CZECH TECHNICAL UNIVERSITY IN PRAGUE

FACULTY OF MECHANICAL ENGINEERING

MATHEMATICS I

November 1, 2026

Experimental version of the lecture notes for the academic year 2026/2027

Based on: *J. Neustupa: Matematika I. Skriptum Strojní fakulty. Vydavatelství ČVUT, Praha 2015*

Contents

I. Linear algebra **5**
I.1 Vector spaces . . . 5
I.2 Matrices and determinants . . . 10
I.3 Systems of linear algebraic equations . . . 19
I.4 * Linear mappings of $\mathbb{R}^n$ into $\mathbb{R}^m$ . . . 24
I.5 Eigenvalues and eigenvectors of square matrices . . . 29
I.6 Survey of equivalent properties of a square matrix . . . 32

II. Analytic geometry in $\mathbb{E}_3$ **33**
II.1 Some basic notions . . . 33
II.2 Straight lines in $\mathbb{E}_3$ . . . 35
II.3 Planes in $\mathbb{E}_3$ . . . 38
II.4 * Quadric surfaces in $\mathbb{E}_3$ . . . 43

III. Differential calculus **50**
III.1 Sequences of real numbers . . . 51
III.2 Functions – basic notions . . . 56
III.3 Some concrete functions . . . 61
III.3 Limit and continuity of a function . . . 66
III.4 Derivative of a function . . . 74
III.5 Applications of derivatives, behavior of a function . . . 82
III.6 The osculating circle, Taylor's theorem . . . 95
III.7 * Parametric representation of functions . . . 98
III.8 * Approximate solution of the nonlinear equation $f(x) = 0$ . . . 102
III.9 * Complex and vector–valued functions of a real variable . . . 105

IV. Indefinite integral **108**
IV.1 Antiderivative, indefinite integral . . . 108
IV.2 Integration by parts . . . 111
IV.3 Integration by substitution . . . 113
IV.4 Integration of simple rational functions . . . 117
IV.5 Integration of functions of the type $\sin^n x \cdot \cos^m x$ . . . 123
IV.6 Integration of some other types of functions . . . 124
IV.7 * Separable differential equations . . . 128

V. Definite (Riemann's) integral **134**
V.1 Historical approach . . . 134
V.2 Definition of the Riemann integral . . . 137
V.3 Important properties of the Riemann integral . . . 139
V.4 Evaluation of the Riemann integral . . . 143
V.5 * Numerical integration . . . 146
V.6 The improper Riemann integral . . . 148
V.7 Some geometric and physical applications of the definite integral . . . 150

Index + Vocabulary **153**

List of selected symbols **159**

Bibliography **160**

Introduction

This text contains an approximate list of definitions and theorems that students meet in the Mathematics I course in their first term of studies at the Faculty of Mechanical Engineering at the Czech Technical University, together with brief remarks, comments and examples. Some proofs and derivations of formulas are also included. These can be regarded as useful exercises leading to a better understanding of the sense and properties of notions that we deal with. The text was not written to be a completely independent textbook, especially due to brief explications and the limited number of solved and unsolved problems. It is important to emphasize that, in order to be well prepared for the examination in Mathematics I, it is necessary to solve problems individually and to think over a large number of examples. Appropriate examples and exercises can be found e.g. in the textbook [NK].

Numbers of paragraphs or sections, whose contents are not actually required for the examination and which are addressed to more interested readers, are marked by the symbol $*$ on the right side above.

We suppose that readers are familiar with the notion of a *set* and that they know the set operations *union*, *intersection*, *complement* and *difference*. Let us remind the reader that an empty set is denoted by $\emptyset$. We can use this notation for describing a set: $M = \{\, x;\ V(x) \,\}$. We read this as: M is the set of all elements x such that $V(x)$ holds. ($V(x)$ is some statement that can be made about x.)

We also assume knowledge of the following notions:

- *mapping* (of a set A to a set B – we use for instance the denotation $F : A \mapsto B$),
- *one–to–one mapping* (also called *injective mapping*),
- *mapping* of the set A *onto* the set B (also called *surjective mapping*),
- *bijective mapping* of the set A onto the set B (a mapping which is injective and surjective),
- *inverse mapping* (we denote this by F_{-1}),
- *composite mapping* (we denote this by $F \circ G$),
- *domain of a mapping* (we denote this by $D(F)$),
- *range of a mapping* (we denote this by $R(F)$).

Further topics which are also assumed to be known from secondary school include some elementary notions of mathematical logic, i.e. a *statement* and operations with statements:

- *negation of a statement* X (we denote this $non\, X$ or $\neg X$),
- *conjunction of statements* X *and* Y (we denote this $X \wedge Y$ and we read it "both X and Y hold" or briefly "X and Y"),
- *alternative of statements* X *and* Y (we denote this $X \vee Y$ and we read it "X or Y"),
- *implication* (we denote this $X \implies Y$ and we read it "X implies Y", "Y follows from X", "if X holds then Y also holds", "X is a sufficient condition for Y", "Y is a necessary condition for X", "Y holds provided that X holds", etc.) and
- *equivalence* (we denote this $X \iff Y$ and we read it "X holds if and only if Y holds", "X is equivalent to Y", "X is a necessary and sufficient condition for Y", "Y is a necessary and sufficient condition for X", etc.).

We shall often use so called quantifiers:

- a *universal quantifier* is denoted by $\forall$ and it can be used for example in the sentence: $\forall\, x \in I : V(x)$ – we read it "for each $x \in I$ the statement $V(x)$ holds" or "each $x \in I$ has the property $V(x)$",

- an *existential quantifier* is denoted by $\exists$ and it can be used for example in this way: $\exists\, x \in I\ :\ V(x)$ – we read it "There exists $x \in I$ such that the statement $V(x)$ holds."

Quantifiers can also be used to create more complicated statements and assertions. Do not underestimate them! Their incorrect usage can entirely change the sense of various statements. You can compare e.g. these two sentences which differ only in the order of the quantifiers: 1) "To every married man there exists a woman who is his wife." 2) "There exists a married man such that every woman is his wife."

If n is a natural number (we use the denotation: $n \in \mathbb{N}$) then the set of all ordered n–tuples of real numbers is denoted by $\mathbb{R}^n$. Thus, $\mathbb{R}^2$ is the set of all ordered pairs of real numbers, $\mathbb{R}^3$ is the set of all ordered 3–tuples of real numbers, etc. An exception is made in the case $n = 1$ where instead of $\mathbb{R}^1$, we write only $\mathbb{R}$. Elements of $\mathbb{R}^n$ are written for example in this way: $[\,a_1; a_2\,]$, $[\,1; 3\,]$ (if $n = 2$), $[\,x_1; x_2; x_3\,]$ (if $n = 3$), $[\,x_1; x_2; \ldots; x_n\,]$, etc. If the distance of any two elements $X = [\,x_1; x_2; \ldots; x_n\,]$ and $Y = [\,y_1; y_2; \ldots; y_n\,]$ from $\mathbb{R}^n$ is defined in such a way that it is equal to

$$d(X, Y) \;=\; \sqrt{(x_1 - y_1)^2 \,+\, (x_2 - y_2)^2 \,+\, \ldots \,+\, (x_n - y_n)^2}$$

then $\mathbb{R}^n$ becomes the so called *n–dimensional Euclidean space*. This is denoted by $\mathbb{E}_n$. Elements of $\mathbb{E}_n$ are often called "points" and the distance of two points X and Y is also denoted by $\|X - Y\|$. $\mathbb{E}_1$ can be imagined as a straight line, $\mathbb{E}_2$ as a plane, etc.

CHAPTER I.

Linear algebra

I.1. Vector spaces

I.1.1. The n–dimensional arithmetic space.

Let us define the sum of any two elements $[x_1; x_2; \ldots; x_n]$, $[y_1; y_2; \ldots; y_n]$ from $\mathbb{R}^n$ by the formula

$$[x_1; x_2; \ldots; x_n] + [y_1; y_2; \ldots; y_n] = [x_1 + y_1; x_2 + y_2; \ldots; x_n + y_n]$$

and the product of any element $[x_1; x_2; \ldots; x_n]$ from $\mathbb{R}^n$ and any real number λ by the formula

$$\lambda \cdot [x_1; x_2; \ldots; x_n] = [\lambda x_1; \lambda x_2; \ldots; \lambda x_n].$$

The set $\mathbb{R}^n$ with these two operations is called the *n–dimensional arithmetic space*. Its elements (i.e. n–tuples of real numbers) are called *arithmetic vectors*.

I.1.2. Vectors in $\mathbb{E}_2$ and in $\mathbb{E}_3$.

Oriented segments AB and CD in $\mathbb{E}_2$ are called *equivalent* if they can be identified by parallel shifting. Each class of all oriented segments in $\mathbb{E}_2$ which are equivalent one to another is called a *vector* in $\mathbb{E}_2$. Any segment from this class is called a *representative* of the vector. The set of all vectors in $\mathbb{E}_2$ is denoted by $V(\mathbb{E}_2)$.

Each vector in $\mathbb{E}_2$ is uniquely defined by any of its representatives. Vectors are denoted by small boldface letters (for example $\vec{u}$, $\vec{v}$, etc.). In a chosen Cartesian coordinate system, every vector can be given by means of its *coordinates*. These are an ordered pair of numbers (in round brackets) which is obtained in such a way that the representative of the vector is chosen to be the oriented segment coming out of the origin, and the coordinates of the end point of this segment are the coordinates of the vector.

If $\vec{u} = (u_1; u_2)$ and $\vec{v} = (v_1; v_2)$ are vectors in $\mathbb{E}_2$ and λ is a real number, then we can define the sum of $\vec{u}$ and $\vec{v}$ and the product of $\vec{u}$ and the number λ by the equalities:

$$\vec{u} + \vec{v} = (u_1 + v_1;\ u_2 + v_2),$$
$$\lambda \cdot \vec{u} = (\lambda u_1;\ \lambda u_2).$$

It can easily be verified that the operations defined above have the properties:

(a) $\mathbb{E}_2$ is closed with respect to both operations. That means: if $\vec{u}, \vec{v} \in V(\mathbb{E}_2)$ and $\lambda \in \mathbb{R}$ then the sum $\vec{u} + \vec{v}$ and the product $\lambda \cdot \vec{u}$ also belong to $V(\mathbb{E}_2)$.

(b) If $\vec{u}$, $\vec{v}$, $\vec{w} \in V(\mathbb{E}_2)$ and α, $\beta \in \mathbb{R}$ then

(b1) $\vec{u} + \vec{v} = \vec{v} + \vec{u}$,

(b2) $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$,

(b3) $1 \cdot \vec{u} = \vec{u}$,

(b4) $\alpha \cdot (\beta \cdot \vec{u}) = (\alpha \cdot \beta) \cdot \vec{u}$,

(b5) $\alpha \cdot (\vec{u} + \vec{v}) = \alpha \cdot \vec{u} + \alpha \cdot \vec{v}$,

(b6) $(\alpha + \beta) \cdot \vec{u} = \alpha \cdot \vec{u} + \beta \cdot \vec{u}$.

(c) There exists the so called *zero vector* $\vec{o} = (0; 0)$ in $V(\mathbb{E}_2)$. If $\vec{u}$ is any vector from $V(\mathbb{E}_2)$ then $\vec{u} + \vec{o} = \vec{u}$.

(d) To every vector $\vec{u} \in V(\mathbb{E}_2)$ there exists a vector $-\vec{u}$ (the so called *opposite vector* to $\vec{u}$) so that $\vec{u} + (-\vec{u}) = \vec{o}$.

Due to property (a), $V(\mathbb{E}_2)$ is said to be *closed* with respect to the two operations "addition of vectors" and "multiplication of vectors by real numbers".

By analogy, it is possible to define $V(\mathbb{E}_3)$ (or even $V(\mathbb{E}_n)$) and operations "addition of vectors" and "multiplication of vectors by real numbers" in this set. These operations in $V(\mathbb{E}_3)$ (or in $V(\mathbb{E}_n)$) also have properties (a) – (d).

I.1.3. Vector spaces.

It can easily be verified that the operations "addition" and "multiplication by real numbers" in the n–dimensional arithmetic space $\mathbb{R}^n$ have properties (a) – (d), too. Thus, the sets $\mathbb{R}^n$, $V(\mathbb{E}_2)$, $V(\mathbb{E}_3)$ and $V(\mathbb{E}_n)$ are in a certain sense similar. Generally, it can be observed that various nonempty sets with the operations "addition" and "multiplication by real numbers" (which satisfy conditions (a) – (d)) appear very often in mathematics and its applications. All these sets are called *vector spaces*.

Elements of concrete vector spaces need not always be classical vectors as is the case in $V(\mathbb{E}_2)$ and in $V(\mathbb{E}_3)$. As examples of further vector spaces, we can mention:

- the set of all polynomials whose degree is less than or equal to n (i.e. functions that have the form $f(x) = a_0 + a_1x + \cdots + a_nx^n$, where $a_0, a_1, \ldots, a_n$ are real numbers),
- the set of all functions defined by the equation $f(x) = a_0 + a_1 \cdot \sin x + a_2 \cdot \cos x$, where a_0, a_1, a_2 are real numbers,
- the set of all sequences of real numbers, etc.

Try to suggest for yourself how it is possible to define the operations "addition" and "multiplication by real numbers" in these spaces so that the operations have properties (a) – (d).

Let us return to the vector space $V(\mathbb{E}_2)$. Two vectors $\vec{u} = (u_1; u_2)$, $\vec{v} = (v_1; v_2)$ from $V(\mathbb{E}_2)$ are equal if and only if the corresponding coordinates are equal, i.e. $u_1 = v_1$, $u_2 = v_2$. This simple assertion follows immediately from properties (a) – (d) from paragraph I.1.2. Analogous assertions also hold in $V(\mathbb{E}_3)$ and $\mathbb{R}^n$.

We are going to study a general vector space V in the next part of this chapter. If the approach seems to be too abstract for you, you can imagine that V coincides for instance with $V(\mathbb{E}_2)$, $V(\mathbb{E}_3)$ or $V(\mathbb{E}_n)$. Elements of V will also be called vectors and they will be denoted in the same way as elements of $V(\mathbb{E}_2)$ or $V(\mathbb{E}_3)$, i.e. by small boldface letters. The zero vector will be again denoted by $\vec{o}$.

I.1.4. Theorem (uniqueness of the zero vector).

There exists only one zero vector in the vector space V.

P r o o f :
Suppose that there are two different zero vectors $\vec{o}_1$ and $\vec{o}_2$ in V. Using property (c) from paragraph I.1.2, we get: $\vec{o}_1 = \vec{o}_1 + \vec{o}_2 = \vec{o}_2 + \vec{o}_1 = \vec{o}_2$. This is in contradiction with the assumption that the vectors $\vec{o}_1$ and $\vec{o}_2$ are different. Hence two different zero vectors in the vector space V cannot exist.

I.1.5. Theorem.

The following identities hold for any vector $\vec{u} \in V$ and any real number α:

$$1) \quad 0 \cdot \vec{u} = \vec{o}, \qquad 2) \quad (-1) \cdot \vec{u} = -\vec{u}, \qquad 3) \quad \alpha \cdot \vec{o} = \vec{o}.$$

(We do not show the proof of this theorem here. Nevertheless, it could easily be done by means of conditions (a) – (d) from I.1.2.)

I.1.6. Linear dependence and independence of vectors.

If $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ is a group of vectors in the vector space V and $\alpha_1, \alpha_2, \ldots, \alpha_n$ are real numbers then the vector

$$\alpha_1 \vec{u}_1 + \alpha_2 \vec{u}_2 + \ldots + \alpha_n \vec{u}_n$$

is called the *linear combination* of the vectors $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$. (The linear combination of vectors from V is a vector that also belongs to V. It is an easy consequence of the statements in item (a) in paragraph I.1.2.)

The group of vectors $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ is called *linearly dependent* if there exist coefficients $\alpha_1, \alpha_2, \ldots, \alpha_n$ such that at least one of them is different from zero and

$$\alpha_1 \vec{u}_1 + \alpha_2 \vec{u}_2 + \ldots + \alpha_n \vec{u}_n = \vec{o}.$$

A group of vectors which is not linearly dependent is called *linearly independent*.

I.1.7. Theorem.

If one of the vectors $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ in the vector space V is equal to the zero vector then the group $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ is linearly dependent.

P r o o f :
Let for instance $\vec{u}_1 = \vec{o}$. Then $1 \cdot \vec{u}_1 + 0 \cdot \vec{u}_2 + \cdots + 0 \cdot \vec{u}_n = \vec{o}$. So we have the linear combination of the vectors $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ which is equal to the zero vector $\vec{o}$ and not all coefficients in this linear combination are zeros. Thus, in accordance with definition I.1.6, the group $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ is linearly dependent.

I.1.8. Theorem.

The group of vectors $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ (where $n > 1$) from the vector space V is linearly dependent if and only if at least one vector from this group can be expressed as a linear combination of the other vectors of the group.

P r o o f :
a) Suppose that the group $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ is linearly dependent. Then there exist coefficients $\alpha_1, \alpha_2, \ldots, \alpha_n$ (such that at least one of them is different from zero – let it be for instance α_1) and $\alpha_1\vec{u}_1 + \alpha_2\vec{u}_2 + \cdots + \alpha_n\vec{u}_n = \vec{o}$. Since $\alpha_1 \neq 0$, we can divide the equality by α_1 and express $\vec{u}_1$:

$$\vec{u}_1 = -\frac{\alpha_2}{\alpha_1}\vec{u}_2 - \frac{\alpha_3}{\alpha_1}\vec{u}_3 - \ldots - \frac{\alpha_n}{\alpha_1}\vec{u}_n .$$

So the vector $\vec{u}_1$ is a linear combination of the other vectors of the group. Similarly, if $\alpha_2 \neq 0$ then it is possible to express $\vec{u}_2$ in the form of a linear combination of the other vectors of the group, etc.

b) Suppose now that for example the vector $\vec{u}_1$ is a linear combination of the other vectors, i.e. there exist coefficients $\beta_2, \ldots, \beta_n$ such that $\vec{u}_1 = \beta_2\vec{u}_2 + \cdots + \beta_n\vec{u}_n$. If we put $\alpha_1 = -1$, $\alpha_2 = \beta_2, \ldots, \alpha_n = \beta_n$ then we can see that at least one of these numbers is nonzero and $\alpha_1\vec{u}_1 + \alpha_2\vec{u}_2 + \cdots + \alpha_n\vec{u}_n = \vec{o}$. Hence the group $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ is linearly dependent.

I.1.9. Theorem.

The group of vectors $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n$ from the vector space V is linearly independent if and only if the vector equation

$$\alpha_1\vec{u}_1 + \alpha_2\vec{u}_2 + \dots + \alpha_n\vec{u}_n = \vec{o} \tag{I.1.1}$$

(for unknowns $\alpha_1, \alpha_2, \dots, \alpha_n$) has only the zero solution $\alpha_1 = 0, \dots, \alpha_n = 0$.
(This theorem is an immediate consequence of the definition of linear dependence and linear independence of vectors from paragraph I.1.6.)

I.1.10. Example.

The group of vectors $(1;1;0)$, $(0;2;3)$ and $(3;5;3)$ in $\mathbb{E}_3$ is linearly dependent. Equation (I.1.1), which in our concrete case has the form

$$\alpha_1 \cdot (1;1;0) + \alpha_2 \cdot (0;2;3) + \alpha_3 \cdot 5,3,5) = (0;0;0),$$

has for instance this nonzero solution: $\alpha_1 = 3$, $\alpha_2 = 1$, $\alpha_3 = -1$. Linear dependence of the given group of vectors follows now from theorem I.1.9.

I.1.11. Theorem.

Let $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n$ be a linearly dependent group of vectors from the vector space V. Then every group of vectors from V which contains the vectors $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n$, is also linearly dependent.

P r o o f :
The group $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n$ is linearly dependent, hence there exist coefficients $\alpha_1, \alpha_2, \dots, \alpha_n$ such that at least one of them is different from zero and $\alpha_1\vec{u}_1 + \alpha_2\vec{u}_2 + \dots + \alpha_n\vec{u}_n = \vec{o}$. Let $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m$ be a group of vectors which contains the vectors $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n$. Suppose that the vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m$ are ordered so that $\vec{v}_1 = \vec{u}_1$, $\vec{v}_2 = \vec{u}_2$, $\dots$, $\vec{v}_n = \vec{u}_n$. Then $\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \dots + \alpha_n\vec{v}_n + 0\cdot\vec{v}_{n+1} + \dots + 0\cdot\vec{v}_m = \vec{o}$. The coefficients $\alpha_1, \alpha_2, \dots, \alpha_n, 0, \dots, 0$ are surely not all equal to zero. So the group $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m$ is linearly dependent.

I.1.12. Dimension of a vector space.

Let n be a natural number. We say that the vector space V is *n–dimensional* (or equivalently: its *dimension* is equal to n – we write $\dim V = n$) if
a) there exists a group of n vectors in V that is linearly independent,
b) each group of more than n vectors from V is linearly dependent.

I.1.13. Basis of a vector space.

Let V be an n–dimensional vector space. Each linearly independent group of n vectors from V is called a *basis* of the space V.

I.1.14. Remark.

The dimension of vector space V equals the maximum number of linearly independent vectors that can be found in V. It also equals the number of vectors in an arbitrary basis of V.

I.1.15. Theorem.

Let $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ be a basis of the vector space V. Then every vector from V can be uniquely expressed as a linear combination of the vectors $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$.

P r o o f :

a) Existence of the expression: Let $\vec{v}$ be an arbitrarily chosen vector from V. The group of vectors $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n, \vec{v}$ is linearly dependent (because it is the group of more than n vectors). Hence there exist coefficients $\alpha_1, \alpha_2, \ldots, \alpha_n, \beta$ such that not all of them are equal to zero and

$$\alpha_1\vec{u}_1 + \alpha_2\vec{u}_2 + \ldots + \alpha_n\vec{u}_n + \beta\vec{v} = \vec{o}.$$

If β were equal to zero then the above equality would imply the linear dependence of the vectors $\vec{u}_1, \vec{u}_2 \ldots, \vec{u}_n$, which would be in contradiction with the assumptions of the theorem. So β is different from zero and it is possible to express $\vec{v}$:

$$\vec{v} = -\frac{\alpha_1}{\beta}\vec{u}_1 - \frac{\alpha_2}{\beta}\vec{u}_2 - \ldots - \frac{\alpha_n}{\beta}\vec{u}_n .$$

b) Uniqueness of the expression: Suppose that $\vec{v}$ can be also written in another way as the linear combination of the vectors $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$: $\vec{v} = c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_n\vec{u}_n$. If we subtract the first expression of $\vec{v}$ from the second one, we obtain:

$$\vec{o} = \left(c_1 + \frac{\alpha_1}{\beta}\right)\vec{u}_1 + \left(c_2 + \frac{\alpha_2}{\beta}\right)\vec{u}_2 + \ldots + \left(c_n + \frac{\alpha_n}{\beta}\right)\vec{u}_n .$$

The linear independence of the vectors $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ implies: $c_1 + \frac{\alpha_1}{\beta} = 0$, $c_2 + \frac{\alpha_2}{\beta} = 0$, $\ldots$, $c_n + \frac{\alpha_n}{\beta} = 0$. So we get equalities $c_1 = -\frac{\alpha_1}{\beta}$, $c_2 = -\frac{\alpha_2}{\beta}$, $\ldots$, $c_n = -\frac{\alpha_n}{\beta}$, which show that both expressions of the vector $\vec{v}$ are same.

I.1.16. Remark.

Theorem I.1.15. can be "reversed". We mean by this that one can also prove correctness of the inverse implication: If V is a vector space and $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ is a group of vectors with the property that each vector from V can be uniquely expressed as a linear combination of the vectors $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ then these vectors form a basis of the space V.

I.1.17. Example.

The vectors $\widehat{\imath} = (1;0;0)$, $\widehat{\jmath} = (0;1;0)$, $\widehat{k} = (0;0;1)$ form the basis of the vector space $V(\mathbb{E}_3)$ because each vector $\vec{a} = (a_1;a_2;a_3) \in V(\mathbb{E}_3)$ can be uniquely written as a linear combination of the vectors $\widehat{\imath}, \widehat{\jmath}, \widehat{k}$: $\vec{a} = a_1 \cdot (1;0;0) + a_2 \cdot (0;1;0) + a_3 \cdot (0;0;1)$. Taking also into account remark I.1.14, we come to the, by no means surprising, assertion that the space $V(\mathbb{E}_3)$ is three–dimensional.

By analogy, the arithmetic vectors $\vec{e}_1 = [1;0;...;0]$, $\vec{e}_2 = [0;1;...;0]$, $\ldots$, $\vec{e}_n = [0;0;...;1]$ form the basis of the space $\mathbb{R}^n$. (So this space was correctly called n–dimensional in paragraph I.1.1).

I.1.18. Remark.

The basis of a vector space is not unique! For example – you can easily verify that two groups of vectors $\widehat{\imath} = (1;0)$, $\widehat{\jmath} = (0;1)$ and $\vec{u} = (2;-1)$, $\vec{v} = (1;1)$ are both bases of the vector space $V(\mathbb{E}_2)$. Moreover, every vector space (with the exception of the so called trivial vector space, which contains only one element – the zero vector) has infinitely many various bases.

If V is an n–dimensional vector space and $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_j$ (where $j < n$) is a linearly independent group of vectors in V then this group can always be filled up to the basis of V by adding appropriate vectors from V.

I.1.19. Subspace of a vector space.

Suppose that W is a subset of vector space V. If W is the vector space (with the same operations "addition" and "multiplication by real numbers" as in V) then we call W the *subspace* of the vector space V.

I.1.20. How to recognize a subspace.

Let W be a subset of a vector space V. We wish to find out whether W is a subspace of V. The operations "addition" and "multiplication by real numbers" are defined in W because these operations are defined in V and $W \subset V$. Thus, W is an individual vector space itself (and consequently, it is the subspace of V) if it is closed with respect to these operations. This means if the sum of two arbitrary vectors from W remains in W and the λ–multiple of an arbitrary vector from W also remains in W (for any $\lambda \in \mathbb{R}$).

I.1.21. Example.

The set of all arithmetic vectors, that can be written in the form $[\alpha; \beta; 0]$ (where α, β are real numbers), is a subspace of $\mathbb{R}^3$.

The set of all arithmetic vectors, that have the form $[\alpha; \beta; \gamma]$ (where α, β, γ are real numbers and $\alpha > 0$), is not a subspace of $\mathbb{R}^3$.

I.1.22.* Remark.

Try to prove these simple assertions:

a) *If $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k$ is a family of vectors from vector space V then the set of all linear combinations of the vectors $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k$ is a subspace of V. This subspace is called the linear hull of the vectors $\vec{u}_1, \ldots, \vec{u}_k$.*

b) *If the family of vectors $\vec{u}_1, \ldots, \vec{u}_k$ is linearly independent then it forms a basis of its linear hull and the dimension of the hull is thus equal to k.*

I.2. Matrices and determinants

I.2.1. Matrices.

A rectangular array of $m \cdot n$ real numbers written in m rows and n columns is called a *matrix* of the type $m \times n$ (read m by n) or shortly an $m \times n$ matrix. The numbers which are contained in the matrix are called its *entries* or its *elements*. Matrices are usually denoted by capital letters and their entries are denoted by the same small letters with two indices. The indices are related to the position of the entry. For example, a_{ij} denotes the entry in the i–th row and j–th column in matrix $\boldsymbol{A}$.

I.2.2. Example.

$$\boldsymbol{A} = \begin{pmatrix} a_{11}, & a_{12}, & \ldots, & a_{1n} \\ a_{21}, & a_{22}, & \ldots, & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1}, & a_{m2}, & \ldots, & a_{mn} \end{pmatrix} \qquad \boldsymbol{B} = \begin{pmatrix} 0, & 2, & 3, & 4, & 8, & -5 \\ 2, & 3, & 5, & -1, & 9, & 17 \\ 3, & -8, & 7, & 6, & -4, & 23 \end{pmatrix}$$

$\boldsymbol{A}$ is the $m \times n$ matrix, $\boldsymbol{B}$ is the 3×6 matrix. If the type of matrix $\boldsymbol{A}$ is known, then $\boldsymbol{A}$ can be written down in a shorter way: $\boldsymbol{A} = (a_{ij})$.

I.2.3. Identity of two matrices.

Two matrices are identical if they are of the same type and if they have the same entries at corresponding positions.

I.2.4. Main diagonal, upper triangular matrix, zero matrix, transposed matrix.

Suppose that $\boldsymbol{A} = (a_{ij})$ is an $m \times n$ matrix.

The entries $a_{11}, a_{22}, \dots$ form the so called *main diagonal* in matrix $\boldsymbol{A}$.

If all entries under the main diagonal are equal to zero, then matrix $\boldsymbol{A}$ is called the *upper triangular matrix*.

A matrix whose all entries are equal to zero is called a *zero matrix*.

The $n \times m$ matrix $\boldsymbol{B} = (b_{ij})$ whose entries satisfy $b_{ij} = a_{ji}$ $(i = 1, \dots, n;\ j = 1, \dots, m)$ is called a *transposed matrix* to matrix $\boldsymbol{A}$. It is denoted by $\boldsymbol{A}^T$. (The 1st column of matrix $\boldsymbol{A}^T$ is identical with the 1st row of matrix $\boldsymbol{A}$, the 2nd column of $\boldsymbol{A}^T$ is equal to the 2nd row of $\boldsymbol{A}$, etc. In other words: the transposed matrix to matrix $\boldsymbol{A}$ can be obtained by turning $\boldsymbol{A}$ over the main diagonal.)

I.2.5. Square matrix, identity matrix.

A matrix with the same number of rows as columns is said to be a *square matrix*.

A square matrix all of whose entries on the main diagonal equal 1 and all of whose other entries are zeros is called the *identity matrix.* It is denoted by $\boldsymbol{E}$:

$$\boldsymbol{E} = \begin{pmatrix} 1, & 0, & \dots, & 0 \\ 0, & 1, & \dots, & 0 \\ \vdots & \vdots & & \vdots \\ 0, & 0, & \dots, & 1 \end{pmatrix}$$

I.2.6. Addition of matrices.

If the matrices $\boldsymbol{A} = (a_{ij})$ and $\boldsymbol{B} = (b_{ij})$ are both $m \times n$ then their *sum* is the $m \times n$ matrix $\boldsymbol{C} = (c_{ij})$ with the entries $c_{ij} = a_{ij} + b_{ij}$ $(i = 1, \dots, m;\ j = 1, \dots, n)$. We use the notation $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$.

I.2.7. Multiplication of matrices by real numbers.

If $\boldsymbol{A} = (a_{ij})$ is an $m \times n$ matrix and $\lambda \in \mathbb{R}$, then the *product of the number λ and matrix $\boldsymbol{A}$* (or in other words the *λ–multiple of matrix $\boldsymbol{A}$*) is the matrix $\boldsymbol{C} = (c_{ij})$ of the same type $m \times n$ with the entries $c_{ij} = \lambda \cdot a_{ij}$ $(i = 1, \dots, n;\ j = 1, \dots, m)$. We use the notation $\boldsymbol{C} = \lambda \cdot \boldsymbol{A}$ or $\boldsymbol{C} = \lambda \boldsymbol{A}$.

I.2.8. Example.

$$\begin{pmatrix} 1, & -3, & 2 \\ 4, & -1, & 3 \end{pmatrix} + \begin{pmatrix} 1, & 2, & 2 \\ -1, & 3, & -8 \end{pmatrix} = \begin{pmatrix} 2, & -1, & 4 \\ 3, & 2, & -5 \end{pmatrix}, \qquad 2 \cdot \begin{pmatrix} 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \end{pmatrix}.$$

I.2.9. Remark.

Matrices of the same type can also be subtracted: *The difference of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$* is the matrix $\boldsymbol{C} = \boldsymbol{A} + (-1) \cdot \boldsymbol{B}$. We write: $\boldsymbol{C} = \boldsymbol{A} - \boldsymbol{B}$.

I.2.10. Multiplication of matrices.

If $\boldsymbol{A} = (a_{ij})$ is an $m \times n$ matrix and $\boldsymbol{B} = (b_{ij})$ is an $n \times p$ matrix then the *product* of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ is the $m \times p$ matrix $\boldsymbol{C} = (c_{ij})$ whose entries satisfy: $c_{ij} = a_{i1} \cdot b_{1j} + a_{i2} \cdot b_{2j} + \cdots + a_{in} \cdot b_{mj}$ $(i = 1, \ldots, m;\ j = 1, \ldots, p)$. We write: $\boldsymbol{C} = \boldsymbol{A} \cdot \boldsymbol{B}$.

I.2.11. Remark.

The definition of the multiplication of matrices seems to be artificial at first sight and so we shall analyze it once more. It can be explained by means of the following new notion:

If $[u_1; u_2; \ldots; u_n]$ and $[v_1; v_2; \ldots; v_n]$ are arithmetic vectors from $\mathbb{R}^n$ then the number $u_1 \cdot v_1 + u_2 \cdot v_2 + \cdots + u_n \cdot v_n$ is called their *scalar product*.

The rows of matrix $\boldsymbol{A}$ can be identified with arithmetic vectors from $\mathbb{R}^n$ (the number of rows is m, and each of them has n entries). Similarly, we can regard the columns of matrix $\boldsymbol{B}$ as arithmetic vectors also from $\mathbb{R}^n$ (their number is p, and each of them has n entries). If you read definition I.2.10 carefully, you can observe that the entry c_{ij} in matrix $\boldsymbol{C}$ is the scalar product of the i–th row of matrix $\boldsymbol{A}$ with the j–th column of matrix $\boldsymbol{B}$.

Matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ can be multiplied (in this order) only if matrix $\boldsymbol{A}$ has the same number of columns as the number of rows of matrix $\boldsymbol{B}$. We can easily recognize whether this is fulfilled: we write down the types of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ (for example $m \times n,\ n \times p$) and the 2nd and 3rd number must be the same. Otherwise matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ cannot be multiplied (in this order).

I.2.12. Example.

Verify for yourself by calculation that it holds:

$$\begin{pmatrix} 3, & 2, & 5 \\ 2, & -4, & 6 \\ 1, & 0, & 0 \end{pmatrix} \cdot \begin{pmatrix} 1, & 5 \\ 3, & -2 \\ -2, & 0 \end{pmatrix} = \begin{pmatrix} -1, & 11 \\ -22, & 18 \\ 1, & 5 \end{pmatrix}, \quad \begin{pmatrix} 3, & 5 \\ 6, & -2 \\ -1, & 0 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 19 \\ 14 \\ -3 \end{pmatrix}.$$

I.2.13. Rules for operations with matrices.

Suppose that $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are matrices and α, β are real numbers. Then each of the equalities

a) $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$, b) $(\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C} = \boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C})$,

c) $\alpha \cdot (\boldsymbol{A} + \boldsymbol{B}) = \alpha \cdot \boldsymbol{A} + \alpha \cdot \boldsymbol{B}$, d) $(\alpha + \beta) \cdot \boldsymbol{A} = \alpha \cdot \boldsymbol{A} + \beta \cdot \boldsymbol{A}$,

e) $\boldsymbol{A} \cdot (\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{A} \cdot \boldsymbol{B} + \boldsymbol{A} \cdot \boldsymbol{C}$, f) $(\boldsymbol{A} \cdot \boldsymbol{B}) \cdot \boldsymbol{C} = \boldsymbol{A} \cdot (\boldsymbol{B} \cdot \boldsymbol{C})$,

g) $\boldsymbol{A} \cdot \boldsymbol{E} = \boldsymbol{A}$, h) $\boldsymbol{E} \cdot \boldsymbol{B} = \boldsymbol{B}$,

i) $(\boldsymbol{A} + \boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$, j) $(\boldsymbol{A} \cdot \boldsymbol{B})^T = \boldsymbol{B}^T \cdot \boldsymbol{A}^T$

holds if the types of matrices are such that the operations on the left hand sides have a sense. Try to prove equalities a) – j) for yourself under this assumption.

Multiplication of matrices is not commutative, i.e. it does not generally hold that $\boldsymbol{A} \cdot \boldsymbol{B} = \boldsymbol{B} \cdot \boldsymbol{A}$! If, for example, $\boldsymbol{A}$ is a 3×5 matrix and $\boldsymbol{B}$ is a 5×7 matrix then the product $\boldsymbol{A} \cdot \boldsymbol{B}$ is a 3×7 matrix, while the product $\boldsymbol{B} \cdot \boldsymbol{A}$ has no sense, it cannot be created. Moreover, even if both products $\boldsymbol{A} \cdot \boldsymbol{B}$ and $\boldsymbol{B} \cdot \boldsymbol{A}$ have a sense, there exist examples when $\boldsymbol{A} \cdot \boldsymbol{B} \neq \boldsymbol{B} \cdot \boldsymbol{A}$.

I.2.14. The rank of a matrix.

The maximum number of linearly independent rows of matrix $\boldsymbol{A}$ (taken as arithmetic vectors) is called the *rank* of matrix $\boldsymbol{A}$. We denote it $\mathrm{r}(\boldsymbol{A})$.

I.2.15. Example.

Let $\begin{pmatrix} 1, & -3, & 2 \\ 4, & -1, & 3 \end{pmatrix} + \begin{pmatrix} 1, & 2, & 2 \\ -1, & 3, & -8 \end{pmatrix} = \begin{pmatrix} 2, & -1, & 4 \\ 3, & 2, & -5 \end{pmatrix}$, $2 \cdot \begin{pmatrix} 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \end{pmatrix}$.

For instance – it can be verified by means of theorem I.1.9 that the first two rows of the matrix are linearly independent. The third row is the linear combination of the first two rows (it is equal to the difference of the 1st row (multiplied by 3) and the 2nd row (multiplied by 2)). Hence all three rows form the linearly dependent group. The maximum number of linearly independent rows is two, and for this reason $\mathrm{r}(\boldsymbol{A}) = 2$.

I.2.16. Remark.

It is natural to put the question whether it is possible to define the rank of a matrix by means of its columns instead of its rows (i.e. as the maximum number of linearly independent columns, if the columns are identified with arithmetic vectors). The answer is simple: YES. The "row definition" and the "column definition" assign to the matrix the same number as its rank. However, we should mention that the exact proof of this assertion is not quite simple.

If we deal with more complicated matrices than matrix $\boldsymbol{A}$ from example I.2.15, we are not able to recognize at first sight which rows form a linearly independent group and conversely, which ones are linear combinations of other rows. That is why we shall study the problem of how to specify the rank of a matrix in greater detail in several following paragraphs.

I.2.17. Theorem.

Let $\boldsymbol{A}$ be an $m \times n$ upper triangular matrix and let all entries on the main diagonal be different from zero. Then the rank of $\boldsymbol{A}$ is equal to the minimum of the numbers m, n.

I.2.18. Example.

Instead of showing a general proof of theorem I.2.17, let us study this special case: Suppose that

$$\boldsymbol{A} = \begin{pmatrix} 3, & 3, & 5, & 0, & -3, & 8 \\ 0, & 1, & 7, & -6, & 4, & 3 \\ 0, & 0, & 5, & 14, & 4, & 2 \end{pmatrix}.$$

Let us verify that the rows of matrix $\boldsymbol{A}$ are linearly independent. We can write the vector equation

$$\alpha \cdot (3;3;5;0;-3;8) + \beta \cdot (0;1;7;-6;4;3) + \gamma \cdot (0;0;5;14;4;2) = (0;0;0;0;0;0).$$

If we write down the corresponding equations for all coordinates, we obtain a system of six linear algebraic equations for three unknowns: α, β and γ. We can easily find that there exists only one solution: $\alpha = \beta = \gamma = 0$. The linear independence of the rows of matrix $\boldsymbol{A}$ now follows from theorem I.1.9. Their number is three, hence $\mathrm{r}(\boldsymbol{A}) = 3$. By analogy, the matrices

$$\begin{pmatrix} 1, & 3, & -2 \\ 0, & 4, & 7 \\ 0, & 0, & 5 \\ 0, & 0, & 0 \end{pmatrix}, \quad \begin{pmatrix} -2, & 3, & 0, & 5 \\ 0, & -1, & 5, & 0 \\ 0, & 0, & 15, & 1 \\ 0, & 0, & 0, & 7 \end{pmatrix}, \quad \begin{pmatrix} 2, & 1 \\ 0, & 4 \end{pmatrix}, \quad \begin{pmatrix} 5 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1, & 2, & -5, & 8 \end{pmatrix}$$

have the ranks 3, 4, 2, 1, 1.

I.2.19. Elementary row and column operations.

If we have to find the rank of a general matrix $\boldsymbol{A}$ which is not triangular, then we can transform the matrix to an upper triangular matrix (with non–zero entries on the main diagonal) using so called elementary row and column operations, which do not change the rank of the matrix, and afterwards we specify the rank by means of theorem I.2.17. We shall use the following elementary row operations:

a) *change of order of rows*,
b) *multiplication of some row by a nonzero number*,
c) *addition to some row of a linear combination of the other rows (specially, addition of a multiple of another row)*,
d) *omission of a row which is a linear combination of the other rows (specially, omission of a row all of whose entries are zeros or omission of a row which is a multiple of another row).*

(All the operations can also be performed with columns. We are not going to prove that these row and column operations do not change the rank of a matrix.)

The procedure of transformation of an arbitrary matrix to an upper triangular matrix (all of whose entries on the main diagonal are different from zero) by means of the elementary row and column operations is called the *Gauss algorithm*. The algorithm is explained in the next example:

I.2.20. Example.

We find the rank of the matrix

$$\boldsymbol{A} = \begin{pmatrix} 2, & -1, & 1, & 8, & 2 \\ 4, & -3, & 5, & 1, & 7 \\ 8, & -6, & 8, & 12, & 12 \\ 6, & -4, & 6, & 9, & 9 \end{pmatrix}.$$

Step 1. Since $a_{11} \neq 0$, we rewrite the 1st row. (If $a_{11} = 0$ then we interchange the rows or columns so that there is a nonzero entry on the position 1,1.) Now we want to get only zeros under the entry a_{11}. That is why we first multiply the 1st row by 2 and subtract it from the 2nd row, then we multiply the 1st row by 4 and subtract it from the 3rd row and finally, we multiply the 1st row by 3 and subtract it from the 4th row. We obtain the matrix:

$$\begin{pmatrix} 2, & -1, & 1, & 8, & 2 \\ 0, & -1, & 3, & -15, & 3 \\ 0, & -2, & 4, & -20, & 4 \\ 0, & -1, & 3, & -15, & 3 \end{pmatrix}.$$

Step 2. We rewrite the 1st row. Since the entry at position 2,2 is nonzero, we also rewrite the 2nd row. Now we want to get only zeros under the entry at position 2,2. Thus, we multiply the 2nd row by 2 and subtract from the 3rd row. Finally, we subtract the 2nd row from the 4th row. We get the matrix:

$$\begin{pmatrix} 2, & -1, & 1, & 8, & 2 \\ 0, & -1, & 3, & -15, & 3 \\ 0, & 0, & 2, & -36, & 0 \\ 0, & 0, & 0, & 0, & 0 \end{pmatrix}.$$

Step 3. The last row contains only zeros, hence we omit it. We obtain the upper triangular matrix:

$$\begin{pmatrix} 2, & -1, & 1, & 8, & 2 \\ 0, & -1, & 3, & -15, & 3 \\ 0, & 0, & 2, & -36, & 0 \end{pmatrix}.$$

Due to theorem I.2.17, the last matrix has the rank 3. Thus, $\mathrm{r}(\boldsymbol{A}) = 3$.

I.2.21. Determinant.

Let $\boldsymbol{A}$ be a square matrix. The *determinant* of matrix $\boldsymbol{A}$ is the number which is denoted by $\det \boldsymbol{A}$ and which is assigned to matrix $\boldsymbol{A}$ in accordance with these rules:

a) If $\boldsymbol{A} = (a)$ is a 1×1 square matrix then $\det \boldsymbol{A} = a$.

b) If $\boldsymbol{A} = (a_{ij})$ is an $n \times n$ square matrix (for $n > 1$) then we choose an arbitrary row of $\boldsymbol{A}$ (let us denote this row as the i–th one) and we put

$$\det \boldsymbol{A} = a_{i1} \cdot A_{i1} + a_{i2} \cdot A_{i2} + \ldots + a_{in} \cdot A_{in}, \tag{I.2.1}$$

where A_{ij} is the so called *co–factor of the entry* a_{ij}. The co–factor is equal to $(-1)^{i+j} \cdot \boldsymbol{A}^*_{ij}$ where $\boldsymbol{A}^*_{ij}$ is the determinant of the $(n-1) \times (n-1)$ square matrix which arises from $\boldsymbol{A}$ by omission the i–th row and the j–th column. ($\boldsymbol{A}^*_{ij}$ is called the *minor*, which is the abbreviation for "minor determinant".)

I.2.22. Remark.

The sum on the right side of (I.2.1) is called the *expansion of the determinant along the i–th row*. It can be proved that the choice of the row along which the determinant is expanded is not important because the result is always the same. The determinant can even be expanded along an arbitrary column. The *expansion of the determinant along the j–th column* is:

$$\det \boldsymbol{A} = a_{1j} \cdot A_{1j} + a_{2j} \cdot A_{2j} + \ldots + a_{nj} \cdot A_{nj}.$$

It can be easily verified that the determinant of a 2×2 square matrix $\boldsymbol{A}$ is:

$$\det \boldsymbol{A} = a_{11}\, a_{22} - a_{12}\, a_{21}.$$

Remember this simple formula !

The expansion of the $n \times n$ determinant along some row or column leads to the expression of this determinant by means of n $(n-1) \times (n-1)$ determinants. Each of these determinants can be further expanded along some of its rows or columns and so the problem is transformed to a question of calculation of $(n-2) \times (n-2)$ determinants. We can proceed in this way until we come to 2×2 determinants (or even 1×1 determinants) which we already know how to compute.

I.2.23. Remark.

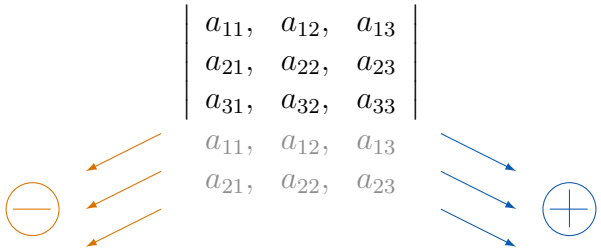
The determinant of matrix $\boldsymbol{A}$ is often written down analogously as matrix $\boldsymbol{A}$, only instead of round brackets we use straight vertical lines.

I.2.24. Saruss' rule.

The determinant of a 3×3 matrix can also be, apart from the expansion along some row or column, computed by the so called "Saruss' rule":

$$\begin{aligned} \det A = {} & a_{11} \cdot a_{22} \cdot a_{33} + a_{12} \cdot a_{23} \cdot a_{31} + a_{13} \cdot a_{21} \cdot a_{32} \\ & - a_{13} \cdot a_{22} \cdot a_{31} - a_{11} \cdot a_{23} \cdot a_{32} - a_{12} \cdot a_{21} \cdot a_{33}. \end{aligned}$$

The correctness of this formula can be verified by comparison with the result that can be obtained by expansion along some row or column. You can easily remember the formula by means of the following scheme:



I.2.25. Problems.

Verify that

a) $\begin{vmatrix} 2, & 5 \\ 3, & 7 \end{vmatrix} = -1,$ b) $\begin{vmatrix} 4, & 8, & 3 \\ 5, & -1, & 0 \\ 3, & 2, & -4 \end{vmatrix} = 215,$ c) $\begin{vmatrix} 4, & 2, & 5, & 0 \\ 2, & -1, & 0, & 2 \\ 3, & 6, & -8, & 2 \\ 7, & 1, & 0, & 1 \end{vmatrix} = -501.$

To compute the last determinant, it is advantageous to use the expansion along the 3rd column. (Why?)

I.2.26. Geometrical meaning of the determinant.

a) Suppose at first that $\boldsymbol{A}$ is a 2×2 square matrix. We can consider its rows to be the vectors $\vec{a}_1$ and $\vec{a}_2$. Let us fill in these two vectors to the parallelogram in plane $\mathbb{E}_2$. You can verify by an easy calculation that the area of the parallelogram is equal to $|\det \boldsymbol{A}|$.

b) Assume now that $\boldsymbol{A}$ is a 3×3 square matrix whose rows are the vectors $\vec{a}_1$, $\vec{a}_2$ and $\vec{a}_3$. Let us fill in these vectors to the parallelogram in space $\mathbb{E}_3$. Then the volume of the parallelogram is equal to $|\det \boldsymbol{A}|$.

c) Generally, let $\boldsymbol{A}$ be an $n \times n$ square matrix whose rows are the vectors $\vec{a}_1, \ldots, \vec{a}_n$. Let us fill in these vectors to the n–dimensional parallelogram in $\mathbb{E}_n$. Then the n–dimensional volume of the parallelogram is $|\det \boldsymbol{A}|$.

Assertions a), b) and c) remain valid if we work with the columns of matrix $\boldsymbol{A}$ instead of its rows.

I.2.27. Important properties of determinants.

Knowledge of the following assertions is very useful for computing determinants. Assume that $\boldsymbol{A}$ is an $n \times n$ square matrix (where $n > 1$).

a) If all entries in some row (or column) of $\boldsymbol{A}$ are zero then $\det \boldsymbol{A} = 0$. (This is seen if we expand the determinant along the zero row or column.)

b) $\det \boldsymbol{A} = \det \boldsymbol{A}^T$

c) Interchanging two rows (or columns) changes the sign of the determinant.

d) If two rows (or columns) are identical, the determinant is zero. (This is the consequence of item c): Interchanging two same columns changes the sign of the determinant. However, the new matrix is identical with $\boldsymbol{A}$, so its determinant is equal to $\det \boldsymbol{A}$. The equality $-\det \boldsymbol{A} = \det \boldsymbol{A}$ implies that $\det \boldsymbol{A} = 0$.)

e) If we multiply some row (or column) of matrix $\boldsymbol{A}$ by a number λ then the determinant of the new matrix is equal to $\lambda \cdot \det \boldsymbol{A}$. (This can be easily proved by expanding the determinant along the multiplied row or column.)

f) If any row (respectively column) of $\boldsymbol{A}$ is a multiple of another row (respectively column) of $\boldsymbol{A}$, the determinant of $\boldsymbol{A}$ is zero. (This is an easy consequence of items d) and e).)

g) If any row (respectively column) of $\boldsymbol{A}$ is a linear combination of the other rows (respectively columns) of $\boldsymbol{A}$, the determinant is zero. (We can prove this if we expand the determinant along the row (respectively column) that is the linear combination of the other rows (respectively columns) and apply the assertions from the items d) and e).)

h) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are $n \times n$ square matrices then $\det(\boldsymbol{A} \cdot \boldsymbol{B}) = \det \boldsymbol{A} \cdot \det \boldsymbol{B}$.

I.2.28. Remark.

The determinant of the $n \times n$ identity matrix (for arbitrary $n \in \mathbb{N}$) is equal to 1.

More generally: *The determinant of a square upper triangular matrix is equal to the product of all entries on the main diagonal.* Try to verify for yourself that this simple assertion is true.

(The expression "square upper triangular matrix" sounds rather strange, however when you read carefully the definition of a square and of an upper triangular matrix, you can see that the adjectives "square" and "upper triangular" are not in contradiction.)

The determinant of an $n \times n$ matrix was defined in paragraph I.2.21 by means of the expansion along some row (or column). Determinants of "smaller" matrices can really be calculated by means of these expansions. However, the calculation of determinants of "larger" matrices by means of the expansions along rows or columns would require an extremely large number of operations. (For instance, in the case of a 100×100 matrix, no modern computer would be able to do it in the epoch of existence of our universe.) Such determinants can be calculated by other (so called numerical) methods. For example: Applying the operations from item h) of paragraph I.2.27, the determinant can be transformed to the determinant of an upper triangular matrix. Then the assertion from the first part of this paragraph can be applied.

I.2.29. Regular and singular matrices.

An $n \times n$ square matrix which has the maximum possible rank (i.e. n) is called a *regular matrix*.
A square matrix which is not regular is called *singular*.

I.2.30. Inverse matrix.

Suppose that $\boldsymbol{A}$ is an $n \times n$ square matrix and $\boldsymbol{E}$ is the $n \times n$ identity matrix. An $n \times n$ matrix $\boldsymbol{A}^{-1}$ is called the *inverse matrix* to matrix $\boldsymbol{A}$ if

$$\boldsymbol{A} \cdot \boldsymbol{A}^{-1} = \boldsymbol{E}.$$

You will see in the following paragraphs that the inverse matrix $\boldsymbol{A}^{-1}$ need not exist to each square matrix $\boldsymbol{A}$!

I.2.31. Theorem – Equivalent characterization of regular matrices

Let $\boldsymbol{A}$ be a square matrix. Then the following statements are equivalent:

a) $\boldsymbol{A}$ is regular.

b) $\det \boldsymbol{A} \neq 0$.

c) The inverse matrix $\boldsymbol{A}^{-1}$ exists.

(We omit the proof of this theorem. The theorem states, among other things, when the inverse matrix does exist.)

I.2.32. Theorem.

If $\boldsymbol{A}$ and $\boldsymbol{B}$ are $n \times n$ regular matrices then matrix $\boldsymbol{A} \cdot \boldsymbol{B}$ is also regular. Moreover, it holds: $(\boldsymbol{A} \cdot \boldsymbol{B})^{-1} = \boldsymbol{B}^{-1} \cdot \boldsymbol{A}^{-1}$.

P r o o f :
Matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are regular, hence their determinants are different from zero. (See theorem I.2.31.) Assertion I.2.27 h) implies: $\det(\boldsymbol{A} \cdot \boldsymbol{B}) = \det \boldsymbol{A} \cdot \det \boldsymbol{B}$, which is different from zero. Using theorem I.2.31, we can see that matrix $\boldsymbol{A} \cdot \boldsymbol{B}$ is regular. The formula for $(\boldsymbol{A} \cdot \boldsymbol{B})^{-1}$ follows from the equalities $(\boldsymbol{A} \cdot \boldsymbol{B}) \cdot (\boldsymbol{B}^{-1} \cdot \boldsymbol{A}^{-1}) = \boldsymbol{A} \cdot (\boldsymbol{B} \cdot \boldsymbol{B}^{-1}) \cdot \boldsymbol{A}^{-1} = \boldsymbol{A} \cdot \boldsymbol{E} \cdot \boldsymbol{A}^{-1} = \boldsymbol{A} \cdot \boldsymbol{A}^{-1} = \boldsymbol{E}$.

I.2.33. Theorem.

If matrix A is regular then matrix A^{-1} is also regular. Moreover, it holds:

a) $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$, b) $\boldsymbol{A} \cdot \boldsymbol{A}^{-1} = \boldsymbol{A}^{-1} \cdot \boldsymbol{A} = \boldsymbol{E}$.

P r o o f :

a) If matrix $\boldsymbol{A}^{-1}$ were singular, it would be $1 = \det \boldsymbol{E} = \det(\boldsymbol{A} \cdot \boldsymbol{A}^{-1}) = \det \boldsymbol{A} \cdot \det \boldsymbol{A}^{-1} = \det \boldsymbol{A} \cdot 0 = 0$, which is impossible. Hence $\boldsymbol{A}^{-1}$ is regular. Further, one has: $(\boldsymbol{A}^{-1})^{-1} = (\boldsymbol{A} \cdot \boldsymbol{A}^{-1}) \cdot (\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A} \cdot (\boldsymbol{A}^{-1} \cdot (\boldsymbol{A}^{-1})^{-1}) = \boldsymbol{A} \cdot \boldsymbol{E} = \boldsymbol{A}$.

b) The formula $\boldsymbol{A} \cdot \boldsymbol{A}^{-1} = \boldsymbol{E}$ is already known. It remains to show that it also holds $\boldsymbol{A}^{-1} \cdot \boldsymbol{A} = \boldsymbol{E}$. Denote $\boldsymbol{B} = \boldsymbol{A}^{-1}$. Obviously, one has: $\boldsymbol{A}^{-1} \cdot \boldsymbol{A} = \boldsymbol{B} \cdot (\boldsymbol{A}^{-1})^{-1} = \boldsymbol{B} \cdot \boldsymbol{B}^{-1} = \boldsymbol{E}$.

I.2.34. Theorem (uniqueness of the inverse matrix).

If a square matrix A has an inverse matrix then the inverse matrix is unique.

P r o o f :
Suppose that both $\boldsymbol{A}_I^{-1}$ and $\boldsymbol{A}_{II}^{-1}$ are inverse matrices to matrix $\boldsymbol{A}$. Then we have:

a) $\boldsymbol{A}_I^{-1} \cdot \boldsymbol{A} \cdot \boldsymbol{A}_{II}^{-1} = (\boldsymbol{A}_I^{-1} \cdot \boldsymbol{A}) \cdot \boldsymbol{A}_{II}^{-1} = \boldsymbol{E} \cdot \boldsymbol{A}_{II}^{-1} = \boldsymbol{A}_{II}^{-1}$,

b) $\boldsymbol{A}_I^{-1} \cdot \boldsymbol{A} \cdot \boldsymbol{A}_{II}^{-1} = \boldsymbol{A}_I^{-1} \cdot (\boldsymbol{A} \cdot \boldsymbol{A}_{II}^{-1}) = \boldsymbol{A}_I^{-1} \cdot \boldsymbol{E} = \boldsymbol{A}_I^{-1}$.

This means that $\boldsymbol{A}_I^{-1} = \boldsymbol{A}_{II}^{-1}$. Thus, the inverse matrix (if it exists) is unique.

I.2.35. Remark.

A practical question is how to compute the inverse matrix $\boldsymbol{A}^{-1}$ to a given regular square matrix $\boldsymbol{A}$. For instance, it can be proved that

$$\boldsymbol{A}^{-1} = \frac{1}{\det \boldsymbol{A}} \begin{pmatrix} A_{11}, & \dots, & A_{1n} \\ \vdots & & \vdots \\ A_{n1}, & \dots, & A_{nn} \end{pmatrix}^T$$

where A_{ij} are co–factors of the entries a_{ij} in matrix $\boldsymbol{A}$ (see paragraph I.2.21, item c). However, the practical application of this formula for larger n is not advantageous, due to the necessity to compute the co–factors A_{ij}, which can be very laborious. There exists a less laborious procedure, based on the same idea as the Gauss algorithm described in I.2.20. It is explained in example 159, pp. 9–10, in the textbook [NK].

I.3. Systems of linear algebraic equations

I.3.1. Basic notions.

The system of equations

$$\begin{array}{ccccccccc} a_{11} \cdot x_1 & + & a_{12} \cdot x_2 & + & \dots & + & a_{1n} \cdot x_n & = & b_1 \\ a_{21} \cdot x_1 & + & a_{22} \cdot x_2 & + & \dots & + & a_{2n} \cdot x_n & = & b_2 \\ & & & & \vdots & & & & \\ a_{m1} \cdot x_1 & + & a_{m2} \cdot x_2 & + & \dots & + & a_{mn} \cdot x_n & = & b_m \end{array} \tag{I.3.1}$$

(where $a_{11}, a_{12}, \dots, a_{mn}, b_1, b_2, \dots, b_m$ are given real numbers and $x_1, x_2, \dots, x_n$ are unknowns) is called the *system of linear algebraic equations*. The matrices

$$\boldsymbol{A} = \begin{pmatrix} a_{11}, & a_{12}, & \dots, & a_{1n} \\ a_{21}, & a_{22}, & \dots, & a_{2n} \\ & & \vdots & \\ a_{m1}, & a_{m2}, & \dots, & a_{mn} \end{pmatrix}, \qquad (\boldsymbol{A} \,|\, \boldsymbol{B}) = \left(\begin{array}{cccc|c} a_{11}, & a_{12}, & \dots, & a_{1n} & b_1 \\ a_{21}, & a_{22}, & \dots, & a_{2n} & b_2 \\ & & \vdots & & \vdots \\ a_{m1}, & a_{m2}, & \dots, & a_{mn} & b_m \end{array}\right)$$

are called the *matrix of the system* (I.3.1) and the *augmented matrix of the system* (I.3.1). If we further denote

$$\boldsymbol{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad \boldsymbol{B} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix},$$

we can write the system (I.3.1) in the much shorter way:

$$\boldsymbol{A} \cdot \boldsymbol{X} = \boldsymbol{B}. \tag{I.3.1}$$

A solution of the system (I.3.1) is every ordered n–tuple of real numbers $x_1, x_2, \dots, x_n$ which satisfies the system. Solutions of the system can also be regarded as arithmetic vectors, i.e. as elements of $\mathbb{R}^n$. We believe that no misunderstanding can arise if the solution taken as an arithmetic vector is denoted in the same way as the solution taken as an $n \times 1$ matrix (i.e. for example by $\boldsymbol{X}$ in both cases).

If all the numbers $b_1, b_2, \dots, b_m$ are equal to zero, then the system (I.3.1) is called *homogeneous*. In the opposite case, system (I.3.1) is called *non–homogeneous*. The homogeneous system can be shortly written as the matrix equation

$$\boldsymbol{A} \cdot \boldsymbol{X} = \boldsymbol{O}. \tag{I.3.2}$$

where $\boldsymbol{O}$ is the $m \times 1$ zero matrix.

A non–homogeneous system need not always have a solution. (Example: $x_1 + x_2 = 1$, $x_1 + x_2 = 2$) On the other hand, an homogeneous system always has at least one (zero) solution $x_1 = \cdots = x_n = 0$. (This solution is called the *trivial solution.*) Of course, in some cases it also has other, non–zero (= non–trivial) solutions.

Two systems of equations are called *equivalent* if they have identical sets of solutions.

Our next task is to learn to find all solutions of the system (I.3.1) and to study the structure of the set of all solutions of the system (I.3.1) (or (I.3.2)).

I.3.2. Gaussian elimination.

This method can be used to solve the system (I.3.1). You will study it in detail in the exercises. However, here are the basic steps:

Step 1. Write down the augmented matrix of the system. The matrix can be transformed to an upper triangular one by means of elementary operations described in items a) – d) in paragraph I.2.19 and in example I.2.20. If possible, avoid interchanges of columns – they correspond to interchanges in the order of unknowns and they are often sources of mistakes for beginners. If you cannot avoid interchanges of columns, then you should write the unknowns above the column which contains coefficients standing at this unknown. The $(n+1)$–th column cannot be interchanged with any other one because it contains right hand sides of the equations in the system (I.3.1). Nevertheless, in what follows we suppose that interchanges of columns were not necessary.

Step 2. Write down the system of equations which corresponds to the last matrix. (It is equivalent to the original system we have to solve.) This system can be successively solved from the last to the first equation. Each of the equations can be regarded as the equation for one unknown only – that unknown with the lowest index.

