Mathematics I

Jiří Neustupa

Part III: **Differential Calculus**

Contents

III. Differential calculus 43

1. Sequences of real numbers 44
2. Functions – basic notions 48
3. Some concrete functions 53
4. Limits and continuity of a function 58
5. Derivative of a function 64
6. Applications of derivatives: intervals of monotonicity and convexity, l'Hospital's rule, osculating circle, curvature 72
7. Local and absolute extreme values, points of inflection 78
8. Asymptotes, behavior of a function 83
9. Taylor's polynomial, Taylor's formula 86
10. Parametrically defined functions 89
11. Approximate solution of a nonlinear equation $f(x) = 0$ 92

III. Differential calculus

The extended set of real numbers. The <u>*extended set of real numbers*</u> is the union of **R** with the two point set that contains the elements called <u>*plus infinity*</u> and <u>*minus infinity*</u> and denoted $+\infty$ (or just ∞) and $-\infty$. We denote the extended set of real numbers $\mathbb{R}^*$. Its elements $-\infty$ and ∞ are called the <u>*improper points*</u>, other elements (i.e. numbers from $\mathbb{R}$) are called the <u>*proper points*</u>. 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
$x + \infty = \infty, \quad x + (-\infty) = -\infty, \quad x - \infty = -\infty,$
$x - (-\infty) = \infty, \quad x/\infty = x/(-\infty) = 0;$

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 $x \in \mathbb{R} - \{0\}$ 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 $(x < 0)$,
$\infty/x = \operatorname{sgn} x \cdot \infty, \quad (-\infty)/x = \operatorname{sgn} x \cdot (-\infty).$

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

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

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

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

It is easy to observe that <u>not</u> 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 <u>*supremum*</u> of set M if

a) $\forall\, x \in M : \ x \leq K$,

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 <u>*least upper bound*</u>, l.u.b. M.

By analogy, we can define the <u>*infimum*</u> of set M. It is denoted by $\inf M$ and it is the greatest of all numbers $L \in \mathbb{R}^*$ such that $\forall\, x \in M : \ x \geq L$. The infimum is also often called the <u>*greatest lower bound*</u> and denoted by 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 $U_\varepsilon(x)$ or simply $U(x)$. (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 $U(x) - \{x\}$. This neighborhood will be denoted by $P(x)$.

A *neighborhood of* ∞ ($\equiv$ the reduced neighborhood of ∞) is any interval (a, ∞) (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 ∞, we can denote it either $U(\infty)$ or $P(\infty)$. Similarly, the neighborhood of $-\infty$ can be denoted by $U(-\infty)$ or by $P(-\infty)$.

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 $P_-(x)$ (respectively $P_+(x)$).

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.

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” or “monotonous”.

A sequence that is both increasing and decreasing does not exist. a sequence that is both non=increasing and non=decreasing is constant.

III.1.3. Limit of a sequence. The number $a \in \mathbb{R}^*$ is called the *limit* of the sequence $\{a_n\}$ if

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

(We read it: For every neighborhood $U(a)$ 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 $a_n \in U(a)$.) The fact that a is the limit of the sequence $\{a_n\}$ is written down in this way: $\lim a_n = a$ or shortly $a_n \to a$.

III.1.4. 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.1.6 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.1.10.

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.1.5. 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 ∞), then we say that this sequence is *convergent*. A sequence which either has no limit or has an infinite limit is called *divergent*.

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

III.1.6. Theorem. *Every sequence has at most one limit.*

P r o o f : We show the 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 $U(a')$ and $U(a'')$ which are disjoint (i.e. their intersection is an empty set). According to (III.1.1), to $U(a')$ there exists $n_1 \in \mathbb{N}$ so that the following implication holds for all $n \in \mathbb{N}$: $n \geq n_1 \implies a_n \in U(a')$. (III.1.1) also implies that to the neighborhood $U(a'')$ there exists $n_2 \in \mathbb{N}$ such that the implication $n \geq n_2 \implies a_n \in U(a'')$ 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 $U(a')$ and also to $U(a'')$. However, this is impossible since $U(a')$ and $U(a'')$ have no common points. This is the desired contradiction.

III.1.7. 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 a <u>*subsequence*</u> of the sequence $\{a_n\}$. This subsequence is shortly denoted by $\{a_{k_n}\}$.

III.1.8. Example. The sequence $\{1/(2n)\}$ (i.e. the sequence $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{6}, \frac{1}{8}, \ldots\}$) is the subsequence of the sequence $\{1/n\}$ (i.e. of the sequence $\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots\}$).

III.1.9. Theorem. *If the sequence $\{a_n\}$ has a limit equal to a then its every subsequence has the same limit a.*

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

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

(III.1.1) implies that to any $U(a)$ there exists $n_0 \in \mathbb{N}$ such that $a_n \in U(a)$ 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 $(m \geq n_1) \implies (a_{k_m} \in U(a))$ is satisfied for $\forall\, m \in \mathbb{N}$.

III.1.10. 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 ∞. 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.1.9, 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.1.11. Theorem. *Let* $\lim a_n = a$ *and* $\lim b_n = b$. *Then the following equalities hold:*

a) $\lim\,(a_n + b_n) \;=\; a + b,$ *b)* $\lim\,(a_n - b_n) \;=\; a - b,$

c) $\lim\,(a_n \cdot b_n) \;=\; a \cdot b,$ *d)* $\lim\,(a_n / b_n) \;=\; a/b,$

(provided that the expressions on the right–hand sides make sense and in the case d) the quotient a_n/b_n makes sense for all $n \in \mathbb{N}$).

III.1.12. Remark. Thus, the right– side cannot contain for example the expressions $\infty+(-\infty)$, $\infty-(+\infty)$, $(\pm\infty)/(\pm\infty)$, $0\cdot(\pm\infty)$ and $a/0$. Moreover, when computing the limit of the quotient 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 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.1.13. Example. $\lim \dfrac{3n^2+2n-1}{2n^2+1000} = \lim \dfrac{3+(2/n)-(1/n^2)}{2+(1000/n^2)} =$

$$= \lim \frac{3\,+\,2/\infty\,-\,1/\infty^2}{2\,+\,1000/\infty^2} = \frac{3+0-0}{2+0} = \frac{3}{2}\,.$$

III.1.14. Example. $\lim\,(\sqrt{n}-\sqrt{2n}) = \lim \dfrac{(\sqrt{n}-\sqrt{2n})\,(\sqrt{n}+\sqrt{2n})}{\sqrt{n}+\sqrt{2n}} =$

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

III.1.15. Theorem. *Let* $\{a_n\}$, $\{b_n\}$, $\{c_n\}$ *be such sequences that* $\lim a_n = \lim c_n = c$ *and* $\forall\, n \in \mathbb{N}:\ \ a_n \leq b_n \leq c_n$. *Then* $\lim b_n = c$.

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

III.1.16. Example. We show that $\lim \sqrt[n]{n} = 1$. Put $\sqrt[n]{n} = 1+\delta_n$. Raising this equality to the n–th power, we obtain:

$$n = 1 + \genfrac{}{}{0pt}{}{n}{1}\delta_n + \genfrac{}{}{0pt}{}{n}{2}\delta_n^2 + \genfrac{}{}{0pt}{}{n}{3}\delta_n^3 + \ldots + \delta_n^n\,.$$

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

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

Obviously, $\lim \sqrt{2/(n-1)} = 0$. Hence due to Theorem III.1.16, it also holds that $\lim \delta_n = 0$. The desired result – $\lim \sqrt[n]{n} = 1$ – now follows from the equality $\sqrt[n]{n} = 1+\delta_n$.

III.1.17. Problems. Evaluate the following limits (or prove that they do not exist):

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

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

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

R e s u l t s : a) ∞, b) 1, c) ∞, d) does not exist, e) 0, f) ∞, g) 0, h) $\frac{1}{2}$, i) 0.

III.1.18. Problems. Try to prove for yourself (e.g. by contradiction – see the proof of Theorem III.1.6) that the following assertions hold:

a) A sequence of non–negative numbers (i.e. a sequence in $[0, \infty)$) cannot have a negative limit.

b) A sequence of non–positive numbers (i.e. a sequence in $(-\infty, 0]$) 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 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 $(0, 2]$ 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 $x \in (0, 2]$;

b) $f(x) = x^2 - 1 \quad$ for $x \in (0, 2]$.

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) makes sense. For example, the function $f(x) = \sqrt{x-2}$ (with no more specifications) has the domain $[2, \infty)$.

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 meaning. (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 $[\,D(f) \cap D(g)\,] - \{x \in D(g);\ g(x) = 0\}$.

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, defined only in the set M, which assigns to each $x \in M$ the same value as the function f, i.e. the value $f(x)$, is said to be the *restriction* of the function f to the set M. It is denoted by $f|_M$. The set of all values of the function f on the set M can be denoted either by $R(f|_M)$ or by $f(M)$.

The graph of function f and a chosen subset $M \subset D(f)$ are sketched on Fig. 14. The graph of the restriction of f to the set M is drawn as a thick curve.

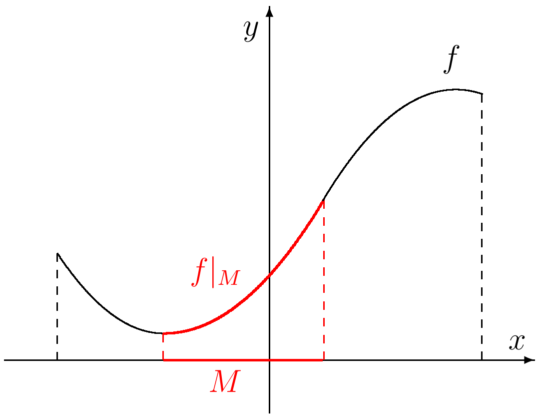


Fig. 14

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.)

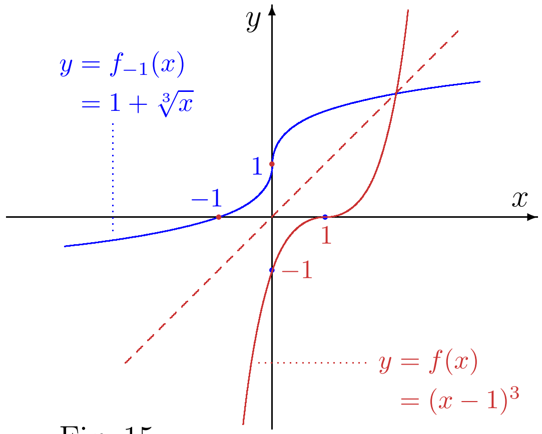


Fig. 15

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 $h = f * g$ or $h = f \circ g$. f is called the *outside function* and g the *inside function*.

III.2.8. Bounded function. Function f is called *bounded above* (or *upper bounded*) if there exists a number $K \in \mathbb{R}$ such that $\forall\, x \in D(f) \;:\; f(x) \le K$. ($K$ is called the "upper bound" of function f.)

By analogy, we say that function f is *bounded below* (or *lower bounded*) if there exists a number $L \in \mathbb{R}$ such that $\forall\, x \in D(f) \;:\; f(x) \ge L$. ($L$ is called the "lower bound" of function f.)

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 $\forall\, x \in M \;:\; f(x) \le K$ (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*.

For example, the restriction of the function $1/x$ to the interval $(-\infty, 0)$ is upper bounded, however not lower bounded. On the other hand, the restriction of the same function to the interval $(0, \infty)$ is lower bounded, but not upper bounded. (See Fig. 21 b.)

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 $\forall\, x \in D(f) \;:\; f(x) \le f(x_0)$. We write: $f(x_0) = \max f$.

By analogy, function f has its *minimum* at the point $x_0 \in D(f)$ if $\forall\, x \in D(f) \;:\; f(x) \ge f(x_0)$. 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 $\forall\, x \in M \;:\; f(x) \le f(x_0)$. We write: $f(x_0) = \max_M f$. Other often used denotations of maximum of function f on the set M are:
$\max\limits_{M} f$, $\max\limits_{x \in M} f(x)$.

Based on the above definitions, think for yourself why the following equalities hold: $\max_M f = \max f|_M$ and $\min_M f = \min f|_M$.

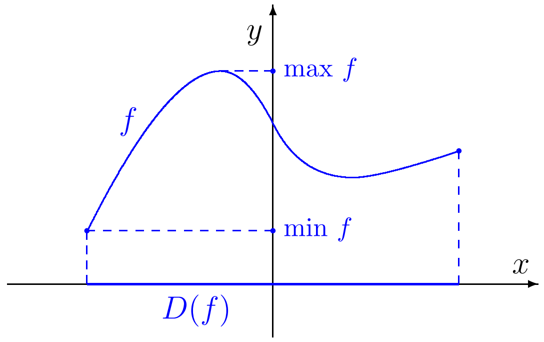


Fig. 16a

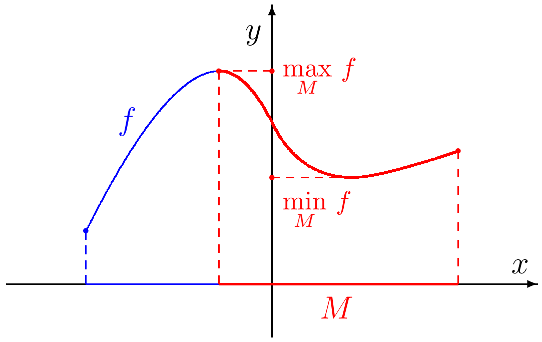


Fig. 16b

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 of the set of values of function f (i.e. of the set $R(f)$) is called the *supremum of function f*. We denote it $\sup f$ (or l.u.b. f, which means the *least upper bound* of f).

