \subseteq \mathbb{R} \to \mathbb{R}$, where $I \subseteq A$, an *antiderivative* of f on I is the function $F : I \to \mathbb{R}$ such that $F'(x) = f(x)$ for all $x \in \mathbb{R}$.

Example 5.1.1.

(a) $F_1(x) = x^2$ *is an antiderivative of* $f_1(x) = 2x$ *on* $\mathbb{R} = (-\infty, +\infty)$*;*

(b) $F_2(x) = \sin x$ *is an antiderivative of* $f_2(x) = \cos x$ *on* $\mathbb{R} = (-\infty, \infty)$*;*

(c) $F_3(x) = \arcsin x$ *is an antiderivative of* $f_3(x) = \frac{1}{\sqrt{1-x^2}}$ *on* $(-1, 1)$*.*

Theorem 5.1.2. *Suppose that F is an antiderivative of a function f on an interval I. Then the general form of the antiderivative of f on the interval I is $F + C$, where C is an arbitrary constant. More precisely, a function G is an antiderivative of f on I if and only if there exists some constant $C \in \mathbb{R}$ such that for all $x \in I$, we have that $G(x) = F(x) + C$.*

Proof. Suppose first that there exists some $C \in \mathbb{R}$ such that for all $x \in I$, we have that $G(x) = F(x) + C$. Then for all $x \in I$, we have that $G'(x) = F'(x) = f(x)$, and so G is an antiderivative of f on I.

Suppose conversely that G is an antiderivative of f on I. Then for $x \in I$, we have that $\frac{d}{dx}\big(G(x) - F(x)\big) = G'(x) - F'(x) = f(x) - f(x) = 0$. So, by Theorem 4.13.1(a), the function $G - F$ is constant on the interval I. Thus, there exists a constant $C \in \mathbb{R}$ such that for all $x \in I$, we have that $G(x) - F(x) = C$, that is, $G(x) = F(x) + C$. $\square$

The indefinite integral. Suppose that a function f has at least one antiderivative, say F, on an interval I. The *indefinite integral* of f on I, denoted by $\int f(x)dx$, is the set of all antiderivatives of f on I. By Theorem 5.1.2, we have that

$$\int f(x)dx \quad = \quad \{F + C \mid C \in \mathbb{R}\}.$$

However, in practice, the set notation is virtually never used in the context of indefinite integrals. Instead, we write

$$\int f(x)dx \quad = \quad F(x) + C.$$

Here, C is understood to be an arbitrary constant. Importantly, the above is supposed to be valid for a fixed interval I (which is either specified explicitly or is clear from context).

Remark: If we happen to know that $F'(x) = f(x)$ for all $x \in I$ (where I is some interval), then Theorem 5.1.2 guarantees that $\int f(x)dx = F(x) + C$ on the interval I. In particular, we have the proposition below.

Proposition 5.1.3. *Let $F : A \subseteq \mathbb{R} \to \mathbb{R}$ be a function, differentiable on some interval $I \subseteq A$. Then*

$$\int F'(x)dx \quad = \quad F(x) + C$$

on the interval I.

Proof. This follows immediately from the fact that F is an antiderivative of F' on the interval I. $\square$

Example 5.1.4.

(a) $\int 2xdx = x^2 + C$, *valid on the interval* $\mathbb{R} = (-\infty, +\infty)$*;*

(b) $\int \cos xdx = \sin x + C$, *valid on the interval* $\mathbb{R} = (-\infty, +\infty)$*;*

(c) $\int \frac{dx}{\sqrt{1-x^2}} = \arcsin x + C$, *valid on the interval* $(-1, +1)$.

Remark: When asked to compute $\int f(x)dx$, we must **include the constant** C! If the constant C is omitted, then our answer is incorrect/incomplete.

