# Analysis-3 lecture schemes
## (with Homeworks)[1]

Csörgő István

November, 2015

[1]A jegyzet az ELTE Informatikai Kar 2015. évi Jegyzetpályázatának támogatásával készült

# Contents

# 1. Lesson 1

## 1.1. The Space $\mathbb{R}^n$

In this book $\mathbb{N}$ denotes the set of positive integers and $\mathbb{N}_0$ denotes the set of nonnegative integers:

$$\mathbb{N} = \{1,\, 2,\, 3,\, \dots\} \qquad \text{and} \qquad \mathbb{N}_0 = \{0,\, 1,\, 2,\, 3,\, \dots\} = \mathbb{N} \cup \{0\}\,.$$

In Linear Algebra we have studied the vector spaces and their special type, the Euclidean spaces. As you remember, $\mathbb{R}^n$ was an important example for real Euclidean space. So every definition and theorem in connection with Euclidean spaces is valid for $\mathbb{R}^n$.

Why $\mathbb{R}^n$ is important in the multivariable analysis? „Multivariable" means, that a multivariable function has a finite number of real variables - say $n$ variables. Thus its domain can be regarded as a collection of ordered $n$-tuples, and forms a subset of $\mathbb{R}^n$. The Reader can consider that how connects the case $n = 1$ to the one-variable analysis studied in the subjects Analysis-1 and Analysis-2.

We review shortly the most important properties of $\mathbb{R}^n$.

For a fixed $n \in \mathbb{N}$ $\mathbb{R}^n$ is the set of all possible ordered $n$-tuples whose terms (components) are in $\mathbb{R}$:

$$\mathbb{R}^n := \{x = (x_1,\, x_2,\, \dots,\, x_n) \mid x_i \in \mathbb{R}\}.$$

Notice that in the case $n = 2$ the notation $(x,\, y)$ is often used instead of $(x_1,\, x_2)$. Similarly in the case $n = 3$ the notation $(x,\, y,\, z)$ may be used instead of $(x_1,\, x_2,\, x_3)$.

We have the following operations in $\mathbb{R}^n$:
Let $x,\, y \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$.

- Addition: $x + y := (x_1 + y_1,\, x_2 + y_2,\, \dots,\, x_n + y_n) \quad (x,\, y \in \mathbb{R}^n)$;
- Scalar Multiplication: $\lambda x := (\lambda x_1,\, \lambda x_2,\, \dots,\, \lambda x_n) \quad (x \in \mathbb{R}^n,\, \lambda \in \mathbb{R})$;
- Scalar Product $\langle x,\, y\rangle := x_1y_1 + x_2y_2 + \dots + x_ny_n = \sum\limits_{i=1}^{n} x_iy_i \quad (x,\, y \in \mathbb{R}^n)$;
- Norm (length):

$$\|x\| := \sqrt{\langle x,\, x\rangle} = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} = \sqrt{\sum_{i=1}^{n} x_i^2} \quad (x \in \mathbb{R}^n);$$

We have learnt the properties of the above operations in Linear Algebra, and we have proved that $\mathbb{R}^n$ is a real Euclidean Space. Consequently it is a Normed Vector Space (see Linear Algebra).

We have defined the distance in $\mathbb{R}^n$ as follows

$$d(x,\,y) := \|x-y\| = \sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\ldots+(x_n-y_n)^2} =$$
$$= \sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \quad (x,\,y\in\mathbb{R}^n)\,.$$

**1.1. Theorem** *[the properties of the distance] Let $X$ be a linear normed space with norm $\|.\|$, especially $X=\mathbb{R}^n$ with the above defined norm. Then*

1. $d(x,\,y)\ge 0 \quad (x,\,y\in X)$. *Furthermore* $d(x,\,y)=0 \Leftrightarrow x=y$
2. $d(x,\,y)=d(y,\,x) \qquad (x,\,y\in X)$
3. $d(x,\,y)\le d(x,\,z)+d(z,\,y) \qquad (x,\,y,\,z\in X)$ *(triangle inequality)*

**Proof.** The first and the second statements are obvious by the axioms of the norm. Let us prove the triangle inequality:

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

□

**1.2. Remark.** If we define a mapping $d : V\times V \to \mathbb{R}$ which satisfies the above properties on a nonempty set $X$, then $X$ is called metric space and the above properties are called the axioms of the metric space. So we have proved that every linear normed space is a metric space with the metric indicated by the norm $d(x,y)=\|x-y\|$.

A lot of properties of $\mathbb{R}^n$ and of functions defined on $\mathbb{R}^n$ are based on the metric structure of $\mathbb{R}^n$, so we could describe them using the notation of the metric: $d(x,y)$. But for simplicity we will use the normed space structure, so the distance will be denoted by $\|x-y\|$ instead of $d(x,y)$. The Reader can consider that a lot of the definitions and statements can be generalized for any metric space.

Let us review some important relations which can be deduced from the Euclidean structure of $\mathbb{R}^n$ (see: Linear Algebra):

1. Cauchy's inequality. For any elements $x,y$ of an Euclidean space holds

$$|\langle x,\,y\rangle|\le\|x\|\cdot\|y\|\,.$$

Here the equality holds if and only if the vectors $x$, $y$ are linearly dependent (parallel). Especially in $\mathbb{R}^n$ it is called Cauchy-Bunyakovsky-Schwarz inequality:

$$(x_1y_1+\cdots+x_ny_n)^2\le(x_1^2+\cdots+x_n^2)\cdot(y_1^2+\cdots+y_n^2) \qquad (x_i,\,y_i\in\mathbb{R})$$

and equality holds if and only if the vectors $(x_1,\ldots,x_n)$ and $(y_1,\ldots,y_n)$ are linearly dependent (parallel).

2. Pythagorean Theorem. If $N \in \mathbb{N}$ and $x_1, \ldots, x_N$ is an orthogonal system in an Euclidean space, especially in $\mathbb{R}^n$ then

$$\|\sum_{i=1}^{N} x_i\|^2 = \sum_{i=1}^{N} \|x_i\|^2 .$$

(The square of the hypotenuse equals to the sum of the squares of the perpendicular sides.)

We will use the following consequence of the Pythagorean theorem. If $x_1, \ldots, x_N$ is an orthogonal system in an Euclidean space and $k \in \{1, 2, \ldots, N\}$ is a fixed index then

$$\|\sum_{i=1}^{N} x_i\|^2 = \sum_{i=1}^{N} \|x_i\|^2 \geq \|x_k\|^2,$$

thus taking square root we got:

$$\|\sum_{i=1}^{N} x_i\| \geq \|x_k\| .$$

Here equality holds if and only if $x_i = 0$ for any $i \neq k$. This inequality expresses that the length of the hypotenuse is at least the length of any perpendicular side.

Combining this result with the triangle inequality we obtain:

$$\|x_k\| \leq \|\sum_{i=1}^{N} x_i\| \leq \sum_{i=1}^{N} \|x_i\| \qquad (k = 1, \ldots, N) . \tag{1.1}$$

Let us apply this result in $\mathbb{R}^n$ as follows. If $e_1, \ldots, e_n$ is the standard (orthonormal) basis

$$e_1 = (1, 0, \ldots, 0), \quad \ldots, \quad e_n = (0, 0, \ldots, 1)$$

in $\mathbb{R}^n$ and $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ then $x$ can be written as the orthogonal sum

$$x = \sum_{i=1}^{n} x_i e_i .$$

Apply the result (1.1) with $N = n$ and with the vectors $x_1 e_1, \ldots, x_n e_n \in \mathbb{R}^n$. Then we obtain on the one hand

$$\|x\| = \|\sum_{i=1}^{n} x_i e_i\| \geq \|x_k e_k\| = |x_k| \cdot \|e_k\| = |x_k| \cdot 1 = |x_k| \qquad (k = 1, \ldots, n) ,$$

on the other hand

$$\|x\| = \|\sum_{i=1}^{n} x_i e_i\| \leq \sum_{i=1}^{n} \|x_i e_i\| = \sum_{i=1}^{n} |x_i| .$$

Thus we have the following result:

$$|x_k| \leq \|x\| \leq \sum_{i=1}^{n} |x_i| \qquad (k = 1, \ldots, n)\,. \tag{1.2}$$

Notice that these inequalities can be deduced in elementary way too:

$$\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2} \geq \sqrt{x_k^2} = |x_k| \qquad (k = 1, \ldots, n)\,.$$

and

$$\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{\sum_{i=1}^{n} |x_i|^2} \leq \sqrt{\left(\sum_{i=1}^{n} |x_i|\right)^2} = \sum_{i=1}^{n} |x_i|\,.$$

## 1.2. $k$-arrays

Studying the higher order derivatives of a multivariate function it will be important some basic knowledge about the $k$-arrays.

**1.3. Definition** Let $n \in \mathbb{N}$ and $k \in \mathbb{N}$. The functions

$$A:\, \{1, \ldots, n\}^k \to \mathbb{R}$$

are called real $k$-arrays with size $n \times n \times \ldots \times n$. Their set is denoted by $\mathbb{R}^{\overset{1}{\smile}{n} \times n \times \ldots \times \overset{k}{\smile}{n}}$ (the number of $n$-s is $k$) or by $\mathbb{R}_k^n$.

The function value $A(j_1, \ldots, j_k) \in \mathbb{R}$ is called the $(j_1, \ldots, j_k)$-th entry of the $k$-array and is denoted by $(A)_{j_1,\ldots,j_k}$ or by $a_{j_1,\ldots,j_k}$.

**1.4. Remarks.**

1. In the definition $\{1, \ldots, n\}^k$ denotes the $k$-times Cartesian product of the set $\{1, \ldots, n\}$:
$$\{1, \ldots, n\}^k = \{1, \ldots, n\} \times \ldots \times \{1, \ldots, n\}\,,$$
that is the set of $k$-long finite sequences $(j_1, \ldots, j_k)$
where $j_1, \ldots, j_k \in \{1, \ldots, n\}$.

2. The 1-arrays are the vectors in $\mathbb{R}^n$. The 2-arrays are the matrices in $\mathbb{R}^{n \times n}$ and can be represented by a square in the plane with size $n \times n$.

3. The 3-arrays can be represented by a cube in the space with size
$n \times n \times n$. A general entry of a 3-array can be written using three indices as $a_{ijk}$.

4. A general $k$-array can be represented by a $k$ dimensional rectangular box whose sides have the lengthes $n$. Here the length means the number of entries in the current direction (dimension).

5. $\mathbb{R}^n_k$ is a vector space over $\mathbb{R}$ and $\dim \mathbb{R}^n_k = n^k$. It follows from the fact that the elements of $\mathbb{R}^n_k$ are functions defined on an $n^k$-element finite set. So $\mathbb{R}^n_k$ is isomorphic with $\mathbb{R}^{n^k}$ as vector space.

**1.5. Definition** The $k$-array $A \in \mathbb{R}^n_k$ is called symmetric if for any permutation $p_1, \ldots, p_k$ of the index system $j_1, \ldots, j_k$ holds

$$(A)_{p_1,\ldots,p_k} = (A)_{j_1,\ldots,j_k} .$$

Notice that here the permutation can be permutation with repetition.

**1.6. Remarks.** Every 1-array is symmetric. The interesting case is – from the point of view of symmetry – the case $k \geq 2$.

<u>The symbol $Ax^k$</u>

In the followings we will generalize the one variable monomials $ax^k$ for $n$-variable.

**1.7. Definition** Let $A \in \mathbb{R}^n_k$ and $x \in \mathbb{R}^n$. Then the symbol $Ax^k$ denotes the real number defined as

$$Ax^k := \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \ldots \sum_{j_k=1}^{n} a_{j_1,\ldots,j_k} \cdot x_{j_1} \cdot x_{j_2} \ldots \cdot x_{j_k} \in \mathbb{R} .$$

The $n$-multiple sum on the right side can be denoted by one sum-symbol where the indices are running – independently of each other – from 1 to $n$:

$$Ax^k = \sum_{j_1,\ldots j_k=1}^{n} a_{j_1,\ldots,j_k} \cdot x_{j_1} \cdot x_{j_2} \ldots \cdot x_{j_k} \in \mathbb{R} .$$

The mapping

$$\mathbb{R}^n \to \mathbb{R}, \qquad x \mapsto Ax^k$$

is called $n$-variable homogeneous polynomial of degree $k$.

**1.8. Remarks.**

1. Let $n = 1$, $A \in \mathbb{R}^1_k$. Denote by $a$ the single entry of $A$ that is $(A)_{1,\ldots,1} = a$. Then for any $x = (x_1) \in \mathbb{R}^1 \cong \mathbb{R}$:

$$Ax^k = \sum_{j_1,\ldots j_k=1}^{1} (A)_{j_1,\ldots,j_k} \cdot x_{j_1} \cdot x_{j_2} \ldots \cdot x_{j_k} = a \cdot \underbrace{x_1}_{1} \cdot x_1 \cdot \ \ldots \ \cdot \underbrace{x_1}_{k} = ax^k ,$$

   thus $Ax^k$ is really a generalization of the one-variable monomial $ax^k$.

2. The $n$-variable homogeneous polynomials of degree 1 are the linear functionals of type $\mathbb{R}^n \to \mathbb{R}$:

$$Ax = \sum_{i=1}^{n} a_i \cdot x_i \quad \text{where} \quad A = (a_1, \ldots, a_n) \in \mathbb{R}^n, \qquad x \in \mathbb{R}^n .$$

3. The $n$-variable homogeneous polynomials of degree 2 are the quadratic forms:

$$Ax^2 = \sum_{i,j=1}^{n} a_{ij} \cdot x_i \cdot x_j \quad \text{where} \quad A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ & & \vdots & \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} \in \mathbb{R}^{n\times n}, \qquad x \in \mathbb{R}^n .$$

4. It is obvious that if $A,\, B \in \mathbb{R}^n_k$ and $\lambda \in \mathbb{R}$ then for any $x \in \mathbb{R}^n$ hold

$$(A+B)x^k = Ax^k + Bx^k , \qquad (\lambda A)x^k = \lambda(Ax^k)$$
$$A(\lambda x)^k = \lambda^k \cdot (Ax^k) .$$

$Ax^k$ with symmetric $k$-array.

Let us discuss another formula for the symbol $Ax^k$ provided that $A \in \mathbb{R}^n_k$ is symmetric (in the sense of Definition 1.5). In this case the sum in the definition of $Ax^k$ contains a lot of identical terms, more precisely, the term

$$a_{j_1,\ldots,j_k} \cdot x_{j_1} \cdot x_{j_2} \cdot \ldots \cdot x_{j_k}$$

equals to the term

$$a_{p_1,\ldots,p_k} \cdot x_{p_1} \cdot x_{p_2} \cdot \ldots \cdot x_{p_k}$$

where $p_1, \ldots, p_k$ is a permutation of $j_1, \ldots, j_k$. If the index system $j_1, \ldots, j_k$ contains $i_1$ times the index 1, $i_2$ times the index 2, …, $i_n$ times the index $n$ where

$$i_1,\ i_2,\ \ldots,\ i_n \in \mathbb{N}_0, \qquad i_1 + i_2 + \ldots + i_n = k ,$$

then the number of the possible permutations (permutation with repetition) is

$$\frac{k!}{i_1! \cdot i_2! \cdot \ldots \cdot i_n!} = \frac{k!}{i!} ,$$

where the meaning of $i!$ will be given in the following definition.

**1.9. Definition** The $n$-dimensional vector $i = (i_1, \ldots, i_n)$ is called multi-index if $i_1, i_2, \ldots, i_n \in \mathbb{N}_0$. Thus the set of the multi-indices is $\mathbb{N}_0^n$. If $i \in \mathbb{N}_0^n$ is a multi-index and $x \in \mathbb{R}^n$ is a vector then the absolute value of $i$, the factorial of $i$ and the power $x^i$ are defined as

$$|i| := i_1 + i_2 + \ldots + i_n , \qquad i! := i_1! \cdot i_2! \cdot \ldots \cdot i_n! , \qquad x^i := x_1^{i_1} \cdot x_2^{i_2} \cdot \ldots \cdot x_n^{i_n}$$

respectively.

Using these abbreviations, the terms of the sum of the definition of $Ax^k$ can be grouped into subsets. A subset will contain the terms whose indices are the permutations of each other, and these terms are equal. Such a subset can be described uniquely by a multi-index in the following way. Let $i \in \mathbb{N}_0^n$, $|i| = k$ be a multi-index and associate to this $i$ the non-increasing index system $j(i)$ where

$$j(i) := (j_1, j_2, \ldots, j_k) := (\underbrace{n, \ldots, n}_{i_n \text{ times}}, \underbrace{n-1, \ldots, n-1}_{i_{n-1} \text{ times}}, \ldots, \underbrace{1, \ldots, 1}_{i_1 \text{ times}}).$$

Denote by $\mathrm{Perm}\,(j(i))$ the set of possible permutations $p = (p_1, \ldots, p_k)$ of the sequence $j(i)$. Thus the identical terms of the above mentioned subset are

$$a_{p_1,\ldots,p_k} \cdot x_{p_1} \cdot x_{p_2} \cdot \ldots \cdot x_{p_k}$$

where $p = (p_1, \ldots, p_k) \in \mathrm{Perm}\,(j(i))$. We remind that the number of elements in $\mathrm{Perm}\,(j(i))$ is $\dfrac{k!}{i!}$.