By analogy, the infimum of the set $R(f)$ is called the *infimum of function f*. We denote it $\inf f$ (or g.l.b. f, which means the *greatest lower bound* of f).

If $M \subset D(f)$, then the supremum of the set of values of function f on set M (i.e. the set $f(M)$) is called the *supremum of function f on set M*. We denote it $\sup_M f$, $\sup\limits_M f$ or $\sup\limits_{x\in M} f(x)$.

By analogy, the infimum of the set $f(M)$ is called the *infimum of function f on set M*. We denote it $\inf_M f$, $\inf\limits_M f$ or $\inf\limits_{x\in M} f(x)$.

Note that we can also always write l.u.b. instead of sup and g.l.b. instead of inf.

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 this example: the function $f(x) = 2x$ has neither maximum, nor minimum, in the open interval $(0,4)$. If you write that $\min_{(0,4)} f = 0$ and $\max_{(0,4)} f = 8$ then it is a mistake, because 0 and 8 do not belong to the set of function values of f in the interval $(0,4)$. (Recall that $\min_{(0,4)} f$, respectively $\max_{(0,4)} f$, if they exist, should be the least, respectively the greatest, function values of the function f in the interval $(0,4)$, which means that they must be the function values.) On the other hand, the supremum and the infimum of the function f in the interval $(0,4)$ exist: $\inf_{(0,4)} f = 0$ and $\sup_{(0,4)} f = 8$.

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 adjectives "monotone" or "monotonous".

A function that is increasing on the whole its domain is shortly called *increasing,* without a closer specification where. We can by analogy introduce the notions of functions *decreasing*, *non-increasing*, *non-decreasing*, etc, without specification where.

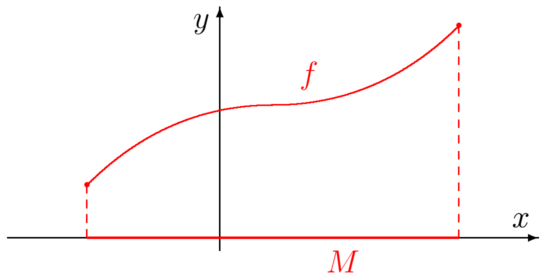


Fig. 17a

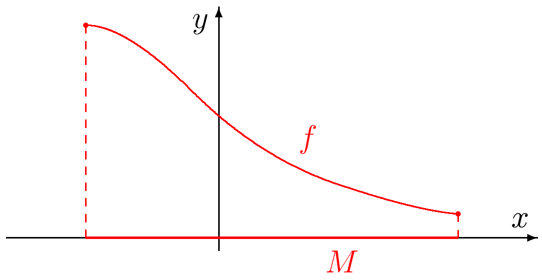


Fig. 17b

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

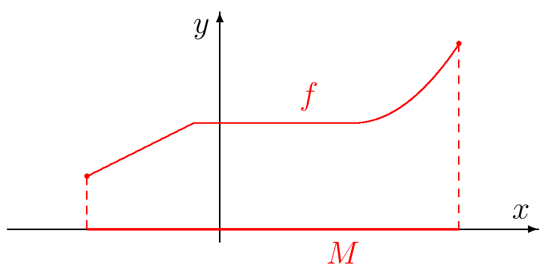


Fig. 18a

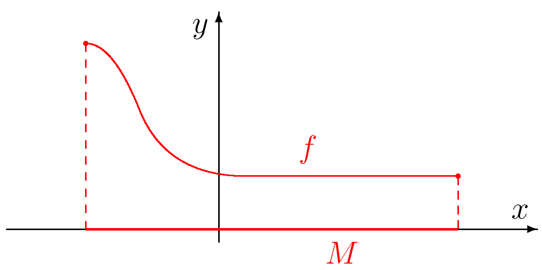


Fig. 18b

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.

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*: $f(x) = \cot\, x$,
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), the exponential and logarithmic functions with base e (where e is the so called Euler number – paragraph III.3.3) and the inverse trigonometric functions (paragraphs III.3.5 – III.3.6). We return to the power function in paragraph III.3.4.

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

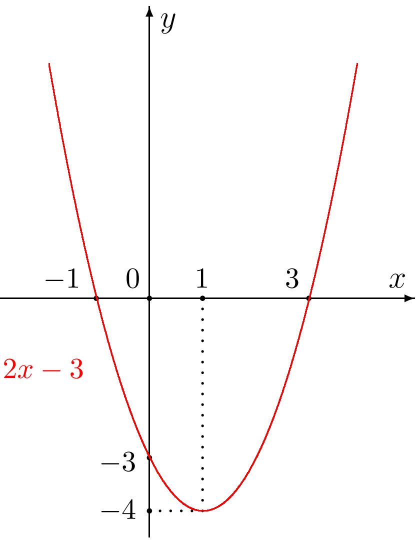
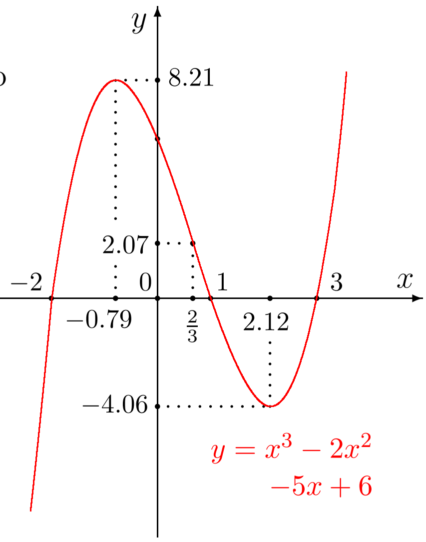


Fig. 19b



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.5.13, 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. Many formulas by which one can express e were discovered in the past. For example:

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

Instead of e^x, we also often 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 $\log_e x$, it is usually denoted by $\ln x$.

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

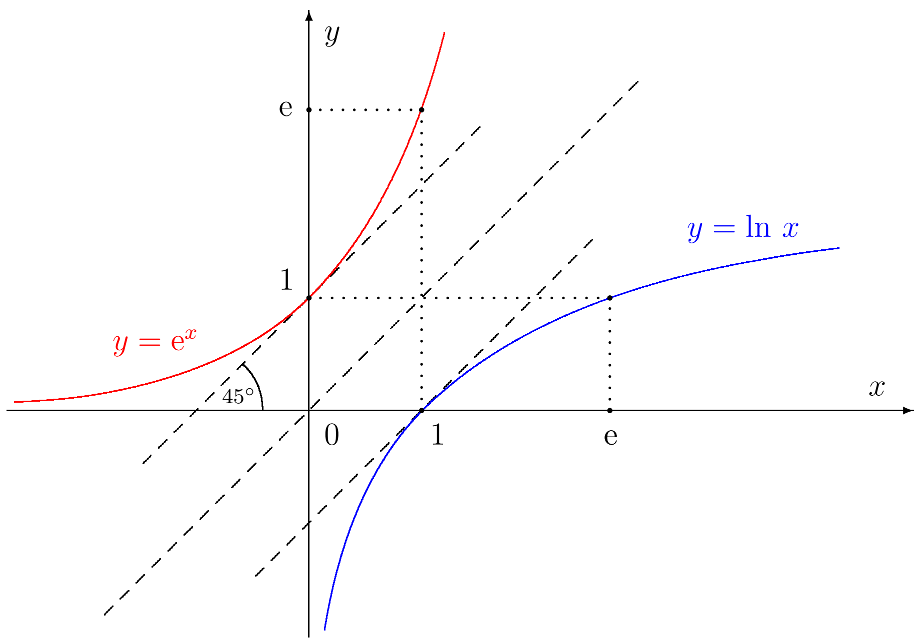


Fig. 20

It follows from the definition of the natural logarithm (i.e. from the fact that it is the inverse function to the exponential function $y = e^x$) that $\ln e^x = x$ for all $x \in \mathbb{R}$, and $e^{\ln x} = x$ 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 $D(f) = (-\infty, \infty) - \{0\}$. (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 $\ln x$ makes 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 $[0, \infty)$.
- For those α that are rational and can be therefore expressed as 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 $x = -8$: $(-8)^{1/3} = \sqrt[3]{-8} = -\sqrt[3]{8} = -2$.

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

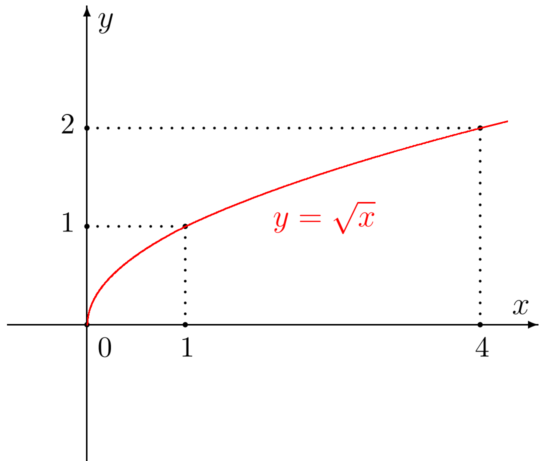


Fig. 21a

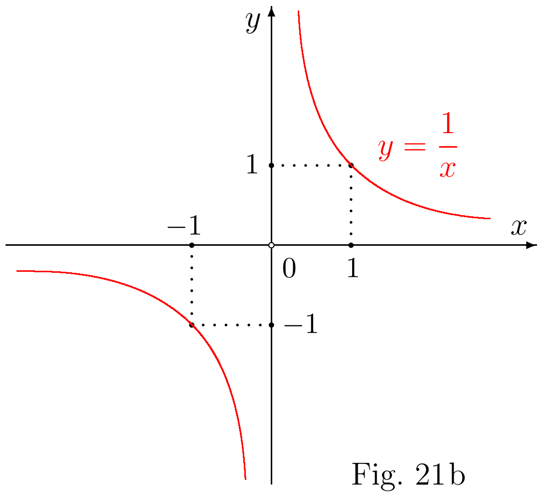


Fig. 21b

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. $a = 2$ or $a = -2$, to the square root of a^2 and the quadratic equation $x^2 = a^2$.

III.3.5. Inverse trigonometric functions $\operatorname{asin} x$ **(arcsine) and** $\operatorname{acos} x$ **(arccosine).** 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 $[-\pi/2, \pi/2]$ is one–to–one. The inverse function to this restriction is called *arcsine*. We denote it asin. The domain of arcsine is the interval $[-1, 1]$ and the range is the interval $[-\pi/2, \pi/2]$. Thus, for $x \in [-\pi/2, \pi/2$ and $y \in [-1, 1]$, we have: $y = \sin x \iff x = \operatorname{asin} y$.

You can see a part of the graph of the function sine (the so called sinusoid) on Fig. 22a. The restriction to the interval $[-\pi/2, \pi/2]$ is drawn thickly. The graph of the restriction and the inverse function asin are drawn in greater detail on Fig. 22b. The graph of the restriction is drawn in red, the graph of asin is drawn in blue.

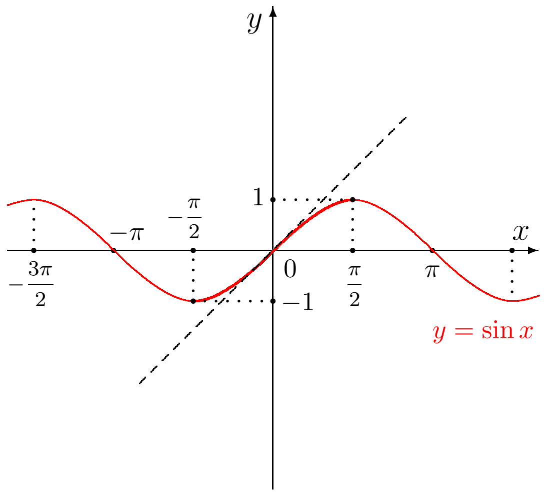


Fig. 22a

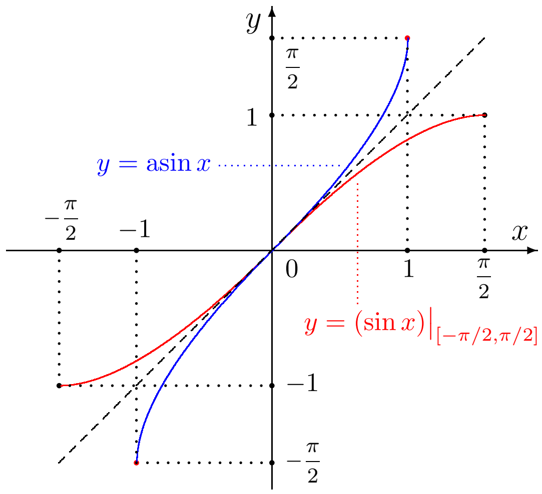


Fig. 22b

Similarly, the inverse function to the restriction of the function cosine to the interval $[0,\pi]$ is called *arccosine*. We denote it acos. Its domain is the interval $[-1,1]$ and the range is the interval $[0,\pi]$. Thus, for $x \in [0,\pi]$ and $y \in [-1,1]$, we have $y = \cos x \iff x = \operatorname{acos} y$.

