



UNIVERSITATEA BABEȘ-BOLYAI  
BABEȘ-BOLYAI TUDOMÁNYEGYETEM  
BABEȘ-BOLYAI UNIVERSITÁT  
BABEȘ-BOLYAI UNIVERSITY  
TRADITIO ET EXCELLENTIA

# Mathematical Analysis

1st Year Computer Science

Mihai Nechita \*

Last updated: December 13, 2023

## Contents

Real numbers	2
Sequences	5
Series. Power series	8
Limits, continuity, differentiability	17
Taylor series	21
Riemann integrals. Improper integrals	25
The Euclidean space $\mathbb{R}^n$	29
Functions of several variables. Limits and continuity	33
Partial derivatives and differentiability in $\mathbb{R}^n$ . Gradient descent	36
Higher order derivatives. Local extremum conditions. Optimization	43

---

\*[math.ubbcluj.ro/~mihai.nechita](mailto:math.ubbcluj.ro/~mihai.nechita). If you find any typos, please let me know on Teams or by email.

## ❖ Real numbers

Let us start with some standard notation:  $\emptyset$  is the empty set;  $\mathbb{N} = \{1, 2, \dots\}$  the set of natural numbers;  $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\} = \{m - n \mid m, n \in \mathbb{N}\}$  the set of integers;  $\mathbb{Q} = \{\frac{m}{n} \mid m, n \in \mathbb{Z}, n \neq 0\}$  the set of rational numbers;  $\mathbb{R}$  the set of real numbers.

You are very much used with the real numbers. However, it is not trivial to define them in a rigorous way, see for example [1] or [2] for different methods of constructing  $\mathbb{R}$  from  $\mathbb{Q}$ . We will straightforwardly start working with the real numbers – for this reason some of their properties, e.g. [definition 1.5](#), will simply be given as definitions.

**Definition 1.1.** Let  $A$  be a subset of  $\mathbb{R}$ , denoted as  $A \subseteq \mathbb{R}$ . We define  $x \in \mathbb{R}$  to be

a *lower bound* for  $A$  if  $x \leq a, \forall a \in A$ ; an *upper bound* for  $A$  if  $x \geq a, \forall a \in A$ .

We define

$lb(A) := \{x \in \mathbb{R} \mid x \leq a, \forall a \in A\}$  the set of lower bounds of  $A$ ,

$ub(A) := \{x \in \mathbb{R} \mid x \geq a, \forall a \in A\}$  the set of upper bounds of  $A$ .

We define  $x \in \mathbb{R}$  to be

the *minimum* of  $A$  if  $x \in lb(A) \cap A$ ; the *maximum* of  $A$  if  $x \in ub(A) \cap A$ ,

denoted by  $\min(A)$ , respectively  $\max(A)$ . In other words, we have that

$\min(A) \in A$  and  $\min(A) \leq a, \forall a \in A$ ;  $\max(A) \in A$  and  $\max(A) \geq a, \forall a \in A$ .

Note that there are sets which do not have minimum or maximum, e.g.  $(0, 1)$ .

**Definition 1.2.** A set  $A \subseteq \mathbb{R}$  is defined to be

- bounded (from) below if  $lb(A) \neq \emptyset$ ;
- bounded (from) above if  $ub(A) \neq \emptyset$ ;
- bounded if it is both bounded below and above;
- unbounded if it is not bounded.

**Definition 1.3.** We say that  $x \in \mathbb{R}$  is the *supremum* of  $A \subseteq \mathbb{R}$ ,  $x := \sup(A)$ , if and only if:

1.  $x \geq a, \forall a \in A$ , that is  $x \in ub(A)$ .
2. if  $u$  is an upper bound for  $A$ , then  $x \leq u$ .

The supremum is *the least upper bound*, i.e.  $\sup(A) := \min(ub(A))$ .

**Definition 1.4.** We say that  $x \in \mathbb{R}$  is the *infimum* of  $A \subseteq \mathbb{R}$ ,  $x := \inf(A)$ , if and only if:

1.  $x \leq a$ ,  $\forall a \in A$ , that is  $x \in lb(A)$ .
2. if  $u$  is a lower bound for  $A$ , then  $x \geq u$ .

The infimum is *the greatest lower bound*, i.e.  $\inf(A) := \max(lb(A))$ .

**Definition 1.5** (Completeness Axiom). Every set  $A \subseteq \mathbb{R}$  that is bounded above has a supremum. Similarly, every set  $A \subseteq \mathbb{R}$  that is bounded below has an infimum.

Note that if  $A$  has a maximum, then  $\sup(A) = \max(A)$ . Similarly, if  $A$  has a minimum, then  $\inf(A) = \min(A)$ . Also, if  $\sup(A) \in A$ , then  $\max(A) = \sup(A)$ .

**Example 1.6.** (a)  $A = \{\frac{1}{n} \mid n \in \mathbb{N}\}$ ,  $\sup(A) = 1 = \max(A)$ ,  $\inf(A) = 0$ ,  $\nexists \min(A)$ .

(b)  $A = \{x \in \mathbb{Q} \mid x^2 \leq 2\}$ ,  $\sup(A) = \sqrt{2}$ ,  $\nexists \max(A)$ ,  $\inf(A) = -\sqrt{2}$ ,  $\nexists \min(A)$ .

**Theorem 1.7.** Let  $A \subseteq \mathbb{R}$  be a bounded set. For  $\sup(A)$  and  $\inf(A)$  the following are true:

$$\forall \varepsilon > 0, \exists x \in A \text{ such that } \sup(A) - \varepsilon < x,$$

$$\forall \varepsilon > 0, \exists x \in A \text{ such that } x < \inf(A) + \varepsilon.$$

*Proof.* By definition, for any  $y < \sup(A)$ , say  $y = \sup(A) - \varepsilon$  with  $\varepsilon > 0$ , we have that  $y \notin ub(A)$ . Hence there exists  $x \in A$  such that  $y < x$ . Similar proof for  $\inf(A)$ .  $\square$

**Proposition 1.8.** Let  $A \subseteq B \subseteq \mathbb{R}$  be (nonempty) bounded sets. Then

$$\inf(B) \leq \inf(A) \leq \sup(A) \leq \sup(B)$$

and

$$\sup(A \cup B) = \max\{\sup(A), \sup(B)\},$$

$$\inf(A \cup B) = \min\{\inf(A), \inf(B)\}.$$

*Proof.* It follows directly from the definitions.  $\square$

**Definition 1.9.** Define the *extended real line*  $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ , where  $\infty$  and  $-\infty$  are such that

$$\forall x \in \mathbb{R}, -\infty < x < \infty.$$

If a set  $A$  is not bounded above, we define  $\sup(A) := \infty$ .

If a set  $A$  is not bounded below, we define  $\inf(A) := -\infty$ .

[Seminar] The empty set  $\emptyset$  is bounded by any number. In  $\overline{\mathbb{R}}$ ,  $\sup(\emptyset) = -\infty$ ,  $\inf(\emptyset) = \infty$ .

**Definition 1.10.** A set  $V \subseteq \mathbb{R}$  is a *neighborhood* (vecinity) of  $x \in \mathbb{R}$  if

$$\exists \varepsilon > 0 \text{ such that } (x - \varepsilon, x + \varepsilon) \subseteq V.$$

A set  $V \subseteq \mathbb{R}$  is a *neighborhood* of  $\infty$  if  $\exists a \in \mathbb{R}$  such that  $(a, \infty) \subseteq V$ .

A set  $V \subseteq \mathbb{R}$  is a *neighborhood* of  $-\infty$  if  $\exists a \in \mathbb{R}$  such that  $(-\infty, a) \subseteq V$ .

We denote all the neighborhoods of  $x$  by  $\mathcal{V}(x) := \{V \subseteq \mathbb{R} \mid V \text{ is a neighborhood of } x\}$ .

**Definition 1.11.** Let  $A \subseteq \mathbb{R}$ . The following set is called the *interior* of  $A$

$$\text{int}(A) := \{x \in \mathbb{R} \mid \exists V \in \mathcal{V}(x) \text{ such that } V \subseteq A\},$$

and the following set is called the *closure* of  $A$

$$\text{cl}(A) := \{x \in \mathbb{R} \mid \forall V \in \mathcal{V}(x), V \cap A \neq \emptyset\}.$$

**Proposition 1.12.** For any  $A \subseteq \mathbb{R}$ , it holds that  $\text{int}(A) \subseteq A \subseteq \text{cl}(A)$ .

*Proof.* To prove that  $\text{int}(A) \subseteq A$  we prove that if  $x \in \text{int}(A)$ , then  $x \in A$ . Let  $x \in \text{int}(A)$ , then  $\exists \varepsilon > 0$  such that  $(x - \varepsilon, x + \varepsilon) \subseteq A$ . Since  $x \in (x - \varepsilon, x + \varepsilon)$ , we have that  $x \in A$ .

To prove that  $A \subseteq \text{cl}(A)$  we show that if  $x \in A$ , then  $x \in \text{cl}(A)$ . Let  $x \in A$ . Then for any  $V \in \mathcal{V}(x)$  it holds that  $x \in V$ , giving that  $x \in V \cap A$ . Hence  $x \in \text{cl}(A)$  since  $V \cap A \neq \emptyset$ .  $\square$

**Definition 1.13.** If  $A = \text{int}(A)$ , then  $A$  is called *open*. If  $A = \text{cl}(A)$ , then  $A$  is called *closed*.

**Remark 1.14.** To prove that a set  $A$  is open, it is sufficient to prove that  $A \subseteq \text{int}(A)$ .

To prove that a set  $A$  is closed, it is sufficient to prove that  $\text{cl}(A) \subseteq A$ .

**Proposition 1.15.** The following statements are true:

- The complement of an open set is closed.
- The complement of a closed set is open.

*Proof.* Let us prove the first statement, the other one being similar. Consider  $A$  an open set, i.e.  $A = \text{int}(A)$ , and denote by  $A^c = \{x \in \mathbb{R} \mid x \notin A\}$  its complement. To prove that  $A^c$  is closed, we prove that  $\text{cl}(A^c) \subseteq A^c$ . Consider  $x \in \text{cl}(A^c)$  and let's assume that  $x \notin A^c$ , i.e.  $x \in A$ , aiming to obtain a contradiction. Since  $A$  is open, there exists  $V \in \mathcal{V}(x)$  such that  $V \subseteq A$ , giving that  $V \cap A^c = \emptyset$ : contradiction with  $x \in \text{cl}(A^c)$ . Hence the assumption  $x \notin A^c$  is false, and we have that if  $x \in \text{cl}(A^c)$ , then  $x \in A^c$ . In other words,  $\text{cl}(A^c) \subseteq A^c$ .  $\square$

**Proposition 1.16.** Any union of open sets is open. Any intersection of closed sets is closed. Any finite intersection of open sets is open. Any finite union of closed sets is closed.

*Proof.* (Optional) Left as extra homework.  $\square$

## ❖ Sequences

A set  $\{x_n \mid n \in \mathbb{N}\}$  is called a sequence and is denoted by  $(x_n)_{n \in \mathbb{N}}$  or simply  $(x_n)$ . A sequence  $(x_n)$  is bounded above (or below) if the set  $\{x_n \mid n \in \mathbb{N}\}$  is bounded above (or below). A sequence  $(x_n)$  is increasing if  $x_{n+1} \geq x_n$ ,  $\forall n \in \mathbb{N}$ , and decreasing if  $x_{n+1} \leq x_n$ ,  $\forall n \in \mathbb{N}$ . A sequence is monotone if it is either increasing or decreasing.

**Definition 2.1.** A sequence  $(x_n)$  has a limit  $\ell \in \overline{\mathbb{R}}$ , and we write  $\lim_{n \rightarrow \infty} x_n = \ell$  or  $x_n \rightarrow \ell$ , if

$$\forall V \in \mathcal{V}(\ell), \exists N_V \in \mathbb{N} \text{ such that } x_n \in V, \forall n \geq N_V.$$

If  $\ell \in \mathbb{R}$ , we say that  $(x_n)$  converges to  $\ell$ :  $\forall \varepsilon > 0$ ,  $\exists N_\varepsilon \in \mathbb{N}$  such that  $|x_n - \ell| < \varepsilon$ ,  $\forall n \geq N_\varepsilon$ .

$x_n \rightarrow \infty$  if  $\forall a > 0$ ,  $\exists N_a \in \mathbb{N}$  such that  $x_n > a$ ,  $\forall n \geq N_a$ .

$x_n \rightarrow -\infty$  if  $\forall a < 0$ ,  $\exists N_a \in \mathbb{N}$  such that  $x_n < a$ ,  $\forall n \geq N_a$ .

**Proposition 2.2.** A sequence  $(x_n)$  converges to  $\ell \in \mathbb{R}$  if and only if  $\lim_{n \rightarrow \infty} |x_n - \ell| = 0$ .

**Proposition 2.3.** Any convergent sequence is bounded.

*Proof.* TBA (left to the reader). □

**Theorem 2.4** (Weierstrass). Any monotone and bounded sequence is convergent.

*Proof.* Assume that the sequence is increasing, for example. Let  $S = \{x_n \mid n \in \mathbb{N}\}$  and consider  $\sup(S) \in \mathbb{R}$  (we know that  $S$  is bounded). From [theorem 1.7](#) we have that

$$\forall \varepsilon > 0, \exists N_\varepsilon \in \mathbb{N} \text{ such that } \sup(S) - \varepsilon < x_{N_\varepsilon}.$$

As  $(x_n)$  is increasing,  $\sup(S) - \varepsilon < x_{N_\varepsilon} \leq x_n$   $\forall n \geq N_\varepsilon$ . Hence  $\sup(S) - x_n < \varepsilon$ ,  $\forall n \geq N_\varepsilon$ .

The sequence converges to  $\sup(S)$  by [definition 2.1](#). Similarly, a decreasing and bounded sequence converges to its infimum. □

**Proposition 2.5.** Any monotone sequence has a limit in  $\overline{\mathbb{R}}$ .

*Proof.* If the sequence is bounded and monotone, then it is convergent by the Weierstrass theorem. If the sequence is unbounded and monotone, then its limit will be infinite. □

**Theorem 2.6** (Squeeze/Sandwich theorem). Let  $(x_n)$ ,  $(y_n)$ ,  $(z_n)$  be sequences for which there is an  $n_0 \in \mathbb{N}$  such that

$$x_n \leq y_n \leq z_n, \forall n \geq n_0,$$

and

$$\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} z_n.$$

Then

$$\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} y_n = \lim_{n \rightarrow \infty} z_n.$$

*Proof.* Let  $\ell := \lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} z_n$  and assume first that  $\ell \in \mathbb{R}$ . Let  $\varepsilon > 0$ . Then

$$\exists N_1 \in \mathbb{N} \text{ such that } |x_n - \ell| < \varepsilon, \forall n \geq N_1$$

and

$$\exists N_2 \in \mathbb{N} \text{ such that } |z_n - \ell| < \varepsilon, \forall n \geq N_2.$$

Taking  $N_\varepsilon := \max\{N_1, N_2\}$ , we have that

$$|y_n - \ell| \leq \max\{|x_n - \ell|, |z_n - \ell|\} < \varepsilon, \forall n \geq N_\varepsilon,$$

hence the conclusion. When  $\ell$  is infinite the proof is similar.  $\square$

**Theorem 2.7** (Cantor's nested intervals). Let  $(a_n)$  be increasing and  $(b_n)$  decreasing such that  $a_n \leq a_{n+1} \leq b_{n+1} \leq b_n, \forall n \in \mathbb{N}$ . Consider the closed intervals  $I_n := [a_n, b_n]$ , with  $I_{n+1} \subseteq I_n$ . If  $\lim_{n \rightarrow \infty} (b_n - a_n) = 0$ , then there exists  $x \in \mathbb{R}$  such that

$$\bigcap_{n=1}^{\infty} I_n = \{x\}.$$

*Proof.* Consider the bounded sets  $A := \{a_n \mid n \in \mathbb{N}\}$  and  $B := \{b_n \mid n \in \mathbb{N}\}$ . For any  $k \in \mathbb{N}$ , we have that

$$a_k \leq \sup(A) \leq b_k$$

and

$$b_k \geq \inf(B) \geq a_k.$$

Hence by the squeeze theorem we have that  $\sup(A) = \inf(B)$  and  $\bigcap_{n=1}^{\infty} I_n = \{\sup(A)\}$ .  $\square$

**Theorem 2.8** (Bolzano-Weierstrass). Any bounded sequence has a convergent subsequence.

*Proof.* Consider the bounded set  $A := \{x_n \mid n \in \mathbb{N}\}$ . Let  $a_1 := \inf(A)$  and  $b_1 := \sup(A)$ , and define  $I_1 := [a_1, b_1]$ . Bisect  $I_1$  and notice that at least one of the two halves must contain infinitely many terms from the sequence. Take  $I_2 := [a_2, b_2]$  to be the half that does. Continuing this procedure we obtain for each  $k \in \mathbb{N}$  an interval  $I_k := [a_k, b_k]$  containing (at least) a term  $x_{n_k} \in A$ , such that  $I_{k+1} \subseteq I_k$  and  $b_k - a_k \rightarrow 0$ .