After these considerations we can write the expression of $Ax^k$ as follows:

$$Ax^k = \sum_{j_1,\ldots j_k=1}^{n} a_{j_1,\ldots,j_k} \cdot x_{j_1} \cdot x_{j_2} \cdot \ldots \cdot x_{j_k} = \sum_{\substack{i \in \mathbb{N}_0^n \\ |i|=k}} \sum_{p \in \mathrm{Perm}\, j(i)} a_{p_1,\ldots,p_k} \cdot x_{p_1} \cdot x_{p_2} \cdot \ldots \cdot x_{p_k} =$$

$$= \sum_{\substack{i \in \mathbb{N}_0^n \\ |i|=k}} \frac{k!}{i!} \cdot a_{j_1,\ldots,j_k} \cdot x_{j_1} \cdot x_{j_2} \cdot \ldots \cdot x_{j_k} = \sum_{\substack{i \in \mathbb{N}_0^n \\ |i|=k}} \frac{k!}{i!} \cdot a_{j_1,\ldots,j_k} \cdot x_1^{i_1} \cdot x_2^{i_2} \cdot \ldots \cdot x_n^{i_n},$$

where $j(i) = (j_1, j_2, \ldots, j_k)$. If we use the notation $a_i := a_{j(i)} = a_{j_1,\ldots,j_k}$, then we obtain the following form (the so called *multi-index form*) for $Ax^k$:

$$Ax^k = \sum_{\substack{i \in \mathbb{N}_0^n \\ |i|=k}} \frac{k!}{i!} \cdot a_i \cdot x^i. \tag{1.3}$$

Note that the number of terms in this sum equals to $\binom{n+k-1}{k}$ (see: combinations with repetition).

<u>The norm of a $k$-array.</u>

Let us introduce the Euclidean vector norm on the vector space $\mathbb{R}_k^n$ (it is isomorphic with $\mathbb{R}^{n^k}$). This will be called the Frobenius-norm (or Euclidean norm).

**1.10. Definition** Let $A \in \mathbb{R}_k^n$. Its Frobenius-norm (or Euclidean norm) is defined as

$$\|A\|_F := \sqrt{\sum_{j_1,\ldots j_k=1}^{n} a_{j_1,\ldots,j_k}^2}$$

**1.11. Remark.** The Frobenius-norm satisfies the axioms of vector norm. Moreover $\mathbb{R}_k^n$ supplied with the Frobenius-norm and $\mathbb{R}^{n^k}$ supplied with the Euclidean norm are isomorphic as linear normed spaces.

The following theorem gives us an upper estimation for $|Ax^k|$.

**1.12. Theorem** *Let $A \in \mathbb{R}_k^n$ and $x \in \mathbb{R}^n$. Then*

$$|Ax^k| \leq \|A\|_F \cdot \|x\|^k .$$

**Proof.** The proof is based on the application of the Cauchy-Bunyakovsky-Schwarz inequality several times.

We will prove the theorem for the case $k = 2$. The general proof is similar to this case and requires mathematical induction.

Suppose that $k = 2$. Then we can use the indices $i$ and $j$ instead of $j_1$ and $j_2$.

$$\begin{aligned}|Ax^2|^2 = (Ax^2)^2 = \left(\sum_{i,j=1}^n a_{ij}x_ix_j\right)^2 &= \left(\sum_{i=1}^n x_i \cdot \sum_{j=1}^n a_{ij}x_j\right)^2 \leq \\ &\leq \left(\sum_{i=1}^n x_i^2\right) \cdot \left(\sum_{i=1}^n \left(\sum_{j=1}^n a_{ij}x_j\right)^2\right) \leq \\ &\leq \|x\|^2 \cdot \sum_{i=1}^n \left(\sum_{j=1}^n a_{ij}^2\right)\left(\sum_{j=1}^n x_j^2\right) = \|x\|^2 \cdot \|A\|_F^2 \cdot \|x\|^2 = \|A\|_F^2 \cdot \|x\|^4 .\end{aligned}$$

So $|Ax^2|^2 \leq \|A\|_F^2 \cdot \|x\|^4$, which implies the statement of the theorem. □

## 1.3. Homeworks

1. The following vectors are given in $\mathbb{R}^4$:

   $$x := (-1,\, 3,\, 5,\, 2) \qquad y := (2,\, -3,\, -1,\, 1).$$

   Determine:

   $$a)\ \ x+y \quad b)\ \ x-y \quad c)\ \ 3x \quad d)\ \ 2x-5y$$

   $$e)\ \ \langle x,\, y\rangle \quad f)\ \ ||x|| \quad g)\ \ d(x,y)$$

2. Let $A \in \mathbb{R}_3^2$ be a symmetric 3-array and $x \in \mathbb{R}^2$. Write $Ax^3$ in the original form (see Defininition 1.7) and in the multi-index form (see formula (1.3)) respectively, and check their identity.

3. Compute the Frobenius-norm of the following 2-arrays:

$$a) \quad A = \begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix} \qquad b) \quad B = \begin{bmatrix} 0 & 4 & -1 \\ -3 & 1 & 2 \\ -2 & 0 & -3 \end{bmatrix}$$

c) Compute $Ax^2$ if $x = (1, 2)$ then check the statement of Theorem 1.12.

d) Compute $Bx^2$ if $x = (1, 1, -1)$ then check the statement of Theorem 1.12.

4. a) Compute the Frobenius-norm of the following 3-array:

$$A \in \mathbb{R}^{2\times2\times2}, \qquad a_{ijk} := i + j - k^2 \qquad (i, j, k = 1,\, 2)\,.$$

b) Compute $Ax^3$ if $x = (-2, 1)$ then check the staement of Theorem 1.12.

# 2. Lesson 2

## 2.1. Balls in $\mathbb{R}^n$

**2.1. Definition** The neighbourhood (or ball or environment) of the point $a \in \mathbb{R}^n$ with radius $r > 0$ is the set

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

**2.2. Remark.** In the case $n = 1$ the ball is the open interval

$$B(a, r) = (a - r,\, a + r) .$$

In the case $n = 2$ the ball is the open circular disk

$$B(a, r) = \{x = (x_1,\, x_2) \in \mathbb{R}^2 \mid (x_1 - a_1)^2 + (x_2 - a_2)^2 < r^2\} .$$

One can easily prove the following basic properties of neighbourhoods:

**2.3. Theorem** *1. If* $0 < r_1 < r_2$ *then* $B(a, r_1) \subset B(a, r_2)$

*2.* $\bigcap\limits_{r \in \mathbb{R}^+} B(a, r) = \{a\}$

*3. ($T_2$-property, separation) Let* $a,\, b \in \mathbb{R}^n$, $a \neq b$ *Then there exists* $r_1 > 0$, $r_2 > 0$ *such that*

$$B(a, r_1) \cap B(b, r_2) = \emptyset$$

**Proof.** We will prove only the third statement of the theorem.

Let $r_1 = r_2 = r = \dfrac{\|a - b\|}{3}$. We will prove that $B(a, r) \cap B(b, r) = \emptyset$.
Assume, on the contrary that $\exists\, x \in B(a, r) \cap B(b, r)$.
Then it holds for such an $x$ that

$$\|x - a\| < \frac{\|a - b\|}{3} \quad \text{and} \quad \|x - b\| < \frac{\|a - b\|}{3} .$$

Using this and the triangle inequality:

$$\|a - b\| = \|a - x + x - b\| \leq \|a - x\| + \|x - b\| = \|x - a\| + \|x - b\| <$$
$$< \frac{\|a - b\|}{3} + \frac{\|a - b\|}{3} = \frac{2}{3}\|a - b\|$$

holds which is a contradiction. □

We add briefly some concepts in connection with the neighbourhoods in different dimensions:

**2.4. Definition** Let $k \in \{1, \ldots, n\}$ and $I := \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$. Suppose that $1 \leq i_1 < i_2 < \ldots < i_k \leq n$. Denote by $i$ the index vector $i = (i_1, i_2, \ldots i_k)$. The set

$$CHP(i) := \{x \in \mathbb{R}^n \mid x_i = 0 \text{ if } i \notin I\} \subseteq \mathbb{R}^n$$

is called $i$-coordinate (or: $(i_1, i_2, \ldots i_k)$-coordinate) hyperplane.

E. g. in $\mathbb{R}^3$ the (1, 2)-coordinate hyperplane is the $xy$-plane, the (2)-coordinate hyperplane is the $y$-axis, etc.

Obviously the set $CHP(i)$ is a $k$-dimensional subspace in $\mathbb{R}^n$ that is isomorphic with $\mathbb{R}^k$ via the following mapping:

$$\varphi : CHP(i) \to \mathbb{R}^k, \quad \varphi(x) := (x_{i_1}, \ldots, x_{i_k}).$$

Thus a point $a \in CHP(i)$ has two kinds of neighbourhoods: in $\mathbb{R}^n$ ($n$-dimensional, denote it by $B(a, r)$) and in $\mathbb{R}^k$ ($k$-dimensional, denote it by $B_i(a, r)$). You can easily prove that

$$B_i(a, r) = B(a, r) \cap CHP(i).$$

If e. g. the point $a$ lies on the $xy$-plane in $\mathbb{R}^3$ then this connection expresses that the intersection of a ball with a plane is a circle.

## 2.2. Topology in $\mathbb{R}^n$

In the previous section we have defined the ball. Using this concept we can define important classes of points in connection of a fixed set.

**2.5. Definition** Let $\emptyset \neq H \subset \mathbb{R}$, $a \in \mathbb{R}^n$. Then

1. $a$ is an interior point of $H$, if $\exists\, r > 0 : \quad B(a, r) \subseteq H$.
2. $a$ is an exterior point of $H$, if $\exists\, r > 0 : \; B(a, r) \cap H = \emptyset$.
   In other words: $\exists\, r > 0 : \; B(a, r) \subseteq \overline{H}$.
   Here $\overline{H}$ denotes the complement of $H$ that is $\overline{H} = \mathbb{R} \setminus H$.
3. $a$ is a boundary point of $H$, if
   $\forall\, r > 0 : \quad B(a, r) \cap H \neq \emptyset$ and $B(a, r) \cap \overline{H} \neq \emptyset$.

**2.6. Remark.** Every interior point lies in $H$, every exterior point lies in $\overline{H}$. But a boundary point can belong to $H$ or to its complement.

**2.7. Definition** 1. The set of the interior points of $H$ is called the interior of $H$ and is denoted by $\operatorname{int} H$. So

$$\operatorname{int} H := \{a \in \mathbb{R} \mid \exists\, r > 0 : \quad B(a, r) \subseteq H\} \subseteq H.$$

2. The set of the exterior points of $H$ is called the exterior of $H$ and is denoted by $\operatorname{ext} H$. So
$$\operatorname{ext} H := \{a \in \mathbb{R} \mid \exists\, r > 0 : \quad B(a, r) \subseteq \overline{H}\} \subseteq \overline{H}.$$

3. The set of the boundary points of $H$ is called the boundary of $H$ and is denoted by $\partial H$. So
$$\partial H := \{a \in \mathbb{R} \mid \forall\, r > 0 : \quad B(a, r) \cap H \neq \emptyset \quad \text{and} \quad B(a, r) \cap \overline{H} \neq \emptyset\} \subset \mathbb{R}.$$

**2.8. Remark.** $\mathbb{R} = \operatorname{int} H \cup \partial H \cup \operatorname{ext} H$ and this is a union of disjoint sets.

You can easily see that $\operatorname{int} H = H \setminus \partial H$, so we obtain the interior of a set if we subtract from the set its boundary. If we add the boundary to the set then we obtain the closure of the set as you see in the following definition.

**2.9. Definition** The set $H \cup \partial H$ is called the closure of $H$ and is denoted by $\operatorname{clos} H$. So $\operatorname{clos} H := H \cup \partial H$.

It is obvious that $\overline{\operatorname{clos} H} = \operatorname{int} \overline{H}$ and $\overline{\operatorname{int} H} = \operatorname{clos} \overline{H}$. This is based on the simple fact that $\partial H = \partial \overline{H}$.

**2.10. Definition** Let $H \subseteq \mathbb{R}^n$. Then

1. $H$ is called an open set $\overset{\text{df}}{\Leftrightarrow} \partial H \subseteq \overline{H}$.

2. $H$ is called a closed set $\overset{\text{df}}{\Leftrightarrow} \partial H \subseteq H$.

**2.11. Remarks.**

1. $H$ is open if and only if it does not contain any boundary point and it is closed if and only if it contains all of its boundary points.

2. $\emptyset$ and $\mathbb{R}^n$ are open and closed sets at the same time. There is no other set in $\mathbb{R}^n$ that is open and closed at the same time.

3. $H$ is open $\Leftrightarrow$ $\overline{H}$ is closed, $H$ is closed $\Leftrightarrow$ $\overline{H}$ is open.

4. $H$ is open $\Leftrightarrow$ $H \subseteq \operatorname{int} H \Leftrightarrow H = \operatorname{int} H$.

5. $H$ is closed $\Leftrightarrow$ $\operatorname{clos} H \subseteq H \Leftrightarrow H = \operatorname{clos} H$.

**2.12. Definition** Let $\emptyset \neq H \subseteq \mathbb{R}^n$. Then $H$ is called bounded if
$$\exists\, M > 0\ \forall\, x \in H : \quad \|x\| \leq M\,.$$

## 2.3. Homeworks

1. Prove the statements 1. and 2. of Theorem 2.3.

2. Let $a \in \mathbb{R}^n$, $r > 0$. Prove that in $\mathbb{R}^n$

   a) $B(a,\, r)$ is an open set.

   b) The set $\{x \in \mathbb{R}^n \mid \|x - a\| \le r\}$ (the closed ball) is a closed set.

   c) The set $\{x \in \mathbb{R}^n \mid \|x - a\| = r\}$ (the sphere) is a closed set.

3. a) Prove that any ball $B(a, r)$ contains $n$ linearly independent vectors.

   b) Prove that for any subspace $W \subsetneqq \mathbb{R}^n$ $\operatorname{int} W = \emptyset$.

   c) Prove that for any subspace $W \subseteq \mathbb{R}^n$ $W$ is a closed set.

4. Determine $\operatorname{int} H$, $\partial H$, $\operatorname{ext} H$ and $\operatorname{clos} H$ if $H \subset \mathbb{R}^2$ and

   a) $H = \{(x, y) \in \mathbb{R}^2 \mid 1 \le x < 3,\ 1 \le y < 2\}$.

   b) $H = \{(x, y) \in \mathbb{R}^2 \mid x \ge 0,\ x^2 + y^2 < 1\}$.

5. Prove that a set $\emptyset \ne H \subset \mathbb{R}^n$ is bounded if and only if it can be covered by a ball that is

$$\exists\, a \in \mathbb{R}^n \text{ and } \exists\, r > 0 : \qquad H \subseteq B(a, r)\,.$$

# 3. Lesson 3

## 3.1. Sequences in $\mathbb{R}^n$

Similarly to number sequences we can define vector sequences in $\mathbb{R}^n$.

**3.1. Definition** A function $a : \mathbb{N} \to \mathbb{R}^n$ is called a vector sequence in $\mathbb{R}^n$. The function value $a(k) \in \mathbb{R}^n$ ordered to the number $k \in \mathbb{N}$ is called the $k$-th term of the sequence. If we want to indicate the components of the $k$-th term then we will rather use the notation $a^{(k)}$ instead of $a(k)$. If we don't want to indicate the components then we can use the usual notation $a_k$ for the $k$-th term.

**3.2. Definition** Let $a^{(k)} \in \mathbb{R}^n$ $(k \in \mathbb{N})$ be a vector sequence in $\mathbb{R}^n$. Then

$$a^{(k)} = (a_1^{(k)}, \, a_2^{(k)}, \, \ldots, \, a_n^{(k)}) \in \mathbb{R}^n \quad (k \in \mathbb{N}) \, .$$

The number sequence $a_i^{(k)} \in \mathbb{R}$ $(k \in \mathbb{N})$ is called the $i$-th coordinate sequence of $(a^{(k)})$ $(i = 1, \ldots, n)$.

**3.3. Definition** The sequence $a^{(k)} \in \mathbb{R}^n$ $(k \in \mathbb{N})$ is called bounded if

$$\exists \, M > 0 \; \forall \, k \in \mathbb{N} : \quad \|a^{(k)}\| \le M \, .$$

It is obvious that the vector sequence $(a^{(k)})$ is bounded if and only if the real number sequence $(\|a^{(k)}\|)$ is bounded. Moreover, using the inequalities (1.2) one can prove, that a vector sequence is bounded if and only if its every coordinate sequence is bounded. That is

$$(a^{(k)}) \text{ is bounded} \quad \Leftrightarrow \quad (a_i^{(k)}) \text{ is bounded} \quad (i = 1, \ldots, n) \, .$$

**3.4. Definition** The vector sequence $a : \mathbb{N} \to \mathbb{R}^n$ is called convergent if

$$\exists A \in \mathbb{R}^n \; \forall \varepsilon > 0 \; \exists N \in \mathbb{N} \; \forall k \ge N : \quad a^{(k)} \in B(A, \varepsilon) \, .$$

The definition can be written using inequalities as follows:

$$\exists A \in \mathbb{R}^n \; \forall \varepsilon > 0 \; \exists N \in \mathbb{N} \; \forall k \ge N : \quad \|a^{(k)} - A\| < \varepsilon \, .$$

A vector sequence is called divergent if it is not convergent.