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

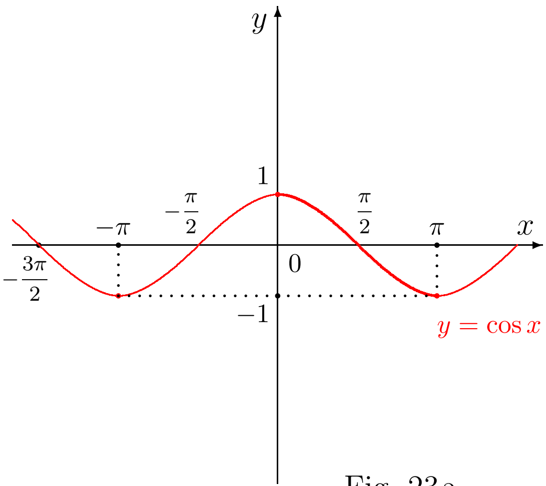


Fig. 23a

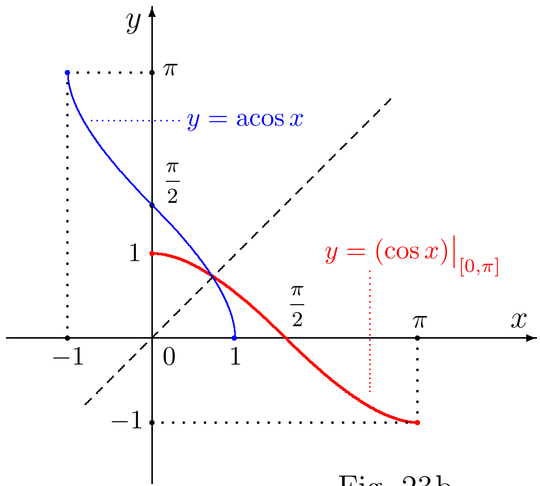
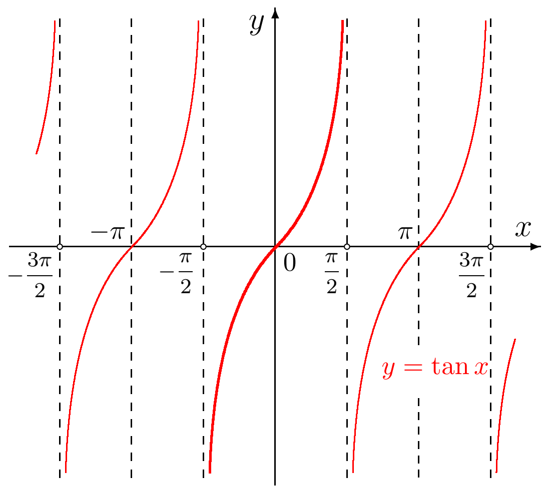


Fig. 23b

III.3.6. Inverse trigonometric functions $\operatorname{atan} x$ **(arctangent) and** $\operatorname{acot} x$ **(arccotangent).** The inverse function to the restriction of the function tangent to the interval $(-\pi/2,\pi/2)$ is called *arctangent* and denoted by atan. Its domain is the interval $(-\infty,\infty)$ and the range is the interval $(-\pi/2,\pi/2)$. Thus, for $y \in (-\infty,\infty)$ and $x \in (-\pi/2,\pi/2)$, we have $y = \tan x \iff x = \operatorname{atan} y$.



Obr. 24a

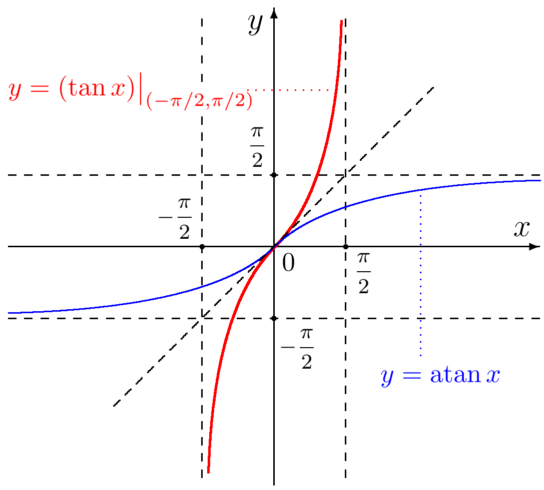


Fig. 24b

A part of the graph of the function tangent with the marked restriction to the interval $(-\pi/2,+\pi/2)$ is drawn on Fig. 24a. The graph of this restriction and the graph of the inverse function arctangent are drawn on Fig. 24b.

The inverse function to the restriction of the function cotangent to the interval $(0, \pi)$ is called *arccotangent* and denoted by acot. 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 inverse function arccotangent yourself .

Note that instead of asin, acos, atan and acot , one can often also find the notation $\sin^{-1}$, $\cos^{-1}$, $\tan^{-1}$ and $\cot^{-1}$ in literature.

III.4. Limit and continuity of a function

III.4.1. Example. The domain of the function $f(x) = (\mathrm{e}^x - 1)/x$ is the set $\mathbb{R} - \{0\}$. 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.4.2. The limit of a function. Assume that $c \in \mathbb{R}^*$ and the domain of function f contains some reduced neighborhood $P(x_0)$. If for each sequence $\{x_n\}$ in $P(x_0)$ 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: $\lim_{x \to x_0} f(x) = a$.

You will see later that the limit from example III.4.1 is indeed: $\lim_{x \to x_0} \dfrac{\mathrm{e}^x - 1}{x} = 1.$

III.4.3. 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 ∞ or $a = \infty$, is called an *improper limit*.

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

III.4.4. Theorem. *Function f can have at any point $x_0 \in \mathbb{R}^*$ at most one limit.*

III.4.5. Example. Function $f(x) = \sin x$ has no limit at ∞. 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 $x_n = \pi/2 + n\pi$, does not satisfy this implication. This sequence has the limit ∞, 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.

III.4.6. Theorem. *Let* $\lim_{x \to x_0} f(x) = a$ *and* $\lim_{x \to x_0} g(x) = b$. *Then*

a) $\lim_{x \to x_0} [f(x) + g(x)] \;=\; a + b,$

b) $\lim_{x \to x_0} [f(x) - g(x)] \;=\; a - b,$

c) $\lim_{x \to x_0} [f(x) \cdot g(x)] \;=\; a \cdot b,$

d) $\lim_{x \to x_0} [f(x)/g(x)] \;=\; a/b,$

if the expressions on the right hand sides make sense.

III.4.7. 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.4.8. Remark. If for example $\lim_{x \to x_0} f(x) \;=\; \lim_{x \to x_0} g(x) \;=\; \infty$ then the limit $\lim_{x \to x_0} f(x)/g(x)$ cannot be computed by Theorem III.4.6 because the expression ∞/∞ makes no sense.

III.4.9. 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 $P(x_0)$, then $\lim_{x \to x_0} f(x)/g(x) = \infty$. The limit of the quotient $f(x)/g(x)$ cannot be evaluated by means of Theorem III.4.7 (because the fraction $a/0$ makes 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 $f(x)/g(x)$ approaches ∞. A similar reasoning can also be used in the case when $a < 0$ or $g(x) < 0$ for all $x \in P(x_0)$.

III.4.10. Example. $\lim_{x \to 0} \dfrac{1}{x^2} = \infty$ – this is the consequence of Remark III.4.9.

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.5.34; you will appreciate it as a useful aid for evaluations of limits of a quotient of two functions.

III.4.11. One–sided limits. Suppose that $x_0 \in \mathbb{R}$ and the domain of function f contains some right neighborhood $P_+(x_0)$. If for every sequence $\{x_n\}$ in $P_+(x_0)$ 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: $\lim\limits_{x \to x_0-} f(x) = a$.

Comparing these definitions with the definition of the "both–sided" limit of a function (paragraph III.4.2), we obtain the following theorem:

III.4.12. 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.4.4 and III.4.7 also hold for one sided limits.

III.4.13. 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 $P(x_0)$ 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.4.14.* 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 $\lim\limits_{x \to x_0} f(x) = a$ can also be defined in this way:

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

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.4.15. Continuity of a function – motivation. You can see graphs of two functions in Fig. 25 a and Fig. 25 b. 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.

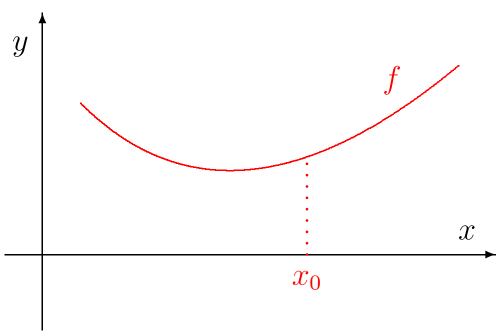


Fig. 25 a

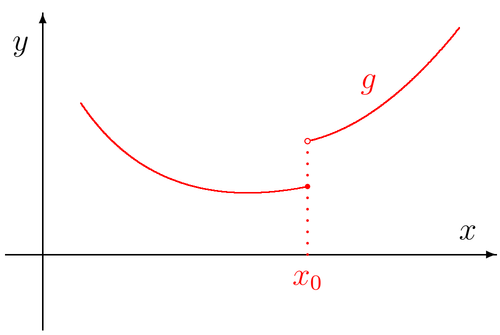


Fig. 25 b

III.4.16. Continuity of a function at the point. We say that function f is <u>*continuous*</u> at the point $x_0 \in D(f)$ if

$$\lim_{x \to x_0} f(x) = f(x_0). \tag{III.4.1}$$

III.4.17. Remark. If you read this definition and definition III.4.2 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 $U(x_0)$).

Function g in Fig. 15 b 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.4.12). Thus the equality (III.4.1) does not hold and, consequently, function g is not continuous at the point x_0.

III.4.18. Right continuity and left continuity. Function f is called <u>*right continuous*</u> (respectively <u>*left continuous*</u>) at the point $x_0 \in D(f)$, if

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

The following assertion is an easy consequence of Theorem III.4.12: *Function f is continuous at point x_0 if and only if it is right and left continuous at this point.*

III.4.19. 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.4.20. 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 <u>*continuous on the interval I*</u>, 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.4.21. 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 $\lim x_n = x_0$. We need to show that $\lim f(x_n) = f(x_0)$, i.e. $\lim (3x_n^2 + 2x_n - 1) = 3x_0^2 + 2x_0 - 1$. By Theorem III.4.7 we get: $\lim (3x_n^2+2x_n-1) = \lim 3x^2+\lim 2x_n-\lim 1 = 3\cdot[\lim x_n]^2+2\cdot\lim x_n-1 = 3x_0^2+2x_0-1$.

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.4.22, III.4.23 and III.4.25.

III.4.22. 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.4.23. Remark. If you sketch the graph of function tangent, you can see that function tangent is continuous in each interval of the type $(-\pi/2+k\pi, \pi/2+k\pi)$ (where

k is an integer). Function tangent is not continuous at the points $\pi/2 + k\pi$ (where k is an integer). However, this is not a contradiction with the assertion of Theorem III.4.22 because the points $\pi/2 + k\pi$ do not belong to the domain of function tangent.

III.4.24. Theorem (on continuity of the sum, difference, product, quotient and 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 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 f/g is also continuous on the interval I.

III.4.25. 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 $f*g$ 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 $f*g$ 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 yourself in appropriate pictures.

III.4.26. 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.4.27. 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.4.28, 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.4.28. 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.4.29. Theorem (on the existence of maximum and minimum). *A function which is continuous on a closed bounded interval $[a,b]$ has its maximum and minimum on this interval. (Thus, $\max_{x\in[a,b]} f(x)$ and $\min_{x\in[a,b]} f(x)$ exist.)*

III.4.30. Theorem. *Let function f be continuous at point x_0. If $f(x_0)>0$ then there exists a neighborhood $U(x_0)$ such that $f(x)>0$ for all $x\in U(x_0)$.*

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 $U(x_0)$, one can find a point x such that $f(x) \le 0$. Since $U(x_0)$ can be taken smaller and smaller, for example $U(x_0) = U_{1/n}(x_0)$, one gets a sequence $\{x_n\}$ such that $[\, x_n \to x_0 \,] \wedge [\, f(x_n) \le 0 \text{ for all } n \in \mathbb{N} \,]$. 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 prove by analogy that *if function f is continuous at point x_0 and $f(x_0) < 0$ then there exists a neighborhood $U(x_0)$ such that $f(x_0) < 0$ for all $x \in U(x_0)$.*

III.4.31. 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.4.16) and from Theorem III.4.23 that if f is any of the functions named in Theorem III.4.23 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.4.32. Theorem (1st theorem on the 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.4.33. 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.4.34. 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.4.35. Remark. Theorems III.4.32 and III.4.34 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.4.36. Example. Evaluate the limit $\lim\limits_{x \to 1+} \operatorname{atan} x/(x-1)$. The inside function $g(x) = x/(x-1)$ has the right–hand limit equal to ∞, 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 $x/(x-1)$ is ∞.) The outside function arc tangent has the limit equal to $\pi/2$ at ∞. Hence we obtain: $\lim\limits_{x \to 1+} \operatorname{atan} x/(x-1) = \pi/2$.

III.4.37. Example. Evaluate the limit $\lim\limits_{x \to 0} (\sin x)/x$. Let us first deal with the right–hand limit. If x belongs to a right neighborhood of 0, for example to the interval $(0, \pi/2)$, 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 $(\sin x)/x$ is "closed" between the function $\cos x$ (from below) and the constant function 1 (from above) for $x \in (0, \pi/2)$. Since $\lim_{x\to 0+} \cos x = 1$ and $\lim_{x\to 0+} 1 = 1$, we also have: $\lim_{x\to 0+} (\sin x)/x = 1$. (See Remark III.4.13.) We can show similarly that the left– limit is also equal to one. Applying Theorem III.4.12, 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.5.34. (It will be based on the so called l'Hospital Rule.) Nevertheless, we regard the procedure used here as instructive, too.

