2 Functions

2.1 Notion of Functions

2.1.1 Definition of a Function

2.1.1.1 Function

If x and y are two variable quantities, and if there is a rule which assigns a unique value of y to a given value of x, then y is called a function of x, using the notation

$$y = f(x)\,. \tag{2.1}$$

The variable x is called the *independent variable* or the *argument* of the function y. The values of x, to which a value of y is assigned, form the *domain* D of the function $f(x)$. The variable y is called the *dependent variable*; the values of y form the *range* W of the function $f(x)$. Functions can be represented by the points (x, y) as curves, or graphs of the function.

2.1.1.2 Real Functions

If both the domain and the range contain only real numbers the function $y = f(x)$ is called a *real function* of a *real variable*.

■ **A:** $y = x^2$ with $D:\ -\infty < x < \infty,\ W:\ 0 \le y < \infty$.

■ **B:** $y = \sqrt{x}$ with $D:\ 0 \le x < \infty,\quad W:\ 0 \le y < \infty$.

2.1.1.3 Functions of Several Variables

If the variable y depends on several independent variables $x_1, x_2, \ldots, x_n$, then the notation

$$y = f(x_1, x_2, \ldots, x_n) \tag{2.2}$$

is used for a function of several variables (see 2.18, p. 118).

2.1.1.4 Complex Functions

If the dependent and independent variables are *complex numbers* w *and* z respectively, then $w = f(z)$ means a *complex function* of a *complex variable*, (see 14.1, p. 731). *Complex-valued functions* $w(x)$ are called complex functions even if they have real arguments x.

2.1.1.5 Further Functions

In different fields of mathematics, for instance in vector analysis and in vector field theory (see 13.1, p. 701), other types of functions are to be considered whose arguments and values are defined as follows:

1. The arguments are real – the function values are vectors.

■ **A:** Vector functions (see 13.1.1, p. 701).

■ **B:** Parameter representations of curves (see 3.6.2, p. 256).

2. The arguments are vectors – the function values are real numbers.

■ Scalar fields (see 13.1.2, p. 702).

3. The arguments are vectors – the function values are vectors.

■ **A:** Vector fields (see 13.1.3, p. 704). ■ **B:** Parametric representations or vector forms of surfaces (see 3.6.3, p. 261).

2.1.1.6 Functionals

If a real number is assigned to every function $x = x(t)$ of a given class of functions, then it is called a *functional*.

■ **A:** If $x(t)$ is a given function which is integrable on $[a, b]$, then $f(x) = \int_a^b x(t)\,dt$ is a linear functional defined on the set of continuous functions x integrable on $[a, b]$ (see 12.5, p. 677).

I.N. Bronshtein et al., *Handbook of Mathematics*,
DOI 10.1007/978-3-662-46221-8_2

■ **B:** Integral expressions in variational problems (see 10.1, p. 610).

2.1.1.7 Functions and Mappings

Suppose there are given two non-empty sets X and Y. A *mapping*, which is denoted by

$$f\colon X \to Y, \tag{2.3}$$

is a rule, by which a uniquely defined element y of Y is assigned to every element x of X. The element y is called the *image* of x, as formula $y = f(x)$. The set Y is called the *image space* or *range* of f, the set X is called the *original space* or *domain* of f.

■ **A:** If both the original and the image spaces are subsets of real numbers, i.e., $X = D \subset \mathbb{R}$ and $Y = W \subset \mathbb{R}$ hold, then (2.3) defines a real function $y = f(x)$ of the real variable x.

■ **B:** If f is a matrix $\mathbf{A} = (a_{ij})$ $(i = 1, 2, \ldots, m; j = 1, 2, \ldots, n)$ of type (m, n) and $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$, then (2.3) defines a mapping from $\mathbb{R}^n$ into $\mathbb{R}^m$. The rule (2.3) is given by the following system of m linear equations:

$$\underline{\mathbf{y}} = \mathbf{A}\underline{\mathbf{x}} \qquad \text{or} \qquad \begin{array}{l} y_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ y_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_m = a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{array},$$

i.e. $\mathbf{A}\underline{\mathbf{x}}$ means the product of the matrix $\mathbf{A}$ and the vector $\underline{\mathbf{x}}$.

Remarks:

1. The notion of mapping is a generalization of the notion of function. So, some mappings are sometimes called functions.

2. The important properties of mappings can be found in 5.2.3, **5.** p. 333.

3. A mapping, which assigns to every element from an abstract space X a unique element usually from a different abstract space Y, is called an *operator*. Here an *abstract space* usually means a function space, since the most important spaces in applications consist of functions. Abstract spaces are for instance linear spaces (see vector spaces 5.3.8, p. 365), metric spaces (see 12.2, p. 662) and normed spaces (see 12.3, p. 669).

2.1.2 Methods for Defining a Real Function

2.1.2.1 Defining a Function

A function can be defined in several different ways, for instance by a table of values, by graphical representation, i.e., by a curve, by a formula, which is called an *analytic expression*, or piece by piece with different formulas. Only such values of the independent variable can belong to the *domain* of an analytic expression for which the function makes sense, i.e., it takes a unique, finite real value. If the domain is not otherwise defined, the domain is considered as the maximal set for which the definition makes sense.

2.1.2.2 Analytic Representation of a Function

Usually the following three forms are in use:

1. Explicit Form:

$$y = f(x). \tag{2.4}$$

■ $y = \sqrt{1 - x^2}, \quad -1 \le x \le 1, \quad y \ge 0$. Here the graph is the upper half of the unit circle centered at the origin.

2. Implicit Form:

$$F(x, y) = 0, \tag{2.5}$$

in the case when there is a unique y which satisfies this equation, or it can be told which solution is considered to be the value of the function.

■ $x^2 + y^2 - 1 = 0, \quad -1 \le x \le +1, \quad y \ge 0$. Here the graph is again the upper half of the unit circle centered at the origin. It should be emphasized that $x^2 + y^2 + 1 = 0$ itself does not define a real function.

3. Parametric Form:

$$x = \varphi(t), \quad y = \psi(t). \tag{2.6}$$

The corresponding values of x and y are given as functions of an auxiliary variable t, which is called a *parameter*. The functions $\varphi(t)$ and $\psi(t)$ must have the same domain. This representation defines a real function only if $x = \varphi(t)$ defines a one-to-one correspondence between x and t.

■ $x = \varphi(t), \quad y = \psi(t)$ with $\varphi(t) = \cos t$ and $\psi(t) = \sin t, \quad 0 \le t \le \pi$. Here the graph is again the upper half of the unit circle centered at the origin.

Remark: Functions given in parametric form sometimes do not have any explicit or implicit parameter-free equation.

■ $x = t + 2\sin t = \varphi(t)$, $y = t - \cos t = \psi(t)$.

Examples for Functions Given Piece by Piece:

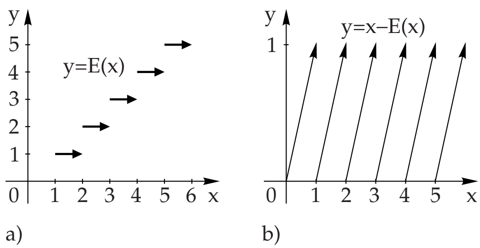

Figure 2.1

■ **A:** $y = E(x) = \operatorname{int}(x) = [x] = n$ for $n \le x < n+1$, n integer.
The function $E(x)$ or $\operatorname{int}(x)$ (read "*integer part of x*") means the greatest integer less than or equal to x.

■ **B:** The function $y = \operatorname{frac}(x) = x - [x]$ (read "*fractional part of x*") gives the difference of x and $[x]$ **(Fig. 2.1b)**. **Fig. 2.1a,b** shows the corresponding graphical representations, where the arrow-heads mean that the endpoints do not belong to the curves.

■ **C:** $y = \begin{cases} x & \text{for } x \le 0, \\ x^2 & \text{for } x \ge 0, \end{cases}$ **(Fig. 2.2a)**.

■ **D:** $y = \operatorname{sign}(x) = \begin{cases} -1 & \text{for } x < 0, \\ 0 & \text{for } x = 0, \\ +1 & \text{for } x > 0, \end{cases}$

(Fig. 2.2b). By $\operatorname{sign}(x)$ (read "signum x"), the *sign function* is denoted.

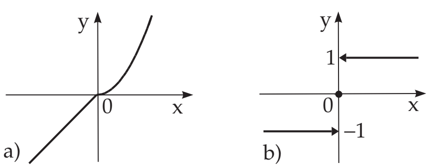

Figure 2.2

2.1.3 Certain Types of Functions

2.1.3.1 Monotone Functions

If a function satisfies the relations

$$f(x_2) \ge f(x_1) \quad \text{or} \quad f(x_2) \le f(x_1), \tag{2.7a}$$

for arbitrary arguments x_1 and x_2 with $x_2 > x_1$ in its domain, then it is called *monotonically increasing* or *monotonically decreasing* **(Fig. 2.3a,b)**.

