Mathematical Analysis

1st Year Computer Science

Mihai Nechita *

Last updated: January 28, 2023

Contents

*ictp.acad.ro/nechita. If you find any typos, please let me know on Teams or at mihai.nechita@ubbcluj.ro

※ Real numbers

Let us start with some standard notation: $\emptyset$ is the empty set; $\mathbb{N} = \{1, 2, \ldots\}$ the set of natural numbers; $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\} = \{m - n \mid m, n \in \mathbb{N}\}$ 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 1.3, 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

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

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

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

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

These are denoted by $\min(A)$ and $\max(A)$.

Note that there are sets which do no 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 (Completeness Axiom)**.** Every set $A \subseteq \mathbb{R}$ that is bounded above has a *least upper bound*, called the *supremum* of A and denoted by $\sup(A)$. Similarly, every set $A \subseteq \mathbb{R}$ that is bounded below has a *greatest lower bound*, called the *infimum* of A and denoted by $\inf(A)$. In other words, we have

$$\sup(A) := \min(ub(A))$$

and

$$\inf(A) := \max(lb(A)).$$

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

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

Definition 1.5. Define the *extended real line* $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$, where ∞ 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 $\inf(A) := -\infty$.

Note that the empty set $\emptyset$ is bounded by any real number and $\sup(\emptyset) = -\infty$, $\inf(\emptyset) = \infty$.

Proposition 1.6. 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. TBA (left to the reader). □

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, $\sup(A)$ is the least upper bound of A. In other words, $\sup(A)$ is an upper bound for A and any number less than $\sup(A)$ is not an upper bound for A. This means that 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 $y = \sup(A) - \varepsilon < x$. The claim for $\inf(A)$ follows similarly. □

Definition 1.8. 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 ∞ if

$$\exists a \in \mathbb{R} \text{ such that } (a, \infty) \subseteq V.$$

A set $V \subseteq \mathbb{R}$ is a *neighborhood* of $-\infty$ if

$$\exists a \in \mathbb{R} \text{ such that } (-\infty, a) \subseteq V.$$

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

Definition 1.9. Let $A \subseteq \mathbb{R}$. The following set is called the *interior* of A

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

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

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

Proposition 1.10. For any $A \subseteq \mathbb{R}$, it holds that $\mathrm{int}(A) \subseteq A \subseteq \mathrm{cl}(A)$.

Proof. To prove that $\mathrm{int}(A) \subseteq A$ we prove that if $x \in \mathrm{int}(A)$, then $x \in A$. Let $x \in \mathrm{int}(A)$, 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 $A \subseteq \mathrm{cl}(A)$ we show that if $x \in A$, then $x \in \mathrm{cl}(A)$. 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 $x \in \mathrm{cl}(A)$ since $V \cap A \neq \emptyset$. □

Definition 1.11. If $A = \mathrm{int}(A)$, then A is called *open*. If $A = \mathrm{cl}(A)$, then A is called *closed*.

Remark 1.12. To prove that a set A is open, it is sufficient to prove that $A \subseteq \mathrm{int}(A)$.
To prove that a set A is closed, it is sufficient to prove that $\mathrm{cl}(A) \subseteq A$.

Proposition 1.13. 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. $A = \mathrm{int}(A)$, 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 $\mathrm{cl}(A^c) \subseteq A^c$. Consider $x \in \mathrm{cl}(A^c)$ 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 $x \in \mathrm{cl}(A^c)$. Hence the assumption $x \notin A^c$ is false, and we have that if $x \in \mathrm{cl}(A^c)$, then $x \in A^c$. In other words, $\mathrm{cl}(A^c) \subseteq A^c$. □

Proposition 1.14. Any union of open sets is open. Any finite intersection of closed sets is closed.

Proof. (Optional) Left to the reader. □

❊ 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 either increasing or decreasing.

Definition 2.1. A sequence (x_n) has a limit $\ell \in \overline{\mathbb{R}}$, and we write $\lim\limits_{n\to\infty} x_n = \ell$ or $x_n \to \ell$, 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 ℓ: $\forall \varepsilon > 0$, $\exists N_\varepsilon \in \mathbb{N}$ such that $|x_n - \ell| < \varepsilon$, $\forall n \geq N_\varepsilon$.

Proposition 2.2. A sequence (x_n) converges to ℓ if and only if $\lim\limits_{n\to\infty} |x_n - \ell| = 0$.

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 $\sup(S) - x_n \leq \varepsilon$, $\forall n \geq N_\varepsilon$. The sequence converges to its supremum. Similarly, a decreasing and bounded sequence converges to its infimum. □

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

Theorem 2.6 (Squeeze 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

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

Then

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

Proof. Let $\ell := \lim_{n\to\infty} x_n = \lim_{n\to\infty} z_n$ and 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

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

□

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 $\lim_{n\to\infty}(b_n - a_n) = 0$, 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 $\bigcap_{n=1}^{\infty} I_n$ contains only the element $\sup(A)$. □

Definition 2.8. For a sequence (x_n) we define the set of its *limit points* by

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

and

$$\liminf_{n\to\infty} x_n := \inf\big(\mathrm{LIM}(x_n)\big),$$
$$\limsup_{n\to\infty} x_n := \sup\big(\mathrm{LIM}(x_n)\big).$$

Proposition 2.9. $\lim_{n\to\infty} x_n = \ell \in \overline{\mathbb{R}}$ if and only if $\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n = \ell$.

Theorem 2.10 (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 $b_k - a_k \to 0$.

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. □

Definition 2.11 (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 2.12. 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. □

Theorem 2.13. A sequence is convergent if and only if it is Cauchy.

Proof. Let's consider first a convergent sequence (x_n) with $x_n \to \ell$. 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}), $x_{n_k} \to x \in \mathbb{R}$. 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 $x_n \to x$. □

Example 2.14. The sequence defined by $x_n = 1 + \frac{1}{2} + \ldots + \frac{1}{n}$ is not convergent. Indeed, one can see, for example, that

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

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 of real numbers

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. Let the partial sum $s_n = 1 + 1 + \frac{1}{2!} + \ldots + \frac{1}{n!}$. Start from $\lim_{n\to\infty} (1 + \frac{1}{n})^n = e$ and expand

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

We have that

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

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

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

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

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

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 $\lim_{n\to\infty} x_n = 0$.

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 $\lim_{n\to\infty} x_n \neq 0$, then the series $\sum_{n=1}^{\infty} x_n$ is divergent.

Example 3.6. Series like $\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \sum_{n=1}^{\infty} (\frac{1}{n} - \frac{1}{n+1}) = 1$ 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) are 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. □

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

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

- 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

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

- 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.

Proof. The idea is that $\sum_{n\geq 1} x_n$ behaves like a geometric series with ratio ℓ. 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 $N_\varepsilon \in \mathbb{N}$ such that

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

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

Note that the Ratio test is *inconclusive* when $\ell = 1$.

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

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

- 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.

Proof. The idea is that $\sum_{n\geq 1} x_n$ behaves like $\sum_{n\geq 1} \ell^n$. Like in the ratio test, the proof uses the comparison test with a geometric series. Left to the reader. □

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 $\frac{x_{n+1}}{x_n} = \frac{x}{n+1} \to 0 < 1$, 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 $\sum_{n=0}^{\infty} 2^n \cdot x_{2^n}$ have the same nature.

Proof. TBA (given during the lectures). □

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 $\sum_{n=1}^{\infty} \frac{2^n}{2^{np}} = \sum_{n=1}^{\infty} \left(2^{1-p}\right)^n$, which converges if and only if $2^{1-p} < 1$, i.e for $p > 1$. □

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

(a) If
$$\lim_{n\to\infty} \left(c_n \frac{x_n}{x_{n+1}} - c_{n+1}\right) > 0,$$
then $\sum_{n\geq 1} x_n$ is convergent.

(b) If $\sum_{n\geq 1} \frac{1}{c_n} = \infty$ and
$$\lim_{n\to\infty} \left(c_n \frac{x_n}{x_{n+1}} - c_{n+1}\right) < 0,$$
then $\sum_{n\geq 1} x_n$ is divergent.