III.4.38. Problems. Evaluate the following limits.

a) $\lim_{x\to\infty} \frac{x^2 - 2x + 100}{3x^2 + 15x - 5}$ b) $\lim_{x\to 3} (x^3 + 2x - 7)$ c) $\lim_{x\to 0+} \frac{1}{\sqrt{x}}$

d) $\lim_{x\to 1-} \operatorname{acos} x$ e) $\lim_{x\to -1} \frac{x^5 + x^3 - x + 1}{x^4 - 2x + 1}$ f) $\lim_{x\to 2} \left(\frac{8}{x^2 - 4} - \frac{2}{x - 2}\right)$

g) $\lim_{x\to -3} \frac{\sqrt{4 + x} - 1}{x + 4}$ h) $\lim_{x\to -\infty} (x^5 - 2x^2)$ i) $\lim_{x\to\infty} (x^5 - 10x^4 + 155)$

j) $\lim_{x\to 0} \ln\left(\frac{x + 1}{x}\right)$ k) $\lim_{x\to\infty} \frac{\sqrt{x + \sqrt{x}} - 1}{\sqrt[3]{x} - \sqrt{x}}$ l) $\lim_{x\to\infty} x \cdot \sin x$

m) $\lim_{x\to -\infty} \operatorname{acot} x$ n) $\lim_{x\to\infty} \frac{\sin x}{x}$ o) $\lim_{x\to\infty} \sqrt{x \cdot (\sqrt{x - 3} - \sqrt{x})}$

R e s u l t s : a) $\frac{1}{3}$, b) 26, c) ∞, d) 0, e) 0, f) $-\frac{1}{2}$, g) 0, h) $-\infty$, i) ∞, 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.5.32–III.5.34.)

III.5. 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 of the 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.5.1. 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 $X_0 = [\,x_0, f(x_0)\,]$. The slope is equal to $\tan\alpha$, where α is the angle between the tangent line and the positive semi-axis x. (See Fig. 26.)

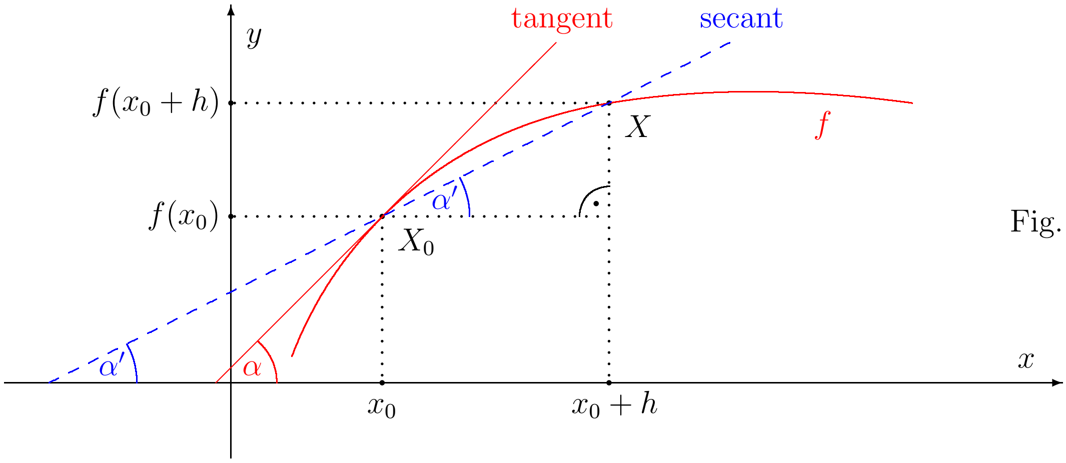


Fig. 26

In order to express $\tan\alpha$ (= the slope of the tangent line), 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 the point x_0+h moves towards the point x_0 then the secant line X_0X approaches the tangent line. Thus, $\tan\alpha$ is the limit of $\tan\alpha'$ for $h \to 0$ (if the limit exists and is finite).

III.5.2. 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 $[\,t_0, t_0+h\,]$ 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.5.3. Derivative of a function. If there exists a finite limit

$$\lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h}, \qquad \text{(III.5.1)}$$

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 that 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: $D(f') \subset D(f)$. (We remind the reader that the symbol "$\subset$" is used in such a way that it also involves the possibility "$=$".)

III.5.4. Remark. If we denote $x = x_0 + h$ then we can also write the limit (III.5.1) 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.5.5. 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 $y - f(x_0) = k \cdot (x - x_0)$.

III.5.6. The velocity of motion. Let us return to the situation described in paragraph III.5.2. 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.5.7. 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.4.12 that $f'(x_0) = k \iff f'_-(x_0) = f'_+(x_0) = k$.

III.5.8. 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.

P r o o f : 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 $U(x_0)$. For $x \in P(x_0)$, we can write: $f(x) = f(x_0) + \{[f(x) - f(x_0)]/(x - x_0)\} \cdot (x - x_0)$. 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.4.6, we get

$$\lim_{x \to x_0} f(x) = \lim_{x \to x_0} \left[f(x_0) + \frac{f(x) - f(x_0)}{x - x_0} \cdot (x - x_0)\right] = f(x_0) + f'(x_0) \cdot 0 = f(x_0).$$

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.5.9. Remark. As an immediate consequence of Theorem III.5.8, 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.5.10. 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) $[k \cdot f]'(x) \;=\; k \cdot f'(x)$,

b) $[f + g]'(x) \;=\; f'(x) + g'(x)$,

c) $[f - g]'(x) \;=\; f'(x) - g'(x)$,

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 f/g also has a derivative at the point x and it can be expressed by the formula:

e) $$\left[\frac{f}{g}\right]'(x) \;=\; \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{g^2(x)}\,.$$

P r o o f : All formulas follow from (III.5.1). We show the derivation of only one of them, for example of the formula from item d).

$$\begin{aligned}
&[f \cdot g]'(x) \;= \\
&= \lim_{h\to 0} \frac{[f \cdot g](x+h) - [f \cdot g](x)}{h} = \lim_{h\to 0} \frac{f(x+h) \cdot g(x+h) - f(x) \cdot g(x)}{h} \\
&= \lim_{h\to 0} \frac{f(x+h) \cdot g(x+h) - f(x) \cdot g(x+h) + f(x) \cdot g(x+h) - f(x) \cdot g(x)}{h} \\
&= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}\, g(x+h) \;+\; \lim_{h\to 0} f(x)\, \frac{g(x+h) - g(x)}{h} \\
&= f'(x) \cdot g(x) \,+\, f(x) \cdot g'(x).
\end{aligned}$$ □

III.5.11. Remark. The formulas a) – e) from Theorem III.5.10 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) $\left(\dfrac{u}{v}\right)' = \dfrac{u'\,v - u\,v'}{v^2}.$

III.5.12. Derivatives of some elementary functions. Evaluating the limit in (III.5.1) and applying the formulas a) – e) from Theorem III.5.10, 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) $[\,\tan\,x\,]' = \dfrac{1}{(\cos\,x)^2}$ f) $[\,\cot\,x\,]' = -\dfrac{1}{(\sin\,x)^2}$

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) makes 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.4.37 (on the limit of the quotient $(\sin\,h)/h$) and moreover, we have also used the fact that $\lim\limits_{h\to 0}(\cos\,h - 1)/h = 0$. This fact can be verified by a method similar to that used in example III.4.37.

III.5.13. Derivative of the exponential function. It was mentioned in paragraph III.2.16 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. $\lim\limits_{h\to 0}(\mathrm{e}^h - 1)/h = 1$. 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 $c\,\mathrm{e}^x$ (where c is a constant), there exist no other functions with this property.

III.5.14. Theorem (on the derivative of a composite function). *Suppose that the function g is differentiable at the point x and the function f is differentiable at the 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.5.15. 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.5.14, we get $h'(x) = f'(g(x)) \cdot g'(x) = 2\sin x \cos x$.

III.5.16. 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.5.17. Derivatives of further elementary functions. Applying Theorems III.5.14, III.5.16 and the formulas for derivatives of the functions $\sin x$, $\cos x$, $\tan x$, $\cot x$ and e^x, we can derive formulas for derivatives of some further elementary functions:

a) $[\,\mathrm{asin}\ x\,]' = \dfrac{1}{\sqrt{1-x^2}}$ for $x \in (-1, 1)$,

b) $[\,\mathrm{acos}\ x\,]' = -\dfrac{1}{\sqrt{1-x^2}}$ for $x \in (-1, 1)$,

c) $[\,\mathrm{atan}\ x\,]' = \dfrac{1}{1+x^2}$ for $x \in (-\infty, \infty)$,

d) $[\,\mathrm{acot}\ x\,]' = -\dfrac{1}{1+x^2}$ for $x \in (-\infty, \infty)$,

e) $[\,\ln x\,]' = \dfrac{1}{x}$ for $x \in (0, \infty)$,

f)] $[\,a^x\,]' = a^x \cdot \ln a$ for $a > 0$ and $x \in (-\infty, \infty)$,

g) $[\,\log_a x\,]' = \dfrac{1}{x \cdot \ln a}$ for $a > 0$, $a \neq 1$ and $x \in (0, \infty)$.

III.5.18. Remark. If function f has a derivative at point x and $f(x) > 0$, then the derivative of the function $\ln f$ at point x can be evaluated by means of the Chain Rule (paragraph III.5.14):

$$[\,\ln f(x)\,]' = \frac{1}{f(x)} \cdot f'(x) = \frac{f'(x)}{f(x)}.$$

The expression f'/f is a so called *logarithmic derivative of function f*.

III.5.19. 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 (compare with paragraph III.2.17):

$$h(x) = \exp[\,\ln(x^2 + 7x - 1)^{\sin x}\,] = \exp[\,\sin x \cdot \ln(x^2 + 7x - 1)\,].$$

This function is defined only for those x where $x^2 + 7x - 1 > 0$. (Otherwise the expression $\ln(x^2+7x-1)$ makes 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 $x_1 = (-7 - \sqrt{53})/2$ and $x_2 = (-7 + \sqrt{53})/2$. For $x \in (-\infty, x_1)$ or $x \in (x_2, \infty)$, we have:

$$h'(x) = \{\exp[\sin x \cdot \ln(x^2 + 7x - 1)]\}' =$$

$$= \exp'[\sin x \cdot \ln(x^2 + 7x - 1)] \cdot [\sin x \cdot \ln(x^2 + 7x - 1)]' =$$

$$= \exp[\sin x \cdot \ln(x^2 + 7x - 1)] \cdot \left[\cos x \cdot \ln(x^2 + 7x - 1) + \sin x \cdot \frac{2x+7}{x^2 + 7x - 1}\right] =$$

$$= (x^2 + 7x - 1)^{\sin x} \cdot \left[\cos x \cdot \ln(x^2 + 7x - 1) + \sin x \cdot \frac{2x+7}{x^2 + 7x - 1}\right].$$

III.5.20. Improper derivative. If the limit (III.5.1) 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 ∞ 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.5.1). 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.5.1)) and the improper derivative (the infinite value of the limit (III.5.1)). The notion "derivative" (without a more detailed specification) will in the following refer only to the proper (i.e. finite) derivative.

III.5.21. 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.5.5.) 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 by means of this formula:

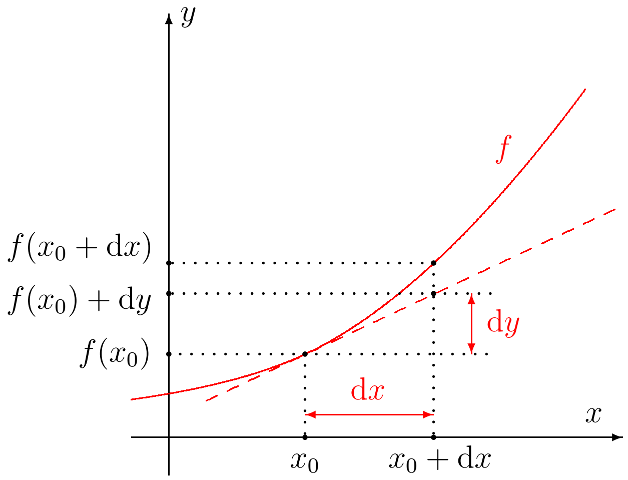


Fig. 27

$$f(x) \doteq f(x_0) + f'(x_0) \cdot (x - x_0).$$

If we write $x = x_0 + \mathrm{d}x$ then $f(x_0 + \mathrm{d}x) \doteq f(x_0) + f'(x_0) \cdot \mathrm{d}x$. (See Fig. 27.) The term $f'(x_0) \cdot \mathrm{d}x$ is the 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 $\mathrm{d}x$. The differential is denoted by $\mathrm{d}y$ or $\mathrm{d}f$. We also often write only x instead of x_0.Using this denotation, we have

$$f(x+\mathrm{d}x) \doteq f(x) + \mathrm{d}y, \quad \text{where} \quad \mathrm{d}y = f'(x)\cdot \mathrm{d}x.$$

The differential is more interesting and more important in the theory of functions of more variables.

III.5.22. Higher order derivatives. The *2nd order derivative* of function f (we denote it f'') is the derivative of the function f'. By analogy, the *3rd order derivative* of function f (we denote it f''') is the derivative of the function f'', etc.

The *n–th order derivative* of function f (for $n \in \mathbb{N}$) is denoted by $f^{(n)}$. Following this notation, we can also write $f^{(0)}$, which denotes the function f itself.

Domains of function f and its derivatives satisfy the inclusions: $D(f) \supset D(f') \supset D(f'') \supset D(f''') \supset \ldots$

The 2nd derivative (= the 2nd order derivative) of the function $y = f(x)$ is also often written down in a way which is consistent with the notation introduced in paragraph III.5.4:

$$\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}, \qquad \frac{\mathrm{d}^2}{\mathrm{d}x^2}\, f, \qquad y'', \qquad \frac{\mathrm{d}^2 y}{\mathrm{d}x^2}.$$

Other higher order derivatives can be denoted and written down analogously.