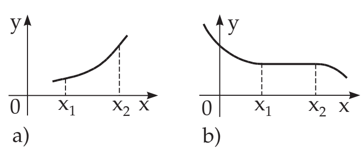

Figure 2.3

If one of the above relations (2.7a) does not hold for every x in the domain of the function, but it is valid, e.g., in an interval or on a half-axis, then the function is called *monotonic in this domain.* Functions satisfying the relations

$$f(x_2) > f(x_1) \quad \text{or} \quad f(x_2) < f(x_1), \tag{2.2}$$

i.e., when the equality never holds in (2.7a), are called *strictly monotonically increasing* or *strictly monotonically decreasing*. In **Fig. 2.3a** there is a representation of a strictly monotonically increasing function; in **Fig. 2.3b** there is the graph of a monotonically decreasing function being constant between x_1 and x_2.

■ $y = e^{-x}$ is strictly monotonically decreasing, $y = \ln x$ is strictly monotonically increasing.

2.1.3.2 Bounded Functions

A function is called *bounded above* if there is a number (called an *upper bound*) such that the values of the function never exceed it. A function is called *bounded below* if there is a number (called a *lower bound*) such that the values of the function are never less than this number. If a function is bounded above and below, it simply is called *bounded*.

■ **A:** $y = 1 - x^2$ is bounded above ($y \le 1$). ■ **B:** $y = e^x$ is bounded below ($y > 0$).

■ **C:** $y = \sin x$ is bounded ($-1 \le y \le +1$). ■ **D:** $y = \dfrac{4}{1+x^2}$ is bounded ($0 < y \le 4$).

2.1.3.3 Extreme Values of Functions

The function $f(x)$ with domain D has an *absolute* or *global maximum* at the point a, if for all $x \in D$

$$f(a) \ge f(x) \tag{2.8a}$$

holds. The function $f(x)$ has a *relative* or *local maximum* at the point a, if the inequality (2.8a) holds only in an environment of the point a, i.e. for all x with $a - \varepsilon < x < a + \varepsilon, \varepsilon > 0, x \in D$.

In analogy the definition of an *absolute* or *global minimum* as well as for a *relative* or *local minimum* can be given, but the inequality (2.8a) is to be replaced by

$$f(a) \le f(x)\,. \tag{2.8b}$$

Remarks:

a) The notions maximum and minimum, are called the *extreme values*, they are not coupled to the differentiability of functions, i.e., they hold also for functions which are not differentiable in some points of the domain. Examples are discontinuities of curves (see **Figs.2.9**, p. 58 and **6.10b,c**, p. 443).

b) Criterions for the determination of extreme values of differentiable functions see in 6.1.5.2, p. 443.

2.1.3.4 Even Functions

Even functions **(Fig. 2.4a)** satisfy the relation

$$f(-x) = f(x). \tag{2.9a}$$

If D is the domain of f, then

$$(x \in D) \Rightarrow (-x \in D) \tag{2.9b}$$

should hold.

■ **A:** $y = \cos x$, ■ **B:** $y = x^4 - 3x^2 + 1$.

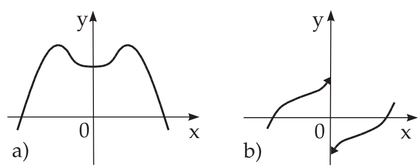

Figure 2.4

2.1.3.5 Odd Functions

Odd functions **(Fig. 2.4b)** satisfy the relation

$$f(-x) = -f(x). \tag{2.10a}$$

If D is the domain of f, then

$$(x \in D) \Rightarrow (-x \in D) \tag{2.10b}$$

should hold.

■ **A:** $y = \sin x$, ■ **B:** $y = x^3 - x$.

2.1.3.6 Representation with Even and Odd Functions

If for the domain D of a function f the condition "from $x \in D$ it follows that $-x \in D$" holds, then f can be written as a sum of an even function g and an odd function u:

$$f(x) = g(x) + u(x) \quad \text{with} \quad g(x) = \frac{1}{2}[f(x) + f(-x)], \quad u(x) = \frac{1}{2}[f(x) - f(-x)]. \tag{2.11}$$

■ $f(x) = e^x = \frac{1}{2}\left(e^x + e^{-x}\right) + \frac{1}{2}\left(e^x - e^{-x}\right) = \cosh x + \sinh x$ (see 2.9.1, p. 89).

2.1.3.7 Periodic Functions

Periodic functions satisfy the relation

$$f(x + T) = f(x), \quad T \text{ const}, \; T \neq 0. \tag{2.12}$$

Obviously, if the above equality holds for some T, it holds for any integer multiple of T. The smallest positive number T satisfying the relation is called the *period* **(Fig. 2.5)**.

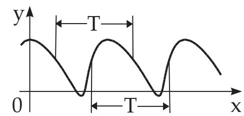

Figure 2.5

2.1.3.8 Inverse Functions

A function $y = f(x)$ with domain D and range W assigns a unique $y \in W$ to every $x \in D$. If reversed, to every $y \in W$ there belongs only one $x \in D$, then the *inverse function* of f can be defined. It is denoted by φ or by f^{-1}. Here f^{-1} is a symbol for a function, not a power of f.
To find the inverse function of f, the variables x and y are interchanged in the formula of f, then y is expressed from $x = f(x)$ in order to get $y = \varphi(x)$. The representations $y = f(x)$ and $x = \varphi(y)$ are equivalent. The following important formulas come from this relation

$$f(\varphi(y)) = y \quad \text{and} \quad \varphi(f(x)) = x. \tag{2.13}$$

The graph of an inverse function $y = \varphi(x)$ is obtained by reflection of the graph of $y = f(x)$ with respect to the line $y = x$ **(Fig. 2.6)**.

■ The function $y = f(x) = e^x$ $(D: -\infty < x < \infty, W: y > 0)$ is equivalent Obviously, every strictly monotonic function has an inverse function.

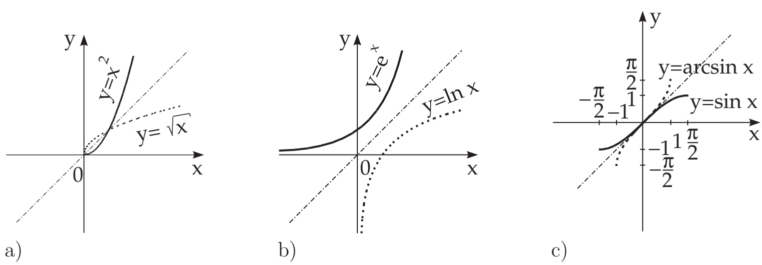

Figure 2.6

Examples of Inverse Functions:

■ **A:**	$y = f(x) = x^2$	with D:	$x \geq 0$,	W: $y \geq 0$;
	$y = \varphi(x) = \sqrt{x}$	with D:	$x \geq 0$,	W: $y \geq 0$.
■ **B:**	$y = f(x) = e^x$	with D:	$-\infty < x < \infty$,	W: $y > 0$;
	$y = \varphi(x) = \ln x$	with D:	$x > 0$,	W: $-\infty < y < \infty$.
■ **C:**	$y = f(x) = \sin x$	with D:	$-\pi/2 \leq x \leq \pi/2$,	W: $-1 \leq y \leq 1$;
	$y = \varphi(x) = \arcsin x$	with D:	$-1 \leq x \leq 1$,	W: $-\pi/2 \leq y \leq \pi/2$.

Remarks:
1. If a function f is strictly monotone in an interval $I \subset D$, then there is an inverse f^{-1} for this interval.
2. If a non-monotone function can be partitioned in strictly monotone parts, then the corresponding inverse exists for each part.

2.1.4 Limits of Functions

2.1.4.1 Definition of the Limit of a Function

The function $y = f(x)$ has the *limit* A at $x = a$

$$\lim_{x \to a} f(x) = A \quad \text{or} \quad f(x) \to A \quad \text{for} \quad x \to a, \tag{2.14}$$

if as x approaches the value a infinitely closely, the value of $f(x)$ approaches the value A infinitely closely. The function $f(x)$ does not have to be defined at a, and even if defined, it does not matter whether $f(a)$ is equal to A.

Precise Definition: The limit (2.14) exists, if for any given positive number ε there is a positive number η such that for every $x \neq a$ belonging to the domain and satisfying the inequality

$$|x - a| < \eta, \tag{2.15a}$$

the inequality

$$|f(x) - A| < \varepsilon \tag{2.15b}$$

holds eventually with the exception of the point a **(Fig. 2.7)**. If a is an endpoint of a connected region, then the inequality $|x - a| < \eta$ is reduced either to $a - \eta < x$ or to $x < a + \eta$ (see also 2.1.4.5).