Proof. TBA (given during the lectures). □

Theorem 3.16 (Raabe-Duhamel)**.** Let $\sum_{n\geq 1} x_n$ be a series with positive terms such that
$$\lim_{n\to\infty} n\left(\frac{x_n}{x_{n+1}} - 1\right) = R.$$

- 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). □

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

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

$$\lim_{n\to\infty} n\Big(\frac{x_n}{x_{n+1}} - 1\Big) = \lim_{n\to\infty} n\Big(\frac{a+n+1}{n+1} - 1\Big) = a.$$

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

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 $\sum_{n\geq 1} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots$ 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 seminar) that

$$\lim_{n\to\infty} 1 + \frac{1}{2} + \ldots + \frac{1}{n} - \ln n = \gamma \in (0,1).$$

Hence

$$\begin{aligned} s_{2n} &= 1 - \frac{1}{2} + \frac{1}{3} - \ldots - \frac{1}{2n} \\ &= 1 + \frac{1}{2} + \ldots + \frac{1}{2n} - 2\Big(\frac{1}{2} + \ldots + \frac{1}{2n}\Big) - \ln(2n) \\ &= \underbrace{1 + \frac{1}{2} + \ldots + \frac{1}{2n} - \ln(2n)}_{\to\gamma} - \underbrace{\Big(1 + \frac{1}{2} + \ldots + \frac{1}{n} - \ln n\Big)}_{\to\gamma} + \ln 2 \to \ln 2. \end{aligned}$$

□

Theorem 3.19 (Leibniz test)**.** Let (x_n) be a decreasing sequence with $x_n \to 0$. 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

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

hence (s_n) is a Cauchy sequence since $|s_{n+p} - s_n|$ can be made arbitrarily small. □

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. □

Theorem 3.22 (Cauchy). Let $\sum_{n\geq 1} x_n$ be an *absolutely convergent series* and let $\sigma : \mathbb{N} \to \mathbb{N}$ 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]. □

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 $\sigma : \mathbb{N} \to \mathbb{N}$ 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]. □

※ 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

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

We denote by A' the set of all the accumulation points of A.

We say that $x_0 \in A$ is an *isolated point* if $x_0 \in A \setminus A'$.

Proposition 4.2. 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 $\lim\limits_{n\to\infty} x_n = x_0$.

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 $x_n \longrightarrow x_0$. If x_0 is infinite, the neighborhoods can be taken $(-\infty, -n)$ or (n, ∞), respectively.

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

Example 4.3. For $A = \{\frac{1}{n} \mid n \in \mathbb{N}\}$, each element $x \in A$ in an isolated point and $A' = \{0\}$.

Definition 4.4. Let $A \subseteq \mathbb{R}$, $f : A \longrightarrow \mathbb{R}$ and $x_0 \in A'$. We say that $\lim\limits_{x\to x_0} f(x) = \ell \in \overline{\mathbb{R}}$ if

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

Remark 4.5 (ε–δ). Let $A \subseteq \mathbb{R}$, $f : A \longrightarrow \mathbb{R}$ and $x_0 \in A'$ finite. If $\lim\limits_{x\to x_0} f(x) = \ell \in \mathbb{R}$, 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 4.6. Let $A \subseteq \mathbb{R}$, $f : A \longrightarrow \mathbb{R}$ and $x_0 \in A'$. Then $\lim\limits_{x\to x_0} f(x) = \ell \in \overline{\mathbb{R}}$ iff

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

Theorem 4.7. Let $A \subseteq \mathbb{R}$, $f : A \longrightarrow \mathbb{R}$, $x_0 \in \mathbb{R}$ s.t. $x_0 \in \big(A \cap (-\infty, x_0)\big)'$ and $x_0 \in \big(A \cap (x_0, \infty)\big)'$. Then

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

Example 4.8. (a) $\operatorname{sgn} : \mathbb{R} \longrightarrow \mathbb{R}$, $\operatorname{sgn}(x) = \begin{cases} +1, & \text{if } x \geq 0 \\ -1, & \text{if } x < 0. \end{cases}$ has no limit at 0.

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

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

Definition 4.9. Let $A \subseteq \mathbb{R}$, $f : A \to \mathbb{R}$ 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 4.10. If $x_0 \in A \cap A'$ is an accumulation point, then f is continuous at x_0 if

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

Remark 4.11. 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.

Definition 4.12. For $f : A \to \mathbb{R}$ 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. $\inf\big(f(A)\big), \sup\big(f(A)\big)$ are finite.

Theorem 4.13. Let $A \subseteq \mathbb{R}$, $f : A \to \mathbb{R}$ 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 $\lim\limits_{n \to \infty} x_n = x_0$, we have that $\lim\limits_{n \to \infty} f(x_n) = f(x_0)$.

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

Theorem 4.15 (Weierstrass)**.** Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then f is bounded and it attains its bounds, i.e. there exist $\min\big(f(A)\big), \max\big(f(A)\big)$.

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 2.10; denote its limit by x. We have that $x_{n_k} \to x$ and f is continuous, hence $f(x_{n_k}) \to f(x)$. But $|f(x_{n_k})| > n_k \to \infty$, 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 $f(x_M) = \sup\big(f(A)\big)$, i.e. $f(x_M) = \max\big(f(A)\big) = \sup\big(f(A)\big)$. From theorem 1.7, we obtain a sequence (x_n) in $[a, b]$ such that $f(x_n) \to \sup\big(f(A)\big)$. 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 $f(x_{n_k}) \to f(x_M)$, but we know that $f(x_{n_k}) \to \sup\big(f(A)\big)$, hence $f(x_M) = \sup\big(f(A)\big)$ and f reaches its upper bound. $\square$

Theorem 4.16 (Intermediate value property)**.** Let $f : [a,b] \to \mathbb{R}$ 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 $f(c) = y$. □

Definition 4.17. Let $A \subseteq \mathbb{R}$, $f : A \to \mathbb{R}$ and $x_0 \in A \cap A'$. The *derivative* of f at x_0 is

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

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

Remark 4.18. $f'(x_0)$ 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 4.19. Let $A \subseteq \mathbb{R}$, $f : A \to \mathbb{R}$ 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 $\lim\limits_{x \to x_0} f(x) = f(x_0) + 0 = f(x_0)$. □

Example 4.20. $f : \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$ is not differentiable in 0 since $\nexists \lim\limits_{x \to 0} \dfrac{|x|}{x}$.

Theorem 4.21 (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)
- $\left(fg\right)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$. (Quotient Rule)
- $(f \circ g)'(x) = f'\big(g(x)\big)g'(x)$. (Chain Rule)

Definition 4.22. $f : A \to \mathbb{R}$ has a local extremum (minimum/maximum) at $x_0 \in A$ if

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

Theorem 4.23 (Fermat)**.** Let $f : (a, b) \to \mathbb{R}$ 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 ≥ 0, the other ≤ 0. Hence $f'(x_0) = 0$. □

Theorem 4.24 (Rolle)**.** Let $f : (a, b) \to \mathbb{R}$ with $f(a) = f(b)$. If 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). □

Theorem 4.25 (Mean value theorem)**.** Let $f : (a, b) \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). Then there exists $c \in (a, b)$ s.t.

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

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

Theorem 4.26. Let $f : (a, b) \to \mathbb{R}$ 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. □

Proposition 4.27 (l'Hôpital's rule)**.** Let I be an open interval, $x_0 \in \overline{\mathbb{R}}$ and $f, g : I \setminus \{x_0\} \to \mathbb{R}$ differentiable. If $\lim\limits_{x\to x_0} f(x) = \lim\limits_{x\to x_0} g(x) = 0$ or $\lim\limits_{x\to x_0} g(x) = \pm\infty$, and $\lim\limits_{x\to x_0} \dfrac{f'(x)}{g'(x)} \in \overline{\mathbb{R}}$, then

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

※ Taylor series and power series

Let $I \subseteq \mathbb{R}$ be an interval, $f : I \to \mathbb{R}$ and $x_0 \in I$ a point where f is differentiable n times ($n \in \mathbb{N}$). Does there exist a polynomial $P : \mathbb{R} \to \mathbb{R}$ 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

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

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

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

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

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

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, $f : I \to \mathbb{R}$ and $x_0 \in I$ a point where f is differentiable n times. The polynomial $T_n : \mathbb{R} \to \mathbb{R}$,

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

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 $f : \mathbb{R} \to \mathbb{R}$, $f(x) = e^x$ and $x_0 = 0$. Then $f(0) = f'(0) = \ldots = f^{(n)}(0) = 1$ and
$$T_n(x) = 1 + x + \frac{x^2}{2} + \ldots + \frac{x^n}{n!}.$$

Definition 5.3. Let $I \subseteq \mathbb{R}$ be an interval, $f : I \to \mathbb{R}$ and $x_0 \in I$ a point where f is differentiable n times. We define $R_n : \mathbb{R} \to \mathbb{R}$,
$$R_n(x) := f(x) - T_n(x)$$
to be the remainder when approximating f by T_n around x_0. Note that (Taylor's formula)
$$f(x) = T_n(x) + R_n(x).$$

Theorem 5.4 (Taylor-Lagrange)**.** Let $I \subseteq \mathbb{R}$ be an interval, $f : I \to \mathbb{R}$ 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
$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \ldots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \frac{f^{(n+1)}(c)}{(n+1)!}(x - x_0)^{n+1}.$$
In other words, $f(x) = T_n(x) + R_n(x)$ with
$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x - x_0)^{n+1}$$
being called *the remainder in Lagrange's form*. □

There are several other forms of the remainder, but we will mostly only use this one. Its main advantage is that assuming that the $(n+1)^{th}$ derivative of f is bounded by $M > 0$,
$$|f(x) - T_n(x)| = |R_n(x)| \leq \frac{M}{(n+1)!}|x - x_0|^{n+1},$$
and we see that if this holds for any n, then $|R_n(x)| \to 0$ as $n \to \infty$.