III.5.23. Leibniz' formula. *The n–th order derivative of the product $f\cdot g$ on the intersection of $D(f^{(n)})$ and $D(g^{(n)})$ can be expressed by the formula:*

$$[\,f\cdot g\,]^{(n)} \;=\; f^{(n)}\cdot g + \binom{n}{1}\cdot f^{(n-1)}\cdot g' + \binom{n}{2}\cdot f^{(n-2)}\cdot g'' + \cdots + \binom{n}{n}\cdot f\cdot g^{(n)}.$$

III.5.24. 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) $\operatorname{asin}\sqrt{x}$	f) $\sin(1-\cos x)$
g) x^x	h) $\ln(\ln x)$	i) $\ln(\operatorname{atan} x)$
j) 2^{2x+1}	k) $(3x-1)^{2001}$	l) $-x\cdot\cot x + \ln(\sin x)$
m) $[\operatorname{asin} x]^2$	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) $\dfrac{1}{\operatorname{atan} x} \cdot \dfrac{1}{x^2+1}$ (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) $\dfrac{x}{(\sin x)^2}$ (for $\sin x > 0$), m) $\dfrac{2 \operatorname{asin} x}{\sqrt{1-x^2}}$ (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 $x > -\frac{1}{2}$.

III.5.25. 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.6. Application of derivatives: intervals of monotonicity and convexity, l'Hospital's rule, osculating circle, curvature

III.6.1. Mean Value Theorem (Lagrange's theorem). *Let function f be continuous on the closed interval $[a,b]$ and let it be differentiable on the open interval (a,b). Then there exists a point $\xi \in (a,b)$ such that*

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

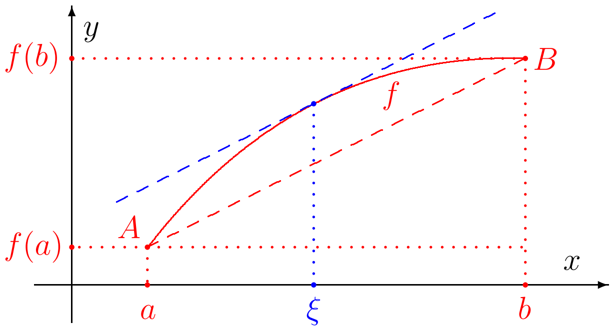


Fig. 28

III.6.2. Remark. The geometric meaning of the Mean Value Theorem is evident from Fig. 28. The theorem basically says that there exists a point ξ between a and b, such that the tangent to the graph of the function f at the point ξ is parallel to the secant line AB.

Lagrange's theorem is usually not immediately applicable to the solution of practical examples. Its great importance lies in the fact that it can be used to derive a number of other theorems, which already have immediate practical application. The first example is the following Theorem III.6.4.

III.6.3. 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 $I = (a,b]$ then $I^\circ = (a,b)$. If $I = [0,1]$, then $I^\circ = (0,1)$, etc.

III.6.4. 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). 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.6.5. 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.6.4.) 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 $(-\infty, -1]$, $[1, \infty)$ and $[-1, 1]$, 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 $(-\infty, -1]$ and $\langle 1, \infty]$.

b) Function f is decreasing on the interval $[-1, 1]$. (See Fig. 29.)

III.6.6. Remark. Students often make this mistake: they forget that Theorem III.6.4 holds on just one interval and they deduce from the continuity of the function on the two intervals $(-\infty, -1]$ and $[1, \infty)$ and positivity of the derivative on $(-\infty, -1)$ and $(1, \infty)$ that f is increasing on the union $(-\infty, -1] \cup [1, \infty)$. However, this is not be true! It only follows from Theorem III.6.4 that f is increasing on each of the intervals $(-\infty, -1]$ and $[1, \infty)$. It is not increasing on the union $(-\infty, -1] \cup [1, \infty)$. (See Fig. 29.)

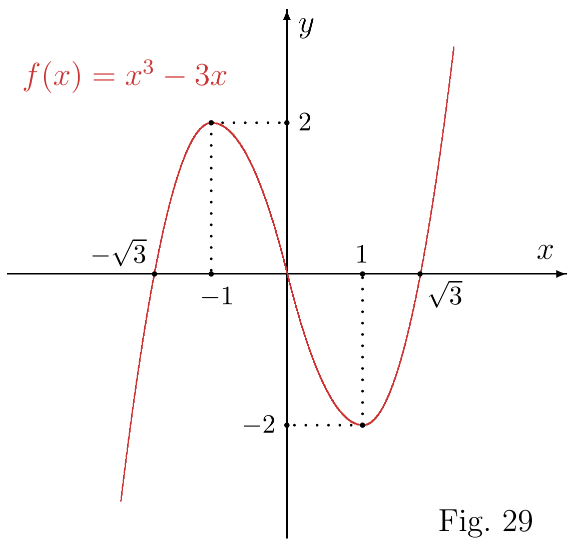


Fig. 29

III.6.7. 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 $[-1, 0]$, $[1, \infty)$ and $(-\infty, -1]$, $[0, 1]$. Thus we may conclude:

a) Function f is increasing on the intervals $[-1, 0]$ and $[1, \infty)$.

b) Function f is decreasing on the intervals $(-\infty, -1]$ and $[0, 1]$. (See Fig. 30.)

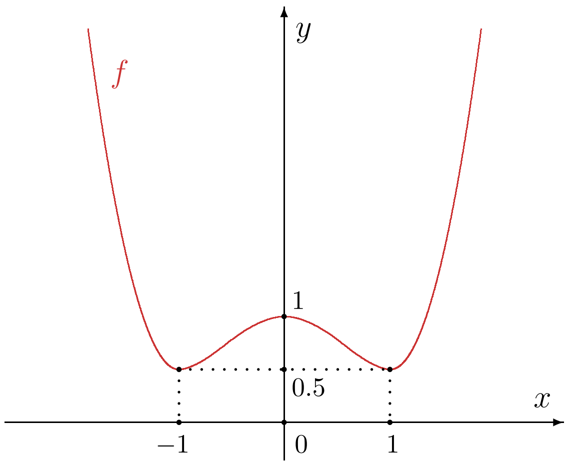


Fig. 30

III.6.8. Convex and concave functions. A function f is said to be *strictly convex on set M* if $M \subset D(f)$ and for any three points x_1, x_2, $x_3 \in M$ such that $x_1 < x_2 < x_3$, the point $Q_2 = [x_2, f(x_2)]$ lies 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 get the definition of a function *strictly concave on set M*.

If we replace the word "below" by the words "below or on", we get the definition of a function *convex on set M*.

If we replace the word "below" by the words "above or on", we get the definition of a function *concave on set M*.

You can see an example of a function strictly vonvex (respectively strictly concave) on Fig. 31 a (respectively 31 b).

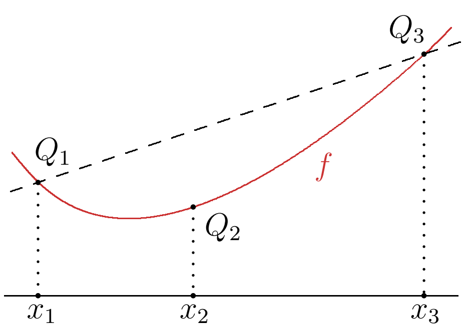


Fig. 31 a

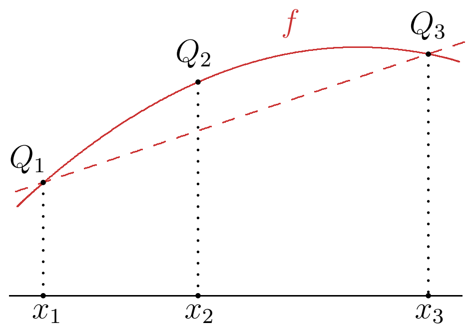


Fig. 31 b

III.6.9. Remark. The strictly convex function is a special case of a convex function and the strictly concave function is a special case of a concave function.

Note that instead of "convex function" (respectively "concave function"), some authors also use the terms "concave up function" (respectively "concave down function").

III.6.10. Remark. The condition that the point $Q_2 = [\,x_2, f(x_2)\,]$ lies 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 analytically 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.6.11. Theorem. *Let the function f be continuous in an interval I. Then the following implications hold:*

a) $f''(x) > 0$ *for all* $x \in I^\circ \implies f$ *is strictly convex in the interval* I,

b) $f''(x) \geq 0$ *for all* $x \in I^\circ \implies f$ *is convex in the interval* I,

c) $f''(x) < 0$ *for all* $x \in I^\circ \implies f$ *is strictly concave in the interval* I,

d) $f''(x) \leq 0$ *for all* $x \in I^\circ \implies f$ *is concave in the interval* I,

e) $f''(x) = 0$ *for all* $x \in I^\circ \implies f$ *is a linear function in the interval* I.

III.6.12. Remark. We omit the proof of Theorem III.6.11. Nevertheless, in order to better understand the sense of the theorem, we at least sketch its idea. Let us deal e.g. with item a): f'' is the first derivative of the function f'. The inequality $f'' > 0$ in I° thus implies that f' is increasing in I°. It means that if one moves in the interval I from the left to the right, the tangent line to the graph of f changes the direction – it leans so that its slope increases. This is, however, possible onlz in the case when the function f is convex in the interval I. (Think about this using Fig. 31 a.)

III.6.13. Example. Find the intervals where the function $f(x) = x^3 - 3x$ is convex or concave.

Solution: It is the same function as in Example III.6.5. The function f is defined and continuous on the interval $(-\infty, \infty)$. Its second derivative is $f''(x) = 6x$ (also for $x \in (-\infty, \infty)$). Solving the inequalities $f''(x) > 0$ and $f''(x) < 0$, we find out that

$$6x > 0 \iff x \in (0, \infty), \qquad 6x < 0 \iff x \in (-\infty, 0).$$

This means, due to Theorem III.6.11, that the function f is strictly convex in the interval $(0, \infty)$ and strictly concave in the interval $(-\infty, 0)$. As the function f is also continuous "up to the end point" 0 on each of these intervals (i.e. it is continuous on the intervals $[0, \infty)$ ans $(-\infty, 0]$, the statements on convexity and concavity can be extended to these intervals. Thus, we come to the conclusion:

a) The function f is strictly convex in the interval $[0, \infty)$.

b) The function f is strictly concave in the interval $(-\infty, 0]$.

III.6.14. Example. Investigate the intervals where the function $f(x) = 1 - x^2 + \frac{1}{2}\,x^4$ is convex or concave.

Solution: f is the same function as in Example III.6.7. Its graph is sketched on Fig. 30. The second derivative is $f''(x) = -2+6x^2$ (for $x \in (-\infty, \infty)$). Solving the inequalities $f''(x) > 0$ and $f''(x) < 0$, we find that

$$-2+6x^2 > 0 \iff x^2 > \tfrac{1}{3} \iff x \in \big(-\infty, -\sqrt{3}/3\big) \text{ or } x \in \big(\sqrt{3}/3, \infty\big),$$
$$-2+6x^2 < 0 \iff x^2 < \tfrac{1}{3} \iff x \in \big(-\sqrt{3}/3, \sqrt{3}/3\big).$$

Moreover, the function is continuous "up to the end points" on each of these intervals, which means that it is continuous on the intervals$(-\infty, -\sqrt{3}/3\big]$, $\big[\sqrt{3}/3, \infty\big)$ and $\big[-\sqrt{3}/3, \sqrt{3}/3\big]$. Thus, we can make the conclusion:

a) The function f is strictly convex in the intervals $(-\infty, -\sqrt{3}/3\big]$ and $\big[\sqrt{3}/3, \infty\big)$.

b) The function f is strictly concave on the interval $\big[-\sqrt{3}/3, \sqrt{3}/3\big]$.

III.6.15. Remark. Be again careful not to confuse the statement ",f is strictly convex on the intervals $\big(-\infty, -\sqrt{3}/3\big]$ and $\big[\sqrt{3}/3, \infty\big)$" (i.e. on each of these intervals) with the statement "f is strictly convex on the union $\big(-\infty, -\sqrt{3}/3\big] \cup \big[\sqrt{3}/3, \infty\big)$" (which is not true).

The next theorem shows how the derivative can also be used when computing the limits leading to the so called indefinite expressions "0/0" or "∞/∞". (The sign of the infinity plays no role.)

III.6.16. Theorem (l'Hospital's Rule). *Suppose that $c \in \mathbb{R}^*$ and the limits $\lim_{x\to c} f(x)$ and $\lim_{x\to c} g(x)$ are either both zero or both are infinite. Then*

$$\lim_{x\to c} \frac{f(x)}{g(x)} = \lim_{x\to c} \frac{f'(x)}{g'(x)},$$

provided that the limit on the right exists. (The same statment also holds for the left limit and the right limit.)

III.6.17. Remark. l'Hospital's Rule says that if the limit of the ration $f(x)/g(x)$ (for $x \to c$) leads to the indefinite expression "0/0" or "∞/∞" and if the limit $\lim_{x\to c} f'(x)/g'(x)$ exists then the limit $\lim_{x\to c} f(x)/g(x)$ also exists and both the limits are equal.

The assumption on the existence of the limit $\lim_{x\to c} f'(x)/g'(x)$ is important, because there exist examples when this limit does not exist and the limit $\lim_{x\to c} f(x)/g(x)$ still exists. (However, its value cannot be calculated by l'Hospital's Rule in such a case.)

III.6.18. Example. Calculate the limit $\lim_{x\to 0} \dfrac{\tan x}{x}$.

This limit cannot be expressed as the ratio of the limits of the numerator and the denominator, because that would yield the indefinite expression $0/0$. Nevertheless, we have:

$$\lim_{x\to 0} \frac{(\tan x)'}{x'} = \lim_{x\to 0} \frac{1/(\cos x)^2)}{1} = \lim_{x\to 0} \frac{1}{(\cos x)^2} = \frac{1}{1^2} = 1.$$

Hence $\lim_{x\to 0} (\tan x)/x$ exists and is equal to 1.