From Cantor's nested intervals [theorem 2.7](#) we have that there exists  $x \in \mathbb{R}$  such that  $\bigcap_{n=1}^{\infty} I_n = \{x\}$ , and hence the subsequence  $(x_{n_k})$  converges to  $x$ .  $\square$

**Definition 2.9.** For a sequence  $(x_n)$  we define the set of its *limit points* by

$$\text{LIM}(x_n) := \{x \in \overline{\mathbb{R}} \mid \text{there exists a subsequence } (x_{n_k}) \text{ s.t. } x_{n_k} \rightarrow x\},$$

and

$$\liminf_{n \rightarrow \infty} x_n := \inf (\text{LIM}(x_n)),$$

$$\limsup_{n \rightarrow \infty} x_n := \sup (\text{LIM}(x_n)).$$

**Example 2.10.** For  $x_n = \frac{(-1)^n n}{n+1}$ ,  $\text{LIM}(x_n) = \{-1, 1\}$ ,  $\liminf_{n \rightarrow \infty} x_n = -1$ ,  $\limsup_{n \rightarrow \infty} x_n = 1$ .

**Proposition 2.11.**  $\lim_{n \rightarrow \infty} x_n = \ell \in \overline{\mathbb{R}}$  if and only if  $\liminf_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} x_n = \ell$ .

**Definition 2.12** (Cauchy sequence). A sequence  $(x_n)$  is called *Cauchy* (or *fundamental*) if

$$\forall \varepsilon > 0, \exists N_\varepsilon \in \mathbb{N} \text{ such that } |x_m - x_n| < \varepsilon, \forall m, n \geq N_\varepsilon.$$

**Proposition 2.13.** Any Cauchy sequence is bounded.

*Proof.* For  $\varepsilon = 1$ , there exists  $N_1 \in \mathbb{N}$  such that  $|x_m - x_n| < 1, \forall m, n \geq N_1$ . In particular,  $|x_n - x_{N_1}| < 1, \forall n \geq N_1$ , hence the terms after index  $N_1$  are bounded. The terms before index  $N_1$  are also bounded since there is a finite number of them. We thus conclude that the entire sequence is bounded.  $\square$

**Theorem 2.14.** A sequence is convergent if and only if it is Cauchy.

*Proof.* Let's consider first a convergent sequence  $(x_n)$  with  $x_n \rightarrow \ell$ . For any  $\varepsilon > 0$ , there exists  $N_\varepsilon \in \mathbb{N}$  such that  $|x_n - \ell| < \frac{\varepsilon}{2}$ , for any  $n \geq N_\varepsilon$ . Then  $|x_m - x_n| \leq |x_m - \ell| + |x_n - \ell| < \varepsilon$ , for any  $n \geq N_\varepsilon$ . Hence the sequence  $(x_n)$  is Cauchy.

Assume now that  $(x_n)$  is a Cauchy sequence. From the previous proposition we have that  $(x_n)$  must be bounded, and thus it has a convergent subsequence  $(x_{n_k}), x_{n_k} \rightarrow x \in \mathbb{R}$ . Let  $\varepsilon > 0$ . There exists thus  $K_\varepsilon \in \mathbb{N}$  such that  $|x_{n_k} - x| < \varepsilon, \forall k \geq K_\varepsilon$ . Also, there exists  $N_\varepsilon \in \mathbb{N}$  such that  $|x_m - x_n| < \varepsilon, \forall m, n \geq N_\varepsilon$ . In particular,  $|x_{n_k} - x_n| < \varepsilon, \forall k, n \geq N_\varepsilon$ . Hence  $|x_n - x| \leq |x_n - x_{n_k}| + |x_{n_k} - x| < 2\varepsilon, \forall n \geq \max\{K_\varepsilon, N_\varepsilon\}$ , meaning that  $x_n \rightarrow x$ .  $\square$

**Example 2.15.** The sequence defined by  $x_n = 1 + \frac{1}{2} + \dots + \frac{1}{n}$  is not convergent. Indeed, one can see, for example, that

$$x_{2n} - x_n = \frac{1}{n+1} + \dots + \frac{1}{2n} > \frac{n}{2n},$$

hence  $x_{2n} - x_n > \frac{1}{2}$  for any  $n \in \mathbb{N}$ . Thus  $(x_n)$  is not convergent since it is not Cauchy.

## ❖ Series. Power series

For a sequence  $(x_n)$ , the sum  $\sum_{n=1}^{\infty} x_n$  is called a *series* and  $s_n := \sum_{k=1}^n x_k$  is called the *partial sum of the series*. The summation in a series can start from any index, not necessarily 0 or 1. A series is also often written as  $\sum_{n \geq 1} x_n$ .

**Definition 3.1.** The series  $\sum_{n=1}^{\infty} x_n$  converges iff the sequence of partial sums  $(s_n)$  converges.

**Example 3.2.** The *geometric series*  $\sum_{n=0}^{\infty} q^n$  converges iff  $|q| < 1$ , with sum  $\frac{1}{1-q}$ .

**Example 3.3.** The *harmonic series*  $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$  diverges since  $(s_n)$  is not a Cauchy sequence.

**Example 3.4** (Euler's number).  $e = \sum_{n=0}^{\infty} \frac{1}{n!}$ .

*Proof.* (Optional) Let  $s_n = 1 + 1 + \frac{1}{2!} + \dots + \frac{1}{n!}$ . Start from  $\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e$  and expand

$$\left(1 + \frac{1}{n}\right)^n = 1 + 1 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \dots + \frac{1}{n!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdot \dots \cdot \left(1 - \frac{n-1}{n}\right) \leq s_n.$$

We have that

$$\left(1 + \frac{1}{n}\right)^n \leq s_n.$$

Consider now an index  $k \geq n$ . We have that

$$\left(1 + \frac{1}{k}\right)^k \geq 1 + 1 + \frac{1}{2!}\left(1 - \frac{1}{k}\right) + \dots + \frac{1}{n!}\left(1 - \frac{1}{k}\right)\left(1 - \frac{2}{k}\right) \cdot \dots \cdot \left(1 - \frac{n-1}{k}\right)$$

and taking  $k \rightarrow \infty$  we obtain that  $e \geq s_n$ . We conclude with the squeeze theorem for

$$\left(1 + \frac{1}{n}\right)^n \leq s_n \leq e,$$

obtaining that the series  $\sum_{n=0}^{\infty} \frac{1}{n!}$  converges and its sum is  $e$ . □

**Proposition 3.5.** If the series  $\sum_{n=1}^{\infty} x_n$  is convergent, then  $\lim_{n \rightarrow \infty} x_n = 0$ .

*Proof.* Consider the partial sum  $s_n$ . We have that  $x_n = s_n - s_{n-1}$ , hence the conclusion. □



It thus follows that if  $\lim_{n \rightarrow \infty} x_n \neq 0$ , then the series  $\sum_{n=1}^{\infty} x_n$  is divergent.

**Example 3.6.** Series like  $\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \sum_{n=1}^{\infty} \left( \frac{1}{n} - \frac{1}{n+1} \right) = 1$  are called *telescoping series*. The partial sum of a telescoping series can be easily computed since after cancellations the only remaining terms are the first one and the last one.

If the sequence  $(x_n)$  has only nonnegative terms  $x_n \geq 0$ , then the sequence of partial sums  $(s_n)$  is increasing. The series  $\sum_{n=1}^{\infty} x_n$  then converges iff  $(s_n)$  is bounded.

**Theorem 3.7** (Comparison test). Let  $\sum_{n=1}^{\infty} x_n, \sum_{n=1}^{\infty} y_n$  be series with nonnegative terms. If there is an  $n_0 \in \mathbb{N}$  such that

$$x_n \leq y_n, \forall n \geq n_0, \text{ then}$$

(a) If  $\sum_{n=1}^{\infty} y_n$  converges, then  $\sum_{n=1}^{\infty} x_n$  also converges.

(b) If  $\sum_{n=1}^{\infty} x_n$  diverges, then  $\sum_{n=1}^{\infty} y_n$  also diverges.

*Proof.* Consider the sequences of partial sums. In case (a), both sequences are bounded. In case (b), both sequences are unbounded.  $\square$

**Example 3.8.** If  $p \leq 1$ , then  $\sum_{n=1}^{\infty} \frac{1}{n^p} = \infty$  since  $\frac{1}{n^p} \geq \frac{1}{n}$  and  $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$ . E.g.  $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} = \infty$ .

**Theorem 3.9.** Let  $\sum_{n=1}^{\infty} x_n, \sum_{n=1}^{\infty} y_n$  be series with nonnegative terms. If

$$\lim_{n \rightarrow \infty} \frac{x_n}{y_n} = \ell, \text{ then}$$

- if  $\ell \in (0, \infty)$ , then the series  $\sum_{n=1}^{\infty} x_n$  and  $\sum_{n=1}^{\infty} y_n$  have the same nature.
- if  $\ell = 0$ , then if the series  $\sum_{n=1}^{\infty} y_n$  converges, the series  $\sum_{n=1}^{\infty} x_n$  also converges.
- if  $\ell = \infty$ , then if the series  $\sum_{n=1}^{\infty} y_n$  diverges, the series  $\sum_{n=1}^{\infty} x_n$  also diverges.

**Theorem 3.10** (Ratio test). Let  $\sum_{n=1}^{\infty} x_n$  be a series with nonnegative terms such that

$$\lim_{n \rightarrow \infty} \frac{x_{n+1}}{x_n} = \ell.$$

- If  $\ell < 1$ , then the series  $\sum_{n=1}^{\infty} x_n$  is convergent.
- If  $\ell > 1$ , then the series  $\sum_{n=1}^{\infty} x_n$  is divergent.

The test is *inconclusive* when  $\ell = 1$ .

*Proof.* The idea is that  $\sum_{n \geq 1} x_n$  behaves like a geometric series with ratio  $\ell$ . We will only give a proof when  $\ell < 1$ , the other case being similar.

Take  $\varepsilon > 0$  such that  $q := \ell + \varepsilon < 1$ . There exists  $N \in \mathbb{N}$  such that

$$\frac{x_{n+1}}{x_n} - \ell < \varepsilon, \forall n \geq N,$$

giving that  $x_{n+1} < x_n \cdot q$ ,  $\forall n \geq N$ . Hence  $x_n < q^{n-N} x_N$ , that is  $x_n < q^n \frac{x_N}{q^N}$ . Since  $q < 1$ , the series converges by comparison with the geometric series  $\sum_{n \geq 1} q^n$ .  $\square$

**Theorem 3.11** (Root test). Let  $\sum_{n=1}^{\infty} x_n$  be a series with nonnegative terms such that

$$\lim_{n \rightarrow \infty} \sqrt[n]{x_n} = \ell.$$

- If  $\ell < 1$ , then the series  $\sum_{n=1}^{\infty} x_n$  is convergent.
- If  $\ell > 1$ , then the series  $\sum_{n=1}^{\infty} x_n$  is divergent.

The test is *inconclusive* when  $\ell = 1$ .

*Proof.* Idea:  $\sum_{n \geq 1} x_n$  behaves like a geometric series with ratio  $\ell$ , as in the ratio test.  $\square$

**Example 3.12.** The series  $\sum_{n \geq 0} \frac{x^n}{n!}$  converges for any  $x \in \mathbb{R}$ . We will see later that  $\sum_{n \geq 0} \frac{x^n}{n!} = e^x$ . We have that  $\frac{x_{n+1}}{x_n} = \frac{x}{n+1} \rightarrow 0 < 1$ , hence the series converges by the ratio test.

**Theorem 3.13** (Cauchy condensation test). Let  $(x_n)$  be a decreasing sequence with  $x_n > 0$ . Then the series  $\sum_{n=1}^{\infty} x_n$  and  $\sum_{n=0}^{\infty} 2^n x_{2^n}$  have the same nature.

*Proof.* Let  $S_n = x_1 + x_2 + \dots + x_n$  and  $T_n = x_1 + 2x_2 + \dots + 2^n x_n$ . Since  $x_n > 0$ , the two series will have the same nature if and only if  $S_n$  and  $T_n$  are both bounded/unbounded.

For any  $n \in \mathbb{N}$  there exists  $k \in \mathbb{N}$  s.t.  $2^k \leq n \leq 2^{k+1} - 1$ . Since  $(x_n)$  is decreasing and positive, we can group the terms in the following ways

$$\begin{aligned} S_n &= x_1 + x_2 + \dots + x_n \leq x_1 + x_2 + \dots + x_{2^{k+1}-1} \\ &\leq x_1 + (x_2 + x_3) + \dots + (x_{2^k} + \dots + x_{2^{k+1}-1}) \\ &\leq T_k, \end{aligned}$$

and

$$\begin{aligned} S_n &= x_1 + x_2 + \dots + x_n \geq x_1 + x_2 + \dots + x_{2^k} \\ &\geq x_1 + x_2 + (x_3 + x_4) + \dots + (x_{2^{k-1}+1} + \dots + x_{2^k}) \\ &\geq \frac{x_1}{2} + \frac{1}{2}T_k. \end{aligned}$$

We obtained that  $0 \leq \frac{1}{2}T_k \leq S_n \leq T_k$ , hence  $(S_n)$  bounded if and only if  $(T_n)$  is bounded.  $\square$

**Example 3.14.** The series  $\sum_{n=1}^{\infty} \frac{1}{n^p}$  converges if and only if  $p > 1$ .

*Proof.* By the Cauchy condensation test,  $\sum_{n=1}^{\infty} \frac{1}{n^p}$  has the same nature as  $\sum_{n=1}^{\infty} \frac{2^n}{2^{np}} = \sum_{n=1}^{\infty} (2^{1-p})^n$ , which converges if and only if  $2^{1-p} < 1$ , i.e for  $p > 1$ .  $\square$

**Theorem 3.15** (Kummer's test). Let  $(x_n)$  be a positive sequence and consider another positive sequence  $(c_n)$ .

(a) If

$$\lim_{n \rightarrow \infty} \left( c_n \frac{x_n}{x_{n+1}} - c_{n+1} \right) > 0,$$

then  $\sum_{n=1}^{\infty} x_n$  is convergent.

(b) If  $\sum_{n=1}^{\infty} \frac{1}{c_n} = \infty$  and

$$\lim_{n \rightarrow \infty} \left( c_n \frac{x_n}{x_{n+1}} - c_{n+1} \right) < 0,$$

then  $\sum_{n=1}^{\infty} x_n$  is divergent.

*Proof.* Let us start with (a). Since that limit is positive, there exist  $r > 0$  and  $n_0 \in \mathbb{N}$  such that

$$c_n x_n - c_{n+1} x_{n+1} \geq r x_{n+1}, \quad \forall n \geq n_0.$$

Denote by  $s_n = x_1 + \dots + x_n$ . Adding all these inequalities for  $k \in \{n_0, \dots, n\}$  we have that

$$c_{n_0} x_{n_0} - c_{n+1} x_{n+1} \geq r(s_{n+1} - s_{n_0}),$$

which gives  $s_{n+1} \leq s_{n_0} + \frac{1}{r} c_{n_0} x_{n_0}$ . Hence  $(s_n)$  is bounded and the series converges.

Let us now consider (b). Since the limit is negative, there exists  $n_0 \in \mathbb{N}$  such that

$$c_n x_n < c_{n+1} x_{n+1}, \quad \forall n \geq n_0.$$

Hence for  $n > n_0$ , we have that  $c_{n_0} x_{n_0} < c_n x_n$ , which gives

$$\frac{1}{c_n} < \frac{1}{c_{n_0} x_{n_0}} x_n.$$

Since  $\sum_{n \geq 1} \frac{1}{c_n} = \infty$ , we conclude that  $\sum_{n \geq 1} x_n = \infty$ .  $\square$

Many convergence tests can be obtained by taking particular sequences in Kummer's test. We will restrict to the following one.

**Theorem 3.16** (Raabe-Duhamel). Let  $\sum_{n \geq 1} x_n$  be a series with positive terms such that

$$\lim_{n \rightarrow \infty} n \left( \frac{x_n}{x_{n+1}} - 1 \right) = R.$$

- If  $R > 1$ , then the series  $\sum_{n=1}^{\infty} x_n$  is convergent.
- If  $R < 1$ , then the series  $\sum_{n=1}^{\infty} x_n$  is divergent.