It can be proved (using the $T_2$-property of the neighbourhoods) that the vector $A$ in the above definition is unique. It is called the limit (or limit vector) of the vector sequence $(a^{(k)})$, and it is denoted in one of the following ways:

$$\lim a = A \, , \quad \lim a^{(k)} = A \, , \quad \lim_{k \to \infty} a^{(k)} = A \, , \quad a^{(k)} \to A \quad (k \to \infty).$$

**3.5. Remark.** If $a : \mathbb{N} \to \mathbb{R}^n$ is a vector sequence and $A \in \mathbb{R}^n$ then $\lim\limits_{k\to\infty} a^{(k)} = A$ is equivalent with

$$\forall \varepsilon > 0 \; \exists N \in \mathbb{N} \; \forall k \geq N : \quad a^{(k)} \in B(A, \varepsilon)\,,$$

or – using inequalities – with

$$\forall \varepsilon > 0 \; \exists N \in \mathbb{N} \; \forall k \geq N : \quad \|a^{(k)} - A\| < \varepsilon\,.$$

The number $N$ is called threshold index to $\varepsilon$.

**3.6. Theorem** *Let $a^{(k)} \in \mathbb{R}^n$ $(k \in \mathbb{N})$ be a vector sequence and $A \in \mathbb{R}^n$ be a vector. Then*

$$\lim_{k\to\infty} a^{(k)} = A \quad \Longleftrightarrow \quad \lim_{k\to\infty} ||a^{(k)} - A|| = 0$$

**Proof.** The proof is obvious, if we consider that

$$||a^{(k)} - A|| = |\,||a^{(k)} - A|| - 0\,|\,.$$

□

In the following theorem we reduce the convergence of a vector sequence back to the convergence of its coordinate sequences.

**3.7. Theorem** *Let $a^{(k)} \in \mathbb{R}^n$ $(k \in \mathbb{N})$ be a vector sequence and $A \in \mathbb{R}^n$ be a vector. Then*

$$\lim_{k\to\infty} a^{(k)} = A \quad \Longleftrightarrow \quad \lim_{k\to\infty} a_i^{(k)} = A_i \quad (i = 1\,, \ldots,\, n)\,.$$

**Proof.** Applying the inequalities (1.2) for the vectors $a^{(k)} - A$ we obtain

$$|a_i^{(k)} - A_i| \leq \|a^{(k)} - A\| \leq \sum_{i=1}^{n} |a_i^{(k)} - A_i| \qquad (i = 1\,, \ldots,\, n)\,,$$

which implies both directions of the statement. □

This reduction to the coordinate sequences makes it possible to prove easily some important basic theorems: connection between the convergence and the boundedness, between the convergence and the algebraic operations, and also to prove the completeness of $\mathbb{R}^n$.

Similarly to the number sequences one can prove that a convergent sequence is bounded (using norm instead of the absolute value). The opposite direction is not true: there exist bounded sequences in $\mathbb{R}^n$ that are divergent. For example if $x \in \mathbb{R}^n \setminus \{0\}$ then the sequence $a_k := (-1)^k \cdot x$ $(k \in \mathbb{N})$ is such a sequence.

**3.8. Theorem** *[Bolzano-Weierstrass] Let $a^{(k)} \in \mathbb{R}^n$ $(k \in \mathbb{N})$ be a bounded vector sequence. Then it has a convergent subsequence.*

**Proof.** For simplicity we will present the proof in the case $n = 2$. The general case can be proved in the same way.

As $(a^{(k)})$ is bounded, its first coordinate sequence $(a_1^{(k)})$ is a bounded real number sequence. Using the Bolzano-Weierstrass theorem in $\mathbb{R}$ (see: Analysis-1) it has a convergent subsequence $(a_1^{(k_m)},\ m \in \mathbb{N})$. But the subsequence $(a_2^{(k_m)},\ m \in \mathbb{N})$ of the second coordinate sequence $(a_2^{(k)})$ is also bounded, so – applying once more the Bolzano-Weierstrass theorem in $\mathbb{R}$ – it has a convergent subsequence $(a_2^{(k_{m_s})},\ s \in \mathbb{N})$. Then the vector sequence

$$a^{(k_{m_s})} = (\, a_1^{(k_{m_s})},\, a_2^{(k_{m_s})}\,) \qquad (s \in \mathbb{N})$$

is obviously convergent. $\square$

## 3.2. Characterization of closed sets with sequences

The closeness of a set in $\mathbb{R}^n$ can be described with vector sequences.

**3.9. Theorem** *Let $\emptyset \neq H \subseteq \mathbb{R}$. Then $H$ is closed if and only if*

$$\forall\, a_k \in H \ (k \in \mathbb{N}) \text{ convergent sequence}: \ \lim_{k\to\infty} a_k \in H\,.$$

**Proof.** $\Rightarrow$:
Let $a_k \in H$ $(k \in \mathbb{N})$ be a convergent sequence and $A := \lim a_k$. We need to prove that $A \in H$.

Suppose indirectly $A \notin H$. Then $A \in \overline{H}$. But $\overline{H}$ is open (because $H$ is closed), therefore

$$\exists\, \varepsilon > 0 : \quad B(A,\, \varepsilon) \subset \overline{H}\,.$$

But to this $\varepsilon$:

$$\exists\, N \in \mathbb{N} \quad \forall\, k \geq N : \quad a_k \in B(A,\, \varepsilon) \subset \overline{H}\,.$$

This is a contradiction: $a_k \in H$ and $a_k \in \overline{H}$ cannot be at the same time.

$\Leftarrow$:
Suppose indirectly that $H$ is not closed. This implies that $\overline{H}$ is not open, thus $\exists\, A \in \overline{H}$ that is not interior point of $\overline{H}$. This means that

$$\forall\, r > 0 : \quad B(a, r) \not\subset \overline{H} \quad \text{that is} \quad B(a, r) \cap H \neq \emptyset\,.$$

Applying this fact for the numbers $r := \dfrac{1}{k}$ $(k \in N)$ we obtain

$$\forall\, k \in \mathbb{N} \quad \exists\, a_k \in B(a,\, \frac{1}{k}) \cap H\,.$$

So we have defined a sequence $a_k \in H$ $(k \in \mathbb{N})$. Since

$$||a_k - A|| < \frac{1}{k} \quad (k \in \mathbb{N})\,,$$

the limit of this sequence is $A$. So by the condition $A \in H$. But $A$ was chosen from the set $\overline{H}$. This is a contradiction. $\square$

The intuitive content of the above theorem is that it is impossible to go out from a closed set via convergence.

## 3.3. Compact sets

**3.10. Definition** Let $\emptyset \neq H \subseteq \mathbb{R}^n$. $H$ is called a compact set if

$$\forall\, a_k \in H \ (k \in \mathbb{N}) \text{ sequence } \quad \exists\, (a_{k_m},\, m \in \mathbb{N}) \text{ subsequence :}$$

$$(a_{k_m},\, m \in \mathbb{N}) \text{ is convergent and } \lim_{m\to\infty} a_{k_m} \in H\,.$$

The $\emptyset$ is called to be compact by definition.

Similarly to the case of compact sets in $\mathbb{R}$ (see: Analysis-2) the following theorem can be proved. We need only use the norm instead of the absolute value.

**3.11. Theorem** *Let $\emptyset \neq H \subseteq \mathbb{R}^n$. Then $H$ is compact if and only if it is closed and bounded.*

**3.12. Remark.** The theorem is not valid in infinite dimensional normed spaces. Every compact set is closed and bounded but there exists a closed and bounded set, that is not compact (see: Functional Analysis).

## 3.4. Homeworks

1. Prove by definition of the convergence, that in $\mathbb{R}^2$

$$\lim_{n\to\infty} \left( \frac{n+1}{2n-3}\,,\ \frac{3n-2}{n+5} \right) = \left( \frac{1}{2},\ 3 \right)$$

   Determine a threshold index to $\varepsilon = 10^{-3}$.

2. Determine the limit of the following sequence in $\mathbb{R}^3$:

$$a_n = \left( \frac{1}{n}\,,\ \left(1+\frac{1}{n}\right)^n\,,\ \frac{2n-1}{3n+7} \right) \qquad (n \in \mathbb{N})\,.$$

3. Using sequences prove that the following set is not closed:

$$H = \{(x,y) \in \mathbb{R}^2 \mid 1 \le x < 3,\ 1 \le y < 2\} \subset \mathbb{R}^2$$

4. Using the definition of compactness prove that the following sets in $\mathbb{R}^2$ are not compact:

   a) $H = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 < 4\}$

   b) $\{(x,y) \in \mathbb{R}^2 \mid 1 \le x \le 3\}$

# 4. Lesson 4

## 4.1. The limit of functions of type $\mathbb{R}^n \to \mathbb{R}^m$

As in the one variable case the concept of limits expresses where tend the function values to if the variable tends to a certain point. The first problem is to discuss the points where the variable can tend. These points are the so called accumulation points of the domain of the function.

**4.1. Definition (Accumulation point)** Let $\emptyset \neq H \subseteq \mathbb{R}^n$ and $a \in \mathbb{R}^n$. We say that $a$ is an accumulation point of $H$ if

$$\forall\, r > 0 : (B(a,r) \setminus \{a\}) \cap H \neq \emptyset\,.$$

The set of accumulation points of $H$ is denoted by $H'$ that is

$$H' := \{a \in \mathbb{R}^n \mid a \text{ is an accumulation point of } H\}\,.$$

The points of the set $H \setminus H'$ are called isolated points of $H$.

**4.2. Definition (isolated point)** Let $\emptyset \neq H \subseteq \mathbb{R}^n$ and $a \in \mathbb{R}^n$. We say that $a$ is an isolated point of $H$ if $a \in H$ and

$$\exists\, r > 0 : \quad B(a,r) \setminus \{a\} \cap H = \emptyset\,.$$

After these preliminaries follows the definition of the limit:

**4.3. Definition** Let $f \in \mathbb{R}^n \to \mathbb{R}^m$, $a \in D'_f$. We say that $f$ has limit at the point $a$ if

$$\exists\, A \in \mathbb{R}^m\ \forall\, \varepsilon > 0\ \exists\, \delta > 0\ \forall\, x \in (B(a,\delta) \setminus \{a\}) \cap D_f : \quad f(x) \in B(A,\varepsilon)\,.$$

Using the $T_2$-property of neighbourhoods it can be proved that the vector $A$ in this definition is unique. This unique $A$ is called the limit of the function $f$ at the point $a$. The notations are:

$$A = \lim_a f, \quad A = \lim_{x \to a} f(x), \quad f(x) \to A\ (x \to a)\,.$$

**4.4. Remark.** Thus the fact $\lim\limits_a f = A$ can be expressed with neighbourhoods:

$$\forall\, \varepsilon > 0\ \exists\, \delta > 0\ \forall\, x \in (B(a,\delta) \setminus \{a\}) \cap D_f : \quad f(x) \in B(A,\varepsilon)\,,$$

and with inequalities:

$$\forall\, \varepsilon > 0\ \exists\, \delta > 0\ \forall\, x \in D_f, 0 < \|x - a\| < \delta : \quad \|f(x) - A\| < \varepsilon\,.$$

**4.5. Examples**

1. (the constant function) Let $f : \mathbb{R}^n \to \mathbb{R}^m,\ f(x) = c$ where $c \in \mathbb{R}^m$ is a fixed vector. Then for any $a \in \mathbb{R}^n$: $\lim\limits_{x\to a} c = c$, because for any $\varepsilon > 0$ any $\delta > 0$ is good:
$$\forall\, x \in \Big(B(a,\delta) \setminus \{a\}\Big) \cap D_f : \quad f(x) = c \in B(c,\varepsilon)\,.$$
2. (identity function) Let $f : \mathbb{R}^n \to \mathbb{R}^n,\ f(x) := x$. Let $a \in \mathbb{R}^n$. Then $\lim\limits_{x\to a} x = a$, because for any $\varepsilon > 0$ let $\delta := \varepsilon$. It will be good, since
$$\forall\, x \in \Big(B(a,\delta) \setminus \{a\}\Big) \cap D_f : \quad f(x) = x \in B(a,\delta) = B(a,\varepsilon)\,.$$
3. (canonical projections) Let $f : \mathbb{R}^n \to \mathbb{R}, f(x) := x_i$ where $i \in \{1,\dots,n\}$ is fixed and $x = (x_1,\dots,x_n)$. Let $a = (a_1,\dots,a_n) \in \mathbb{R}^n$. Then $\lim\limits_{x\to a} f(x) = a_i$ because for any $\varepsilon > 0$ let $\delta := \varepsilon$. This is good since if $0 < \|x-a\| < \delta$, then by (1.2):
$$|f(x) - a_i| = |x_i - a_i| = \|(x-a)_i\| \le \|x-a\| < \delta = \varepsilon\,.$$
Note, that in the case $n = 1$, the projection coincides with the identity.

Similarly to the one variable case it can be proved the theorem of Transference Principle:

**4.6. Theorem** *[Transference Principle for limits] Using the previous notations:*
$$\lim_{x\to a} f(x) = A \quad \Leftrightarrow \quad \forall\, \overbrace{x_k \in D_f \setminus \{a\} \quad (k\in\mathbb{N}),\ \lim x_k = a}^{\textit{allowed sequence}} : \quad \lim f(x_k) = A\,.$$

The most important corollaries of the Transference Principle are – like in the one variable case – the algebraic operations with the limits.

For $m \ge 2$ we can speak about the limits by coordinates. To formulate this statement we have to define the coordinate function.

**4.7. Definition** Let $f \in \mathbb{R}^n \to \mathbb{R}^m$, $f(x) = (f_1(x), \dots f_m(x)) \in \mathbb{R}^m$. Then the function $f_i : D_f \to \mathbb{R}$ is called the $i$-th coordinate function of $f$ $(i = 1, \dots m)$. We often use the notation $f = (f_1, \dots, f_m)$.

Clearly in the case $m = 1$ $f_1 = f$. Using the Transference Principle and the convergence of sequences by coordinates, one can prove the following theorem:

**4.8. Theorem** *[limit by coordinates] Suppose that $m \ge 2$ and let $f \in \mathbb{R}^n \to \mathbb{R}^m,\ a \in D'_f,\ A = (A_1,\dots,A_m) \in \mathbb{R}^m$. Then*
$$\lim_a f = A \quad \Leftrightarrow \quad \lim_a f_i = A_i \quad (i = 1, \dots, m)\,.$$

**4.9. Remark.** With short notation our theorem is:
$$\lim_{x\to a} f(x) = \Big(\lim_{x\to a} f_1(x), \dots, \lim_{x\to a} f_m(x)\Big)\,.$$
Moreover the existence of one side of the above equality implies the existence of the other side.

## 4.2. Limit along a set

Sometimes we approach a point in such way that the variable remains in a fixed given set. In this case we speak about limit along a set. First we define the restriction of a function.

**4.10. Definition** Let $f \in A \to B,\ \emptyset \neq C \subset D_f$. The function

$$f_{|C} : C \to B, \quad f_{|C}(x) := f(x)$$

is called the restriction of $f$ onto the set $C$.

**4.11. Definition (limit along a set)** Let $f \in \mathbb{R}^n \to \mathbb{R}^m, \emptyset \neq H \subseteq \mathbb{R}^n$. Suppose that $a \in \mathbb{R}^n$ and $a \in (H \cap D_f)'$. Then the limit along the set $H$ is defined as follows:

$$\lim_{\substack{x \to a \\ x \in H}} f(x) := \lim_{x \to a} f_{|H \cap \mathcal{D}_f}(x) \,.$$

Since the definition reduces the limit along a set back to the limit of functions, all the statements (Transference Principle, algebraic operations, limit by coordinates) are valid for limits along a set.

**4.12. Remarks.**

1. If $H = D_f$ then $\lim\limits_{\substack{x \to a \\ x \in H}} f(x) = \lim\limits_{x \to a} f(x)$.

2. If $n = 1$ then we may obtain the one-sided limits, that is

   for $H = (-\infty,\, a)$: $\lim\limits_{\substack{x \to a \\ x \in H}} f(x) = \lim\limits_{x \to a-0} f(x)$,

   and for $H = (a, +\infty)$: $\lim\limits_{\substack{x \to a \\ x \in H}} f(x) = \lim\limits_{x \to a+0} f(x)$.

3. If $\exists \lim\limits_{x \to a} f(x) = A$, then for any set $H$ for which $a \in (H \cap \mathcal{D}_f)'$ follows that

   $$\lim_{\substack{x \to a \\ x \in H}} f(x) = A \,.$$

   The practically useful corollary of this statement is, that if we tend to a point „in two different way" and we obtain two different results, then the function has no limit at this point (two-way-method).

4. If $\lim\limits_{\substack{x \to a \\ x \in H}} f(x) = \lim\limits_{\substack{x \to a \\ x \in K}} f(x) = A$ then $\lim\limits_{\substack{x \to a \\ x \in H \cup K}} f(x) = A$.

## 4.3. The continuity of functions of type $\mathbb{R}^n \to \mathbb{R}^m$

We can define the continuity similarly to the one-variable case:

**4.13. Definition** Let $f \in \mathbb{R}^n \to \mathbb{R}^m$, $a \in D_f$. $f$ is continuous at $a$ if
$\forall \varepsilon > 0\ \exists \delta > 0\ \forall x \in B(a,\delta) \cap D_f: \quad f(x) \in B(f(a),\varepsilon)$.
Let us denote the set of functions that are continuous at $a$ by $C(a)$.

From the definition it follows immediately that

– if $a$ is an isolated point of $D_f$ then $f$ is continuous at $a$.

– if $a$ is an accumulation point of $D_f$ then

$$f \quad \text{is continuous at } a \quad \Leftrightarrow \quad \lim_{x\to a} f(x) = f(a).$$

**4.14. Definition** Let $f \in \mathbb{R}^n \to \mathbb{R}^m$. Then $f$ is continuous if it is continuous at every point of its domain, that is

$$\forall\, a \in D_f: \quad f \in C(a).$$

Using the results of Examples 4.5, the constant function, the identity function and the canonical projections are continuous.