III.6.19. Example. We show again an application of l'Hospital's Rule, this time more briefly:

$$\lim_{x\to 0} \frac{x - \sin x}{x^3} = \lim_{x\to 0} \frac{1-\cos x}{3x^2} = \lim_{x\to 0} \frac{\sin x}{6x} = \lim_{x\to 0} \frac{\cos x}{6} = \frac{1}{6}.$$

(Here, we applied l'Hospital's Rule three times.)

III.6.20. Osculating circle, curvature. Suppose that function f has the second derivative $f''(x_0)$, different from zero, at the point x_0. (Then, naturally, f also has the first derivative at the point x_0.) Denote for simplicity $y_0 = f(x_0)$, $y_0' = f'(x_0)$ and $y_0'' = f''(x_0)$.

Let us solve the problem: find the circle $(x-x_c)^2 + (y-y_c)^2 = r^2$, which touches the graph of f at the point $[x_0, y_0]$ and the contact is "of order 2". It means that if we treat the circle as the graph of a function $y(x)$ in the neighborhood of the point $[x_0, y_0]$ then this function has the same value, the same 1st derivative and the same 2nd derivative at the point x_0 as the function f:

$$\text{(a)} \quad y(x_0) = y_0, \qquad \text{(b)} \quad y'(x_0) = y_0', \qquad \text{(c)} \quad y''(x_0) = y_0''.$$

The circle with these properties is called the osculating circle of the graph of f at the point $[x_0, y_0]$. Its center $C = [x_c, y_c]$ is the so called *center of curvature* and the radius r is called the *radius of curvature*. The number $1/r$ is said to be the *curvature* of the function f at the point x_0.

The osculating best approximates (of all possible circles) the graph of the function f in the neighborhood of the point $[x_0, f(x_0)]$.

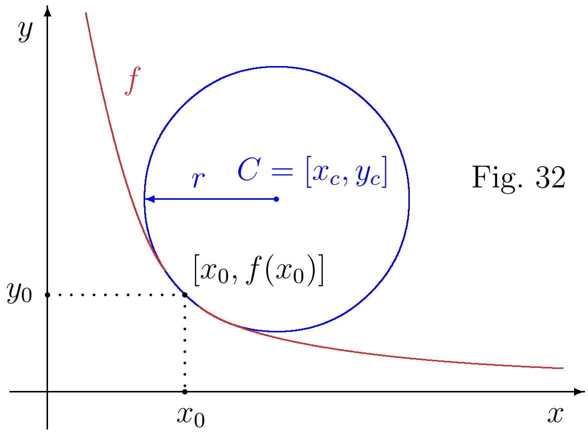


Fig. 32

How to calculate the coordinates of the center of curvature and the radius of curvature? As the graph of the function $y(x)$ coincides with the osculating circle in the neighborhood of the point $[x_0, f(x_0)]$, it satisfies the equation of the osculating circle:

$$\text{(III.6.1)} \qquad \big(x - x_c\big)^2 + \big(y(x) - y_c\big)^2 = r^2.$$

Using condition (a), i.e. substituting $x = x_0$ and $y(x) = y_0$, we obtain the equation

$$\text{(III.6.2)} \qquad (x_0 - x_c)^2 + (y_0 - y_c)^2 = r^2.$$

Differentiating the equation (III.6.1) with respect to x and using condition (b), i.e. substituting $x = x_0$, $y(x) = y_0$ and $y'(x) = y_0'$, we obtain:

$$\text{(III.6.3)} \qquad 2\,(x_0 - x_c) + 2\,(y_0 - y_c)\cdot y_0' = 0.$$

Differentiating the equation (III.6.1) two times with respect to x and applying also condition (c), i.e. substituting $x = x_0$, $y(x) = y_0$, $y'(x) = y_0'$ and $y''(x) = y_0''$, we obtain:

$$2 + 2\,{y_0'}^2 + 2\,(y_0 - y_c)\,y_0'' = 0. \tag{III.6.4}$$

The equations (III.6.2), (III.6.3) and (III.6.4) represent a system of three equations for three unknowns: x_c, y_c and r. Solving this system, we get

$$x_c = x_0 - y_0'\,\frac{1+{y_0'}^2}{y_0''}, \qquad y_c = y_0 + \frac{1+{y_0'}^2}{y_0''}, \qquad r = \frac{(1+{y_0'}^2)^{3/2}}{|y_0''|}.$$

III.7. Local and absolute extreme values, points of inflection

III.7.1. Local extreme values of a function. Suppose that function f is defined in an interval (a, b) containing the point x_0. We say that f has a *local maximum* (respectively *local minimum*) at the point x_0 if there exists a punctured neighborhood $P(x_0)$ of x_0 such that $f(x) \leq f(x_0)$ (respectively $f(x) \geq f(x_0)$) for all $x \in P(x_0)$.

If we replace the non-strict inequalities by the strict ones, we obtain the definitions of the so called *strict local maximum*, respectively the *strict local minimum*.

The local maximum and minimum are together called the *local extreme values*, the strict local maximum and minimum are together called the *strict local extreme values*.

Instead of "extreme values", we san also shortly say "extremes".

Notice that the strict local extremes are special cases of the local extremes. Fig. 33 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 the points x_2, x_4 and x_6.

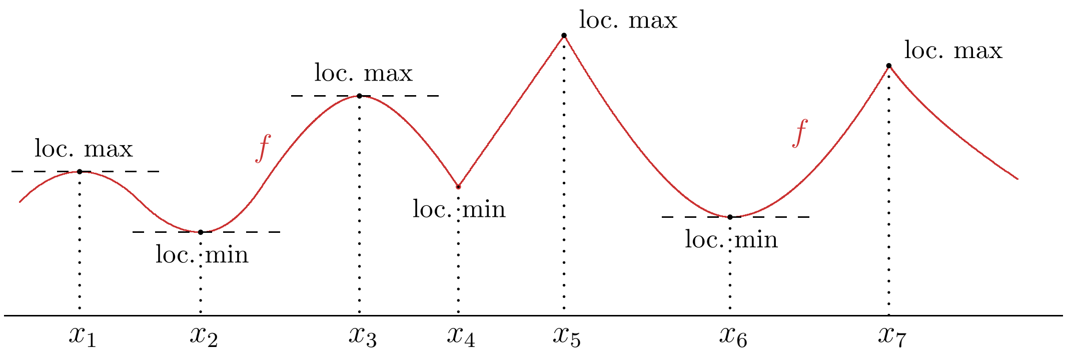


Fig. 33

III.7.2. Remark. To distinguish between extreme values of function f on its whole domain $D(f)$ or on a set $M \subset D(f)$) (defined in paragraph III.2.9) and local extreme values, we often call a maximum (respectively a minimum) of function f on $D(f)$ or on the set $M \subset D(f)$ 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.7.3. 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 $\lim_{h\to 0} [\, f(x_0 + h) - f(x_0)\,]/h = a > 0$. Thus, for each sequence $\{h_n\}$ in $D(f)$ such that $h_n \to 0$, it holds: $\lim [\, f(x_0 + h_n) - f(x_0)\,]/h_n = a > 0$. This means (by definition III.1.4) that to every $U(a)$ there exists $n_0 \in \mathbb{N}$ such that for $n \in \mathbb{N}$, satisfying the inequality $n \geq n_0$, it holds: $[\, f(x_0 + h_n) - f(x_0)\,]/h_n \in U(a)$. If we choose $U(a) = (0, 2a)$ and put $h_n = 1/n$, we get: $0 < [\, f(x_0 + 1/n) - f(x_0)\,]/(1/n) < 2a$. Using only the first part of this inequality (i.e. $0 < \dots$), we can see that $f(x_0 + 1/n) > f(x_0)$ for $n \geq n_0$. However, this means that function f cannot have a local maximum at point x_0. Similarly, by the choice $h_n = -1/n$, we can show that function f cannot have a local minimum at point x_0, either. This is the desired contradiction.

III.7.4. Remark. Remember that the condition $f'(x_0) = 0$ (if the derivative $f'(x_0)$ exists) is only a <u>necessary condition</u> for a local extreme value of function f at point x_0, but <u>it is not a sufficient condition</u>. 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.7.5. Investigation of local extreme values. The only points where function f <u>can</u> have a local extreme value in an interval I are

1) the interior points of interval I where f' is equal to zero (see Fig. 33, the points x_1, x_2 x_3 and x_6), or
2) the interior points of interval I where f' does not exist (see Fig. 33, 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.7.3 and Remark III.7.4.

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.5.8.) 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.5.19. (You will see it later.) Theorem III.5.19 uses the sign of the second derivative of f at the point x_0.

III.7.6. 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.5.10. 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.7.7. Investigation of the absolute (=global) extreme values. The only points where function f can have an absolute extreme value in an interval I are

1) the interior points of interval I where f' is equal to zero, or
2) the interior points of interval I where f' does not exist, or
3) the endpoints of interval I (if interval I is not open).

These cases are illustrated on Fig. 34 a, 34 b and 34 c.

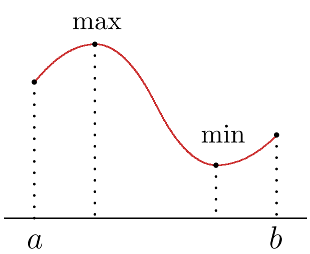


Fig. 34 a

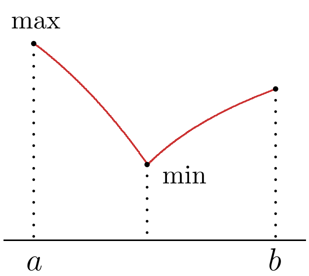


Fig. 34 b

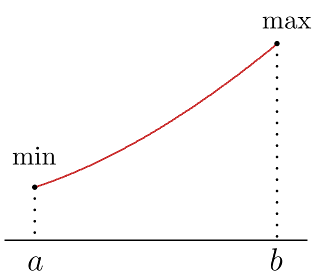


Fig. 34 c

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.4.29. (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.4.29 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.7.8. Example. $f(x) = x^2 + \dfrac{16}{x} - 16,\ \ I = [\,1,4\,]$

Find absolute extreme values of the function f in the interval I (if they exist).

Solution: $[\,1,4\,]$ 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.4.29, the absolute extreme values $\max_{[\,1,4]} f$ and $\min_{[\,1,4]} f$ exist.

Function f is differentiable at all points $x \in [\,1,4\,]$ 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 $[\,1,4]$, 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 $[\,1,4]$. 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 $\max_{[\,1,4]} f = f(2) = -4$ and $\min_{[\,1,4]} f = f(4) = 4$.

Let us now return to the question how to specify the type of a local extreme value – see also Example III.7.6.

III.7.9. Theorem (sufficient conditions for the strict local extreme value).

a) *If $f'(x_0) = 0$ and $f''(x_0) > 0$, then f has a strict local minimum at point x_0.*

b) *If $f'(x_0) = 0$ and $f''(x_0) < 0$, then f has a strict local maximum at point x_0.*

III.7.10. Remark. The proof of Theorem III.7.9 is also omitted. However, the following consideration can contribute to better understanding of the theorem: Suppose that $f')x_0) = 0$, $f''(x_0) > 0$ and in order to exclude complicated situations, we also assume that the second derivative f'' is continuous at the point x_0. The inequality $f''(x_0) > 0$ implies that there exists a neighborhood $U(x_0)$ such that $f''(x) > 0$ for all $x \in U(x_0)$. (This follows from Theorem III.4.30.) This means, by Theorem III.6.11, that the function f is strictly convex in the interval $U(x_0)$. This information and the equality $f'(x_0) = 0$ imply that f has a strict local minimum at the point x_0. (Sketch a picture.)

III.7.11. 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.5.10.) 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.7.12. Remark. If we replace the inequality $f''(x_0) > 0$ by the inequality $f''(x_0) \geq 0$ in the assumptions of Theorem III.7.9, 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.7.13. Point of inflection. Suppose that there exists a tangent to the graph of function 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.

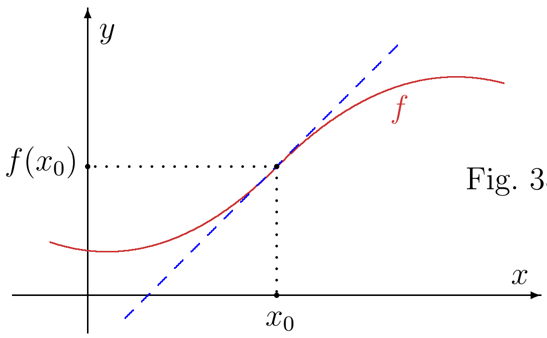


Fig. 35

III.7.14. Example. $x_0 = 0$ 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.7.15. Theorem (necessary condition for the point of inflection). *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.7.16. Remark. If $f''(x_0)$ exists then the condition $f''(x_0) = 0$ is necessary for x_0 to be an inflection point. However, it is not a sufficient condition! This means that one cannot reverse Theorem III.7.15 and to assert that the equality $f''(x_0) = 0$ implies that x_0 is the inflection point. It is evident from this example: The function $f(x) = x^4$ satisfies $f''(0) = 0$, but 0 is not the 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.7.17. Theorem (sufficient conditions for the point of inflection). *If* $f''(x_0) = 0$ *and* $f'''(x_0) \neq 0$, *then* x_0 *is an inflection point of function* f.

III.7.18. 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.5.22). 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: $x_1 = -1/\sqrt{2}$ and $x_2 = 1/\sqrt{2}$. 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) $f'''(x_1) = f'''(-1/\sqrt{2}) = -8/\sqrt{2} \cdot \exp(-0.5) \neq 0$, hence the point x_1 is the inflection point of the function f (by Theorem III.7.17).

b) $f'''(x_2) = f'''(1/\sqrt{2}) = 8/\sqrt{2} \cdot \exp(-0.5) \neq 0$, so x_2 is also the inflection point of the function f. (This follows again from Theorem III.7.17.)