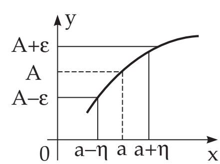

Figure 2.7

2.1.4.2 Definition by Limit of Sequences (see 7.1.2, p. 458)

A function $f(x)$ has the limit A at $x = a$ if for every sequence $x_1, x_2, \ldots, x_n, \ldots$ of the values of x from the domain and converging to a (but being not equal to a), the sequence of the corresponding values of the function $f(x_1), f(x_2), \ldots, f(x_n), \ldots$ converges to A.

2.1.4.3 Cauchy Condition for Convergence

A necessary and sufficient condition for a function $f(x)$ to have a limit at $x = a$ is that for any two values $x_1 \neq a$ and $x_2 \neq a$ belonging to the domain and being close enough to a, the values $f(x_1)$ and $f(x_2)$ are also close enough to each other.

Precise Definition: A necessary and sufficient condition for a function $f(x)$ to have a limit at $x = a$ is that for any given positive number ε there is a positive number η such that for arbitrary values x_1 and x_2 belonging to the domain and satisfying the inequalities

$$0 < |x_1 - a| < \eta \quad \text{and} \quad 0 < |x_2 - a| < \eta, \tag{2.16a}$$

the inequality

$$|f(x_1) - f(x_2)| < \varepsilon \tag{2.16b}$$

holds.

2.1.4.4 Infinity as a Limit of a Function

The symbol

$$\lim_{x \to a} |f(x)| = \infty \tag{2.17}$$

means that as x approaches a, the absolute value $|f(x)|$ does not have an upper bound.

Precise Definition: The equality (2.17) holds if for any given positive number K there is a positive number η such that for any $x \neq a$ from the interval

$$a - \eta < x < a + \eta \tag{2.18a}$$

the corresponding value of $|f(x)|$ is larger than K:

$$|f(x)| > K\,. \tag{2.18b}$$

If all the values of $f(x)$ in the interval

$$a - \eta < x < a + \eta \tag{2.18c}$$

are positive, one writes

$$\lim_{x\to a} f(x) = +\infty; \tag{2.18d}$$

if they are negative, one writes

$$\lim_{x\to a} f(x) = -\infty. \tag{2.18e}$$

2.1.4.5 Left-Hand and Right-Hand Limit of a Function

A function $f(x)$ has a *left-hand limit* A^- at $x = a$, if as x tends to a from the left, the value $f(x)$ tends to A^-:

$$A^- = \lim_{x\to a-0} f(x) = f(a-0). \tag{2.19a}$$

Similarly, a function has a *right-hand limit* A^+ if as x tends to a from the right, the value $f(x)$ tends to A^+:

$$A^+ = \lim_{x\to a+0} f(x) = f(a+0). \tag{2.19b}$$

The equality $\lim\limits_{x\to a} f(x) = A$ is valid only if the left-hand and right-hand limits exist, and they are equal:

$$A^+ = A^- = A. \tag{2.19c}$$

■ The function $f(x) = \dfrac{1}{1+e^{\frac{1}{x-1}}}$ tends to different values from the left and from the right for $x \to 1$: $f(1-0) = 1,\ f(1+0) = 0$ **(Fig. 2.8)**.

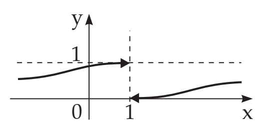

Figure 2.8

2.1.4.6 Limit of a Function as x Tends to Infinity

Case a) A number A is called the limit of a function $f(x)$ as $x \to +\infty$, and one writes

$$A = \lim_{x\to+\infty} f(x) \tag{2.20a}$$

if for any given positive number ε there is a number $N > 0$ such that for every $x > N$, the corresponding value $f(x)$ is in the interval $A - \varepsilon < f(x) < A + \varepsilon$. Analogously

$$A = \lim_{x\to-\infty} f(x) \tag{2.20b}$$

is the limit of a function $f(x)$ as $x \to -\infty$ if for any given positive number ε there is a positive number $N > 0$ such that for any $x < -N$ the corresponding value of $f(x)$ is in the interval $A-\varepsilon < f(x) < A+\varepsilon$.

■ **A:** $\lim\limits_{x\to+\infty} \dfrac{x+1}{x} = 1$, ■ **B:** $\lim\limits_{x\to-\infty} \dfrac{x+1}{x} = 1$, ■ **C:** $\lim\limits_{x\to-\infty} e^x = 0$.

Case b) Assume that for any positive number K, there is a positive number N such that if $x > N$ or $x < -N$ then the absolute value of the function is larger then K. In this case one writes

$$\lim_{x\to+\infty} |f(x)| = \infty \quad \text{or} \quad \lim_{x\to-\infty} |f(x)| = \infty. \tag{2.20c}$$

■ **A:** $\lim\limits_{x\to+\infty} \dfrac{x^3-1}{x^2} = +\infty$, ■ **B:** $\lim\limits_{x\to-\infty} \dfrac{x^3-1}{x^2} = -\infty$,

■ **C:** $\lim\limits_{x\to+\infty} \dfrac{1-x^3}{x^2} = -\infty$, ■ **D:** $\lim\limits_{x\to-\infty} \dfrac{1-x^3}{x^2} = +\infty$.

2.1.4.7 Theorems About Limits of Functions

1. Limit of a Constant Function The limit of a constant function is the constant itself:

$$\lim_{x\to a} A = A. \tag{2.21}$$

2. Limit of a Sum or a Difference If among a finite number of functions each has a limit, then the limit of their sum or difference is equal to the sum or difference of their limits (if this last expression does not contain $\infty - \infty$):

$$\lim_{x\to a} [f(x) + \varphi(x) - \psi(x)] = \lim_{x\to a} f(x) + \lim_{x\to a} \varphi(x) - \lim_{x\to a} \psi(x)\,. \tag{2.22}$$

3. Limit of Products If among a finite number of functions each has a limit, then the limit of their product is equal to the product of their limits (if this last expression does not contain a $0 \cdot \infty$ type):

$$\lim_{x\to a} [f(x)\,\varphi(x)\,\psi(x)] = \left[\lim_{x\to a} f(x)\right] \left[\lim_{x\to a} \varphi(x)\right] \left[\lim_{x\to a} \psi(x)\right]\,. \tag{2.23}$$

4. Limit of a Quotient The limit of the quotient of two functions is equal to the quotient of their limits, in the case when both limits exist and the limit of the denominator is not equal to zero (and this last expression is not an ∞/∞ type):

$$\lim_{x\to a} \frac{f(x)}{\varphi(x)} = \frac{\lim\limits_{x\to a} f(x)}{\lim\limits_{x\to a} \varphi(x)}. \tag{2.24}$$

Also if the denominator is equal to zero, usually one can tell if the limit exists or not, checking the sign of the denominator (the indeterminate form is $0/0$). Similarly, one can calculate the limit of a power by taking a suitable power of the limit (if it is not a 0^0, 1^∞, or ∞^0 type).

5. Pinching If the values of a function $f(x)$ lie between the values of the functions $\varphi(x)$ and $\psi(x)$, i.e., $\varphi(x) < f(x) < \psi(x)$, and if $\lim\limits_{x\to a} \varphi(x) = A$ and $\lim\limits_{x\to a} \psi(x) = A$ hold, then $f(x)$ has a limit, too, and

$$\lim_{x\to a} f(x) = A. \tag{2.25}$$

2.1.4.8 Calculation of Limits

The calculation of the value of a limit can be made by using the 5 described theorems as well as some transformations (see 2.1.4.7).

1. Suitable Transformations

For the calculation of limits the expression is to be transformed into a suitable form. There are several types of recommended transformations in different cases; here are three of them as examples.