*Proof.* Take  $c_n = n$  in Kummer's test ([theorem 3.15](#)).  $\square$

**Example 3.17.** Study the convergence of the series  $\sum_{n \geq 0} \frac{n!}{a(a+1)\dots(a+n)}$ , with  $a > 0$ .

*Proof.* The ratio test is inconclusive since  $\frac{x_{n+1}}{x_n} = \frac{n+1}{a+n+1} \rightarrow 1$ . Let us then try the Raabe-Duhamel test:

$$\lim_{n \rightarrow \infty} n \left( \frac{x_n}{x_{n+1}} - 1 \right) = \lim_{n \rightarrow \infty} n \left( \frac{a+n+1}{n+1} - 1 \right) = a.$$

Hence if  $a > 1$  the series converges; and if  $a < 1$  the series diverges. When  $a = 1$  the series is  $\sum_{n \geq 0} \frac{1}{n+1} = \infty$ .  $\square$

A series  $\sum_{n \geq 1} x_n$  is called an *alternating series* if  $x_n x_{n+1} \leq 0$ ,  $\forall n \in \mathbb{N}$ . A fundamental class of alternating series are series of the form  $\sum_{n \geq 1} (-1)^n a_n$  or  $\sum_{n \geq 1} (-1)^{n+1} a_n$ , with  $a_n > 0$ .

**Example 3.18.** The series  $\sum_{n \geq 1} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$  converges to  $\ln 2$ .

*Proof.* Let us prove convergence by considering the partial sums  $s_n = \sum_{k=1}^n \frac{(-1)^{k+1}}{k}$ . Notice that  $s_{2k+2} - s_{2k} = \frac{1}{2k+1} - \frac{1}{2k+2} > 0$  and that  $s_{2k+3} - s_{2k+1} = \frac{1}{2k+3} - \frac{1}{2k+2} < 0$ . This means that the subsequence  $(s_{2k})$  is increasing, while the subsequence  $(s_{2k+1})$  is decreasing. Notice also that  $s_{2k+1} - s_{2k} = \frac{1}{2k+1}$  and  $s_{2k} < s_{2k+1}$ , so both subsequences are also bounded and converge to the same limit. To find the sum of the alternating series, recall (from the seminar) that

$$\begin{aligned} \lim_{n \rightarrow \infty} 1 + \frac{1}{2} + \dots + \frac{1}{n} - \ln n &= \gamma \in (0, 1), \text{ hence} \\ s_{2n} &= 1 - \frac{1}{2} + \frac{1}{3} - \dots - \frac{1}{2n} = 1 + \frac{1}{2} + \dots + \frac{1}{2n} - 2\left(\frac{1}{2} + \dots + \frac{1}{2n}\right) \\ &= \underbrace{1 + \frac{1}{2} + \dots + \frac{1}{2n} - \ln(2n)}_{\rightarrow \gamma} - \underbrace{\left(1 + \frac{1}{2} + \dots + \frac{1}{n} - \ln n\right)}_{\rightarrow \gamma} + \ln 2 \rightarrow \ln 2. \end{aligned}$$

□

**Theorem 3.19** (Leibniz test). Let  $(x_n)$  be a decreasing sequence with  $x_n \rightarrow 0$ . Then the series  $\sum_{n \geq 1} (-1)^n x_n$  is convergent.

*Proof.* Consider the partial sum  $s_n = \sum_{k=1}^n (-1)^k x_k$ . We will prove that  $(s_n)$  is convergent by showing that it is a Cauchy sequence. For  $n, p \in \mathbb{N}$  consider

$$\begin{aligned} |s_{n+p} - s_n| &= |(-1)^{n+1} x_{n+1} + \dots + (-1)^{n+p} x_{n+p}| \\ &= \underbrace{|x_{n+1} - x_{n+2}|}_{\geq 0} + \underbrace{|x_{n+3} - x_{n+4}|}_{\geq 0} + \dots + (-1)^{p-2} x_{n+p-1} + (-1)^{p-1} x_{n+p} \\ &= x_{n+1} - \underbrace{x_{n+2} + x_{n+3} - x_{n+4}}_{\leq 0} + \dots \pm x_{n+p-1} \mp x_{n+p} \\ &\leq x_{n+1}. \end{aligned}$$

Since  $x_n \rightarrow 0$ ,  $|s_{n+p} - s_n|$  can be made arbitrarily small, so  $(s_n)$  is Cauchy. □

**Definition 3.20.** A series  $\sum_{n \geq 1} x_n$  is called *absolutely convergent* if  $\sum_{n \geq 1} |x_n|$  is convergent.

**Proposition 3.21.** Any absolutely convergent series is also convergent.

*Proof.* If  $\sum_{k=1}^n |x_k|$  gives a Cauchy sequence, then  $\sum_{k=1}^n x_k$  also gives a Cauchy sequence.  $\square$

**Theorem 3.22 (Cauchy).** Let  $\sum_{n \geq 1} x_n$  be an *absolutely convergent series* and let  $\sigma : \mathbb{N} \rightarrow \mathbb{N}$  be a bijection. Then  $\sum_{n \geq 1} x_{\sigma(n)}$  is also absolutely convergent and  $\sum_{n \geq 1} x_{\sigma(n)} = \sum_{n \geq 1} x_n$ . In other words, any rearrangement of an absolutely convergent series has the same sum.

*Proof.* (Optional) See [2][Theorem 7.4.3].  $\square$

**Definition 3.23.** A series  $\sum_{n \geq 1} x_n$  is called *conditionally convergent* (or semi-convergent) if  $\sum_{n \geq 1} x_n$  converges, but  $\sum_{n \geq 1} |x_n|$  diverges.

**Theorem 3.24 (Riemann).** Let  $\sum_{n \geq 1} x_n$  be a *conditionally convergent series* and let  $x \in \overline{\mathbb{R}}$ . Then there exists a bijection  $\sigma : \mathbb{N} \rightarrow \mathbb{N}$  such that  $\sum_{n \geq 1} x_{\sigma(n)} = x$ . In other words, a conditionally convergent series can be rearranged to converge to any value or diverge to  $\pm\infty$ .

*Proof.* (Optional) See [2][Theorem 8.2.8].  $\square$

**Example 3.25.** Rearranging the terms in the alternating harmonic series one can obtain a different sum. Indeed, consider  $\sum_{n \geq 1} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots = \ln 2$ , and reorder the terms in the following way: one positive, two negative. Then

$$1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + \dots = \frac{1}{2} - \frac{1}{4} + \frac{1}{6} - \frac{1}{8} + \dots = \frac{1}{2} \left( 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots \right) = \frac{1}{2} \ln 2.$$

**Definition 3.26.** Let  $(a_n)$  be a sequence of real numbers and let  $c \in \mathbb{R}$ . The series

$$\sum_{n=0}^{\infty} a_n(x - c)^n$$

is called a *power series* centered at  $c$ .

**Theorem 3.27.** Consider the power series  $\sum_{n=0}^{\infty} a_n(x - c)^n$ . There exists a unique  $R \in [0, \infty]$ , called the *radius of convergence* of the power series, such that the power series

- converges absolutely when  $|x - c| < R$ .
- diverges when  $|x - c| > R$ .

**Theorem 3.28.** If the limit

$$\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} = L \in [0, \infty]$$

exists, then the power series  $\sum_{n=0}^{\infty} a_n(x - c)^n$  has the radius of convergence

$$R = \begin{cases} \frac{1}{L}, & \text{if } L \in (0, \infty) \\ 0, & \text{if } L = \infty \\ \infty, & \text{if } L = 0. \end{cases}$$

*Proof.* It follows from the root test for series with positive terms. □

**Corollary 3.29.** If the limit

$$\lim_{n \rightarrow \infty} \frac{|a_{n+1}|}{|a_n|} = L \in [0, \infty]$$

exists, then the power series  $\sum_{n=0}^{\infty} a_n(x - c)^n$  has the radius of convergence

$$R = \begin{cases} \frac{1}{L}, & \text{if } L \in (0, \infty) \\ 0, & \text{if } L = \infty \\ \infty, & \text{if } L = 0. \end{cases}$$

*Proof.* It follows from  $\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|} = \lim_{n \rightarrow \infty} \frac{|a_{n+1}|}{|a_n|} = L$ . □

**Definition 3.30.** The convergence set of a power series is

$$C := \{x \in \mathbb{R} \mid \sum_{n=0}^{\infty} a_n(x - c)^n \text{ converges}\}.$$

**Remark 3.31.** The convergence set  $C$  contains the open interval  $(c - R, c + R)$  and possibly the endpoints  $\{c - R, c + R\}$ .

**Example 3.32.** The power series  $\sum_{n \geq 0} x^n$  has radius of convergence  $R = 1$ , it converges absolutely for  $|x| < 1$  and diverges when  $|x| > 1$  (by the root test or the ratio test). The convergence set is  $(-1, 1)$  and for  $x \in (-1, 1)$  we have that

$$\sum_{n \geq 0} x^n = \frac{1}{1 - x}, \quad \sum_{n \geq 0} (-x)^n = \frac{1}{1 + x}.$$

**Example 3.33.** The power series  $\sum_{n \geq 1} \frac{x^n}{n}$  has radius of convergence  $R = 1$ , it converges absolutely for  $|x| < 1$  and diverges when  $|x| > 1$  (by the root test or the ratio test). Moreover, the series converges for  $x = -1$  (alternating harmonic series) and diverges for  $x = 1$  (harmonic series), hence its convergence set is  $C = [-1, 1)$ .

**Theorem 3.34.** Consider a power series with radius of convergence  $R$ , given by

$$s(x) = \sum_{n=0}^{\infty} a_n (x - c)^n.$$

Then for any  $x \in (c - R, c + R)$ , the power series can be differentiated term by term and

$$s'(x) = \sum_{n=1}^{\infty} n a_n (x - c)^{n-1},$$

and for any  $t \in (c - R, c + R)$  the power series can be integrated term by term

$$\int_c^t s(x) dx = \sum_{n=0}^{\infty} \frac{a_n}{n+1} (t - c)^{n+1}.$$

**Example 3.35.** The power series  $\sum_{n \geq 1} \frac{x^n}{n!}$  converges absolutely for any  $x \in \mathbb{R}$  (ratio test). Let

$\exp(x) := \sum_{n \geq 1} \frac{x^n}{n!}$  and differentiate term by term, then  $\exp'(x) = \exp(x)$  and  $\exp(0) = 1$ .



## ❖ Limits, continuity, differentiability

**Definition 4.1.** Let  $A \subseteq \mathbb{R}$ . We say that  $x_0 \in \overline{\mathbb{R}}$  is an *accumulation point* (or *cluster point*) if

$$\forall V \in \mathcal{V}(x_0), V \cap (A \setminus \{x_0\}) \neq \emptyset.$$

We denote by  $A'$  the set of the accumulation points of  $A$ . We say that  $x_0 \in A$  is an *isolated point* if  $x_0 \in A \setminus A'$ .

**Remark 4.2.**  $\text{cl}(A) = A' \cup \{\text{isolated points}\}$ .

**Proposition 4.3.** Let  $A \subseteq \mathbb{R}$  and  $x_0 \in \overline{\mathbb{R}}$ , then  $x_0 \in A'$  if and only if there exists a sequence  $(x_n)$  in  $A \setminus \{x_0\}$  such that  $\lim_{n \rightarrow \infty} x_n = x_0$ .

*Proof.* Assume that  $x_0 \in A'$ , with  $x_0 \in \mathbb{R}$ , and consider the neighborhoods  $(x_0 - \frac{1}{n}, x_0 + \frac{1}{n})$ . Then each neighborhood must contain an  $x_n \in A \setminus \{x_0\}$  with  $|x_n - x_0| < \frac{1}{n}$ , hence  $x_n \rightarrow x_0$ . If  $x_0$  is infinite, the neighborhoods can be taken  $(-\infty, -n)$  or  $(n, \infty)$ , respectively.

Assume now that there exists a sequence  $(x_n)$  in  $A \setminus \{x_0\}$  such that  $\lim_{n \rightarrow \infty} x_n = x_0$ . Then for any  $V \in \mathcal{V}(x_0)$ , there exists  $N_V \in \mathbb{N}$  such that  $x_n \in V$ , for any  $n \geq N_V$ . In particular,  $x_{N_V} \in V \cap (A \setminus \{x_0\})$ , for any  $V \in \mathcal{V}(x_0)$ , hence  $x_0 \in A'$ .  $\square$

**Example 4.4.** For  $A = \{\frac{1}{n} \mid n \in \mathbb{N}\}$ , each element  $x \in A$  is an isolated point and  $A' = \{0\}$ .

**Definition 4.5.** Let  $A \subseteq \mathbb{R}$ ,  $f : A \rightarrow \mathbb{R}$  and  $x_0 \in A'$ . We say that  $\lim_{x \rightarrow x_0} f(x) = \ell \in \overline{\mathbb{R}}$  if

$$\forall V \in \mathcal{V}(\ell), \exists U \in \mathcal{V}(x_0) \text{ s.t. } f(x) \in V, \forall x \in U \cap (A \setminus \{x_0\}).$$

**Remark 4.6** ( $\varepsilon$ - $\delta$ ). Let  $A \subseteq \mathbb{R}$ ,  $f : A \rightarrow \mathbb{R}$  and  $x_0 \in A'$  finite. If  $\lim_{x \rightarrow x_0} f(x) = \ell \in \mathbb{R}$ , then

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } |f(x) - \ell| < \varepsilon, \forall x \in A \text{ with } |x - x_0| < \delta.$$

**Theorem 4.7.** Let  $A \subseteq \mathbb{R}$ ,  $f : A \rightarrow \mathbb{R}$  and  $x_0 \in A'$ . Then  $\lim_{x \rightarrow x_0} f(x) = \ell \in \overline{\mathbb{R}}$  iff

for any sequence  $(x_n)$  in  $A \setminus \{x_0\}$  with  $\lim_{n \rightarrow \infty} x_n = x_0$ , we have that  $\lim_{n \rightarrow \infty} f(x_n) = \ell$ .

**Theorem 4.8.** Let  $A \subseteq \mathbb{R}$ ,  $f : A \rightarrow \mathbb{R}$ ,  $x_0 \in \mathbb{R}$  s.t.  $x_0 \in (A \cap (-\infty, x_0))'$  and  $x_0 \in (A \cap (x_0, \infty))'$ . Then

$$\lim_{x \rightarrow x_0} f(x) = \ell \text{ iff } \lim_{\substack{x \rightarrow x_0 \\ x < x_0}} f(x) = \lim_{\substack{x \rightarrow x_0 \\ x > x_0}} f(x) = \ell.$$

**Example 4.9.** (a)  $\text{sgn} : \mathbb{R} \rightarrow \mathbb{R}$ ,  $\text{sgn}(x) = \begin{cases} +1, & \text{if } x \geq 0 \\ -1, & \text{if } x < 0. \end{cases}$  has no limit at 0.

(b)  $f : \mathbb{R}^* \rightarrow \mathbb{R}$ ,  $f(x) = \sin(\frac{1}{x})$  has no limit at 0 since  $f(\frac{1}{2n\pi}) = 0$ ,  $f(\frac{1}{2n\pi + \pi/2}) = 1$ .

(c)  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = \begin{cases} 1, & \text{if } x \in \mathbb{Q} \\ 0, & \text{if } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$  has no limit at any  $x \in \mathbb{R}$ .

**Definition 4.10.** Let  $A \subseteq \mathbb{R}$ ,  $f : A \rightarrow \mathbb{R}$  and  $x_0 \in A$ . We say that  $f$  is *continuous* at  $x_0$  if

$$\forall V \in \mathcal{V}(f(x_0)), \exists U \in \mathcal{V}(x_0) \text{ s.t. } f(x) \in V, \forall x \in U \cap A.$$

**Remark 4.11.** If  $x_0 \in A \cap A'$  is an accumulation point, then  $f$  is continuous at  $x_0$  if

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

**Remark 4.12.** If  $x_0$  is an isolated point of  $A$ , then  $\exists U \in \mathcal{V}(x_0)$  with  $U \cap A = \{x_0\}$ , and since  $f(x_0) \in V$ ,  $\forall V \in \mathcal{V}(f(x_0))$ , we have that  $f$  is continuous at  $x_0$ .

**Theorem 4.13.** Let  $A \subseteq \mathbb{R}$ ,  $f : A \rightarrow \mathbb{R}$  and  $x_0 \in A \cap A'$ . The following are equivalent:

- (a)  $f$  is continuous at  $x_0$ .
- (b)  $\forall \varepsilon > 0, \exists \delta > 0$  s.t.  $|f(x) - f(x_0)| < \varepsilon$ ,  $\forall x \in A$  with  $|x - x_0| < \delta$ .
- (c) for any sequence  $(x_n)$  in  $A$  with  $\lim_{n \rightarrow \infty} x_n = x_0$ , we have that  $\lim_{n \rightarrow \infty} f(x_n) = f(x_0)$ .

**Remark 4.14.** Elementary operations – e.g. sums, products or compositions – of continuous functions are continuous (when defined).

**Definition 4.15.** For  $f : A \rightarrow \mathbb{R}$  denote by  $f(A) := \{f(x) \mid x \in A\}$  the image of  $A$ . We say that  $f$  is *bounded* if  $f(A)$  is *bounded*, i.e.  $\inf(f(A)), \sup(f(A))$  are finite.