III.8. Asymptotes, behavior of a function

III.8.1. Asymptotes of the graph of a function. The straight line $y = kx + q$ is a so called <u>*slant asymptote*</u> of the graph of function f for $x \to -\infty$ if $\lim_{x\to-\infty} [\,f(x) - kx - q\,] = 0$.

Similarly, the straight line $y = kx + q$ is a slant asymptote of the graph of function f for $x \to \infty$ if $\lim_{x\to\infty} [\,f(x) - kx - q\,] = 0$.

The straight line $x = x_0$ is called a <u>*vertical asymptote*</u> 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. 36 a, and an example of a vertical asymptote is shown in Fig. 36 b.

(The asymptotes are drawn in dotted lines. The slant asymptote can also be horizontal, i.e. parallel to the x–axis, in a special case.)

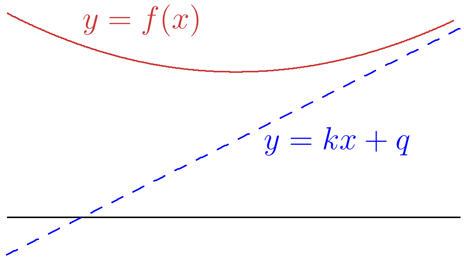


Fig. 36 a

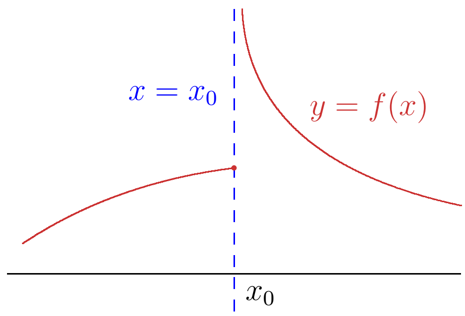


Fig. 36 b

III.8.2. Theorem (necessary and sufficient conditions for a slant asymptote). *The straight line* $y = kx+q$ *is a slant asymptote of the graph of function* f *as* $x \to \infty$ *if and only if*

$$\lim_{x\to\infty} \frac{f(x)}{x} = k \qquad \textit{and} \qquad \lim_{x\to\infty} [\, f(x) - kx \,] = q.$$

The theorem enables us to investigate whether a given function f has a slant asymptote for $x \to \infty$: We calculate the above two limits. If they both exist and are equal to finite numbers k and q then the straight line $y = kx + q$ is the slant asymptote of function f for $x \to \infty$. If some of the limits does not exist or is infinite then the function f does not have a slant asymptote for $x \to \infty$.

The theorem is also valid in the case when ∞ is everywhere replaced by $-\infty$.

III.8.3. Investigating the behavior of function f:

a) • Find the domain of f (if it is not already given).
• Find out whether the function is even, odd or 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 intersects the x–axis, the y–axis, and find maximum intervals where f is positive, respectively negative.

b) • Calculate 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 is increasing, decreasing, non-increasing or non-decreasing).
• Find local extreme values of f.

c) • Calculate the 2nd derivative f'' of function f.
• Find maximum intervals where the function is strictly convex or strictly concave (respectively only convex or concave).
• Find inflection points of f.

d) • Find asymptotes of the graph of function f.
• Sketch the graph of f.

III.8.4. 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 $[\,0,2) \cup (2,\infty)$. The information about its behavior in the set $(-\infty,-2) \cup (-2,0\,]$ follows 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 $D(f)$.) We have:

$$\lim_{x\to 2-} f(x) = \lim_{x\to 2-} \frac{x^3}{4-x^2} = \infty, \qquad \lim_{x\to 2+} f(x) = \lim_{x\to 2+} \frac{x^3}{4-x^2} = -\infty,$$

$$\lim_{x\to\infty} f(x) = \lim_{x\to\infty} \frac{x^3}{4-x^2} = \lim_{x\to\infty} \frac{x}{(4/x^2)-1} = \frac{\infty}{-1} = -\infty,$$

• $f(x) = 0 \iff \dfrac{x^3}{4-x^2} = 0 \iff x = 0,$

$f(x) > 0 \iff x \in (0,2), \qquad f(x) < 0 \iff x \in (2,\infty).$

b) • The derivative of f: $\quad f'(x) = \dfrac{3x^2(4-x^2) - (-2x)\,x^3}{(4-x^2)^2} = \dfrac{x^2(12-x^2)}{(4-x^2)^2},$

$D(f') = D(f),$

• $f'(x) = 0 \iff \dfrac{x^2(12-x^2)}{(4-x^2)^2} = 0 \iff x = 0 \text{ or } x = 2\sqrt{3},$

$f'(x) > 0 \iff x \in (0,2) \cup (2, 2\sqrt{3}), \qquad f'(x) < 0 \iff x \in (2\sqrt{3}, \infty).$

This means that f is increasing on the interval $[0,2)$ and on the interval $(2, 2\sqrt{3}]$ and decreasing in the interval $[2\sqrt{3}, \infty)$.

• 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 $[0,2)$ and due to the fact that it is odd, it is also increasing in the interval $(-2,0]$. Thus, f is increasing on the interval $(-2,2)$.

c) • The second derivative of f is:

$$f''(x) = [f'(x)]' = $$
$$= \frac{(24x - 4x^3)(4-x^2)^2 - 2(4-x^2)(-2x)(12x^2 - x^4)}{(4-x^2)^4} = \frac{8x(12+x^2)}{(4-x^2)^3},$$

$D(f'') = D(f') = D(f),$

• $f''(x) = 0 \iff \dfrac{8x(12+x^2)}{(4-x^2)^3} = 0 \iff x = 0,$

$f''(x) > 0 \iff x \in (0,2), \qquad f''(x) < 0 \iff x \in (2,\infty).$

This implies that the function f is strictly convex in the interval $[0,2)$ and strictly concave on the interval $(2,\infty)$.

• 0 is the inflection point – this follows from the strict convexity of f on $[0,2)$ and the strict concavity of f on $(-2,0]$. (The second is a consequence of the symmetry of f.)

d) • Since f has infinite one–sided limits at the points -2 and 2, 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.6.308 and evaluate the 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 $x \to -\infty$.

The graph of f is sketched on Fig. 37.

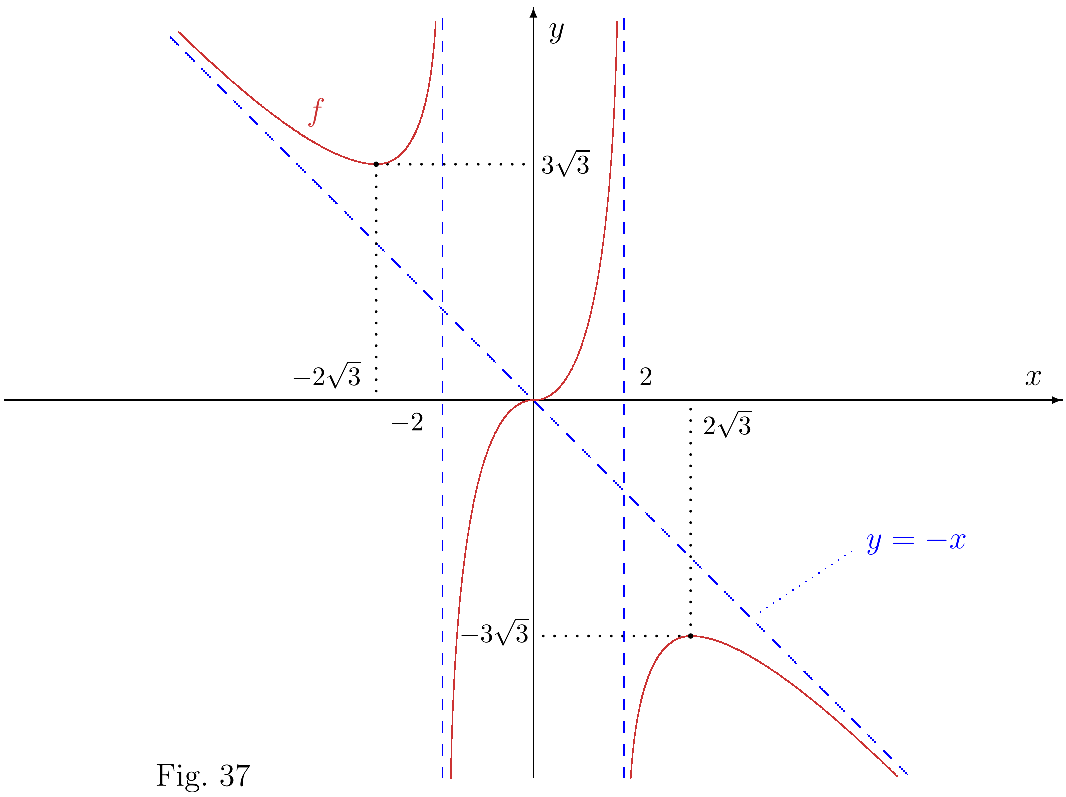


Fig. 37

III.8.5. Problems. Investigate the behavior of the functions

a) $f(x) = x^{2/3} - x$, b) $f(x) = x \cdot \exp(1/x)$, c) $f(x) = x \cdot \exp(-x^2)$,

d) $f(x) = x^2/(x^2 - 4)$, e) $f(x) = 2x^3 + 3x^2 - 12x + 7$.

III.9. Taylor's polynomial, Taylor's theorem

III.9.1. Motivation, derivation of the form of the Taylor polynomial. 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:

$$T_n(x_0)=f(x_0),\quad T_n'(x_0)=f'(x_0),\quad T_n''(x_0)=f''(x_0),\quad \ldots,\quad T_n^{(n)}(x_0)=f^{(n)}(x_0).$$

These are altogether $n+1$ conditions. Substituting here the general form of T_n, we can express by a simple calculation $n+1$ coefficients $a_0, a_1, a_2, \ldots, a_n$:

$$a_0=f(c),\quad a_1=\frac{f'(x_0)}{1!},\quad a_2=\frac{f''(x_0)}{2!},\quad \ldots,\quad a_n=\frac{f^{(n)}(x_0)}{n!}.$$

The polynomial T_n with these coefficients, i.e. the polynomial

$$T_n(x) \;=\; f(x_0) \;+\; \frac{f'(x_0)}{1!}\,(x-x_0) \;+\; \frac{f''(x_0)}{2!}\,(x-x_0)^2 \;+\; \ldots \;+\; \frac{f^{(n)}(x_0)}{n!}\,(x-x_0)^n$$

is called *Taylor's polynomial* of degree n of function f with center x_0. If $x_0=0$ then this polynomial is also often called *Maclaurin's polynomial* of degree n 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.9.2. 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.9.1) $$f(x) \;=\; T_n(x) \;+\; R_{n+1}(x),$$

where

(III.9.2) $$R_{n+1}(x) \;=\; \frac{f^{(n+1)}(\xi)}{(n+1)!}\cdot(x-x_0)^{n+1}.$$

III.9.3. Remark. Formula (III.5.1) is called *Taylor's formula*. It can also be written in the form

$$f(x) \;=\; f(x_0) \;+\; \frac{f'(x_0)}{1!}\,(x-x_0) \;+\; \frac{f''(x_0)}{2!}\,(x-x_0)^2 \;+\; \ldots \;+\; \frac{f^{(n)}(x_0)}{n!}\,(x-x_0)^n \\ +\,R_{n+1}(x).$$

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 *Maclaurin's formula*.

III.9.4. Example (Maclaurin's formula for the exponential function). Maclaurin's formula for the function $f(x)=\mathrm{e}^x$ has the concrete form

$$\mathrm{e}^x \;=\; 1 \;+\; \frac{x}{1!} \;+\; \frac{x^2}{2!} \;+\; \ldots \;+\; \frac{x^n}{n!} \;+\; R_{n+1}(x), \qquad \text{where} \quad R_{n+1}(x)=\frac{\mathrm{e}^{\xi}\,x^{n+1}}{(n+1)!}.$$

III.9.5. Example (Maclaurin's polynomial for the function sine). Calculating the derivatives of the function sine at the point $x_0 = 0$, we find that

a) all derivatives of even orders are equal to zero ,

b) the derivatives of an odd order are equal to 1 or -1 ± 1 and the signs alternate:
$(\sin x)'\big|_{x=0} = \cos 0 = -1, \quad (\sin x)^{(3)}\big|_{x=0} = -\cos 0 = -1,$
$(\sin x)^{(5)}\big|_{x=0} = \cos 0 = 1, \quad$ etc.

Thus, if number n is e.g. even, which means that n has the form $n = 2m$ for appropriate $m \in \mathbb{N}$, then Taylor's polynomial of order n of the function sine at the point 0 is

$$T_n(x) \equiv T_{2m}(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots + (-1)^{m-1}\frac{x^{2m-1}}{(2m-1)!}.$$

The remainder after the n–th term is

$$R_{n+1}(x) \equiv R_{2m+1}(x) = (-1)^m \frac{\cos\xi}{(2m+1)!}\cdot x^{2m+1}.$$

III.9.6. Example (Maclaurin's polynomial for the function cosine). Calculating the derivatives of the function cosine at the point $x_0 = 0$, we find:

a) all derivatives of odd orders are equal to zero,

b) the derivatives of an even order are equal to 1 or -1 and the signs alternate:
$(\cos x)''\big|_{x=0} = -\cos 0 = -1, \quad (\cos x)^{(4)}\big|_{x=0} = \cos 0 = 1,$
$(\cos x)^{(6)}\big|_{x=0} = -\cos 0 = -1, \quad$ etc.