2a) If the last equation has the form $c_n x_n = \gamma$ (where $c_n \neq 0$) then one can use it to express x_n, and substituting its value to the preceding equation, one can get x_{n-1}, etc.

2b) If the last equation has the form $0 = \gamma$ (where $\gamma \neq 0$) then the system has no solution.

2c) If the last equation has the form $c_k x_k + c_{k+1} x_{k+1} + \cdots + c_n x_n = \gamma$ (where $c_k \neq 0$) then $x_{k+1}, x_{k+2}, \ldots, x_n$ can be put equal to parameters which can be denoted for example by $p_1, p_2, \ldots, p_{n-k}$. We can express x_k from this equation. Substituting for x_k, x_{k+1}, $\ldots$, x_n to the preceding equation, we can use it to get x_{k-1}, etc. The system has now infinitely many solutions. Concrete solutions can be obtained by a concrete choice of values of parameters $p_1, p_2, \ldots, p_{n-k}$.

I.3.3. Remark.

Let us first study the homogeneous system (I.3.2). It is obvious that if we solve it by Gaussian elimination, case 2b) cannot appear. (The last column of the augmented matrix contains only zeros, and row operations a) – d) from paragraph I.2.19 cannot affect it. That is also why it is sufficient to work only with the matrix of the system. The augmented matrix is not necessary.) Thus, the homogeneous system is always solvable – it has the unique, trivial, solution in case 2a) or infinitely many solutions in case 2c). The following theorem gives important information on the structure of the set of all solutions of the system (I.3.2).

I.3.4. Theorem.

The set of all solutions of the homogeneous system (I.3.2) *is a subspace of the* n*–dimensional arithmetic space* $\mathbb{R}^n$ *whose dimension is equal to* $(n - \mathrm{r}(\boldsymbol{A}))$.

P r o o f :
Let us first show that the set of all solutions of the system (I.3.2) is a vector space which is a subspace of $\mathbb{R}^n$. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are solutions of (I.3.2) then it holds: $\boldsymbol{A}\cdot(\boldsymbol{X}+\boldsymbol{Y}) = \boldsymbol{A}\cdot\boldsymbol{X}+\boldsymbol{A}\cdot\boldsymbol{Y} = \boldsymbol{O}+\boldsymbol{O} = \boldsymbol{O}$, i.e. $\boldsymbol{X}+\boldsymbol{Y}$ also is a solution of the system (I.3.2). Similarly, if $\boldsymbol{X}$ is a solution of (I.3.2) and λ is a real number then $\boldsymbol{A}\cdot(\lambda\cdot\boldsymbol{X}) = \lambda\cdot(\boldsymbol{A}\cdot\boldsymbol{X}) = \lambda\cdot\boldsymbol{O} = \boldsymbol{O}$, i.e. $\lambda\cdot\boldsymbol{X}$ also

is a solution of (I.3.2). The set of all solutions of the homogeneous system (I.3.2) is a subset of $\mathbb{R}^n$ because its every element belongs to $\mathbb{R}^n$ and it is closed with respect to the operations "addition" and "multiplication by real numbers". Therefore it is a subspace of $\mathbb{R}^n$.

The assertion on the dimension of the subspace follows from theorem I.2.17 and from the procedure described in paragraph I.3.2. (The dimension is equal to the number of parameters $p_1, p_2, \ldots, p_{n-k}$ in paragraph I.3.2 – think over this fact for yourself.)

I.3.5. Example.

Let us solve the system
$$\begin{array}{rcrcrcl} x_1 & + & x_2 & - & 3x_3 & = & 0, \\ 5x_1 & - & 2x_2 & - & 8x_3 & = & 0, \\ 3x_1 & - & 4x_2 & - & 2x_3 & = & 0. \end{array}$$

The matrix of the system is transformed in accordance with instructions from paragraphs I.3.2 and I.2.19:
$$\begin{pmatrix} 1, & 1, & -3 \\ 5, & -2, & -8 \\ 3, & -4, & -2 \end{pmatrix} \sim \begin{pmatrix} 1, & 1, & -3 \\ 0, & -7, & 7 \\ 0, & -7, & 7 \end{pmatrix} \sim \begin{pmatrix} 1, & 1, & -3 \\ 0, & -7, & 7 \\ 0, & 0, & 0 \end{pmatrix} \sim \begin{pmatrix} 1, & 1, & -3 \\ 0, & -7, & 7 \end{pmatrix}.$$

The system of equations corresponding to the last matrix is
$$\begin{array}{rcrcrcl} x_1 & + & x_2 & - & 3x_3 & = & 0, \\ & - & 7x_2 & + & 7x_3 & = & 0. \end{array}$$

We put $x_3 = p$ (where p is a parameter). Then the second equation yields: $x_2 = p$. Substituting for x_3 and x_2 to the first equation, we obtain: $x_1 = 2p$. The solution can be generally expressed: $[x_1; x_2; x_3] = [2p; p; p] = [2; 1; 1]\,p$. Now it is obvious that there exist infinitely many solutions and the set of all solutions forms a 1–dimensional subspace of $\mathbb{R}^3$, whose basis is the arithmetic vector $[2; 1; 1]$. This corresponds to the equality $1 = 3 - 2$ (3 is the number of unknowns, 2 is the rank of the matrix of the system) – see theorem I.3.4.

I.3.6. Remark.

If the rank of matrix $\boldsymbol{A}$ is equal to the number of unknowns (i.e. $k = n$) then the set of all solutions of the homogeneous system (I.3.2) is the subspace of $\mathbb{R}^n$ of the dimension $n - n$, i.e. zero. This subspace contains only one element – the zero element – that is the n–tuple of nothing but zeros. This is exactly the case when the homogeneous system (I.3.2) has only the trivial solution.

Let us deal with the general (possibly non–homogeneous) system (I.3.1) again. The next theorem gives information about how to recognize which of the cases 2a) – 2c) (see paragraph I.3.2) occurs.

I.3.7. Frobenius' theorem.

I. The system of linear algebraic equations $\boldsymbol{A} \cdot \boldsymbol{X} = \boldsymbol{B}$ *(for* n *unknowns) has a solution if and only if* $\mathrm{r}(\boldsymbol{A}) = \mathrm{r}(\boldsymbol{A} \mid \boldsymbol{B})$.

II. If $\mathrm{r}(\boldsymbol{A}) = \mathrm{r}(\boldsymbol{A} \mid \boldsymbol{B}) = n$ *then the solution is unique.*
If $\mathrm{r}(\boldsymbol{A}) = \mathrm{r}(\boldsymbol{A} \mid \boldsymbol{B}) < n$ *then the system has infinitely many solutions.*

I.3.8.* Remark.

Let us analyze in greater detail the last case, i.e. the situation when the ranks of the matrix and the augmented matrix of the system (I.3.1) are both equal to k, where $k < n$. There is a natural question which is the structure of the set of all solutions of (I.3.1). Theorem I.3.4 states that the set of all solutions of the corresponding homogeneous system (I.3.2) forms a vector space (a subspace of $\mathbb{R}^n$) of the dimension $n - k$. If $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n-k}$ is a basis of this subspace then solutions of the homogeneous system (I.3.2) can generally be expressed in the form $c_1\boldsymbol{X}_1 + \cdots + c_{n-k}\boldsymbol{X}_{n-k}$. If some concrete, particular solution $\boldsymbol{Y}$ of the general (possibly non–homogeneous) system (I.3.1) is known, then all solutions of the system (I.3.1) can be expressed in the form:

$$\boldsymbol{X} = c_1\boldsymbol{X}_1 + \ldots + c_{n-k}\boldsymbol{X}_{n-k} + \boldsymbol{Y}. \tag{I.3.3}$$

This means that if $c_1, \ldots, c_{n-k}$ can each run independently each of the others over the set of all real numbers, then $\boldsymbol{X}$ runs over the set of all solutions of the system (I.3.1). This is why $\boldsymbol{X}$ is often called the *general solution* of the system (I.3.1).

I.3.9. Cramer's rule.

Let us now deal with the special case when system (I.3.1) $\boldsymbol{A} \cdot \boldsymbol{X} = \boldsymbol{B}$ is a system of n equations for n unknowns. The matrix of the system is a square matrix. Applying Frobenius' theorem, one can easily obtain this important assertion:

If matrix $\boldsymbol{A}$ of the system of equations (I.3.1) is regular then the system has a *unique* solution.

(Think over this fact and find for yourself reasons why this is true.) In this case, apart from by Gaussian elimination, the solution can also be obtained by means of the formulas

$$x_i = \frac{\det \boldsymbol{A}_i}{\det \boldsymbol{A}} \quad (i = 1, \ldots, n)$$

where $\boldsymbol{A}_i$ is the square matrix which arises from matrix $\boldsymbol{A}$ interchanging the i–th column with the column $\boldsymbol{B}$ of the right hand sides of the equations in (I.3.1).

(To derive this formula, it is possible to use the matrix form of the system: $\boldsymbol{A} \cdot \boldsymbol{X} = \boldsymbol{B}$, the consequent form of the solution: $\boldsymbol{X} = \boldsymbol{A}^{-1} \cdot \boldsymbol{B}$ and the expression of $\boldsymbol{A}^{-1}$ by means of the formula from paragraph I.2.35.)

I.4.* Linear mappings of $\mathbb{R}^n$ into $\mathbb{R}^m$

I.4.1. Motivation.

Suppose that $\boldsymbol{A}$ is the 2×3 matrix: $\boldsymbol{A} = \begin{pmatrix} 2, & 1, & -3 \\ -1, & 5, & -1 \end{pmatrix}$.

The equations

$$\begin{aligned} y_1 &= 2x_1 + x_2 - 3x_3 \\ y_2 &= -x_1 + 5x_2 - x_3 \end{aligned} \tag{I.4.1}$$

define a mapping $L : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$. This mapping assigns to given $X = [x_1; x_2; x_3] \in \mathbb{R}^3$ the image $Y = L(X) = [y_1; y_2] \in \mathbb{R}^2$.

I.4.2. Linear mapping of $\mathbb{R}^n$ into $\mathbb{R}^m$.

A mapping L of the n–dimensional arithmetic space $\mathbb{R}^n$ into the m–dimensional arithmetic space $\mathbb{R}^m$ is said to be *linear* if

a) L is *homogeneous*, i.e. $L(\alpha \cdot X) = \alpha \cdot L(X)$ for all $\alpha \in \mathbb{R}$ and for all points $X \in \mathbb{R}^n$ and

b) L is *additive*, i.e. $L(X_1 + X_2) = L(X_1) + L(X_2)$ for all points $X_1,\ X_2 \in \mathbb{R}^n$.

The linear mapping is also called the *linear transformation* or the *linear operator*.

I.4.3. Linear mapping defined by a matrix.

Similarly as in example I.4.1, one can define a mapping $L : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ by means of a general $m \times n$ matrix $\boldsymbol{A}$:

$$\begin{aligned} y_1 &= a_{11}\,x_1 + a_{12}\,x_2 + \ldots + a_{1n}\,x_n \\ &\vdots \\ y_m &= a_{m1}\,x_1 + a_{m2}\,x_2 + \ldots + a_{mn}\,x_n\,. \end{aligned} \tag{I.4.2}$$

It will be advantageous to write the coordinates of points in $\mathbb{R}^n$ in the form of an $n \times 1$ matrix in this section. Then

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n, \qquad Y = L(X) = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \in \mathbb{R}^m.$$

Equations (I.4.2) can be, in accordance with the rules on multiplication of matrices, written in much simpler form as one matrix equation:

$$Y = \boldsymbol{A} \cdot X. \tag{I.4.3}$$

Since the mapping which is defined by this equation is named L, we can also write: $Y = L(X) = \boldsymbol{A} \cdot X$. It is obvious that

$$L(\alpha \cdot X) = \boldsymbol{A} \cdot (\alpha \cdot X) = \alpha \cdot (\boldsymbol{A} \cdot X) = \alpha \cdot L(X)$$

$$\text{and} \quad L(X_1 + X_2) = \boldsymbol{A} \cdot (X_1 + X_2) = \boldsymbol{A} \cdot X_1 + \boldsymbol{A} \cdot X_2 = L(X_1) + L(X_2)$$

for all $\alpha \in \mathbb{R}$ and for all points $X,\ X_1,\ X_2 \in \mathbb{R}^n$. This means that the mapping L, defined by equations (I.4.2) (or by the equivalent matrix equation (I.4.3)), is a linear mapping of $\mathbb{R}^n$ into $\mathbb{R}^m$. We say that this mapping is *defined by matrix* $\boldsymbol{A}$. On the other hand, matrix $\boldsymbol{A}$ is said to be *the matrix of linear mapping* L.

I.4.4. One–to–one correspondence between linear mappings and matrices.

We have seen that each $m \times n$ matrix $\boldsymbol{A}$ defines a unique linear mapping $L : \mathbb{R}^n \longrightarrow \mathbb{R}^m$. It is natural to ask whether on the other hand every linear mapping of $\mathbb{R}^n$ to $\mathbb{R}^m$ defines a unique $m \times n$ matrix. The next theorem says that the answer is positive:

Theorem (on representation of a linear mapping by a matrix). *If L is a linear mapping of $\mathbb{R}^n$ into $\mathbb{R}^m$ then there exists a unique $m \times n$ matrix $\boldsymbol{A}$ such that $L(X) = \boldsymbol{A} \cdot X$ for all $X \in \mathbb{R}^n$.*

I.4.5. Remark.

There exists the following discord in mathematical terminology: The function $y = kx + q$ (where k, q are coefficients and x is variable) is called "linear" in the mathematical discipline called mathematical analysis. However, looking at this function from the point of view of linear algebra and regarding it as a mapping of $\mathbb{R}^1$ to $\mathbb{R}^1$ we can see that it is linear only if $q = 0$. Otherwise it is neither homogeneous nor additive (see paragraph I.4.1).

I.4.6. Example.

Linear mapping L of the space $\mathbb{R}^n$ to the space $\mathbb{R}^m$ is given by the matrix

$$\boldsymbol{A} = \begin{pmatrix} 2, & 3, & 1 \\ 2, & -1, & 2 \end{pmatrix}.$$

Specify numbers n and m. Choose X in $\mathbb{R}^n$ and find its image $Y = L(X)$ in $\mathbb{R}^m$.

Solution: m is equal to the number of rows of matrix $\boldsymbol{A}$, i.e. $m = 2$. n is equal to the number of columns of matrix $\boldsymbol{A}$, i.e. $n = 3$. L is thus a linear mapping of $\mathbb{R}^3$ into $\mathbb{R}^2$.

Let us choose $X \in \mathbb{R}^3$, for instance

$$X = \begin{pmatrix} 5 \\ 2 \\ 1 \end{pmatrix}.$$

Then it holds:

$$Y = L(X) = \boldsymbol{A} \cdot X = \begin{pmatrix} 2, & 3, & 1 \\ 2, & -1, & 2 \end{pmatrix} \cdot \begin{pmatrix} 5 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 17 \\ 10 \end{pmatrix}.$$

I.4.7. Example.

Does there exist a linear mapping L of the space $\mathbb{R}^2$ into itself such that $L(X_1) = Y_1$ and $L(X_2) = Y_2$? Points X_1, Y_1, X_2 and Y_2 are:

$$X_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \qquad Y_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad X_2 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \qquad Y_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}.$$

If such mapping exists, find its matrix.

Solution: A linear mapping L with required properties exists if and only if there exists a 2×2 matrix $\boldsymbol{A} = (a_{ij})$ such that $Y_1 = \boldsymbol{A} \cdot X_1$ and $Y_2 = \boldsymbol{A} \cdot X_2$. Substituting here the coordinates of the points X_1, Y_1, X_2 and Y_2, we obtain:

$$\begin{array}{rcrcl} 0 & = & 2a_{11} & + & a_{12} \\ 1 & = & 2a_{21} & + & a_{22} \end{array} \qquad \begin{array}{rcrcl} 1 & = & -a_{11} & - & a_{12} \\ -2 & = & -a_{21} & - & a_{22} \end{array}$$

This system for the unknowns $a_{11}, a_{12}, a_{21}, a_{22}$ has a unique solution $a_{11} = 1$, $a_{12} = -2$, $a_{21} = -1$, $a_{22} = 3$. Hence the linear mapping L exists, it is unique and its matrix is

$$\boldsymbol{A} = \begin{pmatrix} 1, & -2 \\ -1, & 3 \end{pmatrix}.$$

I.4.8. Example – change of the basis in $V(\mathbb{E}_3)$, change of the coordinate system in $\mathbb{E}_3$.

Let $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ be a basis in the vector space $V(\mathbb{E}_3)$. Then, in accordance with theorem I.1.15, each vector $\vec{x} \in V(\mathbb{E}_3)$ can be uniquely expressed in the form

$$\vec{x} = x_1\,\vec{e}_1 + x_2\,\vec{e}_2 + x_3\,\vec{e}_3 = \sum_{j=1}^{3} x_j\,\vec{e}_j. \tag{I.4.4}$$

Numbers x_1, x_2, x_3 are said to be the *coordinates of vector $\vec{x}$ with respect to the basis $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$*.

If $\vec{e}_1'$, $\vec{e}_2'$, $\vec{e}_3'$ is another basis in $V(\mathbb{E}_3)$ then

$$\vec{e}_j = a_{1j}\,\vec{e}_1' + a_{2j}\,\vec{e}_2' + a_{3j}\,\vec{e}_3' = \sum_{i=1}^{3} a_{ij}\,\vec{e}_i' \qquad (j = 1, 2, 3), \tag{I.4.5}$$

where a_{1j}, a_{2j}, a_{3j} are the coordinates of the vector $\vec{e}_j$ with respect to the basis $\vec{e}_1'$, $\vec{e}_2'$, $\vec{e}_3'$. If we substitute the expression (I.4.5) of $\vec{e}_j$ to (I.4.4), we obtain:

$$\vec{x} = x_1 \sum_{i=1}^{3} a_{i1}\,\vec{e}_i' + x_2 \sum_{i=1}^{3} a_{i2}\,\vec{e}_i' + x_3 \sum_{i=1}^{3} a_{i3}\,\vec{e}_i' = \sum_{j=1}^{3}\sum_{i=1}^{3} x_j\,a_{ij}\,\vec{e}_i'$$
$$= \sum_{i=1}^{3}\left(\sum_{j=1}^{3} a_{ij}\,x_j\right)\vec{e}_i' = \sum_{i=1}^{3} x_i'\,\vec{e}_i',$$

where

$$x_i' = \sum_{j=1}^{3} a_{ij}\,x_j.$$

Numbers x_1', x_2', x_3' are the coordinates of vector $\vec{x}$ relative to the basis $\vec{e}_1'$, $\vec{e}_2'$, $\vec{e}_3'$. We can see that the mapping which to each given $[x_1; x_2; x_3] \in \mathbb{R}^3$ assigns $[x_1'; x_2'; x_3'] \in \mathbb{R}^3$ is a linear mapping of $\mathbb{R}^3$ into itself, whose matrix is $\boldsymbol{A} = (a_{ij})$. $\boldsymbol{A}$ is said to be the *transition matrix* from the basis $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ to the basis $\vec{e}_1'$, $\vec{e}_2'$, $\vec{e}_3'$.

The result can also be interpreted in this way: The change of coordinates, corresponding to the transition from an affine coordinate system x_1, x_2, x_3 in $\mathbb{E}_3$ (which can be Cartesian, but it is not a condition), whose coordinate axes are oriented by vectors $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$, to a coordinate system x_1', x_2', x_3' (which has the same origin, however the coordinate axes are oriented by vectors $\vec{e}_1'$, $\vec{e}_2'$, $\vec{e}_3'$) is a linear mapping. Its matrix is matrix $\boldsymbol{A}$.

In the special case when both the coordinate systems x_1, x_2, x_3 and x_1', x_2', x_3' are cartesian and they have the same origin, the element a_{ij} in matrix $\boldsymbol{A}$ has this geometrical meaning: $a_{ij} = \cos\,\alpha$ where α is the angle between the nonnegative part of the x_i'–axis and the nonnegative part of the x_j–axis. (I.e. α is the angle between the vectors $\vec{e}_i'$ and $\vec{e}_j$.) This is why we often refer to numbers a_{ij} as *direction cosines* in this case.

I.4.9. The null space (the kernel) of a linear mapping.

Let L be a linear mapping of $\mathbb{R}^n$ to $\mathbb{R}^m$. The set of all points $X \in \mathbb{R}^n$ such that $L(X) = O$ is called the *null space* (or the *kernel*) of mapping L. We denote it $N(L)$ (or $Ker(L)$).

We deal only with linear mappings $L : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ whose domain is the whole space $\mathbb{R}^n$ in this textbook. However, the range $R(L)$ (= the image of $\mathbb{R}^n$) need not generally cover the whole space $\mathbb{R}^m$.

The following theorem gives the information about the structure of the null space $N(L)$ and the range $R(L)$ of the linear mapping L.

I.4.10. Theorem.

Let L be a linear mapping of $\mathbb{R}^n$ to $\mathbb{R}^m$. Then $N(L)$ is a subspace of $\mathbb{R}^n$ and $R(L)$ is a subspace of $\mathbb{R}^m$.

The proof is simple. Think over it for yourself. (You can use paragraph I.1.20.)

I.4.11. A composite mapping.

If L_1 is a linear mapping of $\mathbb{R}^n$ to $\mathbb{R}^m$ (given by an $m \times n$ matrix $\boldsymbol{A}_1$) and L_2 is a linear mapping of $\mathbb{R}^m$ to $\mathbb{R}^k$ (given by a $k \times m$ matrix $\boldsymbol{A}_2$) then one can create a composite mapping $L = L_2 \circ L_1$ of $\mathbb{R}^n$ to $\mathbb{R}^k$. The following theorem tells us more about this mapping.

Theorem (on a composite mapping). *Under the mentioned assumptions L is a linear mapping of $\mathbb{R}^n$ to $\mathbb{R}^k$ whose matrix is $\boldsymbol{A} = \boldsymbol{A}_2 \cdot \boldsymbol{A}_1$.*

I.4.12. Example.

Linear mappings $L_{\boldsymbol{A}}$ and $L_{\boldsymbol{B}}$ have the matrices

$$\boldsymbol{A} = \begin{pmatrix} -1, & 0 \\ 5, & 3 \\ -2, & 6 \end{pmatrix} \text{ and } \boldsymbol{B} = \begin{pmatrix} 0, & 8 \\ 2, & 2 \end{pmatrix}.$$

Which of the two composite mappings $L_{\boldsymbol{A}} \circ L_{\boldsymbol{B}}$, $L_{\boldsymbol{B}} \circ L_{\boldsymbol{A}}$ can be created? Determine the matrix of this composite mapping.

Solution: $\boldsymbol{A}$ is a 3×2 matrix, hence $L_{\boldsymbol{A}}$ is a mapping of $\mathbb{R}^2$ into $\mathbb{R}^3$. $\boldsymbol{B}$ is a 2×2 matrix, therefore $L_{\boldsymbol{B}}$ is a mapping of $\mathbb{R}^2$ to $\mathbb{R}^2$. This shows that the composite mapping $L_{\boldsymbol{A}} \circ L_{\boldsymbol{B}}$ (from $\mathbb{R}^2$ to $\mathbb{R}^3$) has a sense, while the composite mapping $L_{\boldsymbol{B}} \circ L_{\boldsymbol{A}}$ cannot be created. This is also seen from the fact that the matrix product $\boldsymbol{A} \cdot \boldsymbol{B}$ is defined (the result is a 3×2 matrix), while the matrix product $\boldsymbol{B} \cdot \boldsymbol{A}$ has no sense.

Matrix $\boldsymbol{C}$ of the composite mapping $L_{\boldsymbol{A}} \circ L_{\boldsymbol{B}}$ can be computed in this way:

$$\boldsymbol{C} = \boldsymbol{A} \cdot \boldsymbol{B} = \begin{pmatrix} -1, & 0 \\ 5, & 3 \\ -2, & 6 \end{pmatrix} \cdot \begin{pmatrix} 0, & 8 \\ 2, & 2 \end{pmatrix} = \begin{pmatrix} 0, & -8 \\ 6, & 46 \\ 12, & -4 \end{pmatrix}.$$

I.4.13. The inverse mapping.

It generally holds that the inverse mapping L_{-1} to mapping L exists if and only if L is one–to–one.

Let us now deal with the question when a linear mapping L is one–to–one and what form has the inverse mapping L_{-1} in the special case when the matrix of mapping L is a square $n \times n$ matrix. L is thus a linear mapping of $\mathbb{R}^n$ to $\mathbb{R}^n$. The next theorem gives the important information.

Theorem (on the inverse mapping). *Suppose that L is a linear mapping of $\mathbb{R}^n$ to $\mathbb{R}^n$ which is defined by an $n \times n$ matrix $\boldsymbol{A}$. Then L is one–to–one if and only if matrix $\boldsymbol{A}$ is regular. In this case, the inverse mapping L_{-1} is also a linear mapping of $\mathbb{R}^n$ into itself. Its matrix is $\boldsymbol{A}^{-1}$.*

I.4.14. Example.

L is a linear mapping of $\mathbb{R}^2$ to $\mathbb{R}^2$, defined by the matrix $\boldsymbol{A} = \begin{pmatrix} 3, & 1 \\ 0, & 2 \end{pmatrix}$.

Does there exist an inverse mapping to L?
If the inverse mapping exists, find its matrix.

Solution: $\det \boldsymbol{A} = 3 \cdot 2 - 0 \cdot 1 = 6 \neq 0$, hence matrix $\boldsymbol{A}$ is regular and the inverse matrix $\boldsymbol{A}^{-1}$ exists (see theorem I.2.31). Thus, the inverse mapping L_{-1}, also exists and it is a linear mapping of $\mathbb{R}_2$ to $\mathbb{R}_2$. Its matrix is matrix $\boldsymbol{A}^{-1}$. The inverse matrix can be obtained either by means of the formula from paragraph I.2.35 or it can be calculated by the method which is explained in the textbook [NK], pp. 9–10. We show only the result:

$$\boldsymbol{A}^{-1} = \begin{pmatrix} \frac{1}{3}, & -\frac{1}{6} \\ 0, & \frac{1}{2} \end{pmatrix}.$$

I.4.15.* Remark.

The inverse mapping can also exist to the linear mapping $L : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ in the case when $m > n$ and the matrix of the mapping is therefore nor a square matrix. Then the domain of the inverse mapping is not the whole space $\mathbb{R}^m$. However, we do not deal with such mappings in this text.

I.4.16.* Remark.

Assume again that L is a linear mapping of $\mathbb{R}^n$ to $\mathbb{R}^m$ defined by an $m \times n$ matrix $\boldsymbol{A}$. It can be proved that the dimensions of the null space $N(L)$ and the range $R(L)$ of mapping L satisfy the formulas

$$\dim N(L) = n - \mathrm{r}(\boldsymbol{A}), \qquad \dim R(L) = \mathrm{r}(\boldsymbol{A}). \tag{I.4.6}$$

Try to prove the following simple assertions for yourself.

a) The range $R(L)$ of the linear mapping L covers the whole space $\mathbb{R}^m$ if and only if its dimension is the same as the dimension of $\mathbb{R}^m$, i.e. if $\mathrm{r}(\boldsymbol{A}) = m$.

b) Linear mapping L is one–to–one if and only if $\dim N(L) = 0$. (It is, due to the first formula in (I.4.6), satisfied if and only if $\mathrm{r}(\boldsymbol{A}) = n$.)

I.4.17. Remark.

We have in remind that $V(\mathbb{E}_n)$ is the vector space of all (free) vectors from $\mathbb{E}_n$. If we make an agreement that the coordinates of vectors in $V(\mathbb{E}_n)$ and in $V(\mathbb{E}_m)$ will be written as columns (i.e. as $n \times 1$ or $m \times 1$ matrices) then we can define and treat linear mappings of $V(\mathbb{E}_n)$ to $V(\mathbb{E}_m)$ in exactly same way as the linear mappings of $\mathbb{R}^n$ do $\mathbb{R}^m$.

The contents of this section could naturally be extended to the case when L is a linear mapping of a general vector space V to a general vector space W. The theory of such mappings has many applications. However, we do not study the linear mappings on such a general level in this text.

I.5. Eigenvalues and eigenvectors of square matrices

I.5.1. Motivation.

In order to keep a simple notation, it will be again advantageous to identify vectors with n–tuples of numbers, written in a column. Such vectors can be regarded as $n \times 1$ matrices. Thus, we shall denote vectors in the same way as matrices (i.e. by capital letters) in this chapter.

Suppose that $\boldsymbol{A}$ is an $n \times n$ square matrix. In mathematics and its applications, one often solves the question whether there exists a nonzero vector $\boldsymbol{X} \in V(\mathbb{E}_n)$. such that the product $\boldsymbol{A} \cdot \boldsymbol{X}$ is a vector, laying on the same straight line as $\boldsymbol{X}$. The fact that $\boldsymbol{A} \cdot \boldsymbol{X}$ and $\boldsymbol{X}$ lay on the same straight line means that $\boldsymbol{A} \cdot \boldsymbol{X} = \lambda \boldsymbol{X}$ for an appropriate number λ. The important problem is not only to find such vectors $\boldsymbol{X}$, but also the corresponding numbers λ. Solution of this problem plays a big role for instance in the theory of stability of mechanical systems.

The equation $\boldsymbol{A} \cdot \boldsymbol{X} = \lambda \boldsymbol{X}$ (for the unknown λ) can generally have complex solutions. That is why we admit the possibility that λ is a complex number and the entries (coordinates) of vector $\boldsymbol{X}$ are also complex and not only real numbers. On the other hand, the square matrix $\boldsymbol{A}$, we work with in this section is always supposed to have only real entries.

I.5.2. Eigenvalues, eigenvectors.

A complex number λ is called an *eigenvalue* of a square matrix $\boldsymbol{A}$ if there exists a nonzero vector $\boldsymbol{X}$ such that $\boldsymbol{A} \cdot \boldsymbol{X} = \lambda \boldsymbol{X}$. Such a vector $\boldsymbol{X}$ is called an *eigenvector* of matrix $\boldsymbol{A}$ corresponding to the eigenvalue λ.

I.5.3. Remark.

The eigenvector is not determined uniquely, the number of eigenvectors corresponding to the eigenvalue λ is always infinite. Clearly, if $\boldsymbol{A} \cdot \boldsymbol{X} = \lambda \boldsymbol{X}$ (i.e. $\boldsymbol{X}$ is the eigenvector of matrix $\boldsymbol{A}$ corresponding to the eigenvalue λ) and $k \in \mathbb{C}$, $k \neq 0$, then it also holds: $\boldsymbol{A} \cdot (kX) = \lambda(kX)$. This means that kX is also the eigenvector of $\boldsymbol{A}$ corresponding to the eigenvalue λ.

I.5.4. How to find the eigenvalues.

The equation $\boldsymbol{A} \cdot \boldsymbol{X} = \lambda \boldsymbol{X}$ can be written in the equivalent form $\boldsymbol{A} \cdot \boldsymbol{X} - \lambda \boldsymbol{E} \cdot \boldsymbol{X} = \boldsymbol{O}$ or $(\boldsymbol{A} - \lambda \boldsymbol{E}) \cdot \boldsymbol{X} = \boldsymbol{O}$. ($\boldsymbol{E}$ is the $n \times n$ identity matrix and $\boldsymbol{O}$ is the zero vector, i.e. it is the $n \times 1$ zero matrix). The vector equation $(\boldsymbol{A} - \lambda \boldsymbol{E}) \cdot \boldsymbol{X} = \boldsymbol{O}$ can be regarded as the homogeneous system of linear algebraic equations for the unknown components of the vector $\boldsymbol{X}$. This system has a nonzero solution if and only if the rank of the matrix of the system, i.e. the matrix $\boldsymbol{A} - \lambda \boldsymbol{E}$, is less than n. (See theorem I.3.4 and remark I.3.6.) The inequality $\mathrm{r}(\boldsymbol{A} - \lambda \boldsymbol{E}) < n$ means that the matrix $(\boldsymbol{A} - \lambda \boldsymbol{E})$ is singular (see paragraph I.2.29), which is true if and only if

$$\det(\boldsymbol{A} - \lambda \boldsymbol{E}) = 0. \tag{I.5.1}$$

(See theorem I.2.31.) Hence a nonzero vector $\boldsymbol{X}$ satisfying the equation $\boldsymbol{A} \cdot \boldsymbol{X} = \lambda \boldsymbol{X}$ exists if and only if (I.5.1) holds. Equation (I.5.1) (for the unknown λ) is called the *characteristic equation* of matrix $\boldsymbol{A}$. Solving the characteristic equation, we obtain all eigenvalues of matrix $\boldsymbol{A}$.

I.5.5. Example.

Find eigenvalues of the matrix $\boldsymbol{A} = \begin{pmatrix} 3, & 4 \\ 5, & 5 \end{pmatrix}$.

Solution: The characteristic equation of matrix $\boldsymbol{A}$ is

$$\det(\boldsymbol{A}-\lambda\boldsymbol{E}) = \begin{vmatrix} 3-\lambda, & 4 \\ 5, & 5-\lambda \end{vmatrix} = (3-\lambda)\cdot(5-\lambda)-4\cdot 5 = \lambda^2-8\lambda-5 = 0.$$

Its solution is $\lambda_1 = 4+\sqrt{21}$ a $\lambda_2 = 4-\sqrt{21}$.

I.5.6. How to find the eigenvectors.

Eigenvectors corresponding to the eigenvalue λ can be obtained by solving the homogeneous system of linear algebraic equations $(\boldsymbol{A}-\lambda\boldsymbol{E})\cdot\boldsymbol{X} = \boldsymbol{O}$ (for unknown components $x_1,\dots,x_n$ of the vector $\boldsymbol{X}$).

I.5.7. Example.

Find eigenvectors of the matrix $\boldsymbol{A}$ from example I.5.5 which correspond to the eigenvalue $\lambda_1 = 4+\sqrt{21}$.

Solution: Substituting the value $4+\sqrt{21}$ for λ in the vector equation $(\boldsymbol{A}-\lambda\boldsymbol{E})\cdot\boldsymbol{X} = \boldsymbol{O}$ and expressing this equation in coordinates, we obtain the system of linear algebraic equations

$$\begin{aligned} (-1-\sqrt{21})\,x_1 &+ & 4\,x_2 &= 0,\\ 5\,x_1 &+ & (1-\sqrt{21})\,x_2 &= 0. \end{aligned}$$

This system has infinitely many solutions: $x_1 = (1-\sqrt{21})p$, $x_2 = -5p$ (where $p\in\mathbb{C}$). Every vector $\boldsymbol{X}$ with these coordinates x_1, x_2 (where $p\neq 0$ because $\boldsymbol{X}$ cannot be the zero vector) is the eigenvector.

I.5.8. Remark.

If λ is a real eigenvalue of the matrix $\boldsymbol{A}$ then the system of equations $(\boldsymbol{A}-\lambda\boldsymbol{E})\cdot\boldsymbol{X} = \boldsymbol{O}$ (for unknown components of vector $\boldsymbol{X}$) has a matrix all of whose entries are real numbers. Therefore it is possible to find a nonzero solution $\boldsymbol{X}$ of this system with all components also being real numbers. This means that to real eigenvalues there exist real eigenvectors.

I.5.9. Remark.

The eigenvalues and the eigenvectors of square matrices have a series of interesting properties. Their detailed explication, proofs and examples of some applications would go beyond the scope of this text. Nevertheless, we still present several simple assertions in this remark.

a) *The eigenvectors, corresponding to different eigenvalues, are linearly independent.* (Let us show that this statement is true, for simplicity, only for the family of two eigenvectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, corresponding to the eigenvalues λ_1 and λ_2: if the vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are linearly dependent then there exists $k\in\mathbb{R}$, $k\neq 0$, such that $\boldsymbol{X}_2 = k\boldsymbol{X}_1$. Then $\boldsymbol{A}\cdot\boldsymbol{X}_2 = \boldsymbol{A}\cdot k\boldsymbol{X}_1 = k\boldsymbol{A}\cdot\boldsymbol{X}_1 = k\lambda_1\boldsymbol{X}_1 = \lambda_1\boldsymbol{X}_2$. We observe that $\boldsymbol{X}_2$ is an eigenvector, corresponding to the eigenvalue λ_1. This means that $\lambda_2 = \lambda_1$. Thus, on the other hand, if $\lambda_1\neq\lambda_2$ then the eigenvectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ cannot be linearly dependent.)

b) *0 is the eigenvalue of matrix $\boldsymbol{A}$ if and only if $\boldsymbol{A}$ is singular.* ($\boldsymbol{A}$ has the eigenvalue 0 if and only if there existrs a nonzero vector $\boldsymbol{X}$, such that $\boldsymbol{A}\cdot\boldsymbol{X} = 0\cdot\boldsymbol{X} = \boldsymbol{O}$. This vectorial equation can be regarded as a system of n linear algebraic equations with the square matrix $\boldsymbol{A}$. Such a system has a nonzero solution if and only if matrix $\boldsymbol{A}$ is singular – see remark I.3.6.)

Next simple statements are listed without proofs:

c) *If λ is an eigenvalue of matrix $\boldsymbol{A}$ and $\boldsymbol{X}$ is the corresponding eigenvector then $\overline{\lambda}$ is also an eigenvalue of $\boldsymbol{A}$ and $\overline{X}$ is the corresponding eigenvector.*

d) *If λ is an eigenvalue of matrix $\boldsymbol{A}$ and $\boldsymbol{X}$ is the corresponding eigenvector then λ^2 is the eigenvalue of matrix $\boldsymbol{A}^2$ and $\boldsymbol{X}$ is the corresponding eigenvector.*

e) If the inverse matrix $\boldsymbol{A}^{-1}$ exists then λ is an eigenvalue of matrix $\boldsymbol{A}$ if and only if $\frac{1}{\lambda}$ is the eigenvalue of matrix $\boldsymbol{A}^{-1}$. The corresponding eigenvectors are the same.

f)* *If $\boldsymbol{A}$ is a symmetric square matrix then all its eigenvalues are real. The eigenvectors, corresponding to different eigenvalues, are perpendicular.*

I.5.10. Example.

$\boldsymbol{A}$ is a 3×3 square matrix. Decide whether $\boldsymbol{A}$ can have the given eigenvalues, eventually also the given eigenvectors:

a) $2,\ 3$ b) $3,\ 2+i,\ -3-2i$ c) $5+i,\ 5-i,\ 7$

d) $4,\ 3-i,\ 3+i,\ 5$ e) $7,\ 5,\ 1,\ \begin{pmatrix}-1\\2\\3\end{pmatrix},\ \begin{pmatrix}2\\3\\1\end{pmatrix},\ \begin{pmatrix}4\\6\\2\end{pmatrix}$

Solution:

a) Matrix $\boldsymbol{A}$ can have the eigenvalues 2 and 3. However, these need not be all eigenvalues. The characteristic equation of $\boldsymbol{A}$ is a cubic equation and it can have three different roots.

b) If $2+i$ is an eigenvalue of matrix $\boldsymbol{A}$ then $2-i$ must also be an eigenvalue of $\boldsymbol{A}$. (See remark I.5.10, item c).) By analogy, since $-3-2i$ is an eigenvalue, the complex conjugate $-3+2i$ is an eigenvalue, too. However, matrix $\boldsymbol{A}$ cannot have five eigenvalues $3,\ 2\pm i$ and $-3\pm 2i$ because it is a 3×3 matrix and as such, it can have at most three different eigenvalues.

c) $5+i,\ 5-i\ \ 7$ can be eigenvalues of matrix $\boldsymbol{A}$.

d) The given numbers are four, hence the answer is negative.

e) The given numbers can be the eigenvalues of matrix $\boldsymbol{A}$. The numbers are all different and so the eigenvectors should be linearly independent. (See remark I.5.10, item a).) This is not fulfilled because the third vector is a multiple of the second one. Thus, the answer is negative.

I.5.11. Problems.

Find eigenvalues and eigenvectors of the following matrices.

a) $\begin{pmatrix}2, & 1\\1, & 2\end{pmatrix}$; b) $\begin{pmatrix}3, & 4\\5, & 2\end{pmatrix}$; c) $\begin{pmatrix}0, & a\\-a, & 0\end{pmatrix}$ $(a\neq 0)$; d) $\begin{pmatrix}5, & 6, & -3\\-1, & 0, & 1\\1, & 2, & -1\end{pmatrix}$.

Results:

a) $\lambda_1 = 3,\ \boldsymbol{X}_1 = \begin{pmatrix}p\\p\end{pmatrix},\ \lambda_2 = 1,\ \boldsymbol{X}_2 = \begin{pmatrix}p\\-p\end{pmatrix}$

b) $\lambda_1 = 7,\ \boldsymbol{X}_1 = \begin{pmatrix}p\\p\end{pmatrix},\ \lambda_2 = -2,\ \boldsymbol{X}_2 = \begin{pmatrix}-4p\\5p\end{pmatrix}$

c) $\lambda_{1,2} = \pm a\,i,\ \boldsymbol{X}_{1,2} = \begin{pmatrix} p \\ \pm p\,i \end{pmatrix}$

d) $\lambda = 2,\ \boldsymbol{X} = \begin{pmatrix} -2p+q \\ p \\ q \end{pmatrix}$.

(The parameters p, q can be any complex numbers such that the corresponding eigenvector is non–zero.)

I.6. Survey of equivalent properties of a square matrix

We do not explain any new notion or a new method in this chapter. On the other hand, you already know everything which follows. Nevertheless, if you are studying linear algebra for the first time, it is possible that your knowledge is still not enough classified and various assertions and theorems are not mutually interconnected. In order to show clearly the connections, we give a survey of equivalent statements one can make about a square matrix $\boldsymbol{A}$ in this chapter. (This means that for a given square matrix $\boldsymbol{A}$ either all statements are true or all statements are false.)

We assume that $\boldsymbol{A}$ is an $n \times n$ square matrix.
Then the following statements are equivalent:

1. *Matrix $\boldsymbol{A}$ is regular.*
2. $\det \boldsymbol{A} \neq 0$.
   (See theorem I.2.31.)
3. *An inverse matrix $\boldsymbol{A}^{-1}$ exists.*
   (See theorem I.2.31.)
4. *The rank of matrix $\boldsymbol{A}$ is n.*
   (See the definition of a regular matrix in paragraph I.2.29.)
5. *The rows of matrix $\boldsymbol{A}$ are linearly independent.*
   (See the definition of the rank of a matrix in paragraph I.2.14 and statement 4.)
6. *The columns of matrix $\boldsymbol{A}$ are linearly independent.*
   (See remark I.2.16 and statement 4.)
7. *The homogeneous system of linear algebraic equations $\boldsymbol{A} \cdot \boldsymbol{X} = \boldsymbol{O}$ has a unique (i.e. zero) solution.*
   (See remark I.3.6.)
8. *The general (i.e. homogeneous or non–homogeneous) system of linear algebraic equations $\boldsymbol{A} \cdot \boldsymbol{X} = \boldsymbol{B}$ has a unique solution.*
   (See the Frobenius theorem I.3.7.)
9. *0 is not an eigenvalue of matrix $\boldsymbol{A}$.*
   (See remark I.5.9, item b).)

Naturally, the negations of all these statements are also equivalent. Formulate and write all the negations for yourself.

CHAPTER II.

Analytic geometry in $\mathbb{E}_3$

II.1. Some basic notions

II.1.1. Cartesian coordinates in $\mathbb{E}_3$.

To locate points and other objects in space $\mathbb{E}_3$, we use three mutually perpendicular coordinate axes, intersecting at one point. We denote them x, y, z or x_1, x_2, x_3. Their orientation is chosen so that they make a *right handed* system. This means that when you hold your right hand so that the fingers curl from the positive part of the x–axis toward the positive part of the y–axis, your thumb points along the positive part of the z–axis. The intersection of all three axes is called the *origin* of the coordinate system.

To each point in $\mathbb{E}_3$, we can uniquely assign its *Cartesian coordinates* – they are successively the distances of orthogonal projections of the considered point onto the axes x, y, z from the origin, taken with the "+" sign if the projection lays on the positive part of the axis and with the "−" sign in the opposite case.

By analogy, we can uniquely assign the Cartesian coordinates to each free vector in $\mathbb{E}_3$ – we choose the concrete position of the vector so that its initial point is at the origin of the coordinate system and we regard the cartesian coordinates of its end point as the cartesian coordinates of the vector. (See also paragraph I.1.2.)

To distinguish between points and free vectors in $\mathbb{E}_3$, we write the Cartesian coordinates of points in $\mathbb{E}_3$ in brackets (for example $[1;2;3]$) and the Cartesian coordinates of vectors in $\mathbb{E}_3$ in parentheses (for instance $(-1;2;5)$).

II.1.2. The length (the magnitude) of a vector in $\mathbb{E}_3$.

If $\vec{u} = (u_1; u_2; u_3)$ is a vector in $\mathbb{E}_3$ then the number

$$\|\vec{u}\| \;=\; \sqrt{u_1^2 + u_2^2 + u_3^2}$$

is called its *length* (or its *magnitude*).

II.1.3. The scalar product of vectors in $\mathbb{E}_3$.

If $\vec{u} = (u_1; u_2; u_3)$ and $\vec{v} = (v_1; v_2; v_3)$ are vectors in $\mathbb{E}_3$ then the number

$$\vec{u} \cdot \vec{v} \;=\; u_1 \cdot v_1 + u_2 \cdot v_2 + u_3 \cdot v_3.$$

is said to be their *scalar product* (or their *dot product*). (Compare with the scalar product of arithmetic vectors, defined in paragraph I.2.11.)

II.1.4. Theorem.

If $\vec{u}$ and $\vec{v}$ are non–zero vectors in $\mathbb{E}_3$ and φ is the angle between the vectors $\vec{u}$ and $\vec{v}$ then

$$\vec{u} \cdot \vec{v} \;=\; \|\vec{u}\| \cdot \|\vec{v}\| \cdot \cos\,\varphi.$$

P r o o f :

Let us choose an arbitrary point A in $\mathbb{E}_3$ and put $B = A + \vec{u}$ and $C = A + \vec{v}$. Obviously, the vector $B - C$ can also be written as $\vec{u} - \vec{v}$. Applying the cosine theorem to the triangle ABC, we obtain: $\|B - C\|^2 = \|B - A\|^2 + \|C - A\|^2 - 2\,\|B - A\|\,\|C - A\| \cdot \cos\,\varphi$, or $\|\vec{u} - \vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\,\|\vec{u}\|\,\|\vec{v}\| \cdot \cos\,\varphi$. This yields: $(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2 =$

$u_1^2+u_2^2+u_3^2+v_1^2+v_2^2+v_3^2-2\,\|\vec{u}\|\,\|\vec{v}\|\,\cos\varphi$. The desired formula is an easy consequence of this equality.

II.1.5. Remark.

Thus, the non–zero vectors $\vec{u}$, $\vec{v}$ are perpendicular if and only if $\vec{u}\cdot\vec{v}=0$.

II.1.6. Another form of vectors in $\mathbb{E}_3$.

Denote by $\widehat{\imath}=(1;0;0)$, $\widehat{\jmath}=(0;1;0)$, $\widehat{k}=(0;0;1)$. If $\vec{u}=(u_1;u_2;u_3)$ is a vector in $\mathbb{E}_3$ then it can also be written down in this form: $\vec{u}=u_1\cdot\widehat{\imath}+u_2\cdot\widehat{\jmath}+u_3\cdot\widehat{k}$.

II.1.7. The vector product of vectors in $\mathbb{E}_3$.

Let $\vec{u}=(u_1;u_2;u_3)$ and $\vec{v}=(v_1;v_2;v_3)$ be vectors in $\mathbb{E}_3$. The vector

$$\vec{u}\times\vec{v}=\begin{vmatrix}\widehat{\imath}, & \widehat{\jmath}, & \widehat{k}\\ u_1, & u_2, & u_3\\ v_1, & v_2, & v_3\end{vmatrix}=(u_2v_3-u_3v_2)\cdot\widehat{\imath}+(u_3v_1-u_1v_3)\cdot\widehat{\jmath}+(u_1v_2-u_2v_1)\cdot\widehat{k}$$

is called the *vector product* (or the *cross product*) of $\vec{u}$ and $\vec{v}$ (in this order).

II.1.8. Theorem.

If $\vec{u}$ and $\vec{v}$ are non–zero vectors in $\mathbb{E}_3$ and φ is the angle between them then

a) the vector $\vec{u}\times\vec{v}$ is perpendicular to both the vectors $\vec{u}$ and $\vec{v}$;

b) $\|\vec{u}\times\vec{v}\|=\|\vec{u}\|\,\|\vec{v}\|\cdot\sin\varphi$;

We omit the proof of this theorem. However, the assertion a) can easily be verified computing the scalar products $(\vec{u}\times\vec{v})\cdot\vec{u}$, $(\vec{u}\times\vec{v})\cdot\vec{v}$. All of them are equal to zero. The vector product $\vec{u}\times\vec{v}$ is oriented in accordance with the so called right–hand rule: *$\vec{u}\times\vec{v}$ points the way your right thumb points when your fingers curl through the angle between $\vec{u}$ and $\vec{v}$ from $\vec{u}$ to $\vec{v}$.*

II.1.9. The sum of a point and a vector.

If $A=[\,a_1;a_2;a_3\,]$ is a point in $\mathbb{E}_3$ and $\vec{u}=(u_1;u_2;u_3)$ is a vector in $\mathbb{E}_3$, then the *sum of point A and vector $\vec{u}$* is equal to the point $B=[a_1+u_1;a_2+u_2;a_3+u_3\,]$ in $\mathbb{E}_3$. We write: $B=A+\vec{u}$.

On the other hand, the *difference of two points* $B=[\,b_1;b_2;b_3\,]$ and $A=[\,a_1;a_2;a_3\,]$ in $\mathbb{E}_3$ (in this order) is defined to be equal to the vector $\vec{u}=(b_1-a_1;b_2-a_2;b_3-a_3)$. We write: $\vec{u}=B-A$.

II.1.10. The distance between two points.

Remember that if $A=[\,a_1;a_2;a_3\,]$ and $B=[\,b_1;b_2;b_3\,]$ are two points from $\mathbb{E}_3$ then their distance $\|B-A\|$ is

$$\|B-A\|=\sqrt{(a_1-b_1)^2+(a_2-b_2)^2+(a_3-b_3)^2}.$$

(It is the length of the vector $B-A$.)

II.1.11. The distance between a point and a set; the distance between two sets.

If $A=[\,a_1;a_2;a_3\,]$ is a point in $\mathbb{E}_3$ and $M\subset\mathbb{E}_3$, then the *distance between point A and set M* is defined as follows:

$$d(A,M)=\inf\{\,\|A-X\|;\ X\in M\,\}.$$

(See p. 50 for the definition of "inf".) By analogy, the *distance between sets M and N* in $\mathbb{E}_3$ is defined by the formula

$$d(M,N) \;=\; \inf\{\,\|X-Y\|;\; X\in M,\; Y\in N\,\}.$$

II.2. Straight lines in $\mathbb{E}_3$

II.2.1. A straight line in $\mathbb{E}_3$.

If $A=[\,a_1;a_2;a_3\,]$ is a point in $\mathbb{E}_3$ and $\vec{u}=(u_1;u_2;u_3)$ is a non–zero vector in $\mathbb{E}_3$, then the set $p=\{\,X\in\mathbb{E}_3;\;\exists t\in\mathbb{R}:\;X=A+t\cdot\vec{u}\,\}$ is called a *straight line* in $\mathbb{E}_3$. The equation

(II.2.1) $$X \;=\; A \;+\; t\cdot\vec{u};\qquad t\in\mathbb{R}$$

is called the *parametric equation* for straight line p. This equation is also often written down in a coordinate form:

(II.2.2) $$x_1=a_1+t\cdot u_1,\quad x_2=a_2+t\cdot u_2,\quad x_3=a_3+t\cdot u_3;\qquad t\in\mathbb{R}.$$

The straight line p is said to be *parallel* to $\vec{u}$ and on the other hand, $\vec{u}$ is called the *directional vector* of p.

If parameter t in (II.2.1) is taken not from the whole set $\mathbb{R}$, but for example only from the interval $\langle -1;5\rangle$, then (II.2.1) is the parametrization of a line segment. Its end points are $A-\vec{u}$ and $A+5\vec{u}$. Similarly, if t is taken for instance from the interval $\langle 1;+\infty)$ then (II.2.1) parametrizes a half–line, etc.

II.2.2. The distance from a point to a straight line.

Let $p:\;\; X=A+t\cdot\vec{u};\; t\in\mathbb{R}$ be a straight line in $\mathbb{E}_3$ and M be a point in $\mathbb{E}_3$ which does not lie on straight line p. We show two methods for evaluating the distance $d(M,p)$.

1st method: Denote by P the nearest point to M on straight line p. (It can be proved that such a point exists and the vector $M-P$ is perpendicular to p.) Thus, for some (still unknown) value of parameter t, it holds: $P=A+t\cdot\vec{u}$. The value of t can be evaluated from the equation $(M-P)\cdot\vec{u}=0$, i.e. $(M-P)\cdot\vec{u}-(\vec{u}\cdot\vec{u})\,t=0$. Substituting now for t back to the formula $P=A+t\cdot\vec{u}$, we obtain the concrete position of P. The distance $d(M,p)$ is equal to $|P-M|$.

2nd method: AMP is the right triangle. Hence

$$d(M,p)=\|M-P\|=\|M-A\|\,\sin\varphi.$$

It follows from theorem II.1.8 that

$$\|(M-A)\times\vec{u}\|=\|M-A\|\,\|\vec{u}\|\,\sin\varphi.$$

Expressing $\sin\varphi$ from this equality and substituting it to the preceding equality, we obtain:

(II.2.3) $$d(M,p)\;=\;\frac{\|(M-A)\times\vec{u}\|}{\|\vec{u}\|}.$$

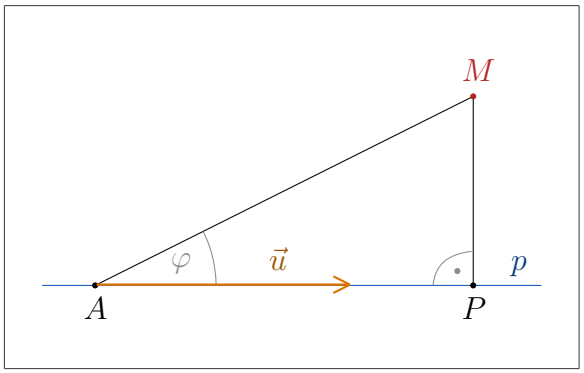


Fig. 1

II.2.3. Two straight lines in space.

Let p and q be two straight lines in $\mathbb{E}_3$ whose parametric equations are

$$p: \quad X = A + t \cdot \vec{u}; \quad t \in \mathbb{R}, \qquad\qquad q: \quad Y = B + s \cdot \vec{v}; \quad s \in \mathbb{R}.$$

We say that p and q are

a) *identical* (if they have infinitely many common points),
b) *intersecting* (if they have just one common point),
c) *parallel*, (if they have no common points and vectors $\vec{u}$, $\vec{v}$ are linearly dependent),
d) *skew*, (if they have no common points and vectors $\vec{u}$, $\vec{v}$ are linearly independent).

Let us now deal with the question how to distinguish between these cases in a particular situation. We look for common points of straight lines p and q. This task can be reformulated: We seek for such values of parameters t, s that using them in the parametric equations for p and q, we get the same point, i.e. $X = Y$. This yields: $A + t \cdot \vec{u} = B + s \cdot \vec{v}$. Rewriting this into coordinates, we obtain the system of three linear equations for two unknowns t, s:

$$a_1 + t \cdot u_1 = b_1 + s \cdot v_1\,, \qquad a_2 + t \cdot u_2 = b_2 + s \cdot v_2\,, \qquad a_3 + t \cdot u_3 = b_3 + s \cdot v_3\,.$$

This is equivalent to

$$\begin{aligned} u_1 \cdot t &- v_1 \cdot s &&= b_1 - a_1\,, \\ u_2 \cdot t &- v_2 \cdot s &&= b_2 - a_2\,, \\ u_3 \cdot t &- v_3 \cdot s &&= b_3 - a_3\,. \end{aligned}$$

If the number of solutions is infinite then the number of common points of straight lines p, q is also infinite and so the straight lines are identical.

If the above system has a unique solution, straight lines p, q have only one common point and so they are intersecting.

If the above system of equations has no solution then straight lines p, q have no common point. Thus, they are either parallel or skew. This depends on directional vectors $\vec{u}$, $\vec{v}$. If they are linearly dependent, straight lines p, q are parallel, otherwise they are skew.

II.2.4. A straight line given by two points.

If A and B are two different points in $\mathbb{E}_3$, then the straight line passing through these points can be described for example by the following parametric equation:

(II.2.4) $$X = A + t \cdot (B - A); \quad t \in \mathbb{R}.$$

II.2.5. Remark.

Suppose that points A, B, C, D, E are different points, all lying on the same straight line p. Put for example $\vec{u} = B - A$, $\vec{v} = 2 \cdot (B - D)$. Then the equations

$$\begin{aligned} X &= A + t \cdot (B - A); && t \in \mathbb{R}, \\ Y &= B + s \cdot (C - A); && s \in \mathbb{R}, \\ Z &= D + r \cdot \vec{u}; && r \in \mathbb{R}, \\ X &= E + \alpha \cdot \vec{v}; && \alpha \in \mathbb{R} \end{aligned}$$

are all parametric equations for line p. (Why?)

II.2.6. A secant of two lines.

A straight line which is intersecting the two lines p and q is called the *secant line* of p and q.

II.2.7. Example.

Lines p, q are given by their parametric equations

$$p: \quad X = A + t\cdot\vec{u}; \quad t\in\mathbb{R}, \qquad\qquad q: \quad X = B + s\cdot\vec{v}; \quad t\in\mathbb{R},$$

where $A=[3;0;-1]$, $\vec{u}=(2;1;1)$, $B=[1;1;1]$, $\vec{v}=(1;3;-2)$. Find a straight line r which intersects both lines p and q and which passes through the point $M=[-7;-6;5]$.

Solution: Denote by P (respectively by Q) the point of intersection of line r with line p (respectively with line q). Then there exist such values of parameters t and s that

$$P = A + t\cdot\vec{u}, \qquad\qquad Q = A + s\cdot\vec{v}. \tag{II.2.5}$$

All of points P, Q, M lie on line r and $M\neq Q$, because point M does not lie on the line q. (This can easily be verified: If $M\in p$ then the vectors $M-B$ and $\vec{v}$ would be linearly dependent. However, this is not true, because $(M-B)=(-8;-7;4)$, $\vec{v}=(1;3;-2)$ and it is seen that none of these vectors is equal to a multiple of the second vector.) Thus, there exists a number $\alpha\in\mathbb{R}$ such that

$$P - M = \alpha\cdot(Q-M).$$

Substituting here the above expression of points P and Q, we obtain the equation:

$$A + t\cdot\vec{u} - M = \alpha\cdot(B + s\cdot\vec{v} - M).$$

Writing this in coordinates and using here known coordinates of points A, B, M and vectors $\vec{u}$, $\vec{v}$, we obtain a system of three linear algebraic equations for the unknowns t, αs, α:

$$2\,t - \alpha s - 8\,\alpha = -10, \qquad t - 3\,\alpha s - 7\,\alpha = -6, \qquad t + 2\,\alpha s + 4\,\alpha = 6.$$

This system has the unique solution $t=2$, $\alpha s=-2$, $\alpha=2$. Hence, we also get: $s=-1$. When used in (II.2.5), this gives: $P=[7;2;1]$, $Q=[0;-2;3]$. A parametric equation for line r can now be written down for example by means of (II.2.4), i.e. as the equation of a straight line passing through two known points P, Q:

$$X = P + \tau\cdot(Q-P); \quad \tau\in\mathbb{R},$$
$$X = [7;2;1] + \tau\cdot(-7;-4;2); \quad \tau\in\mathbb{R}.$$

The above equation can also be written in coordinates:

$$x_1 = 7 - 7\tau, \qquad x_2 = 2 - 4\tau, \qquad x_3 = 1 + 2\tau; \qquad \tau\in\mathbb{R}.$$

II.2.8. The distance between two straight lines.

Suppose that the straight lines $p: X = A + t\cdot\vec{u};\ t\in\mathbb{R}$ and $q: Y = B + s\cdot\vec{v};\ s\in\mathbb{R}$ are parallel or skew.

To find the distance $d(p,q)$, we first find a straight line r which intersects lines p, q and is perpendicular to both of them. Denote by P, Q points of intersection of r with lines p, q. Then $P = A + t\cdot\vec{u}$, $Q = B + s\cdot\vec{v}$. for some values of parameters t, s. Line r is parallel to the vector $Q-P$. The vector $Q-P$ is orthogonal to vectors $\vec{u}$ and $\vec{v}$. So $(Q-P)\cdot\vec{u}=0$, $(Q-P)\cdot\vec{v}=0$. Substituting here from the equations for points P and Q, we obtain (after an easy rearrangement):

$$(\vec{v}\cdot\vec{u})\,s - (\vec{u}\cdot\vec{u})\,t = -(B-A)\cdot\vec{u}, \qquad (\vec{v}\cdot\vec{v})\,s - (\vec{u}\cdot\vec{v})\,t = -(B-A)\cdot\vec{v}.$$

We solve this system of two linear algebraic equations for unknowns t, s and use the values of t, s in the formulas for P and Q. Finally, the distance between lines p and q can be calculated as the distance between points P and Q: $d(p,q) = \|P-Q\|$.

II.2.9. The angle between two straight lines.

Suppose that p, q are straight lines in $\mathbb{E}_3$ whose parametrizations are: $p:\ X = A+t\cdot\vec{u};\ t\in\mathbb{R}$ and $q:\ X = B+s\cdot\vec{v};\ s\in\mathbb{R}$. Denote by φ the angle between vectors $\vec{u}$ and $\vec{v}$. Its cosine is given by formula (II.1.4). The angle φ can take on the values from 0 to π (sketch a picture). The *angle between straight lines p, q* is called the angle $\vartheta\in\left\langle 0;\frac{\pi}{2}\right\rangle$, which is equal to φ if $\varphi\in\left\langle 0;\frac{\pi}{2}\right\rangle$ and it is equal to $\pi-\varphi$ if $\varphi\in\left(\frac{\pi}{2};\pi\right\rangle$. (In other words, the angle between two lines is defined to be the acute or right angle between their directional vectors.) Using (II.1.4), we obtain the formula:

(II.2.6) $$\cos\vartheta = \frac{|\vec{u}\cdot\vec{v}|}{\|\vec{u}\|\cdot\|\vec{v}\|}.$$

II.2.10. Problems.

a) Find the angle between the straight lines AB and CD if $A=[-1;2;0]$, $B=[-2;0;2]$, $C=[4;-4;5]$, $D=[2;4;3]$.

b) Find the distance from the point $M=[1;-6;8]$ to the straight line AB, if $A=[2;1;3]$, $B=[3;-1;6]$.

c) Find the distance between the straight lines $p:\ X=[6;4;3]+t\cdot(1;1;1);\ t\in\mathbb{R}$ and $q:\ X=[7;0;-18]+s\cdot(2;-1;4);\ s\in\mathbb{R}$.