■ **A:** $\lim\limits_{x\to 1} \dfrac{x^3-1}{x-1} = \lim\limits_{x\to 1} (x^2+x+1) = 3.$

■ **B:** $\lim\limits_{x\to 0} \dfrac{\sqrt{1+x}-1}{x} = \lim\limits_{x\to 0} \dfrac{(\sqrt{1+x}-1)(\sqrt{1+x}+1)}{x(\sqrt{1+x}+1)} = \lim\limits_{x\to 0} \dfrac{1}{\sqrt{1+x}+1} = \dfrac{1}{2}.$

■ **C:** $\lim\limits_{x\to 0} \dfrac{\sin 2x}{x} = \lim\limits_{x\to 0} \dfrac{2(\sin 2x)}{2x} = 2 \lim\limits_{2x\to 0} \dfrac{\sin 2x}{2x} = 2\,.$ Here one can refer to the well-known theorem

$$\lim_{\alpha \to 0} \frac{\sin \alpha}{\alpha} = 1 \quad \text{(see (■ A, 2.1.4.9, p. 57))}.$$

2. Bernoulli-l'Hospital Rule

In the case of indeterminate forms like $\frac{0}{0}$, $\frac{\infty}{\infty}$, $0 \cdot \infty$, $\infty - \infty$, 0^0, ∞^0, 1^∞, one often applies the *Bernoulli-l'Hospital rule* (usually called *l'Hospital rule* for short):

Case a) Indeterminate Forms $\frac{\mathbf{0}}{\mathbf{0}}$ or $\frac{\infty}{\infty}$: First, use the theorem only after checking if for $f(x) = \frac{\varphi(x)}{\psi(x)}$ the following conditions are fulfilled.
Suppose $\lim\limits_{x \to a} \varphi(x) = 0$ and $\lim\limits_{x \to a} \psi(x) = 0$ or $\lim\limits_{x \to a} \varphi(x) = \infty$ and $\lim\limits_{x \to a} \psi(x) = \infty$, and suppose that there is an interval containing a such that the functions $\varphi(x)$ and $\psi(x)$ are defined and differentiable in this interval except perhaps at a, and $\psi'(x) \neq 0$ in this interval, and $\lim\limits_{x \to a} \frac{\varphi'(x)}{\psi'(x)}$ exists. Then

$$\lim_{x \to a} f(x) = \lim_{x \to a} \frac{\varphi(x)}{\psi(x)} = \lim_{x \to a} \frac{\varphi'(x)}{\psi'(x)}. \tag{2.26}$$

Remark: If the limit of the ratio of the derivatives does not exist, it does not mean that the original limit does not exist. Maybe it does, but one cannot tell this using l'Hospital's rule.
If $\lim\limits_{x \to a} \frac{\varphi'(x)}{\psi'(x)}$ is still an indeterminate form, and the numerator and denominator satisfy the assumptions of the above theorem, l'Hospital's rule can be used again.

■ $$\lim_{x \to 0} \frac{\ln \sin 2x}{\ln \sin x} = \lim_{x \to 0} \frac{\dfrac{2\cos 2x}{\sin 2x}}{\dfrac{\cos x}{\sin x}} = \lim_{x \to 0} \frac{2 \tan x}{\tan 2x} = \lim_{x \to 0} \frac{\dfrac{2}{\cos^2 x}}{\dfrac{2}{\cos^2 2x}} = \lim_{x \to 0} \frac{\cos^2 2x}{\cos^2 x} = 1.$$

Case b) Indeterminate Form $\mathbf{0} \cdot \infty$: Having $f(x) = \varphi(x)\,\psi(x)$ and $\lim\limits_{x \to a} \varphi(x) = 0$ and $\lim\limits_{x \to a} \psi(x) = \infty$, then in order to use l'Hospital's rule for $\lim\limits_{x \to a} f(x)$ it is to be transformed into one of the forms $\lim\limits_{x \to a} \frac{\varphi(x)}{\frac{1}{\psi(x)}}$ or $\lim\limits_{x \to a} \frac{\psi(x)}{\frac{1}{\varphi(x)}}$, so it is reduced to an indeterminate form $\frac{0}{0}$ or $\frac{\infty}{\infty}$ like in case **a)**.

■ $$\lim_{x \to \pi/2} (\pi - 2x) \tan x = \lim_{x \to \pi/2} \frac{\pi - 2x}{\cot x} = \lim_{x \to \pi/2} \frac{-2}{-\dfrac{1}{\sin^2 x}} = 2.$$

Case c) Indeterminate Form $\infty - \infty$: If $f(x) = \varphi(x) - \psi(x)$ and $\lim\limits_{x \to a} \varphi(x) = \infty$ and $\lim\limits_{x \to a} \psi(x) = \infty$ hold, then this expression can be transformed into the form $\frac{0}{0}$ or $\frac{\infty}{\infty}$ usually in several different ways; for instance as $\varphi - \psi = \left(\frac{1}{\psi} - \frac{1}{\varphi}\right) \Big/ \frac{1}{\varphi\psi}$. Then it is to proceed as in case **a)**.

■ $$\lim_{x \to 1} \left(\frac{x}{x-1} - \frac{1}{\ln x} \right) = \lim_{x \to 1} \left(\frac{x \ln x - x + 1}{x \ln x - \ln x} \right) = \frac{0}{0}.$$ Applying l'Hospital rule twice yields

$$\lim_{x \to 1} \left(\frac{x \ln x - x + 1}{x \ln x - \ln x} \right) = \lim_{x \to 1} \left(\frac{\ln x}{\ln x + 1 - \dfrac{1}{x}} \right) = \lim_{x \to 1} \left(\frac{\dfrac{1}{x}}{\dfrac{1}{x} + \dfrac{1}{x^2}} \right) = \frac{1}{2}.$$

Case d) Indeterminate Forms $\mathbf{0^0}$, $\infty^{\mathbf{0}}$, $\mathbf{1}^\infty$: If $f(x) = \varphi(x)^{\psi(x)}$ and $\lim\limits_{x \to a} \varphi(x) = +0$ and $\lim\limits_{x \to a} \psi(x) =$

0 holds, then first the limit A of $\ln f(x) = \psi(x)\ln\varphi(x)$, is to be calculated, which has the form $0\cdot\infty$ (case **b)**), then $\lim\limits_{x\to a} f(x) = e^A$ holds.

The procedures in the cases ∞^0 and 1^∞ are similar.

■ $\lim\limits_{x\to+0} x^x = X,\ \ln x^x = x\ln x,\ \lim\limits_{x\to+0} x\ln x = \lim\limits_{x\to+0}\dfrac{\ln x}{x^{-1}} = \lim\limits_{x\to+0}(-x) = 0$, i.e., $A = \ln X = 0$, so $X = 1$, and finally $\lim\limits_{x\to+0} x^x = 1$.

3. Taylor Expansion

Besides l'Hospital's rule the expansion of functions of indeterminate form into Taylor series can be applied (see 6.1.4.5, p. 442).

■ $$\lim_{x\to 0}\frac{x-\sin x}{x^3} = \lim_{x\to 0}\frac{x-\left(x-\dfrac{x^3}{3!}+\dfrac{x^5}{5!}-\cdots\right)}{x^3} = \lim_{x\to 0}\left(\frac{1}{3!}-\frac{x^2}{5!}+\cdots\right) = \frac{1}{6}.$$

2.1.4.9 Order of Magnitude of Functions and Landau Order Symbols

Comparing two functions, often their mutual behavior with respect to a certain argument $x = a$ is to be considered. It is also convenient to compare the order of magnitude of the functions.

1. A function $f(x)$ tends to infinity at a at a higher order (or faster rate) than a function $g(x)$ at a, if the quotient $\left|\dfrac{f(x)}{g(x)}\right|$ and the absolute values of $f(x)$ exceed any limit as x tends to a.

2. A function $f(x)$ tends to zero at a at a higher order than a function $g(x)$ at a, if the absolute values of $f(x), g(x)$ and the quotient $\dfrac{f(x)}{g(x)}$ tends to zero as x tends to a.

3. Two functions $f(x)$ and $g(x)$ tend to zero or to infinity at s at the same order of magnitude, if $0 < m < \left|\dfrac{f(x)}{g(x)}\right| < M$ holds for the absolute value of their quotient as x tends to a, where M and m are constants.

4. Landau Order Symbols The mutual behavior of two functions at a point $x = a$ can be described by the *Landau order symbols* O ("big O"), or o ("small o") as follows: If $x \to a$ then

$$f(x) = O(g(x)) \quad\text{means that}\quad \lim_{x\to a}\frac{f(x)}{g(x)} = A \neq 0, \quad A = \text{const}, \tag{2.27a}$$

and

$$f(x) = o(g(x)) \quad\text{means that}\quad \lim_{x\to a}\frac{f(x)}{g(x)} = 0, \tag{2.27b}$$

where $a = \pm\infty$ is also possible. The Landau order symbols have meaning only by assuming x tends to a given a.

■ **A:** $\sin x = O(x)$ for $x \to 0$, because with $f(x) = \sin x$ and $g(x) = x$ holds: $\lim\limits_{x\to 0}\dfrac{\sin x}{x} = 1 \neq 0$, i.e., $\sin x$ behaves like x in the neighborhood of $x = 0$.

■ **B:** For $f(x) = 1-\cos x$ and $g(x) = \sin x$ the function $f(x)$ vanishes with a higher order than $g(x)$: $\lim\limits_{x\to 0}\left|\dfrac{f(x)}{g(x)}\right| = \lim\limits_{x\to 0}\left|\dfrac{1-\cos x}{\sin x}\right| = 0$, i.e., $1-\cos x = o(\sin x)$ for $x \to 0$.