Corollary 5.5 (Local optimality conditions)**.** Let $I \subseteq \mathbb{R}$ be an open interval, $f : I \to \mathbb{R}$ and $x_0 \in I$ a point where f is differentiable n times and
$$f'(x_0) = f''(x_0) = \ldots = f^{(n-1)}(x_0) = 0 \text{ and } f^{(n)}(x_0) \neq 0.$$

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.

Example 5.6 (Convex/concave)**.** Let $f : I \to R$ 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.

Definition 5.7. Let $I \subseteq \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ 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

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

then f can be expanded as a Taylor series around x_0 (also called a Taylor expansion). □

Note that the partial sum of a Taylor series is simply the Taylor polynomial $T_n(x)$, and that a Taylor series converges to $f(x)$ if and only if the remainder $R_n(x) \to 0$ as $n \to \infty$.

Example 5.8. Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = e^x$ and $x_0 = 0$. Then $f(0) = f'(0) = \ldots = f^{(n)}(0) = 1$ and

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

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

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

Since $\frac{|x|^n}{n!} \to 0$ as $n \to \infty$, it follows that e^x can be expanded as a Taylor series around 0:

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

Example 5.9. 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

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

Definition 5.10. 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.

Definition 5.11. The convergence set of a power series is

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

Theorem 5.12. 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 $|x-c| < R$.
- the power series diverges when $|x-c| > R$.

Remark 5.13. The convergence set C contains the open interval $(c-R, c+R)$ and possibly the endpoints $\{c-R, c+R\}$.

Example 5.14. The power series $\sum_{n\geq 1} \frac{x^n}{n}$ converges absolutely for $|x| < 1$ and diverges when $|x| > 1$ (by the ratio test), hence its radius of convergence is $R = 1$. 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 5.15. 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

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

Theorem 5.16. If the limit

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

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 5.17. If the limit

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

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

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

※ Riemann integrals. Improper integrals

[to be added]

※ The Euclidean space $\mathbb{R}^n$

Elements in R^n are vectors with n components. We will write $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ 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 $[\quad]_{m\times n}$ with vectors $[\quad]_{n\times 1}$ to get vectors $[\quad]_{m\times 1}$.

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 $x + y := (x_1 + y_1, \ldots, x_n + y_n)$, and a vector can be multiplied by a scalar $\alpha \in \mathbb{R}$ to get $\alpha x := (\alpha x_1, \ldots, \alpha x_n)$. We will denote by e_i the canonical basis vector with a 1 in the ith component and 0's everywhere else, giving $x = x_1 e_1 + \ldots x_n e_n$.

Definition 7.1. A map $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ 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) $\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle x, z\rangle$, for any $x, y, z \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$.

(c) $\langle x, x\rangle > 0$, for any $x \in \mathbb{R}^n \setminus \{0\}$.

Definition 7.2. The *dot product* of two vectors $x, y \in \mathbb{R}^n$ is given by

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

Note that the dot product is a scalar product. It is the most important scalar product.

In matrix notation, the dot product is written as

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

Definition 7.3. Two vectors $x, y \in \mathbb{R}^n$ are perpendicular (or orthogonal) if $x \cdot y = 0$.

Definition 7.4. A function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ 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).

Remark 7.5. Any scalar product generates a norm on $\mathbb{R}^n$ given by $\|x\| = \sqrt{\langle x, x\rangle}$.

Definition 7.6. The Euclidean norm is generated by the dot product and it is given by

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

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

Theorem 7.7. 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 $x\cdot y = \|x\|\|y\| \cos \angle(x, y)$. ◻

Theorem 7.8 (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. See Seminar 8. ◻

Definition 7.9. A function $d : \mathbb{R}^n \times \mathbb{R}^n \to [0, \infty)$ 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).

Remark 7.10. Any norm generates a distance on $\mathbb{R}^n$ given by $d(x, y) = \|x - y\|$.

Definition 7.11. The Euclidean distance is generated by the Euclidean norm and it is given by

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

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

Neighborhoods. Interior. Closure. Boundary.

Definition 7.12. 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

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

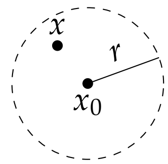


Figure 1: Open ball $B(x_0, r)$.

Definition 7.13. 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 7.14. Let $A \subseteq \mathbb{R}^n$. The following set is called the *interior* of A

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

the following set is called the *closure* of A

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

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

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

Example 7.15. Let $A = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$. Then

$$\begin{aligned} \operatorname{int}(A) &= A, \\ \operatorname{cl}(A) &= \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}, \\ \operatorname{bd}(A) &= \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}. \end{aligned}$$

Definition 7.16. If $A = \operatorname{int}(A)$, then A is called *open*. If $A = \operatorname{cl}(A)$, then A is called *closed*.

Proposition 7.17. For any $A \subseteq \mathbb{R}^n$, it holds that $\operatorname{int}(A) \subseteq A \subseteq \operatorname{cl}(A)$.

Proof. Similar to proposition 1.10. □

Remark 7.18. To prove that a set A is open, it is sufficient to prove that $A \subseteq \operatorname{int}(A)$.
To prove that a set A is closed, it is sufficient to prove that $\operatorname{cl}(A) \subseteq A$.

Proposition 7.19. 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 1.13. □

Definition 7.20. A set $A \subseteq \mathbb{R}^n$ is called *bounded* if there exists $r > 0$ such that

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

Sequences.

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

Definition 7.21. A sequence (x^k) converges to $x \in \mathbb{R}^n$ if $\lim\limits_{k\to\infty} \|x^k - x\| = 0$. We write $\lim\limits_{k\to\infty} x^k = x$.

Example 7.22. Let $x^k = (\frac{1}{k}, \frac{1}{k^2})$, then $\lim\limits_{k\to\infty} x^k = (0, 0)$.

Theorem 7.23. A sequence (x^k) converges to $x \in \mathbb{R}^n$ if and only if $\lim\limits_{k\to\infty} x_i^k = x_i$, $\forall i = \overline{1, n}$.

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

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

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

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

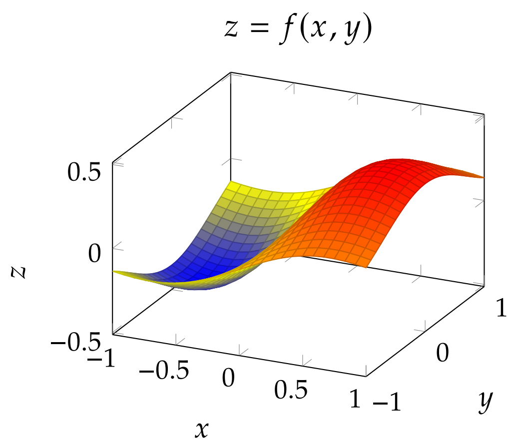
$$\|x^k - x\| = \sqrt{(x_1^k - x_1)^2 + \ldots + (x_n^k - x_n)^2} \to 0,$$

hence $x^k \to x$. □

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 $f : A \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ mapping vectors in $\mathbb{R}^n$ into real numbers. As you already now, when $n = 1$ the graph of a function is a curve in $\mathbb{R}^2$. When $n = 2$, the graph of a function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ 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