Example 5.1.5. *Find the antiderivative of $f(x) = \cos x$ that passes through the point $(\frac{\pi}{2}, 0)$.*

Solution. We know that $\int \cos xdx = \sin x + C$. By solving $\sin(\frac{\pi}{2}) + C = 0$ for C, we get $C = -1$. So, the antiderivative that we need is the function $F(x) = \sin x - 1$. $\square$

Proposition 5.1.6. *For $x \in \mathbb{R} \setminus \{0\}$, we have the formula*

$$\tfrac{d}{dx} \ln |x| \quad = \quad \tfrac{1}{x}.$$

Consequently, we have the formula

$$\textstyle\int \frac{dx}{x} \quad = \quad \ln |x| + C,$$

valid on intervals $(-\infty, 0)$ and $(0, +\infty)$.

Proof. For $x \in (0, +\infty)$, we have that

$$\tfrac{d}{dx} \ln |x| \quad \overset{x>0}{=} \quad \tfrac{d}{dx} \ln x \quad = \quad \tfrac{1}{x}.$$

So, the formula $\int \frac{dx}{x} = \ln |x| + C$ is valid on the interval $(0, +\infty)$.

For $x \in (-\infty, 0)$, we have that

$$\tfrac{d}{dx} \ln |x| \quad \overset{x<0}{=} \quad \tfrac{d}{dx} \ln(-x) \quad = \quad \tfrac{1}{-x} \tfrac{d}{dx}(-x) \quad = \quad \tfrac{1}{-x}(-1) \quad = \quad \tfrac{1}{x}.$$

So, the formula $\int \frac{dx}{x} = \ln |x| + C$ is valid on the interval $(-\infty, 0)$. □

Remark: Indefinite integrals are computed on intervals, not on arbitrary sets. So, the formula $\int \frac{dx}{x} = \ln |x| + C$ from Proposition 5.1.6 is valid on the interval $(-\infty, 0)$, and also on the interval $(0, +\infty)$, as well as for any subinterval of those two intervals. However, the formula is **not** valid for the set $(-\infty, 0) \cup (0, +\infty)$, since this set is not an interval. Let us explain the reasoning. Define the function $F : (-\infty, 0) \cup (0, +\infty) \to \mathbb{R}$ by setting

$$F(x) \quad = \quad \begin{cases} \ln |x| - 1 & \text{if} \quad x < 0 \\ \ln |x| + 1 & \text{if} \quad x > 0 \end{cases}$$

for all $x \in (-\infty, 0) \cup (0, +\infty)$. Then for all $x \in (-\infty, 0) \cup (0, +\infty)$, we do indeed have that $F'(x) = \frac{1}{x}$. However, there is no fixed constant C such that for all $x \in (-\infty, 0) \cup (0, +\infty)$, we have that $F(x) = \ln |x| + C$. The reason for this "pathology" is the fact that $(-\infty, 0) \cup (0, +\infty)$ is not an interval.

5.1.2 A summary of important indefinite integrals

Using the table for derivatives from subsection 4.8, as well as Proposition 5.1.6, we get the following table of indefinite integrals (which should be memorized).

indefinite integral $\int f(x)dx$	constraints on the parameter α	constraints on the variable x
$\int 0dx = C$		
$\int x^\alpha dx = \frac{1}{\alpha+1}x^{\alpha+1} + C$	$\alpha \in \mathbb{R} \setminus \{-1\}$	$x > 0$
$\int \frac{dx}{x} = \ln\lvert x\rvert + C$		$x \neq 0$
$\int e^x dx = e^x + C$		
$\int \sin x dx = -\cos x + C$		
$\int \cos x dx = \sin x + C$		
$\int \frac{dx}{\cos^2 x} = \tan x + C$		$x \neq \frac{2k+1}{2}\pi,\ k \in \mathbb{Z}$
$\int \frac{dx}{\sqrt{1-x^2}} = \arcsin x + C$		$x \in (-1,1)$
$\int \frac{dx}{1+x^2} = \arctan x + C$		

Remark: When our table gives no constraints on the variable x, it means that the formula is valid on the entire interval $\mathbb{R} = (-\infty, +\infty)$.