Results:

a) $\cos\vartheta=\frac{\sqrt{2}}{2}, \vartheta=\frac{\pi}{4}$ b) 4.36 c) 11.83

II.3. Planes in $\mathbb{E}_3$

II.3.1. A plane in $\mathbb{E}_3$.

If $A=[a_1;a_2;a_3]$ is a point in $\mathbb{E}_3$ and $\vec{u}=(u_1;u_2;u_3)$, $\vec{v}=(v_1;v_2;v_3)$ are two linearly independent vectors in $\mathbb{E}_3$, then the set $\sigma=\{X\in\mathbb{E}_3;\ \exists\,\alpha\in\mathbb{R},\ \exists\,\beta\in\mathbb{R}:\ X=A+\alpha\cdot\vec{u}+\beta\cdot\vec{v}\}$ is called the *plane* in $\mathbb{E}_3$. The equation

(II.3.1) $$X = A+\alpha\cdot\vec{u}+\beta\cdot\vec{v};\qquad \alpha,\ \beta\in\mathbb{R}$$

is called the *parametric equation* for plane σ. This equation is often used in the coordinate form:

$$\begin{aligned} x_1 &= a_1+\alpha\cdot u_1+\beta\cdot v_1\\ x_2 &= a_2+\alpha\cdot u_2+\beta\cdot v_2\\ x_3 &= a_3+\alpha\cdot u_3+\beta\cdot v_3\,; \end{aligned}\qquad \alpha,\ \beta\in\mathbb{R}.$$

II.3.2. A plane given by three points.

Let A, B, C be points in $\mathbb{E}_3$ which do not lie on a line. A plane passing through all these points (it will also be called plane ABC) can be described by the parametric equation

(II.3.2) $$X = A+\alpha\cdot(B-A)+\beta\cdot(C-A);\qquad \alpha,\ \beta\in\mathbb{R}.$$

II.3.3. Remark.

Each plane has infinitely many parametric equations. (Compare with lines – see remark II.2.5).

II.3.4. A normal vector.

Let σ be a plane, given by the parametric equation (II.3.1). Every non–zero vector which is perpendicular to plane σ (i.e. perpendicular to vectors $\vec{u}$, $\vec{v}$) is called the *normal vector* to plane σ.

It follows from theorem II.1.11 that for example the vector $\vec{n} = \vec{u} \times \vec{v}$ is the normal vector to plane σ. All other normal vectors to plane σ are non–zero multiples of $\vec{n}$, i.e. they have the form $c \cdot \vec{n}$, where $c \in \mathbb{R}$, $c \neq 0$.

II.3.5. Another analytic description of a plane.

If σ is a plane whose parametric equation is equation (II.3.1) and if its normal vector is $\vec{n} = (n_1; n_2; n_3)$, then any point $X = [x_1; x_2; x_3]$ in $\mathbb{E}_3$ belongs to plane σ if and only if $(X - A) \cdot \vec{n} = 0$. Expressing the scalar product on the left hand side, we get:

$$(x_1 - a_1) \cdot n_1 + (x_2 - a_2) \cdot n_2 + (x_3 - a_3) \cdot n_3 = 0,$$

$$n_1 x_1 + n_2 x_2 + n_3 + q = 0, \tag{II.3.3}$$

where $q = -a_1 n_1 - a_2 n_2 - a_3 n_3$. Equation (II.3.3) is called the *equation for plane* σ.

Equation (II.3.3) can be obtained from the parametric equation for plane σ (written in coordinates) by excluding parameters α and β. Conversely, if plane σ is given by equation (II.3.3) then its parametrization can be obtained for example in such a way that we find three different points A, B, C of σ (i.e. points whose coordinates satisfy equation (II.3.3)) which do not lie on a line, and we then use (II.3.2).

II.3.6. Example.

Let plane σ have the equation $5x_1 - 3x_2 + 7x_3 - 12 = 0$. Then the vector $\vec{n} = (5; -3; 7)$ is the normal vector to this plane.

II.3.7. The distance from a point to a plane.

Assume that plane σ is given by the parametric equation (II.3.1) and M is a point in $\mathbb{E}_3$. We derive a formula for the distance $d(M, \sigma)$.

From the right–angled triangle APM, we obtain: $d(M, \sigma) = \|M - P\| = \|M - A\| \cdot \cos\varphi$. $\cos\varphi$ can be expressed by means of the scalar product of the vectors $\vec{n}$ and $(M - A)$:

$$\cos\varphi = \frac{\|\vec{n} \cdot (M - A)\|}{\|\vec{n}\| \cdot \|M - A\|}$$

($\vec{n}$ is an arbitrary normal vector to plane σ.) Using this in the formula for $d(M, \sigma)$, we get

$$d(M, \sigma) = \frac{|\vec{n} \cdot (M - A)|}{\|\vec{n}\|}. \tag{II.3.4}$$

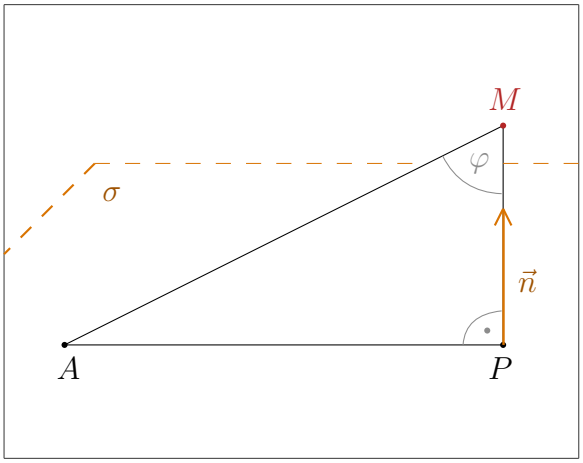


Fig. 2

Suppose now that $M = [m_1; m_2; m_3]$. Then $\vec{n} \cdot (M - A) = n_1 m_1 + n_2 m_2 + n_3 m_3 + q$ (where $q = -n_1 a_1 - n_2 a_2 - n_3 a_3$). Substituting this to (II.3.4) and expressing $|\vec{n}|$ as $\sqrt{n_1^2 + n_2^2 + n_3^2}$, we obtain:

$$d(M, \sigma) = \frac{|n_1 m_1 + n_2 m_2 + n_3 m_3 + q|}{\sqrt{n_1^2 + n_2^2 + n_3^2}}. \tag{II.3.5}$$

This formula is useful especially if plane σ is given by equation (II.3.3).

II.3.8. Remark.

Assume that p is a straight line, given by parametric equations (II.2.2). Excluding parameter t from the three equations (II.2.2), we obtain a system of two equations which can be written in the form

$$n'_1 x_1 + n'_2 x_2 + n'_3 x_3 + q' = 0, \qquad n''_1 x_1 + n''_2 x_2 + n''_3 x_3 + q'' = 0. \tag{II.3.6}$$

Straight lines are often described by such systems of two equations. Each of them is an equation for a plane, and the straight line is the intersection of the two planes.

Conversely, if a straight line is given by two equations of type (II.3.6) then its parametrization can be obtained for example as follows: We solve equations (II.3.6) as two equations for unknowns x_1, x_2, x_3. If the planes described by the equations in (II.3.6) are not parallel (or even identical) then the general solution contains one arbitrary parameter (see remark I.3.8). Writing down the general solution of (II.3.6), we get the parametric equation for the line.

II.3.9. Example.

Find a parametrization of the line given by two equations

$$5x_1 + 7x_2 - 4x_3 + 1 = 0, \qquad 2x_1 + 4x_2 - 4x_3 - 2 = 0.$$

Applying for example Gauss' method of elimination, we express the solution of this system: $x_1 = -2t - 3$, $x_2 = 2t + 2$, $x_3 = t$ $(t \in \mathbb{R})$. These three equations represent parametric equations for the line. Obviously, the parametric representation can also be written in a vector form, as one equation: $X = A + t \cdot \vec{u}$, where $A = [-3; 2; 0]$ and $\vec{u} = (-2; 2; 1)$.

II.3.10. A position of a straight line according to a plane.

Let us deal with a straight line p and a plane σ in space $\mathbb{E}_3$. p and σ can find themselves in three different positions:

a) Line p is a subset of plane σ.
b) Line p intersects plane σ at one point.
c) Line p and plane σ are disjoint, they have no common points. In this case, we say that line p is *parallel* to plane σ.

How to distinguish between the possibilities a), b), c) in a particular case? Let us analyze for example the situation when line p is given by two equations (II.3.6) and plane σ is described by one equation (II.3.3). Equations (II.3.6) and (II.3.3) together form a system of three linear equations for three unknowns x_1, x_2, x_3. We solve this system. There are three possibilities (see Frobenius' theorem): 1) The system has infinitely many solutions. 2) The system has a unique solution. 3) The system has no solution. These possibilities successively correspond to the cases a), b) and c) mentioned above. In case 2) the solution of the system gives coordinates of the point of intersection of line p with plane σ.

We will examine the position of a line according to a plane once again in example II.3.12, the line will be given parametrically this time.

II.3.11. The angle between a straight line and a plane.

Let p be a straight line and σ be a plane in space $\mathbb{E}_3$. If line p is not perpendicular to σ then the *angle between line p and plane σ* is defined to be equal to the angle between line p and its orthogonal projection to plane σ. If line p is perpendicular to plane σ then the angle between them is defined to be equal to $\frac{\pi}{2}$.

How can the angle between line p and plane σ be evaluated in a particular case? First we determine a normal vector $\vec{n}$ to plane σ (see paragraph II.3.4 or possibly also II.3.5, II.3.6). Let q be an arbitrary line parallel to vector $\vec{n}$ (see Fig. 3). Denote by ϑ the angle between lines p and q. $\cos\vartheta$ can be calculated from (II.2.6). The angle between line p and plane σ is equal to $\frac{\pi}{2} - \vartheta$.

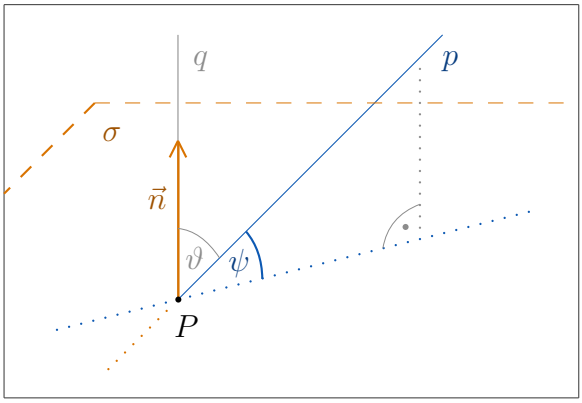


Fig. 3

II.3.12. Example.

Line p is given parametrically: $x_1 = 1 + t$, $x_2 = 2 + t$, $x_3 = 3$. Plane σ is given by the equation $2x_1 + 4x_2 + 4x_3 + 2 = 0$. What is the position of p according to σ and what is the angle between p and σ?

We substitute for x_1, x_2, x_3 from the parametric equations for p to the equation for $sigma$. We obtain the equation for the unknown t: $2 + 2t + 8 + 4t + 12 + 2 = 0$. This equation has a unique solution $t = -4$. As the solution exists and is unique, line p and plane σ have a unique common point. The coordinates of this point can be obtained by the substitution of $t = -4$ to the parametric equations for line p: $x_1 = -3$, $x_2 = -1$, $x_3 = 3$.

For example, the vector $\vec{n} = (2; 4; 4)$ is the normal vector to plane σ. A directional vector of line p can be chosen for example to be equal to the vector $\vec{a} = (1; 1; 0)$. The angle between line p and a line with the directional vector $\vec{n}$ is the angle ϑ whose cosine can be calculated by formula (II.2.6). It is equal to $\frac{\sqrt{2}}{2}$. Since ϑ is the acute angle, the information on its cosine implies: $\vartheta = \frac{\pi}{4}$. Hence the angle between line p and plane σ is the angle $\psi = \frac{\pi}{2} - \vartheta = \frac{\pi}{2} - \frac{\pi}{4} = \frac{\pi}{4}$.

II.3.13. A position of two planes in space.

Two planes in space $\mathbb{E}_3$ can either be *identical* or *intersecting in a line* or *parallel*. If two particular planes are given and one wishes to recognize which of these cases obtains, then the approach one can use depends on the way the planes are given. We show two possible methods in examples II.3.15 and II.3.16.

II.3.14. The angle between two planes.

The *angle between two planes σ and η* is defined to be equal to the angle between two arbitrary lines, the first of which is perpendicular to plane σ, and the second being perpendicular to plane η.

II.3.15. Example.

Analyze the position of plane σ according to plane η and evaluate the angle between them. σ: $x_1 - x_2 + x_3 - 1 = 0$ and η: $2x_1 + x_2 - x_3 - 1 = 0$.

Coordinates of common points of the two planes must satisfy both equations. The equations represent the system of two linear equations for the unknowns x_1, x_2, x_3. We can easily check that the system has the solution $x_1 = \frac{2}{3}$, $x_2 = -\frac{1}{3} + t$, $x_3 = t$; $t \in \mathbb{R}$. These equations for x_1, x_2, x_3 can be regarded as parametric equations of the line which is the intersection of planes σ and η.

The vector $\vec{n}' = (1; -1; 1)$ is the normal vector to plane σ and $\vec{n}'' = (2; 1; -1)$ is the normal vector to plane η. Applying formula (II.2.6), we can now find out that the angle between planes σ and η is equal to $\frac{\pi}{2}$.

II.3.16. Problems.

1) Find the distance from point M to plane σ, if
   a) $M = [7; 0; -1]$, σ: $3x_1 + x_2 - 2x_3 + 5 = 0$,
   b) $M = [2; 3; -1]$, σ: $X = [1; 0; -1] + \alpha \cdot (2; 0; 1) + \beta \cdot (0; 2; 1)$; $\alpha, \beta \in \mathbb{R}$.
2) What is the position of plane σ in relation to plane η and what is the angle between σ and η if
   a) σ: $3x_1 - 6x_2 + 6x_3 + 9 = 0$ and η: $x_1 - 2x_2 + 2x_3 - 3 = 0$,
   b) σ: $2x_1 - x_2 + x_3 - 1 = 0$ and η: $x_1 + x_2 + 2x_3 + 1 = 0$?
3) Find the angle between the line AB (where $A = [2; -1; 2]$ and $B = [1; -1; 1]$) and the plane σ: $x_1 - x_2 - 5 = 0$.
4) Find coordinates of an orthogonal projection of the point $A = [7; 0; -1]$ onto the plane σ: $3x_1 + x_2 - 2x_3 + 5 = 0$.
5) In which position is line p in relation to plane σ, if p : $5x_1 + 7x_2 - 4x_3 + 15 = 0$, $12x_1 - 2x_2 + 8x_3 + 3 = 0$ and σ : $5x_1 + 7x_2 - 4x_3 - 8 = 0$?
6) In which position is the plane ABC in relation to the plane DEF, if $A = [1; 1; 2]$, $B = [1; 1; 4]$, $C = [1; 2; 1]$, $D = [2; 0; -1]$, $E = [2; 1; 1]$, $F = [4; -1; 3]$? Also find the angle between the planes ABC and DEF.

Results: 1a) $2\sqrt{14}$ 1b) $\frac{4}{\sqrt{30}}$ 2a) The planes are parallel and their distance is 2. 2b) The planes intersect in the straight line $X = [0; -1; 0] + t\,(-1; -1; 1)$, the angle between them is $\frac{\pi}{3} = 60°$. 3) $\frac{\pi}{6} = 30°$ 4) $[1; -2; 3]$ 5) Straight line p and plane σ are parallel, their distance is $\frac{23}{\sqrt{90}}$. 6) The planes intersect in the straight line $X = [1; 2; 0] + t(0; 1; 2)$, the angle between them is $36.67°$.

II.4.* Quadric surfaces in $\mathbb{E}_3$

II.4.1. Quadric surfaces in $\mathbb{E}_3$.

From secondary school, you are familiar with conic sections in $\mathbb{E}_2$ and with their equations. These equations are quadratic. By analogy, quadratic equations describe so called *quadric surfaces* in $\mathbb{E}_3$. A general equation for a quadric surface is

$$a_{11}\,x_1^2 + a_{12}\,x_1\,x_2 + a_{13}\,x_1\,x_3 + a_{22}\,x_2^2 + a_{23}\,x_2\,x_3 + a_{33}\,x_3^2 + \\ + b_1\,x_1 + b_2\,x_2 + b_3\,x_3 + c = 0. \tag{II.4.1}$$

(The statement "quadric surface σ is described by equation (II.4.1)" or "(II.4.1) is the equation for quadric surface σ" means that quadric surface σ is the set of all points in $\mathbb{E}_3$ whose coordinates satisfy equation (II.4.1).)

It can be shown that there exists a Cartesian coordinate system x_1', x_2', x_3' in $\mathbb{E}_3$ which is only turned on in relation to the coordinate system x_1, x_2, x_3 and the quadric surface given by (II.4.1) can be described in the new system by an equation which does not contain the mixed products $x_1'\,x_2'$, $x_1'\,x_3'$, $x_2'\,x_3'$.

To avoid complications, we shall further deal only with quadric surfaces whose equations even in the coordinate system x_1, x_2, x_3 do not contain the mixed products $x_1\,x_2$, $x_1\,x_3$, $x_2\,x_3$. This approach corresponds to situations you have been used to from analytic geometry in plane from secondary school. For example, you have mostly described an ellipse in the x, y–plane by an equation which contained the terms x^2 and y^2, but it did not contain the mixed product xy. This means that you have (mostly) restricted yourself to ellipses whose axes are parallel to the x and y axes.

We present only a survey of the names of quadric surfaces which correspond to the most important special cases of equation (II.4.1) in the following. Instead of x_1, x_2, x_3, we use the denotation x, y, z in the rest of this chapter.

II.4.2. Circular quadric surfaces (quadric surfaces of revolution).

A surface which arises by the rotation of a conic section about its axis is called a *circular quadric surface* or a *quadric surface of revolution*. Particularly, the rotation of

- an ellipse about its axis yields ..... *a circular ellipsoid,*
- a parabola about its axis yields ..... *a circular paraboloid,*
- a hyperbola about its non–focal axis yields *a one sheet–circular hyperboloid,*
- a hyperbola about its focal axis yields ..... *a two sheet–circular hyperboloid,*
- two skew lines about the axis of symmetry (lying in their plane) yields ..... *a circular conic surface,*
- two parallel lines about the axis (lying in the middle between them) yields ... *a circular cylindrical surface.*

The first four surfaces are the so called *regular quadric surfaces*, the last two are the so called *singular quadric surfaces*.

II.4.3. How to recognize a circular quadric surface.

If the equation of a quadric surface depends on x and y only through x^2+y^2 (i.e. x and y appear in the equation only as a part of the expression x^2+y^2) then the quadric surface is circular and its axis of revolution is the z–axis.

This holds due to the following reason: $x^2 + y^2$ is the second power of the distance of the point $[x; y; z]$ from the z–axis. Thus, x and y appear in the equation only through the distance of $[x; y; z]$ from the z–axis. The distance is the same on each circle which lies in a plane perpendicular to the z–axis and such that the z–axis passes through its center. This means that either all points of the circle satisfy the equation (and all points of the circle belong to the quadric surface) or no point of the circle satisfy the equation (and therefore no point of the circle belongs to the quadric surface).

Analogous assertions can also be formulated in the cases when the equation of some quadric surface depends on x, z only through $x^2 + z^2$ or if it depends on y, z only through $y^2 + z^2$.

II.4.4. Quadric surfaces in the basic and in a shifted position.

Conic sections in a plane can be in a *basic position* (for example the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$), or in a *shifted position* (as the ellipse $\frac{(x-\alpha)^2}{a^2} + \frac{(y-\beta)^2}{b^2} = 1$). By analogy, quadric surfaces in $\mathbb{E}_3$ can also be in a basic or in a shifted position. We restrict ourselves to quadric surfaces in the basic position in this text. Equations for the same surfaces in the shifted position (shifted by the vector $(\alpha; \beta; \gamma)$) can be obtained from the equations for surfaces in the basic position by replacing x^2 by $(x-\alpha)^2$, y^2 by $(y-\beta)^2$ and z^2 by $(z-\gamma)^2$.

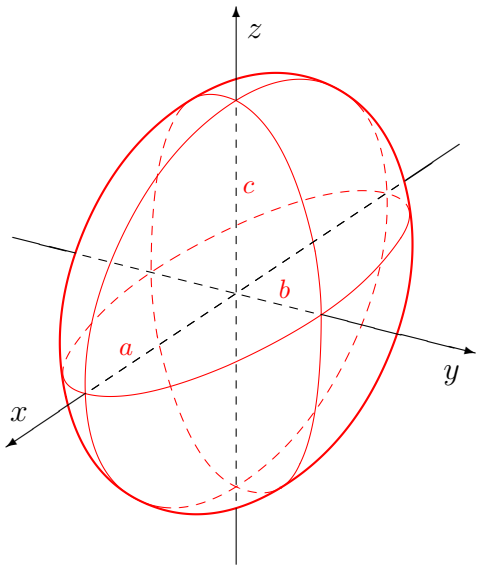


Fig. 4

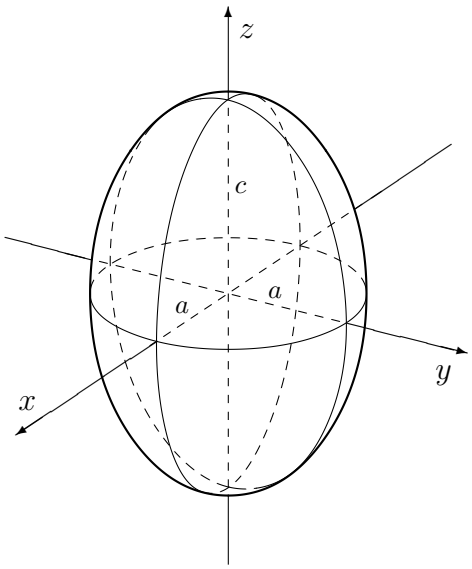


Fig. 5

II.4.5. Ellipsoid.

Let a, b, c be positive numbers. The equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{II.4.2}$$

defines the so called *ellipsoid* with semi–axes a, b, c. The center of the ellipsoid is the point $[0; 0; 0]$. (See Fig. 4.)

If any two of the semi–axes are equal, the ellipsoid is circular (= the ellipsoid of revolution). (See Fig. 5.) If all three semi–axes are equal, the ellipsoid coincides with a sphere.

II.4.6. One sheet – hyperboloid.

Let a, b, c be positive numbers. The equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1 \tag{II.4.3}$$

is the equation for the so called *one sheet – hyperboloid* (= a *hyperboloid of one sheet*). (See Fig. 6.)

If $a = b$ then it is a circular one sheet – hyperboloid (= a one sheet – hyperboloid of revolution) which arises e.g. by revolution of the hyperbola $\frac{x^2}{a^2} - \frac{z^2}{c^2} = 1$ in the x, z–plane about the z–axis.

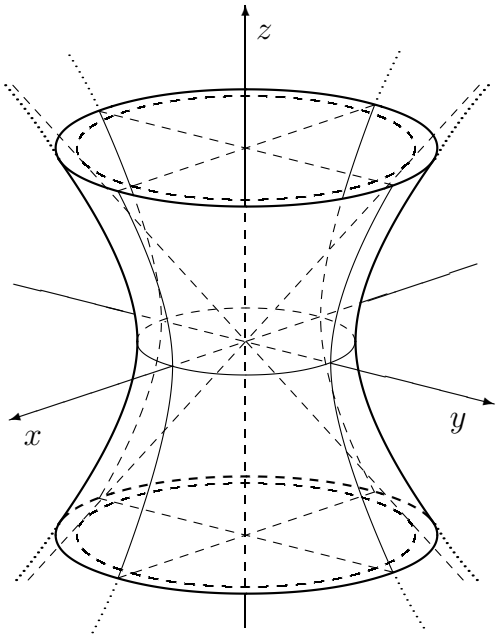


Fig. 6

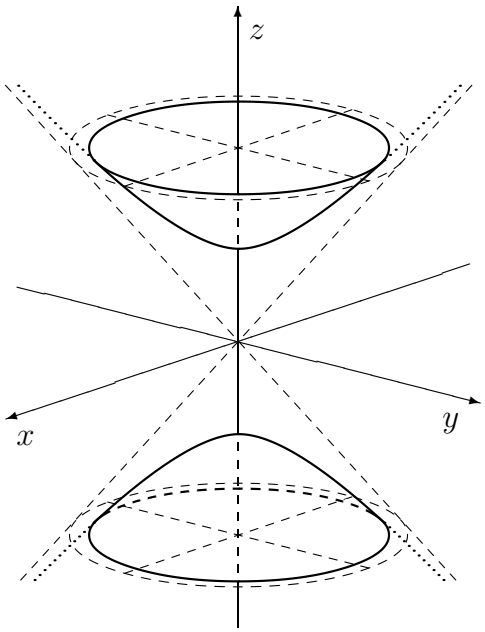


Fig. 7

II.4.7. Two sheet – hyperboloid.

Let a, b, c be positive numbers. The equation

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1 \tag{II.4.4}$$

defines the so called *two sheet – hyperboloid* (= a *hyperboloid of two sheets*). (See Fig. 7.)

If $b = c$ then the hyperboloid is circular and its axis of revolution is the x–axis. It arises e.g. by revolution of the hyperbola $\frac{x^2}{a^2} - \frac{z^2}{c^2} = 1$ in the x, z–plane about the z–axis.

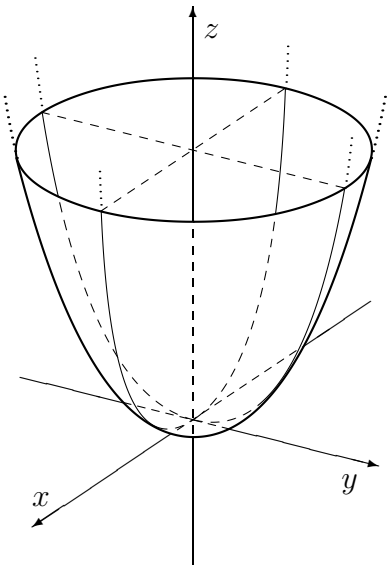


Fig. 8

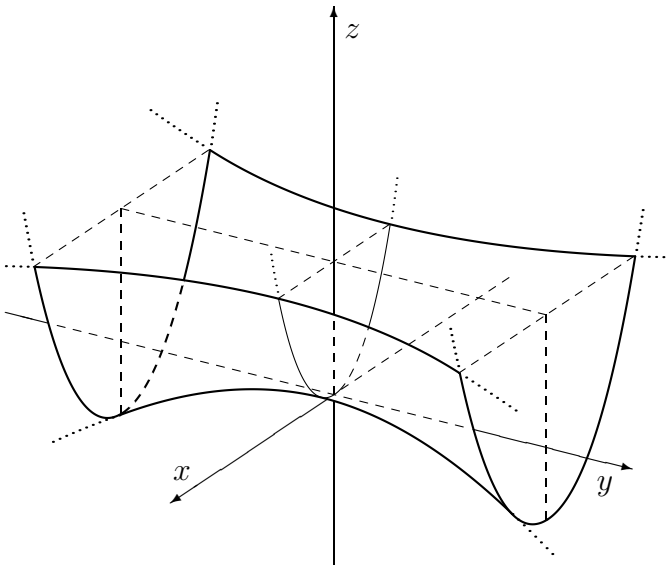


Fig. 9

II.4.8. Paraboloid.

Let a, b be positive numbers. The equation

$$z = \frac{x^2}{a^2} + \frac{y^2}{b^2} \tag{II.4.5}$$

is the equation for the so called *elliptic paraboloid*. Its axis is the z–axis. (See Fig. 8.)

If $a = b$ then the paraboloid is circular (in other words: it is a paraboloid of revolution). It arises e.g. by revolution of the parabola $z = \frac{x^2}{a^2}$ in the x, z–plane about the z–axis.

The equation

$$z = \frac{x^2}{a^2} - \frac{y^2}{b^2} \tag{II.4.6}$$

defines the so called *hyperbolic paraboloid*. (See Fig. 9.) This paraboloid intersects every plane of the type $z = c$ (where $c > 0$) in a hyperbola.

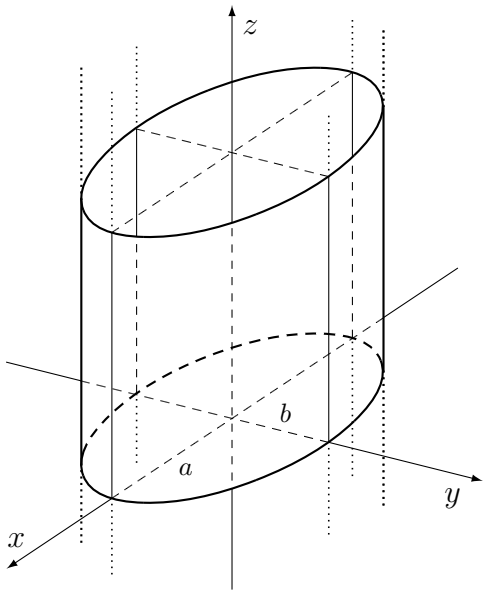


Fig. 10

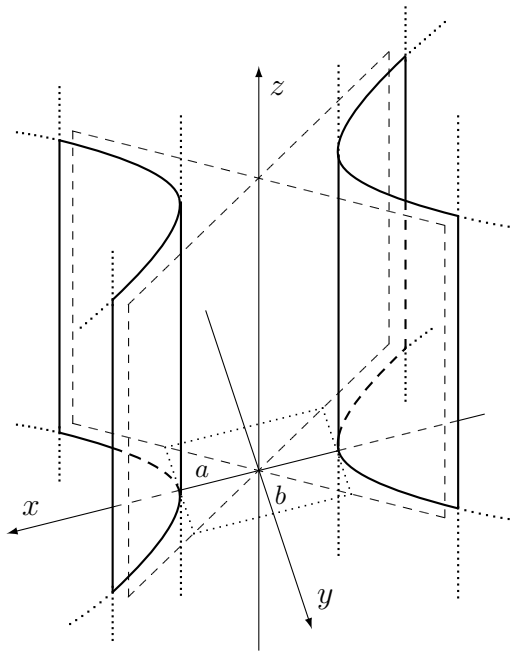


Fig. 11

II.4.9. Cylindrical surface, cylinder.

Let a, b be positive numbers. The equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{II.4.7}$$

defines the so called *elliptic cylindrical surface* (= an *elliptic cylinder*). (See Fig. 10.)

If $a = b$ then it is a circular cylindrical surface (= a cylindrical surface of revolution). This arises e.g. by rotation of the parallel lines $x = a, y = 0$ and $x = -a, y = 0$ about the z–axis.

The equation

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \tag{II.4.8}$$

defines the so called *hyperbolic cylindrical surface* (= a *hyperbolic cylinder*). (See Fig. 11.)

The equation

$$y = k\,x^2 \tag{II.4.9}$$

defines a *parabolic cylindrical surface* (= a *parabolic cylinder*). (See Fig. 12.)

II.4.10. Conic surface, cone.

Let a, b, c be positive numbers. The equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0 \tag{II.4.8}$$

describes the so called *elliptic conic surface* (= an *elliptic cone*).

If $b = c$ then the cone is circular. This arises for example by rotation of the straight lines $\frac{x}{a} = \frac{z}{c}, y = 0$ and $\frac{x}{a} = \frac{z}{c}, y = 0$ about the z–axis.

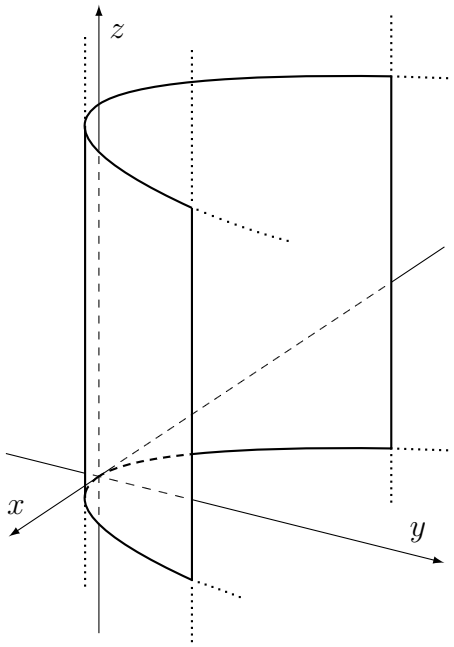


Fig. 12

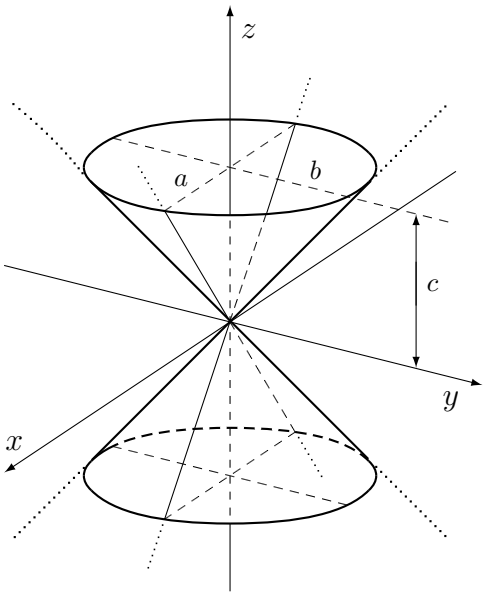


Fig. 13

You will again meet with quadric surfaces in the Mathematics II course, especially in the chapter on surface integrals.

II.4.11. Problems.

The given equations define various quadric surfaces. Recognize their types, give them appropriate names and specify their axes, possibly also semi–axes.

a) $x^2 + \frac{y^2}{4} + 9z^2 = 1$

b) $25\,(x-1)^2 + y^2 + z^2 = 25$

c) $\frac{x^2}{4} + \frac{y^2}{4} - \frac{z^2}{9} = 1$

d) $x^2 - \frac{(y-2)^2}{4} - \frac{(z-4)^2}{4} = 1$

e) $z = x^2 + 4y^2$

f) $y = 1 - 5x^2 - 5(z-2)^2$

g) $y^2 - x^2 = z$

h) $(x+2)^2 - 5z^2 = y$

i) $x^2 + z^2 = 4$

j) $\frac{x^2}{4} + (y+4)^2 = 1$

k) $4x^2 + 9z^2 = 9y^2$

l) $(y-1)^2 + z^2 = x^2$

m) $x^2 - 4z^2 = 1$

n) $\frac{(y+2)^2}{4} - \frac{x^2}{16} = 1$

o) $4x^2 - 9z^2 = 9y^2$

p) $y = x^2 + 2$

q) $x^2 + 4y^2 = 1$

r) $\frac{(x-4)^2}{4} + \frac{y^2}{16} + z^2 = 1$

R e s u l t s :

a) Ellipsoid with center $[0;0]$ and semi–axes 1, 2, $\frac{1}{3}$.

b) Circular ellipsoid with center at the point $[1;0;0]$ and semi–axes 1, 5, 5. Axis of revolution is the x–axis.

c) Circular one–sheet hyperboloid with center at the origin. Axis of revolution is the z–axis.
d) Circular two–sheet hyperboloid with center $[0; 2; 4]$. Axis of revolution is the straight line $y = 2$, $z = 4$, parallel with the x–axis.
e) Elliptic paraboloid with center at the origin and axis z, open in the positive direction of the z–axis.
f) Circular paraboloid with vertex $[0; 1; 2]$, open in the negative direction of the y–axis. Axis of revolution is the straight line $x = 0$, $z = 2$, parallel with the y–axis.
g) Hyperbolic paraboloid with vertex at the origin and axis z.
h) Hyperbolic paraboloid with vertex $[-2; 0; 0]$ and axis y.
i) Circular cylindrical surface. Axis of revolution is the y–axis and radius equals 2.
j) Elliptic cylindrical surface. Its axis is the straight line $x = 0$, $y = 4$, parallel with the z–axis.
k) Elliptic conic surface with vertex at the origin and axis y.
l) Circular conic surface with vertex $[0; 1; 0]$. Axis of revolution is the straight line $y = 1$, $z = 0$, parallel with the x–axis.
m) Hyperbolic cylindrical surface, parallel with the y–axis, symmetric according to this axis.
n) Hyperbolic cylindrical surface, parallel with the z–axis, symmetric according to the straight line $x = 0$, $y = -2$.
o) Circular conic surface with vertex at the origin. Axis of revolution is the x–axis.
p) Parabolic cylindrical surface, parallel with the z–axis. It meets the x, y–plane in the parabola with vertex $[0; 2]$ and axis y. (Its equation in the $x, , y$–plane is identical with the equation of the whole cylindrical surface: $y = x^2 + 2$.)
q) Elliptic cylindrical surface. Its axis is the z–axis.
r) Ellipsoid with center at the point $[4; 0; 0]$ and semi–axes 2, 4, 1.

CHAPTER III.

Differential calculus

The extended set of real numbers. The *extended set of real numbers* is the union of $\mathbb{R}$ with the two point set that contains the elements called *plus infinity*, *minus infinity* and denoted $+\infty$ and $-\infty$. The extended set of real numbers will be denoted by $\mathbb{R}^*$. Its elements $-\infty$ and $+\infty$ will be called the *improper points*, other elements (i.e. numbers from $\mathbb{R}$) will be called the *proper points*. The operations addition, subtraction, multiplication, division and raising to a power, which are well known in $\mathbb{R}$, can be extended in a natural way to $\mathbb{R}^*$:

a) if $x \in \mathbb{R}$ then we put
$x + (+\infty) = +\infty, \quad x + (-\infty) = -\infty, \quad x - (+\infty) = -\infty,$
$x - (-\infty) = +\infty, \quad \frac{x}{+\infty} = \frac{x}{-\infty} = 0;$

b) $(+\infty) + (+\infty) = +\infty, \quad (-\infty) + (-\infty) = -\infty, \quad (+\infty) - (-\infty) = +\infty,$
$(-\infty) - (+\infty) = -\infty,$
$(+\infty) \cdot (+\infty) = +\infty, \quad (+\infty) \cdot (-\infty) = -\infty, \quad (-\infty) \cdot (-\infty) = +\infty;$

c) if $x \in \mathbb{R}\setminus\{0\}$ then we define
$x \cdot (+\infty) = +\infty$ (if $x > 0$) or $x \cdot (+\infty) = -\infty$ (if $x < 0$),
$x \cdot (-\infty) = -\infty$ (if $x > 0$) or $x \cdot (-\infty) = +\infty$ (if $(x < 0$),
$\frac{+\infty}{x} = \operatorname{sgn} x \cdot (+\infty), \quad \frac{-\infty}{x} = \operatorname{sgn} x \cdot (-\infty).$

The operations division by zero, $(+\infty)-(+\infty)$, $(-\infty)-(-\infty)$, $(+\infty)+(-\infty)$, $\frac{\pm\infty}{\pm\infty}$ and $0 \cdot (\pm\infty)$ remain undefined. (We say that they have no sense.)

The set $\mathbb{R}$ is ordered by the relation "$<$" ("less than"). This relation can naturally be extended to $\mathbb{R}^*$: For any $x \in \mathbb{R}$ we define: $-\infty < x$ and $x < +\infty$.

Extreme values of sets in $\mathbb{R}$. If M is a subset of $\mathbb{R}$ then the *maximum* of M is a number $y \in M$ such that $\forall\, x \in M : \ x \le y$. The maximum of the set M is denoted by $\max M$.

By analogy, the *minimum* of set M is a number $z \in M$ such that $\forall\, x \in M : \ x \ge z$. The minimum of the set M is denoted by $\min M$.

It can easily be observed that not every set M in $\mathbb{R}$ must have a maximum and a minimum. (See for example $M = (0; 1)$.)

A generalization of the notion of maximum of set M is a so called supremum of set M. Number $K \in \mathbb{R}^*$ is said to be the *supremum* of set M if

a) $\forall\, x \in M \ : \ x \le K$,
b) K is the least of all numbers with property a).

Supremum of set M is denoted by $\sup M$. Each number K which satisfies condition a) is a so called "upper bound" of set M. This is the reason for another often used name and denotation of the supremum: the *least upper bound*, $\mathrm{l.u.b.}\, M$.

By analogy, we can define the *infimum* of set M. It is denoted by $\inf M$ and it is the greatest of all numbers $L \in \mathbb{R}^*$ such that $\forall\, x \in M \ : \ x \ge L$. The infimum is also often called the *greatest lower bound* and denoted by $\mathrm{g.l.b.}\, M$.

On the contrary to the maximum and minimum which need not exist, it can be proved (not very simply) that every set in $\mathbb{R}$ has a supremum and an infimum. Verify for yourselves that if $max\, M$ exists then $\sup M = \max M$. Similarly, if $\min M$ exists then $\inf M = \min M$.

Neighborhoods of points in $\mathbb{R}^*$. If $x \in \mathbb{R}$, then a *neighborhood* of point x is any interval $(x-\varepsilon; x+\varepsilon)$ where $\varepsilon > 0$. This neighborhood is denoted by $\mathcal{U}_\varepsilon(x)$ or simply $\mathcal{U}(x)$. (The notation is derived from the German name for the neighborhood: die Umgebung.)
A *punctured neighborhood* of the point $x \in \mathbb{R}$, also called the *deleted neighborhood* or *reduced neighborhood*, is every set of the type $\mathcal{U}(x) \setminus \{x\}$. This neighborhood will be denoted by $\mathcal{P}(x)$.

A *neighborhood of* $+\infty$ ($\equiv$ the reduced neighborhood of $+\infty$) is any interval $(a; +\infty)$ (where $a \in \mathbb{R}$). A *neighborhood of* $-\infty$ ($\equiv$ the reduced neighborhood of $-\infty$) is any interval $(-\infty; a)$ (where $a \in \mathbb{R}$). As we do not distinguish between the neighborhood and the reduced neighborhood of $+\infty$, we can denote it either $\mathcal{U}(+\infty)$ or $\mathcal{P}(+\infty)$. Similarly, the neighborhood of $-\infty$ can be denoted by $\mathcal{U}(-\infty)$ or by $\mathcal{P}(-\infty)$.

A *left neighborhood* of point $x \in \mathbb{R}$ is any interval of the type $(x-\varepsilon; x)$, where $\varepsilon > 0$. Similarly, we can define a *right neighborhood* of point $x \in \mathbb{R}$ to be any interval of the type $(x; x+\varepsilon)$ where $\varepsilon > 0$. The left (respectively the right) neighborhood of the point x is denoted by $\mathcal{P}_-(x)$ (respectively $\mathcal{P}_+(x)$).

III.1. Sequences of real numbers

III.1.1. A sequence of real numbers.

A *sequence of real numbers* (shortly a *sequence*) is a mapping of the set of natural numbers $\mathbb{N}$ to the set of real numbers $\mathbb{R}$. A sequence which to every $n \in \mathbb{N}$ assigns the number a_n, is denoted by $\{a_1; a_2; a_3; \dots\}$ or shortly $\{a_n\}$. The number a_n is called the *n–th term* of the sequence $\{a_n\}$. If $M \subset \mathbb{R}$ and $a_n \in M$ for all $n \in \mathbb{N}$ then $\{a_n\}$ is called the *sequence in M*.

III.1.2. Bounded, monotonic and strictly monotonic sequences.

The sequence $\{a_n\}$ is called

a) *bounded above* if there exists $K \in \mathbb{R}$ so that $\forall\, n \in \mathbb{N}:\ a_n \le K$;
b) *bounded below* if there exists $K \in \mathbb{R}$ so that $\forall\, n \in \mathbb{N}:\ a_n \ge K$;
c) *bounded* if it is bounded above and bounded below;
d) *increasing* if $\forall\, n \in \mathbb{N}:\ a_n < a_{n+1}$;
e) *decreasing* if $\forall\, n \in \mathbb{N}:\ a_n > a_{n+1}$;
f) *non–decreasing* if $\forall\, n \in \mathbb{N}:\ a_n \le a_{n+1}$;
g) *non–increasing* if $\forall\, n \in \mathbb{N}:\ a_n \ge a_{n+1}$;
h) *monotonic* if it is non–increasing or non–decreasing;
i) *strictly monotonic* if it is increasing or decreasing.

III.1.3. Remark.

Notice that: An increasing sequence is a special case of a non–decreasing sequence. A decreasing sequence is a special case of a non–increasing sequence. A strictly monotonic sequence is a special case of a monotonic sequence.

Instead of "monotonic", you can also say "monotone".

III.1.4. Limit of a sequence.

The number $a \in \mathbb{R}^*$ is called the *limit* of the sequence $\{a_n\}$ if

(III.1.1) $\quad [\forall\, \mathcal{U}(a)\,] \quad [\exists\, n_0 \in \mathbb{N}] \quad [\forall\, n \in \mathbb{N}\,] : \quad (n \geq n_0) \implies (a_n \in \mathcal{U}(a)\,).$

(We read it: For every neighborhood $\mathcal{U}(a)$ of the point a there exists $n_0 \in \mathbb{N}$ so that for all $n \in \mathbb{N}$ it holds: If $n \geq n_0$, then $a_n \in \mathcal{U}(a).$) The fact that a is the limit of the sequence $\{a_n\}$ is written down in this way: $\lim\limits_{n\to+\infty} a_n = a$ or shortly $a_n \to a$.

III.1.5. Remark.

A limit of the sequence $\{a_n\}$ is a number a such that the elements a_n tend to a (approach a) if the indices n tend to infinity (n approaches infinity). You may have an impression that the statement (III.1.1) is an unnecessarily complicated description of this situation which intuitively seems to be quite clear. However, the reason is that mathematics does not have another, simpler, but also precise expression of the notions "to tend", "to approach". In case you do not understand the statement (III.1.1) immediately after reading it for the first time, don't worry. It may take some time and some thought to understand it well. An attentive study of the proof of Theorem III.1.8 may help. The notion of the limit of a sequence (or of a function – which will be introduced later) is one of the basic notions of mathematical analysis. Its importance is given mainly by the fact that it describes a certain infinite process (of approaching a number). It first appeared in the 16th–17th century. It brought dynamics to mathematical thinking, which had remained strongly under the influence of ancient mathematics, and it also led to ways of dealing with a "mysterious" infinity.

Not every sequence must have a limit! You can see an example of a sequence which has no limit in paragraph III.1.12.

If a sequence has a limit then the limit (its existence as well as its value) is not dependent on the behavior of any initial part of the sequence (containing for example the first one million terms). On the contrary – if the sequence $\{a_n\}$ has the limit a and we modify the sequence so that we change arbitrarily the values of its first one million terms, then the modified sequence will also have a limit equal to a.

III.1.6. A convergent and a divergent sequence.

If a sequence $\{a_n\}$ has a limit from $\mathbb{R}$ (i.e. is not equal to $-\infty$ or $+\infty$), then we say that this sequence is *convergent*. A sequence which either has no limit or has an infinite limit is called *divergent*.

III.1.7. Remark.

If $\lim\limits_{n\to+\infty} a_n = a$ and $a \in \mathbb{R}$ then we also say that *the sequence* $\{a_n\}$ *converges to the number* a.

III.1.8. Theorem.

Every sequence has at most one limit.

P r o o f :
We show a so called "proof by contradiction". We assume that the assertion of the theorem does not hold and by means of further considerations, we derive a contradiction with this assumption.

Thus, the assumption will be shown to be false and so the theorem will be proved.

If the assertion of the theorem does not hold then there exists at least one sequence which has more that one limit. Let us denote this sequence by $\{a_n\}$ and let a' and a'' be its different limits. There exist neighborhoods $\mathcal{U}(a')$ and $\mathcal{U}(a'')$ which are disjoint (i.e. their intersection is an empty set). According to (III.1.1), to $\mathcal{U}(a')$ there exists $n_1 \in \mathbb{N}$ so that the following implication holds for all $n \in \mathbb{N}$: $n \geq n_1 \implies a_n \in \mathcal{U}(a')$. (III.1.1) also implies that to the neighborhood $\mathcal{U}(a'')$ there exists $n_2 \in \mathbb{N}$ such that the implication $n \geq n_2 \implies a_n \in \mathcal{U}(a'')$ holds for all $n \in \mathbb{N}$. This means that if n is so large that it is $\geq n_1$ and $\geq n_2$ then a_n must belong to $\mathcal{U}(a')$ and also to $\mathcal{U}(a'')$. However, this is impossible since $\mathcal{U}(a')$ and $\mathcal{U}(a'')$ have no common points. This is the desired contradiction.

III.1.9. A subsequence.

Let $\{k_n\}$ be an increasing sequence of real numbers. Then the sequence

$$\{a_{k_1};\ a_{k_2};\ a_{k_3};\ \ldots\ ;\ a_{k_n};\ \ldots\}$$

is called the *subsequence* of the sequence $\{a_n\}$. The subsequence is shortly denoted by $\{a_{k_n}\}$.

III.1.10. Example.

The sequence $\left\{\frac{1}{2n}\right\}$ (i.e. the sequence $\left\{\frac{1}{2};\ \frac{1}{4};\ \frac{1}{6};\ \frac{1}{8};\ \ldots\right\}$) is the subsequence of the sequence $\left\{\frac{1}{n}\right\}$ (i.e. of the sequence $\left\{1;\ \frac{1}{2};\ \frac{1}{3};\ \frac{1}{4};\ \frac{1}{5};\ \ldots\right\}$).

III.1.11. Theorem.

If the sequence $\{a_n\}$ has a limit equal to a then its every subsequence also has the same limit with the same value a.

P r o o f :
Suppose that $\lim\limits_{n\to+\infty} a_n = a$, i.e. that (III.1.1) holds. We want to show that $\lim\limits_{n\to+\infty} a_{k_n} = a$, i.e. that the following statement holds:

$$[\forall\, \mathcal{U}(a)]\quad [\exists\, n_1 \in \mathbb{N}]\quad [\forall\, m \in \mathbb{N}]\ :\quad (n \geq n_1) \implies (a_{k_m} \in \mathcal{U}(a)\,).$$

(III.1.1) implies that to any $\mathcal{U}(a)$ there exists $n_0 \in \mathbb{N}$ such that $a_n \in \mathcal{U}(a)$ for all $n \geq n_0$. If we now choose n_1 to be such a natural number that $k_m \geq n_0$ for $m \geq n_1$ then apparently the required implication $(m \geq n_1) \implies (a_{k_m} \in \mathcal{U}(a))$ is satisfied for $\forall\, m \in \mathbb{N}$.

III.1.12. Example.

The sequence $\{(-1)^n \cdot n\}$ has no limit. The subsequence, consisting only of the terms with even indices n (i.e. the sequence $\{2;\ 4;\ 6;\ 8;\ \ldots\}$) has the limit $+\infty$. The "complementary" subsequence, containing only the terms with odd indices n (i.e. the subsequence $\{-1;\ -3;\ -5;\ -7;\ \ldots\}$), has the limit $-\infty$. If the sequence $\{(-1)^n \cdot n\}$ had a limit a then according to Theorem III.1.11, both subsequences would have the same limit a. However, as we see, this is not true.

The rules given by the following theorem play an important role in evaluation of concrete limits. Their proofs are omitted.

III.1.13. Theorem.

Let $\lim\limits_{n\to+\infty} a_n = a$ *and* $\lim\limits_{n\to+\infty} b_n = b$. *Then the following equalities hold:*

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

c) $\lim\limits_{n\to+\infty} (a_n \cdot b_n) = a \cdot b$, *d)* $\lim\limits_{n\to+\infty} \left(\dfrac{a_n}{b_n}\right) = \dfrac{a}{b}$,

(provided that the expressions on the right– hand sides have a sense and in the case d) the quotient $\dfrac{a_n}{b_n}$ *has a sense for all* $n \in \mathbb{N}$*).*

III.1.14. Remark.

Thus, the right– side cannot contain for example the expressions $(+\infty) + (-\infty)$, $(+\infty) - (+\infty)$, $\dfrac{\pm\infty}{\pm\infty}$, $0 \cdot (\pm\infty)$ and $\dfrac{a}{0}$. Moreover, when computing the limit of the quotient $\dfrac{a_n}{b_n}$, we need this quotient to be defined. Since the limit does not depend on the behavior of initial m_0 terms of the sequence (for an arbitrary, but fixed number $m_0 \in \mathbb{N}$ – see Remark III.1.5), from the point of view of the evaluation of the limit the quotient $\dfrac{a_n}{b_n}$ need not necessarily be defined for all $n \in \mathbb{N}$. It is sufficient when it is defined only for $n \geq m_0$.

III.1.15. Example.

$$\lim_{n\to+\infty} \frac{3n^2 + 2n - 1}{2n^2 + 1000} = \lim_{n\to+\infty} \frac{3 + (\frac{2}{n}) - (\frac{1}{n^2})}{2 + (\frac{1000}{n^2})} =$$

$$= \lim_{n\to+\infty} \frac{3 + \frac{2}{+\infty} - \frac{1}{(+\infty)^2}}{2 + \frac{1000}{(+\infty)^2}} = \frac{3 + 0 - 0}{2 + 0} = \frac{3}{2}.$$

III.1.16. Example.

$$\lim_{n\to+\infty} (\sqrt{n} - \sqrt{2n}) = \lim_{n\to+\infty} \frac{(\sqrt{n} - \sqrt{2n})(\sqrt{n} + \sqrt{2n})}{\sqrt{n} + \sqrt{2n}} =$$

$$= \lim_{n\to+\infty} \frac{n - 2n}{\sqrt{n} + \sqrt{2n}} = \lim_{n\to+\infty} \frac{-\sqrt{n}}{1 + \sqrt{2}} = -\infty.$$

III.1.17. Theorem.

Let $\{a_n\}$, $\{b_n\}$, $\{c_n\}$ *be such sequences that* $\lim\limits_{n\to+\infty} a_n = \lim\limits_{n\to+\infty} c_n = c$ *and* $\forall\, n \in \mathbb{N}$: $a_n \leq b_n \leq c_n$. *Then* $\lim\limits_{n\to+\infty} b_n = c$.

The last theorem is often called "the sandwich theorem" – try to guess why.

III.1.18. Example.

We show that $\lim\limits_{n\to+\infty} \sqrt[n]{n} = 1$. Put $\sqrt[n]{n} = 1 + \delta_n$. Raising this equality to the n–th power, we obtain:

$$n = 1 + \binom{n}{1}\delta_n + \binom{n}{2}\delta_n^2 + \binom{n}{3}\delta_n^3 + \ldots + \delta_n^n.$$

Since $\delta_n \geq 0$, we have for $n > 1$:

$$n \geq \binom{n}{2} \delta_n^2 = \frac{n\,(n-1)}{2}\,\delta_n^2 \qquad \Longrightarrow \qquad 0 \leq \delta_n \leq \sqrt{\frac{2}{n-1}}\,.$$

Obviously, $\lim\limits_{n\to+\infty} \sqrt{\dfrac{2}{n-1}} = 0$. Hence due to Theorem III.1.17, it also holds that $\lim\limits_{n\to+\infty} \delta_n = 0$. The desired result – $\lim\limits_{n\to+\infty} \sqrt[n]{n} = 1$ – now follows from the equality $\sqrt[n]{n} = 1 + \delta_n$.

III.1.19. Problems.

Evaluate the following limits (or prove that they do not exist):

a) $\lim\limits_{n\to+\infty} (n^3 - 3\,n^2 - 521\,n)$ b) $\lim\limits_{n\to+\infty} \sqrt[n]{3n}$ c) $\lim\limits_{n\to+\infty} \dfrac{3n^3 - n + 1}{2n^2 + 7n}$

d) $\lim\limits_{n\to+\infty} (-1)^n \dfrac{2n}{n-3}$ e) $\lim\limits_{n\to+\infty} (\sqrt{n+2} - \sqrt{n+1})$ f) $\lim\limits_{n\to+\infty} (\sqrt{n} - \sqrt[3]{3n})$

g) $\lim\limits_{n\to+\infty} \dfrac{\sin n\pi}{n}$ h) $\lim\limits_{n\to+\infty} \dfrac{2\sqrt{n}}{4\sqrt{n} - \sqrt[3]{n}}$ i) $\lim\limits_{n\to+\infty} \dfrac{2^n + (-1)^n}{3^n}$

Results:

a) $+\infty$ b) 1 c) $+\infty$ d) does not exist e) 0

f) $+\infty$ g) 0 h) $\dfrac{1}{2}$ i) 0.

III.1.20. Problems.

Try to prove that the following assertions hold:

a) A sequence of non–negative numbers cannot have a negative limit.

b) A sequence of non–positive numbers cannot have a positive limit.

III.2. Functions – basic notions

III.2.1. The notion of a function.

If $M \subset \mathbb{R}$, then each mapping of M to $\mathbb{R}$ is called a *real function of one real variable* (shortly: a *function*).

Functions will be denoted by letters, as for example f, g, h, φ, ψ, F, G, etc.

III.2.2. Domain, range and graph of a function.

A function is a special case of a mapping and the notions "domain of definition of a mapping" (shortly: "domain of a mapping") and "range of a mapping" are known from secondary school. Hence, the notions "domain of a function") and "range of a function" can also be regarded to be known. In accordance with the denotation which is used in connection with general mappings, $D(f)$ will be the domain of definition and $R(f)$ will be the range of function f.

A *graph* of function f is the set $G(f) = \{\, [x; f(x)\,] \in \mathbb{R}^2;\ x \in D(f)\,\}$.

III.2.3. Remark.

For example, the fact that f is the function defined in the interval $\langle 0; 2\rangle$ which assigns to each x from this interval the value $x^2 - 1$, can be written down in the following ways:

a) $f: \quad y = x^2 - 1 \quad$ for $x \in \langle 0; 2\rangle$;
b) $f(x) = x^2 - 1 \quad$ for $x \in \langle 0; 2\rangle$.

x is called the *dependent variable* or the *argument* of the function f. If we use notation a), we can call y the *independent variable*.

Functions are often defined only by formulas, without an exact specification of their domain of definition. In these cases, the domain is the set of all $x \in \mathbb{R}$ such that the formula (used in the definition of a function) has a sense. For example, the function $f(x) = \sqrt{x-2}$ (with no more specifications) has the domain $\langle 2; +\infty)$.

On an exact level, there is the following difference between f and $f(x)$: f is the denotation of a function, while $f(x)$ is the value of function f at the point x (i.e. $f(x)$ is the number that is assigned by the function f to the number x). By analogy, $f(a)$ is the value of the function f at the point a, $f(2)$ is the value of the function f at the point 2, etc.

However, the reader should be informed that this notation of functions and their values is not used consistently in scientific literature. For example, if one wishes to point out that f is a function of the variable x, then one often speaks about a "function $f(x)$" or about a "function $y = f(x)$" instead of a "function f" only.

Moreover, instead of "the function f defined by the equation $f(x) = x^2$", one often speaks about "the function x^2" only. If there is no danger of confusion, we shall also use this abridged and labor–saving notation.

As you can see, functions can be denoted and written down in various ways and some notations can even have more than one sense. (For instance, as we have already mentioned, $f(x)$ means exactly the value of function f at the point x. However, $f(x)$ can also sometimes denote function f itself.) We believe that this will not cause any problems or misunderstandings. What we have in mind will always be clear from a concrete situation, and from the circumstances in which the notations will be used.

III.2.4. Operations with functions.

A *sum of functions* f and g is a function h such that $h(x) = f(x) + g(x)$ for $x \in D(f) \cap D(g)$. We use the notation: $h = f + g$.

By analogy, we define a *difference* and a *product of functions* f and g. A *quotient of functions* f and g can also be defined similarly – however, its domain is the set $\{ D(f) \cap D(g) \} \setminus \{x \in D(g);\ g(x) = 0\}$.

The *absolute value* (or the *modulus*) *of a function* f is the function h defined by the equation $h(x) = |f(x)|$ for $x \in D(f)$. We use the notation: $h = |f|$.

III.2.5. Restriction of a function.

Suppose that f is a function and $M \subset D(f)$. A function which is defined only on M and which assigns to each $x \in M$ the same value as the function f (i.e. $f(x)$) is called the *restriction* of the function f to the set M and it is denoted by $f|_M$. The set of all values of the function f on the set M can be denoted by two symbols: $R(f|_M)$ or $f(M)$.

The graph of function f and a chosen subset $M \subset D(f)$ are seen on Fig. 14. The graph of the restriction of function f to the set M is drawn more strongly..

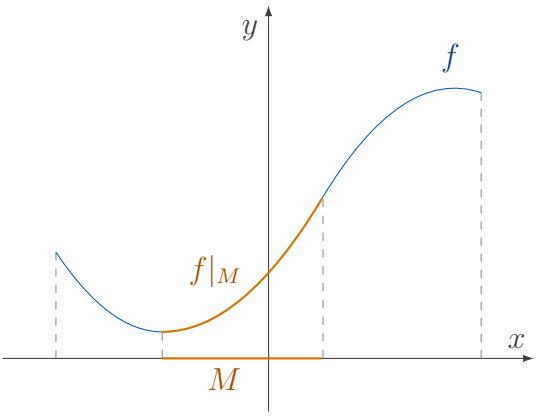


Fig. 14

III.2.6. Inverse function.

A function is a special type of a mapping. Just as there are inverse mappings to one–to–one mappings, there are also so called *inverse functions* to one–to–one functions. (Bear in mind that function f is said to be *one–to–one* if $\forall\, x_1, x_2 \in D(f)\ :\ x_1 \neq x_2 \implies f(x_1) \neq f(x_2)$.) The inverse function to function f will be denoted by f_{-1}. Its domain is $R(f)$, the range is $D(f)$ and $\forall\, x \in D(f)\ :\ y = f(x) \iff x = f_{-1}(y)$.

One can obtain the form of the inverse function from the equation $y = f(x)$ so that one calculates x from this equation (if the calculation is possible). Thus, one obtains the form of the dependence of x on y. For example: if $y = (x-1)^3$ then we calculate $x = 1 + \sqrt[3]{y}$. Interchanging x and y (just because the independent variable is usually denoted by x and the dependent variable by y), we derive that the inverse function to $y = f(x) = (x-1)^3$ (defined for $x \in \mathbb{R}$) is the function $y = f_{-1}(x) = 1 + \sqrt[3]{x}$ (also defined for $x \in \mathbb{R}$). The graphs of both functions f and f_{-1} are sketched on Fig. 15.

The graphs of the functions f and f_{-1} are symmetric with respect to the axis of the 1st and 3rd quadrant. (This follows from the fact that the coordinates of points $[x;y] \in G(f)$ satisfy the equation $y = f(x)$, which is equivalent to $x = f_{-1}(y)$. On the other hand, the coordinates of points $[x;y] \in G(f_{-1})$ satisfy the equation $y = f_{-1}(x)$, which differs from the equation $x = f_{-1}(y)$ only by the interchange of x and y. The interchange of x and y corresponds to the rotation about the straight line $y = x$ in the x, y–plane.)