**4.15. Theorem** *[Transference Principle for continuity] Using our notations:*

$$f \in C(a) \quad \Leftrightarrow \quad \forall\, x_k \in D_f \quad (k \in \mathbb{N}),\ \lim x_k = a: \quad \lim f(x_k) = f(a)\,.$$

The proof of this theorem is similar to the one-variable case.

Using the Transference Principle one can easily see that

1. $$f,\ g \in C(a),\ c \in \mathbb{R} \quad \Rightarrow \quad f+g,\ f-g,\ c\cdot f \in C(a)\,,$$

2. $$g \in C(a),\ f \in C(g(a)) \quad \Rightarrow \quad f \circ g \in C(a)\,,$$

3. $$f = (f_1,\,\dots,\,f_m) \in C(a) \Leftrightarrow f_i \in C(a) \quad (i = 1,\,\dots,\,m)\,.$$

Similarly to the one-variable case, one can prove the most important theorems for continuous functions defined on compact sets.

**4.16. Theorem** *[the compactness of the image]*
*Let $f \in \mathbb{R}^n \to \mathbb{R}^m$ be a continuous function and suppose that $D_f$ is compact. Then $R_f$ is compact.*

Before stating the following theorem, let us define the extreme values of an $n$-variable function:

**4.17. Definition** Let $f \in \mathbb{R}^n \to \mathbb{R}$. The minimum of $f$ is the minimal element of its range (if exists), that is

$$\min f := \min R_f = \min \{f(x) \mid x \in D_f\} = \min_{x \in D_f} f(x)\,.$$

The vector $a \in D_f$ is called the place of the minimum, if $f(a) = \min f$.

Respectively, the maximum of $f$ is the maximal element of its range (if exists), that is

$$\max f := \max R_f = \max \{f(x) \mid x \in D_f\} = \max_{x \in D_f} f(x).$$

The vector $a \in D_f$ is called the place of the maximum, if $f(a) = \max f$. These numbers are called the absolute (or global) extreme values, (absolute (or global) minimum, absolute (or global) maximum) of $f$.

**4.18. Theorem** *[the minimax theorem of Weierstrass] Let $f \in \mathbb{R}^n \to \mathbb{R}$ be a continuous function and $D_f$ be compact. Then $\exists \min f$ and $\exists \max f$.*

The definition of the uniform continuity is defined in a similar way as in the one-variable case:

**4.19. Definition** Let $f \in \mathbb{R}^n \to \mathbb{R}^m$. We say that $f$ is uniformly continuous if

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall x,\, y \in D_f,\ \|x - y\| < \delta : \quad \|f(x) - f(y)\| < \varepsilon\,.$$

**4.20. Theorem** *[theorem of Heine] Let $f \in \mathbb{R}^n \to \mathbb{R}^m$ be a continuous function and $D_f$ be compact. Then $f$ is uniformly continuous.*

## 4.4. Homeworks

1. Determine the following limits if they exist:

*a)* $\displaystyle\lim_{(x,y)\to(0,0)} xy \cdot \frac{x^2 - y^2}{x^2 + y^2}$ *b)* $\displaystyle\lim_{(x,y)\to(0,0)} \frac{3xy\sqrt{x} + xy^2}{x^2 + 2y^2}$

*c)* $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2 + y^2}{\sqrt{x^2 + y^2 + 4} - 2}$ *d)* $\displaystyle\lim_{(x,y)\to(0,0)} \frac{xy}{x - y}\,.$

2. Discuss the continuity of the following $\mathbb{R}^2 \to \mathbb{R}$ type functions:

$$a) \quad f(x,y) = \begin{cases} \dfrac{x-y}{x+y} & \text{if} \quad x+y \neq 0, \\ 0 & \text{if} \quad x+y = 0; \end{cases}$$

$$b) \quad f(x,y) = \begin{cases} \dfrac{x^2y^2}{x^2+y^2} & \text{if} \quad (x,y) \neq (0,0), \\ 0 & \text{if} \quad (x,y) = (0,0). \end{cases}$$

# 5. Lesson 5

## 5.1. The derivative of functions of type $\mathbb{R}^n \to \mathbb{R}^m$

**5.1. Definition** Let $f \in \mathbb{R}^n \to \mathbb{R}^m,\ a \in \operatorname{int} D_f$. We say that $f$ is differentiable at the point $a$ (denoted by $f \in D(a)$) if

$$\exists\, A \in \mathbb{R}^{m\times n}: \quad \lim_{h\to 0} \frac{f(a+h)-f(a)-A\cdot h}{||h||} = 0\,.$$

**5.2. Theorem** *The matrix $A$ in the above definition is unique.*

**Proof.** Suppose that the matrices $A, B \in \mathbb{R}^{m\times n}$ satisfy the definition. In this case

$$\lim_{h\to 0}\left(\frac{f(a+h)-f(a)-Ah}{||h||} - \frac{f(a+h)-f(a)-Bh}{||h||}\right) = 0-0=0\,.$$

After calulations we have

$$\lim_{h\to 0}\frac{(B-A)\cdot h}{||h||} = 0\,.$$

Let $h \to 0$ along the rays of the unit vectors

$$e_j = (0,\ldots,\overset{j}{\smile}{1},\ldots,0), \quad (j=1,\,\ldots,\,n)$$

that is let $h := t\cdot e_j$, where $t>0$, $t\to 0+0$. Since

$$||t\cdot e_j|| = |t|\cdot||e_j|| = t\cdot 1 = t\,,$$

so

$$\begin{aligned}
&\lim_{h\to 0}\frac{(B-A)\cdot h}{||h||} = 0 \quad\Rightarrow\quad \lim_{t\to 0+0}\frac{(B-A)\cdot t\cdot e_j}{||t\cdot e_j||} = 0 \quad\Rightarrow\\
&\quad\Rightarrow\quad \lim_{t\to 0+0}\frac{(B-A)\cdot t\cdot e_j}{t} = 0 \quad\Rightarrow\quad (B-A)\cdot e_j = 0 \quad\Rightarrow\\
&\quad\Rightarrow\quad B-A=0 \quad\Rightarrow\quad B=A.
\end{aligned}$$

□

**5.3. Definition** The matrix $A$ in the above definition is called the derivative (or: derivative matrix) of $f$ at the point $a$ and is denoted by $f'(a)$. So $f'(a) := A$.

**5.4. Remarks.**

1. If $n=m=1$, then we obtain the definition in the case $\mathbb{R}\to\mathbb{R}$, that was studied in Analysis-2.

2. In the case $n = 1$ the derivative can be defined equivalently as it was present in Analysis-2:

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

3. In the case $n \geq 2$ we cannot use the ratio of differences because the division with a vector is undefined.

**5.5. Theorem** *If $f \in D(a)$ then $f \in C(a)$.*

**Proof.**

$$f(a+h) - f(a) = \frac{f(a+h) - f(a) - f'(a) \cdot h}{||h||} \cdot ||h|| + f'(a) \cdot h \to 0 \quad (h \to 0)\,.$$

So $\lim_{h \to 0} f(a+h) = f(a)$ which implies $f \in C(a)$. □

In the following theorem we state some differentiation rules without proof.

**5.6. Theorem** *1. Let $f, g \in \mathbb{R}^n \to \mathbb{R}^m,\ f, g \in D(a)$. Then $f + g \in D(a)$ and*

$$(f+g)'(a) = f'(a) + g'(a)$$

*(in the sense of matrix addition).*

*2. Let $f \in \mathbb{R}^n \to \mathbb{R}^m,\ \lambda \in \mathbb{R},\ f \in D(a)$. Then $\lambda f \in D(a)$ and*

$$(\lambda f)'(a) = \lambda \cdot f'(a)$$

*(in the sense of matrix scalar multiplication).*

*3. (Chain Rule) Let $g \in \mathbb{R}^n \to \mathbb{R}^m,\ g \in D(a),\ f \in \mathbb{R}^m \to \mathbb{R}^p,\ f \in D(g(a))$. Then $f \circ g \in D(a)$ and*

$$(f \circ g)'(a) = f'(g(a)) \cdot g'(a)$$

*(in the sense of matrix multiplication).*

## 5.2. Partial Derivatives

What are the entries of the derivative matrix? This is a natural question. In this section we prepare the answer.

**5.7. Definition** Let $f \in \mathbb{R}^n \to \mathbb{R}^m,\ a = (a_1, \ldots, a_n) \in \operatorname{int} D_f$ and $j \in \{1, \ldots, n\}$. Define the following auxiliary function

$$g_{a,j}(x) := f(a_1, \ldots, a_{j-1}, x, a_{j+1}, \ldots, a_n) \qquad (x \in B(a_j, r))\,,$$

where $r$ denotes the radius for which $B(a,r) \subseteq D_f$.

The column matrix $g'_{a,j}(a_j) \in \mathbb{R}^{m\times 1}$ – if it exists – is called the $j$-th partial derivative (more precisely: the partial derivative by the $j$-th variable) of the function $f$ at the point $a$. Its notations are:

$$\partial_j f(a), \quad \text{or} \quad f'_{x_j}(a), \quad \text{or} \quad \left(\frac{\partial f}{\partial x_j}\right)_{x=a}, \quad \text{or} \quad \left(\frac{\partial f(x)}{\partial x_j}\right)_{x=a}.$$

**5.8. Remarks.**

1. Roughly speaking we can compute the the $j$-th partial derivative, if we fix every variable except the $j$-th one and then differentiate the obtained one-variable function at $a_j$.
2. By the isomorphism between $\mathbb{R}^m$ and $\mathbb{R}^{m\times 1}$ we can say that the partial derivative is a vector in $\mathbb{R}^m$. Especially in the case $m=1$ ($f$ is a scalar-valued function) – because of the isomorphism between $\mathbb{R}$ and $\mathbb{R}^{1\times 1}$ – we can say that the partial derivative is a number. We will use these representations in the followings.

**5.9. Definition** Using the notations of the previous definition, let $D \subseteq \mathbb{R}^n$ denote the set of all points of $D_f$ where the $j$-th partial derivative exists. Suppose that $D \neq \emptyset$. Then the function

$$\partial_j f \;:\; D \to \mathbb{R}^m, \quad a \mapsto \partial_j f(a)$$

is called the $j$-th partial derivative function of $f$.

It is obvious that the partial derivative function is of type $\mathbb{R}^n \to \mathbb{R}^m$, and in the case $m=1$ it is of type $\mathbb{R}^n \to \mathbb{R}$.

## 5.3. Homeworks

1. Using the definition prove that the following functions are differentiable at the given point $(a,b)$, and compute the derivatives:

   a) $f:\mathbb{R}^2 \to \mathbb{R}, \quad f(x,y) = x^3 + xy - 2y, \quad (a,b) = (2,-1)$;
   b) $f:\mathbb{R}^2 \to \mathbb{R}^2, \quad f(x,y) = (x^2y + 5y,\; x^2 - xy) \quad (a,b) = (2,-1)$.

2. Determine the partial derivatives of the following $\mathbb{R}^2 \to \mathbb{R}$ type functions:

$$a)\ f(x,y) = x^2 - 5xy + 3y^2 - 6x + 7y + 8; \qquad b)\ f(x,y) = \arcsin\frac{x}{y};$$

$$c)\ f(x,y) = \frac{xy}{x+y}; \qquad d)\ f(x,y) = \sqrt{x^3 - 5x^2y + y^4};$$

$$e)\ f(x,y) = e^x\cos y - x\ln y; \qquad f)\ f(x,y) = \operatorname{arc\,tg}\frac{1-x}{1-y};$$

$$g)\ f(x,y) = \frac{e^{2x-3y}}{2x-3y}; \qquad h)\ f(x,y) = \frac{x\cdot \operatorname{tg}\, x}{e^{xy}}.$$

# 6. Lesson 6

## 6.1. The entries of the derivative matrix

In the previous lesson we have defined the partial derivatives. Using this concept we can determine the entries of the derivative matrix.

First we will give another form of the partial derivative.

Let $f \in \mathbb{R}^n \to \mathbb{R}^m$, $a = (a_1, \ldots, a_n) \in \operatorname{int} D_f$, $j \in \{1, \ldots, n\}$ and $g_{a,j}$ be the auxiliary function defined in the previous section. Let $F = F_{a,j}$ be the following other auxiliary function:

$$F(t) := f(a + te_j) \qquad (t \in \mathbb{R}, a + te_j \in D_f)\,,$$

where $e_j$ denotes the $j$-th standard unit vector in $\mathbb{R}^n$. Then

$$\begin{aligned} F(t) =& f(a + te_j) = \\ &= f(a_1 + t \cdot 0,\, \ldots,\, a_{j-1} + t \cdot 0,\, a_j + t \cdot 1,\, a_{j+1} + t \cdot 0,\, \ldots,\, a_n + t \cdot 0) = \\ &= f(a_1,\, \ldots,\, a_{j-1},\, a_j + t,\, a_{j+1},\, \ldots,\, a_n) = g_{a,j}(a_j + t)\,, \end{aligned}$$

and $F$ is defined for $|t| < r$, where $r$ denotes the radius, for which $B(a, r) \subseteq D_f$.

Using the Chain Rule we deduce that

$$F'(t) = g'_{a,j}(a_j + t) \cdot 1 \quad \text{consequently} \quad F'(0) = g'_{a,j}(a_j) = \partial_j f(a)\,.$$

After these preliminaries we can discuss the columns of $f'(a)$.

**6.1. Theorem** *Let $f \in \mathbb{R}^n \to \mathbb{R}^m$, $f \in D(a)$, $j \in \{1, \ldots, n\}$. Then $\exists \partial_j f(a)$ and it is identical with the $j$-th column of $f'(a)$.*

**Proof.** $f \in D(a)$ implies that $\lim\limits_{h \to 0} \dfrac{f(a + h) - f(a) - f'(a) \cdot h}{\|h\|} = 0$. Apply this relation with the vectors $h = t \cdot e_j$ where $e_j$ is the $j$-th standard unit vector, and $t \in \mathbb{R}$, $t \to 0$. Then

$$\begin{aligned} 0 &= \lim_{t \to 0} \frac{f(a + t \cdot e_j) - f(a) - f'(a) \cdot t \cdot e_j}{\|t \cdot e_j\|} = \\ &= \lim_{t \to 0} \frac{f(a + t \cdot e_j) - f(a) - f'(a) e_j \cdot t}{|t| \cdot \|e_j\|} = \\ &= \lim_{t \to 0} \frac{f(a + t \cdot e_j) - f(a) - (\text{the } j\text{-th column of } f'(a)) \cdot t}{|t|} \end{aligned}$$

This means – by the definition of the derivative – that $F \in D(0)$, and that

$$\partial_j f(a) = F'(0) = \text{ the } j\text{-th column of } f'(a)\,.$$

$\square$

The following theorem speaks about the rows of the derivative matrix.

**6.2. Theorem** *Let $f \in \mathbb{R}^n \to \mathbb{R}^m$, $f = (f_1, \dots, f_m)$ and $a \in int D_f$. Then*

$$f \in D(a) \quad \Leftrightarrow \quad f_i \in D(a)\ (i = 1, \dots, m).$$

*In this case:*

$$f_i'(a) = \text{ the } i\text{-th row of } f'(a) \quad (i = 1, \dots, m).$$

**Proof.** Let $A \in \mathbb{R}^{m\times n}$. Then the fact

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - A\cdot h}{\|h\|} = 0$$

is equivalent with

$$\lim_{h\to 0} \frac{f_i(a+h) - f_i(a) - (Ah)_i}{\|h\|} = 0$$

(see: limit by coordinates). But $(Ah)_i = (\text{the } i\text{-th row of } A)\cdot h$, so the above relation is equivalent with

$$\lim_{h\to 0} \frac{f_i(a+h) - f_i(a) - (\text{the } i\text{-th row of } A)\cdot h}{\|h\|} = 0.$$

Using these equivalencies in both directions, the statement of the theorem follows immediately. □

Using the two previous theorems we obtain:

$$(f'(a))_{ij} = \text{ the } j\text{-th column of the } i\text{-th row } = \partial_j f_i(a).$$

**6.3. Remark.** The derivative matrix is:

$$f'(a) = \begin{bmatrix} \partial_1 f_1(a) & \partial_2 f_1(a) & \dots & \partial_n f_1(a) \\ \partial_1 f_2(a) & \partial_2 f_2(a) & \dots & \partial_n f_2(a) \\ \vdots & & & \vdots \\ \partial_1 f_m(a) & \partial_2 f_m(a) & \dots & \partial_n f_m(a) \end{bmatrix} \in \mathbb{R}^{m\times n}.$$

**6.4. Definition** Let $f \in \mathbb{R}^n \to \mathbb{R}$, $f \in D(a)$. The vector

$$\operatorname{grad} f(a) := \nabla f(a) := (\partial_1 f(a), \partial_2 f(a), \dots, \partial_n f(a)) \in \mathbb{R}^n$$

is called the gradient vector (or simply: gradient) of $f$ at the point $a$.

Obviously the gradient vector $\nabla f(a)$ is the vector representation of the derivative matrix $f'(a) \in \mathbb{R}^{1\times n}$.

## 6.2. Connection between the derivatives and the partial derivatives

In the previous section we have proved that the differentiability of a function at a point implies the existence of all the partial derivatives at this point. The converse statement is not true as the following example shows.

**6.5. Example**
Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the following function:

$$f(x, y) := \begin{cases} 1 & \text{if } xy = 0 \\ 0 & \text{if } xy \neq 0 \end{cases}$$

Then $\partial_1 f(0, 0) = \partial_2 f(0, 0) = 0$, but $f \notin C(0, 0)$ so $f \notin D(0, 0)$.