Remark: As stated in our remark following Proposition 5.1.6, the formula

$$\int \frac{dx}{x} \quad = \quad \ln|x| + C$$

is valid on the intervals $(-\infty, 0)$ and $(0, +\infty)$. However, the formula is **not** valid on the set $(-\infty, 0) \cup (0, +\infty)$, since that set is not an interval.

Remark: The formula

$$\int \frac{dx}{\cos^2 x} \quad = \quad \tan x + C$$

is valid for any interval of the form $\big(\frac{2k-1}{2}\pi, \frac{2k+1}{2}\pi\big)$, where $k \in \mathbb{Z}$. However, the formula is **not** valid on the set $\mathbb{R} \setminus \{\frac{2k+1}{2}\pi \mid k \in \mathbb{Z}\}$, since this set is **not** an interval.

Remark: For a constant $\alpha \in \mathbb{R}$, the formula

$$\int x^\alpha dx \quad = \quad \frac{1}{\alpha+1}x^{\alpha+1} + C$$

is guaranteed to work on the interval $(0, +\infty)$. However, assuming everything is defined, we do not have to assume that $x > 0$. For instance, formulas $\int x^3 dx = \frac{1}{4}x^4 + C$ and $\int x^{1/3}dx = \frac{3}{4}x^{4/3} + C$ are valid on the interval $\mathbb{R} = (-\infty, +\infty)$, because everything is defined. Meanwhile, the formula $\int x^{-1/5}dx = \frac{5}{4}x^{4/5} + C$ is valid on the intervals $(-\infty, 0)$ and $(0, +\infty)$, because everything is defined on those intervals.[1] For $\alpha = 0$, the formula from the table becomes

$$\int 1dx = x + C \qquad \text{or} \qquad \int dx = x + C,$$

---

[1] Note that $x^{-1/5}$ is undefined for $x = 0$, which is why the formula $\int x^{-1/5}dx = \frac{5}{4}x^{4/5} + C$ is **not** valid on the entire interval $\mathbb{R} = (-\infty, +\infty)$. Once again, we do **not** say that $\int x^{-1/5}dx = \frac{5}{4}x^{4/5} + C$ is valid on the set $(-\infty, 0) \cup (0, +\infty)$, because that set is not an interval. The formula is valid on the interval $(-\infty, 0)$, and also on the interval $(0, +\infty)$.

which is valid on the interval $\mathbb{R} = (-\infty, +\infty)$. Note that for $\alpha = -1$, we have a special formula, namely

$$\int x^{-1} dx \quad = \quad \ln|x| + C,$$

which is valid on the intervals $(-\infty, 0)$ and $(0, +\infty)$.

5.1.3 Addition and scalar multiplication of indefinite integrals

Addition of indefinite integrals. Suppose that functions f_1 and f_2 have antiderivatives F_1 and F_2, respectively, on an interval I, so that $\int f_1(x)dx = F_1(x) + C$ and $\int f_2(x)dx = F_2(x) + C$. Then we have that

$$\begin{aligned} \int f_1(x)dx + \int f_2(x)dx \quad &= \quad \big(F_1(x) + C_1\big) + \big(F_2(x) + C_2\big) \\ &= \quad F_1(x) + F_2(x) + C_1 + C_2 \\ &= \quad F_1(x) + F_2(x) + C. \end{aligned}$$

Here, C_1 and C_2 are arbitrary real constants, and $C_1 + C_2$ yields an arbitrary real constant, which we simply denote by C. Similarly,

$$\int f_1(x)dx - \int f_2(x)dx \quad = \quad F_1(x) - F_2(x) + C.$$

Proposition 5.1.7. *Suppose that functions f_1 and f_2 have antiderivatives on an interval I. Then $f_1 + f_2$ and $f_1 - f_2$ also have antiderivatives on I, and moreover, we have the formula*

$$\int f_1(x) \pm f_2(x)dx \quad = \quad \int f_1(x)dx \pm \int f_2(x)dx,$$

valid on the interval I.