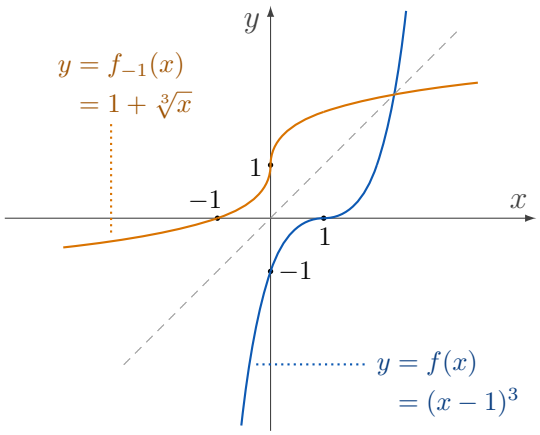


Fig. 15

III.2.7. Composite function.

If f and g are such functions that $R(g) \subset D(f)$, we can define a function h by the equation $h(x) = f(g(x))$ for $x \in D(g)$. The function h is called the *composite function* (of functions f and g). We use the notation $h = f \circ g$. f is called the *outside (external) function* and g the *inside (internal) function*.

III.2.8. Bounded functions.

Function f is called *bounded above* (or *upper bounded*) if there exists a number $K \in \mathbb{R}$ such that $\forall\, x \in D(f) \;:\; f(x) \leq K$. We can define analogously a function *bounded below* (or *lower bounded*). Function f is called *bounded* if it is bounded above and bounded below.

Assume further that $M \subset D(f)$. Function f is called *bounded above on the set* M if there exists a number $K \in \mathbb{R}$ such that $\forall\, x \in M \;:\; f(x) \leq K$ (i.e. if the restriction $f|_M$ is the function which is bounded above). We can similarly define the notion of a function *bounded below on the set M* and the notion of a function *bounded on the set M*.

III.2.9. Extreme values of a function.

We say that function f has its *maximum* at the point $x_0 \in D(f)$ if $\forall\, x \in D(f)$: $f(x) \leq f(x_0)$. We write: $f(x_0) = \max f$.
Analogously, function f has its *minimum* at the point $x_0 \in D(f)$ if $\forall\, x \in D(f)$: $f(x) \geq f(x_0)$. We write: $f(x_0) = \min f$.
The maximum and minimum of function f are both called *extreme* values of f.
Suppose that $M \subset D(f)$. We say that function f has its *maximum on the set* M at the point $x_0 \in M$ if $\forall\, x \in M$: $f(x) \leq f(x_0)$. We write: $f(x_0) = \max_M f$. Other often used denotations of maximum of function f on the set M are: $\max\limits_M f$, $\max\limits_{x\in M} f(x)$.

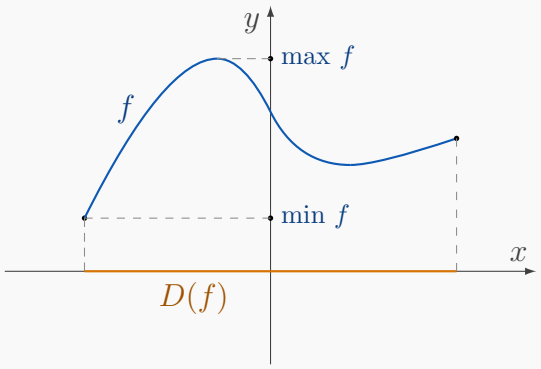


Fig. 16a

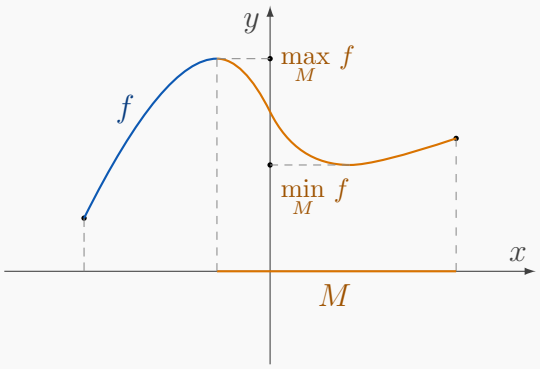


Fig. 16b

By analogy, we can also define the *minimum* of function f on the set M. We denote it: $\min_M f$, $\min\limits_M f$ or $\min\limits_{x\in M} f(x)$.
Using these definitions, one can see that the following inequalities hold:
$\max_M f = \max f|_M$ and $\min_M f = \min f|_M$.
The graph of function f with marked maximum and minimum of this function is sketched on Fig. 16a. The graph of f with a strongly drawn restriction to set M and marked maximum and minimum of f on M are sketched on Fig. 16b.

III.2.10. Supremum and infimum of a function.

The supremum, respectively infimum, of the set of values of function f (i.e. of the set $R(f)$) is called the *supremum*, respectively *infimum*, *of function* f. We use the denotation $\sup f$ (or l.u.b. f, which means the *least upper bound* of f), respectively $\inf f$ (or g.l.b. f, which means the *greatest lower bound* of f).
If $M \subset D(f)$, then the supremum, respectively infimum, of the set $f(M)$ is called the *supremum*, respectively *infimum*, of function f *on set* M. We denote it $\sup_M f$, $\sup\limits_M f$ or $\sup\limits_{x\in M} f(x)$, respectively $\inf_M f$, $\inf\limits_M f$ or $\inf\limits_{x\in M} f(x)$.

III.2.11. Remark.

While the supremum and infimum of function f (in the whole domain of f or only on a set $M \subset D(f)$) always exist, the maximum and minimum of function f (in $D(f)$ or only on $M \subset D(f)$) need not exist. This can be illustrated on the example: $f(x) = x$ for $x \in (0;1)$.

If $\max f$ exists then $\sup f = \max f$. Similarly, if $\min f$ exists then $\inf f = \min f$. (The same assertions also hold for $\max_M f$, $\sup_M f$, $\min_M f$ and $\inf_M f$.)

III.2.12. Monotonic and strictly monotonic functions.

Let f be a function and let $M \subset D(f)$. The function f is called

a) *increasing on M* if $\forall\, x_1, x_2 \in M: \quad x_1 < x_2 \implies f(x_1) < f(x_2)$;
b) *decreasing on M* if $\forall\, x_1, x_2 \in M: \quad x_1 < x_2 \implies f(x_1) > f(x_2)$;
c) *non–increasing on M* if $\forall\, x_1, x_2 \in M: \quad x_1 < x_2 \implies f(x_1) \geq f(x_2)$;
d) *non–decreasing on M* if $\forall\, x_1, x_2 \in M: \quad x_1 < x_2 \implies f(x_1) \leq f(x_2)$;
e) *monotonic on M* if f is non–increasing or non–decreasing on M;
f) *strictly monotonic on M* if f is increasing or decreasing on M.

A function that is increasing on its whole domain of definition is called shortly *increasing* (without a specification where). Similarly, one can introduce the notions of a *decreasing*, *non–increasing*, *non–decreasing*, *monotonic* and *strictly monotonic function*.

Note that instead of "monotonic", many authors also use the adjective "monotone".

The graph of a function, increasing (respectively decreasing) on set M, is sketched on Fig. 17a (respectively Fig. 17b). Fig. 18a (respectively Fig. 18b) shows the graph of a function non-decreasing (respectively non-increasing) on set M.

Notice that every increasing function is also non-decreasing, every decreasing function is also non-increasing and every strictly monotonic function is also monotonic. A function that is both non-increasing and non-decreasing is constant and a function that is both increasing and decreasing does not exist.

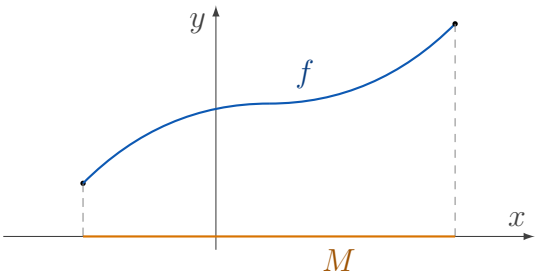


Fig. 17a

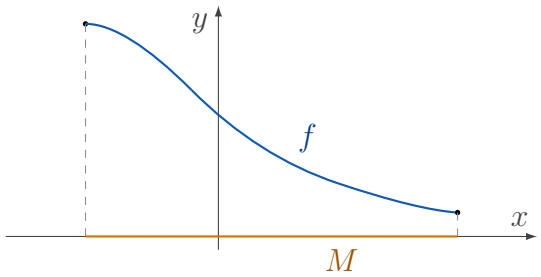


Fig. 17b

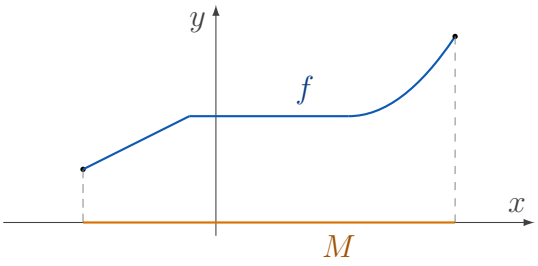


Fig. 18a

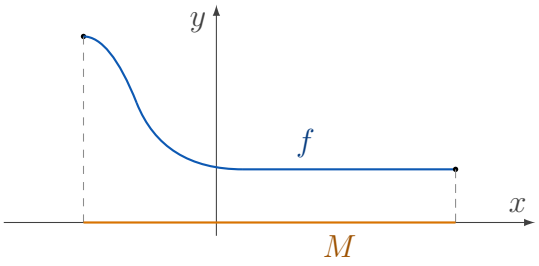


Fig. 18b

A strictly monotonic function is one–to–one, hence the inverse function exists.

III.2.13. Theorem.

If f is an increasing function then the inverse function f_{-1} is also increasing.

An analogous theorem also holds for decreasing functions. Try to prove both theorems (i.e. for functions increasing and decreasing) for yourself .

III.2.14. Even, odd and periodic functions.

Function f is called *even* (respectively *odd*) if $\forall\, x \in D(f) \,:\, -x \in D(f)$ and $f(-x) = f(x)$ (respectively $f(-x) = -f(x)$).

Function f is called *periodic with period ω* if $\forall\, x \in D(f) \,:\, x \pm \omega \in D(f)$ and $f(x \pm \omega) = f(x)$.

The graph of an even function is symmetric with respect to the y–axis and the graph of an odd function is symmetric with respect to the origin. For example, the function $f(x) = x^2$ is even and the function $g(x) = x^3$ is odd.

III.3. Some concrete functions

III.3.1. Basic elementary functions.

The following functions are known from the high school:

a) The *constant function:* $f(x) = c$ (where $c \in \mathbb{R}$),
b) The *linear function*: $f(x) = kx + q$ (where $k,\, q \in \mathbb{R}$ and $k \neq 0$),
c) The *power function*: $f(x) = x^\alpha$ (where $\alpha \in \mathbb{R}$),
d) The function *sine*: $f(x) = \sin x$,
e) The function *cosine*: $f(x) = \cos x$,
f) The function *tangent*: $f(x) = \tan\, x$,
g) The function *cotangent*: $f(x) = \operatorname{cotan}\, x$,
h) The *exponential function with base a* (where $a > 0$): $f(x) = a^x$,
i) The *logarithmic function with base a* (where $a > 0,\ and \neq 1$): $f(x) = \log_a\, x$.

The functions sine, cosine, tangent and cotangent are called *trigonometric functions*. Review for yourself the properties of these functions (i.e. their domains, ranges, graphs, important

formulas, where they are increasing, decreasing, etc., what are their suprema, infima, possibly also maxima and minima, etc.).

The following paragraphs are devoted to some other important functions: the polynomial function (paragraph III.3.2), exponential and logarithmic functions with base e (where e is the so called Euler number – paragraph III.3.3) ans with the inverse trigonometric functions (paragraphs III.3.5 – III.3.6). In paragraph III.3.4, we return to the power function.

III.3.2. Polynomial.

The constant and linear functions are special cases of the so called *polynomial function*. This function is usually shortly called a *polynomial*. *The polynomial of degree* n is a function of the form

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + a_{n-2} x^{n-2} + \ldots + a_1 x + a_0$$

(where a_n, a_{n-1}, $\ldots$, a_0 are real numbers and $a_n \neq 0$.) Particularly, if $n = 1$ then P is the *linear polynomial* (= linear function), if $n = 2$ then P is the so called *quadratic polynomial* (= quadratic function) and if $n = 3$ then P is said to be the *cubic polynomial* (= the cubic function). The graph of the linear function is a straight line, the graph of the quadratic function is a parabola and the graph of the cubic function is the so called cubic parabola. You can see the graph of the quadratic polynomial x^2-2x-3 on Fig. 19a and the graph of the cubic polynomial $x^3 - 2x^2 - 5x + 6$ on Fig. 19b. The graphs intersect the x–axis at points, whose x–coordinates are called the *roots* of the considered polynomials. (The roots can be evaluated by solution of the quadratic equation $x^2-2x-3 = 0$, respectively the cubic equation $x^3-2x^2-5x+6 = 0$.)

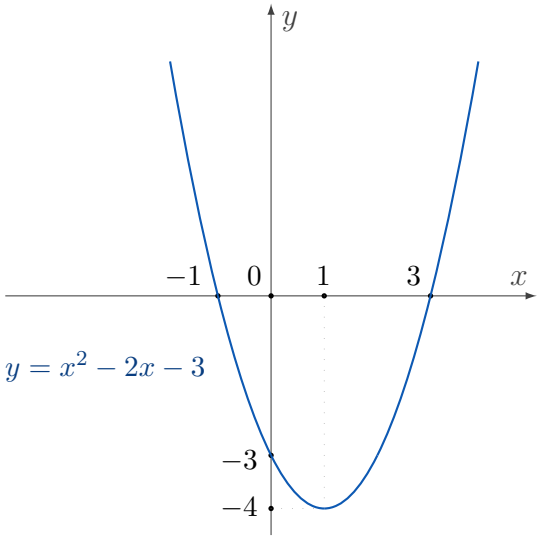


Fig. 19a

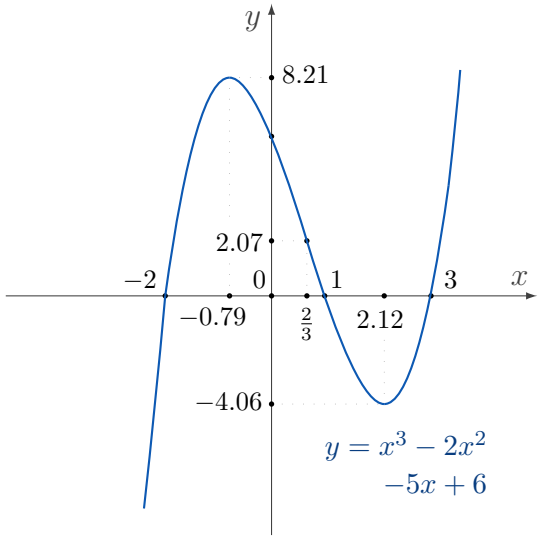


Fig. 19b

III.3.3. The Euler number e. The exponential and logarithmic function with base e.

Due to the reasons, which are explained in paragraph III.4.13, the "most important" of all exponential functions is the function e^x, where e is the so called *Euler number*. This number has the property that the angle between the tangent line to the graph of e^x at the point $x = 0$ and the x–axis is $45°$. The number e is irational and its approximate value is 2.718. It is, similarly

as Ludolf's number π, one of the most frequently used numbers in today's mathematics. In the past many formulas were found by which one can express e. For example:

$$\mathrm{e} = \lim_{n\to+\infty}\left(1+\frac{1}{n}\right)^n \qquad \text{or} \qquad \mathrm{e} = \frac{1}{0!}+\frac{1}{1!}+\frac{1}{2!}+\frac{1}{3!}+\dots.$$

Instead of e^x we often also write $\exp x$.

The logarithmic function with the general base a ($a>0$, $a\neq 1$), i.e. the function $\log_a x$, is defined to be the inverse function to a^x. Particularly, the logarithmic function with base e is the inverse function to e^x. This logarithmic function is called the *natural logarithm* and instead of $\log_{\mathrm{e}} x$, it is usually denoted by $\ln x$.

The graphs of both the functions e^x and $\ln x$ are sketched on Fig. 20.

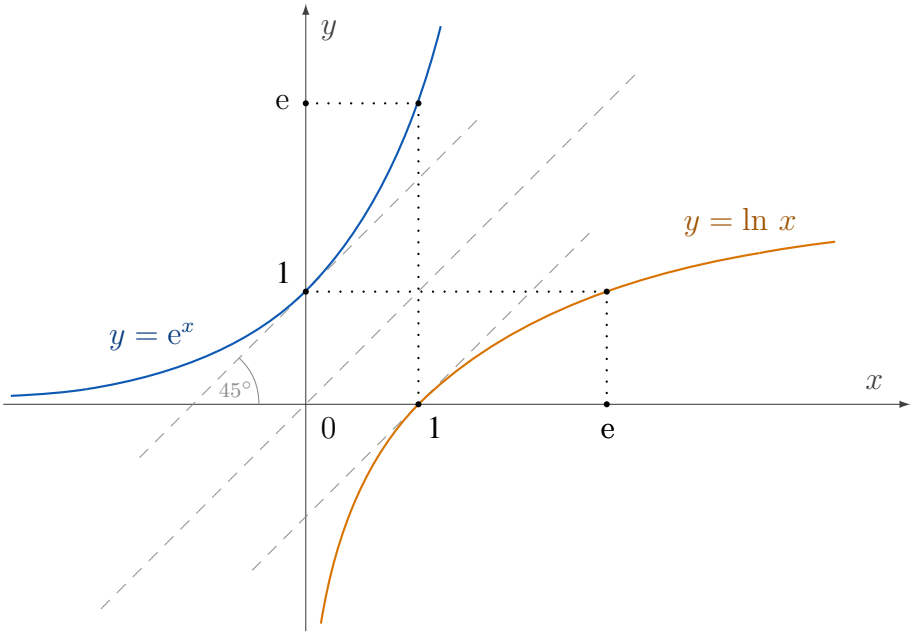


Fig. 20

It follows from the definition of the natural logarithm (i.e. from the fact that it is the inverse function to the exponential function $y=\mathrm{e}^x$) that $\ln \mathrm{e}^x = x$ for all $x\in\mathbb{R}$, and $\mathrm{e}^{\ln x} = x$ for all $x>0$. (This corresponds to the definition, which is mostly used in the high schools: $\ln x$ is such a number that e, raised to power $\ln x$, equals x.)

III.3.4. Power function.

Let us study in greater detail the function $f(x)=x^\alpha$. Its domain as well as its behavior depend essentially on the number α.

- If α is a nonnegative integer then $D(f)=\mathbb{R}$. (The example is the function x^3.)
- If α is a negative integer then $D(f)=(-\infty;+\infty)\setminus\{0\}$. (The example is x^{-2}.)
- If α is not an integer then it is usual in scientific literature to put $D(f)=(0;+\infty)$. The reason is that for α not being an integer, one can define x^a by the formula $x^\alpha = \exp(\ln x^\alpha) = \exp(\alpha\cdot\ln x)$ and $\ln x$ has a sense only for $x>0$.

- However, for some non–integer numbers α, the definition of the function x^α can be extended in a reasonable way: For $\alpha > 0$, we put $0^\alpha = 0$ and the domain of x^α thus becomes the interval $\langle 0; +\infty)$.
- For those α that are rational and can be therefore expressed as $\frac{p}{q}$, where p, q are indivisible and q is odd, one can also define x^α for $x < 0$ – we can show it for example for $\alpha = \frac{1}{3}$ and $x = -8$: $(-8)^{\frac{1}{3}} = \sqrt[3]{-8} = -\sqrt[3]{8} = -2$.

You can see the graphs of two concrete power functions $x^{\frac{1}{2}}$ (the square root) and x^{-1} (the inverse proportionality) on Fig. 21a and Fig. 21b.

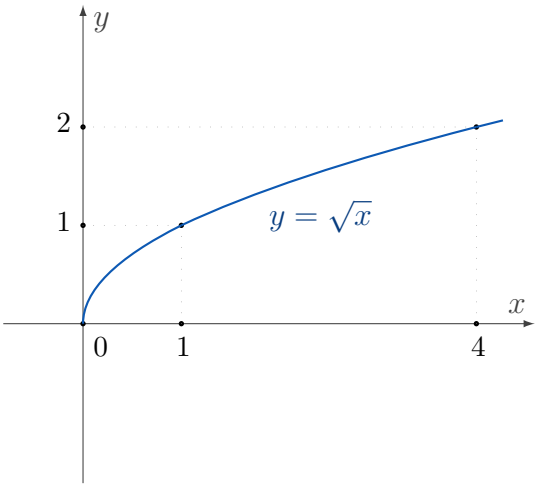


Fig. 21a

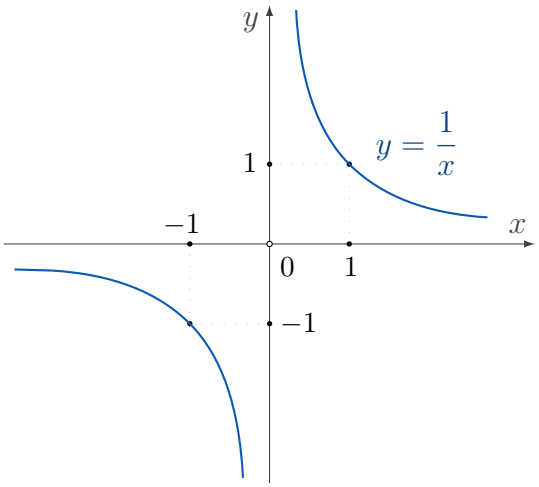


Fig. 21b

In connection with the function "square root" (Fig. 21a), we draw the reader's attention to the difference between the square root of a^2 and solution of the quadratic equation $x^2 = a^2$ (for given $a \neq 0$): While $\sqrt{a^2} = |a|$, the equation $x^2 = a^2$ has two solutions $x_{1,2} = \pm\sqrt{a^2} = \pm|a| = \pm a$. Think for yourself about this simple fact and substitute some concrete values of a, e.g. $a = 2$ and $a = -2$ to the square root and the quadratic equation.

III.3.5. Inverse trigonometric functions $\arcsin x$ (arcussine) and $\arccos x$ (arcuscosine).

The function sine is not one–to–one. Hence the inverse function to sine does not exist. However, the restriction of the function sine to the interval $\left\langle -\frac{\pi}{2}; \frac{\pi}{2} \right\rangle$ is one–to–one. The inverse function to this restriction is called *arcussine* and denoted by $\arcsin$. The domain of arcussine is the interval $\langle -1; 1\rangle$ and the range is the interval $\left\langle -\frac{\pi}{2}; \frac{\pi}{2} \right\rangle$. Thus, for $x \in \left\langle -\frac{\pi}{2}; \frac{\pi}{2} \right\rangle$ and $y \in \langle -1; 1\rangle$ we have: $y = \sin x \iff x = \arcsin y$.

You can see a part of the graph of the function sine (the so called sinusoid) on Fig. 22a, with the restriction to the interval $\left\langle -\frac{\pi}{2}; \frac{\pi}{2} \right\rangle$ drawn thickly. The graph of the restriction and the graph of the function arccussine are drawn in greater detail on Fig. 22b.

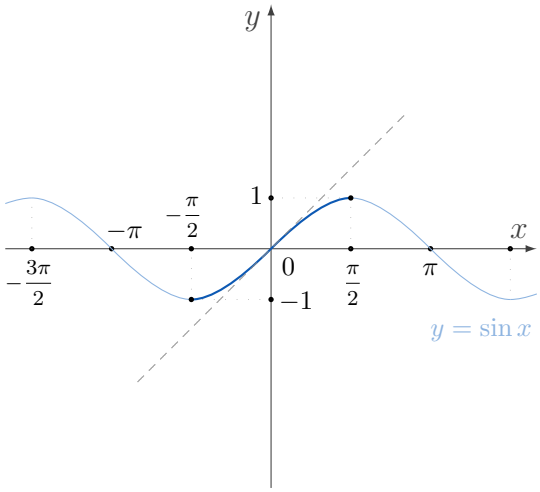


Fig. 22a

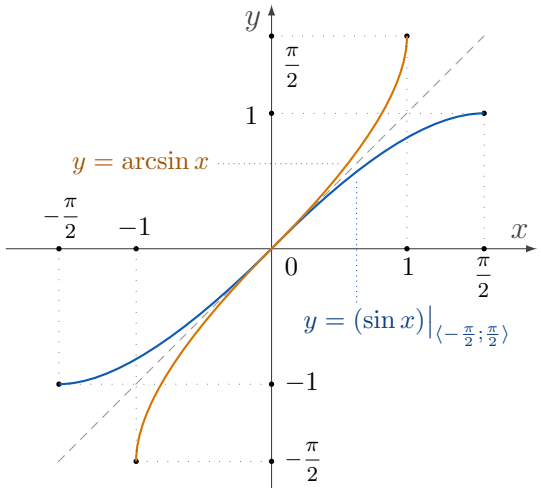


Fig. 22b

Similarly, the inverse function to the restriction of the function cosine to the interval $\langle 0;\pi\rangle$ is called *arcuscosine* and denoted by $\arccos$. Its domain is the interval $\langle -1;1\rangle$ and the range is the interval $\langle 0;\pi\rangle$. Thus, for $x \in \langle 0;\pi\rangle$ and $y \in \langle -1;1\rangle$ we have:
$y = \cos x \iff x = \arccos y$.

A part of the graph of the function cosine with the thickly marked restriction to the interval $\langle 0;\pi\rangle$ is drawn on Fig. 23a. The graph of the restriction and the graph of the inverse function arcuscosine are drawn in greater detail on Fig. 23b.

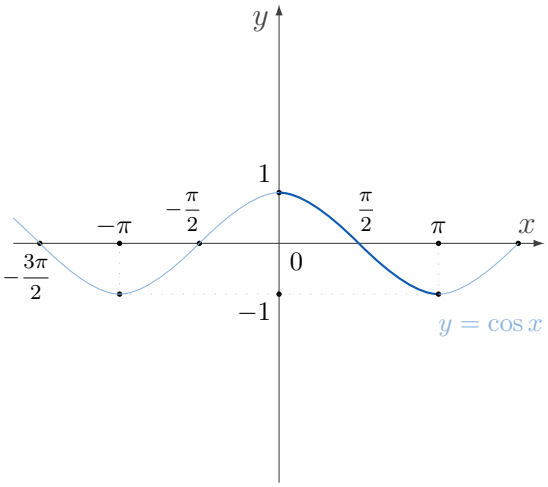


Fig. 23a

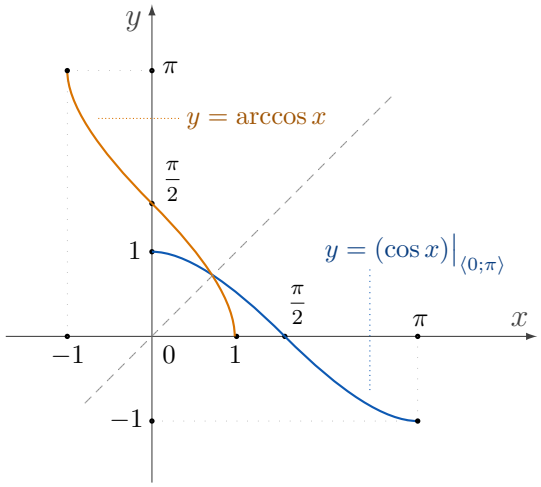


Fig. 23b

III.3.6. Inverse trigonometric functions $\arctan x$ (arcustangent) and $\operatorname{arccotan} x$ (arcuscotangent).

The inverse function to the restriction of the function tangent to the interval $\left(-\frac{\pi}{2};\frac{\pi}{2}\right)$ is called *arcustangent* and denoted by $\arctan$. Its domain is the interval $(-\infty;+\infty)$ and the range is the interval $\left(-\frac{\pi}{2};\frac{\pi}{2}\right)$. Thus, for $y \in (-\infty;+\infty)$ and $x \in \left(-\frac{\pi}{2};+\frac{\pi}{2}\right)$ we have: $y = \tan x \iff$

$x = \arctan y.$

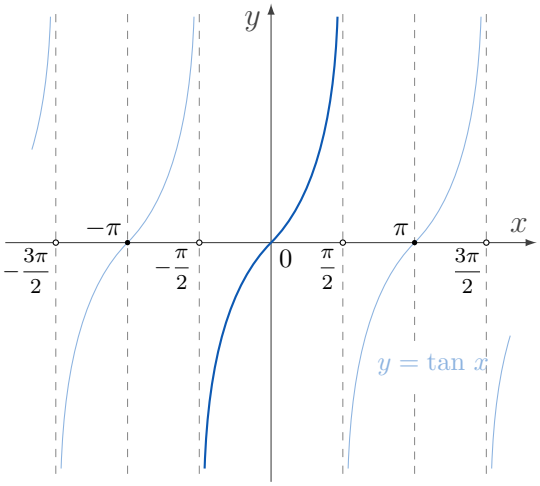


Fig. 24a

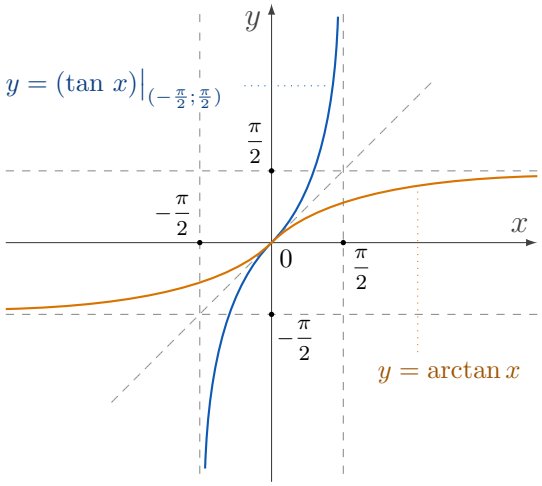


Fig. 24b

A part of the graph of the function tangent with the marked restriction to the interval $\left(-\frac{\pi}{2}; +\frac{\pi}{2}\right)$ is drawn on Fig. 24a. The graph of this restriction and the graph of the inverse function arcustangent are drawn on Fig. 24b.

The inverse function to the restriction of the function cotangent to the interval $(0; \pi)$ is called *arcuscotangent* and denoted by $\operatorname{arccotan}$. Its domain is the interval $(-\infty; +\infty)$ and the range is the interval $(0; \pi)$. Sketch the graph of the considered restriction of the function cotangent and the graph of the inverse function arcuscotangent yourself!

Note that instead of $\arcsin$, $\arccos$, $\arctan$ and $\operatorname{arccotan}$, one can often also find the notation $\sin^{-1}$, $\cos^{-1}$, $\tan^{-1}$ and $\operatorname{cotan}^{-1}$ in literature, which might be confused with reciprocate value and therefore we will not use such notation.

III.3. Limit and continuity of a function

III.3.1. Example.

The domain of the function $f(x) = \dfrac{\mathrm{e}^x - 1}{x}$ is the set $\mathbb{R} \setminus \{0\}$. The following table shows function values of f at some points x:

x	-1	-0.2	-0.05	-0.001	0.001	0.05	0.2	1
$f(x)$	0.6320	0.9060	0.9754	0.9995	1.0005	1.0254	1.1070	1.7182

It is seen from the table that $f(x)$ "approaches" one as x "approaches" zero. This fact, which we have expressed only on an intuitive level so far, can be precisely described by means of the notion of the "limit of a function".

III.3.2. The limit of a function.

Assume that $c \in \mathbb{R}^*$ and the domain of function f contains some reduced neighborhood $\mathcal{P}(x_0)$. If for each sequence $\{x_n\}$ in $\mathcal{P}(x_0)$ the implication

$$x_n \to x_0 \implies f(x_n) \to a,$$

is true, then we say that function f has the *limit* equal to a at point x_0.
We write: $\lim\limits_{x\to x_0} f(x) = a$.

You will see later that the limit from example III.3.1 is indeed: $\lim\limits_{x\to x_0} \dfrac{\mathrm{e}^x - 1}{x} = 1$.

III.3.3. Remark.

Neither the existence nor the value of the limit of function f at the point x_0 depends on whether the point x_0 belongs to $D(f)$. If x_0 belongs to $D(f)$ then the function value $f(x_0)$ does not affect the existence and the value of the limit of f at point x_0. The existence and the value of this limit are exclusively given by the behavior of function f in the reduced neighborhood of point x_0 and not at point x_0 itself!

The limit, whose value is a finite real number, is called a *proper limit*. The limit, whose value is $+\infty$ or $a = \infty$, is called an *improper limit*.

The following theorem is an easy consequence of Theorem III.1.8.

III.3.4. Theorem.

Function f can have at any point $x_0 \in \mathbb{R}^*$ at most one limit.

III.3.5. Example.

Function $f(x) = \sin x$ has no limit at $+\infty$. Actually, if it had a limit (equal a), then for each sequence $\{x_n\}$ in $\mathbb{R}$ the implication $x_n \to \infty \implies \sin x_n \to a$ would have to be true. However, for example the sequence $\{x_n\}$, where $x_n = \dfrac{\pi}{2} + n\pi$, does not satisfy this implication. This sequence has the limit $+\infty$, but the sequence $\{\sin x_n\}$ (i.e. the sequence $\{(-1)^n\}$) has no limit.

It would be very clumsy and inefficient always to evaluate limits from their definition (i.e. by means of limits of sequences). For this reason we will show more effective procedures in this chapter. The following theorem is very important. It concerns the limit of a sum, a difference, a product and a quotient of two functions and it can easily be proved by means of Theorem III.2.13. To save space (and to avoid writing almost the same thing four times), we use the symbol "$\circledast$" which has the meaning of any of the symbols "$+$", "$-$", "$\cdot$" and "$/$" here. (This means that you can replace "$\circledast$" by any of the symbols "$+$", "$-$", "$\cdot$", "$/$" and you obtain a valid theorem.)

III.3.6. Theorem.

Let $\lim\limits_{x\to x_0} f(x) = a$ and $\lim\limits_{x\to x_0} g(x) = b$. Then

$$\lim_{x\to x_0} \big(f(x) \circledast g(x)\big) = a \circledast b$$

(if the expression $a \circledast b$ has a sense).

III.3.7. Example.

$$\lim_{x\to 2}(3x^2+2x-1) = \lim_{x\to 2}(3x^2) + \lim_{x\to 2}(2x) - 1 = 12+4-1 = 15$$

III.3.8. Remark.

If for example $\lim_{x\to x_0} f(x) = \lim_{x\to x_0} g(x) = +\infty$ then the limit $\lim_{x\to x_0} \frac{f(x)}{g(x)}$ cannot be computed by Theorem III.3.6 because the expression $\frac{+\infty}{+\infty}$ has no sense.

III.3.9. Remark.

It can be shown that if $\lim_{x\to x_0} f(x) = a > 0$, $\lim_{x\to x_0} g(x) = 0$ and $g(x) > 0$ for all x from some reduced neighborhood $\mathcal{P}(x_0)$, then $\lim_{x\to x_0} \frac{f(x)}{g(x)} = +\infty$. The limit of the quotient $\frac{f(x)}{g(x)}$ cannot be evaluated by means of Theorem III.3.7 (because the fraction $\frac{a}{0}$ has no sense); nevertheless since $f(x)$ approaches the positive number a as $x \to x_0$ and $g(x)$ approaches 0 from the right side, i.e. from the domain of positive numbers, the quotient $\frac{f(x)}{g(x)}$ approaches $+\infty$. A similar reasoning can also be used in the case when $a < 0$ or $g(x) < 0$ for all $x \in \mathcal{P}(x_0)$.

III.3.10. Example.

$\lim_{x\to 0} \frac{1}{x^2} = +\infty$ – this is the consequence of Remark III.3.9.

Some other methods and examples will be discussed in this chapter after the notion "continuity of a function". Moreover, the so called l'Hospital rule will be explained in paragraph III.5.36; you will appreciate it as a useful aid for evaluations of limits of a quotient of two functions.

III.3.11. One–sided limits.

Suppose that $x_0 \in \mathbb{R}$ and the domain of function f contains some right neighborhood $\mathcal{P}_+(x_0)$. If for every sequence $\{x_n\}$ in $\mathcal{P}_+(x_0)$ the implication

$$x_n \to x_0 \implies f(x_n) \to a,$$

holds, then we say that function f has the *right–hand limit* equal to a at point x_0. We write: $\lim_{x\to x_0+} f(x) = a$.

One can analogously define for $x_0 \in \mathbb{R}$ the notion of the *left–hand limit* of function f at point x_0. We write: $\lim_{x\to x_0-} f(x) = a$.

Comparing these definitions with the definition of the "both–sided" limit of a function (paragraph III.3.2), we obtain the following theorem:

III.3.12. Theorem.

Function f has a limit equal to a at the point $x_0 \in \mathbb{R}$ if and only if it has a right–hand limit and the left–hand limit at point x_0 and they are both equal to a.

Theorems analogous to Theorems III.3.4 and III.3.7 also hold for one sided limits.

III.3.13. Remark.

We can also formulate a theorem analogous to Theorem III.1.17 for the limit of a function. (It can also be called "the sandwich theorem".) Roughly speaking, the theorem says the following: *If the graph of function f is closed between the graphs of functions g and h on some reduced neighborhood $\mathcal{P}(x_0)$ and if both functions g, h have for $x \to x_0$ the same limit equal to a, then function f also has for $x \to x_0$ the limit equal to a.* Try to formulate and to provide a precise proof of this theorem and the version of it that concerns one sided limits!

III.3.14.* Remark.

To conclude the part about the limits of functions, let us return to the definition of the limit once again. There exist more equivalent definitions which have been formulated in the process of development of the differential calculus. Try to verify for yourself that the fact that $\lim_{x\to x_0} f(x) = a$ can also be defined in this way:

$$\forall\, \mathcal{U}(a) \quad \exists\, \mathcal{P}(x_0) \quad \forall\, x \in \mathbb{R} : \quad x \in \mathcal{P}(x_0) \implies f(x) \in \mathcal{U}(a),$$

or in the case when x_0 and a are numbers from $\mathbb{R}$ also in the other way:

$$\forall\, \epsilon > 0 \quad \exists\, \delta > 0 \quad \forall\, x \in \mathbb{R} : \quad 0 < |x - x_0| < \delta \implies |f(x) - a| < \epsilon.$$

III.3.15. Continuity of a function – motivation.

You can see graphs of two functions in Fig. 25a and Fig. 25b. The difference between these graphs is apparent: while the graph of function f can be drawn by one motion of a pen, without lifting it from the paper, the graph of function g is "disconnected" at the point $x = x_0$. Rather than speaking of whether the graph of some function is or is not "connected" at point x_0, one speaks in mathematics about the so called "continuity" or "discontinuity" of the function at point x_0.

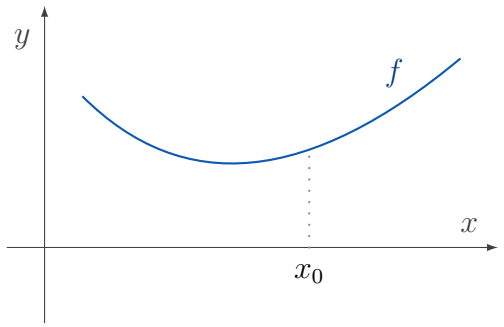


Fig. 25a

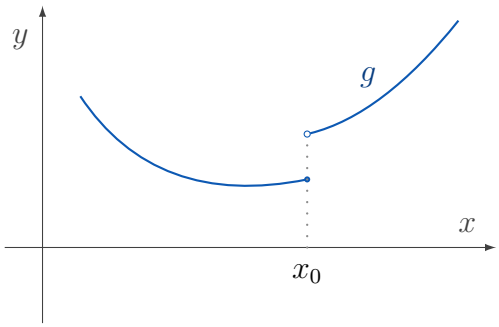


Fig. 25b

III.3.16. Continuity of a function at the point.

We say that function f is *continuous* at the point $x_0 \in D(f)$ if

(III.3.1) $$\lim_{x\to x_0} f(x) = f(x_0).$$

III.3.17. Remark.

If you read this definition and definition III.3.2 carefully, you can see that function f can be continuous at point x_0 only if it is defined in some neighborhood of x_0 (i.e. if $D(f)$ contains some neighborhood $\mathcal{U}(x_0)$).

Function g in Fig. 25b has the right– hand limit different from the left– hand limit at point x_0; hence its "both–sided" limit at point x_0 does not exist (see Theorem III.3.12). Thus the equality (III.3.1) does not hold and, consequently, function g is not continuous at the point x_0.

III.3.18. Right continuity and left continuity.

Function f is called *right continuous* (respectively *left continuous*) at the point $x_0 \in D(f)$, if

$$\lim_{x\to x_0+} f(x) = f(x_0) \qquad \text{(respectively } \lim_{x\to c-} f(x)\, x_0 = f(x_0)\text{)}.$$

The following assertion is an easy consequence of Theorem III.3.12: *Function f is continuous at point x_0 if and only if it is right and left continuous at this point.*

III.3.19. Example.

The function $y = \sqrt{x}$ is not continuous at the point $x_0 = 0$ because it is not defined in any left neighborhood of point x_0. However, this function is right continuous at the point $x_0 = 0$.

III.3.20. Continuity on an interval.

Let I be an interval in $\mathbb{R}$ which is a part of the domain of function f. We say that f is *continuous on the interval I*, if

a) f is continuous in every interior point of the interval I,
b) f is right continuous at the left end point of I (if this point belongs to I),
c) f is left continuous at the right end point of I (if this point belongs to I).

III.3.21. Example.

The function $f(x) = 3x^2 + 2x - 1$ is continuous in $\mathbb{R}$. To verify this fact, it is necessary to show that function f is continuous at each point $x_0 \in \mathbb{R}$ (because $D(f) = \mathbb{R}$). Thus, let x_0 be an arbitrarily chosen point from $\mathbb{R}$ and let $\{x_n\}$ be an arbitrary sequence of numbers from $\mathbb{R}$, such that $\lim_{n\to+\infty} x_n = x_0$. We need to show that $\lim_{n\to+\infty} f(x_n) = f(x_0)$, i.e. $\lim_{n\to+\infty} (3x_n^2 + 2x_n - 1) = 3x_0^2 + 2x_0 - 1$. By Theorem III.3.7 we get: $\lim_{n\to+\infty} (3x_n^2 + 2x_n - 1) =$

$$\lim_{n\to+\infty} 3x_n^2 + \lim_{n\to+\infty} 2x_n - \lim_{n\to+\infty} 1 = 3 \cdot \left(\lim_{n\to+\infty} x_n\right)^2 + 2 \cdot \lim_{n\to+\infty} x_n - 1 = 3x_0^2 + 2x_0 - 1.$$

It would be too laborious if we were always to investigate the continuity of a given function in the way that we have just shown. This can be done much more effectively for example by means of Theorems III.3.22, III.3.23 and III.3.25.

III.3.22. Theorem.

Polynomials, trigonometric functions (i.e. the functions sine, cosine, tangent, cotangent), inverse trigonometric functions (i.e. the functions arc sine, arc cosine, arc tangent, arc cotangent), power function, exponential function and logarithmic functions are all continuous functions in each interval which is a part of their domain.

III.3.23. Remark.

If you sketch the graph of function tangent, you can see that function tangent is continuous in each interval of the type $\left(-\frac{\pi}{2} + k\pi; \frac{\pi}{2} + k\pi\right)$ (where k is an integer). Function tangent is not continuous at the points $\frac{\pi}{2} + k\pi$ (where k is an integer). However, this is not a contradiction

with the assertion of Theorem III.3.22 because the points $\frac{\pi}{2} + k\pi$ do not belong to the domain of function tangent.

III.3.24. Theorem (on continuity of the sum, the difference, the product, the quotient and the absolute value).

If functions f and g are continuous at point c, then also the functions $f+g$, $f-g$, $f\cdot g$, and $|f|$ are continuous at point c. If, in addition, $g(c)\neq 0$ then the function $\frac{f}{g}$ is also continuous at point c.

(This part of the theorem is also valid in the case when we replace "continuity at point c" by "right continuity at point c" or by "left continuity at point c".)

If functions f and g are continuous on the interval I then the functions $f+g$, $f-g$, $f\cdot g$ and $|f|$ are also continuous in the interval I. If, in addition, $g(x)\neq 0$ for all $x\in I$ then the function $\frac{f}{g}$ is also continuous on the interval I.

III.3.25. Theorem (on continuity of a composite function).

If function g is continuous at point x_0 and function f is continuous at point $g(x_0)$ then the composite function $f\circ g$ is continuous at point x_0.

If function g is continuous on the interval I, function f is continuous on the interval J and $g(I)\subset J$ then the composite function $f\circ g$ is also continuous on the interval I.

Continuous functions have many interesting and important properties. We will show at least some of them in the following theorems. The theorems have an understandable geometric meaning. Try to illustrate it on appropriate figures for yourself.

III.3.26. Darboux' theorem.

If function f is continuous on an interval I and x_1, x_2 are any two points from I then to any given number η between $f(x_1)$ and $f(x_2)$ there exists a point ξ between x_1 and x_2 such that $f(\xi)=\eta$.

III.3.27. Remark.

The above theorem is often also called "the intermediate value theorem". It is logical, because the theorem says that if f is continuous on the interval I and v_1, v_2 are its two arbitrary values in I (i.e. $v_1, v_2 \in f(I)$), then f takes on every value between v_1 and v_2 in I.

As an easy consequence of Theorem III.3.28, we can assert: *If f is a continuous function on the interval I then $f(I)$ is also an interval or it is a one point set.* (Thus, the range of function f on I is "connected".)

III.3.28. Theorem (on continuity of the inverse function).

If function f is continuous and one–to–one on the interval I and $f(I)=J$ then the inverse function f_{-1} is continuous on the interval J.

III.3.29. Theorem (on the existence of maximum and minimum).

A function which is continuous on a closed bounded interval $\langle a;b\rangle$ has its maximum and minimum on this interval. (Thus, $\max\limits_{x\in\langle a;b\rangle} f(x)$ and $\min\limits_{x\in\langle a;b\rangle} f(x)$ exist.)

III.3.30. Theorem.

Let function f be continuous at point x_0. If $f(x_0) > 0$ then there exists a neighborhood $\mathcal{U}(x_0)$ such that $f(x) > 0$ for all $x \in \mathcal{U}(x_0)$.

P r o o f :
We show the proof by contradiction, which is simple and illustrative. Suppose first that the theorem is not true. Then in every neighborhood $\mathcal{U}(x_0)$, one can find a point x such that $f(x) \leq 0$. Since $\mathcal{U}(x_0)$ can be taken smaller and smaller, for example $\mathcal{U}(x_0) = \mathcal{U}_{\frac{1}{n}}(x_0)$, one gets a sequence $\{x_n\}$ such that $[x_n \to x_0] \wedge [f(x_n) \leq 0 \text{ for all } n \in \mathbb{N}]$. It follows from the continuity of f at the point x_0 that $f(x_n) \to f(x_0)$. The sequence $\{f(x_n)\}$ is a sequence of non–positive numbers; such a sequence cannot have a positive limit. This is the desired contradiction with the assumption $f(x_0) > 0$. Thus, the theorem is true.

One can analogously prove that *if function f is continuous at point x_0 and $f(x_0) < 0$ then there exists a neighborhood $\mathcal{U}(x_0)$ such that $f(x_0) < 0$ for all $x \in \mathcal{U}(x_0)$.*

III.3.31. Remark.

Let us now turn our attention to evaluations of limits of functions again. It follows immediately from the definition of the notion "continuity at the point" (see paragraph III.3.16) and from Theorem III.3.23 that if f is any of the functions named in Theorem III.3.23 and x_0 is an interior point of the domain of f, then $\lim\limits_{x\to x_0} f(x) = f(x_0)$.

The following theorems can also be often used.

III.3.32. Theorem (1st theorem on limit of a composite function).

Let $\lim\limits_{x\to x_0} g(x) = \lambda$, $\lambda \in \mathbb{R}$ and let function f be continuous at point λ.
Then $\lim\limits_{x\to x_0} f(g(x)) = f(\lambda)$.

III.3.33. Example.

Let us evaluate $\lim\limits_{x\to 0} \exp(1 - x^2)$. The inside function, i.e. the function $g(x) = 1 - x^2$, has the limit equal to 1 at the point $x = 0$. The outside function, i.e. the function $f(y) = \exp y$, is continuous at the point $y = 1$ and its value at this point is $\exp 1 = \mathrm{e}$. So we have: $\lim\limits_{x\to 0} \exp(1 - x^2) = \mathrm{e}$.

III.3.34. Theorem (2nd theorem on the limit of a composite function).

Let $\lim\limits_{x\to x_0} g(x) = +\infty$ (respectively $-\infty$). Let $\lim\limits_{y\to+\infty} f(y) = L$ (respectively $\lim\limits_{y\to-\infty} f(y) = L$). Then $\lim\limits_{x\to x_0} f(g(x)) = L$.

III.3.35. Remark.

Theorems III.3.32 and III.3.34 remain valid even if we modify them in such a way that we replace limits for $x \to x_0$ by one–sided limits, taken for $x \to x_0+$ or $x \to x_0-$.

III.3.36. Example.

Evaluate the limit $\lim\limits_{x\to 1+} \arctan \dfrac{x}{x-1}$. The inside function $g(x) = \dfrac{x}{x-1}$ has the right–hand limit equal to $+\infty$, at point 1. (The numerator x has the limit 1, the denominator $x - 1$ has the

limit 0 and it tends to zero from the right, i.e. from the domain of positive numbers. Hence the limit of $\frac{x}{x-1}$ is $+\infty$.) The outside function arc tangent has the limit equal to $\frac{\pi}{2}$ at $+\infty$. Hence we obtain: $\lim\limits_{x\to 1+} \operatorname{arctan} \frac{x}{x-1} = \frac{\pi}{2}$.

III.3.37. Example.

Evaluate the limit $\lim\limits_{x\to 0} \frac{\sin x}{x}$. Let us first deal with the right–hand limit. If x belongs to a right neighborhood of 0, for example to the interval $\left(0; \frac{\pi}{2}\right)$, then $\sin x \le x$ and $x \le \tan x$. Thus, for these x, we have:

$$\frac{\sin x}{x} \le \frac{x}{x} \le 1, \qquad \frac{\sin x}{x} = \frac{\tan x}{x} \cos x \ge \cos x.$$

So the function $\frac{\sin x}{x}$ is "closed" between the function $\cos x$ (from below) and the constant function 1 (from above) for $x \in \left(0; \frac{\pi}{2}\right)$. Since $\lim\limits_{x\to 0+} \cos x = 1$ and $\lim\limits_{x\to 0+} 1 = 1$, we also have: $\lim\limits_{x\to 0+} \frac{\sin x}{x} = 1$. (See Remark III.3.13.) We can show similarly that the left– limit is also equal to one. Applying Theorem III.3.12, we finally get:

$$\lim_{x\to 0} \frac{\sin x}{x} = 1.$$

We will show another, simpler, method of evaluation of this limit in paragraph III.5.36. (It will be based on the so called l'Hospital Rule.) Nevertheless, we regard the procedure used here as instructive, too.

III.3.38. Problems.

Evaluate the following limits.

a) $\lim\limits_{x\to+\infty} \frac{x^2-2x+100}{3x^2+15x-5}$ b) $\lim\limits_{x\to 3} (x^3+2x-7)$ c) $\lim\limits_{x\to 0+} \frac{1}{\sqrt{x}}$

d) $\lim\limits_{x\to 1-} \arccos x$ e) $\lim\limits_{x\to -1} \frac{x^5+x^3-x+1}{x^4-2x+1}$ f) $\lim\limits_{x\to 2} \left(\frac{8}{x^2-4} - \frac{2}{x-2}\right)$

g) $\lim\limits_{x\to -3} \frac{\sqrt{4+x}-1}{x+4}$ h) $\lim\limits_{x\to -\infty} (x^5-2x^2)$ i) $\lim\limits_{x\to+\infty} (x^5-10x^4+155)$

j) $\lim\limits_{x\to 0} \ln\left(\frac{x+1}{x}\right)$ k) $\lim\limits_{x\to+\infty} \frac{\sqrt{x+\sqrt{x}}-1}{\sqrt[3]{x}-\sqrt{x}}$ l) $\lim\limits_{x\to+\infty} x\cdot\sin x$

m) $\lim\limits_{x\to-\infty} \operatorname{arccotan} x$ n) $\lim\limits_{x\to+\infty} \frac{\sin x}{x}$ o) $\lim\limits_{x\to+\infty} \sqrt{x\cdot\left(\sqrt{x-3}-\sqrt{x}\right)}$

Results:

a) $\frac{1}{3}$ b) 26 c) $+\infty$ d) 0

e) 0 f) $-\frac{1}{2}$ g) 0 h) $-\infty$

i) $+\infty$ j) does not exist k) -1 l) does not exist

m) π n) 0 o) does not exist.

(We will return to evaluation of some types of limits in paragraphs III.5.34–III.5.36.)

III.4. Derivative of a function

If you are an astronomer, it is important for you not only to know the immediate position of the objects you observe, but also to have some information about the rate of change of this position. If you are sitting in a moving car, it is not the velocity itself that causes the power effects you feel, but the changes in velocity. If you own shares in a company, it is not only today's price that interests you, but also the rate of change of the price – i.e. whether and how fast their value is increasing or decreasing. These simple examples can be generalized: important information about a function involves not only its value at one or more points, but also the rate of change (growth or decay) of the value at a given point (or points). The necessity to express the rate of growth or decay of a function leads to the introduction of the notion of the derivative.

We will describe two concrete situations that lead in a natural way to the derivative of a function. However, there exist many more applications and possible interpretations of this notion in miscellaneous scientific disciplines.

III.4.1. Geometrical motivation.

The rate of change of function f at point x_0 can be expressed by the slope of the tangent to the graph of f at the point $X_0 = [\,x_0; f(x_0)\,]$. How to find this slope? We choose another, "variable" point $X = [\,x_0 + h; f(x_0 + h)\,]$ on the graph of f. The straight line X_0X has the slope

$$\tan\,\alpha \;=\; \frac{f(x_0+h)-f(x_0)}{h}.$$

If there exists a limit of this expression as $h \to 0$ then the slope of the tangent to the graph of f at the point X_0 is equal to its value.

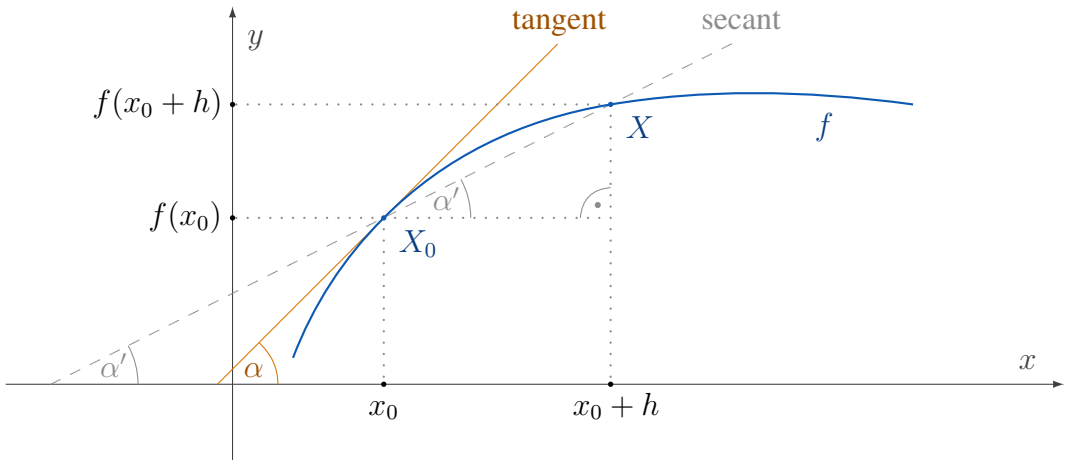


Fig. 26

III.4.2. Physical motivation.

Suppose that a mass point moves on a straight line. The position of the mass point at time t is $s(t)$. The distance which is run by the mass point during the time interval $\langle t_0; t_0 + h\rangle$ is $s(t_0 + h) - s(t_0)$ and so the average velocity of the mass point in this time interval is equal to

$$\frac{s(t_0+h)-s(t_0)}{h}.$$

If there exists a limit of this expression as $h \to 0$ then its value is called the *instantaneous velocity* of the moving point at time t_0.

III.4.3. Derivative of a function.

If there exists a finite limit

$$\lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h}, \tag{III.4.1}$$

then its value is called the *derivative of function f at the point x_0* and it is denoted by $f'(x_0)$.

A function which has a derivative at point x_0 is said to be *differentiable* at x_0.

A function which assigns to each $x \in D(f)$ the derivative $f'(x)$ (if the derivative at the point x exists) is called the *derivative of function f* and it is denoted by f'. The domain of the function f' satisfies: $D(f') \subseteq D(f)$.

III.4.4. Remark.

If we denote $x = x_0 + h$ then we can also write the limit (III.4.1) in the form

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

If function f is given by the equation $y = f(x)$, then the derivative of f is, except for f', often also denoted by the symbols

$$\frac{\mathrm{d}f}{\mathrm{d}x}, \qquad \frac{\mathrm{d}}{\mathrm{d}x} f, \qquad y', \qquad \frac{\mathrm{d}y}{\mathrm{d}x}.$$

III.4.5. The tangent line to the graph of a function.

Suppose that $x_0 \in D(f)$. If function f has the derivative $k = f'(x_0)$ at the point x_0 then the *tangent* to the graph of f at the point $[x_0; f(x_0)]$ is the straight line which is given by the equation $y - f(x_0) = k \cdot (x - x_0)$ that can be rewritten as

$$y = f(x_0) + f'(x_0) \cdot (x - x_0).$$

III.4.6. The velocity of motion.

Let us return to the situation described in paragraph III.4.2. We can see that it is natural to define the velocity of a moving mass point at time t_0 as the derivative of the position function s at time $t = t_0$.

III.4.7. The right derivative and the left derivative.

If there exists a finite one–sided limit

$$\lim_{x\to x_0-} \frac{f(x)-f(x_0)}{x-x_0} \qquad \left(\text{respectively} \qquad \lim_{x\to x_0+} \frac{f(x)-f(x_0)}{x-x_0}\right),$$

then the value of this limit is called the *left derivative* (respectively the *right derivative*) of function f at the point x_0 and we denote it by $f'_-(x_0)$ (respectively by $f'_+(x_0)$).

It follows immediately from Theorem III.3.12 that $f'(x_0) = k \iff f'_-(x_0) = f'_+(x_0) = k$.

III.4.8. Theorem.

If function f has a derivative at the point x_0 then it is continuous at this point.

If function f has a left derivative (respectively a right derivative) at the point x_0 then it is left continuous (respectively right continuous) at this point.

Proof:
We show only the proof of the first part of the theorem. The proof of the second part could be performed analogously.

It follows from the existence of the derivative $f'(x_0)$ that function f is defined in some neighborhood $\mathcal{U}(x_0)$. For $x \in \mathcal{P}(x_0)$, we can write: $f(x) = f(x_0) + \left\{ \frac{f(x)-f(x_0)}{x-x_0} \right\} \cdot (x - x_0)$. The expression in braces (i.e. $\{\dots\}$) approaches $f'(x_0)$ as $x \to x_0$ and $(x - x_0)$ approaches 0 as $x \to x_0$. By Theorem III.3.7, we get

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

This means that function f is continuous at the point x_0.

If no confusion can arise then we shall further write only x instead of x_0.

III.4.9. Remark.

As an immediate consequence of Theorem III.4.8, we can formulate the following assertion: *Let I be an interval with end points a, b and let $a < b$. Let function f be differentiable at each point $x \in (a; b)$, let there exist $f'_+(a)$ (if a belongs to I) and let there exist $f'_-(b)$ (if $b \in I$). Then function f is continuous on the interval I.*

III.4.10. Theorem.

Let functions f and g be differentiable at point x and let $c \in \mathbb{R}$. Then the functions $c \cdot f$, $f + g$, $f - g$ and $f \cdot g$ are also differentiable at point x and the following formulas hold:

a) $(c \cdot f)'(x) = c \cdot f'(x)$, *b)* $(f + g)'(x) = f'(x) + g'(x)$,
$(f - g)'(x) = f'(x) - g'(x)$, *d)* $(f \cdot g)'(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x)$.

If, in addition, $g(x) \neq 0$, then the quotient $\frac{f}{g}$ also has a derivative at the point x and it can be expressed by the formula:

e) $\left(\frac{f}{g}\right)'(x) = \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{g^2(x)}$.

Proof:
All formulas follow from (III.4.1). We show the derivation of only one of them, for example of the formula from item d).

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

$$= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}\, g(x+h) \;+\; \lim_{h\to 0} f(x)\, \frac{g(x+h)-g(x)}{h} \;=$$
$$= \; f'(x)\cdot g(x) \;+\; f(x)\cdot g'(x).$$

III.4.11. Remark.

The formulas a) – e) from Theorem III.4.10 are often written down in such a way that instead of f and g, one uses the denotation u and v and in order to simplify the formulas, one omits (x). Then the formulas have the form:

a) $(c\cdot u)' = c\cdot u'$, b) $(u+v)' = u'+v'$, c) $(u-v)' = u'-v'$,
d) $(u\cdot v)' = u'\,v + u\,v'$, e) $\left(\frac{u}{v}\right)' = \frac{u'\,v - u\,v'}{v^2}$.

III.4.12. Derivatives of some elementary functions.

Evaluating the limit in (III.4.1) and applying the formulas a) – e) from Theorem III.4.10, one can derive the concrete form of the derivatives of some elementary functions:

a) $(\,c\,)' = 0$ (c is the constant function.) b) $(\,x^\alpha\,)' = \alpha\cdot x^{\alpha-1}$ ($\alpha\in\mathbb{R}$, $\alpha\neq 0$)
c) $(\,\sin x\,)' = \cos x$ d) $(\,\cos x\,)' = -\sin x$
e) $(\,\tan\, x\,)' = \dfrac{1}{\cos^2 x}$ f) $(\,\mathrm{cotan}\, x\,)' = -\dfrac{1}{\sin^2 x}$

These formulas hold at all points x from the domain of the function that appears in the formula. The exception is formula b) in the case when $\alpha\in(0;1)$. In this case formula b) has no sense for $x = 0$.