■ **C:** $f(x)$ and $g(x)$ vanish by the same order for $f(x) = 1-\cos x,\ g(x) = x^2$:

$$\lim_{x\to 0}\left|\frac{f(x)}{g(x)}\right| = \lim_{x\to 0}\left|\frac{1-\cos x}{x^2}\right| = \frac{1}{2}\,,\ \text{i.e., } 1-\cos x = O(x^2) \text{ for } x \to 0.$$

5. Polynomial The order of magnitude of polynomials at $\pm\infty$ can be expressed by their degree. So the function $f(x) = x$ has order 1, a polynomial of degree $n+1$ has an order higher by one than a polynomial of degree n.

6. Exponential Function The exponential function e^x tends faster to infinity for $x \to \infty$ more quickly to infinity than any high power x^n (n is a fixed positive number):

$$\lim_{x\to\infty}\left|\frac{e^x}{x^n}\right| = \infty. \tag{2.28a}$$

The proof follows by applying l'Hospital's rule for a natural number n:

$$\lim_{x\to\infty}\frac{e^x}{x^n} = \lim_{x\to\infty}\frac{e^x}{nx^{n-1}} = \ldots = \lim_{x\to\infty}\frac{e^x}{n!} = \infty. \tag{2.28b}$$

7. Logarithmic Function The logarithm tends to infinity more slowly than any small positive power x^α (α is a fixed positive number):

$$\lim_{x\to\infty}\left|\frac{\log x}{x^\alpha}\right| = 0. \tag{2.29}$$

The proof is with the help of l'Hospital's rule.

2.1.5 Continuity of a Function

2.1.5.1 Notion of Continuity and Discontinuity

Most functions occurring in practice are continuous, i.e., for small changes of the argument x a *continuous function* $y(x)$ changes also only a little. The graphical representation of such a function results in a continuous curve. If the curve is broken at some points, the corresponding function is *discontinuous*, and the values of the arguments where the breaks are, are the *points of discontinuity*. **Fig. 2.9** shows the curve of a function, which is *piecewise continuous*. The points of discontinuity are A, B, C, D, E, F and G. The arrow-heads show that the endpoints do not belong to the curve.

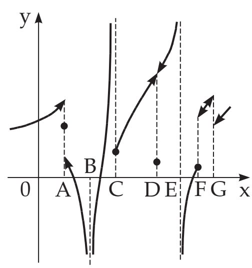

Figure 2.9

2.1.5.2 Definition of Continuity

A function $y = f(x)$ is called *continuous* at the point $x = a$ if

1. $f(x)$ is defined at a;
2. the limit $\lim\limits_{x\to a} f(x)$ exists and is equal to $f(a)$.

This is exactly the case if for an arbitrary $\varepsilon > 0$ there is a $\delta(\varepsilon) > 0$ such that

$$|f(x) - f(a)| < \varepsilon \quad \text{for every } x \text{ with } \quad |x-a| < \delta \tag{2.30}$$

holds.
Here also it is to talk about *one-sided* (*left-* or *right-hand sided*) *continuity*, if instead of $\lim\limits_{x\to a} f(x) = f(a)$ only the one-sided limit $\lim\limits_{x\to a-0} f(x)$ (or $\lim\limits_{x\to a+0} f(x)$) is to be considered and this is equal to the value $f(a)$.

If a function is continuous for every x in a given interval from a to b, then the function is called *continuous in this interval*, which can be open, half-open, or closed (see 1.1.1.3, **3.**, p. 2). If a function is defined and continuous at every point of the numerical axis, it is called *continuous everywhere*.
A function has a *point of discontinuity* at $x = a$, which is an interior point or an endpoint of its domain,

if the function is not defined here, or $f(a)$ is not equal to the limit $\lim\limits_{x\to a} f(x)$, or the limit does not exist. If the function is defined only on one side of $x = a$, e.g., $+\sqrt{x}$ for $x = 0$ and $\arccos x$ for $x = 1$, then it is not a point of discontinuity but it is a termination.
A function $f(x)$ is called *piecewise continuous*, if it is continuous at every point of an interval except at a finite number of points, and at these points it has finite jumps.

2.1.5.3 Most Frequent Types of Discontinuities

1. Values of the Function Tend to Infinity

The most frequent discontinuity occurs if the function tends to $\pm\infty$ (points B, C, and E in **Fig. 2.9**).

■ **A:** $f(x) = \tan x,\ f\left(\frac{\pi}{2} - 0\right) = +\infty,\ f\left(\frac{\pi}{2} + 0\right) = -\infty$. The type of discontinuity (see **Fig. 2.34**, p. 78) is the same as at E in **Fig. 2.9**. For the meaning of the symbols $f(a-0),\ f(a+0)$ see 2.1.4.5, p. 54.

■ **B:** $f(x) = \dfrac{1}{(x-1)^2}\,,\ f(1-0) = +\infty,\ f(1+0) = +\infty$. The type of discontinuity is the same as at the point B in **Fig. 2.9**.

■ **C:** $f(x) = e^{\frac{1}{x-1}}\,,\ f(1-0) = 0,\ f(1+0) = \infty$. The type of discontinuity is the same as at C in **Fig. 2.9**, with the difference that this function $f(x)$ is not defined at $x = 1$.

2. Finite Jump

Passing through $x = a$ the function $f(x)$ jumps from a finite value to another finite value (like at the points A, F, G in **Fig. 2.9**, p. 58): The value of the function $f(x)$ for $x = a$ may not be defined here, as at point G; or it can coincide with $f(a-0)$ or with $f(a+0)$ (point F); or it can be different from $f(a-0)$ and $f(a+0)$ (point A).

■ **A:** $f(x) = \dfrac{1}{1+e^{\frac{1}{x-1}}},\ f(1-0) = 1,\ f(1+0) = 0$ (**Fig. 2.8**, p. 54).

■ **B:** $f(x) = E(x)$ (**Fig. 2.1c**, p. 50) $f(n-0) = n-1,\ f(n+0) = n$ (n integer).

■ **C:** $f(x) = \lim\limits_{n\to\infty} \dfrac{1}{1+x^{2n}}\,,\ f(1-0) = 1,\ f(1+0) = 0,\ f(1) = \dfrac{1}{2}\,.$

3. Removable Discontinuity

Assuming that $\lim\limits_{x\to a} f(x)$ exists, i.e., $f(a-0) = f(a+0)$ holds, but either the function is not defined for $x = a$ or there is $f(a) \neq \lim\limits_{x\to a} f(x)$ (point D in **Fig. 2.9**, p. 58). This type of discontinuity is called *removable*, because defining $f(a) = \lim\limits_{x\to a} f(x)$ the function becomes continuous here. The procedure consists of adding only one point to the curve, or changing the place only of one point at D. The different indeterminate expressions for $x = a$, which have a finite limit examined by l'Hospital's rule or with other methods, are examples of removable discontinuities.

■ $f(x) = \dfrac{\sqrt{1+x}-1}{x}$ is an undetermined $\dfrac{0}{0}$ expression for $x = 0$, but $\lim\limits_{x\to 0} f(x) = \dfrac{1}{2}$; the function

$$f(x) = \begin{cases} \dfrac{\sqrt{1+x}-1}{x} & \text{for } x \neq 0 \\[2ex] \dfrac{1}{2} & \text{for } x = 0 \end{cases}$$

is continuous.

2.1.5.4 Continuity and Discontinuity of Elementary Functions

The elementary functions are continuous on their domains; the points of discontinuity do not belong to their domain. The following theorems hold:

1. Polynomials are continuous everywhere.

2. Rational Functions $\dfrac{P(x)}{Q(x)}$ with polynomials $P(x)$ and $Q(x)$ are continuous everywhere except the points x, where $Q(x) = 0$. If at $x = a$, $Q(a) = 0$ and $P(a) \neq 0$, the function tends to $\pm\infty$ on both sides of a; this point is called a *pole*. The function also has a pole if $P(a) = 0$, but a is a root of the denominator with higher multiplicity than for the numerator (see 1.6.3.1, **2.,** p. 43). Otherwise the discontinuity is removable.

3. Irrational Functions Roots of polynomials are continuous for every x in their domain. At the end of the domain they can terminate by a finite value if the radicand changes its sign. Roots of rational functions are discontinuous for such values of x where the radicand is discontinuous.

4. Trigonometric Functions The functions $\sin x$ and $\cos x$ are continuous everywhere; $\tan x$ and $\sec x$ have infinite jumps at the points $x = \dfrac{(2n+1)\pi}{2}$; the functions $\cot x$ and $\operatorname{cosec} x$ have infinite jumps at the points $x = n\pi$ (n integer).