Notation: It is customary to omit parentheses as in the statement of Proposition 5.1.7 above. For a (potentially complicated) expression $\boxed{}$, we normally write $\int \boxed{} dx$ instead of $\int \big(\boxed{}\big) dx$.

Proof. In what follows, all formulas, and all antiderivatives, are assumed to be on the interval I. Let F_1 and F_2 be antiderivatives of f_1 and f_2, respectively, so that

$$\int f_1(x)dx = F_1(x) + C \qquad \text{and} \qquad \int f_2(x)dx = F_2(x) + C.$$

Note that

$$\tfrac{d}{dx}\big(F_1(x) + F_2(x)\big) \quad \overset{(*)}{=} \quad F_1'(x) + F_2'(x) \quad \overset{(**)}{=} \quad f_1(x) + f_2(x),$$

where (*) follows from the properties of the derivative (and in particular, from Proposition 4.3.3), whereas (**) follows from the fact that F_1 and F_2 are antiderivatives of f_1 and f_2, respectively. Thus, F_1+F_2 is an antiderivative of f_1+f_2, and it follows that

$$\int f_1(x)+f_2(x)dx \quad = \quad F_1(x)+F_2(x)+C \quad = \quad \int f_1(x)dx+\int f_2(x)dx.$$

The formula

$$\int f_1(x)-f_2(x)dx \quad = \quad \int f_1(x)dx-\int f_2(x)dx,$$

is obtained analogously.[2] □

Scalar multiplication of indefinite integrals. Suppose that a function f has an antiderivative F on an interval I, so that $\int f(x)dx = F(x)+C$, valid on I. For a constant $\alpha \in \mathbb{R}\setminus\{0\}$, we have that

$$\alpha\int f(x)dx \quad = \quad \alpha\big(F(x)+C_1\big) \quad = \quad \alpha F(x)+\alpha C_1 \quad = \quad \alpha F(x)+C$$

is valid on the interval I. Here C_1 is an arbitrary constant, which yields an arbitrary constant αC_1 (because $\alpha \neq 0$), which we simply denote by C.

Proposition 5.1.8. *Suppose that a function f has a antiderivative on an interval I, and let $\alpha \in \mathbb{R}\setminus\{0\}$ be a constant. Then αf also has an antiderivative on I, and moreover, we have the formula*

$$\int \alpha f(x)dx \quad = \quad \alpha\int f(x)dx,$$

valid on the interval I.

Proof. In what follows, all formulas, and all antiderivatives, are assumed to be on the interval I. Let F be an antiderivative of f, so that

$$\int f(x)dx \quad = \quad F(x)+C.$$

Note that

$$\tfrac{d}{dx}\big(\alpha F(x)\big) \quad \overset{(*)}{=} \quad \alpha F'(x) \quad \overset{(**)}{=} \quad \alpha f(x),$$

where (*) follows from the properties of the derivative (and in particular, from Proposition 4.3.4), whereas (**) follows from the fact that F is an antiderivative of f. Therefore, αF is an antiderivative of αf, and it follows that

$$\int \alpha f(x)dx \quad = \quad \alpha F(x)+C \quad \overset{\alpha\neq 0}{=} \quad \alpha\int f(x)dx.$$

This completes the argument. □

---

[2] For subtraction, we technically need both Proposition 4.3.3 and Proposition 4.3.4, since we need that

$$\begin{aligned}\tfrac{d}{dx}\big(F_1(x)-F_2(x)\big) \quad &= \quad \tfrac{d}{dx}\big(F_1(x)+(-1)F_2(x)\big)\\ &= \quad F_1'(x)+(-1)F_2'(x)\\ &= \quad F_1'(x)-F_2'(x).\end{aligned}$$

Remark: Suppose that a function f has an antiderivative F on an interval I, so that $\int f(x)dx = F(x) + C$, valid on I. The formula from Proposition 5.1.8 fails for $\alpha = 0$, since we have the following:

- $\int 0f(x)dx = \int 0dx = C$;
- $0\int f(x)dx = 0(F(x) + C) = 0$.