In the following, we present the derivation of one of the formulas a) – f), for example the formula for the derivative of the function sine.

$$(\sin x)' \;=\; \lim_{h\to 0}\frac{\sin(x+h)-\sin x}{h} \;=\; \lim_{h\to 0}\frac{\sin x\cdot\cos\,h+\cos\,x\cdot\sin\,h-\sin x}{h} \;=$$
$$= \;\sin x\cdot\lim_{h\to 0}\frac{\cos\,h-1}{h} \;+\; \cos x\cdot\lim_{h\to 0}\frac{\sin\,h}{h} \;=\; \sin\,x\cdot 0 \;+\; \cos x\cdot 1 \;=\; \cos x$$

We have used the result of example III.3.37 (on the limit of the quotient $\dfrac{\sin\,h}{h}$) and moreover, we have also used the fact that $\lim\limits_{h\to 0}\dfrac{\cos\,h-1}{h} = 0$. This fact can be verified by a method similar to that used in example III.3.37.

III.4.13. The derivative of the exponential function.

It was mentioned in paragraph III.3.3 that of all exponential functions a^x (where $a>0$), the most often used is the function e^x. The reason is the following: The number e was chosen so that for $a = \mathrm{e}$, the tangent to the graph of function a^x at the point $x = 0$ has a slope equal to one. Hence the derivative of the function e^x at the point $x = 0$ is also equal to one, i.e. $\lim\limits_{h\to 0}\dfrac{\mathrm{e}^h-1}{h} = 1$. This has an important consequence for the derivative of the function e^x at an arbitrary point x:

$$(\mathrm{e}^x)' \;=\; \lim_{h\to 0}\frac{\mathrm{e}^{x+h}-\mathrm{e}^x}{h} \;=\; \mathrm{e}^x\lim_{h\to 0}\frac{\mathrm{e}^h-1}{h} \;=\; \mathrm{e}^x\cdot 1 \;=\; \mathrm{e}^x.$$

Thus, the function e^x has the derivative which is equal to the function itself. It can be shown that except for the functions of the type $c \cdot \mathrm{e}^x$ (where c is a constant), there exist no other functions with this property.

III.4.14. Theorem (on the derivative of a composite function).

Suppose that the function g is differentiable at point x and function f is differentiable at point $g(x)$. Then the composite function $y = f(g(x))$ is differentiable at point x and its derivative is

$$y'(x) = f'(g(x)) \cdot g'(x).$$

The above formula is often called the *Chain Rule*. The reason is that it can also be written in the form

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}f}{\mathrm{d}g} \cdot \frac{\mathrm{d}g}{dx}.$$

III.4.15. Example.

Evaluate the derivative of the function $h(x) = (\sin x)^2$. The function h is a composition of two functions: inside $g(x) = \sin x$ and outside $f(y) = y^2$. Thus, $f'(y) = 2y$, i.e. $f'(g(x)) = 2\,g(x) = 2\,\sin x$. Further, one has $g'(x) = \cos x$. By Theorem III.4.14, we get $h'(x) = f'(g(x)) \cdot g'(x) = 2\,\sin x\,\cos x$.

III.4.16. Theorem (on the derivative of an inverse function).

Suppose that there exists an inverse function f_{-1} to function f. If $y = f_{-1}(x)$ and if f has a non–zero derivative at point y, then the inverse function f_{-1} has a derivative at point x. This derivative can be expressed by the formula

$$f'_{-1}(x) = \frac{1}{f'(y)} = \frac{1}{f'(f_{-1}(x))}.$$

III.4.17. Derivatives of further elementary functions.

Applying Theorems III.4.14, III.4.16 and already known formulas for derivatives of the functions $\sin x$, $\cos x$, $\tan x$, $\operatorname{cotan} x$ and e^x, we can derive formulas for derivatives of some further elementary functions:

a) $(\arcsin x)' = \dfrac{1}{\sqrt{1-x^2}}$

b) $(\arccos x)' = -\dfrac{1}{\sqrt{1-x^2}}$

c) $(\arctan x)' = \dfrac{1}{1+x^2}$

d) $(\operatorname{arccotan} x)' = -\dfrac{1}{1+x^2}$

e) $(\ln x)' = \dfrac{1}{x}$

f) $(a^x)' = a^x \cdot \ln a \quad (a > 0)$

g) $(\log_a x)' = \dfrac{1}{x \cdot \ln a} \quad (a > 0, a \neq 1)$

Formulas a) and b) are valid on the interval $(-1; 1)$, c), d) and f) are valid on $(-\infty; +\infty)$ and e) and g) hold on $(0; +\infty)$.

III.4.18. Remark.

If function f has a derivative at point x and if $f(x) > 0$, then the derivative of the function $\ln f$ at point x can be evaluated by means of the Chain Rule (paragraph III.4.14):

$$(\ln f(x))' = \frac{1}{f(x)} \cdot f'(x) = \frac{f'(x)}{f(x)}.$$

The expression $\dfrac{f'}{f}$ is a so called *logarithmic derivative of function f*.

III.4.19. Example.

Evaluate the derivative of the composite function $h(x) = (x^2 + 7x - 1)^{\sin x}$. The function h can be expressed by means of the exponential function and the logarithmic function:

$$h(x) = \exp\left(\ln(x^2 + 7x - 1)^{\sin x}\right) = \exp\left(\sin x \cdot \ln(x^2 + 7x - 1)\right).$$

This function is defined only for those x where $x^2 + 7x - 1 > 0$. (Otherwise the expression $\ln(x^2 + 7x - 1)$ has no sense.) By solution of the inequality $x^2 + 7x - 1 > 0$, we obtain: $x \in (-\infty; x_1)$ or $x \in (x_2; +\infty)$, where $x_1 = \dfrac{-7 - \sqrt{53}}{2}$ and $x_2 = \dfrac{-7 + \sqrt{53}}{2}$. For $x \in (-\infty; x_1)$ or $x \in (x_2; +\infty)$, we have:

$$\begin{aligned}
h'(x) &= \Big(\exp\left(\sin x \cdot \ln(x^2 + 7x - 1)\right)\Big)' = \\
&= \exp'\left(\sin x \cdot \ln(x^2 + 7x - 1)\right) \cdot \left(\sin x \cdot \ln(x^2 + 7x - 1)\right)' = \\
&= \exp\left(\sin x \cdot \ln(x^2 + 7x - 1)\right) \cdot \left(\cos x \cdot \ln(x^2 + 7x - 1) + \sin x \cdot \frac{2x + 7}{x^2 + 7x - 1}\right) = \\
&= (x^2 + 7x - 1)^{\sin x} \cdot \left(\cos x \cdot \ln(x^2 + 7x - 1) + \sin x \cdot \frac{2x + 7}{x^2 + 7x - 1}\right).
\end{aligned}$$

III.4.20. Improper derivative.

If the limit (III.4.1) exists, but its value is infinite, then we say that function f has at point x_0 an *improper derivative*.

For example, the function $f(x) = \operatorname{sgn} x$ (having function values $+1$ for $x > 0$, 0 for $x = 0$ and -1 for $x < 0$) has the improper derivative $+\infty$ at the point $x_0 = 0$. Draw the graph of the function $\operatorname{sgn} x$ and verify it for yourself by evaluating the limit (III.4.1). It is seen from this example that a function can have an improper derivative at some point and it need not be continuous at this point.

If function f is continuous at point x_0 and has an improper derivative at this point, then the tangent to the graph of f at point x_0 is a straight line perpendicular to the x–axis. This straight line has the equation $x = x_0$.

The readers should be aware that it is necessary to distinguish between the notions the derivative (the finite value of the limit (III.4.1)) and the improper derivative (the infinite value of the limit (III.4.1)). The notion "derivative" (without a more detailed specification) will in the following refer only to the proper (i.e. finite) derivative.

III.4.21. Differential of a function.

Suppose that function f is differentiable at point x_0. Then the tangent line to the graph of f at the point $[x_0; f(x_0)]$ has the equation $y = f(x_0) + f'(x_0) \cdot (x - x_0)$. (See paragraph III.4.5.) The linear function, defined by this equation, represents the best linear approximation of function f in the neighborhood of x_0. The function values of f at points x from a small neighborhood of x_0 can be approximately calculated:

$$f(x) \doteq f(x_0) + f'(x_0) \cdot (x - x_0).$$

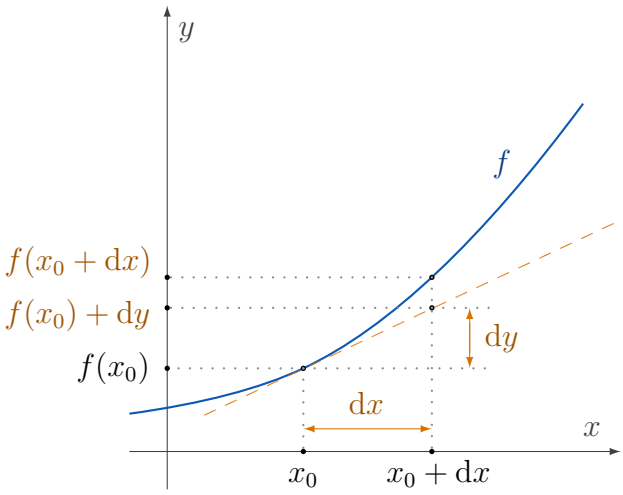


Fig. 27

(See Fig. 27.) Writing $x = x_0 + \Delta x$, we obtain: $f(x + \Delta x) \doteq f(x_0) + f'(x_0) \cdot \Delta x$. The term $f'(x_0) \cdot \Delta x$ is a so called *differential* of function f at the point x_0. You can see that if x_0 is fixed then the differential depends on Δx. The differential is denoted by $\mathrm{d}y$ or $\mathrm{d}f$.

One often writes only x instead of x_0 and $\mathrm{d}x$ instead of Δx. Using this denotation, we have:

$$f(x + \mathrm{d}x) \doteq f(x) + \mathrm{d}y \qquad \text{where } \mathrm{d}y = f'(x) \cdot \mathrm{d}x.$$

The differential is more interesting and more important in the theory of functions of more variables.

III.4.22. Derivatives of higher orders.

The *derivative of the second order* of function f (we denote it f'') is the derivative of the function f'. By analogy, the *derivative of the third order* of function f (we denote it f''') is the derivative of the function f'', etc.

Following this notation, the derivative of the n–th order of function f is denoted by $f^{(n)}$. $f^{(0)}$ denotes the function f itself.

Domains of function f and its derivatives satisfy the inclusions: $D(f) \supseteq D(f') \supseteq D(f'') \supseteq D(f''') \supseteq \dots$

The derivative of the second order of the function $y = f(x)$ is also often written down in a way which is consistent with the denotation introduced in paragraph III.4.4:

$$\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}, \qquad \frac{\mathrm{d}^2}{\mathrm{d}x^2} f, \qquad y'', \qquad \frac{\mathrm{d}^2 y}{\mathrm{d}x^2}.$$

Other higher order derivatives can also be denoted and written down analogously.

III.4.23. Leibniz' formula.

The derivative of the n–th order of the product $f \cdot g$ on the intersection of $D(f^{(n)})$ and $D(g^{(n)})$ can be expressed by the formula:

$$\big(f \cdot g\big)^{(n)} = f^{(n)} \cdot g + \binom{n}{1} \cdot f^{(n-1)} \cdot g' + \binom{n}{2} \cdot f^{(n-2)} \cdot g'' + \dots + \binom{n}{n} \cdot f \cdot g^{(n)}.$$

III.4.24. Problems.

Find derivatives of the following functions. (Don't forget to specify their domains.)

a) $\sqrt{x} \cdot \cos x$ b) $\dfrac{x^2-1}{x^2+1}$ c) $\dfrac{1+\ln x}{x}$

d) $(x^2-1)(x^3+2x-1)$ e) $\arcsin\sqrt{x}$ f) $\sin(1-\cos x)$

g) x^x h) $\ln(\ln x)$ i) $\ln(\arctan x)$

j) 2^{2x+1} k) $(3x-1)^{2001}$ l) $-x \cdot \operatorname{cotan} x + \ln(\sin x)$

m) $\arcsin^2 x$ n) $(\ln x)^x$ o) $x \cdot \sin x^2$

p) $\sqrt{2x+1}$

Results:

a) $\dfrac{1}{2\sqrt{x}}\cos x - \sqrt{x}\cdot\sin x$ (for $x>0$)

b) $\dfrac{2x(x^2+1)-(x^2-1)2x}{(x^2+1)^2}$ (for $x\in\mathbb{R}$)

c) $-\dfrac{\ln x}{x^2}$ (for $x>0$)

d) $2x(x^2+2x-1)+(x^2-1)(3x^2+2)$ (for $x\in\mathbb{R}$)

e) $\dfrac{1}{2\sqrt{x}}\cdot\dfrac{1}{\sqrt{1-x}}$ (for $x\in(0;1)$)

f) $\cos(1-\cos x)\cdot\sin x$ (for $x\in\mathbb{R}$)

g) $x^x\cdot(\ln x+1)$ (for $x>0$)

h) $\dfrac{1}{\ln x}\cdot\dfrac{1}{x}$ (for $x>1$)

i) $\dfrac{1}{\arctan x}\cdot\dfrac{1}{x^2+1}$ (for $x>0$)

j) $2^{(2x+1)}\cdot 2\cdot\ln 2$ (for $x\in\mathbb{R}$)

k) $2001\cdot(3x-1)^{2000}\cdot 3$ (for $x\in\mathbb{R}$)

l) $\dfrac{x}{\sin^2 x}$ (for $\sin x>0$)

m) $\dfrac{2\arcsin x}{\sqrt{1-x^2}}$ (for $x\in(-1;1)$)

n) $(\ln x)^x\cdot\left(\ln(\ln x)+\dfrac{1}{\ln x}\right)$ (for $x>1$)

o) $\sin x^2+2x^2\cdot\cos x^2$ (for $x\in\mathbb{R}$)

p) $\dfrac{1}{\sqrt{2x+1}}$ (for $x>-\dfrac{1}{2}$).

III.4.25. Remark.

Mastering the differentiation of functions (i.e. computation of derivatives of functions) is one of the most important tasks in your first term of study. It is therefore necessary to work individually on a large number of examples on this theme. A lot of appropriate examples can be found e.g. in [NK].

III.5. Applications of derivatives, behavior of a function

III.5.1. Mean Value Theorem (Lagrange's theorem).

Let function f be continuous on the closed interval $\langle a;b\rangle$ and let it be differentiable on the open interval $(a;b)$. Then there exists a point $\xi \in (a;b)$ such that

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

III.5.2. Remark.

The geometric sense of the Mean Value Theorem is obvious from Fig. 28.

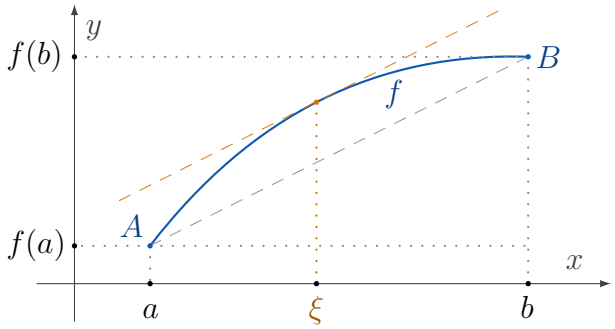


Fig. 28

III.5.3. The interior of an interval.

If I is an interval in $\mathbb{R}$, then the set of all interior points of this interval is called the *interior* of I and it is denoted by I°. Thus if for example $I = (a;b\rangle$ then $I^\circ = (a;b)$. If $I = \langle 0;1\rangle$, then $I^\circ = (0;1)$, etc.

III.5.4. Theorem.

Let f be a continuous function on interval I. Then the following implications hold:

a) $f'(x) > 0$ *for all* $x \in I^\circ \implies$ f *is increasing on interval* I.
b) $f'(x) \geq 0$ *for all* $x \in I^\circ \implies$ f *is non–decreasing on interval* I.
c) $f'(x) < 0$ *for all* $x \in I^\circ \implies$ f *is decreasing on interval* I.
d) $f'(x) \leq 0$ *for all* $x \in I^\circ \implies$ f *is non–increasing on interval* I.
e) $f'(x) = 0$ *for all* $x \in I^\circ \implies$ f *is a constant function on interval* I.

P r o o f :
We present only the proof of implication a) and all other cases are analogous. Let x_1 and x_2 be two arbitrary points from I such that $x_1 < x_2$. It follows from the Mean Value Theorem that there exists a point $\xi \in (x_1;x_2)$ such that $f(x_2) - f(x_1) = f'(\xi)\cdot(x_2 - x_1)$. Since $(x_2 - x_1) > 0$ and $f'(\xi) > 0$ (the last inequality follows from the assumption of item a)), we get: $f(x_2) - f(x_1) > 0$, or $f(x_1) < f(x_2)$. This means that function f is increasing on the interval I. (Compare this with item a) from paragraph III.2.12.)

III.5.5. Example.

Investigate intervals of monotonicity of the function $f(x) = x^3 - 3x$.

Solution: The intervals of monotonicity are the intervals where the function is increasing or decreasing, respectively non–decreasing or non–increasing.

The function f is defined and continuous on the interval $(-\infty; +\infty)$. Its derivative is $f'(x) = 3x^2 - 3$ (also for $x \in (-\infty; +\infty)$). Solving the inequalities $f'(x) > 0$ and $f'(x) < 0$, we find:

$$3x^2 - 3 > 0 \iff x^2 - 1 > 0 \iff x \in (-\infty; -1) \text{ or } x \in (1; +\infty),$$
$$3x^2 - 3 < 0 \iff x^2 - 1 < 0 \iff x \in (-1; 1).$$

This means that the function f is increasing on each of the intervals $(-\infty; -1)$ and $(1; +\infty)$ and decreasing on the interval $(-1; 1)$. (See Theorem III.5.4.) This is, however, not all: as the function f is continuous "up to the end points" -1 and 1 on each of these intervals, i.e. it is continuous on the intervals $(-\infty; -1\rangle$, $\langle 1; +\infty)$ and $\langle -1; 1\rangle$, the statements on the monotonicity can be extended to the intervals, containing the end points. Thus, we may conclude:

a) Function f is increasing on each of the intervals $(-\infty; -1\rangle$ and $\langle 1; +\infty)$.
b) Function f is decreasing on the interval $\langle -1; 1\rangle$. (See Fig. 29.)

III.5.6. Remark.

Students often make the following mistake: They forget that Theorem III.5.4 holds on an interval and they find out that f is continuous for example on two intervals $(-\infty; -1\rangle$ and $\langle 1; +\infty)$ and moreover, $f' > 0$ in the interior of both these intervals. Then they conclude that function f is increasing on the union $(-\infty; -1\rangle \cup \langle 1; +\infty)$. However, this need not be true ! – see e.g. the function $f(x) = x^3 - 3x$. (See Fig. 29.) It follows from Theorem III.5.4 that f is increasing on each of the intervals $(-\infty; -1\rangle$ and $\langle 1; +\infty)$ only. There is no reason for f to be continuous on the union $(-\infty; -1\rangle \cup \langle 1; +\infty)$!

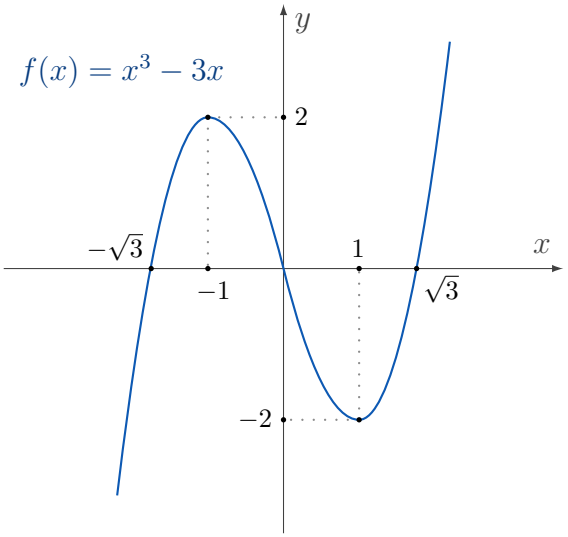


Fig. 29

III.5.7. Example.

Investigate intervals of monotonicity of the function $f(x) = 1 - x^2 + \frac{1}{2}x^4$.

Solution: The function f is defined and continuous on the interval $(-\infty; +\infty)$. Its derivative is $f'(x) = -2x + 2x^3$ (also for $x \in (-\infty; +\infty)$). Solving the inequalities $f'(x) > 0$ and $f'(x) < 0$, we obtain:

$$-2x + 2x^3 > 0 \iff x\,(-1 + x^2) > 0 \iff x \in (-1; 0) \text{ or } x \in (1; +\infty),$$
$$-2x + 2x^3 < 0 \iff x\,(-1 + x^2) < 0 \iff x \in (-\infty; -1) \text{ or } x \in (0; 1).$$

Moreover, f is continuous "up to the end points" on each of these intervals, i.e. it is continuous on the intervals $\langle -1; 0\rangle$, $\langle 1; +\infty)$ and $(-\infty; -1\rangle$, $\langle 0; 1\rangle$. Thus we may conclude:

a) Function f is increasing on the intervals $\langle -1; 0\rangle$ and $\langle 1; +\infty)$.
b) Function f is decreasing on the intervals $(-\infty; -1\rangle$ and $\langle 0; 1\rangle$. (See Fig. 30.)

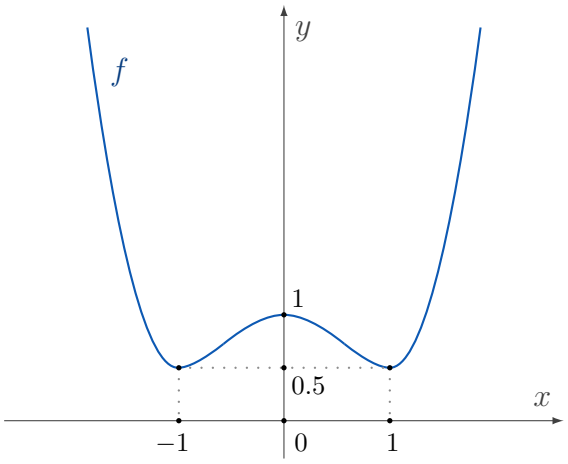


Fig. 30

III.5.8. Local extreme values of a function.

Suppose that function f is defined in some interval $(a; b)$ containing point x_0. We say that f has a *local maximum* (respectively a *local minimum*) at point x_0 if there exists a reduced neighborhood $R(x_0)$ such that $\forall\, x \in R(x_0) \;:\; f(x) \le f(x_0)$ (respectively $\forall\, x \in R(x_0) \;:\; f(x) \ge f(x_0)$).

If we replace the inequalities "$\le$" and "$\ge$" by the sharp inequalities "$<$" and "$>$", we get the definitions of a so called *strict local maximum*, respectively a *strict local minimum*.

Local maximum and local minimum are called *local extreme values*. Strict local maximum and strict local minimum are called *strict local extreme values*.

Instead of "extreme values", we can also shortly say "extremes".

Notice that the strict local extremes are special cases of the local extremes. Fig. 31 shows a function f, which has strict local maxima at the points x_1, x_3, x_5 and x_7 and strict local minima at x_2, x_4 and x_6.

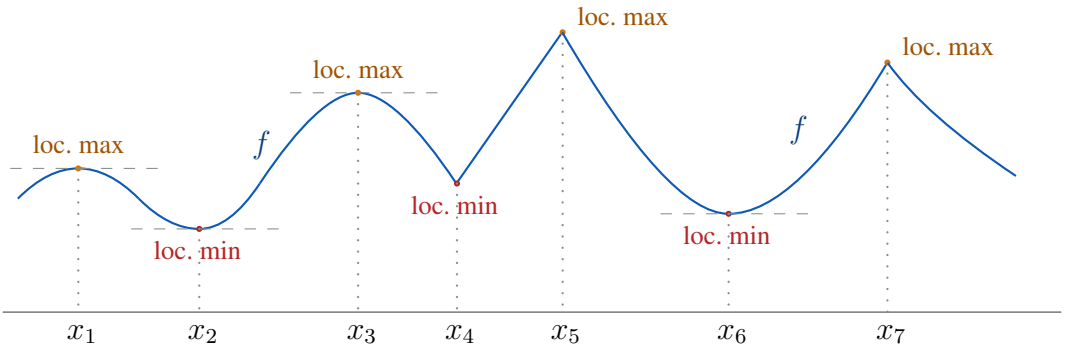


Fig. 31

III.5.9. Remark.

Obviously, strict local extreme values are special cases of local extreme values.

To distinguish between extreme values of function f on its whole domain (defined in paragraph III.2.9) and local extreme values, we often call a maximum (respectively a minimum) of function f (in $D(f)$) an *absolute maximum* (respectively an *absolute minimum*), or sometimes a *global maximum* (respectively a *global minimum*).

The next theorem plays a fundamental role in the investigation of the local extreme values.

III.5.10. Theorem.

If function f has a local extreme value at point x_0 and if f is differentiable at this point then $f'(x_0) = 0$.

P r o o f :
We show the proof by contradiction. Suppose that f has a local extreme value at point x_0, and that the derivative $f'(x_0)$ exists, but it is not equal to zero. Without loss of generality, we can assume that, for example, $f'(x_0) > 0$. We are going to show that this is not possible.

It follows from the inequality $f'(x_0) > 0$ that there exists $a > 0$ such that

$$\lim_{h\to 0} \frac{f(x_0+h) - f(x_0)}{h} = a > 0.$$

Thus, for each sequence $\{h_n\}$ in $D(f)$ such that $h_n \to 0$, it holds:

$$\lim_{n\to+\infty} \frac{f(x_0+h_n) - f(x_0)}{h_n} = a > 0.$$

This means (by definition III.1.4) that to every $\mathcal{U}(a)$ there exists $n_0 \in \mathbb{N}$ such that for $n \in \mathbb{N}$, satisfying the inequality $n \geq n_0$, it holds: $\dfrac{f(x_0+h_n) - f(x_0)}{h_n} \in \mathcal{U}(a)$. If we choose $\mathcal{U}(a) = (0; 2a)$ and put $h_n = \dfrac{1}{n}$, we get:

$$0 < \frac{f\left(x_0 + \frac{1}{n}\right) - f(x_0)}{\frac{1}{n}} < 2a.$$

Using only the first part of this inequality (i.e. $0 < \dots$), we can see that $f\left(x_0 + \dfrac{1}{n}\right) > f(x_0)$ for $n \geq n_0$. However, this means that function f cannot have a local maximum at point x_0. Similarly, by the choice $h_n = -\dfrac{1}{n}$, we can show that function f cannot have a local minimum at point x_0, either. This is the desired contradiction.

This proof is quite instructive. Sketch a graph of a function which has a positive derivative at some point x_0 and follow all steps of the proof in your figure!

III.5.11. Remark – necessary vs. sufficient condition.

Remember that the condition $f'(x_0) = 0$ (if the derivative $f'(x_0)$ exists) is only a necessary condition for the existence of a local extreme value of function f at point x_0, but it is not a sufficient condition. This can be illustrated for instance by a simple example: Function $f(x) = x^3$ has a zero derivative at the point $x_0 = 0$. Nevertheless, it does not have a local extreme value at this point!

III.5.12. More about the local extreme values.

The only points where function f can have a local extreme value in an interval I are
1) interior points of interval I where f' is equal to zero (see Fig. 31, the points x_1, x_2, x_3 and x_6),
2) interior points of interval I where f' does not exist (see Fig. 31, the points x_4, x_5, x_7).

The word "can" is stressed because function f can, however does not necessarily need, have its local extreme values at the mentioned points. This follows from Theorem III.5.10 and Remark III.5.11.

How to find the local extreme values of function f.

a) We find all points where f' equals zero or f' does not exist. These are the only points where function f can take on a local extreme value.
b) We need to check whether function f really has a local extreme value in these points and to specify whether it is a local maximum or a local minimum value. We can apply one of the following procedures:
   b1) Denote by x_0 one of the points from item a). Assume that f is continuous at the point x_0. (This follows e.g. from the existence of the derivative – see Theorem III.4.8.) If we find out, for instance, that f is increasing in some left neighborhood of x_0 and decreasing in some right neighborhood of x_0 then f obviously has a strict local maximum at the point x_0. (Sketch a picture.) On the other hand, if f is decreasing in some left neighborhood of x_0 and increasing in some right neighborhood of x_0 then f has a strict local minimum at x_0.
   b2) To recognize whether function f has a local extreme value at point x_0 and what is its type, one can also apply Theorem III.5.21. (You will see it later.) Theorem III.5.21 uses the sign of the second derivative of f at the point x_0.

III.5.13. Example.

Find local extreme values of the function $f(x) = x^2\,\mathrm{e}^x$.

The domain of the function f is the interval $(-\infty;+\infty)$ and f is differentiable at each point of this interval. Thus, if f has a local extremum at some point x_0, then it must be a local extremum of type 1) from Remark III.5.12. So it must be $f'(x_0) = 0$. f' can easily be expressed: $f'(x) = 2x\,\mathrm{e}^x + x^2\,\mathrm{e}^x = (2+x)\,x\,\mathrm{e}^x$. We put it equal to zero and we get the equation $(2+x)\,x\,\mathrm{e}^x = 0$. This equation has two roots: $x_1 = -2$, $x_2 = 0$. This means that the points -2 and 0 are the only points where f can have a local extremum.

We can apply e.g. the procedure from item b1) in the previous paragraph to check whether the function f really has a local extremum at some of the points -2, 0 and moreover, what kind of local extremum it is. Solving the inequality $f'(x) = (2+x)\,x\,\mathrm{e}^x > 0$, we obtain: $x \in (-\infty;-2)$ or $x \in (0;+\infty)$ and similarly, the inequality $f'(x) = (2+x)\,x\,\mathrm{e}^x < 0$ is satisfied for $x \in (-2;0)$. Thus, the derivative f' is positive on the intervals $(\infty;-2)$ and $(0;+\infty)$. Due to Theorem III.5.4, function f is increasing on each of these intervals. f' is negative on the interval $(-2;0)$, so f is decreasing here. Since f is continuous at the point -2, increasing on its left neighborhood and decreasing on its right neighborhood, it has a strict

local maximum at the point -2. Similarly, f is continuous at the point 0, decreasing on its left neighborhood and increasing on its right neighborhood, so it has a strict local minimum at the point 0.

III.5.14. More about the absolute (= global) extreme values.

The only points where function f can have an absolute extreme value in an interval I are

1) interior points of interval I where f' is equal to zero (see Fig. 32a),
2) interior points of interval I where f' does not exist (see Fig. 32b),
3) endpoints of interval I (if interval I is not open – see Fig. 32c).

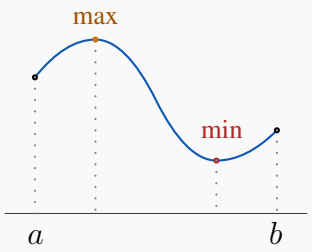


Fig. 32a

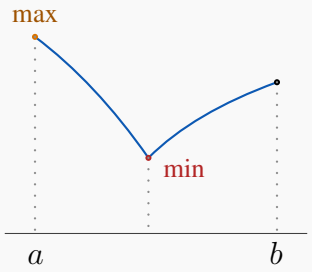


Fig. 32b

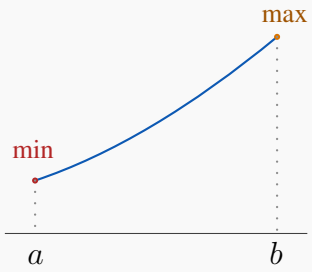


Fig. 32c

How to find the absolute extreme values of function f on interval I.

a) We find all interior points of interval I where f' equals zero or f' does not exist. We add the endpoints of I (if the interval I is not open). These are the only points where f can have an absolute extreme.

b) If we are sure that the absolute extreme values $\max_I f$ and $\min_I f$ exist then we calculate the values of f at the points from item a). The greatest one is $\max_I f$ and the least one is $\min_I f$.
The knowledge about the existence of the absolute extreme values of function f in interval I follows e.g. from Theorem III.3.29. (It says that if interval I is bounded and closed and function f is continuous in I then the absolute extreme values of f in I exist.)
If the assumptions of Theorem III.3.29 are not fulfilled then one has to apply a finer analysis that can vary case from case in order to check whether function f has the absolute extreme values in interval I. It can happen that the absolute extreme values (or at least one of them) do not exist. (See also Remark III.2.11.)

III.5.15. Example.

$f(x) = x^2 + \dfrac{16}{x} - 16,\ I = \langle 1; 4\rangle$

Find absolute extreme values of the function f in the interval I (if they exist).

Solution: $\langle 1; 4\rangle$ is a bounded and closed interval. The function f is continuous in this interval. (It is obvious that f is continuous in $(-\infty; 0) \cup (0; +\infty)$.) Thus, by Theorem III.3.29, the absolute extreme values $\max_{\langle 1;4\rangle} f$ and $\min_{\langle 1;4\rangle} f$ exist.

Function f is differentiable at all points $x \in \langle 1; 4\rangle$ and its derivative is

$$f'(x) = 2x - \frac{16}{x^2}.$$

The equation $f'(x) = 0$ has a unique root: $x = 2$. Adding the endpoints of the interval $\langle 1; 4\rangle$, we obtain the set: $x_1 = 1,\ x_2 = 2,\ x_3 = 4$. These are the only points where function f can have the absolute extreme values in the interval $\langle 1; 4\rangle$. The function values at these points are: $f(x_1) = f(1) = 1,\ f(x_2) = f(2) = -4,\ f(x_3) = f(4) = 4$. The least of them is -4 and the greatest of them is 4. Hence $\max_{\langle 1;4\rangle} f = f(2) = -4$ and $\min_{\langle 1;4\rangle} f = f(4) = 4$.

III.5.16. Functions concave up and concave down.

Function f is called *strictly concave up (strictly convex)* on set M if $M \subset D(f)$ and if for each three points $x_1,\ x_2,\ x_3 \in M$ such that $x_1 < x_2 < x_3$, it holds that: The point $Q_2 = [\,x_2; f(x_2)\,]$ is below the straight line $Q_1 Q_3$, where $Q_1 = [\,x_1; f(x_1)\,]$ and $Q_3 = [\,x_3; f(x_3)\,]$.

If we replace the word "below" by the word above in this definition, we gain a definition of a function which is *strictly concave down (strictly concave)* on the set M.

By analogy, replacing the word "below" by the words below or on (respectively above or on), we get the definitions of a function *concave up (convex)* on the set M (respectively *concave down (concave)* on the set M).

You can see examples of functions that are strictly concave up and strictly concave down in Fig. 33a and Fig. 33b.

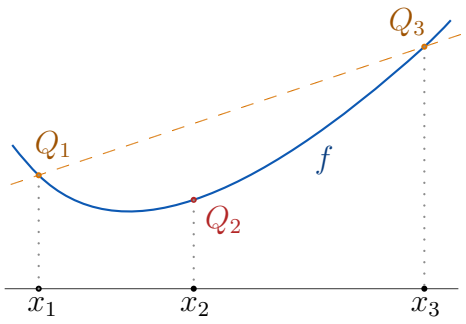


Fig. 33a

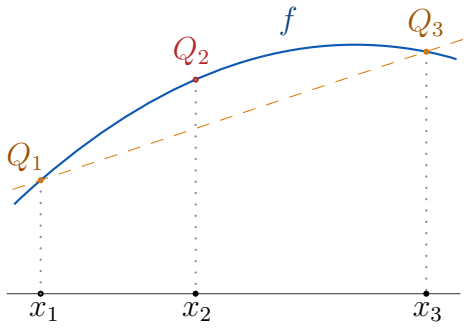


Fig. 33b

III.5.17. Remark.

A function strictly concave up is a special case of a function concave up and a function strictly concave down is a special case of a function concave down.

III.5.18. Remark.

The condition saying that $Q_2 = [\,x_2; f(x_2)\,]$ finds itself below the straight line $Q_1 Q_3$, where $Q_1 = [\,x_1; f(x_1)\,]$ and $Q_3 = [\,x_3; f(x_3)\,]$, can be by computation expressed by the inequality

$$f(x_2) \ < \ f(x_1) + \frac{f(x_3) - f(x_1)}{x_3 - x_1} \cdot (x_2 - x_1).$$

III.5.19. Theorem – Necessary conditions.

Let function f be continuous on interval I. Then the following implications hold:

a) $f''(x) > 0$ *for all* $x \in I^\circ \implies f$ *is strictly concave up on interval* I.
b) $f''(x) \geq 0$ *for all* $x \in I^\circ \implies f$ *is concave up on interval* I.
c) $f''(x) < 0$ *for all* $x \in I^\circ \implies f$ *is strictly concave down on interval* I.
d) $f''(x) \leq 0$ *for all* $x \in I^\circ \implies f$ *is concave down on interval* I.
e) $f''(x) = 0$ *for all* $x \in I^\circ \implies f$ *is a linear function on interval* I.

III.5.20. Remark.

We omit the proof of Theorem III.5.19. Nevertheless, in order to help toward a better understanding of the theorem, we sketch at least its main idea. Let us for instance deal with item a). f'' coincides with the first derivative of the function f'. So if $f'' > 0$ on I°, f' is increasing on I°. This means that if we move in interval I from left to right, the tangent to the graph of f changes its direction – it slants in such a way that its slope increases. However, this is possible only in the case when function f is concave up on interval I. (Think this over by means of Fig. 33a.)

Let us now return to the question how to specify the type of a local extreme value – see also example III.5.13.

III.5.21. Theorem – Sufficient condition.

If $f'(x_0) = 0$ *and* $f''(x_0) > 0$, *then function* f *has a strict local minimum at point* x_0.
If $f'(x_0) = 0$ *and* $f''(x_0) < 0$, *then function* f *has a strict local maximum at point* x_0.

III.5.22. Remark.

The proof of Theorem III.5.19 is also omitted. However, the following consideration can contribute to its understanding: Suppose that $f'(x_0) = 0$, $f''(x_0) > 0$ and to exclude complicated cases, suppose in addition that the second derivative f'' is continuous at point x_0. It follows from the inequality $f''(x_0) > 0$ that there exists a neighborhood $\mathcal{U}(x_0)$ such that $f''(x) > 0$ for all $x \in \mathcal{U}(x_0)$. (We have used Theorem III.3.30.) This means (by Theorem III.5.19) that function f is strictly concave up in the interval $\mathcal{U}(x_0)$. This information together with the equality $f'(x_0) = 0$ leads to the conclusion that f has a strict local minimum at point x_0. (Sketch a picture.)

III.5.23. Example.

We find local extreme values of the function $f(x) = 2x^3 + 3x^2 - 36x + 4$. The domain of f is the interval $(-\infty; +\infty)$ and the function f is differentiable at each point of this interval. Hence it can have the local extreme values only at those points where the derivative is equal to zero. (See Remark III.5.12.) The derivative of f is: $f'(x) = 6x^2 + 6x - 36$. Solving the quadratic equation $6x^2 + 6x - 36 = 0$, we obtain the points $x_1 = -3$ and $x_2 = 2$. The second derivative of the function f is: $f''(x) = 12x + 6$. Substituting the values of x_1 and x_2 to $f''(x)$, we find out that

a) $f''(x_1) = f''(-3) = 12 \cdot (-3) + 6 = -30 < 0$. Thus, function f has a strict local maximum at the point -3.
b) $f''(x_2) = f''(2) = 12 \cdot 2 + 6 = 30 > 0$. So function f has a strict local minimum at the point 2.

III.5.24. Remark.

If we replace the inequality $f''(x_0) > 0$ by the inequality $f''(x_0) \geq 0$ in the assumptions of Theorem III.5.21, then it is not true that we can also replace a "strict local minimum" by a "local minimum" in the statement of the theorem, and the theorem remains valid. On the contrary, the inequality $f''(x_0) \geq 0$ admits the possibility $f''(x_0) = 0$ and the equalities $f'(x_0) = 0$ and $f''(x_0) = 0$ do not allow us to draw any conclusion about the extreme values of function f at point x_0! Think this over in connection with three simple examples: 1) $f(x) = x^4$, 2) $f(x) = -x^4$, 3) $f(x) = x^3$. Put $x_0 = 0$ in all three examples.

III.5.25. Point of inflection.

Suppose that function f is differentiable at point x_0 (and, consequently, there exists a tangent to the graph of f at the point $[\,x_0; f(x_0)\,]$). The tangent divides the x, y plane into two half–planes. If the tangent passes from one half–plane to the other at the point $[\,x_0; f(x_0)\,]$ then x_0 is called the *point of inflection* or the *inflection point* of function f.

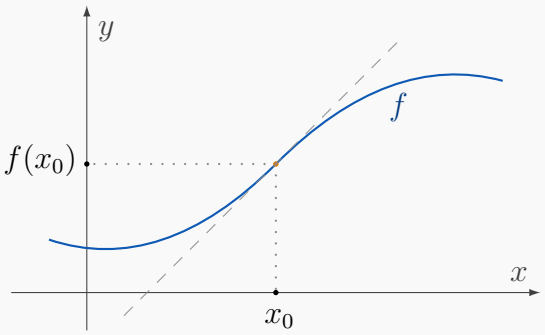


Fig. 34

III.5.26. Example.

0 is a point of inflection of the function $f(x) = x^3 + 1$. Sketch the graph of this function and the tangent to the graph at the point $[\,0; f(0)\,] = [\,0; 1\,]$ for yourself.

III.5.27. Theorem – Necessary condition for inflection point.

If x_0 is an inflection point of function f and if the second derivative $f''(x_0)$ exists, then $f''(x_0) = 0$.

III.5.28. Remark.

If $f''(x_0)$ exists then the condition $f''(x_0) = 0$ is the necessary condition for x_0 to be an inflection point. However, it is not a sufficient condition! This means that it is not possible to overturn Theorem III.5.25 and to assert that the equality $f''(x_0) = 0$ implies that x_0 is the inflection point. It is evident from the example $f(x) = x^4$. This function has the second derivative equal to zero at the point 0, but in spite of this 0 is not a point of inflection of f.

Thus, if we find for a given function f points where f has the second derivative equal to zero, we have only "appropriate candidates" for points of inflection. Then it is necessary to use some other means to find out whether these points are really inflection points. The following theorem. is a useful tool.

III.5.29. Theorem – Sufficient condition for inflection point.

If $f''(x_0) = 0$ and $f'''(x_0) \neq 0$, then x_0 is an inflection point of function f.

III.5.30. Example.

We find inflection points of the function $f(x) = \exp(-x^2)$. Differentiating the function f, we get: $f'(x) = -2x\exp(-x^2)$ and $f''(x) = 2(2x^2-1)\exp(-x^2)$ for all $x \in (-\infty;+\infty)$. Thus, if f has points of inflection, the second derivative of f must be equal to zero at these points (by Theorem III.5.24). Thus we need to solve the equation $f''(x) = 0$, i.e. $2(2x^2-1)\exp(-x^2) = 0$. This equation has two solutions: $x_1 = -\frac{1}{\sqrt{2}}$ and $x_2 = \frac{1}{\sqrt{2}}$. Differentiating the function f once more, we obtain: $f'''(x) = 4(3x-2x^3)\exp(-x^2)$. Substituting here the values of x_1 and x_2, we find out that

a) $f'''(x_1) = f'''\left(-\frac{1}{\sqrt{2}}\right) = -\frac{8}{\sqrt{2}}\cdot\exp(-0.5) \neq 0$, hence the point x_1 is the inflection point of the function f (by Theorem III.5.27).

b) $f'''(x_2) = f'''\left(\frac{1}{\sqrt{2}}\right) = \frac{8}{\sqrt{2}}\cdot\exp(-0.5) \neq 0$, so x_2 is also the inflection point of the function f. (This follows again from Theorem III.5.27.)

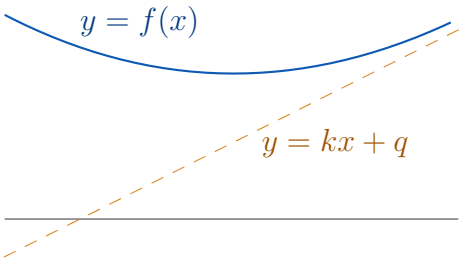


Fig. 35a

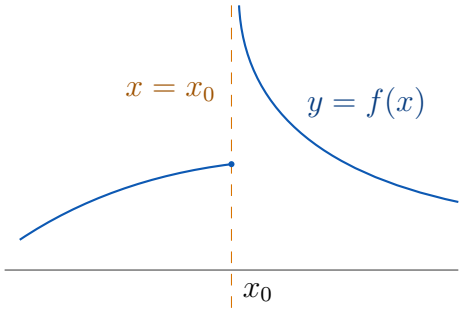


Fig. 35b

III.5.31. Asymptotes to the graph of a function.

The straight line $y = kx+q$ is a so called *slant asymptote* to the graph of function f as $x \to -\infty$ if $\lim_{x\to-\infty}(f(x) - kx - q) = 0$.
Similarly, the straight line $y = kx + q$ is a slant asymptote to the graph of function f as $x \to +\infty$ if $\lim_{x\to+\infty}(f(x) - kx - q) = 0$.
The straight line $x = x_0$ is called a *vertical asymptote* to the graph of function f at point x_0 if at least one of the one–sided limits of f at point x_0 is infinite.

An example of a slant asymptote to the graph of function f as $x \to +\infty$ is seen in Fig. 35a, and an example of a vertical asymptote is shown in Fig. 35b.

III.5.32. Theorem.

The straight line $y = kx+q$ is a slant asymptote of the graph of function f as $x \to +\infty$ if and only if

$$\lim_{x\to+\infty}\frac{f(x)}{x} = k \qquad \text{and} \qquad \lim_{x\to+\infty}(f(x) - kx) = q.$$

(The theorem is also valid in the case when $+\infty$ is everywhere replaced by $-\infty$.)

III.5.33. Behavior of a function.

To investigate the behavior of a function, you can follow this strategy:

a) Specify the domain of f (if it is not already given).
Find out whether the function is even, odd, periodic.
Find intervals of continuity, points of discontinuity and evaluate one–sided limits at end–points of intervals which form the domain of f (possibly also at points of discontinuity of f).
Find points where the graph of f crosses the x–axis, the y–axis, and specify intervals where f is positive, respectively negative.

b) Find the derivative of f. (Don't forget about the domain of the derivative !)
Find maximum intervals where the function is monotonic and specify the type of monotonicity (i.e. whether f rises or falls).
Find local extreme values of f. (To specify their type, you can possibly use the sign of the second derivative.)

c) Find inflection points of f.
Find maximum intervals of concavity of the function f.

d) Find asymptotes of the graph of f.
Sketch the graph of f.

III.5.34. Example.

We investigate the behavior of the function $f(x) = \dfrac{x^3}{4-x^2}$.

a) $D(f) = (-\infty;-2) \cup (-2;2) \cup (2;+\infty)$ (the denominator of the fraction cannot be equal to zero). The function f is odd because for all $x \in D(f)$, it holds: $-x \in D(f)$ and $f(-x) = -f(x)$. Hence the graph of f is symmetric with respect to the origin of the coordinate system. We can therefore study the behavior of f only on the set $\langle 0;2) \cup (2;+\infty)$. Information about its behavior in the set $(-\infty;-2) \cup (-2;0\rangle$ will follow from the mentioned symmetry. The function f is continuous in $D(f)$. (It is a quotient of two continuous functions and the function in the denominator is different from zero in $D(f)$.)
It holds:
$f(0) = 0$,
$$\lim_{x\to 2-} f(x) = \lim_{x\to 2-} \frac{x^3}{4-x^2} = +\infty, \qquad \lim_{x\to 2+} f(x) = \lim_{x\to 2+} \frac{x^3}{4-x^2} = -\infty,$$
$$\lim_{x\to +\infty} f(x) = \lim_{x\to +\infty} \frac{x^3}{4-x^2} = \lim_{x\to +\infty} \frac{x}{(\frac{4}{x^2})-1} = \frac{+\infty}{-1} = -\infty,$$
$$f(x) = 0 \iff \frac{x^3}{4-x^2} = 0 \iff x = 0,$$
$$f(x) > 0 \iff x \in (0;2), \qquad f(x) < 0 \iff x \in (2;+\infty).$$

b) The derivative of f is: $f'(x) = \dfrac{3x^2(4-x^2)-(-2x)x^3}{(4-x^2)^2} = \dfrac{x^2(12-x^2)}{(4-x^2)^2}$,
$D(f') = D(f)$,
$$f'(x) = 0 \iff \frac{x^2(12-x^2)}{(4-x^2)^2} = 0 \iff x = 0 \text{ or } x = 2\sqrt{3},$$
$$f'(x) > 0 \iff x \in (0;2) \cup (2;2\sqrt{3}), \qquad f'(x) < 0 \iff x \in (2\sqrt{3};+\infty).$$
This means that f is increasing on the interval $\langle 0;2)$ and on the interval $(2;2\sqrt{3}\rangle$ and decreasing in the interval $\langle 2\sqrt{3};+\infty)$. There is a strict local maximum at the point $2\sqrt{3}$ and $f(2\sqrt{3}) = -3\sqrt{3}$. Although the derivative of f is equal to zero at the point 0, f has no extreme value at this point. Namely, it is increasing on the interval $\langle 0;2)$ and due

to the fact that it is odd, it is also increasing in the interval $\langle -2; 0\rangle$. Thus, f is increasing on the interval $(-2; 2)$.

c) The second derivative of f is:

$$\begin{aligned} f''(x) &= (f'(x))' = \\ &= \frac{(24x - 4x^3)(4 - x^2)^2 - 2(4 - x^2)(-2x)(12x^2 - x^4)}{(4 - x^2)^4} = \\ &= \frac{8x(12 + x^2)}{(4 - x^2)^3}, \end{aligned}$$

$D(f'') = D(f') = D(f)$,
$f''(x) = 0 \iff \frac{8x(12 + x^2)}{(4 - x^2)^3} = 0 \iff x = 0$,
$f''(x) > 0 \iff x \in (0; 2)$, $\quad f''(x) < 0 \iff x \in (2; +\infty)$.
This implies that the function f is strictly concave up in the interval $\langle 0; 2)$ and strictly concave down on the interval $(2; +\infty)$. 0 is the inflection point – this follows from the strict concavity up of f on $\langle 0; 2)$ and the strict concavity down of f on $(-2; 0\rangle$. (The second is a consequence of the symmetry of f.)

d) Since f has at the points -2 and 2 infinite one–sided limits, the straight lines $x = -2$ and $x = 2$ are vertical asymptotes of f.
Let us now investigate a slant asymptote of f as $x \to +\infty$. We use Theorem III.5.30 and we evaluate limits from this theorem:

$$\lim_{x\to+\infty} \frac{f(x)}{x} = \lim_{x\to+\infty} \frac{x^3}{x(4 - x^2)} = -1 = k,$$

$$\lim_{x\to+\infty} (f(x) - kx) = \lim_{x\to+\infty} \left(\frac{x^3}{4 - x^2} + x\right) = \lim_{x\to+\infty} \frac{4x}{4 - x^2} = 0 = q.$$

Hence the straight line $y = -x$ is a slant asymptote of the graph of f as $x \to +\infty$. Similarly, we can find out that the straight line $y = -x$ is also a slant asymptote of the graph of f as $x \to -\infty$.
You can see the graph of f in Fig. 36.

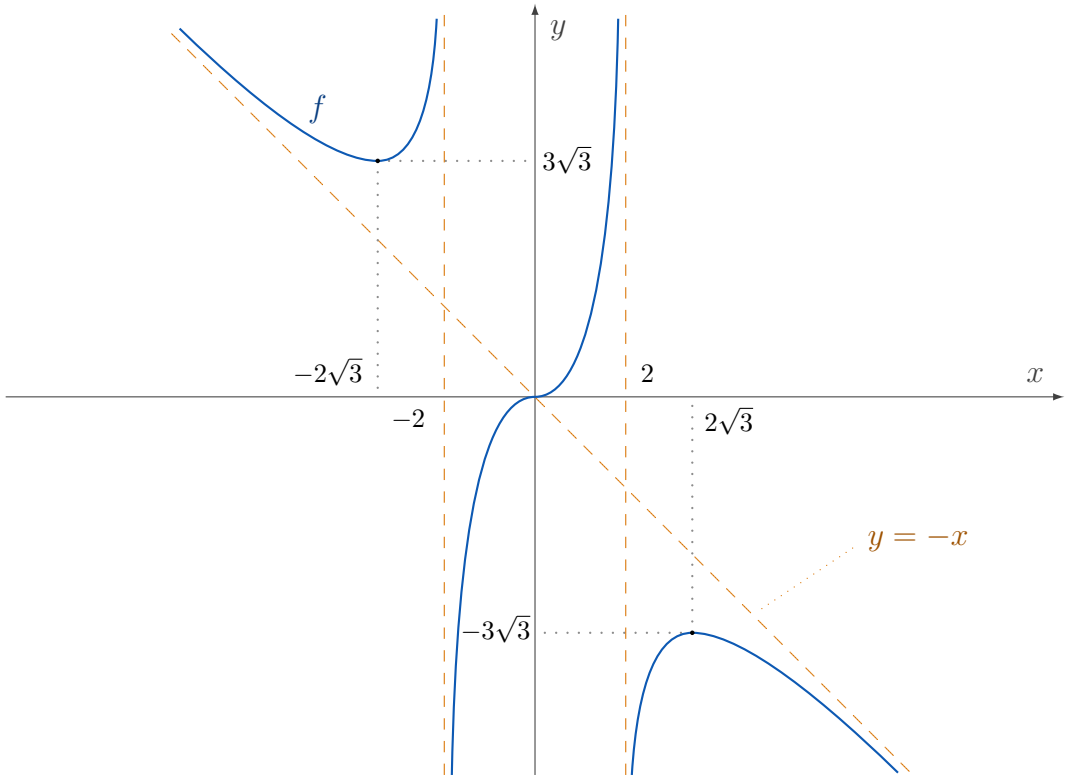


Fig. 36

III.5.35. Problems.

Investigate the behavior of the functions

a) $f(x) = x^{\frac{2}{3}} - x$, b) $f(x) = x \cdot \exp\left(\frac{1}{x}\right)$, c) $f(x) = x \cdot \exp\left(-x^2\right)$,

d) $f(x) = \dfrac{x^2}{x^2-4}$, e) $f(x) = 2x^3 + 3x^2 - 12x + 7$.

The following theorem is a useful aid in computations of limits of quotients of two functions in situations where the computation leads to so called indefinite expressions $\frac{0}{0}$ or $\frac{\pm\infty}{\pm\infty}$.

III.5.36. Theorem (l'Hospital's Rule).

Suppose that $c \in \mathbb{R}^*$ *and the limits* $\lim\limits_{x\to c} f(x)$ *and* $\lim\limits_{x\to c} g(x)$ *are either both equal to zero or are both infinite. Then*

$$\lim_{x\to c} \frac{f(x)}{g(x)} = \lim_{x\to c} \frac{f'(x)}{g'(x)},$$

if the limit on the right– hand side exists.
(The same assertion also holds for the right–hand limits and the left–hand limits.)

III.5.37. Remark.

In other words, the l'Hospital Rule says that if the computation of the limit of $\frac{f(x)}{g(x)}$ (as $x \to c$) leads to the indefinite expression $\frac{0}{0}$ or $\frac{\pm\infty}{\pm\infty}$ and if there exists a limit of $\frac{f'(x)}{g'(x)}$ (as $x \to c$) then the limit $\lim\limits_{x\to c} \frac{f(x)}{g(x)}$ also exists and both limits have the same value.

The assumption concerning the existence of the limit $\lim\limits_{x\to c} \frac{f'(x)}{g'(x)}$ is important, because there are known cases when this limit does not exist, while the limit $\lim\limits_{x\to c} \frac{f(x)}{g(x)}$ does exist. Clearly, its value cannot be found by means of the l'Hospital Rule in such a case.

III.5.38. Example.

Evaluate the limit $\lim\limits_{x\to 0} \left(\frac{\tan x}{x}\right)$. The limit of the quotient $\frac{\tan x}{x}$ cannot be expressed as a quotient of limits, because this leads to the indefinite expression $\frac{0}{0}$. However, the limit of the quotient of the derivatives is:

$$\lim_{x\to 0} \frac{(\tan x)'}{x'} = \lim_{x\to 0} \frac{\frac{1}{\cos^2 x}}{1} = \lim_{x\to 0} \frac{1}{\cos^2 x} = \frac{1}{1^2} = 1.$$

Hence $\lim\limits_{x\to 0} \frac{\tan x}{x}$ also exists and it is also equal to 1.

III.5.39. Example.

We show the application of the l'Hospital Rule once again, but this time more briefly:

$$\lim_{x\to 0} \frac{x-\sin x}{x^3} = \lim_{x\to 0} \frac{1-\cos x}{3x^2} = \lim_{x\to 0} \frac{\sin x}{6x} = \lim_{x\to 0} \frac{\cos x}{6} = \frac{1}{6}.$$

(We have used the l'Hospital Rule three times here.)

III.6. The osculating circle, Taylor's theorem

III.6.1. Motivation.

Suppose that function f is differentiable at point x_0, i.e. it has a derivative $f'(x_0)$ at point x_0. Denote $y_0 = f(x_0)$. You already know that the tangent line $y - y_0 = f'(x_0)\cdot(x - x_0)$ represents the best possible approximation of the graph of function f in a neighborhood of the point $[x_0; y_0]$ of all possible straight lines.

In applications, we are often in a position that we wish to find the best approximation of the graph of function f in the neighborhood of the point $[x_0; y_0]$, however the class of curves we can use consists not of all possible straight lines, but e.g. of all possible circles. This problem is studied in the next paragraph.

III.6.2. The osculating circle, curvature.

Suppose that function f has a second derivative at point x_0. (Then, certainly, $f'(x_0)$ also exists.) Denote for simplicity $y_0 = f(x_0)$, $y_1 = f'(x_0)$ and $y_2 = f''(x_0)$.

Let us solve this problem: Find the circle $(x - x_s)^2 + (y - y_s)^2 = r^2$, which represents the best approximation of the graph of function f in the neighborhood of the point $[x_0; y_0]$. The circle with this property is called the *osculating circle* to the graph of f at the point $[x_0; y_0]$. Its center is called the *center of curvature* and its radius is a so called *radius of curvature*.

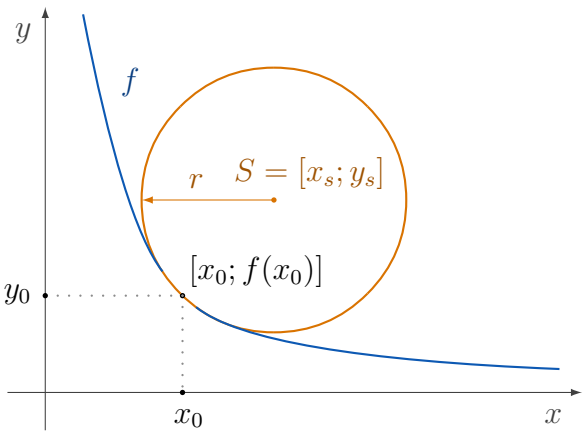


Fig. 37

The circle can be considered to be the graph of a function $y(x)$. The requirement of the best coincidence with the graph of function f in the neighborhood of $[x_0; y_0]$ can be satisfied in such a way that we want function $y(x)$ to have the same value as function f at point x_0 and moreover, to have the same first and second derivative as function f at point x_0. Thus, we obtain the conditions

$$y(x_0) = y_0, \qquad y'(x_0) = y'_0, \qquad y''(x_0) = y''_0\,.$$

Naturally, function $y(x)$ must also satisfy the equation of the circle

$$(x - x_s)^2 + (y(x) - y_s)^2 = r^2. \tag{III.6.1}$$

in the neighborhood of the point $[x_0; y_0]$. Substituting $x = x_0$ and $y(x) = y_0$, we obtain the equation

$$(x_0 - x_s)^2 + (y_0 - y_s)^2 = r^2. \tag{III.6.2}$$

If we differentiate equation (III.6.1) with respect to x and substitute $x = x_0$, $y(x) = y_0$ and $y'(x) = y'_0$, we get

$$2\,(x_0 - x_s) + 2\,(y_0 - y_s)\cdot y'_0 = 0. \tag{III.6.3}$$

Differentiating (III.6.1) two times with respect to x and substituting $x = x_0$, $y(x) = y_0$, $y'(x) = y'_0$ and $y''(x) = y''_0$, we obtain

$$2 + 2\,y_0'^2 + 2\,(y_0 - y_s)\,y''_0 = 0. \tag{III.6.4}$$

Equations (III.6.2), (III.6.3) and (III.6.4) form a system of three equations for three unknowns: x_s, y_s and r. Solving this system, we get the formulas:

$$x_s = x_0 - y'_0\,\frac{1 + y_0'^2}{y''_0}, \qquad y_s = y_0 + \frac{1 + y_0'^2}{y''_0}, \qquad r = \frac{(1 + y_0'^2)^{\frac{3}{2}}}{|y''_0|}.$$

III.6.3. Taylor's theorem – a motivation.

Assume that function f has derivatives up to the order n (inclusively) at point x_0. We look for a polynomial T_n of at most n–th degree which has the form

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

and which is the best approximation of f in the neighborhood of x_0. The requirement of the best approximation is realized in such a way that we want T_n to satisfy:

$$T_n(x_0) = f(x_0), \quad T_n'(x_0) = f'(x_0), \quad T_n''(x_0) = f''(x_0), \quad \dots, \quad T_n^{(n)}(x_0) = f^{(n)}(x_0).$$

These are together $n+1$ conditions. Substituting here the general form of T_n, we can express by a simple calculation $n+1$ coefficients $a_0, a_1, a_2, \dots, a_n$:

$$a_0 = f(c), \quad a_1 = \frac{f'(x_0)}{1!}, \quad a_2 = \frac{f''(x_0)}{2!}, \quad \dots, \quad a_n = \frac{f^{(n)}(x_0)}{n!}.$$

The polynomial T_n with these coefficients, i.e. the polynomial

$$T_n(x) = f(x_0) + \frac{f'(x_0)}{1!}(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \dots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$$

is called *Taylor's polynomial* of the n–th degree of function f at point x_0. In the case that $x_0 = 0$, this polynomial is also often called *Mac Laurin's polynomial* of the n–th degree of function f.

It cannot generally be expected that the equality $f(x) = T_n(x)$ holds quite exactly at the points $x \neq x_0$. Thus, if we use the polynomial $T_n(x)$ instead of $f(x)$, we make a certain error. Let us denote it by $R_{n+1}(x)$. The following theorem provides information that $R_{n+1}(x)$ can be expressed in a certain form. This form can later be used to estimate the magnitude of $R_{n+1}(x)$.

III.6.4. Taylor's Theorem.

Let function f have derivatives up to the order $n+1$ (inclusively) at each point of interval $(a;b)$ and let $x_0 \in (a;b)$. Then to every $x \in (a;b)$ there exists a point ξ between x and x_0 such that

(III.5.1) $$f(x) = T_n(x) + R_{n+1}(x),$$

where

(III.5.2) $$R_{n+1}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \cdot (x-x_0)^{n+1}.$$

III.6.5. Remark.

Formula (III.5.1) is called *Taylor's formula*. It can also be written in the form

$$f(x) = f(x_0) + \frac{f'(x_0)}{1!}(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \dots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + R_{n+1}(x).$$

The term $R_{n+1}(x)$ is called the *remainder* after the n–th term in Taylor's formula. There are more possible ways of expressing the remainder. Formula (III.5.2) presents the so called *Lagrange form of the remainder*.