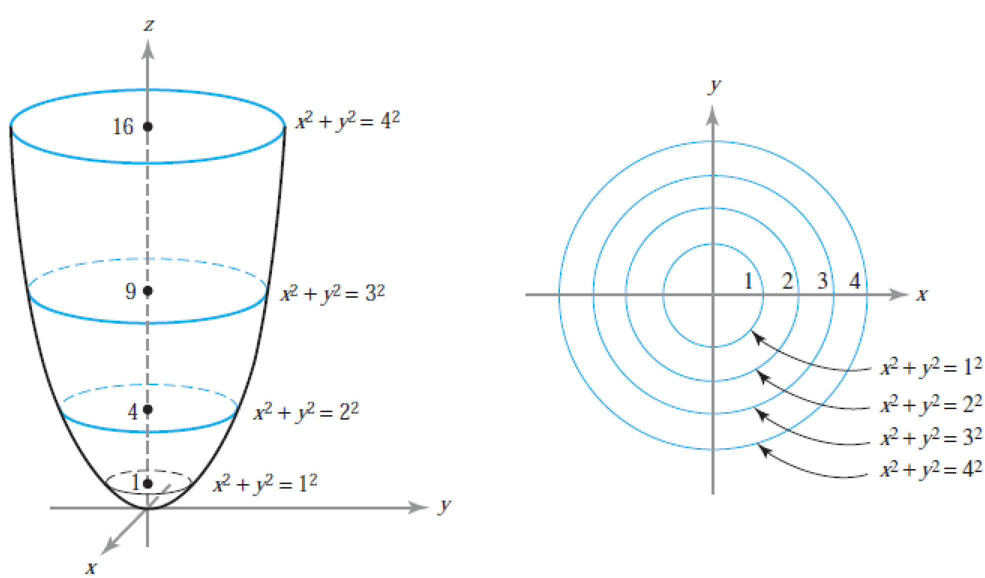


Figure 2: Graph and level curves for $f(x, y) = x^2 + y^2$. Source: [3, page 80].

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 $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ and $x_0 \in A'$. We say that $\lim\limits_{x \to x_0} f(x) = \ell \in \overline{\mathbb{R}}$ if

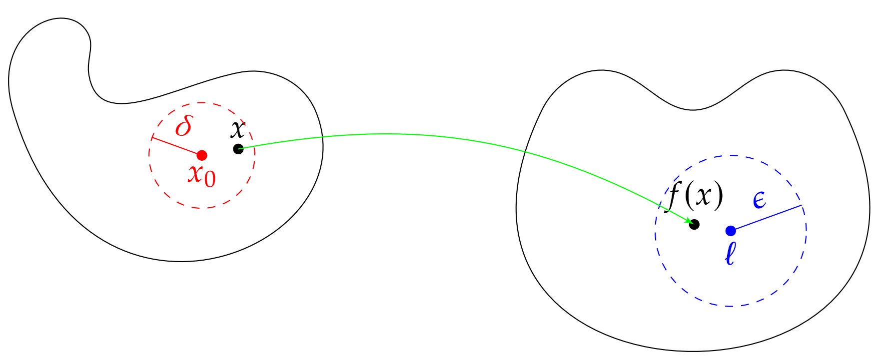
$$\forall V \in \mathcal{V}(\ell),\ \exists U \in \mathcal{V}(x_0) \text{ s.t. } f(x) \in V,\ \forall x \in U \cap \big(A \setminus \{x_0\}\big).$$

Note that $U \in \mathcal{V}(x_0)$ is a set in $\mathbb{R}^n$.

Remark 8.2. Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ and $x_0 \in A'$. We say that $\lim\limits_{x \to x_0} f(x) = \ell \in \mathbb{R}$ 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 $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$, illustrated below. □



Theorem 8.3. Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ and $x_0 \in A'$. Then $\lim\limits_{x \to x_0} f(x) = \ell \in \overline{\mathbb{R}}$ iff

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

□

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

Example 8.4. (a) $\lim\limits_{(x,y)\to(0,0)} \dfrac{x^2}{\sqrt{x^2+y^2}} = 0.$

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

(b) $\lim\limits_{(x,y)\to(0,0)} \dfrac{\sin(x^2+y^2)}{x^2+y^2} = 1.$

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

(c) $\lim\limits_{(x,y)\to(0,0)} \dfrac{xy}{x^2+y^2}$ does not exist.

As $(x,y) \to (0,0)$ 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, $\lim\limits_{(x,mx)\to(0,0)} \dfrac{xy}{x^2+y^2} = \lim\limits_{(x,mx)\to(0,0)} \dfrac{mx^2}{(1+m^2)x^2} = \dfrac{m}{1+m^2}$. 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 $\lim\limits_{(x,y)\to(x_0,y_0)} f(x,y)$ exists, then $\lim\limits_{x\to x_0} \lim\limits_{y\to y_0} f(x,y) = \lim\limits_{y\to y_0} \lim\limits_{x\to x_0} f(x,y)$.

Definition 8.6. Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ 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 $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ and $x_0 \in A \cap A'$ an accumulation point. Then f is continuous at x_0 if

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

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

Theorem 8.9. Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ 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 $\lim\limits_{n\to\infty} x^k = x_0$, we have that $\lim\limits_{k\to\infty} f(x^k) = f(x_0)$.

Example 8.10 (Any norm is continuous)**.** $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \|x\|$ is continuous on $\mathbb{R}^n$. Let $x_0 \in \mathbb{R}^n$ and (x^k) with $x^k \to x_0$. We have that $\|x^k - x^0\| \to 0$. By the triangle inequality $\big|\|x^k\| - \|x^0\|\big| \leq \|x^k - x^0\| \to 0$, hence $\|x^k\| - \|x^0\| \to 0$, i.e $\|x^k\| \to \|x^0\|$.

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

※ Partial derivatives and differentiability in $\mathbb{R}^n$

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

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

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 $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ that has partial derivatives at $x \in A$ with respect to all its variables, the *gradient* at x is given by the vector

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

Example 9.3. For $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = x^2y + y^2$ 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 (Having partial derivatives but discontinuous)**.** Let $f : \mathbb{R}^2 \to \mathbb{R}$,

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

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

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

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

which can also be written as

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

where $R(x - x_0)$ is a remainder.

Definition 9.5. Let $A \subseteq \mathbb{R}^n$ be an open set and $f : A \to \mathbb{R}$. 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.

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

This can also be written as

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

Note that differentiability is equivalent to

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

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

Definition 9.6. Let $A \subseteq \mathbb{R}^n$ be an open set. If $f : A \to \mathbb{R}^m$, $f = (f_1, \ldots, f_m)$, 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.

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

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

Theorem 9.7. Let $A \subseteq \mathbb{R}^n$ be an open set. If $f : A \to \mathbb{R}^m$ 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 $x \to x_0$ we use that $\frac{\|R(x-x_0)\|}{\|x-x_0\|} \to 0$ to obtain that $f(x) \to f(x_0)$. □

Constant functions have zero derivative and linear functions have a constant derivative.

- If $f : \mathbb{R}^n \to \mathbb{R}^m$ is constant, then $Df(x) = 0$ since $f(x) = f(x_0)$ for any $x, x_0 \in \mathbb{R}^n$.
- If $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = a \cdot x$ with $a \in \mathbb{R}^n$, then $Df(x) = a$ since $f(x) - f(x_0) - a \cdot (x - x_0) = 0$.
- If $f : \mathbb{R}^n \to \mathbb{R}^m$, $f(x) = Ax$ 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 and $x \in A$. If $f : A \to \mathbb{R}$ is differentiable at x, then the partial derivatives exist at x and

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

Proof. Differentiability at x gives

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

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

$$\lim_{h_i\to 0} \frac{f(x+h_ie_i) - f(x) - Df(x)\cdot h_ie_i}{h_i} = 0,$$

which gives that

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

hence the ith component of $Df(x)$ is $\frac{\partial f}{\partial x_i}(x)$, which is the ith component of $\nabla f(x)$. □

Theorem 9.9. Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ 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]. □

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 9.10. Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$, $f = (f_1,\ldots,f_m)$ be differentiable at $x \in A$, then

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

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

Theorem 9.11 (Calculus rules)**.** Let $A \subseteq \mathbb{R}^n$ and $f,g : A \to \mathbb{R}$ 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. $\nabla\left(\frac{f}{g}\right)(x) = \dfrac{g(x)\nabla f(x) - f(x)\nabla g(x)}{g^2(x)}$.