Example 5.1.9. *Compute* $\int x + \frac{7}{x}dx$.

Solution. We compute:

$$\int x + \tfrac{7}{x}dx \quad = \quad \int xdx + 7\int \tfrac{dx}{x} \quad = \quad \tfrac{1}{2}x^2 + 7\ln|x| + C.$$

□

5.1.4 The Substitution Rule

The Substitution Rule. *Let* $I, J \subseteq \mathbb{R}$ *be intervals, let* f *be a function that has an antiderivative* F *on the interval* I*, and let* $\varphi : J \to I$ *is a continuously differentiable function.*[3] *Then*

$$\int f\big(\varphi(x)\big)\varphi'(x)dx \quad = \quad F\big(\varphi(x)\big) + C$$

on the interval J.

Proof. By the Chain Rule, we have that

$$\tfrac{d}{dx}F\big(\varphi(x)\big) \quad = \quad F'\big(\varphi(x)\big)\varphi'(x) \qquad \text{for } x \in J.$$

Since F is an antiderivative of f on I, we have that $F'(u) = f(u)$ for all $u \in I$; consequently, $F'(\varphi(x)) = f(\varphi(x))$ for all $x \in J$. It follows that

$$\tfrac{d}{dx}F\big(\varphi(x)\big) \quad = \quad f\big(\varphi(x)\big)\varphi'(x) \qquad \text{for } x \in J.$$

So, $F\big(\varphi(x)\big)$ is an antiderivative of $f\big(\varphi(x)\big)\varphi'(x)$ on the interval J, and it follows that

$$\int f\big(\varphi(x)\big)\varphi'(x)dx \quad = \quad F\big(\varphi(x)\big) + C$$

on the interval J, which is what we needed to show. □

Notation: For $u = \varphi(x)$, where φ is a differentiable function, we write $du(x) = \varphi'(x)dx$, or simply $du = \varphi'(x)dx$. So, under the assumptions of the Substitution Rule, we get the following formula:

$$\int f\big(\color{red}{\varphi(x)}\big)\color{blue}{\varphi'(x)dx} \quad = \quad \int f(\color{red}{u})\color{blue}{du}, \qquad \begin{array}{l} \color{red}{u = \varphi(x)} \\ \color{blue}{du = \varphi'(x)dx.} \end{array}$$

---

[3] So, we are assuming φ is differentiable at all points $x \in J$, and that $\varphi' : J \to \mathbb{R}$ is continuous.

Remark: In applying substitution, the most difficult part tends to be finding the right choice of u. While there are some tricks, there is no algorithm for choosing the right kind of substitution, and so this is largely a matter of trial and error.

Example 5.1.10. *Compute* $\int \frac{x}{1+x^2}dx$.

Solution. We introduce the substitution $u = 1+x^2$ (so, $du = 2xdx$, and therefore $xdx = \frac{1}{2}du$), and we compute:

$$\begin{array}{rcll} \int \frac{x}{1+x^2}dx & = & \frac{1}{2}\int \frac{du}{u} & \begin{array}{l} u = 1+x^2 \\ du = 2xdx \end{array} \\ & = & \frac{1}{2}\ln|u| + C & \\ & = & \frac{1}{2}\ln|1+x^2| + C & \\ & = & \frac{1}{2}\ln(1+x^2) + C & \text{because } 1+x^2 > 0. \end{array}$$

Optional: We can check that our answer is correct by differentiating the answer:

$$\frac{d}{dx}\big(\tfrac{1}{2}\ln(1+x^2)\big) \;=\; \frac{1}{2}\frac{1}{1+x^2}\frac{d}{dx}(1+x^2) \;=\; \frac{1}{2}\frac{1}{1+x^2}\cdot 2x \;=\; \frac{x}{1+x^2}.$$

Since we got the function that we originally integrated, our answer is correct.