**Theorem 4.16** (Weierstrass). Let  $f : [a, b] \rightarrow \mathbb{R}$  be a continuous function. Then  $f$  is bounded and it attains its bounds, i.e. there exist  $\min(f(A)), \max(f(A))$ .

*Proof.* Let us first prove that  $f$  is bounded. Assuming that this is not the case, we have that for any  $n \in \mathbb{N}$  there exists  $x_n \in [a, b]$  such that  $|f(x_n)| > n$ . Since the sequence  $(x_n)$  is bounded, we have that it has a convergent subsequence  $(x_{n_k})$ , see [theorem 2.8](#); denote its limit by  $x$ . We have that  $x_{n_k} \rightarrow x$  and  $f$  is continuous, hence  $f(x_{n_k}) \rightarrow f(x)$ . But  $|f(x_{n_k})| > n_k \rightarrow \infty$ , contradiction. Hence  $f$  is bounded on  $[a, b]$ .

To prove that  $f$  attains its bounds, let's consider the upper bound and show that there exists  $x_M \in [a, b]$  such that  $f(x_M) = \sup(f(A))$ , i.e.  $f(x_M) = \max(f(A)) = \sup(f(A))$ . From [theorem 1.7](#), we obtain a sequence  $(x_n)$  in  $[a, b]$  such that  $f(x_n) \rightarrow \sup(f(A))$ . Since the sequence  $(x_n)$  is bounded, it has a convergent subsequence  $(x_{n_k})$ ; let's call its limit  $x_M \in [a, b]$ . Since  $f$  is continuous, it follows that  $f(x_{n_k}) \rightarrow f(x_M)$ , but we know that  $f(x_{n_k}) \rightarrow \sup(f(A))$ , hence  $f(x_M) = \sup(f(A))$  and  $f$  reaches its upper bound.  $\square$

**Theorem 4.17** (Intermediate value property). Let  $f : [a, b] \rightarrow \mathbb{R}$  be a continuous function. Then  $f$  has the intermediate value property, i.e. if  $y \in \mathbb{R}$  is in between  $f(a)$  and  $f(b)$ , there exists  $c \in (a, b)$  such that  $f(c) = y$ .

*Proof.* Assume that  $f(a) < y < f(b)$  and consider the set  $S := \{x \in [a, b] \mid f(x) \leq y\}$ . Take

$$c := \sup(S)$$

Let  $\varepsilon > 0$ , then  $\exists \delta > 0$  such that  $|f(x) - f(c)| < \varepsilon$ , whenever  $|x - c| < \delta$ . Since  $c = \sup(S)$ , we have from [theorem 1.7](#) that there exists  $x_1 \in S$  such that  $c - \delta < x_1 \leq c$ . From continuity we have that  $f(c) < f(x_1) + \varepsilon \leq y + \varepsilon$ . Also, for  $x_2 \in (c, c + \delta)$ , we have from continuity that  $f(c) > f(x_2) - \varepsilon$ . From the definition of the supremum,  $x_2 \notin S$  hence  $f(x_2) > y$  and  $f(c) > y - \varepsilon$ . We conclude that  $y - \varepsilon < f(c) < y + \varepsilon$ , for any  $\varepsilon > 0$ . Hence  $f(c) = y$ .  $\square$

**Definition 4.18.** Let  $A \subseteq \mathbb{R}$ ,  $f : A \rightarrow \mathbb{R}$  and  $x_0 \in A \cap A'$ . The *derivative* of  $f$  at  $x_0$  is

$$f'(x_0) := \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \in \overline{\mathbb{R}}$$

If  $f'(x_0) \in \mathbb{R}$  (finite) we say that  $f$  is *differentiable* at  $x_0$ .

**Remark 4.19.**  $f'(x_0)$  represents the gradient of the tangent to the curve  $y = f(x)$  at the point  $(x_0, f(x_0))$ . The equation of the tangent is  $f(x) - f(x_0) = f'(x_0)(x - x_0)$ .

**Theorem 4.20.** Let  $A \subseteq \mathbb{R}$ ,  $f : A \rightarrow \mathbb{R}$  and  $x_0 \in A \cap A'$ . If  $f$  is differentiable at  $x_0$ , then  $f$  is continuous at  $x_0$ .

*Proof.* Since  $f(x) = f(x_0) + \frac{f(x) - f(x_0)}{x - x_0}(x - x_0)$ , we have that  $\lim_{x \rightarrow x_0} f(x) = f(x_0) + 0 = f(x_0)$ .  $\square$

**Example 4.21.**  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = |x|$  is not differentiable in 0 since  $\nexists \lim_{x \rightarrow 0} \frac{|x|}{x}$ .

**Theorem 4.22** (Calculus Rules).

- $(cf)'(x) = cf'(x)$ , for any constant  $c \in \mathbb{R}$ .
- $(f + g)'(x) = f'(x) + g'(x)$ .
- $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$ . (Product Rule)
- $\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$ . (Quotient Rule)
- $(f \circ g)'(x) = f'(g(x))g'(x)$ . (Chain Rule)

**Proposition 4.23** (l'Hôpital's rule). Let  $I$  be an open interval,  $x_0 \in \overline{\mathbb{R}}$  and  $f, g : I \setminus \{x_0\} \rightarrow \mathbb{R}$  differentiable. If  $\lim_{x \rightarrow x_0} f(x) = \lim_{x \rightarrow x_0} g(x) = 0$  or  $\pm\infty$ , and  $\lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)} \in \overline{\mathbb{R}}$ , then

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

**Definition 4.24.**  $f : A \rightarrow \mathbb{R}$  has a local extremum (minimum or maximum) at  $x_0 \in A$  if

$$\exists V \in \mathcal{V}(x_0) \text{ s.t. } f(x_0) \leq f(x) \text{ or } f(x_0) \geq f(x), \forall x \in V \cap A.$$

**Theorem 4.25** (Fermat). Let  $f : (a, b) \rightarrow \mathbb{R}$  and  $x_0 \in (a, b)$ . If  $f$  is differentiable at  $x_0$  and  $x_0$  is a local extremum, then  $f'(x_0) = 0$ .

*Proof.* The lateral derivatives at  $x_0$  are equal. Since  $x_0$  is a local extremum, one of them is  $\geq 0$ , the other  $\leq 0$ . Hence  $f'(x_0) = 0$ .  $\square$

**Theorem 4.26** (Rolle). Let  $f : (a, b) \rightarrow \mathbb{R}$  with  $f(a) = f(b)$ . If  $f$  is continuous on  $[a, b]$  and differentiable on  $(a, b)$ , then there exists  $c \in (a, b)$  s.t.  $f'(c) = 0$ .

*Proof.* Since  $f$  is continuous, it is bounded and it attains its bounds. Denote by  $x_m$  and  $x_M$  the minimum and maximum points of  $f$  on  $[a, b]$ . If at least one of  $x_m$  and  $x_M$  belongs to  $(a, b)$ , then  $f'(x_m) = 0$  or  $f'(x_M) = 0$ . Otherwise,  $x_m, x_M \in \{a, b\}$  and  $f(x_m) = f(x_M)$ , hence the function is constant and its derivative is zero on  $(a, b)$ .  $\square$

**Theorem 4.27** (Mean value theorem). Let  $f : (a, b) \rightarrow \mathbb{R}$  be continuous on  $[a, b]$  and differentiable on  $(a, b)$ . Then there exists  $c \in (a, b)$  s.t.

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

*Proof.* Consider the function  $g : (a, b) \rightarrow \mathbb{R}$ ,  $g(x) := f(x) - x \frac{f(b) - f(a)}{b - a}$ . Since  $g(a) = g(b)$ , the conclusion follows from Rolle's theorem.  $\square$

**Theorem 4.28** (Monotony). Let  $f : (a, b) \rightarrow \mathbb{R}$  be differentiable on  $(a, b)$ . Then

$$f \text{ is increasing iff } f' \geq 0,$$

$$f \text{ is decreasing iff } f' \leq 0.$$

*Proof.*  $\Rightarrow$  follows from the definition of the derivative;  $\Leftarrow$  from the mean value theorem.  $\square$

## ❖ Taylor series

Let  $I \subseteq \mathbb{R}$  be an interval,  $f : I \rightarrow \mathbb{R}$  and  $x_0 \in I$  a point where  $f$  is differentiable  $n$  times ( $n \in \mathbb{N}$ ). Does there exist a polynomial  $P : \mathbb{R} \rightarrow \mathbb{R}$  that matches the function  $f$  and all its derivatives up to order  $n$  at the point  $x_0$ ? That is

$$\begin{aligned} P(x_0) &= f(x_0) \\ P'(x_0) &= f'(x_0) \\ P''(x_0) &= f''(x_0) \\ &\vdots \\ P^{(n)}(x_0) &= f^{(n)}(x_0). \end{aligned}$$

Let us look for  $P$  of degree at most  $n$  of the form

$$P(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \dots + a_n(x - x_0)^n.$$

By imposing the conditions at  $x_0$  and differentiating  $P$  we have that

$$P(x_0) = a_0 = f(x_0), P'(x_0) = a_1 = f'(x_0), P''(x_0) = 2a_2 = f''(x_0), \dots, P^{(n)}(x_0) = n!a_n = f^{(n)}(x_0).$$

We thus see that there exists a unique such polynomial  $P$  of degree at most  $n$  given by

$$P(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \dots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n,$$

that matches the function  $f$  and all its derivatives up to order  $n$  at the point  $x_0$ .

**Definition 5.1.** Let  $I \subseteq \mathbb{R}$  be an interval,  $f : I \rightarrow \mathbb{R}$  and  $x_0 \in I$  where  $f$  is differentiable  $n$  times. The polynomial  $T_n : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$T_n(x) := f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \dots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$$

is called the *Taylor polynomial* of degree  $n$  centered around  $x_0$ . □

The Taylor polynomial  $T_n$  gives a good approximation of  $f$  around  $x_0$ , i.e. when  $x \approx x_0$ ,

$$f(x) \approx T_n(x).$$

The simplest approximations are: the *linear approximation* of  $f$  around  $x_0$  given by  $T_1$ , i.e.

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$$

and the *quadratic approximation* of  $f$  around  $x_0$  given by  $T_2$ , i.e.

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2.$$

The closer  $x$  is to  $x_0$  and the higher the degree of  $T_n$  is, the better  $T_n(x)$  approximates  $f(x)$ .

**Example 5.2.** Let  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = e^x$  and  $x_0 = 0$ . Then  $f(0) = f'(0) = \dots = f^{(n)}(0) = 1$  and

$$T_n(x) = 1 + x + \frac{x^2}{2} + \dots + \frac{x^n}{n!}.$$

**Definition 5.3.** Let  $I \subseteq \mathbb{R}$  be an interval,  $f : I \rightarrow \mathbb{R}$  and  $x_0 \in I$  where  $f$  is differentiable  $n$  times. We define  $R_n : \mathbb{R} \rightarrow \mathbb{R}$  to be the remainder when approximating  $f$  by  $T_n$ ,

$$R_n(x) := f(x) - T_n(x).$$

**Theorem 5.4** (Taylor-Lagrange). Let  $I \subseteq \mathbb{R}$  be an interval,  $f : I \rightarrow \mathbb{R}$  differentiable  $n + 1$  times. Then for any  $x, x_0 \in I$ , there exists  $c \in (x_0, x)$  or  $c \in (x, x_0)$  such that

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

called *the remainder in Lagrange's form*. Taylor's formula with Lagrange remainder is

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

**Remark 5.5.** There exist other forms of the remainder, but we will only use this one. Its main advantage is that assuming that all the derivatives of  $f$  are bounded by  $M > 0$ ,

$$|f(x) - T_n(x)| = |R_n(x)| \leq \frac{M}{(n+1)!} |x - x_0|^{n+1} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

**Definition 5.6.** Let  $I \subseteq \mathbb{R}$  be an interval and let  $f : I \rightarrow \mathbb{R}$  be infinitely differentiable. For  $x_0 \in I$  and  $x \in \mathbb{R}$ , the series

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n$$

is called the *Taylor series* of  $f$  around  $x_0$ . If the series converges and

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n,$$

we say that  $f$  can be expanded in a Taylor series around  $x_0$  (also called Taylor expansion).

**Remark 5.7.** The partial sum of a Taylor series is the Taylor polynomial  $T_n(x)$ . A Taylor series converges to  $f(x)$  if and only if the remainder  $f(x) - T_n(x) = R_n(x) \rightarrow 0$  as  $n \rightarrow \infty$ .

**Remark 5.8.** The Taylor series around 0 is called the *MacLaurin series*,

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n.$$

**Example 5.9.** Let  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = e^x$  and  $x_0 = 0$ . Then

$$T_n(x) = 1 + x + \frac{x^2}{2} + \dots + \frac{x^n}{n!}.$$

Consider Taylor's formula  $f(x) = T_n(x) + R_n(x)$  with the Lagrange remainder, for which there exists  $c$  in between 0 and  $x$  such that

$$R_n(x) = \frac{e^c}{(n+1)!} x^{n+1} \rightarrow 0$$

since  $\frac{|x|^n}{n!} \rightarrow 0$  as  $n \rightarrow \infty$ . It follows that  $e^x$  can be expanded as a Taylor series around 0:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \dots + \frac{x^n}{n!} + \dots, \forall x \in \mathbb{R}.$$

**Example 5.10.** The functions  $\sin$  and  $\cos$  can be expanded in a Taylor series around 0.

$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots$$

and

$$\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \dots$$

**Example 5.11.** The function  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$f(x) = \begin{cases} e^{-1/x^2}, & \text{if } x \neq 0 \\ 0, & \text{if } x = 0 \end{cases}$$

is infinitely differentiable at 0, but  $f$  is not expandable in a Taylor series around 0.

**Example 5.12** (Convex/concave). Let  $f : I \rightarrow \mathbb{R}$  be two times differentiable, with a critical point at  $x_0$ , i.e.  $f'(x_0) = 0$ . Then from Taylor's formula we have that

$$f(x) = f(x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + R_2(x).$$

When  $x$  is very close to  $x_0$ , the quadratic approximation is very accurate and the remainder  $R_2(x)$  is very small. Thus the behaviour of  $f(x)$  around  $x_0$  is dictated by the quadratic term  $f''(x_0)(x - x_0)^2$  and we see that:

- If  $f''(x_0) > 0$  (convexity), then  $f(x) > f(x_0)$  and  $x_0$  is a local minimum.
- If  $f''(x_0) < 0$  (concavity), then  $f(x) < f(x_0)$  and  $x_0$  is a local maximum.

**Theorem 5.13** (Local optimality conditions). Let  $I \subseteq \mathbb{R}$  be an open interval,  $f : I \rightarrow \mathbb{R}$  and  $x_0 \in I$  a point where  $f$  is differentiable  $n$  times and

$$f'(x_0) = f''(x_0) = \dots = f^{(n-1)}(x_0) = 0 \text{ and } f^{(n)}(x_0) \neq 0.$$

1. If  $n$  is even and  $f^{(n)}(x_0) > 0$ , then  $x_0$  is a *local minimum* of  $f$ .
2. If  $n$  is even and  $f^{(n)}(x_0) < 0$ , then  $x_0$  is a *local maximum* of  $f$ .
3. If  $n$  is odd, then  $x_0$  is not a local extremum point of  $f$ .

*Proof.* It follows from the Taylor approximation  $f(x) = f(x_0) + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + R_n(x)$ .  $\square$



## ❖ Riemann integrals. Improper integrals

Let  $[a, b]$  be a compact interval and let  $f : [a, b] \rightarrow \mathbb{R}$ . The points  $a = x_0 < x_1 < \dots < x_n = b$  define a partition of the interval  $[a, b]$

$$\mathcal{P} = \{[x_{k-1}, x_k] \mid k = \overline{1, n}\},$$

whose norm is given by

$$\|\mathcal{P}\| = \max_{k=\overline{1, n}} \{x_k - x_{k-1}\}.$$

Consider also a set of intermediate points  $c_k \in [x_k, x_{k-1}]$  attached to the partition  $\mathcal{P}$ .

**Definition 6.1.** For  $f : [a, b] \rightarrow \mathbb{R}$  and a partition  $\mathcal{P}$  of  $[a, b]$ , the Riemann sum is given by

$$\sigma(f, \mathcal{P}) := \sum_{k=1}^n f(c_k)(x_k - x_{k-1}).$$

**Remark 6.2.** The Riemann sum collects the areas of the rectangles defined by the partition  $\mathcal{P}$  (and the intermediate points). In the limit one obtains the area below the curve  $y = f(x)$ .