Theorem 9.12 (Chain rule)**.** Let $g : \mathbb{R}^n \to \mathbb{R}^m$, $f : \mathbb{R}^m \to \mathbb{R}^p$ differentiable at x and $g(x)$, respectively. Then

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

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

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

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

using the triangle inequality we have that

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

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

Example 9.13 (Chain rule)**.** Let $g : \mathbb{R} \to \mathbb{R}^n$, $g = (g_1, \ldots, g_n)$, $f : \mathbb{R}^n \to \mathbb{R}$, $f \circ g : \mathbb{R} \to \mathbb{R}$. With $g'(t) = (g_1'(t), \ldots, g_n'(t))$, we have that

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

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 9.14. Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ 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

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

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

Theorem 9.15. Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ 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

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

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

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

which can be rearranged as

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

□

Proposition 9.16 (Direction of steepest ascent/descent)**.** Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ 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 $v = -\alpha\nabla f(x)$, $\alpha > 0$. □

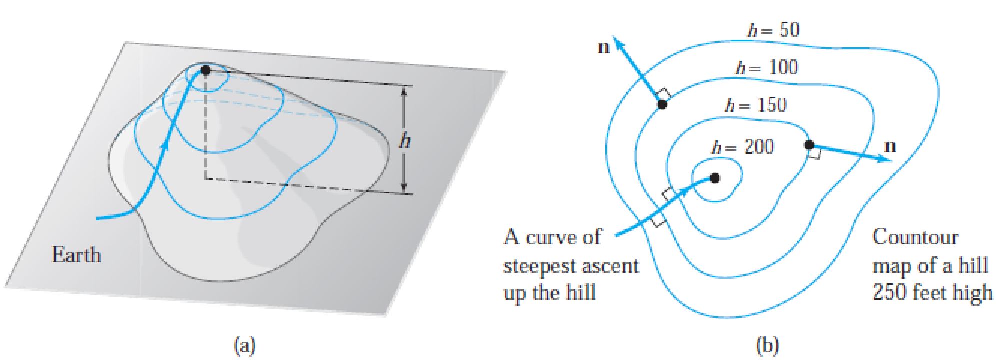


Figure 3: The gradient gives the direction of steepest ascent and is perpendicular to the level curves. Source: [3, page 140].

Proposition 9.17 (Gradient perpendicular to the level set)**.** Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ be differentiable at $x \in A$. Then $\nabla f(x)$ is perpendicular to the level set passing 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. $f(c(t)) =$ 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

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

□

Example 9.18 (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 9.19 (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

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

Theorem 9.20 (Fermat)**.** Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ 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 $\nabla f(x) = 0$. □

※ Higher order derivatives. Local extremum conditions. Applications

The second order partial derivative with respect to x_i is simply

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

and the mixed second order partial derivative is

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

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

$$\begin{aligned} \frac{\partial f}{\partial x} &= y + 2(x+2y), & \frac{\partial^2 f}{\partial x^2} &= 2, & \frac{\partial^2 f}{\partial y \partial x} &= 5 \\ \frac{\partial f}{\partial y} &= x + 4(x+2y), & \frac{\partial^2 f}{\partial y^2} &= 8, & \frac{\partial^2 f}{\partial x \partial y} &= 5 \end{aligned}$$

Theorem 10.2 (Schwarz)**.** If $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ 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. □

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

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

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

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

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

with a remainder $R(x-x_0)$ s.t. $R(x-x_0)/\|x-x_0\|^2 \to 0$ as $x \to x_0$. Note that $\nabla f(x_0)\cdot(x-x_0)$ is linear in $x - x_0$ and $\frac{1}{2}(x - x_0)^T H(x_0)(x - x_0)$ is quadratic.

Proof. Let us first rewrite Taylor's expansion as

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

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

$$g(1) = g(0) + g'(0) + \tfrac{1}{2} g''(0) + R_2, \tag{1}$$

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

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

This gives

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

and from (1) we finally get that

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

□

Local extremum conditions

Corollary 10.5. Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ 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

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

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 $h^T Hh$ 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 $x^T Ax > 0$, $\forall x \in R^n$.
- *negative-definite* if $x^T Ax < 0$, $\forall x \in R^n$.
- *indefinite* if there exist $x_1, x_2 \in \mathbb{R}^n$ s.t. $x_1^T Ax_1 > 0 > x_2^T Ax_2$.

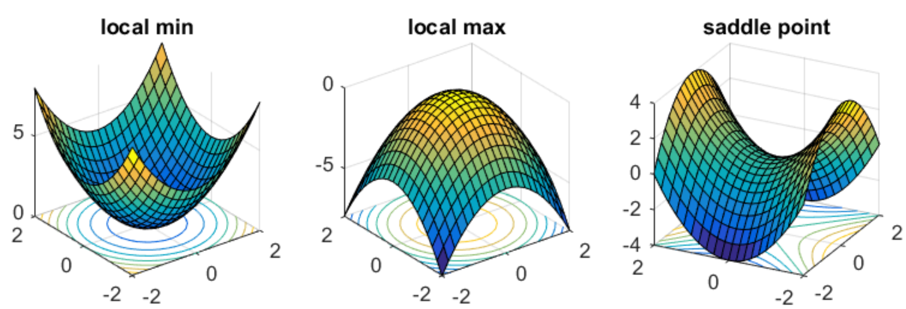
Proposition 10.7. Let A be a symmetric $n \times n$ matrix. Then

- 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]. □

Theorem 10.8 (Local extremum conditions)**.** Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ 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 (minimum in some directions, maximum in others).



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.

***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

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

In other words, we are looking to find the optimal values α, β that minimize the function

$$E = E(\alpha, \beta) = \tfrac{1}{2} \sum_{i=1}^{n} \left(\alpha x_i + \beta - y_i\right)^2 \longrightarrow \min,$$

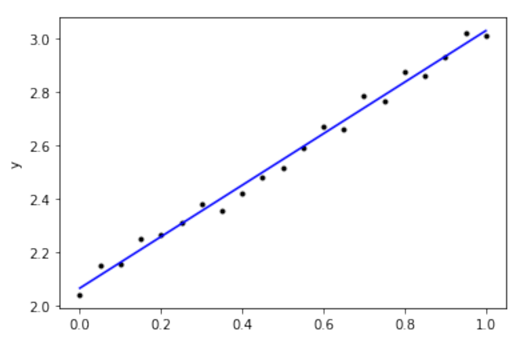
which means we have to look through the critical points

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

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 α, β since $\det = n \sum_{i=1}^{n} x_i^2 > -\left(\sum_{i=1}^{n} x_i\right)^2 > 0$ 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.



***Applications. 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 $f : \mathbb{R}^n \to \mathbb{R}$ is given by $-\nabla f$ (see proposition 9.16). If we want to minimize a function, 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} with $f(x_{k+1}) < f(x_k)$. 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 (learning rate) s_k has to be chosen at every iteration. One way of doing this is called *exact line search*: the optimal step size s_k is given by minimizing the function $\varphi(s) = f(x_{k+1}) = f(x_k - s\nabla f(x_k))$. By the chain rule we have that

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

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. $(x_{k+2} - x_{k+1}) \perp (x_{k+1} - x_k)$.

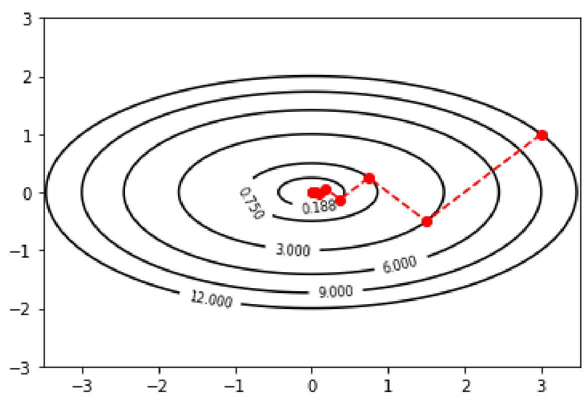


Figure 4: Level curves and gradient descent iterations for example 10.10.

Example 10.10. 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

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

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

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

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 $s = \frac{x_k^2+9y_k^2}{2x_k^2+54y_k^2}$. The fig. 4 shows the gradient descent iterations starting from the initial value $(3, 1)$ converging towards the solution $(0, 0)$. □

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.

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

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

Theorem 10.11. Let $f, g : A \subseteq \mathbb{R}^n \to \mathbb{R}$ be differentiable functions. Let $x_0 \in A$ be a solution to the problem

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

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

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 9.17), 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 $g(c(t)) = 0$ we have by the chain rule that

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

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. $\nabla f(x_0) = \lambda g(x_0)$. □

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.