Remark: Note that we omitted the constant C when we differentiated to check our answer. This is because the derivative of a constant function is zero, and so the constant C would immediately vanish from our calculation.

Remark: Checking the answer is optional because there were no "one way implications" in our calculation. So, unless we made a legitimate mistake in our calculation, our answer will be correct.[4] The purpose of checking the answer in this case is simply to make sure we did not make a mistake. □

Example 5.1.11. *Compute* $\int x^2e^{-x^3}dx$.

Solution. We compute:

$$\begin{array}{rcll} \int x^2e^{-x^3}dx & = & -\frac{1}{3}\int e^u du & \begin{array}{l} u = -x^3 \\ du = -3x^2dx \end{array} \\ & = & -\frac{1}{3}e^u + C & \\ & = & -\frac{1}{3}e^{-x^3} + C. & \end{array}$$

---

[4] This is not always the case when we have "one way implications." For instance, $x = y$ implies $x^2 = y^2$, but the converse need not hold (indeed, $x^2 = y^2$ only implies $|x| = |y|$).

Optional: We check our answer as follows:

$$\tfrac{d}{dx}(-\tfrac{1}{3}e^{-x^3}) \quad = \quad -\tfrac{1}{3}e^{-x^3}(-3x^2) \quad = \quad x^2e^{-x^3}.$$

So, our answer is correct. □

Example 5.1.12. *Let* $a \in (0,1) \cup (1,+\infty)$ *be a constant. Compute* $\int a^x dx$*.*

Solution. We compute:

$$\begin{aligned} \textstyle\int a^x dx &= \textstyle\int e^{x\ln a}dx & & u = x\ln a \\ & & & du = \ln a dx \\ &= \tfrac{1}{\ln a}\textstyle\int e^u du \\ &= \tfrac{1}{\ln a}e^u + C \\ &= \tfrac{1}{\ln a}e^{x\ln a} + C \\ &= \tfrac{a^x}{\ln a} + C. \end{aligned}$$

Optional: We check our answer as follows:

$$\tfrac{d}{dx}\big(\tfrac{a^x}{\ln a}\big) \quad = \quad \tfrac{1}{\ln a}\big(\tfrac{d}{dx}a^x\big) \quad = \quad \tfrac{1}{\ln a}\big(\tfrac{d}{dx}e^{x\ln a}\big) \quad = \quad \tfrac{1}{\ln a}\cdot\underbrace{e^{x\ln a}}_{=a^x}\ln a \quad = \quad a^x.$$

So, our answer is correct. □

Example 5.1.13. *Compute* $\int \tan x dx$*.*

Solution. We compute:

$$\begin{aligned} \textstyle\int \tan x dx &= \textstyle\int \tfrac{\sin x}{\cos x}dx & & u = \cos x \\ & & & du = -\sin x dx \\ &= \textstyle\int \tfrac{-du}{u} \\ &= -\ln|u| + C \\ &= -\ln|\cos x| + C. \end{aligned}$$

Optional: We check our answer as follows:

$$\begin{aligned} \tfrac{d}{dx}\big(-\ln|\cos x|\big) &= -\tfrac{d}{dx}\ln|\cos x| \\ &= -\tfrac{1}{\cos x}\big(\tfrac{d}{dx}\cos x\big) & & \text{by Proposition 5.1.6} \\ & & & \text{and the Chain Rule} \\ &= -\tfrac{1}{\cos x}(-\sin x) \\ &= \tan x. \end{aligned}$$

So, our answer is correct. □

5.1.5 Integration by parts

Integration by parts. *Let function u and v be continuously differentiable on an interval I. Then*

$$\int u(x)dv(x) \quad = \quad u(x)v(x) - \int v(x)du(x)$$

on the interval I.