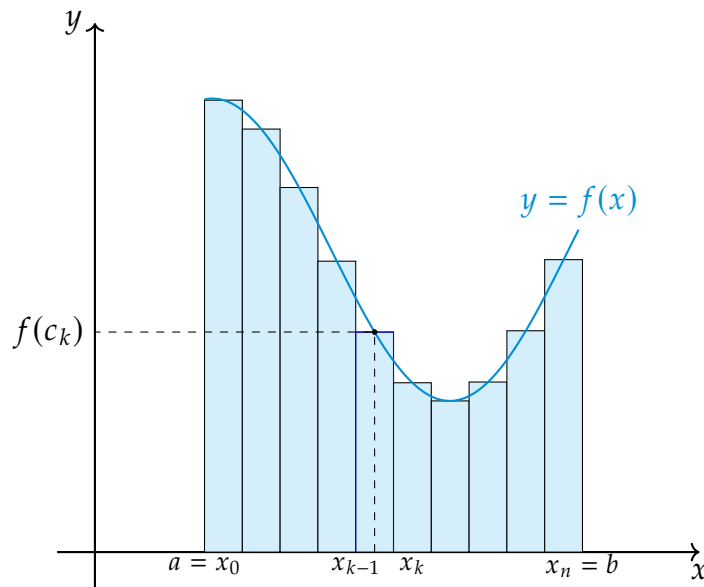


Figure 1: Area under a curve approximated through rectangles. Riemann sum.

**Definition 6.3.** We say that  $f : [a, b] \rightarrow \mathbb{R}$  is *Riemann integrable* if there exists  $I \in \mathbb{R}$  s.t. for any partition  $\mathcal{P}$  of  $[a, b]$  the Riemann sum  $\sigma(f, \mathcal{P})$  converges to  $I$  as  $\|\mathcal{P}\| \rightarrow 0$ , i.e.

$$\lim_{\|\mathcal{P}\| \rightarrow 0} \sigma(f, \mathcal{P}) = I =: \int_a^b f(x) dx.$$

**Proposition 6.4.** Let  $f, g : [a, b] \rightarrow \mathbb{R}$  be Riemann integrable and  $\alpha \in \mathbb{R}$ . Then

- $\int_a^b \alpha f(x) dx = \alpha \int_a^b f(x) dx$ .
- $f + g$  is Riemann integrable and  $\int_a^b (f(x) + g(x)) dx = \int_a^b f(x) dx + \int_a^b g(x) dx$ .
- If  $f \leq g$ , then  $\int_a^b f(x) dx \leq \int_a^b g(x) dx$ .

**Proposition 6.5.** Let  $f : [a, b] \rightarrow \mathbb{R}$  be Riemann integrable and  $c \in (a, b)$ . Then

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx.$$

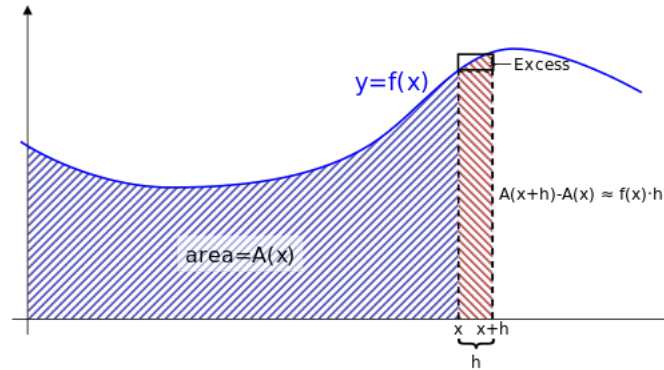


Figure 2: The derivative of the area function  $A$  is  $f$ . Source: wikipedia.

**Theorem 6.6.** Let  $f : [a, b] \rightarrow \mathbb{R}$  be Riemann integrable. Then the function  $A : [a, b] \rightarrow \mathbb{R}$ ,  $A(x) := \int_a^x f(t) dt$  is continuous. Furthermore, if  $f$  is continuous, then  $A$  is differentiable and  $A'(x) = f(x)$ .

**Theorem 6.7** (Fundamental theorem of calculus). Let  $f : [a, b] \rightarrow \mathbb{R}$  be Riemann integrable and  $F : [a, b] \rightarrow \mathbb{R}$  an antiderivative (primitive) of  $f$ , i.e.  $F' = f$ , then

$$\int_a^b f(x) dx = F(b) - F(a).$$

**Definition 6.8** (Trapezium rule). Let  $f : [a, b] \rightarrow \mathbb{R}$  be Riemann integrable and consider  $a = x_0 < x_1 < \dots < x_n = b$ . The area below the curve  $y = f(x)$  can be approximated by

$$\sum_{k=1}^n \frac{f(x_{k-1}) + f(x_k)}{2} (x_k - x_{k-1}).$$

Note that  $\frac{f(x_{k-1})+f(x_k)}{2}(x_k-x_{k-1})$  is the area of the trapezium determined by  $x_{k-1}, x_k, f(x_{k-1}), f(x_k)$ . In the case of a uniform partition with  $x_k - x_{k-1} = \frac{b-a}{n}, \forall k \in \overline{1, n}$ , we have that

$$\int_a^b f(x) dx \approx \frac{b-a}{n} \left( \frac{1}{2}f(a) + \sum_{i=1}^{n-1} f(x_i) + \frac{1}{2}f(b) \right).$$

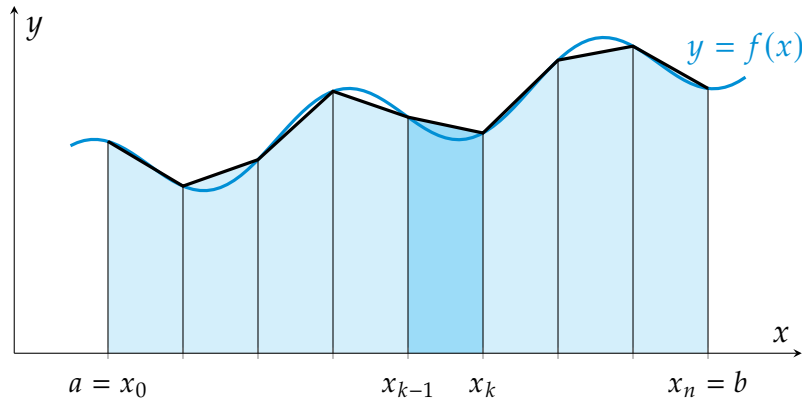


Figure 3: Trapezium rule.

**Definition 6.9.** Let  $a, b \in \mathbb{R}$ . If the following limits exist, we define the *improper integrals*

- If  $f : [a, \infty) \rightarrow \mathbb{R}$  is Riemann integrable on any compact interval in the domain,

$$\int_a^\infty f(x) dx := \lim_{t \rightarrow \infty} \int_a^t f(x) dx.$$

- If  $f : [a, b) \rightarrow \mathbb{R}$  is Riemann integrable on any compact interval included in the domain,

$$\int_a^{b-0} f(x) dx := \lim_{t \nearrow b} \int_a^t f(x) dx.$$

- If  $f : (a, b] \rightarrow \mathbb{R}$  is Riemann integrable on any compact interval included in the domain,

$$\int_{a+0}^b f(x) dx := \lim_{t \searrow a} \int_t^b f(x) dx.$$

The notation  $\int_a^{b-0} \dots, \int_{a+0}^b \dots$  emphasizes that the integrals are improper, but we can also simply write  $\int_a^b \dots$  even when dealing with an improper integral.

**Definition 6.10.** We say that an improper integral is convergent if it is finite (finite limit).

Note that an improper integral represents the area of an infinite region.

**Example 6.11.** Let  $a > 0$  and  $p \in \mathbb{R}$ . The improper integral

$$\int_a^\infty \frac{1}{x^p} dx$$

converges when  $p > 1$  and diverges when  $p \leq 1$ . Indeed, for  $p = 1$  the integral diverges ( $\ln(\infty)$ ) and for  $p \neq 1$ ,

$$\int_a^\infty \frac{1}{x^p} dx = \lim_{t \rightarrow \infty} \int_a^t f(x) dx = \lim_{t \rightarrow \infty} \frac{t^{-p+1}}{-p+1} - \frac{a^{-p+1}}{-p+1},$$

which converges when  $-p+1 < 0$ , i.e.  $p > 1$ , and diverges when  $p < 1$ .

**Example 6.12.** Let  $0 < a < b$  and  $p \in \mathbb{R}$ . The improper integrals

$$\int_a^b \frac{1}{(b-x)^p} dx, \int_a^b \frac{1}{(x-a)^p} dx$$

converge when  $p < 1$  and diverge when  $p \geq 1$ . Indeed, for  $p = 1$  the integrals diverge ( $\ln(0)$ ) and for  $p \neq 1$  the first integral, for example, is

$$\int_a^b \frac{1}{(b-x)^p} dx = -\lim_{t \nearrow b} \frac{(b-t)^{-p+1}}{-p+1} + \frac{(b-a)^{-p+1}}{-p+1},$$

which converges when  $-p+1 > 0$ , i.e.  $p < 1$ , and diverges when  $p > 1$ .

**Theorem 6.13.** Let  $a < b \leq \infty$  and  $f, g : [a, b) \rightarrow [0, \infty)$ . If there exists  $c \in (a, b)$  s.t.  $f(x) \leq g(x), \forall x \geq c$ , then

- If  $\int_a^b g(x) dx$  converges, then  $\int_a^b f(x) dx$  converges.
- If  $\int_a^b f(x) dx$  diverges, then  $\int_a^b g(x) dx$  diverges.
- If  $\lim_{x \nearrow b} \frac{f(x)}{g(x)} \in (0, \infty)$ , then  $\int_a^b f(x) dx$  and  $\int_a^b g(x) dx$  have the same nature.

**Theorem 6.14** (Integral test for series). Let  $f : [1, \infty) \rightarrow [0, \infty)$  be decreasing, then  $\int_1^\infty f(x) dx$  and  $\sum_{n=1}^\infty f(n)$  have the same nature.

*Proof.* Let  $N \in \mathbb{N}$  and write  $\int_1^N f(x) dx = \sum_{n=1}^{N-1} \int_n^{n+1} f(x) dx$ . Since  $f$  is decreasing we have that

$$\sum_{n=1}^{N-1} f(n+1) \leq \int_1^N f(x) dx \leq \sum_{n=1}^{N-1} f(n).$$

The conclusion follows by letting  $N \rightarrow \infty$  and using the comparison test.  $\square$

## ❖ The Euclidean space $\mathbb{R}^n$

Elements in  $\mathbb{R}^n$  are vectors with  $n$  components. We will write  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$  most of the time, apart from situations where matrices will also be involved – in this case we

will adopt the linear algebra notation of writing  $x \in \mathbb{R}^n$  as a column vector  $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{n \times 1}$ , since this allows to multiply matrices  $\begin{bmatrix} & \end{bmatrix}_{m \times n}$  with vectors  $\begin{bmatrix} \end{bmatrix}_{n \times 1}$  to get vectors  $\begin{bmatrix} \end{bmatrix}_{m \times 1}$ .

As you've seen in your Algebra course,  $\mathbb{R}^n$  is a vector space: two vectors  $x, y \in \mathbb{R}^n$  can be added component wise  $x + y := (x_1 + y_1, \dots, x_n + y_n)$ , and a vector can be multiplied by a scalar  $\alpha \in \mathbb{R}$  to get  $\alpha x := (\alpha x_1, \dots, \alpha x_n)$ . We will denote by  $e_i$  the canonical basis vector with a 1 in the  $i$ th component and 0's everywhere else, giving  $x = x_1 e_1 + \dots + x_n e_n$ .

**Definition 7.1.** A map  $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  is called a *scalar product* (or *inner product*) if

- (a)  $\langle x, y \rangle = \langle y, x \rangle$ , for any  $x, y \in \mathbb{R}^n$ .
- (b)  $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$ , for any  $x, y, z \in \mathbb{R}^n$  and  $\alpha, \beta \in \mathbb{R}$ .
- (c)  $\langle x, x \rangle > 0$ , for any  $x \in \mathbb{R}^n \setminus \{0\}$ .

**Example 7.2.** A symmetric positive definite matrix  $M \in \mathbb{R}^{n \times n}$  defines a scalar product  $\langle x, y \rangle = x^T M y$ .

**Definition 7.3.** The *dot product* of two vectors  $x, y \in \mathbb{R}^n$  is given by

$$x \cdot y := x_1 y_1 + \dots + x_n y_n.$$

The dot product is the most important scalar product. In matrix notation, it is written as

$$x \cdot y = x^T y = [x_1 \dots x_n]_{1 \times n} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1} = x_1 y_1 + \dots + x_n y_n.$$

**Definition 7.4.** Two vectors  $x, y \in \mathbb{R}^n$  are perpendicular (or orthogonal) iff  $x \cdot y = 0$ .

**Definition 7.5.** A function  $\| \cdot \| : \mathbb{R}^n \rightarrow [0, \infty)$  is called a *norm* if

- (a)  $\|x\| = 0$  if and only if  $x = 0$ .
- (b)  $\|\alpha x\| = |\alpha| \|x\|$ , for any  $x \in \mathbb{R}^n$  and  $\alpha \in \mathbb{R}$ .
- (c)  $\|x + y\| \leq \|x\| + \|y\|$ , for any  $x, y \in \mathbb{R}^n$  (triangle inequality).

**Proposition 7.6.** Any scalar product generates a norm on  $\mathbb{R}^n$  given by  $\|x\| = \sqrt{\langle x, x \rangle}$ .

**Theorem 7.7** (Cauchy-Schwarz inequality). For any  $x, y \in \mathbb{R}^n$  it holds that

$$|\langle x, y \rangle| \leq \|x\| \|y\|.$$

Here the norm  $\|\cdot\|$  is generated by the scalar product  $\langle \cdot, \cdot \rangle$ .

*Proof.* Consider  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(t) = \langle tx + y, tx + y \rangle = \|tx + y\|^2 \geq 0$ . Since  $f(t) = t^2\|x\|^2 + 2t\langle x, y \rangle + \|y\|^2$  is a quadratic in  $t$ , we must have that  $\Delta = 4\langle x, y \rangle^2 - 4\|x\|^2\|y\|^2 \leq 0$   $\square$

**Definition 7.8.** The Euclidean norm is generated by the dot product and it is given by

$$\|x\| = \sqrt{x \cdot x} = \sqrt{x_1^2 + \dots + x_n^2}.$$

This represents the length of the vector  $x \in \mathbb{R}^n$  measured using the Euclidean norm.

**Theorem 7.9.** For  $n \in \{2, 3\}$  the dot product of  $x, y \in \mathbb{R}^n$  is

$$x \cdot y = \|x\| \|y\| \cos \angle(x, y).$$

*Proof.* Consider the triangle with sides determined by the vectors  $x, y$  and  $x - y$ . From the cosine rule we have that

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2\|x\| \|y\| \cos \angle(x, y).$$

Since  $\|x - y\|^2 = (x - y) \cdot (x - y) = x \cdot x + y \cdot y - 2x \cdot y$ , we obtain that  $x \cdot y = \|x\| \|y\| \cos \angle(x, y)$ .  $\square$

**Example 7.10.** (a)  $\|x\|_1 := |x_1| + \dots + |x_n|$  is a norm (so-called Manhattan norm).

(b)  $\|x\|_p := (|x_1|^p + \dots + |x_n|^p)^{\frac{1}{p}}$ ,  $p > 1$ , is a norm.

(c)  $\|x\|_\infty := \max\{|x_1|, \dots, |x_n|\}$  is a norm.

**Definition 7.11.** A function  $d : \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0, \infty)$  is called a *distance* (or *metric*) if

(a)  $d(x, y) = 0$  if and only if  $x = y$ .

(b)  $d(x, y) = d(y, x)$ , for any  $x, y \in \mathbb{R}^n$ .

(c)  $d(x, z) \leq d(x, y) + d(y, z)$ , for any  $x, y, z \in \mathbb{R}^n$  (triangle inequality).

**Proposition 7.12.** Any norm generates a distance on  $\mathbb{R}^n$  given by  $d(x, y) = \|x - y\|$ .

**Definition 7.13.** The Euclidean distance is generated by the Euclidean norm and it is given by

$$d(x, y) = \|x - y\| = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}.$$

We will be using the Euclidean norm and distance, unless we specify otherwise.

**Neighborhoods. Interior. Closure. Boundary.**

**Definition 7.14.** A set  $A \subseteq \mathbb{R}^n$  is called *bounded* if there exists  $r > 0$  such that

$$\|x\| \leq r, \forall x \in A.$$

**Definition 7.15.** Let  $x_0 \in \mathbb{R}^n$  and  $r > 0$ . The open ball of centre  $x_0$  and radius  $r$  is given by

$$B(x_0, r) := \{x \in \mathbb{R}^n \mid \|x - x_0\| < r\},$$

and the closed ball of centre  $x_0$  and radius  $r$  is given by

$$\overline{B}(x_0, r) := \{x \in \mathbb{R}^n \mid \|x - x_0\| \leq r\}.$$

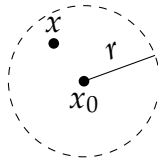


Figure 4: Open ball  $B(x_0, r)$ .