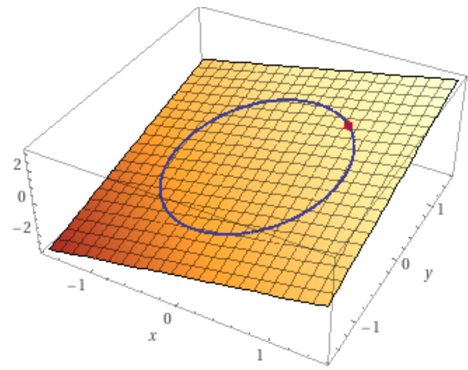


Figure 5: Sketch for example 10.12.

Example 10.12. 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. 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:

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

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. □

※ Double integrals

Rectangular domains. We start by defining double integrals on very simple domains – rectangles.

Let $A = [a,b]\times[c,d]$ be a rectangle and consider a partition of it into smaller rectangles

$$\mathcal{P} = \{A_{ij} \mid A_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j],\ i = \overline{1,m},\ j = \overline{1,n}\},$$

where $a = x_0 < x_1 < \ldots < x_m = b$ and $c = y_0 < y_1 < \ldots < y_n = d$. The norm of the partition is given by $\|\mathcal{P}\| := \max\left\{\max_{i=\overline{1,m}}\{x_i - x_{i-1}\}, \max_{j=\overline{1,n}}\{y_i - y_{i-1}\}\right\}$. We also consider a set of intermediate points $(x^*_{ij}, y^*_{ij}) \in A_{ij}$ attached to the partition $\mathcal{P}$.

Definition 11.1. For a function $f : A \to \mathbb{R}$ and a partition $\mathcal{P}$, the Riemann sum is given by

$$\sigma(f,P) := \sum_{i=1}^{m}\sum_{j=1}^{n} f(x^*_{ij}, y^*_{ij})(x_i - x_{i-1})(y_j - y_{j-1}).$$

Remark 11.2. The Riemann sum collects the volumes of the parallelepipeds defined by the partition $\mathcal{P}$ (and the intermediate points). In the limit one obtains the volume of the solid below the surface $z = f(x,y)$.

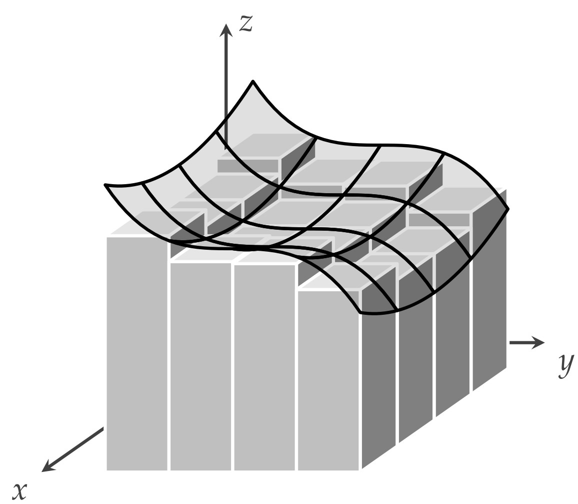


Figure 6: Volume of a solid approximated through parallelepipeds.

Definition 11.3. Let $A = [a,b]\times[c,d]$ and $f : A \to \mathbb{R}$. We say that f is Riemann integrable if there exists $I \in \mathbb{R}$ s.t. for any partition $\mathcal{P}$ of A the Riemann sum $\sigma(f,P)$ converges to I as $\|\mathcal{P}\| \to 0$, i.e.

$$\lim_{\|\mathcal{P}\|\to 0} \sigma(f,P) = I =: \iint\limits_A f(x,y)\,\mathrm{d}x\mathrm{d}y.$$

Proposition 11.4. Let $f, g : A \to \mathbb{R}$ be Riemann integrable and $\alpha \in \mathbb{R}$. Then

- $\iint\limits_A \alpha f(x,y)\,\mathrm{d}x\mathrm{d}y = \alpha \iint\limits_A f(x,y)\,\mathrm{d}x\mathrm{d}y$.
- $f + g$ is Riemann integrable and
$$\iint\limits_A \big(f(x,y) + g(x,y)\big)\,\mathrm{d}x\mathrm{d}y = \iint\limits_A f(x,y)\,\mathrm{d}x\mathrm{d}y + \iint\limits_A g(x,y)\,\mathrm{d}x\mathrm{d}y.$$
- If $f(x,y) \leq g(x,y)$, $\forall (x,y) \in A$, then $\iint\limits_A f(x,y)\,\mathrm{d}x\mathrm{d}y \leq \iint\limits_A g(x,y)\,\mathrm{d}x\mathrm{d}y$.

Proposition 11.5. Let $A_1, A_2 \subset \mathbb{R}^2$ s.t. $A = A_1 \cup A_2$ and $\mathrm{int}A_1 \cap \mathrm{int}A_2 = \emptyset$. Then
$$\iint\limits_A f(x,y)\,\mathrm{d}x\mathrm{d}y = \iint\limits_{A_1} f(x,y)\,\mathrm{d}x\mathrm{d}y + \iint\limits_{A_2} f(x,y)\,\mathrm{d}x\mathrm{d}y.$$

The next fundamental result indicates *how to compute* double integrals: by interated integrals. Note that the order of integration does not matter.

Theorem 11.6 (Fubini)**.** Let $A = [a,b] \times [c,d]$ and $f : A \to \mathbb{R}$ be Riemann integrable. Then
$$\iint\limits_A f(x,y)\,\mathrm{d}x\mathrm{d}y = \int_a^b \int_c^d f(x,y)\,\mathrm{d}y\mathrm{d}x = \int_c^d \int_a^b f(x,y)\,\mathrm{d}x\mathrm{d}y.$$

Proof. We will only give a simplified proof, for which we assume that f is continuous. We consider the iterated integral and we aim to prove that it equals the double integral. Integrating first with respect to y, we define
$$F(x) := \int_c^d f(x,y)\,\mathrm{d}y = \sum_{j=1}^{n} \int_{y_{j-1}}^{y_j} f(x,y)\,\mathrm{d}y,$$
where $c = y_0 < y_1 < \ldots < y_n = d$. Since f is continuous we can apply the mean value theorem for each integral
$$\int_{y_{j-1}}^{y_j} f(x,y)\,\mathrm{d}y = f(x, y_j^*)(y_j - y_{j-1}), \text{ with } y_j^* \in [y_{j-1}, y_j]$$
and obtain that
$$F(x) = \sum_{j=1}^{n} f(x, y_j^*)(y_j - y_{j-1}).$$

Considering now the whole iterated integral and using the usual 1d Riemann sum

$$\int_a^b \int_c^d f(x,y)\,\mathrm{d}y\mathrm{d}x = \int_a^b F(x)\,\mathrm{d}x = \lim_{n\to\infty} \sum_{i=1}^{n} F(x_i^*)(x_i - x_{i-1}),$$

where $a = x_0 < x_1 < \ldots < x_n = b$ and $x_i^* \in [x_i - x_{i-1}]$. We have that

$$\begin{aligned}\int_a^b \int_c^d f(x,y)\,\mathrm{d}y\mathrm{d}x &= \int_a^b F(x)\,\mathrm{d}x = \lim_{n\to\infty} \sum_{i=1}^{n} F(x_i^*)(x_i - x_{i-1}) \\ &= \lim_{n\to\infty} \sum_{i=1}^{n}\sum_{j=1}^{n} f(x_j^*, y_j^*)(y_j - y_{j-1})(x_i - x_{i-1}) \\ &= \iint\limits_A f(x,y)\,\mathrm{d}x\mathrm{d}y,\end{aligned}$$

where in the end we used the 2d Riemann sum that converges to the double integral. By a similar reasoning we also have that $\int_c^d \int_a^b f(x,y)\,\mathrm{d}x\mathrm{d}y = \iint\limits_A f(x,y)\,\mathrm{d}x\mathrm{d}y$. □

Example 11.7. Let $R = [-1,1]\times[0,1]$ and consider $\iint\limits_R (x^2+y^2)\,\mathrm{d}x\mathrm{d}y$. By Fubini's theorem

$$\iint\limits_R (x^2+y^2)\,\mathrm{d}x\mathrm{d}y = \int_{-1}^1 \int_0^1 (x^2+y^2)\,\mathrm{d}y\mathrm{d}x = \int_{-1}^1 (x^2+\frac{1}{3})\mathrm{d}x = \frac{4}{3}.$$

There is a particular kind of functions for which the double integral is very easy to compute, namely separable functions $f(x,y) = g(x)h(y)$. In this case the double integral is simply the product of two separate simple integrals.