Notation: To simplify notation, we may write the formula above simply as

$$\int udv \quad = \quad uv - \int vdu$$

Proof. By the Product Rule, we have that

$$(u(x)v(x))' \quad = \quad u'(x)v(x) + u(x)v'(x).$$

When we integrate, we get that

$$\int (u(x)v(x))'dx \quad = \quad \int u'(x)v(x)dx + \int u(x)v'(x)dx,$$

or equivalently,

$$\int u(x)v'(x)dx \quad = \quad \int (u(x)v(x))'dx - \int u'(x)v(x)dx.$$

Now, by Proposition 5.1.3, we have that $\int (u(x)v(x))'dx = u(x)v(x) + C$. However, both $\int u(x)v'(x)dx$ and $\int u'(x)v(x)dx$ have arbitrary additive constants built into them, and so we can simply write

$$\int u(x)v'(x)dx \quad = \quad u(x)v(x) - \int u'(x)v(x)dx,$$

that is,

$$\int u(x)dv(x) \quad = \quad u(x)v(x) - \int v(x)du(x).$$

This completes the argument. □

Remark: As with substitution, the hardest part of using integration by parts is figuring out how to choose u and v. The rule of thumb is that we want v to be easy to integrate, and we want du to be simpler than u. Apart from that, this is largely a matter of trial and error.

Remark: When integrating by parts, we choose suitable u and dv, and we compute du and v. For the remainder of this section, we will color code this as follows:

$$\begin{array}{ll} u = ??, & v = ?? \\ du = ??, & dv = ?? \end{array}$$

where the question marks are replaced by the appropriate formulas. Here, we choose the red terms, and based on those, we compute the blue terms.

Example 5.1.14. *Compute* $\int xe^x dx$.

Solution. We compute:

$$\begin{array}{rcll} \int xe^x dx & = & xe^x - \int e^x dx & \begin{array}{ll} u = x, & v = e^x \\ du = dx, & dv = e^x dx \end{array} \\ & = & xe^x - e^x + C. & \end{array}$$

Optional: We check our answer as follows:

$$\tfrac{d}{dx}\big(xe^x - e^x\big) \quad = \quad \big(e^x + xe^x\big) - e^x \quad = \quad xe^x.$$

So, our answer is correct. ☐

Example 5.1.15. *Compute* $\int x^2 \sin x dx$.

Solution. In this example, we perform integration by parts twice, as follows:

$$\begin{array}{rll} & \int x^2 \sin x dx & \begin{array}{ll} u_1 = x^2, & v_1 = -\cos x \\ du_1 = 2x dx, & dv_1 = \sin x dx \end{array} \\ = & -x^2 \cos x + 2\int x \cos x dx & \begin{array}{ll} u_2 = x, & v_2 = \sin x \\ du_2 = dx, & dv_2 = \cos x dx \end{array} \\ = & -x^2 \cos x + 2(x \sin x - \int \sin x dx) & \\ = & -x^2 \cos x + 2x \sin x + 2\cos x + C. & \end{array}$$

Optional: We check our answer as follows:

$$\begin{array}{rl} & \tfrac{d}{dx}\big(-x^2 \cos x + 2x \sin x + 2\cos x\big) \\ = & \big(-2x\cos x - x^2(-\sin x)\big) + \big(2\sin x + 2x\cos x\big) + 2(-\sin x) \\ = & -2x\cos x + x^2 \sin x + 2\sin x + 2x\cos x - 2\sin x \\ = & 2x\sin x. \end{array}$$

So, our answer is correct. ☐

Recursive formulas. We now take a look at a couple of more complicated examples, which rely on recursive formulas.

Example 5.1.16. *Compute* $\int e^x \cos x dx$.