**Definition 7.16.** A set  $V \subseteq \mathbb{R}^n$  is a *neighborhood (vecinity)* of  $x \in \mathbb{R}^n$  if

$$\exists r > 0 \text{ such that } B(x, r) \subseteq V.$$

We denote all the neighborhoods of  $x$  by  $\mathcal{V}(x) := \{V \subseteq \mathbb{R}^n \mid V \text{ is a neighborhood of } x\}$ .

**Definition 7.17.** Let  $A \subseteq \mathbb{R}^n$ . The following set is called the *interior* of  $A$

$$\text{int}(A) := \{x \in \mathbb{R}^n \mid \exists V \in \mathcal{V}(x) \text{ such that } V \subseteq A\},$$

the following set is called the *closure* of  $A$

$$\text{cl}(A) := \{x \in \mathbb{R}^n \mid \forall V \in \mathcal{V}(x), V \cap A \neq \emptyset\},$$

and the following set is called the *boundary* of  $A$

$$\text{bd}(A) := \{x \in \mathbb{R}^n \mid \forall V \in \mathcal{V}(x), V \cap A \neq \emptyset \text{ and } V \cap A^c \neq \emptyset\}.$$

**Example 7.18.** Let  $A = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$ . Then

$$\text{int}(A) = A,$$

$$\text{cl}(A) = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\},$$

$$\text{bd}(A) = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}.$$

**Proposition 7.19.** For any  $A \subseteq \mathbb{R}^n$ , it holds that  $\text{cl}(A) = A \cup \text{bd}(A)$ .

**Definition 7.20.** If  $A = \text{int}(A)$ , then  $A$  is called *open*. If  $A = \text{cl}(A)$ , then  $A$  is called *closed*.

**Proposition 7.21.** For any  $A \subseteq \mathbb{R}^n$ , it holds that  $\text{int}(A) \subseteq A \subseteq \text{cl}(A)$ .

*Proof.* Similar to [proposition 1.12](#). □

**Remark 7.22.** To prove that a set  $A$  is open, it is sufficient to prove that  $A \subseteq \text{int}(A)$ .

To prove that a set  $A$  is closed, it is sufficient to prove that  $\text{cl}(A) \subseteq A$ .

**Proposition 7.23.** The following statements are true:

- The complement of an open set is closed.
- The complement of a closed set is open.

*Proof.* Similar to [proposition 1.15](#). □

### Sequences.

A sequence  $(x^k)$  in  $\mathbb{R}^n$  indexed by  $k \in \mathbb{N}$  has vector elements  $x^1, x^2, \dots, x^k, \dots$ . Notice that the index  $k$  appears as superscript (in order to avoid confusion with the coordinates of the vectors).

**Definition 7.24.** A sequence  $(x^k)$  converges to  $x \in \mathbb{R}^n$  if  $\lim_{k \rightarrow \infty} \|x^k - x\| = 0$ . We write  $\lim_{k \rightarrow \infty} x^k = x$ .

**Example 7.25.** Let  $x^k = (\frac{1}{k}, \frac{k}{k+1})$ , then  $\lim_{k \rightarrow \infty} x^k = (0, 1)$ .

**Theorem 7.26.** A sequence  $(x^k)$  converges to  $x \in \mathbb{R}^n$  if and only if  $\lim_{k \rightarrow \infty} x_i^k = x_i, \forall i = \overline{1, n}$ .

*Proof.* Consider first  $i \in \{1, \dots, n\}$ . We have that

$$|x_i^k - x_i| = \sqrt{(x_i^k - x_i)^2} \leq \sqrt{(x_1^k - x_1)^2 + \dots + (x_n^k - x_n)^2} = \|x^k - x\|,$$

hence if  $(x^k)$  converges to  $x \in \mathbb{R}^n$ , i.e.  $\|x^k - x\| \rightarrow 0$ , then  $|x_i^k - x_i| \rightarrow 0$  and  $x_i^k \rightarrow x_i$ .

Let us now prove the converse statement and assume that  $\lim_{k \rightarrow \infty} x_i^k = x_i, \forall i = \overline{1, n}$ . Then

$$\|x^k - x\| = \sqrt{(x_1^k - x_1)^2 + \dots + (x_n^k - x_n)^2} \rightarrow 0,$$

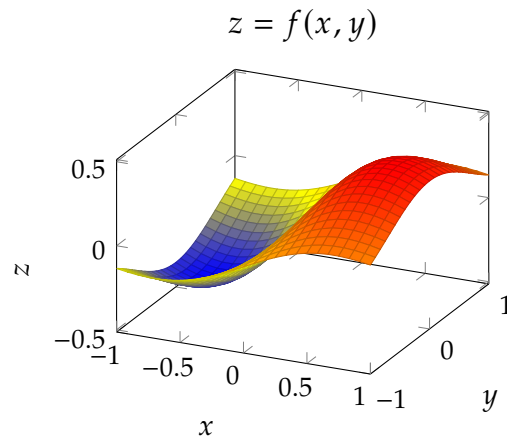
hence  $x^k \rightarrow x$ . □

Note that this is telling us that a sequence of vectors converges if and only if the components of the vectors converge, respectively.



## ❖ Functions of several variables. Limits and continuity

We will now introduce functions of several variables, focusing on those having real (scalar) values. This means we will mostly consider functions  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  mapping vectors in  $\mathbb{R}^n$  into real numbers. As you already know, when  $n = 1$  the graph of a function is a curve in  $\mathbb{R}^2$ . When  $n = 2$ , the graph of a function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  is given by points with coordinates  $(x, y, f(x, y))$  – this represents a surface in  $\mathbb{R}^3$  (an example is shown in the figure below).



What about when  $n \geq 3$ ? The graph of the function,  $\{(x, f(x)) \in \mathbb{R}^{n+1} \mid x \in A \subseteq \mathbb{R}^n\}$ , would be a set in  $\mathbb{R}^{n+1}$  and we are able to visualize only its projections in lower dimensional spaces ( $\mathbb{R}^3$  or  $\mathbb{R}^2$ ). Apart from the graph, another way of visualizing a function is through

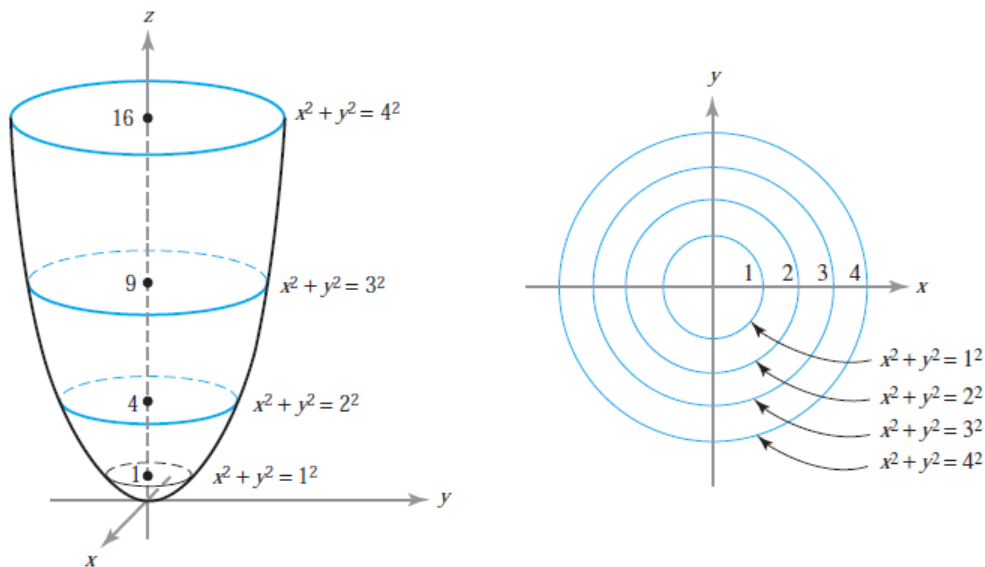


Figure 5: Graph and level curves for  $f(x, y) = x^2 + y^2$ . Source: [3, page 80].

its level sets, which are given by

$$L_c := \{x \in A \subseteq \mathbb{R}^n \mid f(x) = c\},$$

for a constant  $c \in \mathbb{R}$ . If  $n = 2$ , the set  $L_c = \{(x, y) \in A \mid f(x, y) = c\}$  describes a *level curve* (see the figure above). If  $n = 3$ , the set  $L_c = \{(x, y, z) \in A \mid f(x, y, z) = c\}$  describes a *level surface*.

**Limits of functions of several variables. Continuity.**

Using neighbourhoods in  $\mathbb{R}^n$ , we can define limits and continuity exactly as in [section 4](#).

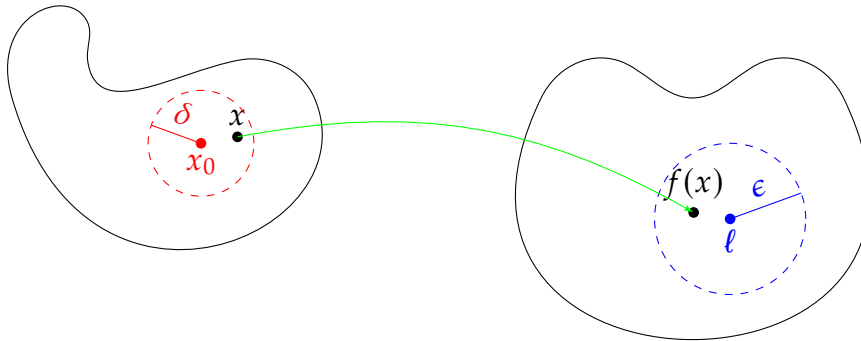
**Definition 8.1.** Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  and  $x_0 \in A'$ . We say that  $\lim_{x \rightarrow x_0} f(x) = \ell \in \overline{\mathbb{R}}$  if

$$\forall V \in \mathcal{V}(\ell), \exists U \in \mathcal{V}(x_0) \text{ s.t. } f(x) \in V, \forall x \in U \cap (A \setminus \{x_0\}).$$

**Remark 8.2.** Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  and  $x_0 \in A'$ . We say that  $\lim_{x \rightarrow x_0} f(x) = \ell \in \mathbb{R}$  if

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } |f(x) - \ell| < \varepsilon, \forall x \in A \text{ with } \|x - x_0\| < \delta.$$

A similar definition can be given when  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ , illustrated below. □



**Theorem 8.3.** Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  and  $x_0 \in A'$ . Then  $\lim_{x \rightarrow x_0} f(x) = \ell \in \overline{\mathbb{R}}$  if and only if

for any sequence  $(x^k)$  in  $A \setminus \{x_0\}$  with  $\lim_{k \rightarrow \infty} x^k = x_0$ , we have that  $\lim_{k \rightarrow \infty} f(x^k) = \ell \in \overline{\mathbb{R}}$ . □

Let us now consider some limits in  $\mathbb{R}^2$  and explore some methods of computing them.

**Example 8.4.** (a)  $\lim_{(x,y) \rightarrow (0,0)} \frac{x^2}{\sqrt{x^2 + y^2}} = 0.$

We will try a simple strategy: to bound the function and use the squeeze theorem. Since  $0 \leq \frac{x^2}{\sqrt{x^2 + y^2}} \leq \frac{x^2 + y^2}{\sqrt{x^2 + y^2}} = \sqrt{x^2 + y^2}$  and  $\sqrt{x^2 + y^2} \rightarrow 0$ , as  $(x, y) \rightarrow (0, 0)$ , we have that the limit is zero.

$$(b) \quad \lim_{(x,y) \rightarrow (0,0)} \frac{\sin(x^2 + y^2)}{x^2 + y^2} = 1.$$

Here we can use a remarkable limit since  $t := x^2 + y^2 \rightarrow 0$  as  $(x, y) \rightarrow (0, 0)$ , hence the limit equals  $\lim_{t \rightarrow 0} \frac{\sin t}{t} = 1$ .

$$(c) \quad \lim_{(x,y) \rightarrow (0,0)} \frac{xy}{x^2 + y^2} \text{ does not exist.}$$

As  $(x, y) \rightarrow (0, 0)$  we are having points in  $\mathbb{R}^2$  that converge towards the origin. The points can approach the origin along any path – if the limit exists, we will always get the same thing. One important strategy is thus to approach the origin along different paths: if the function converges to different values, then the limit doesn't exist! The simplest paths we could consider are lines that pass through the origin, i.e. points  $(x, mx)$ . For our example,  $\lim_{(x,mx) \rightarrow (0,0)} \frac{xy}{x^2 + y^2} = \lim_{(x,mx) \rightarrow (0,0)} \frac{mx^2}{(1 + m^2)x^2} = \frac{m}{1 + m^2}$ . The limit value depends on the gradient  $m$ , e.g. for  $m = 0$  we get 0 and for  $m = 1$  we get  $1/2$ , so the limit does not exist.

**Remark 8.5.** If  $\lim_{(x,y) \rightarrow (x_0,y_0)} f(x, y)$  exists, then  $\lim_{x \rightarrow x_0} \lim_{y \rightarrow y_0} f(x, y) = \lim_{y \rightarrow y_0} \lim_{x \rightarrow x_0} f(x, y)$ .

**Definition 8.6.** Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  and  $x_0 \in A$ . We say that  $f$  is continuous at  $x_0$  if

$$\forall V \in \mathcal{V}(f(x_0)), \exists U \in \mathcal{V}(x_0) \text{ s.t. } f(x) \in V, \forall x \in U \cap A.$$

**Remark 8.7.** Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  and  $x_0 \in A \cap A'$  an accumulation point. Then  $f$  is continuous at  $x_0$  if and only if

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

**Remark 8.8.** If  $x_0 \in A$  is an isolated point, then  $f$  is continuous at  $x_0$ .

**Theorem 8.9.** Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  and  $x_0 \in A \cap A'$ . The following are equivalent:

- (a)  $f$  is continuous at  $x_0$ .
- (b)  $\forall \varepsilon > 0, \exists \delta > 0$  s.t.  $|f(x) - f(x_0)| < \varepsilon, \forall x \in A$  with  $\|x - x_0\| < \delta$ .
- (c) for any sequence  $(x^k)$  in  $A$  with  $\lim_{k \rightarrow \infty} x^k = x_0$ , we have that  $\lim_{k \rightarrow \infty} f(x^k) = f(x_0)$ .

**Proposition 8.10** (Any norm is continuous).  $f : \mathbb{R}^n \rightarrow \mathbb{R}, f(x) = \|x\|$  is continuous on  $\mathbb{R}^n$ .

*Proof.* Let  $x_0 \in \mathbb{R}^n$  and  $(x^k)$  with  $x^k \rightarrow x_0$ . We have that  $\|x^k - x_0\| \rightarrow 0$ . By the triangle inequality  $|\|x^k\| - \|x_0\|| \leq \|x^k - x_0\| \rightarrow 0$ , hence  $\|x^k\| \rightarrow \|x_0\|$ , i.e.  $\|x^k\| \rightarrow \|x_0\|$ .  $\square$

**Theorem 8.11** (Weierstrass). Let  $A \subseteq \mathbb{R}^n$  be closed and bounded, and  $f : A \rightarrow \mathbb{R}$  a continuous function. Then  $f$  is bounded and it attains its bounds, i.e. there exist  $\min(f(A)), \max(f(A))$ .  $\square$

## ❖ Partial derivatives and differentiability in $\mathbb{R}^n$ . Gradient descent

**Definition 9.1.** Let  $A \subseteq \mathbb{R}^n$  be an open set and  $f : A \rightarrow \mathbb{R}$ . The *partial derivative* of  $f$  with respect to  $x_i$  at the point  $x = (x_1, \dots, x_n) \in A$  is given by

$$\begin{aligned} \frac{\partial f}{\partial x_i}(x) &= \partial_i f(x) := \lim_{h \rightarrow 0} \frac{f(x_1, \dots, x_{i-1}, x_i + h, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_n)}{h} \\ &= \lim_{h \rightarrow 0} \frac{f(x + he_i) - f(x)}{h}. \end{aligned}$$

Note that  $\frac{\partial f}{\partial x_i}$  is the derivative of  $f$  with respect to  $x_i$ , with the other variables held fixed.

**Definition 9.2.** For a function  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  that has partial derivatives at  $x \in A$  with respect to all its variables, the *gradient* at  $x$  is given by the vector  $\nabla f(x) \in \mathbb{R}^n$ ,

$$\nabla f(x) := \left( \frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right).$$

**Example 9.3.** For  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ ,  $f(x, y) = x^2y + y^2$  we have that  $\nabla f(x, y) = (\frac{\partial f}{\partial x}(x, y), \frac{\partial f}{\partial y}(x, y)) = (2xy, x^2 + 2y)$ .