Corollary 11.8. Let $f : A = [a,b]\times[c,d] \to \mathbb{R}$ be Riemann integrable. If $f(x,y) = g(x)h(y)$, then

$$\iint\limits_A f(x,y)\,\mathrm{d}x\mathrm{d}y = \int_a^b g(x)\,\mathrm{d}x \int_c^d h(y)\,\mathrm{d}y.$$

Proof. By Fubini's theorem

$$\iint\limits_A g(x)h(y)\,\mathrm{d}x\mathrm{d}y = \int_a^b g(x) \underbrace{\int_c^d h(y)\,\mathrm{d}y}_{\text{constant}}\,\mathrm{d}x = \Big(\int_a^b g(x)\,\mathrm{d}x\Big)\Big(\int_c^d h(y)\,\mathrm{d}y\Big).$$

□

More general domains. Based on the definition of the double integral on rectangles, let us now define the double integral on more general domains. For this let $D \subset \mathbb{R}^2$ be a bounded set.

Definition 11.9. We say that $f : D \to \mathbb{R}$ is Riemann integrable on D if there exists a rectangle $A \subset \mathbb{R}^2$ s.t. $D \subseteq A$ and the extension function $\overline{f} : A \to \mathbb{R} = \begin{cases} f(x), & \text{if } x \in D \\ 0, & \text{if } x \in A \setminus D \end{cases}$ is Riemann integrable on A. Then $\iint\limits_D f(x,y)\,\mathrm{d}x\mathrm{d}y = \iint\limits_A \overline{f}(x,y)\,\mathrm{d}x\mathrm{d}y$.

An important class of domains consists of domains that have four sides – two straight and two curves – which can be described by letting one variable run in an interval and bounding the other variable by two functions. Such domains are called *simple* and the double integral can be computed through iterated integrals using Fubini's theorem.

Definition 11.10. A set $D \subset \mathbb{R}^2$ is called

- *simple with respect to the* y-axis if there exist continuous functions $\varphi_1, \varphi_2 : [a,b] \to \mathbb{R}$
$$D = \{(x,y) \in \mathbb{R}^2 \mid a \le x \le b,\ \varphi_1(x) \le y \le \varphi_2(x)\}.$$
- *simple with respect to the* x-axis if if there exist continuous functions $\psi_1, \psi_2 : [c,d] \to \mathbb{R}$
$$D = \{(x,y) \in \mathbb{R}^2 \mid c \le y \le d,\ \psi_1(y) \le x \le \psi_2(y)\}.$$
- *simple* if it is simple with respect to both x-axis and y-axis.

Domains that are simple w.r.t to the y-axis or x-axis are also called y-simple or x-simple.

Figure 7: Simple domains

Theorem 11.11. Let $D \subset \mathbb{R}^2$ be a bounded set and $f : D \to \mathbb{R}$ Riemann integrable on D.

- If D is y-simple, then
$$\iint\limits_D f(x,y)\,\mathrm{d}x\mathrm{d}y = \int_a^b \int_{\varphi_1(x)}^{\varphi_2(x)} f(x,y)\,\mathrm{d}y\mathrm{d}x.$$
- If D is x-simple, then
$$\iint\limits_D f(x,y)\,\mathrm{d}x\mathrm{d}y = \int_c^d \int_{\psi_1(y)}^{\psi_2(y)} f(x,y)\,\mathrm{d}x\mathrm{d}y.$$

- If D is simple, then

$$\iint\limits_D f(x,y)\,\mathrm{d}x\mathrm{d}y = \int_a^b \int_{\varphi_1(x)}^{\varphi_2(x)} f(x,y)\,\mathrm{d}y\mathrm{d}x = \int_c^d \int_{\psi_1(y)}^{\psi_2(y)} f(x,y)\,\mathrm{d}x\mathrm{d}y.$$

  Note that simple domains allow changing the order of integration.

Example 11.12. Let $D = \{(x,y) \in \mathbb{R}^2 \mid 0 \le x \le 1,\ x \le y \le \sqrt{x}\}$. Compute $\iint\limits_D \frac{e^y}{y}\,\mathrm{d}x\mathrm{d}y$. Notice that the domain D is y-simple and we could try to compute the integral in an iterated way (Fubini) – first w.r.t to y, then w.r.t x. However, this order of integration puts us in the situation of finding a primitive of $\frac{e^y}{y}$ and we get stuck – there is no *elementary* primitive for this function. So let us try to change the order of integration. For this, we write the domain D as x-simple: $D = \{(x,y) \in \mathbb{R}^2 \mid 0 \le y \le 1,\ y^2 \le x \le \sqrt{y}\}$. Then

$$\iint\limits_D \frac{e^y}{y}\,\mathrm{d}x\mathrm{d}y = \int_0^1 \int_{y^2}^{y} \frac{e^y}{y}\,\mathrm{d}x\mathrm{d}y = \int_0^1 \frac{e^y}{y}(y - y^2)\,\mathrm{d}y = \int_0^1 (e^y - ye^y)\,\mathrm{d}y.$$

Changing the order of integration worked – the new integral is much easier to compute! Integrating by parts we finally get $e - 2$.

Figure 8: Simple domain. Changing the order of integration.

Change of variables in double integrals

Linear algebra recap. Let us start by recalling some basic things from linear algebra and the geometry of linear transformations. This will provide the fundamental tools for doing a change of variables (or change of coordinates).

Let A be a 2×2 matrix and consider a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$, $T(x) = Ax$. A linear transformation maps parallelograms into parallelograms. Indeed, consider a parallelogram with edges u and v. This is given by linear combinations $\alpha u + \beta v$ with $\alpha, \beta \in [0,1]$. Applying the linear transformation T, we have that $A(\alpha u + \beta v) = \alpha Au + \beta Av$. This represents a parallelogram with edges Au and Av. So parallelograms $\longmapsto$ parallelograms.

How does a linear transformation change area? It scales it with the factor $|\det(A)|$.

The unit square with area 1 is mapped into a parallelogram with area $|\det(A)|$. For a domain D mapped into $T(D)$, $\mathrm{Area}\big(T(D)\big) = |\det(A)|\mathrm{Area}\big(D\big)$.

Figure 9: Linear transformation. The determinant gives the area scaling factor.

Change of variables. Let $D \subseteq \mathbb{R}^2$ be a domain in the xy plane. We want to make a change of variables from xy to uv, writing $x = x(u,v)$ and $y = y(u,v)$. Consider a domain

D^* in the uv plane and a map $T : D^* \to D$ bijective and of class C^1 (differentiable and with continuous derivatives) such that $(x, y) = T(u, v)$. The question is: how are Area(D) and Area(D^*) related as we change coordinates from xy to uv?

$$\iint_D 1 \, dx dy = \iint_{D^*} .?. \, du dv.$$

As discussed above, if T is a linear map with $T(x) = Ax$ then Area(D) = $|\det(A)|$Area(D^*), which can be written as

$$\iint_D 1 \, dx dy = \iint_{D^*} |\det(A)| \, du dv.$$

In T is not linear, we can take its linear approximation (the differential) around a point and for $u = u_0 + \Delta u$ and $v = v_0 + \Delta v$ use that

$$T(u, v) \approx T(u_0, v_0) + T'(u_0, v_0) \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix},$$

where T' is the differential of T, namely the Jacobian matrix

$$T' = J = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix}.$$

Consider a partition of the domain D^* in the uv plane into small rectangles R^* with sides $\Delta u, \Delta v$. Area(R^*) = $\Delta u \Delta v$. The image $R = T(R^*)$ of a rectangle R^* can be a complicated domain with curved boundaries – we will aim to *approximate* its area. By taking the linear approximation of T, we approximate the region $R = T(R^*)$ with a parallelogram $T'(R^*)$ given by the linear transformation T'. The sides of the parallelogram are $T' \begin{bmatrix} \Delta u \\ 0 \end{bmatrix} = \Delta u \begin{bmatrix} \frac{\partial x}{\partial u} \\ \frac{\partial y}{\partial u} \end{bmatrix}$ and $T' \begin{bmatrix} 0 \\ \Delta v \end{bmatrix} = \Delta v \begin{bmatrix} \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial v} \end{bmatrix}$. More importantly, the area of the parallelogram $T'(R^*)$ approximates the area of the region $R = T(R^*)$, namely

$$\text{Area}(R) \approx \text{Area}\big(T'(R^*)\big) = |\det(J)|\text{Area}(R^*) = |\det(J)|\Delta u \Delta v,$$

where $\det(J)$ is the determinant of the Jacobian matrix J. As the size of the rectangle R^* goes to zero, in the limit we have that

$$\text{Area}(D) = \iint_D 1 \, dx dy = \iint_{D^*} |\det(J)| \, du dv. \tag{2}$$