Using some further assumptions the existence of partial derivatives can imply the differentiability as will be stated – without proof – in the following theorem.

**6.6. Theorem** *Let $f \in \mathbb{R}^n \to \mathbb{R}^m$, $a \in int D_f$. Suppose that*

1. $\exists r > 0\ \forall x \in B(a, r) :\ \exists\ \partial_j f(x)\ (j = 1, \dots, n)$ *and*
2. $\partial_f \in C(a)$.

*Then $f \in D(a)$.*

## 6.3. Directional Derivatives

The partial derivatives can be regarded as the derivatives of the functions restricted to a line through the point $a$, and parallel with one of the coordinate axis. If we take a general direction instead of the directions of coordinate axis then we obtain the concept of directional derivative.

**6.7. Definition** Let $f \in \mathbb{R}^n \to \mathbb{R}^m$, $a \in \operatorname{int} D_f$, $e \in \mathbb{R}^n$, $\|e\| = 1$. Let $F = F_{a,e}$ be the following other auxiliary function:

$$F(t) := f(a + te) \qquad (t \in \mathbb{R}, a + te \in D_f)$$

Then the directional derivative of $f$ at the point $a$ along the direction $e$ is defined as

$$\partial_e f(a) := F'(0) = \left(\frac{d}{dt} f(a + te)\right)_{t=0} .$$

In many cases the directional derivative can be computed with the help of the derivative.

**6.8. Theorem** *Let $f \in \mathbb{R}^n \to \mathbb{R}^m$, $f \in D(a)$. Then for any $e \in \mathbb{R}^n$, $\|e\| = 1$:*

$$\partial_e f(a) = f'(a) \cdot e$$

*in the sense of matrix-vector product.*

**Proof.** Let $g(t) = a + te$ ($t \in \mathbb{R}$). Then $g'(t) = e$ and using the Chain Rule we obtain

$$\begin{aligned} \partial_e f(a) = \left( \frac{d}{dt} f(a+te) \right)_{t=0} &= \left( f'(a+te) \cdot g'(t) \right)_{t=0} = \\ = \left( f'(a+te) \cdot e \right)_{t=0} &= f'(a) \cdot e \,. \end{aligned}$$

□

**6.9. Remark.** The direction is often given not by a unit vector but by another way (see Homework 3.). In this cases we have to determine a unit vector that shows in the given direction.

## 6.4. Homeworks

1. Discuss the differentiability of the function (i.e. at which point of its domain it is differentiable, and what are the derivatives at these points):

$$f : \mathbb{R}^2 \to \mathbb{R}, \qquad f(x,y) = \sqrt[3]{xy}$$

2. Prove that the following function is differentiable, and determine its derivative:

$$f : \mathbb{R}^2 \to \mathbb{R}^2, \qquad f(x,y) = (\operatorname{arc\,tg} \frac{x}{x^2+y^2+1}, \ \cos(x^3 - 4xy)) \,.$$

3. Determine the directional derivatives in the following case:

$$f : \mathbb{R}^2 \to \mathbb{R}, \qquad f(x,y) = x^3 + xy - 2y^3$$

at the point $P_0(2,1)$ along the following directions

$a)$ $v = (-2,3)$; $b)$ $\alpha = 330°$;

$c)$ The direction from $A(1,0)$ to $B(4,4)$.

# 7. Lesson 7

## 7.1. Higher order derivatives

In this section we will study the higher order derivatives of functions of type $\mathbb{R}^n \to \mathbb{R}$. Similarly to the one-variable case the second order derivative is defined as the derivative of the derivative function.

**7.1. Definition** Let $f \in \mathbb{R}^n \to \mathbb{R},\ a \in \operatorname{int} D_f$. We say that $f$ is 2 times differentiable at $a$ (its notation is: $f \in D^2(a)$) if

$$\exists\, r > 0\ \forall\, x \in B(a,r): \quad f \in D(x) \quad \text{and} \quad f' \in D(a)\,.$$

The derivative function $f'$ can be regarded as a vector valued function with the coordinate functions $\partial_j f$ $(j = 1, \ldots, n)$. So the equivalent definition of the 2 times differentiability can be given as follows.

**7.2. Definition** Let $f \in \mathbb{R}^n \to \mathbb{R},\ a \in \operatorname{int} D_f$. We say that $f$ is 2 times differentiable at $a$ if

$$\exists\, r > 0\ \forall\, x \in B(a,r): \quad f \in D(x) \quad \text{and} \quad \partial_j f \in D(a) \quad (j = 1, \ldots, n)\,.$$

Suppose that $f \in D^2(a)$. Since its derivative function is

$$f' = (\partial_1 f,\, \partial_2 f,\, \ldots,\, \partial_n f) : \mathbb{R}^n \to \mathbb{R}^n\,,$$

its second derivative is the derivative matrix of $f'$ at $a$:

$$f''(a) = \left(f'\right)'(a) = \begin{bmatrix} \partial_1\partial_1 f(a) & \partial_2\partial_1 f(a) & \ldots & \partial_n\partial_1 f(a) \\ \partial_1\partial_2 f(a) & \partial_2\partial_2 f(a) & \ldots & \partial_n\partial_2 f(a) \\ \vdots & & & \vdots \\ \partial_1\partial_n f(a) & \partial_2\partial_n f(a) & \ldots & \partial_n\partial_n f(a) \end{bmatrix} \in \mathbb{R}^{n\times n}\,.$$

This matrix is called the Hesse-matrix of $f$ at $a$. The $ij$-th entry of the Hesse-matrix is

$$\left(f''(a)\right)_{ij} = \partial_j\partial_i f(a) \quad (i,j = 1, \ldots, n)\,.$$

**7.3. Definition** The entries of the Hesse-matrix $f''(a)$ are called the second order partial derivatives of $f$ at $a$.

**7.4. Remark.** We have defined the second order partial derivatives only for the functions that are 2 times differentiable at $a$. The concept of the second order partial derivative can be defined in more general case but we will use it only for 2 times differentiable functions.

**7.5. Theorem** *[Theorem of Young] If $f \in D^2(a)$, then the Hesse-matrix $f''(a)$ is symmetric that is*

$$\partial_j \partial_i f(a) = \partial_i \partial_j f(a) \quad (i, j = 1, \ldots, n)\,.$$

**7.6. Definition** Let $f \in \mathbb{R}^n \to \mathbb{R}$ and suppose that the set

$$D_{f''} := \left\{ x \in \operatorname{int} D_f \mid f \in D^2(x) \right\}$$

is nonempty. Then the function

$$f'' \,:\, D_{f''} \to \mathbb{R}^{n \times n}, \quad x \mapsto f''(x)$$

is called the second derivative function (or simply: the second derivative) of $f$. The functions

$$\partial_j \partial_i f \,:\, D_{f''} \to \mathbb{R}, \quad x \mapsto \partial_j \partial_i f(x) \quad (i, j = 1, \ldots, n)$$

are called the second order partial derivative functions (or simply: the second order partial derivatives) of $f$.

If we want to define the 3 times differentiability of $f$ at a point $a \in \operatorname{int} D_f$ (denoted by $f \in D^3(a)$ ) then we have to suppose that $f''$ is differentiable at $a$. Since the coordinate functions of $f''$ are the second order partial derivatives, so

$$f \in D^3(a) \;\Leftrightarrow\; f'' \in D(a) \quad \Leftrightarrow \quad \partial_j \partial_i f \in D(a) \quad (i, j = 1, \ldots, n)\,.$$

**7.7. Definition** Suppose that $f \in D^3(a)$. Then the numbers

$$\partial_k \partial_j \partial_i f(a) \quad (i, j, k = 1, \ldots, n)$$

are called the 3-rd order partial derivatives of $f$ at $a$. The 3-array with entries

$$\left(f'''(a)\right)_{ijk} = \partial_k \partial_j \partial_i f(a) \quad (i, j, k = 1, \ldots, n)$$

is called the 3-rd order derivative of $f$ at $a$.

In similar way – recursively – we can define the $k$-th derivative and the $k$-th order partial derivatives for $k = 4, 5, \ldots$. Thus these concepts are defined for any $k \in \mathbb{N}$. We agree that the 0-th derivative is the function itself.

Some notations:

- $f \in D^k(a)$: $f$ is $k$ times differentiable at $a$.
- $f^{(k)}(a)$: the $k$-th derivative of $f$ at $a$.
- $\partial_{j_k} \partial_{j_{k-1}} \ldots \partial_{j_2} \partial_{j_1} f(a) \quad (j_1, \ldots, j_k = 1, \ldots, n)$:
  the $k$-th order partial derivatives of $f$ at $a$. Their number is $n^k$.

So $f^{(k)}(a) \in \mathbb{R}^{n\times n\times \ldots \times n} = \mathbb{R}^n_k$ is a $k$-array with the entries

$$\left(f^{(k)}(a)\right)_{j_1,\ldots,j_k} = \partial_{j_k}\partial_{j_{k-1}}\,\ldots\,\partial_{j_2}\partial_{j_1} f(a) \quad (j_1,\,\ldots,\,j_k = 1,\,\ldots,\,n)\,. \tag{7.1}$$

Applying several times the Theorem of Young we obtain the following theorem.

**7.8. Theorem** *If $f \in D^k(a)$ then the $k$-array $f^{(k)}(a)$ is symmetric in the following sense.*

*Let $j_1,\,\ldots,\,j_k \in \{1,\,\ldots,\,n\}$ and the finite sequence $p_1,\,\ldots,\,p_k$ be a permutation (may be: permutation with repetition) of the finite sequence $j_1,\,\ldots,\,j_k$. Then*

$$\left(f^{(k)}(a)\right)_{p_1,\ldots,p_k} = \left(f^{(k)}(a)\right)_{j_1,\ldots,j_k}$$

*that is*

$$\partial_{p_k}\partial_{p_{k-1}}\,\ldots\,\partial_{p_2}\partial_{p_1} f(a) = \partial_{j_k}\partial_{j_{k-1}}\,\ldots\,\partial_{j_2}\partial_{j_1} f(a)\,.$$

**7.9. Definition** Let $f \in \mathbb{R}^n \to \mathbb{R}$, $a \in \operatorname{int} D_f$. We say that $f$ is $k$ times continuously differentiable at $a$ (its notation is: $f \in C^k(a)$) if

$$\exists\, r > 0\ \forall\, x \in B(a,r): \quad f \in D^k(x) \quad \text{and} \quad f^{(k)} \in C(a)\,.$$

Naturally, the definition is equivalent with the continuity of the $k$-th order partial derivative functions at $a$.

## 7.2. Taylor's Formula

**7.10. Definition (Taylor's polynomial)** Let $m \in \mathbb{N} \cup \{0\}$, $f \in \mathbb{R}^n \to \mathbb{R}$, $f \in D^m(a)$. The $n$-variable polynomial

$$T_m(x) := f(a) + \frac{f'(a)(x-a)}{1!} + \frac{f''(a)(x-a)^2}{2!} + \ldots + \frac{f^{(m)}(a)(x-a)^m}{m!} =$$
$$= f(a) + \sum_{k=1}^{m} \frac{f^{(k)}(a)(x-a)^k}{k!} \qquad (x \in \mathbb{R}^n)$$

is called the $m$-th Taylor-polynomial of $f$ at the center $a$.

**7.11. Remarks.**

1. For the meaning of the terms of the above sum we remind the Reader of the definition of the symbol $Ax^k$ where $A$ is a $k$-array and $x$ is a vector (see: Definition 1.7 in Lesson 1).

2. It is obvious that the degree of $T_m$ is at most $m$ and that $T_m(a) = f(a)$.

In the followings we will use the abbreviation $h = x - a$. To approximate $f$ with the help of its Taylor-polynomials and to prove the Taylor Formula, we need the concept of the line segment in $\mathbb{R}^n$ and a theorem about the $k$-th derivative of an auxiliary function.

**7.12. Definition** Let $a \in \mathbb{R}^n,\ h \in \mathbb{R}^n \setminus \{0\}$. The set

$$[a,\, a+h] := \{a + th \in \mathbb{R}^n \mid 0 \le t \le 1\} \subset \mathbb{R}^n$$

is called a closed line segment in $\mathbb{R}^n$. $a$ is the starting point and $a+h$ is the terminal point of the line segment.

**7.13. Theorem** *Let $f \in \mathbb{R}^n \to \mathbb{R}$, $k \in \mathbb{N} \cup \{0\}$. Let the closed line segment $[a,\, a+h]$ be a subset of $\operatorname{int} D_f$ and suppose that $f$ is $k$ times differentiable at any point of $[a,\, a+h]$. Let*

$$F \in \mathbb{R} \to \mathbb{R}, \quad F(t) := f(a+th) \qquad (t \in \mathbb{R},\, a+th \in D_f)\,.$$

*Then $[0,\,1] \subset \operatorname{int} D_f$, $F$ is $k$ times differentiable at any point of the closed interval $[0,\,1]$ and*

$$F^{(k)}(t) = f^{(k)}(a+th)h^k \quad (t \in [0,\,1]) \tag{7.2}$$

*in the sense of the symbol $Ax^k$ (see Definition 1.7 in Lesson 1).*

**Proof.** The precise proof of this formula requires mathematical induction. For simplicity we will prove it only for the cases $n = 1$ and $n = 2$. One will see from these two cases the general inductional step.

*Proof of* (7.2) *in the case $k = 1$:*
$F$ is a composition of functions $f$ and $t \mapsto a + th$. Using the Chain Rule we obtain:

$$F'(t) = f'(a+th) \cdot h = \sum_{j=1}^{n} \partial_j f(a+th) \cdot h_j = f'(a+th)h^1\,.$$

*Proof of* (7.2) *in the case $k = 2$ (using that it is true for $k = 1$):*
$F'$ is an $n$-term sum of functions $t \mapsto \partial_j f(a+th) \cdot h_j$. The $j$-th term is a scalar multiple of the composition of functions $\partial_j f$ and $t \mapsto a + th$. Applying the previous result (case $k = 1$) for $\partial_j f$ instead of $f$ we obtain:

$$F''(t) = (F'(t))' = \frac{d}{dt}\left(\sum_{j=1}^{n} \partial_j f(a+th) \cdot h_j\right) = \sum_{j=1}^{n} h_j \cdot \frac{d}{dt}\left(\partial_j f(a+th)\right) =$$

$$= \sum_{j=1}^{n} h_j \cdot \sum_{i=1}^{n} \partial_i \partial_j f(a+th) \cdot h_i = \sum_{i,j=1}^{n} \partial_i \partial_j f(a+th) \cdot h_i \cdot h_j = f''(a+th)h^2\,.$$

□

**7.14. Theorem** *[Taylor's formula]*

*Let $f \in \mathbb{R}^n \to \mathbb{R}$, $a \in D_f$, $h \in \mathbb{R}^n$, $h \neq 0$, $m \in \mathbb{N} \cup \{0\}$. Suppose that $f$ is $m+1$ times differentiable at any point of the closed line segment*

$$[a,\, a+h] := \{a + th \in \mathbb{R}^n \mid 0 \leq t \leq 1\}\,.$$

*(Remember that it requires $[a,\, a+h] \subseteq int\, D_f$).*

*Then there exists $\vartheta \in \mathbb{R}$, $0 < \vartheta < 1$ such that*

$$f(a+h) - T_m(a+h) = \frac{f^{(m+1)}(a+\vartheta h)h^{m+1}}{(m+1)!}$$

*that is*

$$f(a+h) = f(a) + \sum_{k=1}^{m} \frac{f^{(k)}(a)h^k}{k!} + \frac{f^{(m+1)}(a+\vartheta h)h^{m+1}}{(m+1)!} \tag{7.3}$$

**Proof.** Let us define the auxiliary function

$$F \in \mathbb{R} \to \mathbb{R}, \quad F(t) := f(a+th) \qquad (t \in \mathbb{R},\, a+th \in D_f)\,.$$

Then – by the previous theorem – $F$ satisfies the assumptions of the one-variable Taylor Formula (see: Analysis-2) at the center 0. Applying the one-variable Taylor Formula for approximation of $F(1)$ we have:

$$\begin{aligned} \exists\, \vartheta \in (0,\, 1): \quad F(1) &= \sum_{k=0}^{m} \frac{F^{(k)}(0)}{k!} \cdot (1-0)^k + \frac{F^{(m+1)}(\vartheta)}{(m+1)!} \cdot (1-0)^{m+1} = \\ &= F(0) + \sum_{k=1}^{m} \frac{F^{(k)}(0)}{k!} + \frac{F^{(m+1)}(\vartheta)}{(m+1)!}\,. \end{aligned} \tag{7.4}$$

Using

$$F(1) = f(a+h), \quad F(0) = f(a)$$

and the result of the previous theorem for $F^{(k)}$ we have

$$\begin{gathered} F^{(k)}(0) = f^{(k)}(a+0h)h^k = f^{(k)}(a)h^k \qquad (k = 1,\, \ldots,\, m) \\ \text{and} \quad F^{(m+1)}(\vartheta) = f^{(m+1)}(a+\vartheta h)h^{m+1}\,. \end{gathered}$$

Substituting this result into (7.4) we obtain the statement of the theorem. $\square$

The multi-index form of Taylor's Formula

Since $f^{(k)}(a)$ and $f^{(m+1)}(a+\vartheta h)$ are symmetric $k$-arrays, we can apply the multi-index form of $Ax^k$ (see in Lesson 1). Using (7.1) we have:

$$f^{(k)}(a)h^k = \sum_{\substack{i\in\mathbb{N}_0^n \\ |i|=k}} \frac{k!}{i!}\cdot (f^{(k)}(a))_i \cdot h^i =$$

$$= \sum_{\substack{i\in\mathbb{N}_0^n \\ |i|=k}} \frac{k!}{i!}\cdot (f^{(k)}(a))_{\underbrace{n,\dots,n}_{i_n \text{ times}},\underbrace{n-1,\dots,n-1}_{i_{n-1} \text{ times}},\dots,\underbrace{1,\dots,1}_{i_1 \text{ times}}} \cdot h^i =$$

$$= \sum_{\substack{i\in\mathbb{N}_0^n \\ |i|=k}} \frac{k!}{i!}\cdot \underbrace{\partial_1\dots\partial_1}_{i_1 \text{ times}}\underbrace{\partial_2\dots\partial_2}_{i_2 \text{ times}}\dots\underbrace{\partial_n\dots\partial_n}_{i_n \text{ times}} f(a)\cdot h^i =$$

$$= \sum_{\substack{i\in\mathbb{N}_0^n \\ |i|=k}} \frac{k!}{i!}\cdot \partial^i f(a)\cdot h^i\,,$$

where

$$\partial^i f(a) := \underbrace{\partial_1\dots\partial_1}_{i_1 \text{ times}}\underbrace{\partial_2\dots\partial_2}_{i_2 \text{ times}}\dots\underbrace{\partial_n\dots\partial_n}_{i_n \text{ times}} f(a)\,.$$

$f^{(m+1)}(a+\vartheta h)$ can be rewritten into similar form.