**Example 9.4** (With partial derivatives but discontinuous). Let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ ,

$$f(x, y) = \begin{cases} \frac{xy}{x^2 + y^2}, & (x, y) \neq (0, 0) \\ 0, & (x, y) = (0, 0). \end{cases}$$

Since  $f(x, 0) - f(0, 0) = 0$ ,  $\frac{\partial f}{\partial x}(0, 0) = \lim_{x \rightarrow 0} \frac{f(x, 0) - f(0, 0)}{x} = 0 = \frac{\partial f}{\partial y}(0, 0)$ , so  $f$  has partial derivatives zero at  $(0, 0)$ . But  $\lim_{(x, mx) \rightarrow (0, 0)} \frac{mx^2}{(m^2 + 1)x^2} = \frac{m}{(m^2 + 1)}$  depends on  $m$ , so  $f$  doesn't have a limit at  $(0, 0)$ , which means that  $f$  is discontinuous at  $(0, 0)$ .  $\square$

As the above example shows, a function that has partial derivatives at a point doesn't have to be continuous. This means that if we want to have good properties for differentiable functions, we have to find a better way of defining differentiability.

Let us recall an important idea for differentiable functions in  $\mathbb{R}$ :  $f(x_0) + f'(x_0)(x - x_0)$  is the linear approximation to  $f(x)$ . This comes from the definition of the derivative

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{x - x_0} = 0,$$

which can also be written as

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + R(x - x_0), \text{ with } \frac{R(x - x_0)}{x - x_0} \rightarrow 0,$$

where  $R(x - x_0)$  is the remainder of the linear approximation.

**Definition 9.5.** Let  $A \subseteq \mathbb{R}^n$  be an open set and  $f : A \rightarrow \mathbb{R}$ . We say that  $f$  is *differentiable* at  $x_0 \in A$  if there exists a vector  $Df(x_0) \in \mathbb{R}^n$ , called the *differential/derivative* of  $f$  at  $x_0$ , s.t.

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0) - Df(x_0) \cdot (x - x_0)}{\|x - x_0\|} = 0.$$

Here  $Df(x_0) \cdot (x - x_0)$  is the dot product of two vectors. This can also be written as

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0) - Df(x_0) \cdot h}{\|h\|} = 0.$$

Note that differentiability is equivalent to

$$f(x) = f(x_0) + Df(x_0) \cdot (x - x_0) + R(x - x_0), \text{ with } \frac{R(x - x_0)}{\|x - x_0\|} \rightarrow 0,$$

where  $R(x - x_0)$  is the remainder of the linear approximation.

**Definition 9.6.** Let  $A \subseteq \mathbb{R}^n$  be an open set. If  $f : A \rightarrow \mathbb{R}^m$ ,  $f = (f_1, \dots, f_m)$ , then  $f$  is differentiable at  $x_0$  if there exists a matrix  $Df(x_0) \in \mathbb{R}^{m \times n}$ , called the *differential/derivative* of  $f$  at  $x_0$ , s.t.

$$\lim_{x \rightarrow x_0} \frac{\|f(x) - f(x_0) - Df(x_0)(x - x_0)\|_m}{\|x - x_0\|_n} = 0.$$

Here  $Df(x_0)(x - x_0)$  is a matrix–vector product:  $[ \quad ]_{m \times n} [ \quad ]_{n \times 1} = [ \quad ]_{m \times 1}$ .

**Example 9.7.** Constant functions have zero derivative and linear functions have a constant derivative.

- If  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is constant, then  $Df(x) = 0$  since  $f(x) = f(x_0)$  for any  $x, x_0 \in \mathbb{R}^n$ .
- If  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f(x) = a \cdot x$  with  $a \in \mathbb{R}^n$ , then  $Df(x) = a$  since  $f(x) - f(x_0) - a \cdot (x - x_0) = 0$ .
- If  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $f(x) = Ax$  with  $A \in \mathbb{R}^{m \times n}$ , then  $Df(x) = A$ ;  $f(x) - f(x_0) - A(x - x_0) = 0$ .

**Theorem 9.8.** Let  $A \subseteq \mathbb{R}^n$  be an open set. If  $f : A \rightarrow \mathbb{R}^m$  is differentiable at  $x_0$ , then  $f$  is continuous at  $x_0$ .

*Proof.* Since  $f$  is differentiable,  $f(x) = f(x_0) + Df(x_0)(x - x_0) + \frac{R(x - x_0)}{\|x - x_0\|} \|x - x_0\|$ . Letting  $x \rightarrow x_0$  we use that  $\frac{\|R(x - x_0)\|}{\|x - x_0\|} \rightarrow 0$  to obtain that  $f(x) \rightarrow f(x_0)$ .  $\square$

**Theorem 9.9.** Let  $A \subseteq \mathbb{R}^n$  be an open set and  $x \in A$ . If  $f : A \rightarrow \mathbb{R}$  is differentiable at  $x$ , then the partial derivatives exist at  $x$  and

$$Df(x) = \nabla f(x) = \left( \frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right).$$

*Proof.* Differentiability at  $x$  gives

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x) - Df(x) \cdot h}{\|h\|} = 0.$$

Let us take the vector  $h$  in the direction of  $e_i$ , with a non-zero value on the  $i$ th component only, i.e.  $h = (0, \dots, 0, h_i, 0, \dots, 0) = h_i e_i$ . Then we have that

$$\lim_{h_i \rightarrow 0} \frac{f(x + h_i e_i) - f(x) - Df(x) \cdot h_i e_i}{|h_i|} = 0 = \lim_{h_i \rightarrow 0} \frac{f(x + h_i e_i) - f(x) - Df(x) \cdot h_i e_i}{h_i},$$

which gives that

$$\frac{\partial f}{\partial x_i}(x) = \lim_{h_i \rightarrow 0} \frac{f(x + h_i e_i) - f(x)}{h_i} = Df(x) \cdot e_i = Df(x)_i,$$

hence the  $i$ th component of  $Df(x)$  is  $\frac{\partial f}{\partial x_i}(x)$ , which means that  $Df(x) = \nabla f(x)$ .  $\square$

**Theorem 9.10.** Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  and  $x \in A$ . If all the partial derivatives exist and are continuous at  $x$ , then  $f$  is differentiable at  $x$ .

*Proof.* (Optional) See [1][Theorem 9.21].  $\square$

It is possible for a function to have partial derivatives, but not be differentiable if the partial derivatives are not continuous. The function in [example 9.4](#) has partial derivatives at  $(0, 0)$ , but it is discontinuous there, so it is not differentiable at that point.

**Theorem 9.11.** Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $f = (f_1, \dots, f_m)$  be differentiable at  $x \in A$ , then

$$Df(x) = J = \begin{bmatrix} \nabla f_1(x) \\ \vdots \\ \nabla f_m(x) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \dots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \dots & \frac{\partial f_m}{\partial x_n}(x) \end{bmatrix}_{m \times n}.$$

This matrix is called the *Jacobian matrix* and is typically denoted by  $J$ .

**Theorem 9.12** (Calculus rules). Let  $A \subseteq \mathbb{R}^n$  and  $f, g : A \rightarrow \mathbb{R}$  differentiable at  $x \in A$ . Then

1.  $\nabla(f + g)(x) = \nabla f(x) + \nabla g(x)$ .
2.  $\nabla(fg)(x) = g(x)\nabla f(x) + f(x)\nabla g(x)$ .
3.  $\nabla\left(\frac{f}{g}\right)(x) = \frac{g(x)\nabla f(x) - f(x)\nabla g(x)}{g^2(x)}$ .

**Theorem 9.13** (Chain rule). Let  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $f : \mathbb{R}^m \rightarrow \mathbb{R}^p$  differentiable at  $x$  and  $g(x)$ , respectively. Then

$$D(f \circ g)(x) = Df(g(x))Dg(x).$$

In terms of matrix dimensions:  $[ ]_{p \times n} = [ ]_{p \times m} [ ]_{m \times n}$ .

*Proof.* (Optional) Considering  $E(h) := \|f(g(x+h)) - f(g(x)) - Df(g(x))Dg(x)h\|$  we aim to prove that  $\lim_{h \rightarrow 0} \frac{E(h)}{\|h\|} = 0$ . Since

$$\begin{aligned} E(h) &= \|f(g(x+h)) - f(g(x)) - Df(g(x))(g(x+h) - g(x)) \\ &\quad + Df(g(x))(g(x+h) - g(x)) - Df(g(x))Dg(x)h\|, \end{aligned}$$

using the triangle inequality we have that

$$\begin{aligned} E(h) &\leq \|f(g(x+h)) - f(g(x)) - Df(g(x))(g(x+h) - g(x))\| \\ &\quad + \|Df(g(x))\| \|g(x+h) - g(x) - Dg(x)h\|. \end{aligned}$$

At this point, after a few intermediate steps, one can now divide by  $\|h\|$ , take  $h \rightarrow 0$  and use the differentiability of  $f$  at  $g(x)$ , and of  $g$  at  $x$ , together with the fact that  $\|Df(g(x))\|$  is bounded and independent of  $h$ .  $\square$

**Example 9.14** (Chain rule). Let  $g : \mathbb{R} \rightarrow \mathbb{R}^n$ ,  $g = (g_1, \dots, g_n)$ ,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \circ g : \mathbb{R} \rightarrow \mathbb{R}$ . With  $g'(t) = (g'_1(t), \dots, g'_n(t))$ , we have that

$$\begin{aligned} (f \circ g)'(t) &= \nabla f(g(t)) \cdot g'(t) \\ &= \sum_{i=1}^n \frac{\partial f}{\partial x_i}(g(t)) \cdot g'_i(t) \end{aligned}$$

If  $n = 2$  and  $g(t) = (x(t), y(t))$ , then  $(f \circ g)(t) = f(x(t), y(t)) = h(t)$  and

$$\frac{dh}{dt} = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt}.$$

**Definition 9.15.** Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  and a vector  $v \in \mathbb{R}^n$ . The derivative of  $f$  in the direction of  $v$  at  $x \in A$  (*directional derivative*) is given by

$$Df_v(x) := \lim_{h \rightarrow 0} \frac{f(x + hv) - f(x)}{h}.$$

Note that here  $h$  is a scalar. The directional derivative  $Df_v(x)$  is also denoted by  $\partial_v f(x)$ .

**Theorem 9.16.** Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  and  $v \in \mathbb{R}^n$ . If  $f$  is differentiable at  $x \in A$ , then

$$Df_v(x) = \nabla f(x) \cdot v.$$

*Proof.* From differentiability we have that

$$\lim_{h \rightarrow 0} \frac{f(x + hv) - f(x) - \nabla f(x) \cdot hv}{\|hv\|} = 0.$$

Since  $\|hv\| = |h|\|v\|$ , this gives that

$$\lim_{h \rightarrow 0} \frac{f(x + hv) - f(x) - \nabla f(x) \cdot hv}{h} = 0,$$

which can be rearranged as

$$Df_v(x) = \lim_{h \rightarrow 0} \frac{f(x + hv) - f(x)}{h} = \nabla f(x) \cdot v.$$

□

**Theorem 9.17** (Fermat). Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  be differentiable at  $x \in A$ . If  $x$  is a local extremum, then it is a critical point, i.e.  $\nabla f(x) = 0$ .

*Proof.*  $x$  is an extremum in every direction, thus  $0 = Df_v(x) = \nabla f(x) \cdot v$  for every  $v \in \mathbb{R}^n$  (including the canonical vectors  $e_i$ ). This gives that  $\nabla f(x) = 0$ . □

**Proposition 9.18** (Direction of steepest ascent/descent). Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  be differentiable at  $x \in A$  with  $\nabla f(x) \neq 0$ . Then

- $\nabla f(x)$  gives the direction of fastest increase (steepest ascent).
- $-\nabla f(x)$  gives the direction of fastest decrease (steepest descent).

*Proof.* Since  $Df_v(x) = \nabla f(x) \cdot v$ , by the Cauchy-Schwarz inequality we have that

$$-\|\nabla f(x)\|\|v\| \leq Df_v(x) \leq \|\nabla f(x)\|\|v\|,$$

with the maximum obtained for  $v = \alpha \nabla f(x)$ , the minimum for  $v = -\alpha \nabla f(x)$ ,  $\alpha > 0$ . □

**Proposition 9.19** (Gradient perpendicular to the level set). Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  be differentiable at  $x \in A$ . Then  $\nabla f(x)$  is perpendicular to the level set containing  $x$ , i.e. if  $v$  is a tangent vector to the level set then  $Df_v(x) = \nabla f(x) \cdot v = 0$ .

*Proof.* Let  $c(t)$ ,  $t \geq 0$  be a path on the level set, i.e.  $f(c(t)) = \text{constant}$ , that starts from  $x = c(0)$ . Let  $v = c'(0)$  be the tangent vector to the path at  $t = 0$ . By the chain rule

$$0 = \left. \frac{d}{dt} f(c(t)) \right|_{t=0} = \nabla f(c(0)) \cdot c'(0) = \nabla f(x) \cdot v.$$

□



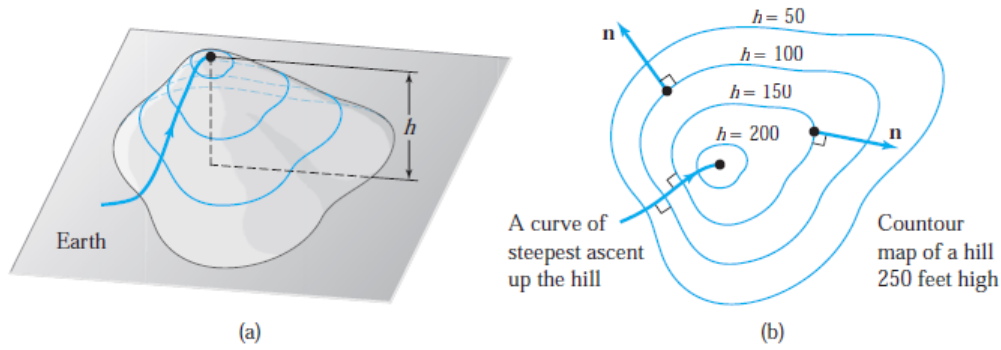


Figure 6: The gradient is the steepest ascent direction and is perpendicular to the level curves. Source: [3, page 140].

**Example 9.20** (Tangent line to a level curve). Consider a level curve  $L = \{(x, y) \in \mathbb{R}^2 \mid f(x, y) = c\}$  and a point  $(x_0, y_0)$  on it. Take the tangent line at that point. If  $(x, y)$  is a point on the tangent line, then the gradient is perpendicular to the vector  $(x - x_0, y - y_0)$ ,

$$\nabla f(x_0, y_0) \cdot (x - x_0, y - y_0) = 0,$$

hence the equation of the tangent line is given by

$$\frac{\partial f}{\partial x}(x_0, y_0) \cdot (x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0) \cdot (y - y_0) = 0.$$

**Example 9.21** (Tangent plane to a level surface). Consider a level curve  $L = \{(x, y, z) \in \mathbb{R}^3 \mid f(x, y, z) = c\}$  and a point  $(x_0, y_0, z_0)$  on it. Take the tangent plane at that point. If  $(x, y, z)$  is a point on the tangent plane, then

$$\nabla f(x_0, y_0, z_0) \cdot (x - x_0, y - y_0, z - z_0) = 0,$$

hence the equation of the tangent plane is given by

$$\frac{\partial f}{\partial x}(x_0, y_0, z_0) \cdot (x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0, z_0) \cdot (y - y_0) + \frac{\partial f}{\partial z}(x_0, y_0, z_0) \cdot (z - z_0) = 0.$$

**Gradient descent.** One of the most important optimization methods is based on the fundamental idea that the direction of steepest descent on the graph of a function  $f$  is given by  $-\nabla f$  (see [proposition 9.18](#)). If we want to minimize a function  $f$ , this naturally suggests an iterative method in which from a current position  $x_k$  we move in the direction of  $-\nabla f(x_k)$  with a step size  $s_k > 0$  in order to get to the next position  $x_{k+1}$  such that  $f(x_{k+1}) < f(x_k)$ . The gradient descent method starts from an initial value  $x_0 \in \mathbb{R}^n$  and then for  $k \geq 0$  takes

$$x_{k+1} = x_k - s_k \nabla f(x_k).$$

The step size (also called learning rate)  $s_k$  has to be chosen at every iteration. One way of doing this is called *exact line search*: we look for the optimal step size  $s_k$  that minimizes the function  $\varphi(s) = f(x_{k+1}) = f(x_k - s\nabla f(x_k))$ . By the chain rule we have that

$$\varphi'(s) = \nabla f(x_{k+1}) \cdot \frac{d}{ds} x_{k+1} = \nabla f(x_{k+1}) \cdot (-\nabla f(x_k)).$$

Since  $\varphi'(s_k) = 0$  for the optimal step size  $s_k$ , we see that  $\nabla f(x_{k+1}) \cdot \nabla f(x_k) = 0$ , which means that two consecutive search directions are perpendicular, i.e.  $(x_{k+2} - x_{k+1}) \perp (x_{k+1} - x_k)$ .