If $x_0 = 0$ then formula (III.5.1) is also called *Mac Laurin's formula*.

III.6.6. Example.

Mac Laurin's formula for the function $f(x) = \mathrm{e}^x$ has the concrete form

$$\mathrm{e}^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} + R_{n+1}(x), \qquad \text{where} \quad R_{n+1}(x) = \frac{\mathrm{e}^\xi\, x^{n+1}}{(n+1)!}.$$

Mac Laurin's formula for the function sine and for $n = 2m$, where $m \in \mathbb{N}$, has the form

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots + (-1)^{m-1}\frac{x^{2m-1}}{(2m-1)!} + R_{2m+1}(x),$$

where $R_{2m+1}(x) = (-1)^m \dfrac{\cos \xi}{(2m+1)!} \cdot x^{2m+1}$.

Mac Laurin's formula for the function cosine and for $n = 2m+1$, where $m \in \mathbb{N}$, has the form

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots + (-1)^m \frac{x^{2m}}{(2m)!} + R_{2m+2}(x),$$

where $R_{2m+2}(x) = (-1)^{m+1} \dfrac{\cos \xi}{(2m+2)!} \cdot x^{2m+2}$.

III.6.7. Example.

Using Mac Laurin's formula for the function $f(x) = \mathrm{e}^x$, evaluate the Euler number e with a maximum error 10^{-2}. The substitution $x = 1$ into the expression of e^x in example III.6.6 yields:

$$\mathrm{e} = 1 + \frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{1}{n!} + R_{n+1}(1), \qquad \text{where} \quad R_{n+1}(1) = \frac{\mathrm{e}^\xi}{(n+1)!}.$$

The only information we have on ξ says that it is a number from the interval $(0;1)$. We can now ask how large n must be so that $|\,R_{n+1}(1)\,| < 10^{-2}$. Since $0 < \mathrm{e}^\xi < \mathrm{e}^1 < 3$, we have:

$$|\,R_{n+1}(1)\,| = \frac{\mathrm{e}^\xi}{(n+1)!} < \frac{3}{(n+1)!}.$$

By a simple calculation, we find that the choice $n = 5$ is satisfactory, because $\dfrac{3}{(n+1)!} = \dfrac{3}{6!} = \dfrac{3}{720} = \dfrac{1}{240} < 10^{-2}$. This means that the number e can be expressed with an error less than 10^{-2} in the following way:

$$\mathrm{e} \doteq 1 + \frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{1}{5!} = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} = \frac{386}{120}.$$

III.7.* Parametric representation of functions

III.7.1. Motivation.

It often happens in miscellaneous computations that we seek for a function $y = f(x)$, however, we do not find it in an explicit form, and instead we obtain separate expressions of x and y in dependence on a parameter. For instance:

(III.7.1) $$x = t^2, \qquad y = 3t - 1; \qquad t \in (-\infty; +\infty).$$

There arise natural questions: How to recognize whether equations (III.7.1) really define a function y of the variable x? What is its domain, its range, its behavior, etc.? Let us first study these questions on a general level. We therefore suppose that we have general equations

(III.7.2) $$x = \varphi(t), \qquad y = \psi(t); \qquad t \in M$$

instead of the concrete equations (III.7.1). If function φ is one–to–one in set M then it takes on each value x at a unique point $t \in M$. Thus, there exists a unique value $y = \psi(t)$ that corresponds to x. This recipe defines a function which assigns to every $x \in R(\varphi)$ a value $y \in R(\psi)$.

Conversely, if function φ is not one–to–one in set M then it takes on some value x at least at two different points $t_1, t_2 \in M$. There exist two values $y_1 = \psi(t_1)$ and $y_2 = \psi(t_2)$ that correspond to x in this way. The values y_1, y_2 can be the same in extraordinary cases, though this cannot be generally expected. The assignment $x \to y$, defined in this way, need not be a function because it can assign more than one value of y to a single value of x.

III.7.2. Parametric representation of a function.

Assume that functions φ, ψ are defined in set M and function φ is one–to–one. Then equations (III.7.2) define a function f which expresses the dependence of y on x. Its value $f(x)$ can be obtained as follows: As function φ is one–to–one, there exists an inverse function φ_{-1} and the equality $x = \varphi(t)$ is fulfilled if and only if $t = \varphi_{-1}(x)$. Substituting this for t in the equation $y = \psi(t)$, we get: $y = f(x) = \psi(\varphi_{-1}(x))$.

The domain of the function f is a set of all x such that $t \in M$ can be expressed in the form $t = \varphi_{-1}(x)$. Obviously, this is the set $D(\varphi_{-1})$, which is identical with $R(\varphi)$. Thus, $D(f) = R(\varphi)$.

The range of f is the set of those y which can be expressed in the form $y = \psi(t)$ for some $t \in M$. This is the set $R(\psi)$. Thus, we have: $R(f) = R(\psi)$.

Function f is said to be *defined parametrically* by equations (III.7.2). Equations (III.7.2) are called the *parametric representation* of function f.

III.7.3. Remark.

The explicit presentation of the dependence of y on x, i.e. $y = f(x) = \psi(\varphi_{-1}(x))$ mostly cannot be used in practice because the inverse function φ_{-1}, although it exists, cannot be reasonably expressed by a formula. (The example: $\varphi(t) = \mathrm{e}^t + t$; $t \in (-\infty; +\infty)$.)

III.7.4. Remark.

The following consideration can contribute to the understanding of the notion of a function which is defined parametrically: Equations (III.7.2) describe a curve C in the plane $\mathbb{E}_2$. C consists of all points $[x; y] = [\varphi(t); \psi(t)]$ for $t \in M$. If function φ is one–to–one then there cannot exist two different points $[x; y_1]$ and $[x; y_2]$ on C with the same first and different second coordinates. This means that the curve C can be regarded as a graph of a function y of the variable x. That function which is defined parametrically by equations (III.7.2).

Conversely, if function φ is not one–to–one then C need not be a graph of any function y of the variable x. This is the case of the function $x = \varphi(t) = t^2$ from equations (III.7.1). This function φ is not one–to–one in the interval $(-\infty; +\infty)$. The curve C described by equations (III.7.1) is a parabola with the equation $x = \frac{1}{9}(y+1)^2$. (We can easily get this equation if we express t from the second equation in (III.7.1) and use this t in the first equation.) Sketch this parabola for yourself ! The axis of the parabola is the straight line $y = -1$. It is obvious that the parabola is not a graph of a function y of the variable x and consequently, equations (III.7.1) are not a parametric representation of such a function.

The function $x = \varphi(t) = t^2$ is not one–to–one in the interval $(-\infty; +\infty)$. Nevertheless, it is one–to–one in each of the intervals $I_1 = (-\infty; 0\rangle$ and $I_2 = \langle 0; +\infty)$. Hence the equations (III.7.1) define parametrically a function $y = f_1(x)$ if we take $t \in I_1$ and they define parametrically another function $y = f_2(x)$ if t is taken only from I_2. The graph of f_1 is the lower branch of the parabola $x = \frac{1}{9}(y+1)^2$ (corresponding to $y \le -1$) and the graph of f_2 is the upper branch of the parabola, corresponding to $y \ge -1$.

III.7.5. Theorem (on continuity of a parametrically defined function).

Let M be an interval, let functions φ and ψ be continuous on M and let function φ be one–to–one in M. Then the function $y = f(x)$, which is defined parametrically by equations (III.7.2), *is continuous on its domain $D(f)$.*

III.7.6. Remark.

Theorem III.7.5 is an easy consequence of the representation $y = f(x) = \psi(\varphi_{-1}(x))$, the theorem on continuity of a composite function (see paragraph III.3.25) and the theorem on continuity of an inverse function (see paragraph III.3.28).

III.7.7. The derivative of a parametrically defined function.

In addition to all assumptions from paragraph III.7.2, let us suppose that both functions φ and ψ have derivatives $\dot\varphi$ and $\dot\psi$ in the set $M_1 \subset M$ and moreover, $\dot\varphi(t) \neq 0$ for all $t \in M_1$. Applying the theorem on the derivative of a composite function and the theorem on the derivative of an inverse function, we obtain for $x \in R(\varphi|_{M_1})$:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f'(x) = \big(\psi(\varphi_{-1}(x))\big)' = \dot\psi(\varphi_{-1}(x)) \cdot \varphi'_{-1}(x) = \dot\psi(\varphi_{-1}(x)) \cdot \frac{1}{\dot\varphi(\varphi_{-1}(x))} = \frac{\dot\psi(t)}{\dot\varphi(t)}.$$

Thus, function f is differentiable in the set $R(\varphi|_{M_1}) \equiv \varphi(M_1)$ and its derivative f' can be parametrically represented by the equations

$$x = \varphi(t), \qquad \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\dot\psi(t)}{\dot\varphi(t)}; \qquad t \in M_1. \tag{III.7.3}$$

III.7.8. Remark.

In scientific literature, you can often find equations (III.7.3) written down in such a way that the second equation has only y instead of $\frac{\mathrm{d}y}{\mathrm{d}x}$ on the left– hand side. This is a purely formal matter – a dependent variable as well as an independent variable can be denoted by various symbols. Hence the dependent variable in equations (III.7.3) can be denoted by y, y' and also by $\frac{\mathrm{d}y}{\mathrm{d}x}$. The last possibility is perhaps less usual, but we regard it as more instructive in a given situation.

III.7.9. Example.

Verify whether the equations $x = 2t - \sin t$, $y = 1 + \cos t$ (for $t \in \langle 0; 2\pi \rangle$) define parametrically a function $y = f(x)$. In a positive case, decide about its continuity, specify its domain, its range and find intervals of monotonicity.

Solution: The function $\varphi(t) = 2t - \sin t$ has the positive derivative $2 - \cos t$ in the interval $\langle 0; 2\pi \rangle$, hence it is increasing (and consequently also one–to–one) in this interval. This means that the given equations define parametrically a function $y = f(x)$. Both functions $\varphi(t) = 2t - \sin t$ and $\psi(t) = 1 + \cos\, t$ are continuous in the interval $\langle 0; 2\pi \rangle$. Therefore, by Theorem III.7.5, the parametrically defined function $y = f(x)$ is also continuous on its domain.

The range of the function $\varphi(t) = 2t - \sin t$ on the interval $\langle 0; 2\pi \rangle$ is the interval $\langle 0; 4\pi \rangle$. (This is apparent from the fact that φ is continuous and increasing in $\langle 0; 2\pi \rangle$, its value at the point 0 is 0 and its value at the point 2π is 4π.) Thus, $D(f) = R(\varphi) = \langle 0; 4\pi \rangle$.

The range of the function $\psi(t) = 1 + \cos t$ on the interval $\langle 0; 2\pi \rangle$ is the interval $\langle 0; 2 \rangle$. Thus, $R(f) = R(\psi) = \langle 0; 2 \rangle$.

Both functions $\varphi(t) = 2t - \sin t$ and $\psi(t) = 1 + \cos t$ are differentiable in the interval $\langle 0; 2\pi \rangle$ and $\dot\varphi(t) = 2 - \cos t \neq 0$ for all $t \in \langle 0; 2\pi \rangle$. Substituting to equations (III.7.3), we get the following parametric representation of the derivative of function f:

$$x = 2t - \sin t, \qquad \frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{\sin t}{2 - \cos t}; \qquad t \in \langle 0; 2\pi \rangle.$$

The values of $f'(x)$ are given by the second equation. So it holds:

$$f'(x) > 0 \iff -\frac{\sin t}{2-\cos t} > 0 \iff -\sin t > 0 \iff t \in (\pi; 2\pi) \iff x \in (2\pi; 4\pi),$$

$$f'(x) < 0 \iff -\frac{\sin t}{2-\cos t} < 0 \iff -\sin t < 0 \iff t \in (0; \pi) \iff x \in (0; 2\pi).$$

Hence f is decreasing on the interval $\langle 0; 2\pi\rangle$ and increasing on the interval $\langle 2\pi; 4\pi\rangle$. It has an absolute minimum at the point 2π, where $f(2\pi) = \psi(\pi) = 0$. It has an absolute maximum at the points 0 and 4π, where $f(0) = \psi(0) = 2$ and $f(4\pi) = \psi(2\pi) = 2$.

III.7.10. The second derivative of a parametrically defined function.

In addition to the assumptions from paragraph III.7.2, let us suppose that functions φ and ψ have second derivatives $\ddot{\varphi}$, $\ddot{\psi}$ in the set $M_2 \subset M$ and $\dot{\varphi}(t) \neq 0$ for all $t \in M_2$. The equations (III.7.2) define parametrically a function $y = f(x)$. Its derivative is represented parametrically by equations (III.7.3). The second derivative can be obtained as a first derivative of the first derivative, i.e. to express it, we apply to equations (III.7.3) the same procedure as we applied to equations (III.7.2) in paragraph III.7.7. If we denote by ϑ the function on the right hand side of the second equation in (III.7.3) (i.e. $\vartheta(t) = \dfrac{\dot{\psi}(t)}{\dot{\varphi}(t)}$), we obtain the following parametric representation of the function f'':

$$x = \varphi(t), \qquad \frac{\mathrm{d}^2y}{\mathrm{d}x^2} = \frac{\dot{\vartheta}(t)}{\dot{\varphi}(t)}; \qquad t \in M_2.$$

Substituting $\dot{\vartheta}(t) = \dfrac{\ddot{\psi}(t)\cdot\dot{\varphi}(t) - \dot{\psi}(t)\cdot\ddot{\varphi}(t)}{[\dot{\varphi}(t)]^2}$ to the second of the above equations, we get:

$$f'' : \qquad x = \varphi(t), \qquad \frac{\mathrm{d}^2y}{\mathrm{d}x^2} = \frac{\ddot{\psi}(t)\cdot\dot{\varphi}(t) - \dot{\psi}(t)\cdot\ddot{\varphi}(t)}{[\,\dot{\varphi}(t)\,]^3}; \qquad t \in M_2.$$

Other higher order derivatives of a parametrically defined function can be expressed analogously.

III.7.11. Problems.

Verify that the given equations define parametrically functions $y = f(x)$. Are these functions continuous? Specify their domains, their ranges and find a parametric representation of their first and second derivatives.

a) $x = t^3 + t + 1, \quad y = t^4 + 2t + 2; \quad t \in (-\infty; +\infty)$,

b) $x = \cos^3 t, \quad y = \sin^3 t; \quad t \in (0; \dfrac{\pi}{2})$.

Results:

a) f is continuous, $D(f) = (-\infty; +\infty)$, $R(f) = (2^{-\frac{4}{3}} - 2^{\frac{2}{3}} + 2; +\infty)$,

$$f' : \quad x = t^3 + t + 1, \quad \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{4t^3+2}{3t^2+1}; \qquad t \in (-\infty; +\infty),$$

$$f'' : \quad x = t^3 + t + 1, \quad \frac{\mathrm{d}^2y}{\mathrm{d}x^2} = \frac{12\,(t^4+t^2-t)}{3t^2+1}; \quad t \in (-\infty; +\infty).$$

b) f is continuous, $D(f) = \langle 0; 1\rangle$, $R(f) = \langle 0; 1\rangle$,

$$f' : \quad x = \cos^3 t, \quad \frac{\mathrm{d}y}{\mathrm{d}x} = -\tan t; \qquad t \in (0; \tfrac{\pi}{2}),$$
$$f'' : \quad x = \cos^3 t, \quad \frac{\mathrm{d}^2y}{\mathrm{d}x^2} = \frac{1}{3\sin t\,\cos^4 t}; \quad t \in (0; \tfrac{\pi}{2}).$$

III.8.* Approximate solution of the nonlinear equation $f(x) = 0$

III.8.1. The root of the equation $f(x) = 0$.

Let f be a function. Every point $\xi \in D(f)$ such that $f(\xi) = 0$ is called the *root* of the equation $f(x) = 0$.

III.8.2. Motivation.

The equation $\mathrm{e}^x - 5 + x = 0$ has one root in the interval $(-\infty; +\infty)$. Try to verify this for yourself ! (Hint: The equation can be written in the equivalent form $\mathrm{e}^x = 5 - x$. It follows immediately from the behavior of the functions e^x and $5 - x$ that their graphs cross at just one point. Sketch these graphs !) However, you will not succeed in expressing "the unknown" x from the considered equation. Its analytic solution is impossible.

III.8.3. Remark.

You learned a series of methods for solving various types of equations at secondary school. These methods mostly led to a so called analytic expression of roots. This means that the roots were given by some formulas and their numerical value could be obtained by performing a finite number of arithmetic operations. In many, indeed in most cases, however, this is impossible. Simple equations can usually be solved analytically, while slightly more complicated cases mostly cannot.

We are going to explain two so called approximate methods of solution in the following paragraphs. It is characteristic of these methods that they enable us to express the solution only approximately, but with an arbitrarily small error. The maximum admissible value of the error can usually be chosen before the beginning of the computation. This is quite satisfactory for practical purposes – you can compare it with the situation where the equation can be solved analytically, but the formula that represents the solution contains some square or higher order roots. The value of the root is often an irrational number, so it can also be determined e.g. by the decimal expansion only approximately. (See for instance the situation when the root of an equation is $\sqrt{3} + 5$.)

Approximate methods mostly require performance of a higher number of arithmetic operations. Thus, an effective realization of these methods is possible only on computers. Approximate methods are also often called numerical methods.

III.8.4. Successive approximations, iterative sequence, error estimate.

Methods of solution of the equation $f(x) = 0$ that we are going to explain in next paragraphs are based on the construction of so called *successive approximations*. We choose in some way (respecting instructions of a used particular method) an *initial approximation* x_0 and then (also in accordance with the instructions of the method) we construct further approximations $x_1, x_2,$

etc. The sequence $\{x_n\}$ is called the *iterative sequence*. Methods based on the construction of an iterative sequence are called *iterative methods*.

The sketched approach has a sense only if $\lim_{n\to+\infty} x_n = \xi$, where ξ is a root of the equation $f(x) = 0$. The reason is that in this case, computing further and further approximations, we usually get nearer and nearer to the exact solution – to the root ξ. Thus, an important part of every iterative method is not only an instruction how to choose an initial approximation x_0 and how to construct further approximations $x_1, x_2, \ldots$, but also an information when (i.e. under which conditions) the iterative sequence converges to the root ξ of the equation $f(x) = 0$.

Every procedure must be sometimes finished. This means that we cannot proceed with the construction of successive approximations to infinity, we must content ourselves with approximations x_n up to some index n. However, how to choose the index of the last approximation in a particular case? This is closely connected with a required accuracy we want to solve the equation $f(x) = 0$ with. Most iterative methods contain as their parts estimates of the type $|x_n - \xi| \le \gamma_n$, where $\gamma_n \to 0$ for $n \to +\infty$ and the methods enable us to specify the number γ_n. Such estimates are called *error estimates*. They tell us that if we replace an exact root ξ of the equation $f(x) = 0$ by an approximate solution x_n, we make an error at most γ_n. So, when we are in the situation that we wish to solve the equation $f(x) = 0$ with an error not exceeding a given positive number ϵ, we compute the approximations to the index n which is so large that $\gamma_n \le \epsilon$. Then we can be satisfied with the approximation x_n because the error estimate yields $|x_n - \xi| \le \gamma_n \le \epsilon$ and so we can regard x_n as an approximate solution of the equation $f(x) = 0$.

It is necessary to mention that computations are sometimes made without error estimates. We simply decide to be satisfied for example with the approximation x_{100} and proclaim it an approximate solution. Nevertheless, it is obvious that this approach is not as correct as if an error estimate is used.

III.8.5. Separation of a root.

By the *separation of a root* we understand the specification of an interval $\langle a; b\rangle$ such that the equation $f(x) = 0$ has a unique root ξ in $\langle a; b\rangle$. To separate the roots of the equation $f(x) = 0$, we often use theorems III.3.26 and III.5.4.

III.8.6. Example.

Let us separate the roots of the equation $\ln x - 2x + 7 = 0$. The function $f(x) = \ln x - 2x + 7$ is defined in the interval $(0; +\infty)$, where it is also continuous and its derivative is $f'(x) = \dfrac{1}{x} - 2$. Moreover,

$$\lim_{x\to 0+} f(x) = -\infty, \qquad \lim_{x\to+\infty} f(x) = -\infty.$$

You can easily verify that the derivative $f'(x)$ is positive for $x \in (0;\, 0.5)$, equal to zero for $x = 0.5$ and negative for $x \in (0.5; +\infty)$. Hence function f is increasing in the interval $(0;\, 0.5\rangle$ and decreasing in the interval $\langle 0.5; +\infty)$. It has a strict local maximum at the point 0.5 and $f(0.5) = 6 - \ln 2 > 0$.

Let us now choose $a > 0$ sufficiently small, for example $a = 0.0001$ and let us show that one root of the equation $f(x) = 0$ is separated in the interval $\langle a;\, 0.5\rangle$.

a) Existence of a root: We already know that f is continuous in the interval $\langle a;\, 0.5\rangle$, $f(a) = f(0.0001) = \ln 0.0001 - 0.0002 + 7 = (-4)\cdot \ln 10 - 0.0002 + 7 < (-4)\cdot 2 - 0.0002 + 7 < 0$ and $f(0.5) > 0$. Since 0 is between $f(a)$ and $f(0.5)$, it follows from Theorem III.3.26 that there exists such a point ξ_1 between a and 0.5 that $f(\xi_1) = 0$.

b) Uniqueness of the root: Function f is increasing in $\langle a;\, 0.5\rangle$, so it takes on every its value in this interval only once. This implies the uniqueness of the point $\xi_1 \in \langle a;\, 0.5\rangle$ such that $f(\xi_1) = 0$.

It can be proved in a similar way that if one chooses $b > 0$ large enough, for example $b = 10$, then there exists another root ξ_2 of the equation $f(x) = 0$ in the interval $\langle 0.5;\, b\rangle$. ξ_1 and ξ_2 are the only roots of the equation $f(x) = 0$.

III.8.7. The Cut and Try Method.

Suppose that function f is continuous and strictly monotonic in the interval $\langle a; b\rangle$ and $f(a) \cdot f(b) < 0$. These assumptions guarantee the existence of a unique root ξ of the equation $f(x) = 0$ in $\langle a; b\rangle$ and moreover, the iterative sequence whose construction is described in the following converges to ξ.

Choice of the initial approximation: Put $x_0 = \dfrac{a+b}{2}$.

Calculation of further approximations: If $f(x_0) \cdot f(b) < 0$ then $\xi \in (x_1; b\rangle$. Therefore we change a and we put $a = x_0$. If $f(x_0) \cdot f(b) \geq 0$ then $\xi \in \langle a; x_0\rangle$ and we change b: we put $b = x_0$. Further, we put $x_1 = \dfrac{a+b}{2}$. Similarly, we obtain x_2, x_3, etc. (Illustrate the procedure on an appropriate picture for yourself!)

The error estimate: Denote by L the length of the interval $\langle a; b\rangle$ at the beginning of the calculation. Since $\xi \in \langle a; b\rangle$, $|x_0 - \xi| \leq \dfrac{L}{2}$. The length of the "variable" interval $\langle a; b\rangle$ (where the root ξ is separated) decreases by one half at each step. Hence

$$|x_n - \xi| \leq \frac{L}{2^{n+1}}.$$

III.8.8. Newton's Method.

Suppose that

a) function f has a second derivative $f''(x)$ at each point $x \in \langle a; b\rangle$ and $f''(x)$ does not change its sign in $\langle a; b\rangle$.
b) $f'(x) \neq 0$ for all $x \in \langle a; b\rangle$,
c) $f(a) \cdot f(b) < 0$.

It can be proven under these assumptions that the equation $f(x) = 0$ has a unique root ξ in $\langle a; b\rangle$ and the iterative sequence, constructed in accordance with rules described in the following, converges to ξ.

Choice of the initial approximation: The initial approximation x_0 can be chosen to be equal to an arbitrary point of the interval $\langle a; b\rangle$ such that $f(x_0) \cdot f''(x_0) \geq 0$. (Among others, this inequality is satisfied by one of the points a and b.)

Calculation of further approximations: To approximate the curve $y = f(x)$ in the neighborhood of the point $[x_0; f(x_0)]$, we use a tangent line to the graph of f at this point. The point where this line crosses the x–axis is called x_1. Similarly, the point where the tangent line to the graph of f at $[x_1; f(x_1)]$ crosses the x–axis is the next approximation x_2, etc. (Sketch a picture for yourself!) This procedure can easily be expressed by computation. Suppose that you already know the approximation x_n and you wish to find the next approximation x_{n+1}. The equation for the tangent line to the graph of f at the point $[x_n; f(x_n)]$ is $y = f(x_n) + f'(x_n) \cdot (x - x_n)$. (See paragraph III.4.5.) $y = 0$ corresponds to $x = x_{n+1}$. So we get the equation $0 = f(x_n) + f'(x_n) \cdot (x_{n+1} - x_n)$, which yields:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

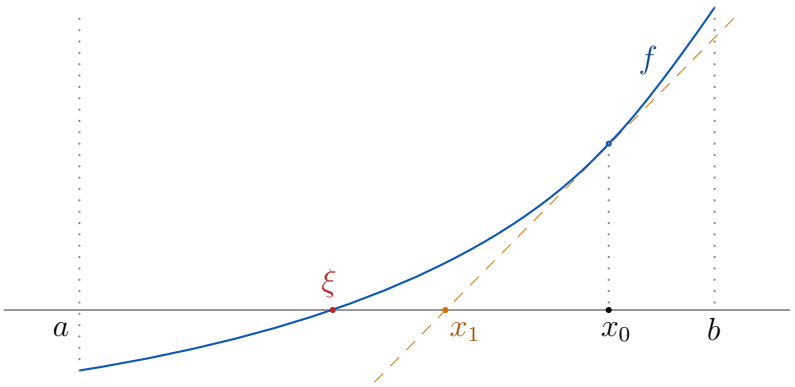


Fig. 38

The error estimate: It follows from the Mean Value Theorem (Theorem III.5.1), applied on the interval with end points x_n and ξ, that there exists η between x_n and ξ such that $f(x_n) - f(\xi) = f'(\eta) \cdot (x_n - \xi)$. However, $f(\xi)$ equals to zero (because ξ is the root of the equation $f(x) = 0$). Thus, we have $x_n - \xi = \dfrac{f(x_n)}{f'(\eta)}$ and this further implies that

$$|\, x_n - \xi \,| \leq \frac{|\, f(x_n) \,|}{m},$$

where $m = \min_{\eta \in \langle a;b \rangle} |\, f'(\eta) \,|$.

III.8.9. Remark.

You will meet other approximate methods of solution of the equation $f(x) = 0$ in your future studies of subject Numerical Mathematics. You will also learn how to work on an approximate solution of systems of generally nonlinear algebraic equations for a larger number of unknowns.

III.9.* Complex and vector–valued functions of a real variable

III.9.1. A complex function of a real variable.

If $M \subset \mathbb{R}$ then each mapping of M to $\mathbb{C}$ (where $\mathbb{C}$ is the domain of complex numbers) is called a *complex function of a real variable* (or shortly a *complex function*).

Every complex function F can be expressed in the form $F = f + \mathrm{i}\, g$ where f and g are real functions (see paragraph III.2.1). Function f is called the *real part* of the complex function F (we write: $f = Re\, F$) and function g is called the *imaginary part* of the complex function F (we denote it: $g = Im\, F$).

III.9.2. The limit of a complex function.

We say that a complex function $F = f + \mathrm{i}\, g$ has a limit L $(= a + \mathrm{i}\, b)$ at point x_0 if $\lim\limits_{x \to x_0} f(x) = a$ and $\lim\limits_{x \to x_0} g(x) = b$. We write: $\lim\limits_{x \to x_0} F(x) = L$.

By analogy, we can define the *left–hand limit* and the *right–hand limit*.

III.9.3. Continuity of a complex function.

A complex function $F = f + \mathrm{i}\, g$ is said to be *continuous at point x_0* if both functions f and g are continuous at x_0.

The *left continuity at point x_0*, the *right continuity at point x_0* and the *continuity on interval I* of a complex function can also be defined analogously.

III.9.4. Derivative of a complex function.

If functions f and g are both differentiable at point x_0 then we say that the complex function $F = f + \mathrm{i}\, g$ is also *differentiable at point x_0*. Its derivative at point x_0 is the complex number $F'(x_0) = f'(x_0) + \mathrm{i}\, g'(x_0)$.

One can also define, by analogy, the *left derivative* and the *right derivative*.

III.9.5. A vector–valued function of a real variable.

If $M \subset \mathbb{R}$ then every mapping of M to $V(\mathbb{E}_3)$ is called a *vector–valued function of a real variable* (or briefly a *vector–valued function* or a *vector function*).

Each vector function $\vec{f}$ can be written in the form $\vec{f} = (u; v; w)$ where u, v, w, are "scalar" functions (see paragraph III.2.1). Functions u, v, w are called the *component functions* of vector function $\vec{f}$.

Vector function $\vec{f} = (u; v; w)$ can also be written down in the form $\vec{f} = u\,\widehat{\imath} + v\,\widehat{\jmath} + w\,\widehat{k}$, where $\widehat{\imath}$, $\widehat{\jmath}$, $\widehat{k}$ are unit vectors oriented in accordance with the coordinate axes: $\widehat{\imath} = (1; 0; 0)$, $\widehat{\jmath} = (0; 1; 0)$, $\widehat{k} = (0; 0; 1)$.

III.9.6. The limit of a vector function.

If $\lim\limits_{x\to x_0} u(x) = \alpha$, $\lim\limits_{x\to x_0} v(x) = \beta$, and $\lim\limits_{x\to x_0} w(x) = \gamma$ then we say that the vector function $\vec{f} = (u; v; w)$ has the limit $\vec{L} = (\alpha; \beta; \gamma)$ as x approaches x_0. We write: $\lim\limits_{x\to x_0} \vec{f}(x) = \vec{L}$.

The *left limit* and the *right limit* can be defined similarly.

III.9.7. Continuity of a vector function.

We say that a vector function $\vec{f} = (u; v; w)$ is *continuous at point x_0* if all component functions u, v, w are continuous at point x_0.

The notions of the *left continuity at point x_0*, the *right continuity at point x_0* and the *continuity on interval I* of a vector function can also be defined analogously.

III.9.8. Derivative of a vector function.

The *derivative of a vector function* $\vec{f} = (u; v; w)$ at point x_0 is the vector denoted by $\vec{f}'(x_0)$ that satisfies: $\vec{f}'(x_0) = (u'(x_0); v'(x_0); w'(x_0)) = u'(x_0)\,\widehat{\imath} + v'(x_0)\,\widehat{\jmath} + w'(x_0)\,\widehat{k}$ (if the derivatives $u'(x_0)$, $v'(x_0)$ and $w'(x_0)$ exist).

The notions of the *left derivative* and the *right derivative* can be defined by analogy.

III.9.9. Remark.

A vector function which has another number of component functions and its limit, continuity and derivative can be introduced in the same way.

III.9.10. Example.

If a point moves in space and its position at time t is given by the position vector of point $X \in \mathbb{R}^3$: $X = P(t) = [3\cos t;\ 3\sin t;\ 4t]$, then the instantaneous velocity (vector) of the point is $\dot{P}(t) = (3\cos t;\ -3\sin t;\ 4)$, its instantaneous acceleration is $\ddot{P}(t) = (-3\sin t;\ -3\cos t;\ 0)$.

III.9.11. Problems.

Find $\vec{f}'$ and $\vec{f}''$, if vector function $\vec{f}$ has the form

a) $\vec{f}(x) = (2-3x)\,\widehat{\imath} + (x^2-5)\,\widehat{\jmath} + (2x-7)\,\widehat{k}$,

b) $\vec{f}(t) = (\sin^2 t;\ \tan t;\ \ln t - t)$.

Results: a) $\vec{f}'(x) = -3\widehat{\imath} + 2x\,\widehat{\jmath} + 2\,\widehat{k}$, $\vec{f}''(x) = 2\,\widehat{\jmath}$ (for $x \in (-\infty;+\infty)$),

b) $\vec{f}'(t) = \left(2\sin t\cos t;\ \dfrac{1}{\cos^2 t};\ \dfrac{1}{t} - 1\right)$,

$\vec{f}''(t) = \left(2\cos^2 t - 2\sin^2 t;\ \dfrac{2\sin t}{\cos^3 t};\ -\dfrac{1}{t^2}\right)$

(for $t \in \left(0;\dfrac{\pi}{2}\right) \cup \left(\dfrac{\pi}{2};\dfrac{3\pi}{2}\right) \cup \left(\dfrac{3\pi}{2};\dfrac{5\pi}{2}\right) \cup \ldots$).

III.9.12. Remark.

To evaluate limits and derivatives of a sum, a difference, a product and a quotient of complex functions (respectively a sum, a difference and a product of vector functions), we can use the same rules as in the case of real functions.

Vector functions can be multiplied in the same way as vectors – i.e. there exists a scalar product and a vector product of vector functions. The rule for differentiation of a product as it is known from calculus of real functions (see paragraph III.4.10 d)) can also be applied to both types of products of vector functions:

If vector functions $\vec{f}$ and $\vec{g}$ have derivatives at point x_0 then the products $\vec{f}\cdot\vec{g}$ and $\vec{f}\times\vec{g}$ also have derivatives at point x_0 and it holds:

$$\begin{aligned}(\vec{f}\cdot\vec{g})'(x_0) &= \vec{f}\,'(x_0)\cdot\vec{g}(x_0) + \vec{f}(x_0)\cdot\vec{g}\,'(x_0),\\ (\vec{f}\times\vec{g})'(x_0) &= \vec{f}\,'(x_0)\times\vec{g}(x_0) + \vec{f}(x_0)\times\vec{g}\,'(x_0).\end{aligned}$$

CHAPTER IV.

Indefinite integral

IV.1. Antiderivative, indefinite integral

IV.1.1. Antiderivative.

Let I be an interval with end points a, b (where $a < b$; numbers a, b can also be infinite). A function F is called an *antiderivative* to function f on interval I if

a) $F'(x) = f(x)$ for all $x \in (a; b)$,
b) $F'_+(a) = f(a)$ (if point a belongs to interval I),
c) $F'_-(b) = f(b)$ (if point b belongs to interval I).

IV.1.2. Remark.

Instead of "F is an antiderivative to function f on interval I", we shall sometimes write only "$F' = f$ in I".

Almost the whole part IV of this textbook deals with the question how to find an antiderivative to a given function f. The antiderivative is important mainly because it enables to evaluate a so called definite integral of function f. (See part V.) Moreover, the solution of many differential equations also leads to the problem of finding the antiderivative. (See section IV.7.)

To find an antiderivative to a given function f is an "inverse" problem to the differentiation of function f. We already know that not every function has a derivative. Similarly, it is also not true that an antiderivative automatically exists to every given function. The following theorem gives a sufficient condition for the existence of an antiderivative to f on interval I.

IV.1.3. Theorem (on the existence of an antiderivative).

If function f is continuous on interval I then there exists an antiderivative to f on I.

IV.1.4. Remark.

The other question is whether an antiderivative to function f on interval I (if it exists) is unique. The following theorem says that it is not unique. There exist infinitely many antiderivatives, but any two of them differ at most in an additive constant.

IV.1.5. Theorem.

a) *If F is an antiderivative to function f on interval I and C is an arbitrary real constant then function G defined by the formula $G(x) = F(x) + C$ (for $x \in I$) is also an antiderivative to function f on interval I.*
b) *If F and G are two antiderivatives to function f on interval I then there exists $C \in \mathbb{R}$ such that $G(x) = F(x) + C$ for all $x \in I$.*

Proof:

a) If $F' = f$ in I, then also $G' = (F + C)' = F' + C' = F' = f$ in I.
b) If $F' = f$ in I and $G' = f$ in I, then $(G - F)' = G' - F' = f - f = 0$ in I. Hence the function $H = G - F$ has a zero derivative in interval I and so it is a constant function in interval I. (See theorem III.5.4, item e).) Thus, there exists $C \in \mathbb{R}$ such that

$H(x) = G(x) - F(x) = C$ for all $x \in I$. This means that $G(x) = F(x) + C$ for all $x \in I$.

IV.1.6. Theorem.

If F and G are antiderivatives to functions f and g on interval I, then $F + G$ is an antiderivative to $f + g$ on interval I.
If F is an antiderivative to f on interval I and α is a real number, then $\alpha \cdot F$ is an antiderivative to $\alpha \cdot f$ on interval I.

Proof:
This follows immediately from theorem III.4.9, parts a) and b).

IV.1.7. Indefinite integral.

The *indefinite integral* of function f on interval I is the set of all antiderivatives to function f on interval I. This set is denoted by

$$\int f(x)\,\mathrm{d}x, \quad x \in I \qquad \text{or briefly only} \qquad \int f(x)\,\mathrm{d}x, \quad \int f\,\mathrm{d}x.$$

If F is an antiderivative to function f on interval I, then the set $\int f(x)\,\mathrm{d}x$ contains only function F and all other functions that differ from F at most in an additive constant on I. This can be expressed by the equation

$$\int f(x)\,\mathrm{d}x = F(x) + C, \qquad x \in I.$$

The function f is called the *integrand* and the constant C is called the *constant of integration*.

It follows from the definition of the indefinite integral that the integral $\int f(x)\,\mathrm{d}x$ in interval I exists if and only if function f has an antiderivative in interval I. Due to theorem IV.1.3, the continuity of function f in I is a sufficient condition for the existence.

IV.1.8. Theorem.

If the indefinite integrals of functions f and g exist on interval I and $\alpha \in \mathbb{R}$ then

(IV.1.1) $$\int \big(f(x) + g(x)\big)\,\mathrm{d}x = \int f(x)\,\mathrm{d}x + \int g(x)\,\mathrm{d}x, \qquad x \in I$$

(IV.1.2) $$\int \alpha f(x)\,\mathrm{d}x = \alpha \int f(x)\,\mathrm{d}x, \qquad x \in I.$$

This theorem directly follows from theorem IV.1.6. Equality (IV.1.1) is in fact the equality of sets of functions. It says:

a) An arbitrary function from the set on the left hand side, i.e. from $\int (f(x) + g(x))\,\mathrm{d}x$, can be expressed as a sum of two functions from the sets on the right hand side, i.e. a function from the set $\int f(x)\,\mathrm{d}x$ and a function from the set $\int g(x)\,\mathrm{d}x$.
b) On the other hand, the sum of two arbitrary functions from the sets on the right hand side, i.e. a function from $\int f(x)\,\mathrm{d}x$ and a function from $\int g(x)\,\mathrm{d}x$, belongs to the set on the left hand side, i.e. to $\int (f(x) + g(x))\,\mathrm{d}x$.

The precise sense of equality (IV.1.2) can be explained similarly.

IV.1.9. Corollary.

Theorem IV.1.8 can be generalized: *If functions* $f_1, \dots, f_n$ *have indefinite integrals in interval* I *and* $\alpha_1, \dots, \alpha_n$ *are real numbers then*

$$\int \left(\alpha_1 f_1(x) + \cdots + \alpha_n f_n(x)\right) \mathrm{d}x = \alpha_1 \int f_1(x)\,\mathrm{d}x + \cdots + \alpha_n \int f_n(x)\,\mathrm{d}x, \quad x \in I.$$

IV.1.10. Table of basic indefinite integrals.

Basic formulas for some indefinite integrals can simply be obtained by "inverting" the formulas for differentiation of elementary functions (see paragraphs III.4.10 and III.4.15). Each of these formulas holds on an arbitrary interval I which is contained in the domain of the integrand.

a) $\int 0\,\mathrm{d}x = C$,

b) $\int x^\alpha\,\mathrm{d}x = \frac{x^{\alpha+1}}{\alpha+1} + C \quad (\alpha \neq -1)$,

c) $\int \cos x\,\mathrm{d}x = \sin x + C$,

d) $\int \sin x\,\mathrm{d}x = -\cos x + C$,

e) $\int \frac{1}{\cos^2 x}\,\mathrm{d}x = \tan x + C$,

f) $\int \frac{1}{\sin^2 x}\,\mathrm{d}x = -\operatorname{cotan} x + C$,

g) $\int \frac{1}{\sqrt{1-x^2}}\,\mathrm{d}x = \arcsin x + C$,

h) $\int \frac{1}{1+x^2}\,\mathrm{d}x = \arctan x + C$,

i) $\int a^x\,\mathrm{d}x = \frac{a^x}{\ln a} + C \quad (a > 0,\ a \neq 1)$,

j) $\int \mathrm{e}^x\,\mathrm{d}x = \mathrm{e}^x + C$,

k) $\int \frac{1}{x}\,\mathrm{d}x = \ln|x| + C$.

IV.1.11. Remark.

It is not quite clear at first sight how formula k) follows from the formula for differentiation of the function $\ln x$ (i.e. from formula f) in paragraph III.4.15). $|x| = x$ in the interval $(0; +\infty)$, so the validity of k) is obvious in this interval. However, since $|x| = -x$ in the interval $(-\infty; 0)$, we also have

$$(\ln|x|)' = \frac{1}{|x|} \cdot |x|' = \frac{1}{(-x)} \cdot (-x)' = \frac{1}{(-x)} \cdot (-1) = \frac{1}{x} \quad \text{for } x \in (-\infty; 0).$$

This confirms the validity of formula k) in the interval $(-\infty; 0)$, too.

IV.1.12. Remark.

Using the formulas for derivatives of functions arc sine and arc cosine (from paragraph III.4.15), we can also express the integrals from items g) and h) in paragraph IV.1.10 as follows:

g)* $\int \frac{1}{\sqrt{1-x^2}}\,\mathrm{d}x = -\arccos x + C$,

h)* $\int \frac{1}{1+x^2}\,\mathrm{d}x = -\operatorname{arccotan} x + C$.

This means that the difference $(\arcsin x - (-\arccos x))$ is a constant function in the interval $(-1; +1)$ and similarly, the difference $(\arctan x - (-\operatorname{arccotan} x))$ is a constant function in the interval $(-\infty; +\infty)$.

IV.1.13. Remark.

Integration is the name given to the procedure leading to an antiderivative or to an indefinite integral of a given function f (which is practically the same thing, since in order to write down an indefinite integral it is sufficient to know any of the antiderivatives to f).

While the differentiation of a given function can be more or less well algorithmized (it usually suffices to choose and apply appropriate differentiation formulas), this cannot be said of the integration. There exists no general algorithm which will always lead to a desired result and which will provide a rule for the integration of any function. Conversely, there exist a lot of relatively simple functions whose antiderivatives exist, but they cannot be expressed in a "closed form", i.e. by a formula that prescribes the performance of a finite number of operations and allows us to use only elementary functions in them. (An example of such a function is $f(x) = \mathrm{e}^{-x^2}$ in the interval $(-\infty; +\infty)$.) A large range of various methods have been developed which enable us to find indefinite integrals for various classes of "similar" functions. This was an object of interest of many mathematicians especially in the 18th and 19th Centuries. Many of these methods are simple, while others are quite complicated, involving many relatively artificial substitutions, etc. Nowadays, these complicated procedures have rather lost their importance. This has been caused mainly by the development of computers, which make it possible to compute so called definite integrals and to solve differential equations (which are the fields where indefinite integrals are mostly used) by various approximate methods. Nevertheless, this does not mean that we can entirely rely on computers and need not learn to integrate at all! (Children still learn – and will continue to learn – their multiplication tables, even though computers can multiply numbers faster and more reliably.) This is why we show several simple methods for integration in the rest of this chapter.

IV.1.14. Example.

$$\int (x^3 - 3x^2 + 4x - 1)\,\mathrm{d}x = \int x^3\,\mathrm{d}x - 3\int x^2\,\mathrm{d}x + 4\int x\,\mathrm{d}x - \int 1\,\mathrm{d}x = \\ = \frac{1}{4}x^4 - x^3 + 2x^2 - x + C.$$

The method used in this example is called *term by term integration* or sometimes *integration by decomposition*. The integrand was decomposed into the sum (or the difference) of more simpler functions, each of which could be integrated individually. The correctness of this approach follows from theorem IV.1.8 and from corollary IV.1.9.

Constants of integration need not be added separately to each indefinite integral in the decomposition shown above. All such constants can be associated to one which can be denoted e.g. by C and added at the end.

IV.2. Integration by parts

IV.2.1. Motivation.

We have derived the rule for differentiation of a product of two functions in paragraph III.4.6. This rule can be symbolically written as follows:

$$(u \cdot v)' = u' \cdot v + u \cdot v' \quad \Longrightarrow \quad \begin{cases} u \cdot v' = (u \cdot v)' - u' \cdot v, \\ u' \cdot v = (u \cdot v)' - u \cdot v'. \end{cases}$$

IV.2.2. Theorem (on integration by parts).

If functions u and v have continuous derivatives on interval I then the following formula holds on I :

$$\int u(x)\cdot v'(x)\,\mathrm{d}x \;=\; u(x)\cdot v(x) \;-\; \int u'(x)\cdot v(x)\,\mathrm{d}x.$$

IV.2.3. Example.

$$\int x\sin x\,\mathrm{d}x = \begin{vmatrix} u = x, & v' = \sin x \\ u' = 1, & v = -\cos x \end{vmatrix} = -x\cos x - \int(-\cos x)\,\mathrm{d}x =$$
$$= -x\cos x + \sin x + C, \qquad x \in \mathbb{R}.$$

IV.2.4. Example.

$$\int (x^2-3x+1)\mathrm{e}^x\,\mathrm{d}x = \begin{vmatrix} u = x^2-3x+1, & v' = \mathrm{e}^x \\ u' = 2x-3, & v = \mathrm{e}^x \end{vmatrix} =$$
$$= (x^2-3x+1)\mathrm{e}^x - \int(2x-3)\mathrm{e}^x\,\mathrm{d}x = \begin{vmatrix} u = 2x-3, & v' = \mathrm{e}^x \\ u' = 2, & v = \mathrm{e}^x \end{vmatrix} =$$
$$= (x^2-3x+1)\mathrm{e}^x - (2x-3)\mathrm{e}^x + \int 2\mathrm{e}^x\,\mathrm{d}x = (x^2-5x+6)\mathrm{e}^x + C, \qquad x \in \mathbb{R}.$$

IV.2.5. Example.

$$\int \ln x\,\mathrm{d}x = \int 1\cdot\ln x\,\mathrm{d}x = \begin{vmatrix} u = \ln x, & v' = 1 \\ u' = \dfrac{1}{x}, & v = x \end{vmatrix} = x\ln x - \int x\cdot\frac{1}{x}\,\mathrm{d}x =$$
$$= x\ln x - \int 1\,\mathrm{d}x = x\ln x - x + C, \qquad x > 0.$$

IV.2.6. Example.

$$\int \mathrm{e}^x\sin x\,\mathrm{d}x = \begin{vmatrix} u = \sin x, & v' = \mathrm{e}^x \\ u' = \cos x, & v = \mathrm{e}^x \end{vmatrix} = \mathrm{e}^x\sin x - \int \mathrm{e}^x\cos x\,\mathrm{d}x =$$
$$= \begin{vmatrix} u = \cos x, & v' = \mathrm{e}^x \\ u' = -\sin x, & v = \mathrm{e}^x \end{vmatrix} = \mathrm{e}^x\sin x - \mathrm{e}^x\cos x + \int \mathrm{e}^x(-\sin x)\,\mathrm{d}x =$$
$$= \mathrm{e}^x\sin x - \mathrm{e}^x\cos x - \int \mathrm{e}^x\sin x\,\mathrm{d}x.$$

If we denote the integral by $\mathcal{I}$, this gives

$$\mathcal{I} = \mathrm{e}^x\sin x - \mathrm{e}^x\cos x - \mathcal{I}.$$

Thus,

$$\mathcal{I} = \int \mathrm{e}^x\sin x\,\mathrm{d}x = \frac{1}{2}\left(\mathrm{e}^x\sin x - \mathrm{e}^x\cos x\right) + C, \qquad x \in \mathbb{R}.$$

IV.2.7.* Example.

$$\int \frac{1}{(1+x^2)^n}\,\mathrm{d}x = \int 1\cdot\frac{1}{(1+x^2)^n}\,\mathrm{d}x = \left|\begin{matrix} u = \dfrac{1}{(1+x^2)^n}, & v' = 1 \\ u' = -\dfrac{2nx}{(1+x^2)^{n+1}}, & v = x \end{matrix}\right| =$$

$$= \frac{x}{(1+x^2)^n} - \int x\left(-\frac{2nx}{(1+x^2)^{n+1}}\right)\,\mathrm{d}x =$$

$$= \frac{x}{(1+x^2)^n} + 2n\int\frac{1+x^2}{(1+x^2)^{n+1}}\,\mathrm{d}x - 2n\int\frac{1}{(1+x^2)^{n+1}}\,\mathrm{d}x.$$

Denote $\mathcal{I}_n = \int \frac{1}{(1+x^2)^n}\,\mathrm{d}x$. Applying integration by parts, we obtain the equation

$$\mathcal{I}_n = \frac{x}{(1+x^2)^n} + 2n\cdot\mathcal{I}_n - 2n\cdot\mathcal{I}_{n+1}\,.$$

This leads to the formula $\mathcal{I}_{n+1} = \frac{1}{2n}\cdot\frac{x}{(1+x^2)^n} + \frac{2n-1}{2n}\mathcal{I}_n\,.$

Writing this for example with $n=1$ and substituting for $\mathcal{I}_1$ from IV.1.9 h), we get:

$$\mathcal{I}_2 = \frac{1}{2}\frac{x}{1+x^2} + \frac{1}{2}\int\frac{1}{1+x^2}\,\mathrm{d}x = \frac{1}{2}\frac{x}{1+x^2} + \frac{1}{2}\arctan x + C; \qquad x\in\mathbb{R}.$$

IV.2.8. Problems.

Find the integrals:

a) $\int x\cdot\mathrm{e}^x\,\mathrm{d}x$, b) $\int x\cdot\cos x\,\mathrm{d}x$, c) $\int x^2\cdot\cos x\,\mathrm{d}x$,

d) $\int x^2\cdot\ln x\,\mathrm{d}x$, e) $\int x\cdot\operatorname{arccotan} x\,\mathrm{d}x$, f) $\int \mathrm{e}^x\cdot\cos x\,\mathrm{d}x$.

Results:

a) $\mathrm{e}^x\cdot(x-1)+C$ (for $x\in\mathbb{R}$)

b) $x\cdot\sin x+\cos x+C$ (for $x\in\mathbb{R}$)

c) $x^2\cdot\sin x+2x\cos x-2\sin x+C$ (for $x\in\mathbb{R}$)

d) $\frac{1}{3}x^3\left(\ln x-\frac{1}{3}\right)+C$ (for $x>0$)

e) $\frac{1}{2}(x^2+1)\cdot\operatorname{arccotan} x+\frac{1}{2}x+C$ (for $x\in\mathbb{R}$)

f) $\frac{1}{2}\mathrm{e}^x(\sin x+\cos x)+C$ (for $x\in\mathbb{R}$).

IV.3. Integration by substitution

IV.3.1. Motivation.

Analogously, as integration by parts represents the reverse procedure to the differentiation of a product of two functions, integration by substitution is a method which arose by inverting the Chain Rule (see theorem III.4.14).

IV.3.2. Theorem.

Suppose that function $f(x)$ *is continuous on interval* I. *Suppose further that function* $x = g(t)$ *is differentiable on interval* J *and it maps* J *onto* I. *Then*

(IV.3.1) $$\int f(x)\,\mathrm{d}x = \int f(g(t))\cdot g'(t)\,\mathrm{d}t \qquad \text{for } x\in I,\ t\in J,\ x=g(t).$$

IV.3.3. Remark

The indefinite integral on the left hand side of (IV.3.1) is the set of all primitive functions to function $f(x)$ on interval I. Similarly, the indefinite integral on the right hand side is the set of all primitive functions to function $f(g(t))\cdot g'(t)$ on interval J. Thus, equality (IV.3.1) is the equality between two sets of functions. It looks strange at the first sight, because the set on the left hand side contains functions defined on interval I, while the set on the right hand side contains functions defined on interval J. However, it is necessary to understand the equality in such a way that if one replaces x by $g(t)$ in every function in the set on the left hand side, then both sets coincide.

Formula (IV.3.1) can be used in two situations:

1) We wish to evaluate the integral on the right hand side and we transform it to the integral on the left hand side. (This has a sense if the integral on the right hand side cannot be computed directly and the integral on the left hand side is simpler.)
2) We wish to find the integral on the left hand side and we transform it to the integral on the right hand side. (This has a sense if the integral on the right hand side is simpler.)

We describe both possibilities in so called first and second substitution methods.

IV.3.4. The 1st method of substitution.

Assume that all assumptions of theorem IV.3.2 are fulfilled. Suppose that we wish to evaluate the integral $\int f(g(t))\cdot g'(t)\ \mathrm{d}t$ on the right hand side of (IV.3.1). We put $g(t)=x$ and $g'(t)\,\mathrm{d}t=\mathrm{d}x$. So we obtain the integral appearing on the left hand side of (IV.3.1). We evaluate this integral (on interval I). Finally, replacing x by $g(t)$, we obtain the desired integral $\int f(g(t))\cdot g'(t)\,\mathrm{d}t$.

IV.3.5. Example.

$$\int \sin^3 t\ \cos t\,\mathrm{d}t = \left|\begin{matrix}\sin t = x\\ \cos t\,\mathrm{d}t = \mathrm{d}x\end{matrix}\right| = \int x^3\,\mathrm{d}x = \frac{1}{4}x^4 + C = \frac{1}{4}\sin^4 t + C.$$

The assumptions of theorem IV.3.2 are fulfilled, because the function $g(t)=\sin t$ is differentiable in $J=(-\infty;+\infty)$, maps this interval onto $I=\langle -1;+1\rangle$ and function $f(x)=x^3$ is continuous on interval I.

IV.3.6. Remark.

When we use the 1st method of substitution then the exact specification of interval I which is the range of $g|_J$ is not necessary. In many cases, it is easier to verify that g maps J into (and not necessarily onto) some interval I' where f is continuous. Then f is also continuous in the interval $I=R(g|_J)$, because $I\subset I'$.

Thus, as a part of the verification of the assumptions of theorem IV.3.2 in example IV.3.5, it is sufficient to find out that function g maps interval J into the interval $I = (-\infty; +\infty)$ where the function $f(x) = x^3$ is continuous.

IV.3.7. Remark.

We can use various denotations for the variable in integration. It is usually denoted by x, but this is not a rule. Thus, if we start with an integral of the type $\int f(g(x)) \cdot g'(x)\ \mathrm{d}x$ then the new variable can be denoted for example by s. Using the substitution $s = g(x)$, $\mathrm{d}s = g'(x)\ \mathrm{d}x$, we transform this integral to $\int f(s)\ \mathrm{d}s$. Get used to various denotations of variables!

IV.3.8. The 2nd method of substitution.

Suppose that the assumptions of theorem IV.3.2 are fulfilled and moreover, function g is strictly monotone in interval J. Assume that we wish to compute the integral $\int f(x)\ \mathrm{d}x$ in interval I. We put $x = g(t)$ and $\mathrm{d}x = g'(t) \cdot \mathrm{d}t$. Thus, we transform the problem to the computation of the integral $\int f(g(t)) \cdot g'(t)\ \mathrm{d}t$ in interval J. We evaluate this integral, replace t by $g_{-1}(x)$ and we obtain the expression of the original integral $\int f(x)\ \mathrm{d}x$.

IV.3.9. Remark.

The application of the 2nd method requires the inverse substitution $t = g_{-1}(x)$ at the very end of the computation. The assumption about the strict monotonicity of function g on interval J guarantees the existence of the inverse function g_{-1} to g.

IV.3.10. Example.

Evaluate $\int \sqrt{9 - x^2}\ \mathrm{d}x$. The integrand is defined and continuous in the interval $\langle -3; 3\rangle$. Hence we will evaluate the integral on this interval. We use the substitution $x = g(t) = 3\sin t$ (for $t \in \left\langle -\frac{\pi}{2}; \frac{\pi}{2} \right\rangle$). Function g is increasing (and therefore strictly monotone) on the interval $\left\langle -\frac{\pi}{2}; \frac{\pi}{2} \right\rangle$, it is differentiable here and it maps this interval onto the interval $\langle -3; 3\rangle$. Thus, function g satisfies the assumptions of the second substitution method. The change of variable gives

$$\int \sqrt{9 - x^2}\ \mathrm{d}x = \left| \begin{matrix} x = 3\sin t \\ \mathrm{d}x = 3\cos t\ \mathrm{d}t \end{matrix} \right| = \int \sqrt{9 - 9\sin^2 t}\ 3\cos t\ \mathrm{d}t = 9\int |\cos t| \cos t\ \mathrm{d}t \overset{*}{=}$$

$$\overset{*}{=} 9\int \cos^2 t\ \mathrm{d}t = \frac{9}{2}\int \big(1 + \cos(2t)\big)\ \mathrm{d}t = \frac{9}{2}\left(t + \frac{\sin(2t)}{2}\right) + C = \frac{9}{2}\left(t + \sin t \cos t\right) + C =$$

$$= \left| \begin{matrix} t = \arcsin\left(\frac{x}{3}\right) \\ \sin t = \frac{x}{3}, \quad \cos t = \sqrt{1 - \frac{x^2}{9}} \end{matrix} \right| = \frac{9}{2}\left(\arcsin\left(\frac{x}{3}\right) + \frac{x}{3}\sqrt{1 - \frac{x^2}{9}}\ \right) + C.$$

*) Function cosine is nonnegative in the interval $\left\langle -\frac{\pi}{2}; \frac{\pi}{2} \right\rangle$, hence $|\cos t| = \cos t$ for all t from this interval.

IV.3.11. Example.

Evaluate the integral $\int \frac{1}{1-x}\,\mathrm{d}x$. The domain of the function $\frac{1}{1-x}$ is the union $(-\infty;1)\cup(1;+\infty)$. The function $\frac{1}{1-x}$ is continuous on the two intervals $(-\infty;1)$ and $(1;+\infty)$, hence an antiderivative and the indefinite integral exist on each of them.

$$\int \frac{1}{1-x}\,\mathrm{d}x = \left|\begin{matrix} x = 1-u \\ \mathrm{d}x = -\mathrm{d}u \end{matrix}\right| = -\int \frac{1}{u}\,\mathrm{d}u = -\ln|u| + C = -\ln|1-x| + C.$$

Verify for yourself that all assumptions of the 2nd method of substitution are fulfilled here.

IV.3.12. Problems.

a) $\int \frac{x}{\sqrt[3]{1-x^2}}\,\mathrm{d}x$ b) $\int \frac{\cos x}{\sin^3 x}\,\mathrm{d}x$ c) $\int \frac{\exp\sqrt{x}}{\sqrt{x}}\,\mathrm{d}x$

d) $\int \frac{\sqrt{\arctan x}}{1+x^2}\,\mathrm{d}x$ e) $\int \frac{\ln x - 3}{x\sqrt{\ln x}}\,\mathrm{d}x$ f) $\int \mathrm{e}^{-x}\,\mathrm{d}x$

g) $\int \frac{\mathrm{e}^x}{1+\mathrm{e}^{2x}}\,\mathrm{d}x$ h) $\int (1-2x)^{1000}\,\mathrm{d}x$ i) $\int \frac{1}{\sqrt{1-2x^2}}\,\mathrm{d}x$

j) $\int \frac{1}{\sqrt{2^x+1}}\,\mathrm{d}x$ k) $\int \frac{\sqrt{1-x}}{x}\,\mathrm{d}x$ l) $\int \frac{\mathrm{d}x}{(1+x^2)\arctan x}$

Results:

a) $-\frac{3}{4}(1-x^2)^{\frac{2}{3}} + C$ (on intervals $(-\infty;-1)$, $(-1;1)$, $(1;+\infty)$)

b) $-\frac{1}{2\sin^2 x} + C$ (on each interval $(k\pi;\,(k+1)\pi)$, k being an integer)

c) $2\exp\sqrt{x} + C$ (on $(-\infty;+\infty)$)

d) $\frac{2}{3}(\arctan x)^{\frac{3}{2}} + C$ (on $(-\infty;+\infty)$)

e) $\frac{2}{3}\sqrt{\ln x}\,(\ln x - 9) + C$ (on $(1;+\infty)$)

f) $-\mathrm{e}^{-x} + C$ (on $(-\infty;+\infty)$)

g) $\arctan(\mathrm{e}^x) + C$ (on $(-\infty;+\infty)$)

h) $-\frac{1}{2002}(1-2x)^{1001} + C$ (on $(-\infty;+\infty)$)

i) $\frac{\arcsin(\sqrt{2}x)}{\sqrt{2}} + C$ (on $\left(-\frac{1}{\sqrt{2}};\frac{1}{\sqrt{2}}\right)$)

j) $\frac{1}{\ln 2}\ln\left(\frac{\sqrt{2^x+1}-1}{\sqrt{2^x+1}+1}\right) + C$
(on $(-\infty;+\infty)$)

k) $2\sqrt{1-x} + \ln\left|\frac{\sqrt{1-x}-1}{\sqrt{1-x}+1}\right| + C$ (on $(-\infty;1)$)

l) $\ln|\arctan x| + C$ (on intervals $(-\infty;0)$ and $(0;+\infty)$).

IV.4. Integration of simple rational functions

IV.4.1. Rational function.

A *rational function* is a function of the type $\frac{P}{Q}$, where P and Q are polynomials and the degree of polynomial Q is greater than or equal to one.

IV.4.2. Remark.

We are going to explain how it is possible to integrate rational functions which have a polynomial of the degree at most three in the denominator. The procedure of integration of other rational functions is similar, though it can be technically more complicated. Example IV.6.6 involves the integration of a rational function with a polynomial of the fourth degree in the denominator.

The basic idea is to write the rational function as a sum of terms, so called partial fractions, each of which can be relatively simply integrated. The procedure starts with the decomposition of the polynomial in the denominator (i.e. polynomial Q) to the product of factors, which are linear, in some cases also quadratic, polynomials.

Linear factors appearing in the decomposition have a close relation to the roots of polynomial Q: Q contains the linear factor $(x-\alpha)$ if and only if α is the root of Q.

A quadratic factor appears in the decomposition of Q only if it has no real factors, i.e. if it has no real roots and its discriminant is negative – see cases Ia) and IIa) in the following paragraph.

IV.4.3. The decomposition of a polynomial of at most the third degree.

I) Suppose first that Q is a quadratic polynomial, i.e.

$$Q(x) = q_0\,x^2 + q_1\,x + q_2 \qquad \text{(where } q_0 \neq 0\text{)}.$$

There are three possibilities:

Ia) Q has no real root and so it has no real linear factor.

Ib) Q has a unique (repeated) real root α and it can be decomposed as follows: $Q(x) = q_0\,(x-\alpha)^2$.