Let us substitute these results into (7.3):

$$f(a+h) = f(a) + \sum_{k=1}^{m}\frac{f^{(k)}(a)h^k}{k!} + \frac{f^{(m+1)}(a+\vartheta h)h^{m+1}}{(m+1)!} =$$

$$= f(a) + \sum_{k=1}^{m}\frac{1}{k!}\cdot\sum_{\substack{i\in\mathbb{N}_0^n \\ |i|=k}}\frac{k!}{i!}\cdot\partial^i f(a)\cdot h^i + \frac{1}{(m+1)!}\cdot\sum_{\substack{i\in\mathbb{N}_0^n \\ |i|=m+1}}\frac{(m+1)!}{i!}\cdot\partial^i f(a+\vartheta h)\cdot h^i\,.$$

Simplifying with $k!$ and with $(m+1)!$ we obtain the multi-index form of Taylor's Formula:

$$f(a+h) = f(a) + \sum_{k=1}^{m}\sum_{\substack{i\in\mathbb{N}_0^n \\ |i|=k}}\frac{\partial^i f(a)}{i!}\cdot h^i + \sum_{\substack{i\in\mathbb{N}_0^n \\ |i|=m+1}}\frac{\partial^i f(a+\vartheta h)}{i!}\cdot h^i\,.$$

The Mean Value Theorem

As an important corollary of the Taylor Formula we state the following theorem.

**7.15. Theorem** *[Mean Value Theorem]*

*Let $f\in\mathbb{R}^n\to\mathbb{R}$, $a\in D_f$, $h\in\mathbb{R}^n$, $h\neq 0$. Suppose that $f$ is differentiable at any point of the closed line segment $[a,\,a+h]$ (it requires $[a,\,a+h]\subseteq int\,D_f$). Then*

$$\exists\,\vartheta\in(0,1):\quad f(a+h)-f(a) = f'(a+\vartheta h)\cdot h\,.$$

**Proof.** Apply Taylor's Formula for $m=0$. □

## 7.3. Homeworks

1. Determine the all order partial derivatives of the following function and check the theorem of Young:

$$f:\mathbb{R}^2\to\mathbb{R}, \qquad f(x,y)=3xy^3+2x^2y-xy\,.$$

2. Let $a\in\mathbb{R}$ be a nonzero constant. Prove that the function

$$u(x,t)=\frac{1}{2a\cdot\sqrt{\pi t}}\cdot e^{-\frac{x^2}{4a^2t}} \qquad (x\in\mathbb{R},\ t>0)$$

satisfies the following partial differential equation

$$\partial_2 u=a^2\cdot\partial_1\partial_1 u\,.$$

3. Using the Taylor-formula rearrange the following polynomial by the powers $(x+1)^i(y-1)^j$:

$$f(x,y)=x^3+x^2y-2xy^2-xy+y \qquad ((x,y)\in\mathbb{R}^2)\,.$$

4. Determine the second order Taylor-polynomial of the function

$$f(x,y)=\frac{\cos x}{\cos y} \qquad ((x,y)\in\mathbb{R}^2,\ \cos y\neq 0)\,.$$

if the center is the origin.

# 8. Lesson 8

## 8.1. Local extreme values: the First Derivative Test

In connection with the Weierstrass-theorem (see: Theorem 4.18) we have defined the (global or absolute) extreme values of a function. Now we will discuss the so called local extrema.

**8.1. Definition** Let $f \in \mathbb{R}^n \to \mathbb{R}$, $a \in D_f$. We say that $f$ has at $a$

1. local minimum $\overset{\text{df}}{\Leftrightarrow} \exists\, r > 0 \;\forall\, x \in B(a, r) \cap D_f : \; f(x) \geq f(a)$;
2. strict local minimum $\overset{\text{df}}{\Leftrightarrow} \exists\, r > 0 \;\forall\, x \in B(a, r) \cap D_f \setminus \{a\} : \; f(x) > f(a)$;
3. local maximum $\overset{\text{df}}{\Leftrightarrow} \exists\, r > 0 \;\forall\, x \in B(a, r) \cap D_f : \; f(x) \leq f(a)$;
4. strict local maximum $\overset{\text{df}}{\Leftrightarrow} \exists\, r > 0 \;\forall\, x \in B(a, r) \cap D_f \setminus \{a\} : \; f(x) < f(a)$;

Here $a$ is the place of the local extremum and $f(a)$ is the local extreme value.

**8.2. Theorem** *[First Derivative Test] Let $f \in \mathbb{R}^n \to \mathbb{R}$, $f \in D(a)$ and suppose that $f$ has a local extremum at $a$.*

*Then $f'(a) = 0$ (or: $\nabla f(a) = 0$).*

**Proof.** Let us introduce the following auxiliary functions for $j \in \{1, \ldots, n\}$:

$$g_j(u) := f(a_1, \ldots, a_{j-1}, u, a_{j+1}, \ldots a_n) \qquad (u \in \mathbb{R},\; (a_1, \ldots, a_{j-1}, u, a_{j+1}, \ldots a_n) \in D_f).$$

Since $f$ has local extremum at $a$, then $g_j$ has the same type of local extremum at $a_j$. Applying the First Derivative Test for one-variable functions (see: Analysis-2) and the definition of the partial derivative we have

$$\partial_j f(a) = g_j'(a_j) = 0 \quad (j = 1, \;, \ldots, n)\,.$$

Consequently $f'(a) = \big[\partial_1 f(a) \ldots \partial_n f(a)\big] = 0.$ ☐

**8.3. Remarks.**

1. The above theorem is often mentioned as the first order necessary condition of local extrema.
2. The equation $f'(x) = 0$ is equivalent with the scalar equation system

$$\begin{gathered}\partial_1 f(x_1, x_2, \ldots, x_n) = 0 \\ \vdots \\ \partial_n f(x_1, x_2, \ldots, x_n) = 0\end{gathered}$$

   The roots of this system are called the stationary points of $f$.

## 8.2. Quadratic Forms

To formulate the second order conditions of the local extrema we need a short study of quadratic forms.

**8.4. Definition** Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix, that is a symmetric 2-array. The function

$$Q : \mathbb{R}^n \to \mathbb{R}, \qquad Q(x) := Ax^2 = \sum_{i,j=1}^{n} a_{ij} \cdot x_i \cdot x_j$$

is called quadratic form determined by the symmetric matrix $A$. $A$ is called the matrix of $Q$.

**8.5. Remarks.**

1. The connection between the $n \times n$ symmetric matrices and the quadratic forms is one-to-one.
2. The quadratic forms are exactly the homogeneous $n$-variable polynomials. This means that they are polynomials whose each term is of second degree.
3. From the definition it follows immediately that

$$Q(\lambda x) = \lambda^2 \cdot Q(x) \quad (\lambda \in \mathbb{R},\, x \in \mathbb{R}^n)\,.$$

**8.6. Theorem** *Let $Q : \mathbb{R}^n \to \mathbb{R}$ be a quadratic form. Then there exist constants $m$, $M \in \mathbb{R}$ such that*

$$m \cdot \|x\|^2 \le Q(x) \le M \cdot \|x\|^2 \qquad (x \in \mathbb{R}^n)\,.$$

**Proof.** The quadratic forms – because of their construction – are continuous functions. Let us restrict $Q$ to the compact set

$$H := \{x \in \mathbb{R}^n \mid \|x\| = 1\}\,,$$

and apply the Weierstrass minimax theorem (Theorem 4.18). Thus $Q_{|H}$ attains its minimal and maximal values. Denote by $m$ the minimal value and by $M$ the maximal value of $Q_{|H}$. Let $x \in \mathbb{R}^n$, $x \neq 0$. Then $\dfrac{x}{\|x\|} \in H$, so

$$m \le Q\left(\frac{x}{\|x\|}\right) \le M \quad \text{that is} \quad m \le \frac{1}{\|x\|^2} \cdot Q(x) \le M$$

which implies immediately the statement of the theorem for $x \neq 0$. The case $x = 0$ is trivial. □

**8.7. Remark.** It can be proved (see: Linear Algebra) that the above defined $m$ is the minimal and $M$ is the maximal eigenvalue of the matrix $A$.

In the next part we classify the quadratic forms by the signs of their values.

**8.8. Definition** Let $Q : \mathbb{R}^n \to \mathbb{R}$ be a quadratic form represented by the symmetric matrix $A \in \mathbb{R}^{n \times n}$. We say that $Q$ is

(a) positive definite if $\forall\, x \in \mathbb{R}^n \setminus \{0\} : \quad Q(x) > 0$,

(b) negative definite if $\forall\, x \in \mathbb{R}^n \setminus \{0\} : \quad Q(x) < 0$,

(c) positive semidefinite if $\forall\, x \in \mathbb{R}^n : \quad Q(x) \geq 0$,

(d) negative semidefinite if $\forall\, x \in \mathbb{R}^n : \quad Q(x) \leq 0$,

(e) indefinite, if $\exists\, x, y \in \mathbb{R}^n : \quad Q(x) > 0,\ Q(y) < 0$.

**8.9. Remarks.**

1. Since a quadratic form can be uniquely represented by a symmetric matrix, the above classification means the classification of symmetric matrices at the same time.

2. Every positive definite quadratic form is positive semidefinite and every negative definite quadratic form is negative semidefinite.

3. The constant 0 function is a quadratic form (represented by the 0 matrix). It is positive and negative semidefinite at the same time. Apart from this case the set of $n$-variable quadratic forms bursts into three disjoint classes: positive semidefinite, negative semidefinite, indefinite.

4. It is obvious that if $Q$ is positive definite then both the constants in theorem 8.6 are positive: $m,\ M > 0$.

Let us study the classification of 2-variable quadratic forms. A 2-variable quadratic form is given by a symmetric matrix of size $2 \times 2$.

**8.10. Theorem** *[classification of the 2-variable quadratic forms]*

*Let* $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \in \mathbb{R}^{2\times 2}$ *and* $Q : \mathbb{R}^n \to \mathbb{R}$ *be the quadratic form given by* $A$*, that is*

$$Q(x_1, x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2 \qquad (x = (x_1, x_2) \in \mathbb{R}^2)\,.$$

*Then* $Q$ *is*

- *positive definite if* $\det A = ac - b^2 > 0$ *and* $a > 0$*,*
- *negative definite if* $\det A = ac - b^2 > 0$ *and* $a < 0$*.*
  *(The case* $\det A = ac - b^2 > 0$ *and* $a = 0$ *is impossible.)*

- *indefinite if* $\det A = ac - b^2 < 0$.
- *semidefinite but not definite if* $\det A = ac - b^2 = 0$.

*The semidefinite case is in detail as follows. Suppose that* $\det A = ac - b^2 = 0$. *Then* $Q$ *is*

- *positive semidefinite but not positive definite if* $a > 0$ *or if* $a = 0, c > 0$,
- *negative semidefinite but not negative definite if* $a < 0$ *or if* $a = 0, c < 0$,
- *the identical* 0*-function if* $a = c = 0$.

**Proof.** The proof is based on the following elementary identities:

$$Q(x_1, x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2 = \begin{cases} \dfrac{(ax_1 + bx_2)^2 + (ac - b^2)x_2^2}{a} & \text{if } a \neq 0, \\ \dfrac{(bx_1 + cx_2)^2 + (ac - b^2)x_1^2}{c} & \text{if } c \neq 0, \\ 2bx_1x_2 & \text{if } a = c = 0. \end{cases}$$

Using these identities one can easily discuss the sign of the values of $Q$. □

## 8.3. Homeworks

1. Using the first order condition of local extremum and the Weierstrass's minimax theorem solve the following absolute extreme value problems in $\mathbb{R}^2$:

   a) $f(x, y) = x^2 + y^2 - xy \quad (0 \le x \le 4,\ 0 \le y \le x)$;
   b) $f(x, y) = x^2 - 2xy + 2y \quad (0 \le x \le 2,\ 0 \le y \le 3)$;
   c) $f(x, y) = x^2 - y^2 - x \quad (x \ge 0,\ y \ge 0,\ x^2 + y^2 \le 1)$.

2. Rotate a rectangular region around one of its sides. In which case will have the resulted cylinder the maximal volume if the perimeter of the rectangle is a given number $a > 0$. Give the sizes of the rectangle in the answer.

3. Write the matrices of the following quadratic forms. Determine in which class of definiteness they are.

   a) $Q(x, y) = 3x^2 - 4xy + 4y^2$
   b) $Q(x, y) = x^2 + 5xy + 4y^2$
   c) $Q(x, y) = x^2 - 2xy + y^2$

# 9. Lesson 9

## 9.1. Local extreme values: the Second Derivative Test

In this section we will prove two theorems – that use second order derivatives – in connection with the local extrema.

**9.1. Theorem** *[the definite case] Let $f \in \mathbb{R}^n \to \mathbb{R},\ f \in C^2(a),\ f'(a) = 0$. Then*

*(a) If $f''(a)$ is positive definite then $f$ attains local minimum at $a$;*

*(b) If $f''(a)$ is negative definite then $f$ attains local maximum at $a$;*

**Proof.** It is enough to prove part (a). The part (b) can be reduced back to (a) applying it for $-f$.

To prove part (a) we will show that

$$\exists\, \delta > 0\ \forall\, h \in \mathbb{R}^n \setminus \{0\},\ \|h\| < \delta :\ f(a+h) - f(a) > 0\,.$$

Denote by $r$ the radius of a ball with center $a$ which is a subset of $D_f$ (this $r$ exists because $a$ is an interior point of $D_f$). Let $h \in \mathbb{R}^n \setminus \{0\}$, $\|h\| < r$. Then we can apply the Taylor's formula with $m = 1$. So $\exists\, \vartheta = \vartheta(h) \in \mathbb{R}$, $0 < \vartheta < 1$ such that

$$f(a+h) = f(a) + \frac{f'(a)h^1}{1!} + \frac{f''(a+\vartheta h)h^2}{2!}$$

Since $f'(a) = 0$ then we have

$$f(a+h) - f(a) = \frac{f''(a+\vartheta h)h^2}{2!}\,.$$

Let us smuggle the term $\dfrac{f''(a)h^2}{2!}$ in this formula:

$$f(a+h) - f(a) = \frac{f''(a)h^2}{2!} + \frac{\left(f''(a+\vartheta h) - f''(a)\right)h^2}{2!}\,,$$

that is

$$f(a+h) - f(a) = \frac{1}{2} \cdot \left(Ah^2 + B(h)h^2\right) \qquad (h \in \mathbb{R}^n,\ 0 < \|h\| < r)$$

where

$$A = f''(a) \quad \text{and} \quad B(h) = f''(a+\vartheta h) - f''(a)\,.$$

Since $A$ is positive definite so

$$\exists\, m > 0\ \forall\, h \in \mathbb{R}^n : \quad Ah^2 \geq m \cdot \|h\|^2\,.$$

Using the continuity of $f''$ at $a$ we obtain $\lim\limits_{h\to 0} B(h) = 0$ so

$$\exists\, \delta > 0 : \quad \delta < r \text{ and } \forall\, h \in \mathbb{R}^n,\ \|h\| < \delta : \quad ||B(h)||_F < \frac{m}{2}$$

where $||B(h)||_F$ denotes the Frobenius-norm of the 2-array $B(h)$ (see Definition 1.10 in Lesson 1). Hence – using the norm-estimation in Theorem 1.12 in Lesson 1 – we obtain

$$|B(h)h^2| \leq ||B(h)||_F \cdot \|h\|^2 < \frac{m}{2} \cdot \|h\|^2 \qquad \text{that is}$$

$$-\frac{m}{2} \cdot \|h\|^2 \leq B(h)h^2 \leq \frac{m}{2} \cdot \|h\|^2 .$$

Using the left hand side inequality and the previous results we have
for $h \in \mathbb{R}^n,\ 0 < \|h\| < \delta$:

$$f(a+h) - f(a) = \frac{1}{2} \cdot \left(Ah^2 + B(h)h^2\right) \geq \frac{1}{2} \cdot \left(m\|h\|^2 - \frac{m}{2}\|h\|^2\right) = \frac{m}{4}\|h\|^2 > 0 .$$

□

**9.2. Theorem** *[the indefinite case] Let $f \in \mathbb{R}^n \to \mathbb{R},\ f \in C^2(a),\ f'(a) = 0$. If $f''(a)$ is indefinite then $f$ has no local extremum at $a$.*

**Proof.**

The first part of the previous proof is independent of the type of $f''(a)$. So we have

$$f(a+h) - f(a) = \frac{1}{2} \cdot \left(Ah^2 + B(h)h^2\right) \qquad (h \in \mathbb{R}^n,\ 0 < \|h\| < r)$$

where $r > 0$ is the radius for which $B(a, r) \subseteq D_f$, and

$$A = f''(a) \quad \text{and} \quad B(h) = f''(a + \vartheta h) - f''(a) .$$

Denote by $Q$ the quadratic form defined by $A$. Now $A$ is indefinite, so
$\exists\, x, y \in \mathbb{R}^n : \quad Q(x) > 0,\ Q(y) < 0$. Naturally $x \neq 0$ and $y \neq 0$.