Solution. We compute:

$$\begin{array}{rcl} I & := & \int e^x \cos x dx \\ & = & e^x \sin x - \int e^x \sin x dx. \end{array} \qquad \begin{array}{ll} u = e^x, & v = \sin x \\ du = e^x dx, & dv = \cos x dx \end{array}$$

Now, let us try to compute $\int e^x \sin x dx$:

$$\begin{array}{rcl} \int e^x \sin x dx & = & -e^x \cos x + \int e^x \cos x dx \\ & = & -e^x \cos x + I. \end{array} \qquad \begin{array}{ll} u = e^x, & v = -\cos x \\ du = e^x dx, & dv = \sin x dx \end{array}$$

So now we have that

$$\begin{array}{rcl} I & = & e^x \sin x - \int e^x \sin x dx \\ & = & e^x \sin x - \big(- e^x \cos x + I\big) \\ & = & e^x(\sin x + \cos x) - I. \end{array}$$

Here, we must be careful: I is an indefinite integral, and so it has a "built in" arbitrary additive constant C, and this constant C need not be the same on both sides of the equation! So, when we solve for I, we account for the constant as follows:

$$2I \quad = \quad e^x(\sin x + \cos x) + C.$$

This yields

$$I \quad = \quad \tfrac{1}{2}e^x(\sin x + \cos x) + C,$$

and we are done.

Remark: In the last line of the computation above, we write C, and not $\frac{1}{2}C$. This is because C is an arbitrary constant, which means that $\frac{1}{2}C$ is also an arbitrary constant. So, we may just as well write C, and not the more complicated $\frac{1}{2}C$.

Optional: We check our answer as follows:

$$\tfrac{d}{dx}\big(\tfrac{1}{2}e^x(\sin x + \cos x)\big) \quad = \quad \tfrac{1}{2}e^x(\sin x + \cos x) + \tfrac{1}{2}e^x(\cos x - \sin x) \quad = \quad e^x \cos x.$$

So, our answer is correct. □

Example 5.1.17. *Show that for all positive integers n, we have that*

$$\int \ln^n x dx \quad = \quad x \ln^n x - n \int \ln^{n-1} x dx.$$

Then, using the recursive formula above, compute $\int \ln^3 x dx$.

Notation: For $n \in \mathbb{N}_0$ and $x \in (0, +\infty)$, we write $\ln^n x := (\ln x)^n$.

Solution. First, for a positive integer n, we have:

$$\begin{aligned} \int \ln^n x dx &= x \ln^n x - \int x \tfrac{n \ln^{n-1} x}{x} dx \\ &= x \ln^n x - n \int \ln^{n-1} x dx, \end{aligned} \qquad \begin{array}{ll} u = \ln^n x, & v = x \\ du = \tfrac{n \ln^{n-1} x}{x} dx, & dv = dx \end{array}$$

which proves our recursive formula. Using this formula, we now compute:

$$\begin{aligned} \int \ln^3 x dx &= x \ln^3 x - 3 \int \ln^2 x dx \\ &= x \ln^3 x - 3\Big(x \ln^2 x - 2 \int \ln x dx\Big) \\ &= x \ln^3 x - 3x \ln^2 x + 6 \int \ln x dx \\ &= x \ln^3 x - 3x \ln^2 x + 6\Big(x \ln x - \int dx\Big) \\ &= x \ln^3 x - 3x \ln^2 x + 6x \ln x - 6x + C. \end{aligned}$$

Optional: We check our answer as follows:

$$\begin{aligned} & \tfrac{d}{dx}\big(x \ln^3 x - 3x \ln^2 x + 6x \ln x - 6x\big) \\ = \;& \ln^3 x + x \cdot 3 \ln^2 x \cdot \tfrac{1}{x} - 3 \ln^2 x - 3x(2 \ln x) \cdot \tfrac{1}{x} + 6(\ln x) + 6x \cdot \tfrac{1}{x} - 6 \\ = \;& \ln^3 x + 3 \ln^2 x - 3 \ln^2 x - 6 \ln x + 6 \ln x + 6 - 6 \quad = \quad \ln^3 x. \end{aligned}$$

So, our answer is correct. □