5. Inverse Trigonometric Functions The functions $\arctan x$ and $\operatorname{arccot} x$ are continuous everywhere, $\arcsin x$ and $\arccos x$ terminate at the end of their domain because of $-1 \leq x \leq +1$, and they are continuous here from one side.

6. Exponential Functions e^x or a^x with $a > 0$ They are continuous everywhere.

7. Logarithmic Function $\log x$ with Arbitrary Positive Base The function is continuous for all positive x and terminates at $x = 0$ because of $\lim\limits_{x\to+0} \log x = -\infty$ by a right-sided limit.

8. Composite Elementary Functions The continuity is to be checked for every point x of every elementary function containing in the composition (see also continuity of composite functions in 2.1.5.5, **2.**, p. 61).

■ Find the points of discontinuity of the function $y = \dfrac{e^{\frac{1}{x-2}}}{x \sin \sqrt[3]{1-x}}$. The exponent $\dfrac{1}{x-2}$ has an infinite jump at $x = 2$; for $x = 2$ also $e^{\frac{1}{x-2}}$ has an infinite jump: $\left(e^{\frac{1}{x-2}}\right)_{x=2-0} = 0$, $\left(e^{\frac{1}{x-2}}\right)_{x=2+0} = \infty$. The function y has a finite denominator at $x = 2$. Consequently, at $x = 2$ there is an infinite jump of the same type as at point C in **Fig. 2.9**, p. 58.
For $x = 0$ the denominator is also zero, just like for the values of x, for which $\sin \sqrt[3]{1-x}$ is equal to zero. These last ones correspond to the roots of the equation $\sqrt[3]{1-x} = n\pi$ or $x = 1 - n^3\pi^3$, where n is an arbitrary integer. The numerator is not equal to zero for these numbers, so at the points $x = 0$, $x = 1$, $x = 1 \pm \pi^3$, $x = 1 \pm 8\pi^3$, $x = 1 \pm 27\pi^3, \ldots$ the function has the same type of discontinuity as the point E in **Fig. 2.9**, p. 58.

2.1.5.5 Properties of Continuous Functions

1. Continuity of Sum, Difference, Product and Quotient of Continuous Functions

If $f(x)$ and $g(x)$ are continuous on the interval $[a, b]$, then $f(x) \pm g(x)$, $f(x)\, g(x)$ are also continuous, and if $g(x) \neq 0$ on this interval, then $\dfrac{f(x)}{g(x)}$ is also continuous.

2. Continuity of Composite Functions $y = f(u(x))$

If $u(x)$ is continuous at $x = a$ and $f(u)$ is continuous at $u = u(a)$ then the composite function $y = f(u(x))$ is continuous at $x = a$, and

$$\lim_{x\to a} f(u(x)) = f\left(\lim_{x\to a} u(x)\right) = f(u(a)) \tag{2.31}$$

is valid. This means that a continuous function of a continuous function is also continuous.

Remark: The converse sentence is not valid. It is possible that the composite function of discontinuous functions is continuous.

3. Bolzano Theorem

If a function $f(x)$ is continuous on a finite closed interval $[a, b]$, and $f(a)$ and $f(b)$ have different signs, then $f(x)$ has at least one root in this interval, i.e., there exists at least one interior point of this interval c such that:

$$f(c) = 0 \quad \text{with} \quad a < c < b. \tag{2.32}$$

The geometric interpretation of this statement is that the graph of a continuous function can go from one side of the x-axis to the other side only if the curve has an intersection point with the x-axis.

4. Intermediate Value Theorem

If a function $f(x)$ is continuous on an interval, and it has different values A and B, at the points a and b of this interval, where $a < b$, i.e.,

$$f(a) = A, \quad f(b) = B, \quad A \neq B, \tag{2.33a}$$

then for any value C between A and B there is at least one point c between a and b such that

$$f(c) = C, \quad (a < c < b, \quad A < C < B \ \text{or} \ A > C > B). \tag{2.33b}$$

In other words: The function $f(x)$ takes every value between A and B on the interval (a, b) at least once. Or: The continuous image of an interval is an interval.

5. Existence of an Inverse Function

If a one-to-one function is continuous on an interval, it is strictly monotone on this interval.
If a function $f(x)$ is continuous on a connected domain I, and it is strictly monotone increasing or decreasing, then for this $f(x)$ there also exists a continuous, strictly monotone increasing or decreasing inverse function $\varphi(x)$ (see also 2.1.3.8, p. 52), which is defined on domain II given by the values of $f(x)$ (**Fig. 2.10**).

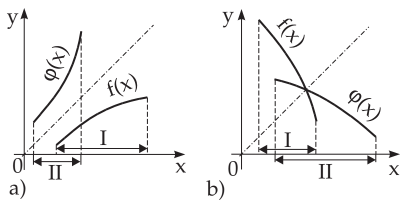

Figure 2.10

Remark: In order to make sure that the inverse function of $f(x)$ is continuous, $f(x)$ must be continuous on an interval. Supposing only that the function is strictly monotonic on an interval, and continuous at an interior point c, and $f(c) = C$, then the inverse function exists, but may be not continuous at C.

6. Theorem About the Boundedness of a Function

If a function $f(x)$ is continuous on a finite, closed interval $[a, b]$ then it is bounded on this interval, i.e., there exist two numbers m and M such that

$$m \le f(x) \le M \quad \text{for} \quad a \le x \le b\,. \tag{2.34}$$

7. Weierstrass Theorem

If the function $f(x)$ is continuous on the finite, closed interval $[a, b]$ then $f(x)$ has an *absolute maximum* M and an *absolute minimum* m, i.e., there exists in this interval at least one point c and at least one point d such that for all x with $a \le x \le b$:

$$m = f(d) \le f(x) \le f(c) = M. \tag{2.35}$$

The difference between the greatest and smallest value of a continuous function is called its *variation* in the given interval. The notion of variation can be extended to the case when the function does not have any greatest or smallest value.

2.2 Elementary Functions

Elementary functions are defined by formulas containing a finite number of operations on the independent variable and constants. The operations are the four basic arithmetical operations, taking powers and roots, the use of an exponential or a logarithm function, or the use of trigonometric functions or inverse trigonometric functions. To distinguish are *algebraic* and *transcendental* elementary functions. As another type of function, can be defined the *non-elementary functions* (see for instance 8.2.5, p. 513).

2.2.1 Algebraic Functions

In an *algebraic function* the argument x and the function y are connected by an *algebraic equation*. It has the form

$$p_0(x) + p_1(x)y + p_2(x)y^2 + \ldots + p_n(x)y^n = 0 \tag{2.36}$$

where p_0, p_1,..., p_n are polynomials in x.

■ $3xy^3 - 4xy + x^3 - 1 = 0$, i.e., $p_0(x) = x^3 - 1$, $p_1(x) = -4x$, $p_2(x) = 0$, $p_3(x) = 3x$.

If it is possible to solve an algebraic equation (2.36) for y, then there is one of the following types of the simplest algebraic functions.

2.2.1.1 Polynomials

Performing only addition, subtraction and multiplication on the argument x then:

$$y = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0. \tag{2.37}$$

holds. In particular one can distinguish $y = a$ as a *constant*, $y = ax + b$ as a *linear function*, and $y = ax^2 + bx + c$ as a *quadratic function*.

2.2.1.2 Rational Functions

A rational function can always be written in the form of the ratio of two polynomials:

$$y = \frac{a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0}{b_m x^m + b_{m-1}x^{m-1} + \ldots + b_0}. \tag{2.38a}$$

The special case

$$y = \frac{ax + b}{cx + d} \tag{2.38b}$$

is called a *homographic or linear fractional function*.

2.2.1.3 Irrational Functions

Besides the operations enumerated for rational functions, the argument x also occurs under the radical sign.

■ **A:** $y = \sqrt{2x + 3}$, ■ **B:** $y = \sqrt[3]{(x^2 - 1)\sqrt{x}}$.

2.2.2 Transcendental Functions

Transcendental functions cannot be given by an algebraic equation like (2.36). In the following paragraphs the simplest elementary transcendental functions are introduced.

2.2.2.1 Exponential Functions

The variable x or an algebraic function of x is in the exponent of a constant base (see 2.6.1, p. 72).

■ **A:** $y = e^x$, ■ **B:** $y = a^x$, ■ **C:** $y = 2^{3x^2-5x}$.

2.2.2.2 Logarithmic Functions

The function is the logarithm with a constant base of the variable x or an algebraic function of x (see 2.6.2, p. 73).

■ **A:** $y = \ln x$, ■ **B:** $y = \lg x$, ■ **C:** $y = \log_2(5x^2 - 3x)$.

2.2.2.3 Trigonometric Functions

The variable x or an algebraic function of x occurs under the symbols sin, cos, tan, cot, sec, cosec (see 2.7, p. 76).