First we will work with $x$. The values of $Q$ along the line $E_1 = \{h = tx \mid t \in \mathbb{R}\}$ are

$$Q(h) = Q(tx) = Q\left(t\|x\| \cdot \frac{x}{\|x\|}\right) = t^2\|x\|^2 \cdot Q\left(\frac{x}{\|x\|}\right) = m_1 \cdot \|tx\|^2 = m_1 \cdot \|h\|^2$$

where

$$m_1 := Q\left(\frac{x}{\|x\|}\right) = \frac{1}{\|x\|^2} \cdot Q(x) > 0 \qquad \text{independently of } h .$$

Hence we can apply the considerations of the proof of the previous theorem, so

$$\exists\, \delta_1 > 0 : \quad f(a+h) - f(a) \geq \frac{m_1}{4}\|h\|^2 > 0$$

$$\text{that is} \quad f(a+h) > f(a) \qquad (h \in E_1,\ 0 < \|h\| < \delta_1) .$$

Then working with $y$ we can deduce in similar way, that along the line

$E_2 = \{h = ty \mid t \in \mathbb{R}\}$:

$$\exists\, \delta_2 > 0: \quad f(a+h) - f(a) \le \frac{m_2}{4}\|h\|^2 < 0$$
$$\text{that is} \quad f(a+h) < f(a) \qquad (h \in E_2,\ 0 < \|h\| < \delta_2)$$

where

$$m_2 := Q\left(\frac{y}{\|y\|}\right) = \frac{1}{\|y\|^2} \cdot Q(y) < 0 \qquad \text{independently of } h\,.$$

Since any neighbourhood of $a$ contains points on the lines $E_1$ and $E_2$ with norm less than $\min\{\delta_1,\, \delta_2\}$ and different from $a$, it follows that $f$ has no local extreme value at $a$. □

**9.3. Corollary. (Second Order Necessary Condition)** Let $f \in \mathbb{R}^n \to \mathbb{R}$, $f \in C^2(a),\ f'(a) = 0$. If $f$ has local extremum at $a$ then $f''(a)$ is semidefinite.

## 9.2. Homeworks

1. Determine the local extrema and the places of local extrema of the following $\mathbb{R}^2 \to \mathbb{R}$ type functions:

   a) $f(x,y) = y^3 - x^2 - 4y^2 + 2xy$

   b) $f(x,y) = f(x,y) = x^4 - 4xy + y^4$

   c) $f(x,y) = x^2 + xy + y^2 + \dfrac{8}{x} + \dfrac{8}{y}$;

   d) $f(x,y) = x^4 + y^4 - x^2 - 2xy - y^2$

2. Let $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$ be the following functions:

$$f(x,y) = x^4 + y^2 \qquad\qquad g(x,y) = x^3 + y^2 \qquad\qquad ((x,y) \in \mathbb{R}^2)\,.$$

   a) Show that $f'(0,0) = g'(0,0) = 0$.

   b) Neither the theorem 9.1 nor the theorem 9.2 can be applied for $f$ and $g$ at the point $(0,0)$.

   c) $f$ has local extreme value at $(0,0)$.

   d) $g$ has no local extreme value at $(0,0)$.

# 10. Lesson 10

## 10.1. Multiple integrals over intervals

In Analysis-2 we have studied the Riemann-integral over intervals in $\mathbb{R}$. Now we will follow a similar way to construct the integral over $n$-dimensional intervals. Since the precise discussion requires long and complicated proofs, a lot of proofs will be omitted.

**10.1. Definition** Let $n \in \mathbb{N}$ and $a_k,\ b_k \in \mathbb{R}$, $a_k < b_k$ $(k = 1, \dots n)$. The set

$$I := [a_1,\, b_1] \times [a_2,\, b_2] \times \ldots \times [a_n,\, b_n] = \\ = \{x = (x_1, \dots, x_n) \in \mathbb{R}^n \mid a_k \leq x_k \leq b_k \ (k = 1, \dots, n)\} \subset \mathbb{R}^n$$

is called an $n$-dimensional interval (or: $n$-dimensional box).

The measure (or: $n$-dimensional volume) of $I$ is defined as

$$\mu(I) := \prod_{k=1}^{n} (b_k - a_k)\,.$$

The diameter (or: length of diagonal) of $I$ is defined as

$$d(I) := \sqrt{\sum_{k=1}^{n} (b_k - a_k)^2}\,.$$

Remember (see: Analysis-2) that in one dimensional case the partition $P$ of an interval $[a,\, b]$ into $n$ closed subintervals is a finite number set $\{x_0,\, x_1,\, \dots,\, x_n\}$, where

$$a = x_0 < x_1 < x_2 < \ldots < x_n = b\,.$$

Let us modify a bit the notation $P$. Denote by $P$ *the set of subintervals* instead of the set of the divisor points. That is let the partition $P$ be as follows:

$$P := \{[x_{i-1},\, x_i] \mid i = 1, \dots, n)\}\,.$$

The set of all partitions of the interval $[a,\, b]$ is denoted by $\mathcal{P}[a,\, b]$.

After these review and preliminaries we can define the partition of an $n$-dimensional interval.

**10.2. Definition** Let

$$I = [a_1,\, b_1] \times [a_2,\, b_2] \times \ldots \times [a_n,\, b_n]$$

be an interval in $\mathbb{R}^n$. Let $P_k \in \mathcal{P}[a_k,\, b_k]$ be partitions „along the $k$-th coordinate direction” $(k = 1, \dots, n)$. Then the interval set

$$P := \{J_1 \times \ldots \times J_n \mid J_k \in P_k \ (k = 1, \dots, n)\}$$

is called a partition of $I$.

The norm of $P$ is the longest diagonal of the subintervals that is

$$||P|| := \max\{d(J) \mid J \in P\}\,.$$

The set of all partitions of $I$ is denoted by $\mathcal{P}(I)$.

One can easily see that for every $\delta > 0$ there exists a partition $P$ „finer" than $\delta$ that is $||P|| < \delta$.

**10.3. Definition** Let $I \subset \mathbb{R}^n$ be an interval and $f : I \to \mathbb{R}$ be a bounded function and $P \in \mathcal{P}(I)$. Let

$$m_J := \inf\{f(x) \mid x \in J\}, \qquad M_J := \sup\{f(x) \mid x \in J\} \qquad (J \in P)\,.$$

We introduce the following sums:

a) lower sum: $s(f,\, P) := \sum\limits_{J \in P} m_J \cdot \mu(J)$,

b) upper sum: $S(f,\, P) := \sum\limits_{J \in P} M_J \cdot \mu(J)$.

**10.4. Theorem** *If $P,\, Q \in \mathcal{P}(I)$, $P \subseteq Q$ then*

$$s(f,\, P) \le s(f,\, Q) \quad \textit{and} \quad S(f,\, P) \ge S(f,\, Q)\,.$$

**10.5. Corollary.** If $P,\, Q \in \mathcal{P}(I)$ then $s(f,\, P) \le S(f,\, Q)$. Hence follows that the set of the lower sums is bounded above and the set of the upper sums is bounded below.

**10.6. Definition** The number $I_*(f) := \sup\{s(f,\, P) \mid P \in \mathcal{P}(I)\}$ is called the lower integral of $f$. Respectively the number $I^*(f) := \inf\{S(f,\, P) \mid P \in \mathcal{P}(I)\}$ is called the upper integral of $f$.

**10.7. Definition** A function $f : I \to \mathbb{R}$ is called to be Riemann-integrable if it is bounded and $I_*(f) = I^*(f)$. This common value of the lower and upper integral is called the Riemann-integral of $f$.

We will use simply „integrable" and „integral" instead of „Riemann-integrable" and „Riemann-integral" respectively since no other integral concept occurs in our subject.

The definition can be extended easily to the case when the domain of $f$ is wider than $I$:

**10.8. Definition** Let $f \in \mathbb{R}^n \to \mathbb{R}$, $I \subseteq D_f$ be an interval. We say that $f$ is integrable over the interval $I$ if the restricted function $f_{|I}$ is integrable. The integral of $f$ over $I$ is defined as the integral of $f_{|I}$ and is denoted as follows:

$$\int\limits_I f, \qquad \int\limits_I f(x)\; dx\,.$$

The set of integrable functions over $I$ is denoted by $R(I)$.

**10.9. Examples**

1. Let $c \in \mathbb{R}$ be fixed and $f(x) := c \ (x \in I)$ be the constant function. Then for any partition $P \in \mathcal{P}(I)$ $m_J = M_J = c$ thus

$$s(f,\, P) := \sum_{J \in P} c \cdot \mu(J) = c \cdot \sum_{J \in P} \mu(J) = c \cdot \mu(I)\,,$$

which implies that $I_*(f) = c \cdot \mu(I)$.

On the other hand

$$S(f,\, P) := \sum_{J \in P} c \cdot \mu(J) = c \cdot \sum_{J \in P} \mu(J) = c \cdot \mu(I)\,,$$

which implies that $I^*(f) = c \cdot \mu(I)$.

So

$$\int\limits_I f(x)\, dx = I_*(f) = I^*(f) = c \cdot \mu(I)\,.$$

## 10.2. Properties of the integral

In this section the theorems are stated without proofs.

**10.10. Theorem** *[Addition] Let $I \subseteq \mathbb{R}^n$ be an interval, $f, g \in R(I)$. Then*

$$f + g \in R(I) \quad \text{and} \quad \int\limits_I (f+g) = \int\limits_I f + \int\limits_I g\,.$$

**10.11. Theorem** *[Constant Multiple] Let $I \subseteq \mathbb{R}^n$ be an interval, $f \in R(I)$, $c \in \mathbb{R}$. Then*

$$cf \in R(I) \quad \text{and} \quad \int\limits_I cf = c \cdot \int\limits_I f\,.$$

**10.12. Theorem** *[Interval Additivity]*
*Let $p \in \{1,\, \dots,\, n\}$ and $c \in \mathbb{R}$, $a_k < c < b_k$. Let*

$$I' := [a_1,\, b_1] \times \ldots \times [a_{p-1},\, b_{p-1}] \times [a_p,\, c] \times [a_{p+1},\, b_{p+1}] \times \ldots \times [a_n,\, b_n]$$

*and*

$$I'' := [a_1,\, b_1] \times \ldots \times [a_{p-1},\, b_{p-1}] \times [c,\, b_p] \times [a_{p+1},\, b_{p+1}] \times \ldots \times [a_n,\, b_n]\,.$$

*Then*

$$f \in R(I) \quad \Leftrightarrow \quad f \in R(I') \quad \text{and} \quad f \in R(I'')\,.$$

*In this case:*

$$\int\limits_I f = \int\limits_{I'} f + \int\limits_{I''} f\,.$$

**10.13. Corollary.** 1. Applying several times the interval additivity we obtain for a partition $P \in \mathcal{P}(I)$ that

$$f \in R(I) \quad \Leftrightarrow \quad \forall\, J \in P: \; f \in R(J)\,. \quad \text{in this case:} \quad \int\limits_I f = \sum_{J \in P} \int\limits_J f\,.$$

2. If $f \in R(I)$ then for every subinterval $K \subseteq I$: $f \in R(K)$.

**10.14. Theorem** *[Monotonicity]*
*Let $f, g \in R(I)$. Suppose that $f(x) \le g(x)$ $(x \in I)$. Then*

$$\int\limits_I f \le \int\limits_I g\,.$$

**10.15. Theorem** *[„Triangle" inequality] Let $f \in R(I)$. Then $|f| \in R(I)$ and*

$$\left| \int\limits_I f \right| \le \int\limits_I |f|\,.$$

**10.16. Theorem** *[Mean value Theorem]*
*Let $f, g \in R(I)$, $g(x) \ge 0$ $(x \in I)$. Let*

$$m := \inf\{f(x) \mid x \in I\}, \qquad M := \sup\{f(x) \mid x \in I\}\,.$$

*Then*

$$m \cdot \int\limits_I g \le \int\limits_I fg \le M \cdot \int\limits_I g\,.$$

*Moreover if $f$ is continuous on $I$ then*

$$\exists\, \xi \in I: \qquad \int\limits_I fg = f(\xi) \cdot \int\limits_I g\,.$$

**10.17. Theorem** *Let $I \subseteq \mathbb{R}^n$ be an interval and $f : I \to \mathbb{R}$ be a continuous function. Then $f$ is integrable.*

## 10.3. Computation of the integral over intervals

In this section we will show how the integral can be computed via reduction back to one-variable integrals.

**10.18. Theorem** *Let*

$$I = [a_1,\, b_1] \times [a_2,\, b_2] \times \ldots \times [a_n,\, b_n]$$

*be an interval in* $\mathbb{R}^n$, $x \in \mathbb{R}^n$ *and* $p \in \{1, \ldots, n\}$. *Define the interval* $I^{(p)} \subset \mathbb{R}^{n-1}$ *and the vector* $x^{(p)} \in \mathbb{R}^{n-1}$ *as follows:*

$$I^{(p)} := [a_1, b_1] \times \ldots \times [a_{p-1}, b_{p-1}] \times [a_{p+1}, b_{p+1}] \times \ldots \times [a_n, b_n]$$

*and*

$$x^{(p)} := (x_1, \ldots x_{p-1}, x_{p+1}, \ldots, x_n)\,.$$

*Let* $f \in R(I)$ *and suppose that for any fixed* $t \in [a_p, b_p]$ *the functions*

$$\varphi_{p,t} : I^{(p)} \to \mathbb{R}, \qquad \varphi_{p,t}(x^{(p)}) := f(x_1, \ldots x_{p-1}, t, x_{p+1}, \ldots, x_n)$$

*are all integrable. Then*

$$\int_I f = \int_{a_p}^{b_p} \left( \int_{I^{(p)}} \varphi_{p,t} \right) dt\,.$$

**10.19. Corollary.** If the function $f$ is continuous on $I$, then the assumptions of the above theorem are satisfied. Applying the theorem $n-1$ times, the integral can be reduced into $n$ one-variable integrals, for example:

$$\int_I f(x_1, \ldots, x_n) dx = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \ldots \int_{a_n}^{b_n} f(x_1, \ldots, x_n)\; dx_n \ldots dx_2\, dx_1\,.$$

The number of the possible reductions is $n!$. A possible reduction is named as an order of integration. In this sense we can say that the integral can be evaluated in $n!$ order. If we write the equality of two possible evaluation order then we say that we have interchanged the order of integration.

Suppose that the function $f$ is „product of one-variable functions” in the following sense

$$f(x_1, x_2, \ldots, x_n) = g_1(x_1) \cdot g_2(x_2) \cdot \ \ldots\ \cdot g_n(x_n)$$

where the functions $g_k \in \mathbb{R} \to \mathbb{R}$ are continuous on $[a_k, b_k]$, then

$$\int_I f = (\int_{a_1}^{b_1} g_1) \cdot (\int_{a_2}^{b_2} g_2) \cdot \ \ldots\ \cdot (\int_{a_n}^{b_n} g_n)\,.$$

This case is called „separable case”. We will prove the above identity only for $n = 2$. The other cases are similar.

Let us see the special cases $n = 2$ and $n = 3$ of the integration process.