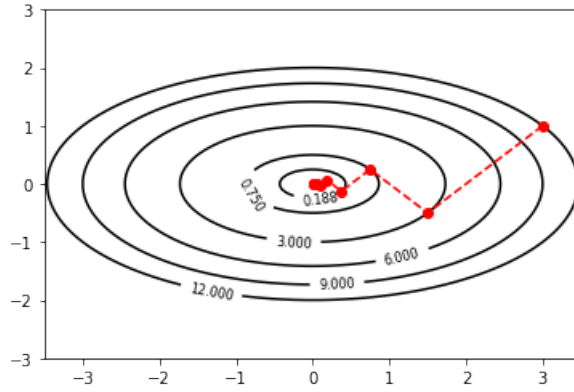


Figure 7: Level curves and gradient descent iterations for [example 9.22](#).

**Example 9.22.** Consider the quadratic function  $f(x, y) = x^2 + 3y^2$  which has a unique global minimum at the origin  $(0, 0)$ . The gradient is given by  $\nabla f(x, y) = (2x, 6y)$ .

Gradient descent: starting from an initial value  $(x_0, y_0)$  consider the sequence

$$(x_{k+1}, y_{k+1}) = (x_k, y_k) - s\nabla f(x_k, y_k),$$

that is

$$x_{k+1} = (1 - 2s)x_k, \quad y_{k+1} = (1 - 6s)y_k.$$

The step size is determined using exact line search. We look for the optimal step size (learning rate)  $s > 0$  by minimizing the function

$$\varphi(s) = f(x_{k+1}, y_{k+1}) = (1 - 2s)^2 x_k^2 + 3(1 - 6s)^2 y_k^2 \rightarrow \min.$$

We want

$$\varphi'(s) = 0 \text{ with } \varphi'(s) = -4(1 - 2s)x_k^2 - 36(1 - 6s)y_k^2,$$

hence we obtain the optimal step size  $s = \frac{x_k^2 + 9y_k^2}{2x_k^2 + 54y_k^2}$ . In [fig. 7](#) we have the gradient descent iterations starting from the initial value  $(3, 1)$  and converging towards the solution  $(0, 0)$ .  $\square$

## ❖ Higher order derivatives. Local extremum conditions. Optimization

The second order partial derivative with respect to  $x_i$  is simply

$$\frac{\partial}{\partial x_i} \left( \frac{\partial f}{\partial x_i} \right) =: \frac{\partial^2 f}{\partial x_i^2} =: \partial_i^2 f$$

and the mixed second order partial derivative is

$$\frac{\partial}{\partial x_i} \left( \frac{\partial f}{\partial x_j} \right) =: \frac{\partial^2 f}{\partial x_i \partial x_j} = \partial_{i,j}^2 f.$$

**Example 10.1.** For  $f(x, y) = x^2y + (x + 2y)^3$  we have that

$$\begin{aligned} \frac{\partial f}{\partial x} &= 2xy + 3(x + 2y)^2, & \frac{\partial^2 f}{\partial x^2} &= 2y, & \frac{\partial^2 f}{\partial y \partial x} &= 2x + 12(x + 2y) \\ \frac{\partial f}{\partial y} &= x^2 + 6(x + 2y)^2, & \frac{\partial^2 f}{\partial y^2} &= 24(x + 2y), & \frac{\partial^2 f}{\partial x \partial y} &= 2x + 12(x + 2y) \end{aligned}$$

**Theorem 10.2** (Schwarz). If  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  has continuous second order partial derivatives, then if  $i \neq j$

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

*Proof.* (Optional) See [3][page 151]. The proof is based on the mean value theorem.  $\square$

**Definition 10.3.** For  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  the *Hessian matrix* is defined by

$$H(x) = D^2f(x) = D(\nabla f)(x) = \begin{bmatrix} \nabla \left( \frac{\partial f}{\partial x_1} \right) \\ \nabla \left( \frac{\partial f}{\partial x_2} \right) \\ \vdots \\ \nabla \left( \frac{\partial f}{\partial x_n} \right) \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}_{n \times n}.$$

If the second order derivatives are continuous, then the Hessian matrix  $H(x)$  is symmetric.

**Theorem 10.4** (Taylor). Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  be of class  $C^2$  (twice differentiable, with continuous second order partial derivatives) and  $x_0 \in A$ . Then we have that

$$f(x) = f(x_0) + \nabla f(x_0) \cdot (x - x_0) + \frac{1}{2}(x - x_0)^T H(x_0)(x - x_0) + R(x - x_0),$$

with a remainder  $R(x - x_0)$  s.t.  $R(x - x_0)/\|x - x_0\|^2 \rightarrow 0$  as  $x \rightarrow x_0$ . Taylor's expansion can also be written as

$$f(x_0 + h) = f(x_0) + \nabla f(x_0) \cdot h + \frac{1}{2}h^T H(x_0)h + R(h).$$

Note that  $\nabla f(x_0) \cdot h$  is a linear function in  $h$  and  $\frac{1}{2}h^T H(x_0)h$  is a quadratic in  $h$ .

*Proof.* Consider the function  $g : [0, 1] \rightarrow \mathbb{R}$ ,  $g(t) = f(x_0 + th)$ . Using the classical 1d Taylor's expansion ([theorem 5.4](#))

$$g(1) = g(0) + g'(0) + \frac{1}{2}g''(0) + R_2, \quad (1)$$

where  $R_2$  is a remainder. Note that  $g(1) = f(x_0 + h)$  and  $g(0) = f(x_0)$ . Using the chain rule we have that

$$g'(t) = \nabla f(x_0 + th) \cdot h,$$

hence

$$g'(0) = \nabla f(x_0) \cdot h$$

and

$$\begin{aligned} g''(t) &= \frac{d}{dt} \left( \nabla f(x_0 + th) \cdot h \right) = \frac{d}{dt} \left( \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x_0 + th) h_i \right) = \sum_{i=1}^n h_i \frac{d}{dt} \left( \frac{\partial f}{\partial x_i}(x_0 + th) \right) \\ &= \sum_{i=1}^n h_i \underbrace{\left( \nabla \left( \frac{\partial f}{\partial x_i}(x_0 + th) \right) \cdot h \right)}_{(H(x_0 + th)h)_i} = h \cdot (H(x_0 + th)h) = h^T H(x_0 + th)h. \end{aligned}$$

This gives

$$g''(0) = h^T H(x_0)h$$

and from (1) we finally get that

$$f(x_0 + h) = f(x_0) + \nabla f(x_0) \cdot h + \frac{1}{2}h^T H(x_0)h + R(h).$$

□

### *Local extremum conditions*

**Corollary 10.5.** Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  be of class  $C^2$  with  $\nabla f(x_0) = 0$ . Then

- If  $h^T H(x_0)h > 0$ ,  $\forall h \in \mathbb{R}^n$ , then  $x_0$  is a local minimum.
- If  $h^T H(x_0)h < 0$ ,  $\forall h \in \mathbb{R}^n$ , then  $x_0$  is a local maximum.

*Proof.* Using that  $\nabla f(x_0) = 0$ , we have from Taylor's theorem that for small  $h$

$$f(x_0 + h) \approx f(x_0) + \frac{1}{2}h^T H(x_0)h.$$

If  $h^T H(x_0)h > 0$ , then  $f(x_0 + h) > f(x_0)$ , so  $x_0$  is a local minimum. The other case is similar. □

We see that Hessian matrices  $H$  for which the quadratic expression  $h^T H h$  is either positive or negative play a crucial role in determining if a critical point is a local minimum or maximum. We now recall the following related results from linear algebra.

**Definition 10.6.** An  $n \times n$  matrix  $A$  is called:

- *positive-definite* if  $x^T A x > 0, \forall x \in \mathbb{R}^n$ .
- *negative-definite* if  $x^T A x < 0, \forall x \in \mathbb{R}^n$ .
- *indefinite* if there exist  $x_1, x_2 \in \mathbb{R}^n$  s.t.  $x_1^T A x_1 > 0 > x_2^T A x_2$ .

**Proposition 10.7.** Let  $A$  be a symmetric  $n \times n$  matrix. Then

- $A$  is positive definite if and only if its eigenvalues are positive.
- $A$  is negative definite if and only if its eigenvalues are negative.
- $A$  is indefinite if and only if it has both positive and negative eigenvalues.

*Proof.* Can be found in any linear algebra textbook, see e.g. [5][Section I.7].  $\square$

**Theorem 10.8** (Local extremum conditions). Let  $f : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  be of class  $C^2$  and  $x \in A$  a critical point with  $\nabla f(x) = 0$ .

- If  $H(x)$  is positive definite (positive eigenvalues), then  $x$  is a local minimum.
- If  $H(x)$  is negative definite (negative eigenvalues), then  $x$  is a local maximum.
- If  $H(x)$  is indefinite, then  $x$  is called a saddle point.

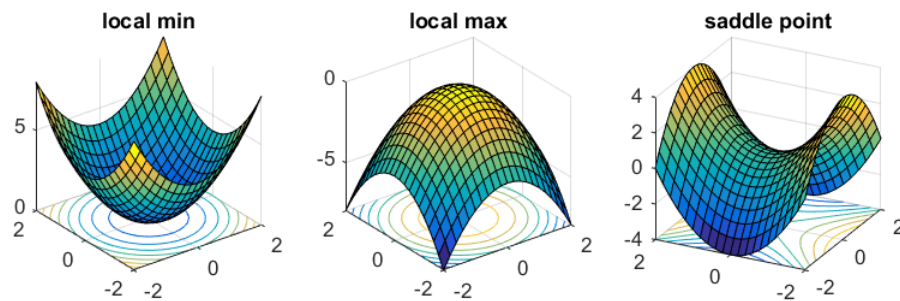


Figure 8: Examples of critical points: local minimum, local maximum and saddle point.

**Example 10.9.** The function  $f(x, y) = x^2 - y^2$  has a unique critical point because  $\nabla f(x, y) = (2x, -2y) = (0, 0)$  only at  $(0, 0)$ . This critical point is a saddle point since the Hessian matrix  $H(0, 0) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$  has eigenvalues 2 and  $-2$ . Note that a saddle point is a minimum in some directions, and a maximum in others.

**Applications. Linear regression.** Consider  $n$  data points  $(x_i, y_i)$ ,  $i = \overline{1, n}$ , with distinct  $x_i$ 's for which we want to find the line of best fit  $y = f(x) = \alpha x + \beta$  that minimizes the least squares error

$$E = \frac{1}{2} \sum_{i=1}^n |y_i - f(x_i)|^2 \rightarrow \min.$$

In other words, we are looking to find the optimal values  $\alpha, \beta$  that minimize the function

$$E = E(\alpha, \beta) = \frac{1}{2} \sum_{i=1}^n (\alpha x_i + \beta - y_i)^2 \rightarrow \min,$$

which means we have to look through the critical points

$$\begin{aligned} 0 = \frac{\partial E}{\partial \alpha} &= \sum_{i=1}^n x_i(\alpha x_i + \beta - y_i) \implies \alpha \sum_{i=1}^n x_i^2 + \beta \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \\ 0 = \frac{\partial E}{\partial \beta} &= \sum_{i=1}^n (\alpha x_i + \beta - y_i) \implies \alpha \sum_{i=1}^n x_i + \beta n = \sum_{i=1}^n y_i. \end{aligned}$$

These two simultaneous equations can be written as the linear system

$$\begin{bmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{bmatrix},$$

which has a unique solution  $\alpha, \beta$  since  $\det = n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2 > 0$  by the Cauchy-Schwarz inequality. Notice that the matrix of the linear systems is actually the Hessian matrix of  $E(\alpha, \beta)$  and that it is positive definite (since its trace and determinant are positive, the eigenvalues must be positive), which means that the critical point is indeed a minimum point.

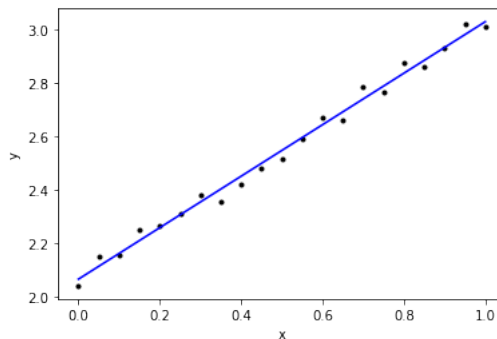


Figure 9: Example of linear regression.

**Optimization with constraints. Lagrange multipliers.** We have seen how to find the extremum points of a differentiable function by using [theorem 10.8](#) – we take the

critical points (where the gradient vanishes) and find if they are local minima/maxima by checking if the Hessian matrix is positive/negative definite.

Let us now consider the problem in which we want to optimize (minimize/maximize) a function and there is a constraint that must be satisfied, i.e.

$$\begin{aligned} &\text{optimize } f(x) \\ &\text{subject to } g(x) = c. \end{aligned}$$

We are thus looking for the minimum/maximum of  $f$  on the level set  $S = \{x \mid g(x) = c\}$ .

**Theorem 10.10.** Let  $f, g : A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  be differentiable functions. Let  $x_0 \in A$  be a solution to the problem

$$\begin{aligned} &\text{optimize } f(x) \\ &\text{subject to } g(x) = c. \end{aligned}$$

The solution  $x_0$  is a critical point of the so-called Lagrangian function

$$L(x, \lambda) = f(x) - \lambda(g(x) - c),$$

meaning that there exists  $\lambda \in \mathbb{R}$  (Lagrange multiplier) s.t.  $\nabla f(x_0) = \lambda \nabla g(x_0)$ .

*Proof.* Consider the level set  $S = \{x \in A \mid g(x) = c\}$  and an arbitrary path  $c(t)$  in  $S$  with  $c(0) = x_0$  and  $c'(0) = v$ . Since the gradient is perpendicular to the level set (see [proposition 9.19](#)), we have that  $\nabla g(x_0) \cdot v = 0$ . Since  $x_0$  is a local extremum of  $f$  along a path  $c(t)$ , we have that  $\frac{d}{dt}f(c(t)) \Big|_{t=0} = 0$  we have by the chain rule that

$$0 = \frac{d}{dt}f(c(t)) \Big|_{t=0} = \nabla f(x_0) \cdot v,$$

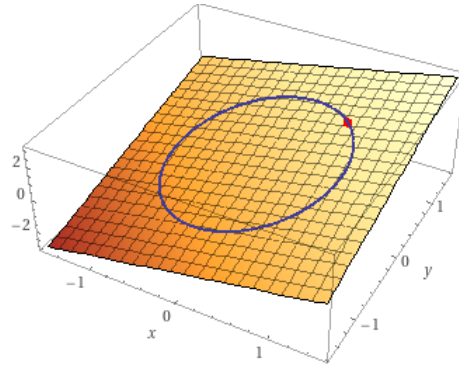
hence  $\nabla f(x_0) \cdot v = \nabla g(x_0) \cdot v$  for an arbitrary direction  $v$ , which means that there exists  $\lambda \in \mathbb{R}$  s.t.  $\nabla f(x_0) = \lambda \nabla g(x_0)$ .  $\square$

The idea for constrained optimization is thus to consider the Lagrangian function

$$L(x, \lambda) = f(x) - \lambda(g(x) - c)$$

and its critical points for which  $\nabla_x L = 0$  and  $\frac{\partial L}{\partial \lambda} = 0$ . This gives a system  $(n + 1)$  equations and  $(n + 1)$  unknowns.

**Example 10.11.** Maximize  $f(x, y) = x + y$  subject to the constraint  $x^2 + y^2 = 1$ . This means finding the highest point on the plane  $z = x + y$  subject to  $(x, y)$  being on the unit circle.

Figure 10: Sketch for [example 10.11](#).

Consider the Lagrangian  $L(x, y, \lambda) = f(x, y) + \lambda(x^2 + y^2 - 1) = x + y + \lambda(x^2 + y^2 - 1)$  and look for the critical points:

$$\begin{aligned} 0 &= \frac{\partial L}{\partial x} = 1 + 2\lambda x, \\ 0 &= \frac{\partial L}{\partial y} = 1 + 2\lambda y, \\ 0 &= \frac{\partial L}{\partial \lambda} = x^2 + y^2 - 1. \end{aligned}$$

From the first two equations we get that  $x = y = -\frac{1}{2\lambda}$  and from the third one that  $\lambda^2 = \frac{1}{2}$ . Hence  $\lambda = \pm \frac{1}{\sqrt{2}}$  and the critical points of  $L$  are  $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, -\frac{1}{\sqrt{2}})$  and  $(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, \frac{1}{\sqrt{2}})$ . By evaluation the function  $f$  at the  $xy$ -coordinates of the two critical points, we see that the first one is a maximum and the second one is a minimum.  $\square$



## References

- [1] W. Rudin, *Principles of Mathematical Analysis 3rd ed*, McGraw-Hill, 1976.
- [2] T. Tao, *Analysis I*, Springer, 2016.
- [3] J.E. Marsden, A. Tromba, *Vector Calculus 6th ed*, W.H. Freeman and Company, 2012
- [4] M. Oberguggenberger, A. Ostermann, *Analysis for Computer Scientists*, Springer, 2018
- [5] G. Strang, *Linear Algebra and Learning from Data*, Wellesley Cambridge Press, 2019