■ **A:** $y = \sin x$, ■ **B:** $y = \cos(2x + 3)$, ■ **C:** $y = \tan\sqrt{x}$.

In general, the argument of a trigonometric function is not only an angle or a circular arc as in the geometric definition, but an arbitrary quantity. The trigonometric functions can be defined in a purely analytic way without any geometry. For instance one can represent them by an expansion in a series, or, e.g., the sin function as the solution of the differential equation $\dfrac{d^2y}{dx^2} + y = 0$ with the initial values $y = 0$ and $\dfrac{dy}{dx} = 1$ at $x = 0$. The numerical value of the argument of the trigonometric function is equal to the *arc* in units of radians. When dealing with trigonometric functions, the argument is considered to be given in *radian measure* (see 3.1.1.5, p. 131).

2.2.2.4 Inverse Trigonometric Functions

The variable x or an algebraic function of x is in the argument of the inverse trigonometric functions (see 2.8, p. 85) arcsin, arccos, etc.

■ **A:** $y = \arcsin x$, ■ **B:** $y = \arccos\sqrt{1-x}$.

2.2.2.5 Hyperbolic Functions

(see 2.9, p. 89).

2.2.2.6 Inverse Hyperbolic Functions

(see 2.10, p. 93).

2.2.3 Composite Functions

Composite functions are all possible compositions of the above algebraic and transcendental functions, i.e., if a function has another function as an argument.

■ **A:** $y = \ln\sin x$, ■ **B:** $y = \dfrac{\ln x + \sqrt{\arcsin x}}{x^2 + 5e^x}$.

Such composition of a finite number of elementary functions again yields an elementary function. The examples **C** in the previous types of functions are also composite functions.

2.3 Polynomials

2.3.1 Linear Function

The graph of the *linear function*

$$y = ax + b \tag{2.39}$$

(polynomial of degree 1) is a *line* **(Fig. 2.11a)**. The *proportional factor* is denoted by a, the crossing point of the line and the axis of the ordinate by b.

For $a > 0$ the function is monotone increasing, for $a < 0$ it is monotone decreasing; for $a = 0$ it is a polynomial of degree zero, i.e., it is a constant function. The intercepts are at $A\left(-\frac{b}{a},0\right)$ and $B(0,b)$ (for details see 3.5.2.6, **1.**, p. 195). With $b=0$ *direct proportionality*

$$y = ax; \tag{2.40}$$

holds, graphically it is a line running through the origin **(Fig. 2.11b)**.

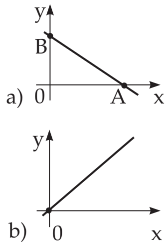

Figure 2.11

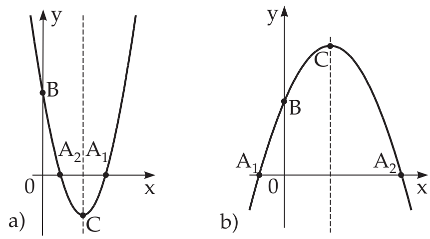

Figure 2.12

2.3.2 Quadratic Polynomial

The *polynomial of second degree*

$$y = ax^2 + bx + c \tag{2.41}$$

(quadratic polynomial) defines a *parabola* with a vertical axis of symmetry at $x = -\frac{b}{2a}$ **(Fig. 2.12)**. For $a > 0$ the function is first monotone decreasing, it has a minimum, then it is monotone increasing. For $a < 0$ first it is monotone increasing, it has a maximum, then it is monotone decreasing. In the case $b^2 - 4ac > 0$: The intersection points A_1, A_2 with the x-axis are $\left(\frac{-b \pm \sqrt{b^2-4ac}}{2a}, 0\right)$, the intersection point B with the y-axis, is at $(0,c)$. In the case $b^2 - 4ac = 0$ there is one intersection point (contact point) with the x-axis. In the case $b^2 - 4ac < 0$ there is no intersection point. The extremum point of the curve is at $C\left(-\frac{b}{2a}, \frac{4ac-b^2}{4a}\right)$ (for more details about the parabola see 3.5.2.10, p. 204).

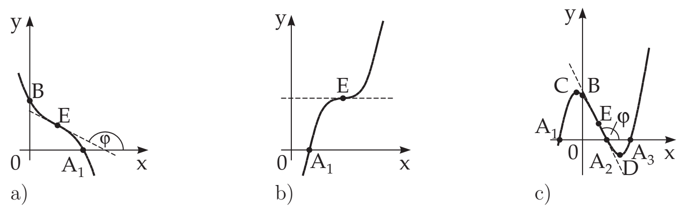

Figure 2.13

2.3.3 Cubic Polynomials

The *polynomial of third degree*

$$y = ax^3 + bx^2 + cx + d \tag{2.42}$$

defines a *cubic parabola* **(Fig. 2.13a,b,c)**. Both the shape of the curve and the behavior of the function depend on a and the discriminant $\Delta = 3ac - b^2$. If $\Delta \geq 0$ holds **(Fig. 2.13a,b)**, then for $a > 0$ the function is monotonically increasing, and for $a < 0$ it is decreasing. If $\Delta < 0$ the function has exactly one local minimum and one local maximum **(Fig. 2.13c)**. For $a > 0$ the value of the function rises from $-\infty$ until the maximum, then falls until the minimum, then it rises again to $+\infty$; for $a < 0$ the value of the function falls from $+\infty$ until the minimum, then rises until the maximum, then it falls again to $-\infty$. The intersection points with the x-axis are at the values of the real roots of (2.42) for $y = 0$. The function can have one, two (then there is a point where the x-axis is the tangent line of the curve) or three real roots: A_1, A_2 and A_3. The intersection point with the y-axis is at $B(0, d)$, the extreme points of the curve C and D, if any, are at $\left(-\frac{b \pm \sqrt{-\Delta}}{3a}, \frac{d + 2b^3 - 9abc \pm (6ac - 2b^2)\sqrt{-\Delta}}{27a^2}\right)$.

The inflection point which is also the center of symmetry of the curve is at $E\left(-\frac{b}{3a}, \frac{2b^3 - 9abc}{27a^2} + d\right)$.

At this point the tangent line has the slope $\tan\varphi = \left(\frac{dy}{dx}\right)_E = \frac{\Delta}{3a}$.

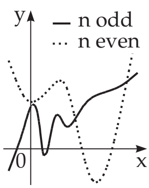

Figure 2.14

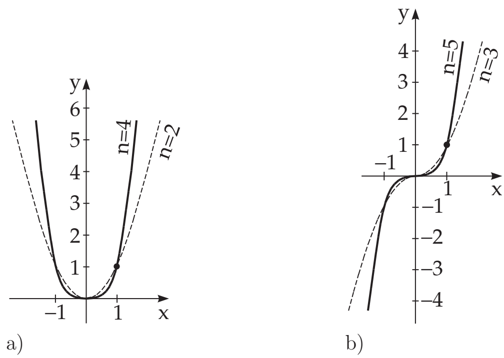

Figure 2.15

2.3.4 Polynomials of n-th Degree

The *integral rational function of n-th degree*

$$y = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0 \qquad (2.43)$$

defines a *curve of n-th degree* or *n-th order* (see 3.5.2.5, p. 195) of *parabolic type* **(Fig. 2.14)**.

Case 1, n odd: For $a_n > 0$ the value of y changes continuously from $-\infty$ to $+\infty$, and for $a_n < 0$ from $+\infty$ to $-\infty$. The curve can intersect or contact the x-axis up to n times, and there is at least one intersection point (for the solution of an equation of n-th degree see 1.6.3.1, p. 43 and 19.1.2, p. 952). The function (2.43) has none or an even number up to $n - 1$ of extreme values, where minima and maxima occur alternately; the number of inflection points is odd and is between 1 and $n - 2$. There are no asymptotes or singularities.

Case 2, n even: For $a_n > 0$ the value of y changes continuously from $+\infty$ through its minimum until $+\infty$ and for $a_n < 0$ from $-\infty$ through its maximum until $-\infty$. The curve can intersect or contact the x-axis up to n times, but it is also possible that it never does that. The number of extrema is odd, and maxima and minima alternate; the number of inflection points is even, and it can also be zero. There are no asymptotes or singularities.

Before sketching the graph of a function, it is recommended first to determine the extreme points, the inflection points, the values of the first derivative at these points, then to sketch the tangent lines at these points, and finally to connect these points continuously.

2.3.5 Parabola of n-th Degree

The graph of the function

$$y = ax^n \tag{2.44}$$

where $n > 0$, integer, is a *parabola of n-th degree, or of n-th order* **(Fig. 2.15)**.

1. Special Case $a = 1$: The curve $y = x^n$ goes through the point $(0, 0)$ and $(1, 1)$ and contacts or intersects the x-axis at the origin. For even n the curve is symmetric with respect to the y-axis, and with a minimum at the origin. For odd n the curve is symmetric with respect to the origin, and it has an inflection point there. There is no asymptote.