Ic) Q has two distinct real roots α, β and it can be decomposed as follows: $Q(x) = q_0\,(x-\alpha)\,(x-\beta)$.

II) Suppose now that Q is a cubic polynomial, i.e.

$$Q(x) = q_0\,x^3 + q_1\,x^2 + q_2\,x + q_3 \qquad \text{(where } q_0 \neq 0\text{)}.$$

There are four possibilities:

IIa) Q has a unique single real root α. In this case Q can be written in the form $Q(x) = q_0\,(x-\alpha)\,(x^2+rx+s)$, where (x^2+rx+s) is a quadratic polynomial which is has no real factors (its discriminant is negative).

IIb) Q has a unique triple real root α. Then $Q(x) = q_0\,(x-\alpha)^3$.

IIc) Q has a single real root α and a double real root β. Then Q can be written as follows: $Q(x) = q_0\,(x-\alpha)\,(x-\beta)^2$.

IId) Q has three distinct real roots α, β, γ. Then $Q(x) = q_0\,(x-\alpha)\,(x-\beta)\,(x-\gamma)$.

Sketch for yourself the graph of polynomial Q in all the above cases!

IV.4.4. Example.

Decompose the polynomial

$$Q(x) \;=\; 2x^3 \,-\, 4x^2 \,-\, 20x \,-\, 50.$$

to the product of linear or quadratic factors (which are further irreducible in the real domain).

First we need to find the real roots of this polynomial. There exist general formulas (so called Cardan's formulas) which enable us to express all the roots of a cubic polynomial (see e.g. [Re]). However, these formulas are not simple and so we try to find at least one root simply by guessing it. (School examples are usually chosen so that cubic polynomials have at least one "nice" root – like 0, ± 1, ± 2, etc., and so you have a good chance to guess it.) Substituting successively the numbers 0, ± 1 ± 2, $\ldots$ to Q, we find out that $Q(5) = 0$. Thus, $\alpha = 5$ is the real root of Q. Hence Q is divisible by the factor $(x-5)$. The division gives:

$$Q(x) \,:\, (x-5) \;=\; (2x^3 - 4x^2 - 20x - 50) \,:\, (x-5) \;=\; 2\,(x^2 + 3x + 5).$$

The discriminant of the quadratic polynomial x^2+3x+5 is negative (it equals -11), therefore the polynomial has only complex roots and it has no real linear factors. Thus, the desired decomposition of polynomial Q is:

$$Q(x) \;=\; 2\,(x-5)\,(x^2+3x+5).$$

This corresponds to case IIa) in the previous paragraph IV.4.3.

IV.4.5. The partial fraction decomposition of a rational function with a quadratic polynomial in the denominator.

Suppose that P is a polynomial of the degree less than 2 (i.e. it is either a constant function or it is a linear polynomial). Suppose further that polynomials P and Q are not divisible by the same linear factor. (Otherwise the rational function $\dfrac{P}{Q}$ could be cancelled by this linear factor and the resulting function would be of another, simpler, type.) In this paragraph, we show how it is possible to write $\dfrac{P}{Q}$ in the form of a sum of simpler, so called *partial fractions*. The form of the sum depends on the factors of Q. That is why we successively discuss the cases Ia), Ib) and Ic) from paragraph IV.4.3 and we always show the corresponding decomposition of the rational function $\dfrac{P}{Q}$.

Ia) $\dfrac{P(x)}{Q(x)} \;=\; \dfrac{P(x)}{q_0x^2+q_1x+q_2}$ ..... In this case the rational function $\dfrac{P(x)}{Q(x)}$ is already a partial fraction.

Ib) $\dfrac{P(x)}{Q(x)} \;=\; \dfrac{P(x)}{q_0\,(x-\alpha)^2} \;=\; \dfrac{A_1}{(x-\alpha)} \;+\; \dfrac{A_2}{(x-\alpha)^2}$

Ic) $\dfrac{P(x)}{Q(x)} \;=\; \dfrac{P(x)}{q_0\,(x-\alpha)\,(x-\beta)} \;=\; \dfrac{A}{(x-\alpha)} \;+\; \dfrac{B}{(x-\beta)}$

There are several methods for calculating coefficients A_1 and A_2 (respectively A and B). One of them is shown in the following example.

IV.4.6. Example.

Find the partial fraction decomposition of the rational function

$$\frac{P(x)}{Q(x)} \;=\; \frac{2x-7}{x^2-5x+6}.$$

Polynomial Q can be written as a product of two linear factors as follows: $Q(x) = x^2 - 5x + 6 = (x-2)(x-3)$. Using formula Ic) from the preceding paragraph, we can write:

$$\frac{P(x)}{Q(x)} = \frac{2x-7}{x^2-5x+6} = \frac{A}{x-2} + \frac{B}{x-3}.$$

The sum on the right hand side can be transformed to the common denominator:

$$\frac{A}{x-2} + \frac{B}{x-3} = \frac{A(x-3) + B(x-2)}{(x-2)(x-3)} = \frac{(A+B)x - (3A+2B)}{(x-2)(x-3)}.$$

The last fraction is identical with $\dfrac{P(x)}{Q(x)}$. Comparing the numerators, we obtain: $(A+B)x - (3A+2B) = 2x - 7$. Two polynomials are equal if and only if all their coefficients at the corresponding powers of the independent variable (i.e. x) are equal. In our case, the comparison of the coefficients yields:

$$A + B = 2, \qquad 3A + 2B = -7.$$

This is a system of two linear algebraic equations for two unknowns A and B. Solving this system, we obtain: $A = 3$, $B = -1$. Thus, the decomposition we wished to find is:

$$\frac{2x-7}{x^2-5x+6} = \frac{3}{x-2} - \frac{1}{x-3}.$$

The method, explained in this example, is called "the method of undetermined coefficients". (The undetermined coefficients were A and B.)

IV.4.7. The partial fraction decomposition of a rational function with a cubic polynomial in the denominator.

Suppose that P is a polynomial of the degree less than 3 (i.e. it is either a constant function or a linear polynomial or a quadratic polynomial) and Q is a polynomial of the degree 3. Assume further that polynomials P and Q have no common factor which could be used to cancel the rational function $\dfrac{P}{Q}$. We show how it is possible to decompose $\dfrac{P}{Q}$ to the sum of partial fractions. The form of the partial fractions is closely related to the decomposition of polynomial Q, explained in paragraph IV.4.3. This is why we discuss all four cases IIa), IIb), IIc) and IId) from paragraph IV.4.3 and we always show the corresponding partial fraction decomposition of the rational function $\dfrac{P}{Q}$.

IIa) $$\frac{P(x)}{Q(x)} = \frac{P(x)}{q_0(x-\alpha)(x^2+rx+s)} = \frac{A}{x-\alpha} + \frac{Bx+C}{x^2+rx+s}$$

IIb) $$\frac{P(x)}{Q(x)} = \frac{P(x)}{q_0(x-\alpha)^3} = \frac{A_1}{x-\alpha} + \frac{A_2}{(x-\alpha)^2} + \frac{A_3}{(x-\alpha)^3}$$

IIc) $$\frac{P(x)}{Q(x)} = \frac{P(x)}{q_0(x-\alpha)(x-\beta)^2} = \frac{A}{x-\alpha} + \frac{B_1}{x-\beta} + \frac{B_2}{(x-\beta)^2}$$

IId) $$\frac{P(x)}{Q(x)} = \frac{P(x)}{q_0(x-\alpha)(x-\beta)(x-\gamma)} = \frac{A}{x-\alpha} + \frac{B}{x-\beta} + \frac{C}{x-\gamma}$$

The coefficients A, B, C (respectively A_1, A_2, A_3, respectively A, B_1, B_2) can be evaluated similarly as in example IV.4.6 in particular cases. You can also see this in the following example.

IV.4.8. Example.

Decompose to the sum of partial fractions the rational function

$$\frac{P(x)}{Q(x)} = \frac{10x^2 + 4x}{2x^3 - 4x^2 - 20x - 50}.$$

The decomposition of polynomial Q is: $Q(x) = 2\,(x-5)\,(x^2+3x+5)$ (see example IV.4.4). Using formula IIa) from the previous paragraph, we get:

$$\frac{P(x)}{Q(x)} = \frac{10x^2 + 4x}{2x^3 - 4x^2 - 20x - 50} = \frac{A}{x-5} + \frac{Bx + C}{x^2+3x+5}.$$

Transforming the sum on the right hand side to the same denominator and making simple rearrangements, we obtain:

$$\frac{A}{x-5} + \frac{Bx + C}{x^2+3x+5} = \frac{(A+B)\,x^2 + (3A-5B+C)\,x + (5A-5C)}{(x-5)\,(x^2+3x+5)}.$$

The last fraction is identical with $\frac{P(x)}{Q(x)}$. The comparison of the numerators gives: $(A+B)\,x^2 + (3A-5B+C)\,x + (5A-5C) = 10x^2 + 4x$. If we further compare the coefficients at the same powers of x, we obtain a system of three linear algebraic equations for three unknowns A, B, C:

$$A + B = 5, \qquad 3A - 5B + C = 2, \qquad 5A - 5C = 0.$$

The system has a unique solution: $A = 3$, $B = 2$, $C = 3$. Thus, the desired decomposition is:

$$\frac{10x^2 + 4x}{2x^3 - 4x^2 - 20x - 50} = \frac{3}{x-5} + \frac{2x + 3}{x^2+3x+5}.$$

IV.4.9. The integration of a rational function.

If we have to integrate a rational function $\frac{P}{Q}$ (where Q is at most a cubic polynomial), we follow these steps:

a) If the degree of P is greater than or equal to the degree of Q then we divide P by Q. Suppose that the resulting polynomial is P_1 and the remainder is P_2. Hence $\frac{P}{Q}$ can be written in the form $P_1 + \frac{P_2}{Q}$, where the degree of P_2 is already less than the degree of Q. The integration of polynomial P_1 is already simple and the quotient $\frac{P_2}{Q}$ can be integrated in the way described in the next item.

b) If the degree of P is less than the degree of Q then we decompose the rational function $\frac{P}{Q}$ to the sum of partial fractions and we integrate this sum "term by term". The integration of the two types of partial fractions is explained in paragraphs IV.4.11 and IV.4.12.

IV.4.10. Remark.

Let us discuss in this paragraph a simple case when polynomial Q in the denominator of the rational function $\frac{P}{Q}$ is linear, i.e. its degree equals 1. If the degree of polynomial P equals zero

(i.e. P is a constant) then the rational function $\frac{P}{Q}$ is already a partial fraction. If the degree of P is greater than zero, then dividing P by Q we transform $\frac{P}{Q}$ to the form $P_1 + \frac{P_2}{Q}$, where P_1 is a polynomial and P_2 is a constant. $\frac{P_2}{Q}$ is a partial fraction. The integration of P_1 is obvious and the integration of $\frac{P_2}{Q}$ is described in paragraph IV.4.11.

Thus, the case when the degree polynomial Q equals one is simple and therefore we do not further pay more attention to this case.

IV.4.11. The integration of partial fractions of the type $\frac{A}{(x-\alpha)^n}$.

The substitution is applied directly in the computation:

$$\int \frac{A}{(x-\alpha)^n}\,\mathrm{d}x = \left|\begin{matrix} t = x-\alpha \\ \mathrm{d}t = \mathrm{d}x \end{matrix}\right| = \int \frac{A}{t^n}\,\mathrm{d}t.$$

The transformed integral can be evaluated by means of formula b) from paragraph IV.1.9 (for $n \neq 1$) or by means of formula k) from the same paragraph (for $n = 1$). Replacing t by $x-\alpha$ in the result, we obtain the expression of the integral $\int \frac{A}{(x-\alpha)^n}\,\mathrm{d}x$ on two intervals: $(-\infty;\alpha)$ and $(\alpha;+\infty)$.

IV.4.12. The integration of partial fractions of the type $\frac{Bx+C}{x^2+rx+s}$.

Suppose that the quadratic polynomial x^2+rx+s cannot be decomposed in the real domain (otherwise $\frac{Bx+C}{x^2+rx+s}$ would not be a partial fraction). Bear in mind that we can recognize this for example in such a way that the discriminant of the polynomial x^2+rx+s is negative. Before we explain the integration of a general function $\frac{Bx+C}{x^2+rx+s}$, we will illustrate the procedure on a concrete example. Since C already denotes one of the coefficients in the partial fraction, the constant of integration will be denoted by c. Completing the square, we obtain

$$\int \frac{2x+3}{x^2+2x+5}\,\mathrm{d}x = \int \frac{2x+3}{(x+1)^2+4}\,\mathrm{d}x = \left|\begin{matrix} u = x+1 \\ x = u-1, \quad \mathrm{d}x = \mathrm{d}u \end{matrix}\right| = \int \frac{2u+1}{u^2+4}\,\mathrm{d}u =$$

$$= \left|\begin{matrix} u = 2v \\ \mathrm{d}u = 2\mathrm{d}v \end{matrix}\right| = \frac{1}{2}\int \frac{4v+1}{v^2+1}\,\mathrm{d}v = \int \frac{2v}{v^2+1}\,\mathrm{d}v + \frac{1}{2}\int \frac{1}{v^2+1}\,\mathrm{d}v =$$

$$= \ln(v^2+1) + \frac{1}{2}\arctan v + c = \left|v = \frac{u}{2} = \frac{x+1}{2}\right| =$$

$$= \ln\left(\frac{(x+1)^2}{4}+1\right) + \frac{1}{2}\arctan\left(\frac{x+1}{2}\right) + c.$$

Here $(v^2+1)' = 2v$, so the first of the two integrals is a logarithm.

The partial fraction $\frac{2x+3}{x^2+2x+5}$ is a continuous function on the interval $(-\infty;+\infty)$, hence its antiderivative and the indefinite integral exist on this interval.

Let us now return to the general partial fraction $\dfrac{Bx+C}{x^2+rx+s}$. The fraction can be written in the form

$$\frac{Bx+C}{x^2+rx+s} = R\,\frac{2x+r}{x^2+rx+s} + S\,\frac{1}{x^2+rx+s}.$$

The values of R and S can be specified by comparing the coefficients at the same powers of x, similarly as the values of A, B and C in paragraph IV.4.8. The first fraction on the right hand side is of the type $\dfrac{f'}{f}$ and so its integral is:

$$\int \frac{2x+r}{x^2+rx+s}\,\mathrm{d}x = \left|\begin{matrix} u = x^2+rx+s \\ \mathrm{d}u = (2x+r)\,\mathrm{d}x \end{matrix}\right| = \int \frac{1}{u}\,\mathrm{d}u = \ln|x^2+rx+s| + c, \qquad x \in \mathbb{R}.$$

The second fraction can be integrated in the way which was already explained on a concrete example in this paragraph. We obtain:

$$\int \frac{1}{x^2+rx+s}\,\mathrm{d}x = \left|\begin{matrix} t = \dfrac{2x+r}{\sqrt{-D}} \\ \mathrm{d}x = \dfrac{\sqrt{-D}}{2}\,\mathrm{d}t \end{matrix}\right| = \frac{2}{\sqrt{-D}}\int \frac{1}{1+t^2}\,\mathrm{d}t =$$

$$= \frac{2}{\sqrt{-D}}\arctan\left(\frac{2x+r}{\sqrt{-D}}\right) + c, \qquad x \in \mathbb{R}.$$

(We remind that the discriminant $D = r^2 - 4s$ is supposed to be negative in this paragraph.) You can either remember the procedure of integration (illustrated on the concrete example) or the above final formula.

IV.4.13. Problems.

Find the integrals

a) $\displaystyle\int \frac{x^3-3x^2+2}{2x+1}\,\mathrm{d}x$,

b) $\displaystyle\int \frac{x}{x^3-1}\,\mathrm{d}x$,

c) $\displaystyle\int \frac{x^3+1}{x^3-x^2}\,\mathrm{d}x$,

d) $\displaystyle\int \frac{5x-4}{x^3-x^2-2x}\,\mathrm{d}x$.

Results:

a) $\dfrac{1}{6}x - \dfrac{7}{8}x^2 + \dfrac{7}{8}x + \dfrac{9}{16}\ln|2x+1| + c$ (for $x \in \left(-\infty; -\dfrac{1}{2}\right)$ and $x \in \left(-\dfrac{1}{2}; +\infty\right)$)

b) $\dfrac{1}{3}\ln\dfrac{|x-1|}{\sqrt{x^2+x+1}} + \dfrac{1}{\sqrt{3}}\arctan\dfrac{2x+1}{\sqrt{3}} + c$ (for $x \in (-\infty; 1)$ and $x \in (1; +\infty)$)

c) $x + \dfrac{1}{x} + \ln\dfrac{(x-1)^2}{|x|} + c$ (for $x \in (-\infty; 0)$, $x \in (0; 1)$ and $x \in (1; +\infty)$)

d) $\ln\dfrac{x^2\,|x-2|}{|x+1|^3} + c$ (for $x \in (-\infty; -1)$, $x \in (-1; 0)$, $x \in (0; 2)$ and $x \in (2; +\infty)$).

IV.5. Integration of functions of the type $\sin^n x \cdot \cos^m x$

IV.5.1. The integration of functions $\sin^n x \cdot \cos^m x$, where at least one of the numbers m, n is odd.

Let m, n be such integers that at least one of them is odd. Suppose for example that m is odd. (If m is even and n is odd then the procedure is analogous.) m can be expressed in the form $m = 2k+1$ where k is an appropriate integer. Then the function $\sin^n x \cdot \cos^m x$ can be written in the form:

$$\begin{aligned}\sin^n x \cdot \cos^m x &= \sin^n x \cdot \cos^{2k+1} x = \sin^n x \cdot (\cos^2 x)^k \cdot \cos x = \\ &= \sin^n x \cdot (1-\sin^2 x)^k \cdot \cos x.\end{aligned}$$

The integral $\int \sin^n x \cdot \cos^m x \, \mathrm{d}x$ can be evaluated by means of the substitution $\sin x = t$:

$$\int \sin^n x \cos^m x \, \mathrm{d}x = \int \sin^n x (1-\sin^2 x)^k \cos x \, \mathrm{d}x = \left| \begin{matrix} t = \sin x \\ \mathrm{d}t = \cos x \, \mathrm{d}x \end{matrix} \right| =$$
$$= \int t^n (1-t^2)^k \, \mathrm{d}t.$$

IV.5.2. Example.

$$\int \sin^2 x \cos^5 x \, \mathrm{d}x = \int \sin^2 x (1-\sin^2 x)^2 \cos x \, \mathrm{d}x = \left| \begin{matrix} t = \sin x \\ \mathrm{d}t = \cos x \, \mathrm{d}x \end{matrix} \right| = \int t^2 (1-t^2)^2 \, \mathrm{d}t =$$
$$= \int (t^2 - 2t^4 + t^6) \, \mathrm{d}t = \frac{t^3}{3} - \frac{2t^5}{5} + \frac{t^7}{7} + C = \frac{\sin^3 x}{3} - \frac{2}{5}\sin^5 x + \frac{\sin^7 x}{7} + C, \qquad x \in \mathbb{R}.$$

IV.5.3. Example.

$$\int \frac{1}{\sin x} \, \mathrm{d}x = \int \frac{\sin x}{1-\cos^2 x} \, \mathrm{d}x = \left| \begin{matrix} t = \cos x \\ \mathrm{d}t = -\sin x \, \mathrm{d}x \end{matrix} \right| = -\int \frac{1}{1-t^2} \, \mathrm{d}t =$$
$$= -\frac{1}{2}\int \frac{1}{1-t} \, \mathrm{d}t - \frac{1}{2}\int \frac{1}{1+t} \, \mathrm{d}t = \frac{1}{2}\ln|1-t| - \frac{1}{2}\ln|1+t| + C =$$
$$= \frac{1}{2}\ln|1-\cos x| - \frac{1}{2}\ln|1+\cos x| + C = \frac{1}{2}\ln\left|\frac{1-\cos x}{1+\cos x}\right| + C.$$

The result holds on each interval $(k\pi; (k+1)\pi)$, where $k \in \mathbb{Z}$.

IV.5.4. The integration of functions $\sin^n x \cdot \cos^m x$, both the numbers m, n are even.

The integrand can be rewritten by means of the formulas

$$\sin^2 x = \frac{1-\cos(2x)}{2}, \qquad \cos^2 x = \frac{1+\cos(2x)}{2}.$$

We also use the substitution $2x = t$. Thus, we decompose the integrand to the sum of functions "of type IV.5.1" (i.e. functions satisfying the assumptions from paragraph IV.5.1) and functions "of type IV.5.4" (i.e. functions satisfying the assumptions from this paragraph). The integration of the functions "of type IV.5.1" is already known. The functions "of type IV.5.4" can be rewritten again by means of the above formulas. We can continue doing this until all remaining functions are "of type IV.5.1".

IV.5.5. Example.

$$\int \sin^4 x \,\mathrm{d}x = \int \left(\frac{1-\cos(2x)}{2}\right)^2 \mathrm{d}x = \int \frac{1-2\cos(2x)+\cos^2(2x)}{4}\,\mathrm{d}x = \left|\begin{matrix} t = 2x \\ \mathrm{d}x = \frac{1}{2}\,\mathrm{d}t \end{matrix}\right| =$$

$$= \frac{1}{8}\int (1-2\cos t+\cos^2 t)\,\mathrm{d}t = \frac{t}{8} - \frac{\sin t}{4} + \frac{1}{16}\int (1+\cos(2t))\,\mathrm{d}t = \left|\begin{matrix} u = 2t \\ \mathrm{d}t = \frac{1}{2}\,\mathrm{d}u \end{matrix}\right| =$$

$$= \frac{t}{8} - \frac{\sin t}{4} + \frac{1}{32}\int (1+\cos u)\,\mathrm{d}u = \frac{t}{8} - \frac{\sin t}{4} + \frac{u}{32} + \frac{\sin u}{32} + C = \left|\begin{matrix} t = 2x \\ u = 2t = 4x \end{matrix}\right| =$$

$$= \frac{3}{8}x - \frac{1}{4}\sin(2x) + \frac{1}{32}\sin(4x) + C, \quad x \in \mathbb{R}.$$

IV.5.6. Problems.

Evaluate the integrals

$$\text{a)} \int \sin^7 x \,\mathrm{d}x, \qquad \text{b)} \int \cos^6 x \,\mathrm{d}x, \qquad \text{c)} \int \frac{\cos^3 x}{\sin^2 x}\,\mathrm{d}x.$$

Results:

a) $-\cos x + \cos^3 x - \frac{3}{5}\cos^5 x + \frac{1}{7}\cos^7 x + C$ (for $x \in (-\infty; +\infty)$)

b) $\frac{5}{16}x + \frac{1}{4}\sin(2x) + \frac{3}{64}\sin(4x) - \frac{1}{48}\sin^3(6x) + C$ (for $x \in (-\infty; +\infty)$)

c) $-\frac{1}{\sin x} - \sin x + C$ (for $x \in (k\pi; (k+1)\pi)$; k is an arbitrary integer).

IV.6. Integration of some other types of functions

IV.6.1. Rational function of two variables.

We shall several times use the symbol $R(u, v)$ in this section. It will denote a so called *rational function of two variables* u and v, i.e. a ratio of two polynomials $P(u, v)$ and $Q(u, v)$.

You will learn more about functions of several variables in the second term in Mathematics. For our purposes, it will be sufficient to know that a polynomial of two variables u and v is a sum of a finite number of terms, each of which has the form $k\,u^m\,v^n$ where k is a real constant and m, n are nonnegative integers.

Examples of rational functions of two variables are:

$$R(u,v) = \frac{2\,u + v^2 + 2}{u - 1}, \qquad R(u,v) = \frac{1}{2+u}, \qquad R(u,v) = \frac{u - v}{u^2 + v^2}.$$

IV.6.2.* Integrals of the type $\int R(\sin x,\, \cos x)\,\mathrm{d}x$.

Integrals of this type can be transformed by means of the substitution

$$t = \tan\left(\frac{x}{2}\right) \tag{IV.6.1}$$

to integrals of rational functions. Let us explain this in greater detail. Using the known formulas for trigonometric functions, we obtain:

$$\cos^2\left(\frac{x}{2}\right) = \frac{1}{1+\tan^2\left(\frac{x}{2}\right)}, \qquad \sin^2\left(\frac{x}{2}\right) = 1 - \cos^2\left(\frac{x}{2}\right) = \frac{\tan^2\left(\frac{x}{2}\right)}{1+\tan^2\left(\frac{x}{2}\right)},$$

$$\cos x = \cos^2\left(\frac{x}{2}\right) - \sin^2\left(\frac{x}{2}\right) = \frac{1-\tan^2\left(\frac{x}{2}\right)}{1+\tan^2\left(\frac{x}{2}\right)} = \frac{1-t^2}{1+t^2},$$

$$\sin x = 2\sin\left(\frac{x}{2}\right)\cdot\cos\left(\frac{x}{2}\right) = \frac{2\tan\left(\frac{x}{2}\right)}{1+\tan^2\left(\frac{x}{2}\right)} = \frac{2\,t}{1+t^2}.$$

Differentiating the equation $\tan\left(\frac{x}{2}\right) = t$, we further get:

$$\frac{1}{2\cos^2\left(\frac{x}{2}\right)}\,\mathrm{d}x = \mathrm{d}t, \qquad \text{i.e.} \quad \frac{1}{2}\left(1+\tan^2\left(\frac{x}{2}\right)\right)\mathrm{d}x = \mathrm{d}t.$$

This implies: $\mathrm{d}x = \dfrac{2}{1+\tan^2\left(\frac{x}{2}\right)}\,\mathrm{d}t = \dfrac{2}{1+t^2}\,\mathrm{d}t.$

If we now replace $\sin x$, $\cos x$ and $\mathrm{d}x$ in the integral $\int R(\sin x,\,\cos x)\,\mathrm{d}x$ by the above expressions, we obtain an integral of a rational function (of variable t). Integrals of rational functions are dealt with in Section IV.4 in this text.

The substitution (IV.6.1) can be used only on an interval where the function $\tan\left(\frac{x}{2}\right)$ is defined, i.e. on each interval $(-\pi + 2k\pi;\, \pi + 2k\pi)$, where k is an integer. (Sometimes we evaluate the integral on an even smaller interval. This depends on a concrete form of an integrated rational function.)

IV.6.3.* Example.

$$\int \frac{1-\sin x}{1+\cos x}\,\mathrm{d}x = \left|\begin{array}{ll} t = \tan\left(\frac{x}{2}\right), & \mathrm{d}x = \frac{2}{1+t^2}\,\mathrm{d}t \\ \sin x = \frac{2t}{1+t^2}, & \cos x = \frac{1-t^2}{1+t^2} \end{array}\right| = \int \frac{1-\frac{2t}{1+t^2}}{1+\frac{1-t^2}{1+t^2}}\,\frac{2}{1+t^2}\,\mathrm{d}t =$$

$$= \int \frac{1-2t+t^2}{1+t^2}\,\mathrm{d}t = \int\left(1-\frac{2t}{1+t^2}\right)\mathrm{d}t = t - \ln(1+t^2) + C =$$

$$= \tan\left(\frac{x}{2}\right) - \ln\left(1+\tan^2\left(\frac{x}{2}\right)\right) + C.$$

The integrand is a continuous function on each of the intervals $(-\pi + 2k\pi;\ \pi + 2k\pi)$, where k is an integer. Hence a primitive function and the indefinite integral exist on each of these intervals. The intervals are identical with those where we can use substitution (IV.6.1). Thus, the obtained expression of the integral is valid on each of the intervals $(-\pi + 2k\pi;\ \pi + 2k\pi)$.

IV.6.4. Remark.

Before starting to compute any integral which is not quite simple, we recommend that you think at least a short while over the possibilities you have and then to try to choose the simplest one. For instance, the integral

$$\int \frac{\cos x}{1 - \sin x}\,\mathrm{d}x$$

is an integral "of type IV.6.2", so it can be evaluated by means of substitution (IV.6.1). However, there also exists another substitution which is much simpler. Can you discover it?

IV.6.5. Integrals of the type $\displaystyle\int R\left(x,\ \sqrt{\frac{ax+b}{cx+d}}\right)\,\mathrm{d}x$**.**

Suppose that $ad - bc \neq 0$. (This inequality guarantees that the linear polynomials $ax + b$ and $cx + d$ have no common factor. Otherwise the fraction $\dfrac{ax+b}{cx+d}$ could be cancelled by this factor and the result would be a constant function.) We can use the substitution

(IV.6.2) $$t = \sqrt{\frac{ax+b}{cx+d}}.$$

This can be used to express x and $\mathrm{d}x$. Substituting this to the integral we want to compute, we obtain an integral of a rational function (of variable t). The procedure is shown in example IV.6.6. Substitution (IV.6.2) can be used on each interval which is a subset of the domain of the integrand. In addition to others, the following inequality must hold on such an interval: $\dfrac{ax+b}{cx+d} \geq 0$.

The integral $\int R(x, \sqrt{ax+b})\,\mathrm{d}x$ is a special case of the above integral. (It corresponds to the values $c = 0$, $d = 1$.) To evaluate it, we can thus use the substitution $t = \sqrt{ax+b}$.

IV.6.6. Example.

Evaluate $\displaystyle\int \frac{1}{x}\sqrt{\frac{1-x}{1+x}}\,\mathrm{d}x$.

The fraction under the square root must be nonnegative. Solving the inequality $\dfrac{1-x}{1+x} \geq 0$, we get the condition: $x \in (-1; 1\rangle$. Moreover, there is also a function $\dfrac{1}{x}$, involved in the integral. This implies another condition: $x \neq 0$. Hence the integrand is defined on intervals $(-1; 0)$ and $(0; 1\rangle$. Since it is continuous on each of these intervals, an antiderivative and the indefinite integral exist on each of them. The computation is

$$\int \frac{1}{x}\sqrt{\frac{1-x}{1+x}}\,\mathrm{d}x = \left|\begin{array}{l} t = \sqrt{\dfrac{1-x}{1+x}} \\ x = \dfrac{1-t^2}{1+t^2} \\ \mathrm{d}x = -\dfrac{4t}{(1+t^2)^2}\,\mathrm{d}t \end{array}\right| = \int \frac{t^2+1}{t^2-1}\,t\,\frac{4t}{(t^2+1)^2}\,\mathrm{d}t =$$

$$= \int \frac{4t^2}{(t-1)(t+1)(t^2+1)}\,\mathrm{d}t = \int \frac{\mathrm{d}t}{t-1} - \int \frac{\mathrm{d}t}{t+1} + 2\int \frac{\mathrm{d}t}{t^2+1} =$$
$$= \ln\left|\frac{t-1}{t+1}\right| + 2\arctan t + C = \ln\left|\frac{\sqrt{1+x}-\sqrt{1-x}}{\sqrt{1+x}+\sqrt{1-x}}\right| + 2\arctan\sqrt{\frac{1-x}{1+x}} + C$$

(for $x \in (-1;0)$ or for $x \in (0;1\rangle$). The rational function of the variable t, which we obtained after the application of the substitution, contains the polynomial of the fourth degree in the denominator. This goes beyond the scope of Section IV.4. Nevertheless, its decomposition is quite analogous:

$$\frac{4\,t^2}{(t-1)(t+1)(t^2+1)} = \frac{A}{t-1} + \frac{B}{t+1} + \frac{C_1 t + C_2}{t^2+1}.$$

The coefficients A, B, C_1 and C_2 can be evaluated by the same method as the numbers A, B and C in paragraph IV.4.8. We obtain: $A = 1$, $B = -1$, $C_1 = 0$, $C_2 = 2$.

IV.6.7.* Integrals of the type $\int R(x, \sqrt{ax^2+bx+c})\,\mathrm{d}x$.

We suppose that $a \neq 0$, otherwise the integral would be a special case of the integral solved in paragraph IV.6.4. We shall distinguish between two cases:

a) $\underline{a < 0}$ Since the term under the square root must be nonnegative, the quadratic polynomial ax^2+bx+c must have two different real roots α, β. Let us choose their denotation in such a way that $\alpha < \beta$. Then $ax^2+bx+c \geq 0$ for $x \in \langle\alpha;\beta\rangle$. For those $x \in (\alpha;\beta\rangle$, we can write:

$$\sqrt{ax^2+bx+c} = \sqrt{a\,(x-\alpha)(x-\beta)} = \sqrt{-a\,(x-\alpha)(\beta-x)} =$$
$$= \sqrt{-a}\,(x-\alpha)\sqrt{\frac{\beta-x}{x-\alpha}}.$$

The integral of this function is the integral treated in paragraph IV.6.4. It can be transformed to the integral of a rational function by means of the substitution (IV.6.2), i.e.

$$t = \sqrt{\frac{\beta-x}{x-\alpha}}.$$

b) $\underline{a > 0}$ In this case, we can use a so called Euler substitution

(IV.6.3) $$\sqrt{ax^2+bx+c} = t + \sqrt{a}\,x.$$

Raising the two sides of the equality (IV.6.3) to the second power, we can express x and $\mathrm{d}x$. The Euler substitution transforms the integral we wish to evaluate to an integral of a rational function of variable t. The procedure is shown in example IV.6.8. The substitution (IV.6.3) can be used on each interval which is a subset of the domain of the integrand.

IV.6.8.* Example.

Evaluate $\displaystyle\int \frac{\mathrm{d}x}{x\sqrt{x^2+2x-1}}$.

Since the quadratic polynomial x^2+2x-1 is under the square root in the denominator, the inequality $x^2+2x-1 > 0$ must be satisfied in the domain of the integrand. Moreover, the integrand also contains x in the denominator. This gives the condition $x \neq 0$. It can easily be verified that maximum intervals where the two conditions are fulfilled are $(-\infty;-1-\sqrt{2})$ and $(-1+\sqrt{2};+\infty)$. The integrand is continuous on each of them, hence a primitive function and the indefinite integral also exist on each of them.

The Euler substitution gives

$$\int \frac{\mathrm{d}x}{x\sqrt{x^2+2x-1}} = \left| \begin{array}{l} \sqrt{x^2+2x-1} = x+t \\ x = \dfrac{t^2+1}{2(1-t)} \\ \mathrm{d}x = \dfrac{-t^2+2t+1}{2(1-t)^2}\,\mathrm{d}t \end{array} \right| = \int \frac{2(1-t)}{t^2+1} \, \frac{1}{\frac{t^2+1}{2(1-t)}+t} \, \frac{-t^2+2t+1}{2(1-t)^2}\,\mathrm{d}t =$$

$$= \int \frac{2}{1+t^2}\,\mathrm{d}t = 2\arctan t + C = 2\arctan\left(\sqrt{x^2+2x-1}-x\right)+C.$$

The result holds for $x \in (-\infty; -1-\sqrt{2})$ or $x \in (-1+\sqrt{2}; +\infty)$.

IV.7.* Separable differential equations

Differential equations form an extensive part of mathematics, with a lot of applications in engineering, physics, chemistry and other scientific disciplines. We introduce some basic notions and show the so called "method of separation of variables" for the solution. Since the method is based mainly on integration, we present it as a part of Chapter IV dealing with indefinite integrals, and we regard it as an illustrative example of applications of theory of indefinite integrals.

You will learn more about differential equations in Mathematics III in your third term of study.

IV.7.1. The differential equation of the 1st order.

The equation

(IV.7.1) $$F(x, y, y') = 0$$

is called the *differential equation of the 1st order*. The function $y(x)$ is the unknown in equation (IV.7.1).

A function of three variables x, y and y' appears on the right hand side of equation (IV.7.1). As was mentioned in paragraph IV.6.1, you will learn more about such functions in Mathematics II in the second term. In this section, it will be sufficient for our purposes to regard $F(x, y, y')$ as an algebraic expression which can contain both x, y and y'. The example is $F(x, y, y') = y' - 2y\sqrt{x^2-1}$.

Equation (IV.7.1) is called "differential" because it contains not only the unknown function y, but also its derivative. The denotation "of the 1st order" means that the highest derivative of the unknown function appearing in the equation is the derivative of the 1st order.

IV.7.2. A solution, a maximum solution.

The *solution* of differential equation (IV.7.1) in interval I is a function y which satisfies equation (IV.7.1) at each point of interval I. This means that if I is an interval with end points a, b ($a < b$,) then

a) $y'(x) = f(x, y(x))$ for all $x \in (a; b)$,
b) $y'_+(a) = f(a, y(a))$, if a belongs to interval I,
c) $y'_-(b) = f(b, y(b))$, if b belongs to interval I.

If y is a solution of differential equation (IV.7.1) in interval I, z is a solution of the same equation in interval J, $I \subset J$, $I \neq J$ and $y(x) = z(x)$ for all $x \in I$, then the solution z is called the *extension* of the solution y. A solution which cannot be extended any more (i.e. no extension of it exists) is called the *maximum solution*.

IV.7.3. Example.

The function

$$y(x) = 2\sqrt{x-1} \qquad (\text{for } x \in (2;+\infty)\,)$$

is a solution of the differential equation

$$y' = \frac{2}{y} \tag{IV.7.2}$$

in interval $(2;+\infty)$. It is not a maximum solution, because the function

$$z(x) = 2\sqrt{x-1} \qquad (\text{for } x \in (1;+\infty)\,)$$

is the extension of the solution y. Function z is already the maximum solution, because as a solution of equation (IV.7.2) it cannot be extended any more.

IV.7.4. Remark.

Function z from the preceding example is not the only maximum solution of equation (IV.7.2). It can easily be verified that if A is an arbitrary real number then the functions

$$u(x) = 2\sqrt{x-A} \qquad (\text{for } x \in (A;+\infty)\,),$$
$$v(x) = -2\sqrt{x-A} \qquad (\text{for } x \in (A;+\infty)\,)$$

are also the maximum solutions of (IV.7.2). Thus, equation (IV.7.2) has infinitely many maximum solutions. (Sketch their graphs for yourself.)

Choose for instance $x_0 = 5$ and look at the values of maximum solutions of equation (IV.7.2) at this point. If $A \geq 5$ then point x_0 does not belong to the domain of u and v. If $A = 4$, $A = 3$, etc., we obtain:

$$A = 4 \ \ldots\ u(5) = 2,\ v(5) = -2, \qquad A = 3 \ \ldots\ u(5) = 2\sqrt{2},\ v(5) = -2\sqrt{2}, \qquad \text{etc.}$$

We can see that we can distinguish between different maximum solutions by means of their values at a chosen point x_0. (This is not a general rule, but it is true in the case of differential equation (IV.7.2).) When solving a differential equation, it is therefore natural to seek for a solution which takes on a chosen (or a given) value y_0 at a chosen (or a given) point x_0.

IV.7.5. The Cauchy problem.

The problem of finding a solution y of differential equation (IV.7.1) which satisfies the condition $y(x_0) = y_0$ is called the *Cauchy problem* for differential equation (IV.7.1).

IV.7.6. Remark.

The condition $y(x_0) = y_0$ is called the *initial condition* for the following reason: Differential equations can be used for modelling processes running in time and then the independent variable x plays the role of time. Thus, we often solve problems of the following type: *What is the behavior of the process described by equation* (IV.7.1) *at times* $x > x_0$, *if we know its state at an initial time instant* x_0*?* Certainly, we can also ask what preceded the state at time x_0, i.e. we can study the behavior of the process at times $x < x_0$. However, problems of this type are not very frequent (or were not so frequent in the past when the terminology was created).

IV.7.7. Separable differential equations, solution by separation of variables.

In the rest of this chapter, we will deal with a special case of differential equation (IV.7.1), i.e. the case when the right hand side is equal to the product of a function depending only on x and a function depending only on y:

(IV.7.3) $$y' = g(x) \cdot h(y).$$

Differential equation (IV.7.3) is called a *differential equation with separable variables* or shortly a *separable differential equation*.

Assume that function $g(x)$ is continuous on interval I and function $h(y)$ is continuous on interval J. Differential equation (IV.7.3) will be solved for $x \in I$.

Denote by M a set of those points y^* from interval J, where $h(y^*) = 0$. We can easily check (by the substitution to equation (IV.7.3)) that the constant function defined by the equation $y(x) = y^*$ (for all $x \in I$) is a solution of equation (IV.7.3) on interval I. Thus, the computation of zero points of function $h(y)$ in interval J leads to constant solutions of equation (IV.7.3).

Suppose further that interval J was chosen so that it contains no zero points of function $h(y)$, i.e. $h(y) \neq 0$ for all $y \in J$. Hence we can divide equation (IV.7.3) by $h(y)$. Moreover, we write the derivative y' in the form $\frac{\mathrm{d}y}{\mathrm{d}x}$ and we multiply equation (IV.7.3) formally by $\mathrm{d}x$. We obtain:

$$\frac{\mathrm{d}y}{h(y)} = g(x)\,\mathrm{d}x.$$

The left hand side of the above equation depends only on y and the right hand side depends only on x. Thus, we have separated the variables. This fact provided the name for the whole method that we are just explaining: the *method of separation of variables*. Further, we add signs of integrals in front of the left hand side as well as in front of the right hand side and we compute the obtained integrals. If $H(y)$ is a primitive function to $\frac{1}{h(y)}$ on interval J and $G(x)$ is a primitive function to $g(x)$ on interval I, then the evaluation of the indefinite integrals leads to the equation:

$$H(y) = G(x) + C.$$

(Constants which arise on both the sides can be coupled to one which can be written for instance on the right hand side only.) The function $H(y)$ is strictly monotone on interval J because it has a continuous and non–zero derivative $\frac{1}{h(y)}$. Therefore an inverse function to $H(y)$ in interval J exists. Using the inverse function, we get:

(IV.7.4) $$y = H_{-1}(G(x) + C).$$

Function y is a solution of differential equation (IV.7.3) in each interval $I_1 \subset I$ such that $G(x) + C$ belongs to the domain of function H_{-1} for all $x \in I_1$.

The explained procedure is not quite correct from the formal point of view because we separate the "non–separable" symbol $\frac{\mathrm{d}y}{\mathrm{d}x}$ denoting the derivative of function y, we treat it as a fraction and we multiply equation (IV.7.3) by the "auxiliary" term $\mathrm{d}x$. Nevertheless, this procedure leads to good results and you can easily remember it.

Different constants C in formula (IV.7.4) lead to different solutions of differential equation (IV.7.3). Although it does not hold generally, in many cases formula (IV.7.4) involves all solutions of equation (IV.7.3) on interval I. (This means that every solution of equation (IV.7.3) on interval I can be represented by formula (IV.7.4) for some appropriate $C \in \mathbb{R}$.) For this reason it is often said that formula (IV.7.4) represents a so called *general solution* of differential equation (IV.7.3) on interval I. To distinguish between the general solution and a concrete solution (corresponding to a concrete value of the constant C in (IV.7.4)), the latter is called a *particular solution*.

IV.7.8. Remark.

Note that the problem of evaluating the indefinite integral $\int g(x)\,\mathrm{d}x$ in interval I can also be formulated as follows: *Find all solutions of the differential equation* $y' = g(x)$ *on interval* I.

IV.7.9. Example.

Find a maximum solution of the Cauchy problem given by the differential equation $y' = 5xy$ and by the initial condition $y(0) = 2$.

The right hand side is the product of the function $g(x) = 5x$ and the function $h(y) = y$. Both the functions are continuous on the interval $(-\infty; +\infty)$. Solving the equation $h(y) = 0$, we obtain the unique zero point of the function $h(y)$: $y^* = 0$. Thus the differential equation $y' = 5xy$ has only one constant solution: $y(x) = 0$ which is defined in the whole interval $I = (-\infty; +\infty)$.

The function $h(y)$ is different from zero in the intervals $J_1 = (-\infty; 0)$ and $J_2 = (0; +\infty)$. We can therefore take y from J_1 or from J_2 and we can solve the differential equation by separation of variables. We write the derivative y' in the form $\dfrac{\mathrm{d}y}{\mathrm{d}x}$ and we multiply the equation formally by $\mathrm{d}x$. Further, we divide the equation by y. This leads to an equation with separated variables x and y. Then we add the signs of integrals. Successively, we obtain:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 5xy, \qquad \frac{\mathrm{d}y}{y} = 5x\,\mathrm{d}x, \qquad \int \frac{\mathrm{d}y}{y} = \int 5x\,\mathrm{d}x.$$

Evaluating the integrals on the left hand side and the right hand side of the equation, we get:

$$\ln|y| = \frac{5}{2}x^2 + C.$$

C can be written in the form $\ln|K|$, where $K \in \mathbb{R}$, $K \neq 0$. The last equation gives:

$$|y| = \exp\left(\frac{5}{2}x^2 + \ln|K|\right) = |K \cdot \exp\left(\frac{5}{2}x^2\right)|.$$

This yields: $y = \pm K \cdot \exp\left(\dfrac{5}{2}x^2\right)$. Since K can take on positive as well as negative values, the symbol $\pm$ can be omitted.

We have obtained the solution in the form $y = K \cdot \exp\left(\dfrac{5}{2}x^2\right)$ under the assumption that $K \neq 0$. However, for $K = 0$ the formula gives the already discovered zero solution. Thus, we can see that the condition $K \neq 0$ was connected only with the method we used for solving the equation, but it need not be taken into account in the final formula expressing the solution. Thus, the general solution of the differential equation is:

$$y = K \cdot \exp\left(\frac{5}{2}x^2\right); \qquad K \in \mathbb{R}.$$

Now we use the general solution and we write it with the values $x = 0$ and $y = 2$, following from the initial condition. We obtain: $2 = K$. Substituting this value of K back to the general solution, we obtain the desired solution of the Cauchy problem:

$$y = 2\exp\left(\frac{5}{2}x^2\right); \qquad x \in (-\infty; +\infty).$$

IV.7.10. Remark.

Differential equations we have to solve are rarely written in the form (IV.7.3). For instance, equation $xy' - y = 0$ does not have this form although the equation can easily be transformed

to the form (IV.7.3): $y' = \dfrac{y}{x}$.

If we solve the equation $y' = \dfrac{y}{x}$ by the separation of variables, we obtain the general form of a maximum solution: $y = K \cdot x$ for $x \in (-\infty; 0)$ or for $x \in (0; +\infty)$ ($K \in \mathbb{R}$). The point $x = 0$ must be omitted because zero cannot appear in the denominator of the fraction on the right hand side.

Let us now return to the original equation $xy' - y = 0$. Here we have no reason to exclude the point $x = 0$ from the domain of the solution and so the general form of a maximum solution is $y = K \cdot x$ for $x \in (-\infty; +\infty)$ ($K \in \mathbb{R}$).

You can derive the conclusion from the above example: If the differential equation does not have the form (IV.7.3) and you transform it to this form by means of arithmetic operations, you must pay attention to the domain of the solution.

IV.7.11. Example.

A body whose mass is m moves on a straight line. The resistance of the environment is proportional to the speed of motion, the coefficient of proportionality is k. Moreover, a constant force f, parallel with the motion, acts onto the body. The known velocity at the time $t = 0$ is v_0. What is the velocity $v(t)$ of the motion at a general time t?

Solution: Let $a(t)$ be the acceleration of the body. Second Newton's law says that

(IV.7.5) $$m\,a(t) \;=\; F(t)$$

where $F(t)$ is a total force acting on the body, parallel with the direction of the motion. This force can be expressed: $F(t) = f - k\,v(t)$. (The term $k\,v(t)$ must be written with the "$-$" sign because the resistance, e.g. the friction, acts opposite the direction of the motion.)

You already know, e.g. from paragraphs III.4.2 and III.4.6, that the velocity $v(t)$ is the derivative of the position $s(t)$. Analogously, the acceleration $a(t)$ can be defined to be the derivative of the velocity $v(t)$. If we denote the derivative with respect to time, as it is usual in physics, by the dot, we can write: $a(t) = \dot{v}(t)$. Substituting this and the expression of the force $F(t)$ to equation (IV.7.5), we obtain the differential equation

(IV.7.6) $$m\,\dot{v}(t) \;=\; f - k\,v(t).$$

This is the differential equation with separable variables. You can easily verify that the general solution is

(IV.7.7) $$v(t) \;=\; \frac{f}{k} \;-\; \frac{C}{k}\,\mathrm{e}^{-\frac{kt}{m}}.$$

Using the formula (IV.7.7) with the concrete values $t = 0$ and $v(0) = v_0$, we can obtain the equation which enables to evaluate C: $C = f - kv_0$. Substituting this C back to the general solution (IV.7.7), we get the desired particular solution:

$$v(t) = \frac{f}{k} + \left(v_0 - \frac{f}{k}\right)\mathrm{e}^{-\frac{kt}{m}}.$$

IV.7.12. Problems.

Find maximum solutions of the Cauchy problems:

a) $x^2y' + y^2 = 0,\ y(-1) = 1;$

b) $2y'\sqrt{x} = y,\ y(4) = 1;$

c) $y' = (2y + 1)\ \mathrm{cotan}\ x,\ y\left(\dfrac{\pi}{4}\right) = \dfrac{1}{2};$

d) $y' = x^2 + 3\,\sin^5 x\,\cos x,\ y(0) = 2.$

R e s u l t s :

a) $y = -x$ for $x \in (-\infty; +\infty)$

b) $y = \exp(\sqrt{x} - 2)$ for $x \in (0; +\infty)$

c) $y = 2 \sin^2 x - \frac{1}{2}$ for $x \in (0; \pi)$

d) $y = \frac{1}{3} x^3 + \frac{1}{2} \sin^6 x + 2$ for $x \in (-\infty; +\infty)$. (The variables are already separated in the given equation. Thus, in order to solve it, you only need to integrate it.)

CHAPTER V.

Definite (Riemann's) integral

V.1. Historical approach

V.1.1. Motivation, introduction.

Imagine that f is a continuous non–negative function in a closed interval $\langle a; b\rangle$. We want to evaluate the area of region R which is bounded by the graph of function f, by the x–axis and by the straight lines $x = a$ and $x = b$. (See Fig. 39.)

If function f is non–negative and constant in the whole interval $\langle a; b\rangle$, taking the value h, then the problem is simple. Region R is a rectangle whose sides have the lengths $b - a$ and h. The area of such a rectangle is $A(R) = (b - a) \cdot h$.

Similarly, if function f is non–negative and linear (i.e. $f:\ y = kx + q$) in the interval $\langle a; b\rangle$ then the problem is also simple. Region R is a trapezoid whose base has the length $b - a$ and the two vertical sides have the lengths $ka + q$ and $kb + q$. Sketch a picture and evaluate the area of R for yourself.

V.1.2. Historical introduction of the definite integral.

The general problem is more complicated and it was solved in the 17th century by Isaac Newton (1642–1727) and Gottfried Wilhelm Leibniz (1646–1716). Their fundamental idea was to divide the interval $\langle a; b\rangle$ into infinitely many "infinitely short" sub–intervals. A typical sub–interval has the end points x and $x + \mathrm{d}x$ where $x \in \langle a; b)$ and $\mathrm{d}x$ is an "infinitely small" positive number. The region, bounded above by the graph of function f on the interval $\langle x; x + \mathrm{d}x\rangle$ and bounded below by the segment which coincides with this interval on the x–axis, is "infinitely narrow" and function f can therefore be considered to be constant, taking the value $f(x)$, on the whole interval $\langle x; x + \mathrm{d}x\rangle$. Thus, the region is an "infinitely narrow" rectangle whose sides have the lengths $\mathrm{d}x$ and $f(x)$. Its area is $\mathrm{d}s = f(x)\,\mathrm{d}x$.

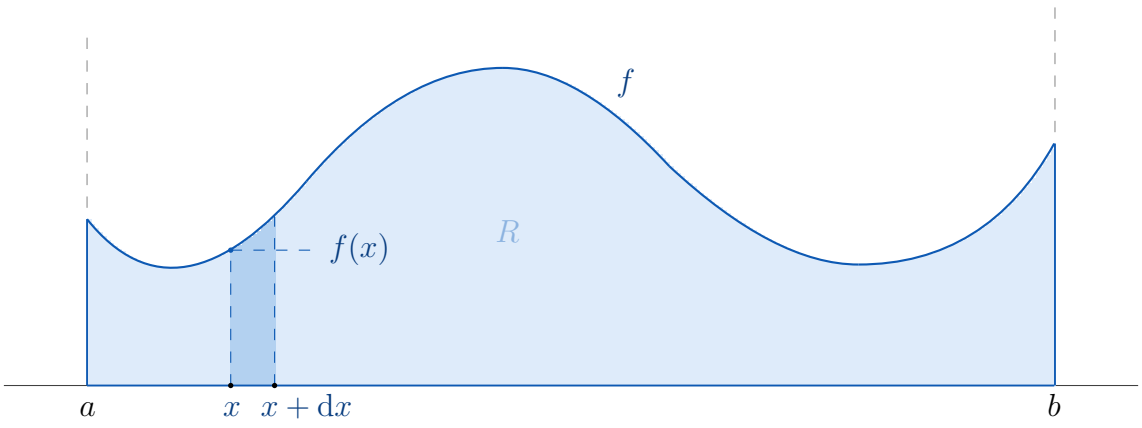


Fig. 39

The area of the whole region R equals the sum of all "infinitely small" numbers $\mathrm{d}s$. The sum was originally denoted by S, however this letter was later stretched and replaced by the sign of integral $\int$. The area of region R can be written as follows:

$$A(R) = \int_a^b \mathrm{d}s = \int_a^b f(x)\,\mathrm{d}x. \tag{V.1.1}$$

The expression on the right hand side is called the *definite integral* of function f on the interval $\langle a;b\rangle$ and the numbers a and b are called the *limits* of the integral: a is the *lower limit* and b is the *upper limit*.

The formulas for the area of a square, a rectangle, a trapezoid, a parallelogram, a disc, etc., are well known. However, an analogous simple formula, expressing the area of a general region of type R (see Fig. 39) does not exist. We must therefore understand formula (V.1.1) so that it defines the area of region R and not only tells us how to evaluate something that we have already defined in another way.

The definite integral also has a sense for functions whose values at some (or at all) points are negative. However, the integral expresses the area between the graph of function f and the x–axis with the minus sign on those intervals where f is negative.

V.1.3. A remark on the physical sense of the definite integral.

Many physical problems naturally also lead to the definite integral. For instance, imagine that a bar covers the interval $\langle a;b\rangle$ on the x–axis. The longitudinal density of the bar is ρ. (The longitudinal density expresses the mass per unit of length. It is given in $\mathrm{kg}\cdot\mathrm{m}^{-1}$.) ρ can vary along the bar, e.g. due to the variable intersection of the bar or because the material the bar is made of varies. Thus, ρ is generally a function of x. We wish to express the total mass of the bar.

As in paragraph V.1.2, we divide the interval $\langle a;b\rangle$ into infinitely many "infinitely small" parts. A typical part is the interval $\langle x;x+\mathrm{d}x\rangle$. Since this interval is "infinitely short", the density ρ can be considered to be constant and equal to $\rho(x)$ on the whole interval $\langle x;x+\mathrm{d}x\rangle$. Then the mass of the part of the bar which covers this interval is $\mathrm{d}m = \rho(x)\,\mathrm{d}x$. The total mass of the bar is the sum of all masses $\mathrm{d}m$ of all "infinitely short" parts of the bar whose length is $\mathrm{d}x$:

$$m = \int_a^b \mathrm{d}m = \int_a^b \rho(x)\,\mathrm{d}x. \tag{V.1.2}$$

V.1.4. A remark on the introduction of the definite integral.

The conception of the definite integral described in the preceding paragraphs is not sufficiently rigorous and clear from the today's point of view. The main reason is that it uses the notion of an "infinitely small", but at the same time positive, number $\mathrm{d}x$. In fact, such a number does not exist. (Assume for a while that $\mathrm{d}x$ with the required properties exists. What value it can have? For example, $\mathrm{d}x = 10^{-2}$? Definitely not, because 10^{-3} is positive and less. Or $\mathrm{d}x = 10^{-6}$? But 10^{-7} is also positive and less than 10^{-6}. You can see that the idea of an infinitely small positive number always leads to a contradiction.) The next reason why the introduction of the definite integral from paragraph V.1.2 is not quite correct is that it does not explain what one can understand under the "sum of infinitely many infinitely small numbers".

Thus, in order to give a correct sense to the definite integral, it is necessary

a) to write a rigorous definition,

b) to derive main properties of the integral (to specify for which functions its construction works; in other words: for which functions the integral exists, etc.)

c) and the last but not least question is how the value of the integral can be evaluated.

V.1.5. The definite integral as a function of the upper limit, the Newton–Leibniz formula.

Answers to the above questions were paradoxically found in the reverse order. Newton and Leibniz were again the first to recognize that the evaluation of the area of the region between the graph of function f and the x–axis is an inverse problem to the differentiation:
Let f be a continuous function in $\langle a;b\rangle$ and $x \in \langle a;b\rangle$. Denote by $A(x)$ the area of the region between the graph of function f and the x–axis on the interval $\langle a;x\rangle$, i.e.

$$A(x) = \int_a^x f(t)\,\mathrm{d}t. \tag{V.1.3}$$

Then, and this is very important, the derivative of A equals f in $\langle a;b\rangle$. Thus, A is an antiderivative to f in $\langle a;b\rangle$! Moreover, it follows directly from (V.1.3) that $A(a) = 0$ and $A(b) = \int_a^b f(x)\,\mathrm{d}x$. Hence we obtain the formula

$$\int_a^b f(x)\,\mathrm{d}x = A(b) - A(a). \tag{V.1.4}$$

Formula (V.1.4) is known as the Newton–Leibniz formula. If we take into account that any other antiderivative F to f on $\langle a;b\rangle$ has a form $F = A + C$ where C is a constant then we can easily verify that $F(b) - F(a) = A(b) - A(a)$. This says that formula (V.1.4) remains valid even if the right hand side equals $F(b) - F(a)$ where F is an arbitrary antiderivative to f on $\langle a;b\rangle$ and not only that particular one which was defined by (V.1.3). We shall again deal with the Newton–Leibniz formula in greater detail in Section V.4.

V.1.6. Linearity of the definite integral.

It is natural to expect from the way in which the definite integral was introduced (we avoid the word "defined", because no precise definition has been shown yet) that apart from other things, it will have the following properties:

1. $\displaystyle\int_a^b \big(\,f(x)+g(x)\,\big)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + \int_a^b g(x)\,\mathrm{d}x,$
2. $\displaystyle\int_a^b c\cdot f(x)\,\mathrm{d}x = c\int_a^b f(x)\,\mathrm{d}x$ (if c is a constant) and
3. $\displaystyle\int_a^b f(x)\,\mathrm{d}x = \int_a^c f(x)\,\mathrm{d}x + \int_c^b f(x)\,\mathrm{d}x$ (if $c \in \langle a;b\rangle$).

We shall discuss these and other properties in greater detail in Section V.3.

V.1.7. Who invented calculus?

Though calculus is the result of a long development that began already in ancient Greece, and many scientists had worked on finding tangents or calculating areas (let us mention e.g. P. Fermat 1601–1665, R. Descartes 1596–1650 and B. Cavalieri 1598–1647), I. Newton (1642–1727) and G. W. Leibniz (1646–1716) were probably the first who recognized that the processes of differentiation (i.e. calculation of the slope of the tangent) and integration (i.e. evaluation of the area) are mutually inverse. However, since they lived at the same time, the question of priority led to bitter controversies. Newton probably arrived at his results earlier, but Leibniz published his conception first. Leibniz was accused of copying Newton's work, and this led to a split in the mathematical world for more than 150 years. While Newton's disciples, mostly British, pursued his approach, Leibniz's followers, mostly Germans and French, were developing his methods. Due to Leibniz's precise notation and rigorous formalism, their contribution to the further development of calculus and its applications was perhaps greater than that of

their British counterparts. Today's historians and mathematicians agree with the opinion that Newton's and Leibniz's inventions were simultaneous but independent.

V.1.8. A remark on definitions of the definite integral.

The introduction of the definite integral on the intuitive level, explained in paragraph V.1.2, was considered to be relatively satisfactory for a certain time. However, the increasing importance of calculus and the range of its applications in physics and other disciplines required a correct and rigorous definition of the definite integral. Let us mention three such definitions that have appeared during the last several centuries.

The first definition is due to Newton. Newton took formula (V.1.4) as a base and suggested that the definite integral of function f on the interval $\langle a; b\rangle$ would be understood to be the difference $F(b) - F(a)$, where F is an antiderivative to f on $\langle a; b\rangle$. However, this definition rather hazes the original geometric and physical problems which had initially motivated the introduction of the definite integral.

A further definition which was very successful was given by George Riemann (1822–1866). We shall explain this definition in section V.2. The definite integral defined in the way proposed by Riemann is usually called the *Riemann integral*. We shall deal only with this type of definite integral in this textbook.

Today's applied mathematics mostly uses the so called Lebesgue definite integral. This integral represents a generalization of the Riemann integral, however its theory requires a longer explication and goes beyond the scope of this text.

Although we have explained that the notion of an infinitely small positive number $\mathrm{d}x$ does not coincide with today's conception of real numbers, we are not going to release this notion completely. By analogy with our derivation of the formula for the mass of an inhomogeneous bar in paragraph V.1.3, we can further follow the same approach and derive, e.g., the formulas for the length of the graph of a function $y = f(x)$, for the position of the center of mass or for the static moment of a mass curve, etc. These applications of the definite integral will be shown in Section V.7.

V.2. Definition of the Riemann integral

Suppose that

a) $\langle a; b\rangle$ is a closed, bounded and non–empty interval,
b) f is a bounded function in the interval $\langle a; b\rangle$.

You will see in this section that the definition of the Riemann integral of function f on the interval $\langle a; b\rangle$ is based on the partition of $\langle a; b\rangle$ to many smaller sub–intervals, on the construction of a Riemann sum which is an approximation of the area of the region R between the graph of function f, the x–axis and the straight lines $x = a$ and $x = b$ and finally on the consideration of the limit of the sum as the partition of $\langle a; b\rangle$ becomes finer and finer.

V.2.1. The partition of an interval.

Let $a < b$. A system of points $x_0, x_1, \ldots, x_n$ in the interval $\langle a; b\rangle$ such that $a = x_0 < x_1 < \cdots < x_n = b$ is called the *partition* of $\langle a; b\rangle$. If the partition is named D then we write:

$$D: \quad a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b. \tag{V.2.1}$$

The *norm* of partition D is the number $\|D\| = \max_{i=1,\ldots,n}(x_i - x_{i-1})$. (In other words, $\|D\|$ is the length of the largest of the sub–intervals $\langle x_0; x_1\rangle, \langle x_1; x_2\rangle, \ldots, \langle x_{n-1}; x_n\rangle$. $\|D\|$ provides information on how fine the partition D is.)

V.2.2. Riemann's sum and its limit.

Suppose that f is a bounded function in the interval $\langle a; b\rangle$ and P is a partition of $\langle a; b\rangle$ given by (V.2.1). Denote by Δx_i the length of the i-th sub–interval $\langle x_{i-1}; x_i\rangle$. (I.e. $\Delta x_i = x_i - x_{i-1}$.)