Theorem 11.13 (Change of variables)**.** Let $D, D^* \subseteq \mathbb{R}^2$ and $T : D^* \to D$ bijective and of class C^1 with the Jacobian J. Then for any Riemann integrable $f : D \to \mathbb{R}$, we have that

$$\iint_D f(x,y)\,\mathrm{d}x\mathrm{d}y = \iint_{D^*} f(x(u,v), y(u,v))|\det(J)|\,\mathrm{d}u\mathrm{d}v.$$

Polar coordinates. One of the most important changes of coordinates is given by

$$x = r\cos\theta,$$
$$y = r\sin\theta,$$

having the Jacobian matrix

$$J = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix} = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}$$

and the Jacobian determinant $|\det(J)| = |r\cos^2\theta + r\sin^2\theta| = r$, that is

$$|\det(J)| = r.$$

Changing from a domain D in the xyplane to a domain D^* in the $r\theta$ plane, we have that

$$\iint_D f(x,y)\,\mathrm{d}x\mathrm{d}y = \iint_{D^*} f(r\cos\theta, r\sin\theta)\, r\,\mathrm{d}r\mathrm{d}\theta. \tag{3}$$

Example 11.14. Let $D = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \le R^2\}$. Compute $\iint_D e^{x^2+y^2}\,\mathrm{d}x\mathrm{d}y$.

Using the polar coordinates (r,θ) we write $x = r\cos\theta, y = r\sin\theta$ and $(x,y) \in D$ for $r \in [0,R], \theta \in [0,2\pi]$. This gives a rectangle $D^* = [0,R] \times [0,2\pi]$ in the $r\theta$ plane.

$$\iint_D e^{x^2+y^2}\,\mathrm{d}x\mathrm{d}y = \int_0^R \int_0^{2\pi} e^{r^2} r\,\mathrm{d}\theta\mathrm{d}r = 2\pi \int_0^R re^{r^2}\,\mathrm{d}r = \pi\big(e^{R^2} - 1\big).$$

Example 11.15 (Gaussian integral $\int_{-\infty}^{\infty} e^{-x^2}\,\mathrm{d}x$)**.** To compute this integral it is easier to consider a double integral. Let $D = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \le R^2\}$.

$$\iint_D e^{-(x^2+y^2)}\,\mathrm{d}x\mathrm{d}y = \int_0^R \int_0^{2\pi} e^{-r^2} r\,\mathrm{d}\theta\mathrm{d}r = 2\pi \int_0^R re^{-r^2}\,\mathrm{d}r = \pi\big(-e^{-R^2} + 1\big).$$

Letting now $R \to \infty$ we obtain that

$$\pi = \iint_{\mathbb{R}^2} e^{-(x^2+y^2)}\,\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{\infty} e^{-x^2}\,\mathrm{d}x \int_{-\infty}^{\infty} e^{-y^2}\,\mathrm{d}y = \left(\int_{-\infty}^{\infty} e^{-x^2}\,\mathrm{d}x\right)^2,$$

hence we obtain the Gaussian integral

$$\int_{-\infty}^{\infty} e^{-x^2}\,\mathrm{d}x = \sqrt{\pi}.$$

※ Triple integrals

The triple Riemann integral can be defined using Riemann sums, in a similar way to the double integral (with parallelepipeds instead of rectangles). Everything that we have seen about double integrals can be generalized to triple integrals. We will simply state the basic results.

Let $A = [a_1, a_2] \times [b_1, b_2] \times [c_1, c_2]$ and $f : A \to \mathbb{R}$ be Riemann integrable. Then

$$\begin{aligned}\iiint\limits_A f(x,y,z)\,\mathrm{d}x\mathrm{d}y\mathrm{d}z &= \int_{a_1}^{a_2}\int_{b_1}^{b_2}\int_{c_1}^{c_2} f(x,y,z)\,\mathrm{d}z\mathrm{d}y\mathrm{d}x\\ &= \int_{c_1}^{c_2}\int_{b_1}^{b_2}\int_{a_1}^{a_2} f(x,y,z)\,\mathrm{d}x\mathrm{d}y\mathrm{d}z\\ &= \ldots\text{(all possible orders)}.\end{aligned}$$

If $D = \{(x,y,z) \in \mathbb{R}^3 \mid a \le x \le b,\ \varphi_1(x) \le y \le \varphi_2(x),\ \psi_1(x,y) \le z \le \psi_2(x,y)\}$, then

$$\iiint\limits_D f(x,y,z)\,\mathrm{d}x\mathrm{d}y\mathrm{d}z = \int_a^b \int_{\varphi_1(x)}^{\varphi_2(x)} \int_{\psi_1(x,y)}^{\psi_2(x,y)} f(x,y,z)\,\mathrm{d}z\mathrm{d}y\mathrm{d}x.$$

In symmetric elementary regions the order of integration can be changed.

When changing variables in triple integrals, we should now have in mind that 3×3 matrices map parallelepipeds to parallelepipeds with the *volume* scaling with the determinant (absolute value). Let $D, D^* \subseteq \mathbb{R}^3$ and $T : D^* \to D$ bijective and of class C^1 with the Jacobian J. Then for any Riemann integrable $f : D \to \mathbb{R}$, we have that

$$\iiint\limits_D f(x,y,z)\,\mathrm{d}x\mathrm{d}y\mathrm{d}z = \iiint\limits_{D^*} f(u,v,w)|\det(J)|\,\mathrm{d}u\mathrm{d}v\mathrm{d}w,$$

with the Jacobian matrix

$$J = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w}\\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w}\\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w}\end{bmatrix}$$

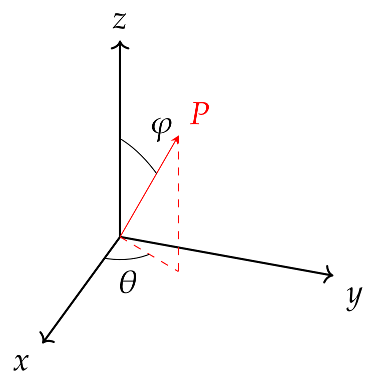
Cylindrical coordinates.

$$\begin{aligned} x &= r\cos\theta,\\ y &= r\sin\theta,\\ z &= z,\end{aligned}$$

having the Jacobian matrix

$$J = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial z} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial w} \end{bmatrix} = \begin{bmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and the Jacobian determinant $|\det(J)| = |r\cos^2\theta + r\sin^2\theta| = r$.
Spherical coordinates.



$$x = (r\sin\varphi)\cos\theta,$$
$$y = (r\sin\varphi)\sin\theta,$$
$$z = r\cos\varphi,$$

having the Jacobian matrix

$$J = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \varphi} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \varphi} \end{bmatrix} = \begin{bmatrix} \sin\varphi\cos\theta & -r\sin\varphi\sin\theta & r\cos\varphi\cos\theta \\ \sin\varphi\sin\theta & r\sin\varphi\cos\theta & r\cos\varphi\sin\theta \\ \cos\varphi & 0 & -r\sin\varphi \end{bmatrix}$$

and the Jacobian determinant

$$|\det(J)| = r^2\sin\varphi.$$

The ball $B(0,R)$ is given by $r \in [0,R], \theta \in [0,2\pi], \varphi \in [0,\pi]$ in spherical coordinates.

Example 12.1. The volume of the ball of radius R is given by

$$\iiint\limits_{B(0,R)} 1\,\mathrm{d}x\mathrm{d}y\mathrm{d}z = \int_0^R \int_0^{2\pi} \int_0^{\pi} r^2 \sin\varphi\,\mathrm{d}\varphi\mathrm{d}\theta\mathrm{d}r = \frac{R^3}{3}2\pi \underbrace{\int_0^{\pi} \sin\varphi\,\mathrm{d}\varphi}_{=2} = \frac{4\pi R^3}{3}.$$

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