Case $n = 2$ (double integral)

Let $I = [a, b] \times [c, d] \subset \mathbb{R}^2$ be an interval and the function $f$ be continuous on $I$. Then

$$\iint_I f(x, y)\ d(x, y) = \int_a^b \int_c^d f(x, y)\ dy\ dx = \int_c^d \int_a^b f(x, y)\ dx\ dy\,.$$

In the separable case let $f(x, y) = g(x) \cdot h(y)$ with continuous $g$ and $h$. Then

$$\iint_I f(x, y)\ d(x, y) = \iint_I g(x)h(y)\ d(x, y) =$$

$$= \int_a^b \int_c^d g(x)h(y)\ dy\ dx = \int_a^b g(x) \cdot \left( \int_c^d h(y)\ dy \right) dx =$$

$$= \left( \int_c^d h(y)\ dy \right) \cdot \int_a^b g(x)\ dx = \left( \int_a^b g(x)\ dx \right) \cdot \left( \int_c^d h(y)\ dy \right).$$

Case $n = 3$ (triple integral)

Let $I = [a, b] \times [c, d] \times [p, q] \subset \mathbb{R}^3$ be an interval and the function $f$ be continuous on $I$. Then

$$\iiint_I f(x, y, z)\ d(x, y, z) = \int_a^b \int_c^d \int_p^q f(x, y, z)\ dz\ dy\ dx =$$

$$= \int_c^d \int_p^q \int_a^b f(x, y, z)\ dx\ dz\ dy = \ldots \text{ (6 possibilities)}.$$

In the separable case let $f(x, y, z) = g(x) \cdot h(y) \cdot k(z)$ with continuous $g$, $h$ and $k$. Then

$$\iiint_I f(x, y, z)\ d(x, y, z) = \iiint_I g(x)h(y)k(z)\ d(x, y, z) =$$

$$= \left( \int_a^b g(x)\ dx \right) \cdot \left( \int_c^d h(y)\ dy \right) \cdot \left( \int_p^q k(z)\ dz \right).$$

## 10.4. Homeworks

Compute the following integrals

1. $$\iint_H 2x^2 + 3xy + 4y^2 \; d(x,y) \quad \text{where} \quad H = [1;\, 2] \times [0;\, 3] \subset \mathbb{R}^2$$

2. $$\iint_H e^{x+y} \; d(x,y) \quad \text{where} \quad H = [1;\, 4] \times [1;\, 2] \subset \mathbb{R}^2$$

3. $$\iiint_H 2x - 4y + 6z - 3 \; d(x,y,z) \quad \text{where} \quad H = [0;\, 2] \times [0;\, 1] \times [0;\, 3] \subset \mathbb{R}^3$$

4. $$\iiint_H xy^2z^3 \; d(x,y,z) \quad \text{where} \quad H = [1;\, 2] \times [0;\, 1] \times [0;\, 2] \subset \mathbb{R}^3$$

# 11. Lesson 11

## 11.1. Integration over bounded sets

**11.1. Definition** Let $\emptyset \neq H \subset \mathbb{R}^n$ be a bounded set, $f \in \mathbb{R}^n \to \mathbb{R}$ and suppose that the restriction $f_{|H}$ is a bounded function. Since $H$ is bounded then there exists an interval $I \subset \mathbb{R}^n$ such that $H \subseteq I$. Define the following function:

$$\widetilde{f}(x) := \begin{cases} f(x) & \text{if} \quad x \in H \\ \\ 0 & \text{if} \ \ x \in I \setminus H \end{cases}$$

$f$ is called integrable over $H$ if $\widetilde{f}$ is integrable. In this case the integral of $f$ over $H$ is defined as

$$\int\limits_H f := \int\limits_I \widetilde{f} \,.$$

One can easily see that the integrability of $f$ over $H$ and the value of the integral $\int_H f$ are independent of choosing $I$.

The set of functions that are intgrable over $H$ is denoted by $R(H)$.

**11.2. Definition** Let $\emptyset \neq H \subset \mathbb{R}^n$ be a bounded set. The set $H$ is called measurable (more precisely: Jordan-measurable) if the constant 1 function is integrable over $H$. In this case the number

$$\mu(H) := \int\limits_H 1 \, dx$$

is called the $n$-dimensional measure (more precisely: $n$-dimensional Jordan-measure) or simply the measure of $H$.

As a generalization of Theorem 10.17 it can be proved that the continuous functions are integrable over a compact measurable set.

**11.3. Theorem** *Let $\emptyset \neq H \subset \mathbb{R}^n$ be a compact and measurable set.*

*Then $C(H) \subseteq R(H)$ that is every continuous function is integrable over $H$.*

## 11.2. Computation of the integral over normal regions

The normal regions are the most simple regions after the intervals. Here – as we did earlier – we will use the notation

$$x^{(p)} := (x_1, \dots x_{p-1}, x_{p+1}, \dots, x_n) \in \mathbb{R}^{n-1} \,,$$

where $x \in \mathbb{R}^n$ and $p \in \{1, \dots, n\}$.

**11.4. Definition** Let $\emptyset \neq T \subset \mathbb{R}^{n-1}$ be a compact and measurable set.

Let $\varphi$, $\psi$ be functions for which hold

$$\varphi : T \to \mathbb{R}, \quad \psi : T \to \mathbb{R}, \quad \varphi(t) \leq \psi(t) \quad (t \in T)\,.$$

Then the set

$$H := \{x \in \mathbb{R}^n \mid x^{(p)} \in T,\ \varphi(x^{(p)}) \leq x_p \leq \psi(x^{(p)})\} \subset \mathbb{R}^n$$

is called $x^{(p)}$-normal region (sometimes it is called $x_p$-normal region).

**11.5. Theorem** *Using the above notations the followings are true:*

1. *$H$ is compact and measurable.*
2. *If $f \in \mathbb{R}^n \to \mathbb{R}$ and $f \in C(H)$ then*

$$\int_H f(x)\ dx = \int_T \left( \int_{\varphi(x^{(p)})}^{\psi(x^{(p)})} f(x_1, \ldots x_{p-1}, x_p, x_{p+1}, \ldots, x_n)\ dx_p \right)\ dx^{(p)}$$

**11.6. Corollary.** If $H \subset \mathbb{R}^n$ is a compact measurable set and the nonnegative function $f \in \mathbb{R}^n \to \mathbb{R}$ is continuous on $H$, then let

$$R := \{(x,t) \in \mathbb{R}^{n+1} \mid x \in H,\ 0 \leq t \leq f(x)\}$$

be the $(n+1)$-dimensional region „under the graph of $f$". Then $R$ is an $x$-normal region in $\mathbb{R}^{n+1}$ so it is compact and measurable set. Its measure is

$$\mu(R) = \int_R 1\ d(x,t) = \int_H \left( \int_0^{f(x)} 1\ dt \right)\ dx =$$
$$= \int_H 1 \cdot (f(x) - 0)\ dx = \int_H f(x)\ dx\,.$$

So the value of the integral gives us the measure of the region under the graph of the function. This is the geometrical meaning of the integral, that is familiar from the one-variable case.

Let us see the special cases of normal regions.

Case $n = 2$ (double integral)

Let $T = [a,b] \subset \mathbb{R}$ be a closed bounded interval, $\varphi$, $\psi : [a,b] \to \mathbb{R}$ be continuous functions with $\varphi(u) \leq \psi(u)$ $(u \in [a,b])$. Then the $x$-normal region is:

$$H = \{(x,y) \in \mathbb{R}^2 \mid x \in [a,b],\ \varphi(x) \leq y \leq \psi(x)\} \subset \mathbb{R}^2\,,$$

and for every function $f \in \mathbb{R}^2 \to \mathbb{R}$ that is continuous on $H$ holds

$$\iint\limits_H f(x,y)\; d(x,y) = \int\limits_a^b \int\limits_{\varphi(x)}^{\psi(x)} f(x,y)\; dy\; dx\,.$$

Furthermore the $y$-normal region is:

$$H = \{(x,y) \in \mathbb{R}^2 \mid y \in [a,b],\; \varphi(y) \le x \le \psi(y)\} \subset \mathbb{R}^2\,,$$

and for every function $f \in \mathbb{R}^2 \to \mathbb{R}$ that is continuous on $H$ holds

$$\iint\limits_H f(x,y)\; d(x,y) = \int\limits_a^b \int\limits_{\varphi(y)}^{\psi(y)} f(x,y)\; dx\; dy\,.$$

Case $n = 3$ (triple integral)

Let $\emptyset \ne T \subset \mathbb{R}^2$ be a compact measurable set, $\varphi,\, \psi : T \to \mathbb{R}$ be continuous functions with $\varphi(u,v) \le \psi(u,v)$ $((u,v) \in T)$. Then the $xy$-normal region is:

$$H = \{(x,y,z) \in \mathbb{R}^3 \mid (x,y) \in T,\; \varphi(x,y) \le z \le \psi(x,y)\} \subset \mathbb{R}^3\,,$$

and for every function $f \in \mathbb{R}^3 \to \mathbb{R}$ that is continuous on $H$ holds

$$\iiint\limits_H f(x,y,z)\; d(x,y,z) = \iint\limits_T \int\limits_{\varphi(x,y)}^{\psi(x,y)} f(x,y,z)\; dz\; d(x,y)\,.$$

Furthermore the $yz$-normal region is:

$$H = \{(x,y,z) \in \mathbb{R}^3 \mid (y,z) \in T,\; \varphi(y,z) \le x \le \psi(y,z)\} \subset \mathbb{R}^3\,,$$

and for every function $f \in \mathbb{R}^3 \to \mathbb{R}$ that is continuous on $H$ holds

$$\iiint\limits_H f(x,y,z)\; d(x,y,z) = \iint\limits_T \int\limits_{\varphi(y,z)}^{\psi(y,z)} f(x,y,z)\; dx\; d(y,z)\,.$$

Finally the $xz$-normal region is:

$$H = \{(x,y,z) \in \mathbb{R}^3 \mid (x,z) \in T,\; \varphi(x,z) \le y \le \psi(x,z)\} \subset \mathbb{R}^3\,,$$

and for every function $f \in \mathbb{R}^3 \to \mathbb{R}$ that is continuous on $H$ holds

$$\iiint\limits_H f(x,y,z)\; d(x,y,z) = \iint\limits_T \int\limits_{\varphi(x,z)}^{\psi(x,z)} f(x,y,z)\; dy\; d(x,z)\,.$$

## 11.3. Homeworks

Compute the integrals of the given functions $f$ over the given regions $H$:

1. $f(x,y) = x^2 + y^2, \quad H = \{(x,y) \in \mathbb{R}^2 \mid 0 \le x \le 1,\ x^2 \le y \le \sqrt{x}\};$
2. $f(x,y) = 2y + x + 2, \quad H = \{(x,y) \in \mathbb{R}^2 \mid 1 \le x \le 3,\ 0 \le y \le \dfrac{1}{x}\};$
3. $f(x,y) = x^2 + y, \quad H = \{(x,y) \in \mathbb{R}^2 \mid 0 \le y \le 1,\ y^2 \le x \le \sqrt{y}\};$
4. $f(x,y,z) = x - 2y + 4z$, the region $H$ is the polyhedron determined by the planes $x = 0$, $y = 0$, $z = 0$, $x + y + z = 1$;
5. $f(x,y,z) = x^2 + 2y + z^2$, the region $H$ is the polyhedron determined by the planes $x = 0$, $y = 0$, $z = 0$, $x + z = 2$, $y = 2$;

# 12. Lesson 12

## 12.1. Integral Transformation

In this section we give the theorem that is the multivariate analogy of the „Change of Variables” in one-dimensional integrals. The proof is complicated, therefore it will be omitted.

**12.1. Theorem** *[Integral Transformation] Let $\emptyset \neq T \subset \mathbb{R}^n$ be a bounded measurable set. Let $\Phi \in \mathbb{R}^n \to \mathbb{R}^n$ be a function. Suppose that*

*a) $\Phi$ is continuously differentiable on $clos T$ that is there exists an open set $G \subseteq \mathbb{R}^n$ for which $clos T \subset G$ and $f$ is continuously differentiable on $G$.*

*b) The restriction of $f$ to $int T$ that is the function $f_{|int\, T}$ is one-to-one.*

*Denote by $\Phi[T]$ the picture of $T$ by $\Phi$ that is*

$$\Phi[T] = \{\Phi(t) \in \mathbb{R}^n \mid t \in T\}\,.$$

*Then $\Phi[T]$ is bounded and measurable in $\mathbb{R}^n$ and for any function $f \in \mathbb{R}^n \to \mathbb{R}$ which is integrable over $\Phi[T]$ holds*

$$\int\limits_{\Phi[T]} f = \int\limits_{T} (f \circ \Phi) \cdot |\det \Phi'|\,.$$

**12.2. Remark.** The above equation can be written using variables as follows:

$$\int\limits_{H} f(x)dx = \int\limits_{T} f(\Phi(t)) \cdot |\det \Phi'(t)|\, dt$$

where $H = \Phi[T]$.

## 12.2. Double integral in polar coordinates

The polar transformation in the plane is a special case of the integral transformation. Let

$$\Phi : \mathbb{R}^2 \to \mathbb{R}^2, \qquad \Phi(r, \varphi) := (r \cos\varphi,\, r \sin\varphi) \qquad ((r, \varphi) \in \mathbb{R}^2).$$

Then $\Phi$ is continuously differentiable everywhere and it is one-to-one on $\operatorname{int} T$ if

$$T \subset [0, +\infty) \times [\alpha,\, \alpha + 2\pi]$$

where $\alpha$ is a fixed real number. In most cases $\alpha = 0$ or $\alpha = -\pi$.

Let us compute the determinant of $\Phi'$:

$$\Phi'(r,\varphi) = \begin{vmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{vmatrix} = r\cos^2\varphi + r\sin^2\varphi = r\cdot(\cos^2\varphi + \sin^2\varphi) = r\,.$$

Applying the integral transformation formula we obtain:

$$\iint\limits_H f(x,y)\ d(x,y) = \iint\limits_T f(r\cos\varphi,\, r\sin\varphi)\cdot r\ d(r,\varphi)$$

where $H = \Phi[T]$.

## 12.3. Triple integral in cylindrical coordinates

The cylindrical transformation in the space is the following mapping:

$$\Phi : \mathbb{R}^3 \to \mathbb{R}^3, \qquad \Phi(r,\,\varphi,\,h) := (r\cos\varphi,\, r\sin\varphi,\, h) \qquad ((r,\varphi,h)\in\mathbb{R}^3).$$

Obviously $\Phi$ is continuously differentiable everywhere and it is one-to-one on $\operatorname{int} T$ if

$$T \subset [0,\,+\infty)\times[\alpha,\,\alpha+2\pi]\times\mathbb{R}$$

where $\alpha$ is a fixed real number. Mainly $\alpha = 0$ or $\alpha = -\pi$.

The determinant of $\Phi'$ can be computed by expansion along its last row:

$$\Phi'(r,\varphi,h) = \begin{vmatrix} \cos\varphi & -r\sin\varphi & 0 \\ \sin\varphi & r\cos\varphi & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1\cdot\begin{vmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{vmatrix} = r\,.$$

Applying the integral transformation formula we obtain:

$$\iiint\limits_H f(x,y,z)\ d(x,y,z) = \iiint\limits_T f(r\cos\varphi,\, r\sin\varphi,\, h)\cdot r\ d(r,\varphi,h)$$

where $H = \Phi[T]$.

## 12.4. Triple integral in polar coordinates

The polar (or: spherical) transformation in the space is the following mapping:

$$\Phi : \mathbb{R}^3 \to \mathbb{R}^3, \qquad \Phi(r,\,\varphi,\,\vartheta) := (r\sin\vartheta\cos\varphi,\, r\sin\vartheta\sin\varphi,\, r\cos\vartheta) \qquad ((r,\varphi,\vartheta)\in\mathbb{R}^3).$$

Obviously $\Phi$ is continuously differentiable everywhere and – using geometrical ideas – it is one-to-one on $\operatorname{int} T$ if

$$T \subset [0,\,+\infty)\times[\alpha,\,\alpha+2\pi]\times[0,\,\pi]$$

where $\alpha$ is a fixed real number. Mostly $\alpha = 0$ or $\alpha = -\pi$.

Compute the determinant of $\Phi'$ by expansion along its last row:

$$\Phi'(r,\varphi,\vartheta) = \begin{vmatrix} \sin\vartheta\cos\varphi & -r\sin\vartheta\sin\varphi & r\cos\vartheta\cos\varphi \\ \sin\vartheta\sin\varphi & r\sin\vartheta\cos\varphi & r\cos\vartheta\sin\varphi \\ \cos\vartheta & 0 & -r\sin\vartheta \end{vmatrix} =$$

$$= (\cos\vartheta)\cdot \begin{vmatrix} -r\sin\vartheta\sin\varphi & r\cos\vartheta\cos\varphi \\ r\sin\vartheta\cos\varphi & r\cos\vartheta\sin\varphi \end{vmatrix} + (-r\sin\vartheta)\cdot \begin{vmatrix} \sin\vartheta\cos\varphi & -r\sin\vartheta\sin\varphi \\ \sin\vartheta\sin\varphi & r\sin\vartheta\cos\varphi \end{vmatrix} =$$

$$= (\cos\vartheta)\cdot(-r^2\sin^2\varphi\sin\vartheta\cos\vartheta - r^2\cos^2\varphi\sin\vartheta\cos\vartheta) -$$

$$- (r\sin\vartheta)\cdot(r\sin^2\vartheta\cos^2\varphi + r\sin^2\vartheta\sin^2\varphi) =$$

$$= -(\cos\vartheta)\cdot r^2\sin\vartheta\cos\vartheta - (r\sin\vartheta)\cdot r\sin^2\vartheta =$$

$$= -r^2\sin\vartheta\cos^2\vartheta - r^2\sin\vartheta\sin^2\vartheta = -r^2\sin\vartheta\,.$$

Applying the integral transformation formula we obtain:

$$\iiint\limits_H f(x,y,z)\ d(x,y,z) = \iiint\limits_H f(r\sin\vartheta\cos\varphi,\ r\sin\vartheta\sin\varphi,\ r\cos\vartheta)\cdot r^2\sin\vartheta\ d(r,\varphi,\vartheta)$$

where $H = \Phi[T]$.

## 12.5. Homeworks

Compute the following integrals using integral transformations.

1.
$$\iint\limits_H \sin(x^2+y^2)\ d(x,y)$$

   where $H \subset \mathbb{R}^2$ is given by the inequalities $x \ge 0$, $y \ge 0$, $x^2+y^2 \le 1$.

2.
$$\iiint\limits_H x^2+y^2\ \ d(x,y,z)$$

   where $H$ is the part of the cylinder $x^2+y^2 = 4$ which is between the planes $z = 0$ and $z = 8$.

3.
$$\iiint\limits_H x^2 + y^2 \ \ d(x, y, z)$$

where $H$ is a right circular cone standing on the $xy$-plane. The basic circle of this cone is the unit circle of the $xy$-plane, the height of the cone is 5 units.

4.
$$\iiint\limits_H x^2yz \ \ d(x, y, z)$$

where $H$ is that part of the unit ball, which lies in the positive octant ($x \geq 0$, $y \geq 0$, $z \geq 0$) of the coordinate system.