Let us choose a point ζ_i in each of the intervals $\langle x_{i-1}; x_i\rangle$. Denote by V the system of all chosen points $\zeta_1 \in \langle x_0; x_1\rangle$, $\zeta_2 \in \langle x_1; x_2\rangle$, $\ldots$, $\zeta_n \in \langle x_{n-1}; x_n\rangle$.

Then the sum

$$s(f, D, V) = \sum_{i=1}^{n} f(\zeta_i) \cdot \Delta x_i$$

is called *Riemann's sum* of function f on the interval $\langle a; b\rangle$ associated with partition D and system V.

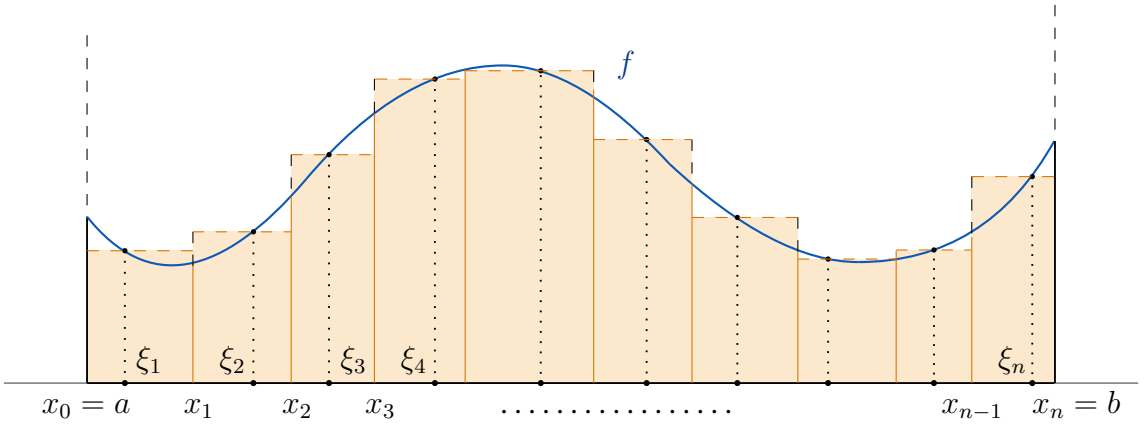


Fig. 40

We say that number S is a *limit of Riemann's sums* $s(f, D, V)$ *as* $\|D\| \to 0+$ if to each $\epsilon > 0$ there exists $\delta > 0$ such that for all partitions D of $\langle a; b\rangle$ and for all systems V of the points $\zeta_i \in \langle x_{i-1}; x_i\rangle$ the implication

$$\|D\| < \delta \implies |s(f, D, V) - S| < \epsilon$$

holds. We write:

(V.2.2)
$$\lim_{\|D\|\to 0+} s(f, D, V) = S.$$

V.2.3. The Riemann integral.

If the limit (V.2.2) exists then function f is said to be *integrable* in the interval $\langle a; b\rangle$ and the number S is called the *Riemann integral* of function f in $\langle a; b\rangle$. The integral is denoted

$$\int_a^b f(x)\,\mathrm{d}x \quad \text{or} \quad \int_a^b f\,\mathrm{d}x.$$

Numbers a and b are called the *limits* of the integral. a is the *lower limit* and b is the *upper* limit. The integrated function is called the *integrand*.

Instead of saying that function f is integrable in $\langle a; b\rangle$, we can also say that the *Riemann integral* $\int_a^b f(x)\,\mathrm{d}x$ *exists*.

The Riemann integral is often called the *definite integral*. (We know from paragraph V.1.8 that there exist several types of definite integrals. However, this cannot lead to a misunderstanding because the only type of definite integral we deal with in this text is the Riemann integral.)

The variable in the Riemann integral can also be denoted by other letters than x and so the integral can be written down as $\int_a^b f(t)\,\mathrm{d}t$, $\int_a^b f(s)\,\mathrm{d}s$ etc.

V.2.4. The area of the region between the graph of function f and the x–axis.

If f is a non–negative integrable function in the interval $\langle a;b\rangle$ then the *area* of region R that is bounded by the graph of f, the x–axis and the straight lines $x=a$ and $x=b$ is defined to be the value of the integral $\int_a^b f(x)\,\mathrm{d}x$.

By analogy, if function f is non–positive and integrable in $\langle a;b\rangle$ then the area of region R between the graph of f, the x–axis and the straight lines $x=a$ and $x=b$ is defined to be equal to $-\int_a^b f(x)\,\mathrm{d}x$.

Think for yourself over this fact: In a general case when f is an integrable function in $\langle a;b\rangle$ which can be positive as well as negative, the integral $\int_a^b f(x)\,\mathrm{d}x$ is equal to the sum of the areas of all parts of the plane between the graph of f and the x–axis where the contributions of the parts below the x–axis are taken with the minus sign. (Sketch a picture.)

V.2.5. The extension of the definition of the Riemann integral.

If f is an integrable function in the interval $\langle a;b\rangle$ then we put

$$\int_b^a f(x)\,\mathrm{d}x = -\int_a^b f(x)\,\mathrm{d}x.$$

Specially, we also put $\int_a^a f(x)\,\mathrm{d}x = 0$.

V.2.6. The mean value of function f in the interval $\langle a;b\rangle$.

Suppose that f is an integrable function in the interval $\langle a;b\rangle$. Then the number

$$\mu = \frac{1}{b-a}\int_a^b f(x)\,\mathrm{d}x$$

is said to be the *mean value* of function f in the interval $\langle a;b\rangle$.

The geometric interpretation of the mean value is easy: Suppose for simplicity that function f is non–negative in $\langle a;b\rangle$. Then the mean value of f in $\langle a;b\rangle$ is such a number μ that region R between the graph of f, the x–axis and the straight lines $x=a$ and $x=b$ has the same area as the rectangle whose sides are $b-a$ and μ. (Sketch a picture.)

V.3. Important properties of the Riemann integral

Let us have in mind that the Riemann integral has been constructed from the beginning only for bounded functions and on a bounded interval $\langle a;b\rangle$.

Most of the bounded functions you will meet in applications will be integrable. Nevertheless, there also exist "bad" functions for which the limit of the Riemann sums (V.2.2) in the interval $\langle a;b\rangle$ does not exist. Such functions are not integrable in $\langle a;b\rangle$, which means that their

Riemann integral in $\langle a; b\rangle$ does not exist. The border between integrable and non–integrable functions is, in general, not easy to describe. In order to have a tool which can enable us to recognize some integrable functions, we show in the next theorem and in remark V.3.2 sufficient conditions which guarantee the integrability of a function f in $\langle a; b\rangle$.

Bear in mind that that the statements "function f is integrable in the interval $\langle a; b\rangle$" and "the Riemann integral $\int_a^b f(x)\,\mathrm{d}x$ exists" say exactly the same.

V.3.1. Theorem (on the existence of the Riemann integral).

Let function f be continuous in the interval $\langle a; b\rangle$. Then f is integrable in $\langle a; b\rangle$.

V.3.2. Remark.

Theorem V.3.1 is satisfactory in most practical cases. However, its generalization is also sometimes useful:

Let function f be bounded and piecewise continuous in the interval $\langle a; b\rangle$. Then f is integrable in $\langle a; b\rangle$.

(Function f is said to be *piecewise continuous* in the interval $\langle a; b\rangle$ if $\langle a; b\rangle$ can be divided into a finite number of sub–intervals such that f is continuous in the interior of each of them.)

V.3.3. Theorem.

a) If function f is integrable in the interval $\langle a; b\rangle$ and $\langle c; d\rangle \subset \langle a; b\rangle$ then f is integrable in the interval $\langle c; d\rangle$, too.

b) If functions f and g are both integrable in the interval $\langle a; b\rangle$ then their product $f \cdot g$ is also an integrable function in $\langle a; b\rangle$.

c) If f is an integrable function in the interval $\langle a; b\rangle$ and a function g differs from f in $\langle a; b\rangle$ only at a finite number of points then g is also an integrable function in $\langle a; b\rangle$ and

$$\int_a^b g(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x.$$

Assertion a) easily follows from the definition of the Riemann integral.

Assertion b) represents a statement about the integrability of the product of the two functions f and g. However, bear in mind that this does not mean that $\int_a^b f \cdot g\,\mathrm{d}x = \left(\int_a^b f\,\mathrm{d}x\right) \cdot \left(\int_a^b g\,\mathrm{d}x\right)$!

Assertion c) states that the change of function values of function f at a finite number of points has no influence on the existence or on the value of the integral $\int_a^b f\,\mathrm{d}x$. In other words: Neither the existence nor the value of the integral $\int_a^b f\,\mathrm{d}x$ depends on the values of f at a finite number of points. Thus, function f need not even be defined at a finite number of points of the interval $\langle a; b\rangle$ and this has no influence on the existence or on the value of the integral $\int_a^b f\,\mathrm{d}x$. In particular, it plays no role whether the integral is considered on an open or on a closed interval.

V.3.4. Upper and lower bounds of the Riemann integral.

It can easily be deduced from the definition of the Riemann integral and its geometric interpretation (see paragraph V.2.4) that if function f, integrable in $\langle a; b\rangle$, satisfies

$$m \le f(x) \le M \quad \text{for all } x \in \langle a; b\rangle,$$

then $$m \cdot (b-a) \;\le\; \int_a^b f(x)\,\mathrm{d}x \;\le\; M \cdot (b-a).$$

Since this holds with an arbitrary lower bound m and an arbitrary upper bound M of function f on $\langle a;b\rangle$, we can replace m by the greatest lower bound (i.e. by $\inf_{x\in\langle a;b\rangle} f(x)$) and M by the least upper bound (i.e. by $\sup_{x\in\langle a;b\rangle} f(x)$) and the inequality still holds.

In particular, if f has its minimum and maximum in the interval $\langle a;b\rangle$ then

(V.3.1) $$\min_{x\in\langle a;b\rangle} f(x) \cdot (b-a) \;\le\; \int_a^b f(x)\,\mathrm{d}x \;\le\; \max_{x\in\langle a;b\rangle} f(x) \cdot (b-a).$$

The next theorem also easily follows from the definition of the Riemann integral and its geometric sense.

V.3.5. Theorem.

If functions f and g are both integrable in $\langle a;b\rangle$ and $g(x) \le f(x)$ for all $x \in \langle a;b\rangle$ then

$$\int_a^b g(x)\,\mathrm{d}x \;\le\; \int_a^b f(x)\,\mathrm{d}x.$$

Particularly, if $f(x) \ge 0$ for all $x \in \langle a;b\rangle$ then $\int_a^b f(x)\,\mathrm{d}x \ge 0$.

The next theorem confirms the validity of the formulas from paragraph V.1.6.

V.3.6. Theorem. (Linearity of the Riemann integral.)

If f and g are integrable functions in $\langle a;b\rangle$ and $\alpha \in \mathbb{R}$ then

$$\int_a^b (f(x)+g(x))\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + \int_a^b g(x)\,\mathrm{d}x \quad\text{and}\quad \int_a^b \alpha\cdot f(x)\,\mathrm{d}x = \alpha\int_a^b f(x)\,\mathrm{d}x.$$

(The same property is already known from the theory of the indefinite integral – see theorem IV.1.8.)

V.3.7. Theorem. (Additivity of the Riemann integral with respect to the interval.)

If the integrals $\int_a^c f\,\mathrm{d}x$ and $\int_c^b f\,\mathrm{d}x$ exist then

$$\int_a^c f(x)\,\mathrm{d}x + \int_c^b f(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x.$$

V.3.8. Theorem. (The Riemann integral as a function of the upper limit.)

Suppose that function f is integrable in the interval $\langle a;b\rangle$. Then

a) the function $A(x) = \int_a^x f(t)\,\mathrm{d}t$ is continuous in $\langle a;b\rangle$,

b) the identity

(V.3.2) $$A'(x) \;=\; \frac{\mathrm{d}}{\mathrm{d}x}\int_a^x f(t)\,\mathrm{d}t \;=\; f(x)$$

holds at all points $x \in (a;b)$ where function f is continuous.

The correctness of assertion a) follows (at least intuitively) from the geometric sense of the Riemann integral (see paragraph V.2.4). Try to sketch a picture and think it over.

Formula (V.3.2) can be proved in this way:

$$A'(x) = \lim_{h\to 0} \frac{A(x+h)-A(x)}{h} = \lim_{h\to 0} \frac{1}{h}\left(\int_a^{x+h} f(t)\,\mathrm{d}t - \int_a^x f(t)\,\mathrm{d}t\right) =$$

$$= \lim_{h\to 0} \frac{1}{h}\int_x^{x+h} f(t)\,\mathrm{d}t = \lim_{h\to 0} \mu(h)$$

where $\mu(h)$ is the mean value of function f in the interval with the end points x and $x+h$. The continuity of function f at point x and (V.3.1) imply that $\mu(h) \to f(x)$ for $h \to 0$. So we obtain (V.3.2).

V.3.9. Remark.

We can deduce from theorem V.3.8 that if function f is continuous in the interval I and $a \in I$ then $A(x) = \int_a^x f(t)\,\mathrm{d}t$ is an antiderivative to f in I.

V.3.10. Remark.

The following generalization of part b) of theorem V.3.8 is often useful in applications of the definite integral:

If function f is continuous in interval I and $a(x)$, $b(x)$ are differentiable functions in interval J such that their values belong to the interval I then

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_{a(x)}^{b(x)} f(t)\,\mathrm{d}t = f(b(x))\cdot b'(x) - f(a(x))\cdot a'(x) \quad \text{for } x \in J.$$

V.3.11. Problems.

Do the following Riemann integrals exist?

a) $\int_{-2}^{1} \frac{x+1}{x^2-x-6}\,\mathrm{d}x$ b) $\int_1^2 \frac{\ln x}{x}\,\mathrm{d}x$ c) $\int_0^1 \frac{\sin x}{x}\,\mathrm{d}x$

d) $\int_{-1}^{5} \mathrm{e}^{-x}\,\mathrm{d}x$ e) $\int_{-2.5}^{3} \frac{x}{\ln(x+3)}\,\mathrm{d}x$ f) $\int_{-2}^{-1} \frac{x^2+1}{x^3-2x^2+x}\,\mathrm{d}x$

Results:

a) no b) yes c) yes d) yes e) no f) yes.

V.3.12. Problems.

Find $G'(x)$ if function G is defined by the following integrals.

a) $G(x) = \int_{\frac{1}{x}}^{\sqrt{x}} \cos(t^2)\,\mathrm{d}t\,;\ x>0$ b) $G(x) = \int_0^{2x} \frac{\sin t}{t}\,\mathrm{d}t$

c) $G(x) = \int_x^0 \sqrt{1+t^4}\,\mathrm{d}t$ d) $G(x) = \int_{x^2}^{x^3} \ln t\,\mathrm{d}t\,;\ x>0$

Results:

a) $G'(x) = \frac{\cos x}{2\sqrt{x}} + \frac{\cos\frac{1}{x^2}}{x^2}$ (for $x>0$) b) $2\,\frac{\sin 2x}{2x}$

c) $-\sqrt{1+x^4}$ d) $(9x^2-4x)\ln x$ (for $x>0$).

V.4. Evaluation of the Riemann integral

We arrive at an important point in the theory of the Riemann integral which is the question of its evaluation. The next theorem is often called, due to its importance, the Fundamental Theorem of Integral Calculus.

V.4.1. Theorem.

If f is a continuous function in interval $\langle a;b\rangle$ and F is an antiderivative to f in $\langle a;b\rangle$ then

$$\int_a^b f(x)\,\mathrm{d}x \;=\; F(b)-F(a). \tag{V.4.1}$$

Formula (V.4.1) is called the *Newton–Leibniz formula*. The difference $F(b)-F(a)$ is usually written down in a shorter way: $F(b)-F(a)=\big[\,F\,\big]_a^b$.

The proof of theorem V.4.1 is easy: Function $A(x)=\int_a^x f(t)\,\mathrm{d}t$ is an antiderivative to f in $\langle a;b\rangle$, too. Two antiderivatives can differ at most by an additive constant (theorem IV.1.5). Hence there exists a constant C such that $F=G+C$ in $\langle a;b\rangle$. This means that $F(a)=A(a)+C=C$ (because $A(a)=0$) and $F(b)=A(b)+C=A(b)+F(a)$. This implies: $\int_a^b f(t)\,\mathrm{d}t=A(b)=F(b)-F(a)$.

The Newton–Leibniz formula connects the definite integral with the indefinite integral: If you know the indefinite integral of function f in the interval $\langle a;b\rangle$ then you know all antiderivatives to f in $\langle a;b\rangle$. You can therefore choose any of them, use it in the Newton–Leibniz formula and you obtain the value of the definite integral of function f in $\langle a;b\rangle$. The fact that the indefinite integral and the antiderivative play such an important role in the evaluation of the definite integral is one of the main reasons why you learned to calculate indefinite integrals in Chapter IV.

We have already several times repeated that all antiderivatives to f in $\langle a;b\rangle$ differ at most by an additive constant. Thus, if you choose for instance the antiderivative $F+C$ (where C is a constant) instead of F and use it in the Newton–Leibniz formula, you obtain

$$\int_a^b f(x)\,\mathrm{d}x \;=\; \big[F+C\big]_a^b \;=\; (F(b)+C)-(F(a)+C) \;=\; F(b)-F(a).$$

The result is the same as in (V.4.1). This confirms that you can indeed use an arbitrary antiderivative to f in order to evaluate the definite integral $\int_a^b f\,\mathrm{d}x$.

V.4.2. Example.

$$\int_0^{\pi}\sin x\,\mathrm{d}x=\big[-\cos x\,\big]_0^{\pi}=(-\cos\,\pi)-(-\cos\,0)=2\,.$$

The next two theorems show that the two methods of integration, by parts and by substitution, known from the theory of the indefinite integral, can also be directly applied to the definite integral.

V.4.3. Theorem. (Integration by parts.)

Suppose that functions u and v have continuous derivatives in the interval $\langle a;b\rangle$. Then

$$\int_a^b u\cdot v'\,\mathrm{d}x \;=\; \big[\,u\cdot v\,\big]_a^b \;-\; \int_a^b u'\cdot v\,\mathrm{d}x. \tag{V.4.2}$$

V.4.4. Example.

$$\int_0^2 \mathrm{e}^{2x} x \,\mathrm{d}x = \left| \begin{matrix} u = x, & v' = \mathrm{e}^{2x} \\ u' = 1, & v = \frac{1}{2}\mathrm{e}^{2x} \end{matrix} \right| = \left[\frac{1}{2} x \mathrm{e}^{2x}\right]_0^2 - \frac{1}{2}\int_0^2 \mathrm{e}^{2x}\,\mathrm{d}x = \mathrm{e}^4 - \left[\frac{1}{4}\mathrm{e}^{2x}\right]_0^2 = \frac{3}{4}\mathrm{e}^4 + \frac{1}{4}.$$

V.4.5. Theorem. (Integration by substitution.)

Let function g have a continuous derivative in the interval $\langle a; b\rangle$ and map $\langle a; b\rangle$ into an interval J. Let function f be continuous in J. Then

$$\int_a^b f(g(x))\, g'(x)\,\mathrm{d}x = \int_{g(a)}^{g(b)} f(s)\,\mathrm{d}s. \tag{V.4.3}$$

Formula (V.4.3) can be used in two ways:

a) We want to evaluate the integral on the left hand side and we transform it to the integral on the right hand side (if the integral on the right hand side is simpler) or

b) we want to evaluate the integral on the right hand side and we transform it to the integral on the left hand side (if the integral on the left hand side is simpler).

V.4.6. Example.

Evaluate the integral $\displaystyle\int_0^{\frac{\pi}{2}} \sin^2 x \cdot \cos x \,\mathrm{d}x$.

If we put $\langle a; b\rangle = \left\langle 0; \frac{\pi}{2}\right\rangle$, $s = g(x) = \sin x$, $f(s) = s^2$, $J = (-\infty; +\infty)$ then all the assumptions of theorem V.4.5 are fulfilled. Moreover, $g(0) = \sin 0 = 0$ and $g\left(\frac{\pi}{2}\right) = \sin\left(\frac{\pi}{2}\right) = 1$. Applying formula (V.4.3), we obtain:

$$\int_0^{\frac{\pi}{2}} \sin^2 x \cos x \,\mathrm{d}x = \left| \begin{matrix} s = \sin x, & x = 0 \implies s = 0 \\ \mathrm{d}s = \cos x\,\mathrm{d}x, & x = \frac{\pi}{2} \implies s = 1 \end{matrix} \right| = \int_0^1 s^2\,\mathrm{d}s = \left[\frac{s^3}{3}\right]_0^1 = \frac{1}{3}.$$

V.4.7. Example.

Evaluate the integral $\displaystyle\int_0^2 \sqrt{4 - x^2}\,\mathrm{d}x$.

This integral can be considered to be the integral on the right hand side of (V.4.3) (with the variable denoted by x instead of s). The function $f(x) = \sqrt{4 - x^2}$ is continuous in $\langle 0; 2\rangle$, hence the integral exists. Put $x = g(t) = 2\sin t$, $\mathrm{d}x = g'(t)\,\mathrm{d}t = 2\cos t\,\mathrm{d}t$. Then $g(a) = 2\sin a = 0$ and $g(b) = 2\sin b = 2$. We can therefore choose $a = 0$ and $b = \frac{\pi}{2}$. All the assumptions of theorem V.4.5 are now fulfilled and so, applying formula (V.4.3), we get:

$$\int_0^2 \sqrt{4 - x^2}\,\mathrm{d}x = \left| \begin{matrix} x = 2\sin t, & x = 0 \implies t = 0 \\ \mathrm{d}x = 2\cos t\,\mathrm{d}t, & x = 2 \implies t = \frac{\pi}{2} \end{matrix} \right| = \int_0^{\frac{\pi}{2}} \sqrt{4 - 4\sin^2 t}\; 2\cos t\,\mathrm{d}t =$$

$$= 4\int_0^{\frac{\pi}{2}} \cos^2 t\,\mathrm{d}t = \int_0^{\frac{\pi}{2}} 2(1 + \cos(2t))\,\mathrm{d}t = \left[2t + \sin(2t)\right]_0^{\frac{\pi}{2}} = \pi.$$

V.4.8. Remark.

If you need to evaluate the Riemann integral of function f in the interval $\langle a; b\rangle$ and you consider applying integration by parts or by a substitution then you have two possibilities:

1) You can use theorem V.4.3 or theorem V.4.5. Thus, you transform the integral to another (simpler) integral and you also deal with the limits of the integral. This approach was explained in examples V.4.5, V.4.6 a V.4.7.
2) You can first calculate the integral as an indefinite integral and then you can apply the Newton–Leibniz formula (V.4.1) on the interval $\langle a; b\rangle$.

   In order to show this approach in greater detail, we evaluate the integral from example V.4.6 once again, this time by the method we are just explaining. Thus, let us begin with the indefinite integral $\int \sin^2 x \, \cos x \, \mathrm{d}x$. We obtain

$$\int \sin^2 x \cos x \, \mathrm{d}x = \left| \begin{matrix} s = \sin x \\ \mathrm{d}s = \cos x \, \mathrm{d}x \end{matrix} \right| = \int s^2 \, \mathrm{d}s = \frac{s^3}{3} + C = \frac{\sin^3 x}{3} + C.$$

   Using formula (V.4.1), we obtain: $\int_0^{\frac{\pi}{2}} \sin^2 x \, \cos x \, \mathrm{d}x = \left[\frac{1}{3} \sin^3 x \right]_0^{\frac{\pi}{2}} = \frac{1}{3}$.
   You will see after having solved a large number of problems individually that the first method, based on the direct application of the integration by parts or the integration by substitution is usually technically easier and less laborious.

V.4.9. Problems.

Evaluate the following integrals.

a) $\int_{-1}^{1} (3x^2 - 4x + 7) \, \mathrm{d}x$ b) $\int_0^1 (8t^3 - 12t^2 + 5) \, \mathrm{d}t$ c) $\int_1^2 \frac{4}{s^2} \, \mathrm{d}s$

d) $\int_1^{27} x^{-\frac{4}{3}} \, \mathrm{d}x$ e) $\int_0^2 \sqrt{4 - s^2} \, \mathrm{d}s$ f) $\int_0^1 u \arctan u \, \mathrm{d}u$

g) $\int_0^{\pi} \cos^2 x \, \mathrm{d}x$ h) $\int_0^1 9w \, \mathrm{e}^{3w} \, \mathrm{d}w$ i) $\int_{\frac{1}{\mathrm{e}}}^{\mathrm{e}} |\ln r| \, \mathrm{d}r$

j) $\int_0^1 \sqrt{\frac{x}{4 - x}} \, \mathrm{d}x$ k) $\int_0^{\frac{\pi}{2}} \sin^3 \varphi \, \cos^2 \varphi \, d\varphi$ l) $\int_0^{\frac{\pi}{2}} \frac{\mathrm{d}\theta}{3 + 2 \sin \theta}$

Evaluate the area of the domain between the graph of function f and the x–axis.

m) $f(x) = x^2 - 4x + 3, \ 0 \le x \le 3$ n) $f(x) = 1 - \left(\frac{x^2}{4} \right), \ -2 \le x \le 3$

o) $f(x) = 5 - 5x^{\frac{2}{3}}, \ -1 \le x \le 8$ p) $f(x) = 1 - \sqrt{x}, \ 0 \le x \le 4$

Find the mean value of function f on a given interval.

q) $f(x) = \sqrt{3x}$ on $\langle 0; 3\rangle$ r) $f(x) = \sqrt{ax}$ on $\langle 0; a\rangle$

s) $f(x) = mx + b$ on $\langle -1; 1\rangle$ t) $f(x) = mx + b$ on $\langle -k; k\rangle$

Results:

a) 18 b) 4 c) 2

d) 2 e) π f) $\frac{1}{4}\pi - \frac{1}{2}$

g) $\frac{1}{2}\pi$ h) $2\mathrm{e}^3 + 1$ i) $2(1 - \mathrm{e}^{-1})$

j) $\frac{2\pi}{3} - \sqrt{3}$ k) $\frac{2}{15}$ l) $\frac{2}{\sqrt{5}} \cdot \arctan\left(\frac{1}{\sqrt{5}}\right)$

m) 0 n) $\frac{41}{12}$ o) -58

p) $-\frac{4}{3}$ q) 2 r) $\frac{2}{3}a$

s) b t) b.

V.5.* Numerical integration

It is well known from the theory of the indefinite integral that an antiderivative to a given function f often exists, but it cannot be obtained by standard integration procedures and it cannot be expressed in a "closed form", i.e. by a formula which prescribes a finite number of operations. By analogy, it often happens that the Riemann integral $\int_a^b f\,\mathrm{d}x$ exists, but it cannot be evaluated by a standard integration based on the application of the antiderivative and the Newton–Leibniz formula. However, there exist approximate methods (often also called numerical methods), which enable us to evaluate the integral only approximately, but with an error as small as we wish. We shall explain two such methods in this section. All these methods require the performance of a relatively high number of arithmetic operations in order to achieve higher accuracy (i.e. a smaller error). They can therefore be practically realized only on computers.

Both the methods are based on the partition

$$D: \quad a = x_0 < x_1 < x_2 \ldots < x_{n-1} < x_n = b$$

of the interval $\langle a; b\rangle$ onto n sub–intervals $\langle x_{k-1}; x_k\rangle$ $(k = 1, 2, \ldots, n)$ of the same length h. Hence

$$h = \frac{b-a}{n} \qquad \text{a} \qquad x_k = a + k \cdot h \quad (k = 1, 2, \ldots, n).$$

We shall denote $y_k = f(x_k)$.

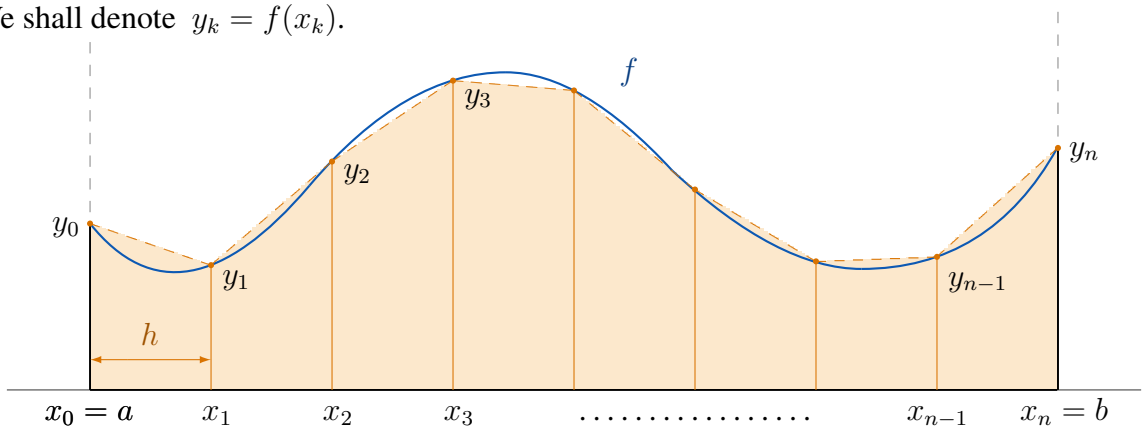


Fig. 41

V.5.1. Trapezoidal Rule.

We approximate function f by a linear function on each sub–interval $\langle x_{k-1}; x_k\rangle$. The linear function is uniquely determined by the requirement that its graph (a straight line) passes through two chosen points. Let us choose the points $[x_{k-1}; y_{k-1}]$ and $[x_k; y_k]$. Then the considered linear function is: $y = y_{k-1} + \dfrac{y_k - y_{k-1}}{h} \cdot (x - x_{k-1})$. Its integral in the interval $\langle x_{k-1}; x_k\rangle$ (let us denote it by I_k) represents the area of the trapezoid (see Fig. 41) and so $I_k = h \cdot \dfrac{y_{k-1} + y_k}{2}$. If we add all the numbers $I_1,\ I_2,\ \ldots\ ,\ I_n$, we obtain

$$L_n = \frac{h}{2} \cdot (\, y_0 + 2y_1 + 2y_2 + \cdots + 2y_{n-1} + y_n\,). \tag{V.5.1}$$

L_n is the approximate value of the Riemann integral $\int_a^b f\,\mathrm{d}x$. The geometric sense of L_n is shown in Fig. 41 – it is the sum of the areas of n trapezoids constructed on the intervals $\langle x_0; x_1\rangle$, $\langle x_1; x_2\rangle\ \ldots\ , \langle x_{n-1}; x_n\rangle$.

It can naturally be expected that the finer the partition of the interval $\langle a; b\rangle$, the better is the approximation of the Riemann integral $\int_a^b f\,\mathrm{d}x$. In other words, the accuracy of the approximation should increase with increasing n (i.e. with decreasing h). This expectation is correct. It can be proved that if function f has a continuous second derivative f'' in $\langle a; b\rangle$ and M_2 is the maximum of $|f''|$ in $\langle a; b\rangle$ then the following error estimate holds:

$$\left| L_n - \int_a^b f(x)\,\mathrm{d}x \right| \le \frac{b-a}{12}\, h^2\, M_2. \tag{V.5.2}$$

(Estimate (V.5.2) is called an error estimate because it provides an upper estimate of the error we make if we replace the exact value of the integral $\int_a^b f\,\mathrm{d}x$ by its approximate value L_n.)

V.5.2. Simpson's Rule.

Let us now choose an integer n so that it is even and let us approximate function f by a quadratic polynomial on each of the sub–intervals $\langle x_0; x_2\rangle$, $\langle x_2; x_4\rangle$, $\ldots$, $\langle x_{n-2}; x_n\rangle$. The quadratic polynomial on the sub–interval $\langle x_{k-2}; x_k\rangle$ $(k = 2, 4, \ldots, n)$ is uniquely determined if we require that its graph (a parabola) passes through three chosen points. Let the three chosen points be $[x_{k-2}; y_{k-2}]$, $[x_{k-1}; y_{k-1}]$, $[x_k; y_k]$. The coefficients and the integral of such a quadratic polynomial in the interval $\langle x_{k-2}; x_k\rangle$ can be relatively easily evaluated – you can verify for yourself that the integral is $I_k = h \cdot \dfrac{y_{k-2} + 4y_{k-1} + y_k}{3}$. Summing all the numbers $I_2,\ I_4,\ \ldots\ ,\ I_n$, we obtain

$$S_n = \frac{h}{3} \cdot (y_0 + 4y_1 + 2y_2 + 4y_3 + \cdots + 2y_{n-2} + 4y_{n-1} + y_n\,). \tag{V.5.3}$$

S_n is an approximate value of the Riemann integral $\int_a^b f\,\mathrm{d}x$. It can be proved that if the fourth derivative $f^{(4)}$ of function f is continuous in $\langle a; b\rangle$ and M_4 is the maximum of $|f^{(4)}|$ in $\langle a; b\rangle$ then the following error estimate holds:

$$\left| S_n - \int_a^b f\,\mathrm{d}x \right| \le \frac{b-a}{180}\, h^4\, M_4. \tag{V.5.4}$$

V.6. The improper Riemann integral

We have assumed that

(V.6.1) $\langle a; b\rangle$ is a bounded interval and

(V.6.2) f is a bounded function in this interval

in the definition of the Riemann integral $\int_a^b f\,\mathrm{d}x$ (see Section V.2). However, there often arises the necessity to evaluate definite integrals where either the domain of integration (i.e. an interval) or the integrand (i.e. a function), or both, are unbounded. Such integrals are called improper integrals. We will explain the definition of the improper Riemann integral in this section. Let us begin with a simple example.

V.6.1. Example.

The Riemann integral $\int_1^{+\infty} \frac{1}{x}\,\mathrm{d}x$ does not exist because the domain of integration is not a bounded interval. However, the question of the area of the region

$$R = \left\{[x; y] \in \mathbb{E}_2;\ x \in \langle 1; +\infty),\ 0 \le y \le \frac{1}{x}\right\}$$

is quite reasonable. (R is a subset of plane $\mathbb{E}_2$, corresponding to $x \in \langle 1; +\infty)$, bounded above by the graph of the function $\frac{1}{x}$ and bounded below by the x–axis. Sketch a picture.) The answer can be found in this way: Let us choose $t \in \langle 1; +\infty)$. The area of the region

$$R(t) = \left\{[x; y] \in \mathbb{E}_2;\ x \in \langle 1; t\rangle,\ 0 \le y \le \frac{1}{x}\right\},$$

i.e. of the part of the plane which corresponds to $x \in \langle 1; t\rangle$ and which is bounded above by the graph of the function $\frac{1}{x}$ and bounded below by the x–axis, is

$$s(R(t)) = \int_1^t \frac{1}{x}\,\mathrm{d}x = \big[\ \ln x\ \big]_1^t = \ln t - \ln 1 = \ln t.$$

The area of the whole region R can now be naturally defined by the equation

$$s(R) = \lim_{t\to+\infty} s(R(t)) = \lim_{t\to+\infty} \ln t = +\infty.$$

This approach is the motivation for the following definition.

V.6.2. The improper Riemann integral with the singular upper limit.

Assume that function f is defined in the interval $\langle a; b)$ and it is integrable in each interval $\langle a; t\rangle$ (for $a \le t < b$). If

a) at least one of the conditions (V.6.1) and (V.6.2) is not fulfilled and

b) the limit

$$\lim_{t\to b-} \int_a^t f(x)\,\mathrm{d}x,$$

exists

then the value of the limit is called the *improper Riemann integral* with the *singular upper limit*.

The improper Riemann integral is denoted in the same way as the "usual" Riemann integral, i.e. $\int_a^b f(x)\,\mathrm{d}x$. Thus, we can write:

$$\int_a^b f(x)\,\mathrm{d}x \;=\; \lim_{t\to b-} \int_a^t f(x)\,\mathrm{d}x.$$

The improper Riemann integral with the *singular lower limit* can be defined by analogy. (Try to write the definition for yourself.)

While the value of the "proper" Riemann integral can only be a finite number, the value of an improper Riemann integral can be finite or infinite. If the improper integral $\int_a^b f\,\mathrm{d}x$ is finite then we say that the integral *converges*. If $\int_a^b f\,\mathrm{d}x = \pm\infty$, we say that the integral *diverges*.

V.6.3. Example.

Let us evaluate the improper Riemann integral $\displaystyle\int_{-5}^{2} \frac{1}{\sqrt{2-x}}\,\mathrm{d}x$.

The function $f(x) = \dfrac{1}{\sqrt{2-x}}$ is unbounded in the interval $\langle -5; 2\rangle$. (This is due to its behavior in the neighborhood of the upper limit 2, namely $\lim\limits_{x\to 2-} \dfrac{1}{\sqrt{2-x}} = +\infty$.) Hence the Riemann integral of function f in $\langle -5; 2\rangle$ does not exist. However, f is a continuous function in the interval $\langle -5; 2)$ and so it is integrable in each interval $\langle -5; t\rangle$ (where $-5 \le t < 2$). The antiderivative to f in $\langle -5; 2)$ is $F(x) = -2\sqrt{2-x}$. Thus,

$$\int_{-5}^{t} \frac{1}{x}\,\mathrm{d}x \;=\; F(t) - F(1) \;=\; -2\sqrt{2-t} + 2\sqrt{2-(-5)} \;=\; -2\sqrt{2-t} + 2\sqrt{7}.$$

Since $\lim\limits_{t\to 2-} -2\sqrt{2-t} + 2\sqrt{7} = 2\sqrt{7}$, we have: $\displaystyle\int_{-5}^{2} \frac{1}{\sqrt{2-x}}\,\mathrm{d}x = 2\sqrt{7}$.

V.6.4. The improper Riemann integral with both singular limits.

The definition of the improper Riemann integral can be extended to the cases when both the limits are singular: Assume that $c \in (a; b)$ and the two integrals $\int_a^c f\,\mathrm{d}x$ and $\int_c^b f\,\mathrm{d}x$ exist (the first as the improper integral with the singular lower limit a and the second as the improper integral with the upper limit b). If the sum of the integrals has a sense (i.e. if it is not for instance $-\infty + \infty$) then we put

$$\int_a^b f(x)\,\mathrm{d}x \;=\; \int_a^c f(x)\,\mathrm{d}x + \int_c^b f(x)\,\mathrm{d}x.$$

The integral $\int_a^b f\,\mathrm{d}x$ is an improper integral with both singular limits.

The definition can be further extended to situations when some points inside the interval $(a; b)$ are singular. However, we do not deal with such cases in this textbook.

V.6.5. Remark.

If the improper integral $\int_a^b f\,\mathrm{d}x$ exists and the interval $(a; b)$ is unbounded then we sometimes speak about the *improper integral due to the limit*. If function f is unbounded then we speak about the *improper integral due to the function*. These two effects can also be combined (the function is unbounded in the neighborhood of one limit and the second limit is infinite) or can be cumulated (the function is unbounded in the neighborhood of an infinite limit).

V.6.6. The evaluation of the improper Riemann integral.

We shall assume for simplicity that f is a continuous function in the interval (open or closed) with the end points a, b (where $a < b$) in this paragraph.

According to theorem IV.1.3, function f has an antiderivative in the interval between a and b. Let us denote it by F. (There exist in fact infinitely many antiderivatives and they all differ by additive constants – see theorem IV.1.5. However, we shall need only one of them.) It can be shown that the integral $\int_a^b f\,\mathrm{d}x$ exists as an improper integral and

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{t\to b-} F(t) - \lim_{t\to a+} F(t), \tag{V.6.1}$$

if the right hand side has a sense (i.e. if both the limits exist and their difference has a sense).

V.6.7. Example.

The integral $\int_0^{+\infty} \frac{1}{\sqrt{x}}\,\mathrm{d}x$ is an improper integral with both singular limits: 0 is a singular limit because the function $\frac{1}{\sqrt{x}}$ is not bounded in its right neighborhood and $+\infty$ is a singular limit because it is not a finite number. The antiderivative to $\frac{1}{\sqrt{x}}$ in $(0;+\infty)$ is for instance the function $2\sqrt{x}$. Then

$$\int_0^{+\infty} \frac{1}{\sqrt{x}}\,\mathrm{d}x = \lim_{t\to+\infty} 2\sqrt{t} - \lim_{t\to 0+} 2\sqrt{t} = +\infty - 0 = +\infty.$$

V.6.8. Problems.

Evaluate the improper integrals.

a) $\int_0^{+\infty} \frac{\mathrm{d}x}{1+x^2}$ b) $\int_{-\infty}^{+\infty} \frac{\mathrm{d}x}{4+x^2}$ c) $\int_{-\infty}^{-2} \frac{\mathrm{d}x}{x^2}$

d) $\int_2^{+\infty} \frac{\mathrm{d}x}{x^2-1}$ e) $\int_0^{+\infty} x^2\,\mathrm{e}^{-x}\,\mathrm{d}x$ f) $\int_0^5 \frac{1}{\sqrt{x}}\,\mathrm{d}x$

g) $\int_{-\infty}^{+\infty} \frac{\mathrm{d}x}{x^2+4x+9}$ h) $\int_0^{\infty} x\,\mathrm{e}^{-x^2}\,\mathrm{d}x$ i) $\int_1^5 (x-1)^a\,\mathrm{d}x\,;\ a>-1$

Results:

a) $\frac{\pi}{2}$ b) $\frac{\pi}{2}$ c) $\frac{1}{2}$ d) $-\frac{1}{2}\ln\frac{1}{3}$

e) 2 f) $2\sqrt{5}$ g) $\frac{\pi}{\sqrt{20}}$ h) $\frac{1}{2}$

i) $\frac{4^{a+1}}{a+1}$.

V.7. Some geometric and physical applications of the definite integral

The simple geometric and physical sense of the definite integral has already been discussed in paragraphs V.1.1, V.1.3 and V.2.4. We will show more possible applications of the definite integral in geometry and physics in this section.

We assume that f is a non–negative and continuous function in the interval $\langle a; b\rangle$ everywhere in this section.

V.7.1. The volume of a circular body.

Let us denote, as in paragraph V.1.1, by R the region which is bounded by the graph of function f, by the x–axis and by the straight lines $x = a$, $x = b$. (See Fig. 39.) The revolution of region R about the x–axis leads to a circular body. Its volume is:

$$V = \pi \int_a^b f^2(x)\,\mathrm{d}x. \tag{V.7.1}$$

We can arrive at formula (V.7.1) by a similar consideration as in paragraph V.1.2, where we have derived the expression of the area of region R by the definite integral of function f. The interval $\langle a; b\rangle$ can be divided into infinitely many "infinitely short" parts. A typical part is the interval with the end points x and $x + \mathrm{d}x$ where $x \in \langle a; b)$. The region between the graph of function f on the interval $\langle x; x + \mathrm{d}x\rangle$ and the line segment with the end points x and $x + \mathrm{d}x$ on the x–axis is an "infinitely narrow" rectangle whose sides have the lengths $\mathrm{d}x$ and $f(x)$. Its revolution about the x–axis leads to an "infinitely thin" cylinder whose volume is $\mathrm{d}V = \pi f^2(x)\,\mathrm{d}x$. The volume of the whole circular body is the sum of infinitely many "infinitely small" numbers $\mathrm{d}V$. Thus, we obtain formula (V.7.1).

Other formulas, shown in this section, can be obtained by analogy. This is why we do not show the ideas which stand in the background.

V.7.2. The area of a circular surface.

The revolution of the graph of function f about the x–axis creates a circular surface. Its area is

$$s = 2\pi \int_a^b f(x)\sqrt{(f'(x))^2 + 1}\,\mathrm{d}x.$$

V.7.3. The length of a curve.

The graph of function f in the interval $\langle a; b\rangle$ is a curve. Its length is

$$\ell = \int_a^b \sqrt{(f'(x))^2 + 1}\,\mathrm{d}x.$$

V.7.4. The static moments, the center of mass and the moments of inertia of a curve.

Suppose that the curve from the preceding paragraph (i.e. the graph of function f) is covered by a mass with a constant longitudinal density ρ. The total mass of the curve is in this case equal to the product of ρ with the length of the curve:

$$m = \rho\,\ell = \rho \int_a^b \sqrt{(f'(x))^2 + 1}\,\mathrm{d}x.$$

The so called static moments of the curve with respect to the x– and y–axes are the integrals

$$m_x = \rho \int_a^b f(x)\sqrt{(f'(x))^2 + 1}\,\mathrm{d}x, \qquad m_y = \rho \int_a^b x\sqrt{(f'(x))^2 + 1}\,\mathrm{d}x.$$

The coordinates $x_{\scriptscriptstyle T}$ and $y_{\scriptscriptstyle T}$ of the center of mass of the considered curve satisfy:

$$x_{\scriptscriptstyle T} = \frac{m_x}{m}, \qquad y_{\scriptscriptstyle T} = \frac{m_y}{m}.$$

The so called moments of inertia of the curve according to the x– and y–axes are

$$J_x = \rho \int_a^b f^2(x)\sqrt{(f'(x))^2 + 1}\,\mathrm{d}x, \qquad J_y = \rho \int_a^b x^2\sqrt{(f'(x))^2 + 1}\,\mathrm{d}x.$$

V.7.5. The static moments, the center of mass and the moments of inertia of a planar region.

Let R be a region from paragraph V.7.1. Suppose that R is covered by a mass with a constant planar density ρ. The total mass of region R is

$$m = \rho \int_a^b f(x)\,\mathrm{d}x.$$

The static moments of region R with respect to the x– and y–axes are

$$m_x = \frac{1}{2}\,\rho \int_a^b f^2(x)\,\mathrm{d}x, \qquad m_y = \rho \int_a^b x\,f(x)\,\mathrm{d}x.$$

The coordinates $x_{_T}$ and $y_{_T}$ of the center of mass of region R satisfy:

$$x_{_T} = \frac{m_x}{m}, \qquad y_{_T} = \frac{m_y}{m}.$$

The moments of inertia of the region with respect to the x– and y–axes are

$$J_x = \frac{1}{3}\,\rho \int_a^b f^3(x)\,\mathrm{d}x, \qquad J_y = \frac{1}{2}\,\rho \int_a^b x^2\,f(x)\,\mathrm{d}x.$$

V.7.6. The static moments and the center of mass of a circular body.

Suppose that the circular body created by region R revolving about the x–axis contains a mass which is distributed with a constant volume density ρ. The total mass of the body is equal to the product of ρ and the volume of the body:

$$m = \rho\,V = \rho \int_a^b f^2(x)\,\mathrm{d}x.$$

The static moments of the body according to the xy–, xz– and yz–planes are

$$m_{xy} = 0, \qquad m_{xz} = 0, \qquad m_{yz} = \pi\rho \int_a^b x\,f^2(x)\,\mathrm{d}x.$$

The coordinates $x_{_T}$, $y_{_T}$ and $z_{_T}$ of the center of mass of the body satisfy:

$$x_{_T} = \frac{m_{yz}}{m}, \qquad y_{_T} = \frac{m_{xz}}{m} = 0, \qquad z_{_T} = \frac{m_{xy}}{m} = 0.$$

The moment of inertia of the body with respect to the axis of revolution x is

$$J_x = \frac{1}{2}\,\rho\,\pi \int_a^b f^4(x)\,\mathrm{d}x.$$

V.7.7. Remark.

You will see in the Mathematics II course how we can evaluate the lengths of more general curves, the areas of more general surfaces and the volumes of more general bodies. You will also learn how we can evaluate the mass, the static moments, the moments of inertia and the coordinates of the center of mass of curves, surfaces or bodies in the case of a generally non–constant density.

Index + Vocabulary

alternative of statements	alternativa výroků 3
angle	úhel
– between a line and a plane	– mezi přímkou a rovinou 41
– between two lines	– mezi dvěma přímkami 38
– between two planes	– mezi dvěma rovinami 42
antiderivative	primitivní funkce 108
area	obsah, plošný obsah 139, 151
asymptote	asymptota
– slant	– šikmá 91
– vertical	– svislá 91
basis	báze 8
Cauchy problem	Cauchyova úloha 129
center of curvature	střed křivosti 96
co–factor	algebraický doplněk 15
cone, conic surface	kužel, kuželová plocha 47
– elliptic	– eliptický, eliptická 47
conjunction of statements	konjunkce výroků 3
constant of integration	integrační konstanta 109
continuity of a function	spojitost funkce
– at a point	– v bodě 67
– left	– zleva 68
– on an interval	– na intervalu 70
– right	– zprava 70
coordinates	souřadnice
– Cartesian	– kartézské 33
– of the center of mass	– težiště 151
curvature	křivost 96
cut and try method	metoda půlení intervalu 104
cylinder	válec 47
– circular, of revolution	– rotační 47
– elliptic	– eliptický 47
– hyperbolic	– hyperbolický 47
– parabolic	– parabolický 47
derivative	derivace 75
– of a higher order	– vyššího řádu 80
– improper	– nevlastní 79
– left	– zleva 75
– logarithmic	– logaritmická 79
– right	– zprava 75
– second	– druhá 80
determinant	determinant 15
difference of two points	rozdíl dvou bodů 34
differential	diferenciál 80
dimension	dimenze 8

distance
– between a point and a line . . . . . .
– between a point and a plane . . . . . .
– between a point and a set . . . . . .
– between two lines . . . . . .
– between two points . . . . . .
– between two sets . . . . . .
domain of a function . . . . . .
eigenvalue of a square matrix . . . . . .
eigenvector of a square matrix . . . . . .
elementary row and column operations . . . . . .
ellipsoid . . . . . .
– circular, of revolution . . . . . .
entry of a matrix . . . . . .
equation
– characteristic . . . . . .
– differential . . . . . .
– – of the 1st order . . . . . .
– – separable . . . . . .
– for a plane . . . . . .
equivalence . . . . . .
equivalent systems . . . . . .
error estimate . . . . . .
Euler number . . . . . .
expansion of the determinant . . . . . .
extreme value
– absolute . . . . . .
– global . . . . . .
– local . . . . . .
– strict local . . . . . .
formula
– Leibniz' . . . . . .
– Mac Laurin's . . . . . .
– Newton–Leibniz . . . . . .
– Taylor's . . . . . .
function . . . . . .
– bounded . . . . . .
– complex . . . . . .
– component . . . . . .
– composite . . . . . .
– concave down . . . . . .
– concave up . . . . . .
– constant . . . . . .
– continuous . . . . . .
– decreasing . . . . . .
– defined parametrically . . . . . .

vzdálenost
– bodu od přímky 35
– bodu od roviny 39
– bodu od množiny 34
– dvou přímek 38
– dvou bodů 4, 34
– dvou množin 34
definiční obor funkce 56
vlastní číslo čtvercové matice 29
vlastní vektor čtvercové matice 29

elementární úpravy matice 14
elipsoid 44
– rotační 44
prvek matice 12
rovnice
– charakteristická 29, 31
– diferenciální 128
– prvního řádu 128
– separovatelná 130
– roviny 39
ekvivalence 3
ekvivalentní soustavy 20
odhad chyby 102, 104
Eulerovo číslo 62
rozvoj determinantu 15
extrém
– absolutní 85
– globální 85
– lokální 84
– ostrý lokální 84
vzorec
– Leibnizův 80
– Mac Laurinův 97
– Newtonův–Leibnizův 143
– Taylorův 97
funkce 61
– omezená 58
– komplexní 105
– souřadnicová 106
– složená 58
– konkávní 88
– konvexní 88
– konstantní 61
– spojitá 70, 71
– klesající 60
– definovaná parametricky 99

– even
– exponential
– increasing
– integrable
– inverse
– – trigonometric
– linear
– logarithmic
– monotonic
– non–decreasing
– non–increasing
– odd
– periodic
– power
– rational
– strictly monotonic
– trigonometric
– vector–valued
Gauss' algorithm
Gaussian elimination
graph of a function
greatest lower bound
hyperboloid
– circular, of revolution
– one–sheet
– two–sheet
implication
infimum
initial approximation
initial condition
integral
– definite
– improper
– indefinite
– Riemann's
integrand
integration
– by parts
– by substitution
interior of an interval
kernel of a linear transformation
least upper bound
Leibniz, G.W.
length
– of a vector
– of a curve
limit

– sudá 61
– exponenciální 61
– rostorucí 60
– integrovatelná 138
– inverzní 57
– cyklometrická 61
– lineární 61
– logaritmická 61
– mnotónní 60
– neklesající 60
– nerostoucí 60
– lichá 61
– periodická 61
– mocninná 61
– racionální 117
– ryze monotónní 60
– goniometrická 61
– vektorová 106
Gaussův algoritmus 14
Gaussova eliminace 20
graf funkce 59
infimum 50, 59
hyperboloid
– rotační 45
– jednodílný 45
– dvoudílný 45, 49
implikace 3
infimum 50, 59
počáteční aproximace 102
počáteční podmínka 129
integrál
– určitý 135, 139
– nevlastní 148
– neurčitý 109
– Riemannův 138
integrand 109
integrace 109
– per–partes 111
– substitucí 113
vnitřek intervalu 82
jádro lineárního zobrazení 26
supremum 50, 59
Leibniz, G.W. 134, 135
délka
– vektoru 38, 39
– křivky 151
limita

– improper
– left–hand
– lower
– of a function
– of a sequence
– of integration
– of Riemann's sum
– one–sided
– proper
– right–hand
– upper
linear combination
linear hull
linear operator
linear transformation
– composite
– inverse
linearly dependent vectors
linearly independent vectors
main diagonal
matrix
– augmented
– identity
– inverse
– of a linear transformation
– of a system of equations
– regular
– singular
– square
– transposed
– upper triangular
– zero
maximum
– absolute
– global
– local
– of a set
– strict local
minimum
– absolute
– global
– local
– of a set
– strict local
mean value
Mean Value Theorem
moments

– nevlastní 67
– zleva 68
dolní mez 138
– funkce 62
– posloupnosti 52
integrační mez 136
– Riemannova součtu 138
– jednostranná 68
– vlastní 67
– zprava 68
horní mez 135
lineární kombinace 7
lineární obal 10
lineární operátor 24
lineární zobrazení (transformace) 24
– složené 27
– inverzní 32
lineárně závislé vektory 7
lineárně nezávislé vektory 8
hlavní diagonála 11
matice 11
– rozšířená 19
– jednotková 11
– inverzní 18
– lineárního zobrazení 24
– soustavy rovnic 19
– regulární 18
– singulární 18
– čtvercová 11
– transponovaná 11
– horní trojúhelníková 11
– nulová 11
maximum
– absolutní 85
– globální 85
– lokální 84
– množiny 58
– ostré lokální 84
minimum
– absolutní 85
– globální 85
– lokální 84
– množiny 58
– ostré lokální 84
střední hodnota 139
věta o střední hodnotě 82
momenty

– of inertia
– static
natural logarithm
neighborhood
– left
– reduced
– right
Newton, I.
Newton's method
norm of a partition
null space of a linear transformation
origin of a system of coordinates
osculating circle
paraboloid
– circular, of revolution
– elliptic
– hyperbolic
parametric equation
– for a straight line
– for a plane
partial fraction
partition of an interval
point of inflection
polynomial
– cubic
– Mac Laurin's
– of degree n
– quadratic
– Taylor's
product
– cross
– dot
– scalar
– vector
quadric surface
– circular, of revolution
– in the basic position
– in a shifted position
– regular
– singular
quantifier
– existential
– universal
radius of curvature
range of a function
rank of a matrix
remainder

– setrvačnosti 151
– statické 151
přirozený logaritmus 63
okolí 51
– levé 51
– prstencové 51
– pravé 51
Newton, I. 134, 135
Newtonova metoda 104
norma dělení 137
nulový prostor lineárního zobrazení 26
počátek soustavy souřadnic 33
osculační kružnice 96
paraboloid 46, 49
– rotační 46
– eliptický 47
– hyperbolický 47
parametrická rovnice
– přímky 35
– roviny 39
částečný zlomek 118
dělení intervalu 137
inflexní bod 90
polynom 62
– kubický 62
– Mac Laurinův 97
– stupně n 62
– kvadratický 62
– Taylorův 97
součin
– vektorový 34
– skalární 33
– skalární 33
– vektorový 39
kvadrika 43
– rotační 43
– v základní poloze 44
– v posunuté poloze 44
– regulární 43
– singulární 43
kvantifikátor
– existenční 4
– všeobecný 3
poloměr křivosti 96
obor hodnot funkce 56
hodnost matice 14
zbytek 97

restriction of a function . . . . . restrikce (zúžení) funkce 57
Riemann, G. . . . . Riemann, G. 137
Riemann's sum . . . . . Riemannův součet 138
root of an equation . . . . . kořen rovnice 102
rule pravidlo
– Cramer's . . . . . – Cramerovo 23
– l'Hospital' . . . . . – l'Hospitalovo 94
– Saruss' . . . . . – Sarussovo 15
– Simpson's . . . . . – Simpsonovo 147
– trapezoidal . . . . . – lichoběžníkové 146
secant line . . . . . sečna 36
semi–axes . . . . . poloosy 48
sequence . . . . . posloupnost 52
– bounded . . . . . – omezená 58
– convergent . . . . . – konvergentní 52
– decreasing . . . . . – klesající 60
– divergent . . . . . – divergentní 52
– increasing . . . . . – rostoucí 60
– iterative . . . . . – iterační 103
– monotonic . . . . . – monotónní 60
– non–decreasing . . . . . – neklesající 60
– non–increasing . . . . . – nerostoucí 60
– strictly monotonic . . . . . – ryze monotónní 60
solution řešení
– general . . . . . – obecné 23, 130
– maximum . . . . . – maximální 128
– of a differential equation . . . . . – diferenciální rovnice 128
– particular . . . . . – partikulární 130
– trivial . . . . . – triviální 20
space prostor
– arithmetic . . . . . – aritmetický 5
– Euclidean . . . . . – Euclidův 4
– vector . . . . . – vektorový 5
straight line . . . . . přímka 35
straight lines přímky
– identical . . . . . – totožné 41
– intersecting . . . . . – různoběžné 41
– parallel . . . . . – rovnoběžné 40
– skew . . . . . – mimoběžné 36
subsequence . . . . . vybraná posloupnost 53
subspace . . . . . podprostor 10
successive approximations . . . . . postupné aproximace 102
supremum . . . . . supremum 50, 59
surface plocha
– conic . . . . . – kuželová 47
– – elliptic . . . . . – – eliptická 47
– cylindrical . . . . . – válcová 47

– – elliptic ....
– – hyperbolic ....
– – parabolic ....
– quadric ....
– – circular, of revolution ....
system of linear algebraic equations ....

– homogeneous ....
– nonhomogeneous ....
tangent line ....
theorem
– Darboux's ....
– Frobenius's ....
– fundamental of integral calculus ....
– Lagrange's ....
– on existence of absolute extrema ....
– Taylor's ....
type of a matrix ....
variable
– dependent ....
– independent ....
vector
– arithmetic ....
– directional ....
– normal ....
– opposite ....
– zero ....
velocity of motion ....
volume of a circular body ....

– – eliptická 47
– – hyperbolická 47
– – parabolická 47
kvadrika 43
– – rotační 43
soustava lineárních algebraických rovnic 19
– homogenní 20
– nehomogenní 20
tečna 75
věta
– Darbouxova 71
– Frobeniusova 22
– základní integrálního počtu 143
– Lagrangeova 82
– o existenci absolutních extrémů 71
– Taylorova 97
typ matice 11
proměnná
– závisle 56
– nezávisle 56
vektor
– aritmetický 5
– směrový 35
– normálový 39
– opačný 5
– nulový 5
rychlost pohybu 75
objem rotačního tělesa 151

List of selected symbols

$F : A \longrightarrow B$ 3
$\Longrightarrow, \Longleftrightarrow$ 3
$\forall, \exists$ 3
$\mathbb{R}^n, \mathbb{E}_n$ 4
$d(X,Y), \|X-Y\|$ 4, 35
$V(\mathbb{E}_2), V(\mathbb{E}_3)$ 5
$\vec{o}$ 5
$\dim V$ 8
$\det \boldsymbol{A}$ 15
$\boldsymbol{E}$ 11
$\boldsymbol{A}^{-1}$ 18
$(\boldsymbol{A} \,|\, \boldsymbol{B})$ 19
$\|\vec{u}\|$ 38
$\vec{u} \cdot \vec{v}$, $\vec{u} \times \vec{v}$ 39
$\mathbb{R}^*$ 52
$\mathcal{U}_\varepsilon(x)$, $\mathcal{U}(x)$ 51
$\mathcal{P}(x), \mathcal{P}_+(x), \mathcal{P}_-(x)$. 51
max, min 59, 60
sup, l.u.b. 50, 59
inf, g.l.b. 50, 59
$\{a_n\}$ 53
lim 54, 63
$a_n \to a$ 52
$D(f), H(f), G(f)$ 61
f_{-1} 57
$f \circ g$, $f * g$ 58
f', $\dfrac{\mathrm{d}f}{\mathrm{d}x}$ 75
f'', $\dfrac{\mathrm{d}^2 f}{\mathrm{d}x^2}$ 80
$f^{(n)}$, $\dfrac{\mathrm{d}^n f}{\mathrm{d}x^n}$ 80
$T_n(x)$, $R_{n+1}(x)$ .. 97
$\int f(x)\,\mathrm{d}x$ 109
$\int_a^b f(x)\,\mathrm{d}x$. 136, 139
D, $\|D\|$ 137

Bibliography

[BC] B.Budínský, J.Charvát: *Matematika I.* SNTL/Alfa, Praha 1987. (A detailed and understandable textbook, originally written for faculties of civil engineering. In Czech.)

[FT] R.L.Finney, G.B.Thomas, Jr.: *Calculus.* Addison–Wesley Publishing Company, New York, 1990, 1994, etc. (A basic textbook used especially at American universities.)

[J1] V.Jarník: *Diferenciální počet I.* Nakladatelství ČSAV, Praha 1946, 1951, 1953, 1955, 1963, etc. (A fundamental and detailed textbook on differential calculus. In Czech.)

[J2] V.Jarník: *Integrální počet I.* Nakladatelství ČSAV, Praha 1948, 1954, 1956, 1963, etc. (A fundamental and detailed textbook on integral calculus. In Czech.)

[NK] J.Neustupa, S.Kračmar: *Problems in Mathematics I.* Editorial Center of the CTU, Praha 1998. (Book of problems, useful for the exercises and recommended seminars in Mathematics I and for individual study.)

[Ne] J.Neustupa: *Matematika I.* Editorial Center of the CTU, Praha 2002. (A Czech version of this textbook.)

[Re] K.Rektorys: *Survey of Applied Mathematics.* SNTL, Praha 1998 (An extensive survey of classical parts of applied mathematics.)

[NK-CZ] J. Neustupa, S. Kračmar: *Sbírka příkladů z matematiky I,* Ediční středisko ČVUT, Praha 2003, 2006.

[Ne-EN] J. Neustupa: *Mathematics I,* Ediční středisko ČVUT, Praha 2004.

[Re-CZ] K. Rektorys: *Přehled užité matematiky,* SNTL, Praha 1963, 1968.