Thus, if number n is e.g. odd, which means that $n = 2m+1$ for an appropriate $m \in \mathbb{N}$, then Taylor's polynomial of order n of the function cosine at the point 0 has the form

$$T_n(x) \equiv T_{2m+1}(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots + (-1)^m\frac{x^{2m}}{(2m)!}.$$

The remainder after the n–th term is

$$R_{n+1}(x) \equiv R_{2m+2}(x) = (-1)^m \frac{\cos\xi}{(2m+2)!}\cdot x^{2m+2}.$$

III.9.7. Example. We calculate the Euler number e with the maximum error $\frac{1}{100}$. Substituting $x = 1$ to Maclaurin's formula for e^x from Example III.9.4, we get:

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{1}{n!} + R_{n+1}(1), \qquad \text{where} \quad R_{n+1}(1) = \frac{e^\xi}{(n+1)!}.$$

Replacing the number e by the sum $1 + \ldots + 1/n!$, the error is equal to $R_{n+1}(1)$. The above expression of $R_{n+1}(1)$ contains the number ξ, which is not precisely determined. The only information following from Taylor's theorem is $\xi \in (0,1)$. Now, the question is how large does n have to be for the remainder to satisfy the estimate $|R_{n+1}(1)| < \frac{1}{100}$. Since $0 < e^\xi < e^1 < 3$, the remainder can be estimated:

$$|\,R_{n+1}(1)\,| \;=\; \frac{e^{\xi}}{(n+1)!} \;<\; \frac{3}{(n+1)!}.$$

A simple calculation shows that if $n = 5$ then $3/(n+1)! = 3/6! = \frac{3}{720} = \frac{1}{240} < \frac{1}{100}$. Hence it is sufficient to put $n = 5$. Then the number e, expressed with an error less than $\frac{1}{100}$, is

$$e \doteq 1 + \frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{1}{5!} = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} = \frac{326}{120}.$$

III.10. Parametrically defined functions

III.10.1. 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.10.1) $$x = t^2, \qquad y = 3t - 1; \qquad t \in (-\infty, \infty).$$

There arise natural questions: How to recognize whether equations (III.10.1) 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.10.2) $$x = \varphi(t), \qquad y = \psi(t); \qquad t \in M$$

instead of the concrete equations (III.10.1). 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.10.2. Parametric definition of a function. Assume that functions φ, ψ are defined in set M and function φ is one–to–one. Then equations (III.10.2) 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.10.2). Equations (III.10.2) are called the *parametric equations* of function f.

III.10.3. 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.10.4. Remark. In order to better understand the notion of a parametrically defined function, consider again the equations (III.10.2). They describe a curve C in plane $\mathbb{E}_2$. This curve 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. It means that the curve C can be considered as a graph of a function y of the variable x. That function which is defined parametrically by equations (III.10.2).

On the contrary, 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.10.1). This function φ is not one–to–one in the interval $(-\infty, \infty)$. The curve C described by equations (III.10.1) 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.10.1) 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.10.1) are not parametric equations 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 $I_1 = (-\infty, 0]$ and $I_2 = [0, \infty)$. Hence the equations (III.10.1) define parametrically a function $y = f_1(x)$ if we consider $t \in I_1$ and they define parametrically another function $y = f_2(x)$ if $t \in I_2$. The graph of f_1 is the lower branch of the parabola $x = \frac{1}{9}(y+1)^2$ (corresponding to $y \leq -1$) and the graph of f_2 is the upper branch of the parabola, corresponding to $y \geq -1$.

III.10.5. 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.19.2), is continuous on its domain $D(f)$.*

III.10.6. Remark. Theorem III.10.5 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.10.7. The derivative of a parametrically defined function. In addition to all assumptions from paragraph III.10.2, 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 $x \in R(\varphi|_{M_1})$:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f'(x) =$$

$$= \; [\,\psi(\varphi_{-1}(x))\,]' \; = \; \dot\psi(\varphi_{-1}(x)) \cdot \varphi'_{-1}(x) \; = \; \dot\psi(\varphi_{-1}(x)) \cdot \frac{1}{\dot\varphi(\varphi_{-1}(x))} \; = \; \frac{\dot\psi(t)}{\dot\varphi(t)}.$$

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

(III.10.3) $$x \; = \; \varphi(t), \qquad \frac{\mathrm{d}y}{\mathrm{d}x} \; = \; \frac{\dot\psi(t)}{\dot\varphi(t)}; \qquad t \in M_1.$$

III.10.8. Remark. In scientific literature, you can often find equations (III.10.3) written down in such a way that the second equation has only y instead of $\mathrm{d}y/\mathrm{d}x$ 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.10.3) can be denoted by y, y' and also by $\mathrm{d}y/\mathrm{d}x$. The last possibility is perhaps less usual, but we regard it as more instructive in a given situation.

III.10.9. Example. Verify whether the equations $x = 2t - \sin t$, $y = 1 + \cos t$ (for $t \in [\,0, 2\pi\,]$) 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 $[\,0, 2\pi\,]$, 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 $[\,0, 2\pi\,]$. Therefore, by Theorem III.10.5, 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 $[\,0, 2\pi\,]$ is the interval $[\,0, 4\pi\,]$. (This is apparent from the fact that φ is continuous and increasing in $[\,0, 2\pi\,]$, its value at the point 0 is 0 and its value at the point 2π is 4π.) Thus, $D(f) = R(\varphi) = [\,0, 4\pi\,]$.

The range of the function $\psi(t) = 1 + \cos t$ on the interval $[\,0, 2\pi\,]$ is the interval $[\,0, 2\,]$. Thus, $R(f) = R(\psi) = [\,0, 2\,]$.

Both functions $\varphi(t) = 2t - \sin t$ and $\psi(t) = 1 + \cos t$ are differentiable in the interval $[\,0, 2\pi\,]$ and $\dot\varphi(t) = 2 - \cos t \neq 0$ for all $t \in [\,0, 2\pi\,]$. Substituting to equations (III.10.3), we get the parametric equations for the derivative of function f:

$$x \; = \; 2t - \sin t, \qquad \frac{\mathrm{d}y}{\mathrm{d}x} \; = \; -\frac{\sin t}{2 - \cos t}; \qquad t \in [\,0, 2\pi\,].$$

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 $[\,0, 2\pi\,]$ and increasing on the interval $[\,2\pi, 4\pi\,]$. 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.10.10. The second derivative of a parametrically defined function. In addition to the assumptions from paragraph III.10.2, 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.10.2) define parametrically a function $y = f(x)$. Its derivative is given parametrically by equations (III.10.3). The second derivative is the first derivative of the first derivative, i.e. to express it, we apply the same procedure as in paragraph III.10.7, this time to equations (III.10.3) . If we denote by ϑ the function on the right hand side of the second equation in (III.10.3) (i.e. $\vartheta(t) = \dot{\psi}(t)/\dot{\varphi}(t)$), we obtain the following parametric representation of the function f'':

$$x = \varphi(t), \qquad \frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = \frac{\dot{\vartheta}(t)}{\dot{\varphi}(t)}; \qquad t \in M_2.$$

Substituting $\dot{\vartheta}(t) = [\,\ddot{\psi}(t) \cdot \dot{\varphi}(t) - \dot{\psi}(t) \cdot \ddot{\varphi}(t)\,] \,/\, [\,\dot{\varphi}(t)\,]^2$ to the second of the above equations, we get:

$$f'' : \qquad x = \varphi(t), \qquad \frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = \frac{\ddot{\psi}(t) \cdot \dot{\varphi}(t) - \dot{\psi}(t) \cdot \ddot{\varphi}(t)}{[\,\dot{\varphi}(t)\,]^3}; \qquad t \in M_2.$$

Other higher order derivatives of a parametrically defined function can be expressed similarly.

III.10.11. 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) $x = \cos^3 t, \quad y = \sin^3 t; \quad t \in (0, \pi/2)$.

Results: a) f is continuous, $D(f) = (-\infty, \infty)$, $R(f) = (2^{-4/3} - 2^{2/3} + 2, \infty)$,

$$f' : \qquad x = t^3 + t + 1, \qquad \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{4t^3 + 2}{3t^2 + 1}; \qquad t \in (-\infty, \infty),$$

$$f'' : \qquad x = t^3 + t + 1, \qquad \frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = \frac{12\,(t^4 + t^2 - t)}{3t^2 + 1}; \qquad t \in (-\infty, \infty).$$

b) f is continuous, $D(f) = [\,0, 1\,]$, $R(f) = [\,0, 1\,]$,

$$f' : \qquad x = \cos^3 t, \qquad \frac{\mathrm{d}y}{\mathrm{d}x} = -\tan t; \qquad t \in (0, \pi/2),$$

$$f'' : \qquad x = \cos^3 t, \qquad \frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = \frac{1}{3 \sin t \cos^4 t}; \qquad t \in (0, \pi/2).$$

III.11. Approximate solution of the nonlinear equation $f(x) = 0$

III.11.1. 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.11.2. 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.11.3. 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.11.4. 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 makes 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 $x_1, x_2, \dots$, 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 the 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 $\gamma_n \leq \epsilon$. Then we can be satisfied with the approximation x_n because the error estimate yields $|x_n - \xi| \leq \gamma_n \leq \epsilon$ 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.11.5. Separation of a root. By the *separation of a root* we understand the specification of an interval $[a, b]$ such that the equation $f(x) = 0$ has a unique root ξ in $[a, b]$. To separate the roots of the equation $f(x) = 0$, we often use theorems III.4.26 and III.6.4.

III.11.6. 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 $f'(x) = 1/x - 2$. 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 $(0,\, 0.5]$ and decreasing in the interval $[0.5, \infty)$. 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 $[a,\, 0.5]$.

a) Existence of a root: We already know that f is continuous in the interval $[a,\, 0.5]$, $f(a) = f(0.0001) = \ln 0.0001 - 0.0002 + 7 = (-4) \cdot \ln 10 - 0.0002 + 7 < (-4) \cdot 2 - 0.0002 + 7 < 0$ and $f(0.5) > 0$. Since 0 is between $f(a)$ and $f(0.5)$, it follows from Theorem III.4.26 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 $[a,\, 0.5]$, so it takes on every its value in this interval only once. This implies the uniqueness of the point $\xi_1 \in [a,\, 0.5]$ 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 $[0.5,\, b]$. ξ_1 and ξ_2 are the only roots of the equation $f(x) = 0$.

III.11.7. The cut–and–try method. Suppose that function f is continuous and strictly monotonic in the interval $[a, b]$ and $f(a) \cdot f(b) < 0$. These assumptions guarantee the existence of a unique root ξ of the equation $f(x) = 0$ in $[a, b]$ and moreover, the iterative sequence whose construction is described in the following converges to ξ.

Choice of the initial approximation: Put $x_0 = (a + b)/2$.

Calculation of further approximations: If $f(x_0) \cdot f(b) < 0$ then $\xi \in (x_1, b]$. Therefore we change a and we put $a = x_0$. If $f(x_0) \cdot f(b) \geq 0$ then $\xi \in [a, x_0]$ and we change b: we put $b = x_0$. Further, we put $x_1 = (a + b)/2$. Similarly, we obtain x_2, x_3, etc. (Illustrate the procedure on an appropriate picture for yourself!)

Error estimate: Denote by L the length of the interval $[a, b]$ at the beginning of the calculation. Since $\xi \in [a,b]$, $|x_0 - \xi| \le L/2$. The length of the "variable" interval $[a,b]$ (where the root ξ is separated) decreases by one half at each step. Hence

$$|x_n - \xi| \;\le\; L/2^{n+1}.$$

III.11.8. Newton's method. Suppose that

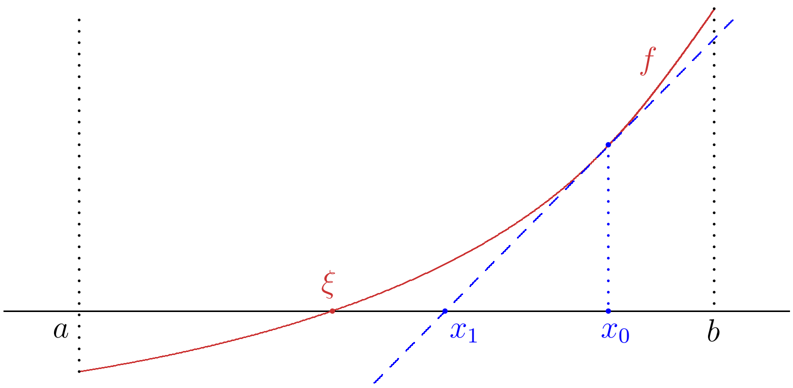
a) function f has a second derivative $f''(x)$ at each point $x \in [a,b]$ and $f''(x)$ does not change its sign in $[a,b]$.

b) $f'(x) \ne 0$ for all $x \in [a,b]$,

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 $[a,b]$ and the iterative sequence, constructed in accordance with rules described in the following, converges to ξ.

Fig. 38



Choice of the initial approximation: The initial approximation x_0 can be chosen to be equal to an arbitrary point of the interval $[a,b]$ such that $f(x_0) \cdot f''(x_0) \ge 0$. (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.5.5.) $y = 0$ 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:

$$x_{n+1} \;=\; x_n \;-\; \frac{f(x_n)}{f'(x_n)}.$$

Error estimate: It follows from the Mean Value Theorem (Theorem III.6.1), 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 $x_n - \xi = f(x_n)/f'(\eta)$ and this further implies that

$$|\, x_n - \xi \,| \ \le \ \frac{|\, f(x_n) \,|}{m},$$

where $m \ = \ \min_{\eta \in [\,a,b\,]} |\, f'(\eta) \,|$.

III.11.9. 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.