2. General Case $a \neq 0$: The curve of $y = ax^n$ can be got from the curve of $y = x^n$ by stretching the ordinates by the factor $|a|$. For $a < 0$ the curve $y = |a|x^n$ is to be reflected with respect to the x-axis.

2.4 Rational Functions

2.4.1 Special Fractional Linear Function (Inverse Proportionality)

The graph of the function

$$y = \frac{a}{x} \tag{2.45}$$

is an *equilateral hyperbola*, whose asymptotes are the coordinate axes **(Fig. 2.16)**. The point of discontinuity is at $x = 0$ with $y = \pm\infty$. If $a > 0$ holds, then the function is strictly monotone decreasing in the interval $(-\infty, 0)$ with values from 0 to $-\infty$ and also strictly monotone decreasing in the interval $(0, +\infty)$ with values from $+\infty$ to 0 (curve in the first and third quadrants). If $a < 0$, then the function is increasing in the interval $(-\infty\,, 0)$ with values from 0 to $+\infty$ and also increasing in the interval $(0, +\infty)$ with values from $-\infty$ to 0 (dotted curve in the second and fourth quadrants). The vertices A and B are at $\left(\pm\sqrt{|a|}, +\sqrt{|a|}\right)$ and $\left(\pm\sqrt{|a|}, -\sqrt{|a|}\right)$ with the same sign for $a > 0$ and with different sign for $a < 0$. There are no extreme values (for more details about hyperbolas see 3.5.2.9, p. 201).

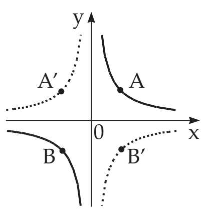

Figure 2.16

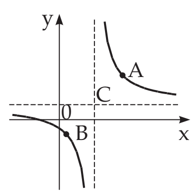

Figure 2.17

2.4.2 Linear Fractional Function

The graph of the function

$$y = \frac{a_1x + b_1}{a_2x + b_2} \quad \left(a_2 \neq 0\,,\ \Delta = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1b_2 - b_1a_2 \neq 0\right) \tag{2.46}$$

is an *equilateral hyperbola*, whose asymptotes are parallel to the coordinate axes **(Fig. 2.17)**. The center is at $C\left(-\frac{b_2}{a_2}, \frac{a_1}{a_2}\right)$. The parameter a in the equality (2.45) corresponds here to $-\frac{\Delta}{{a_2}^2}$ with $\Delta = \begin{vmatrix} a_1 \, b_1 \\ a_2 \, b_2 \end{vmatrix}$. The vertices of the hyperbola $A \neq 0$ and B are at $\left(-\frac{b_2 \pm \sqrt{|\Delta|}}{a_2}, \frac{a_1 + \sqrt{|\Delta|}}{a_2}\right)$ and $\left(-\frac{b_2 \pm \sqrt{|\Delta|}}{a_2}, \frac{a_1 - \sqrt{|\Delta|}}{a_2}\right)$, where for $\Delta < 0$ the same signs are taken, for $\Delta > 0$ the different ones. The point of discontinuity is at $x = -\frac{b_2}{a_2}$. For $\Delta < 0$ the values of the function are decreasing from $\frac{a_1}{a_2}$ to $-\infty$ and from $+\infty$ to $\frac{a_1}{a_2}$. For $\Delta > 0$ the values of the function are increasing from $\frac{a_1}{a_2}$ to $+\infty$ and from $-\infty$ to $\frac{a_1}{a_2}$. There is no extremum.

2.4.3 Curves of Third Degree, Type I

The graph of the function

$$y = a + \frac{b}{x} + \frac{c}{x^2} \quad \left(= \frac{ax^2 + bx + c}{x^2}\right) \quad (b \neq 0, \;\; c \neq 0) \tag{2.47}$$

(Fig. 2.18) is a *curve of third degree* (type I). It has two asymptotes $x = 0$ and $y = a$ and it has two branches. One of them corresponds to the monotone changing of y while it takes its values between a and $+\infty$ or $-\infty$; the other branch goes through three characteristic points: the intersection point with the asymptote $y = a$ at $A\left(-\frac{c}{b}, a\right)$, an extreme point at $B\left(-\frac{2c}{b}, a - \frac{b^2}{4c}\right)$ and an inflection point at $C\left(-\frac{3c}{b}, a - \frac{2b^2}{9c}\right)$. The positions of the branches depend on the signs of b and c, and there are four cases **(Fig. 2.18)**. The intersection points D, E with the x-axis, if any, are for $a \neq 0$ at $\left(\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, 0\right)$, for $a = 0$ at $\left(-\frac{c}{b}, a\right)$; their number can be two, one (the x-axis is a tangent line) or none, depending on whether $b^2 - 4ac > 0, \; = 0$ or < 0 holds.

For $b = 0$ the function (2.47) becomes the function $y = a + \frac{c}{x^2}$ (see **(Fig. 2.21)** the reciprocal power), and for $c = 0$ it becomes the homographic function $y = \frac{ax + b}{x}$, as a special case of (2.46).

2.4.4 Curves of Third Degree, Type II

The graph of the function

$$y = \frac{1}{ax^2 + bx + c} \quad (a \neq 0) \tag{2.48}$$

is a *curve of third degree* (type II) which is symmetric about the vertical line $x = -\frac{b}{2a}$ and the x-axis is its asymptote **(Fig. 2.19)**, because $\lim\limits_{x \to \pm\infty} y = 0$. Its shape depends on the signs of a and $\Delta = 4ac - b^2$. From the two cases $a > 0$ and $a < 0$ only the first one is considered, because reflecting the curve of $y = \frac{1}{(-a)x^2 - bx - c}$ with respect to the x-axis one gets the second one.

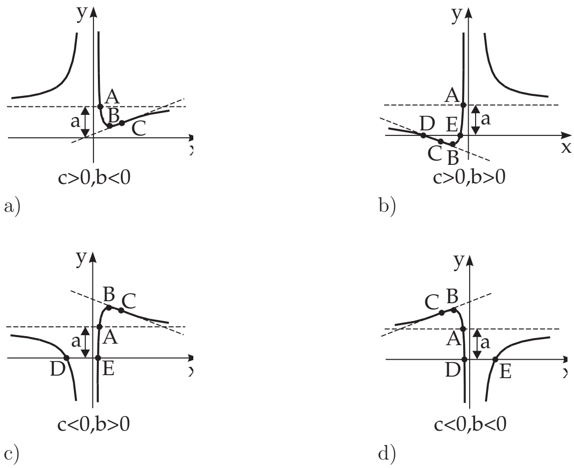

Figure 2.18

Case a) $\boldsymbol{\Delta > 0}$: The function is positive and continuous for arbitrary values of x and it is increasing on the interval $(-\infty\,, -\dfrac{b}{2a})$. Here it takes its maximum, $\dfrac{4a}{\Delta}$, then it is decreasing again in the interval $(-\dfrac{b}{2a}\,, \infty)$. The extreme point A of the curve is at $\left(-\dfrac{b}{2a}\,, \dfrac{4a}{\Delta}\right)$, the inflection points B and C are at $\left(-\dfrac{b}{2a} \pm \dfrac{\sqrt{\Delta}}{2a\sqrt{3}}\,, \dfrac{3a}{\Delta}\right)$; and for the corresponding slopes of the tangent lines (*angular coefficients*) we get $\tan\varphi = \mp a^2 \left(\dfrac{3}{\Delta}\right)^{3/2}$ **(Fig. 2.19a)**.

Case b) $\boldsymbol{\Delta = 0}$: The function is positive for arbitrary values of x, its value rises from 0 to $+\infty$, at $x = -\dfrac{b}{2a} = x_0$ it has a point of discontinuity (a pole), where $\lim\limits_{x \to x_0} y = +\infty$. Then its value falls from here back to 0 **(Fig. 2.19b)**.

Case c) $\boldsymbol{\Delta < 0}$: The value of y rises from 0 to $+\infty$, at the point of discontinuity it jumps to $-\infty$, and rises to the maximum, then falls back to $-\infty$; at the other point of discontinuity it jumps to $+\infty$, then it falls to 0. The extreme point A of the curve is at $\left(-\dfrac{b}{2a}\,, \dfrac{4a}{\Delta}\right)$. The points of discontinuity are at $x = \dfrac{-b \pm \sqrt{-\Delta}}{2a}$ **(Fig. 2.19c)**.

2.4.5 Curves of Third Degree, Type III

The graph of the function

$$y = \frac{x}{ax^2 + bx + c} \quad (a \neq 0, b \neq 0, c \neq 0) \tag{2.49}$$

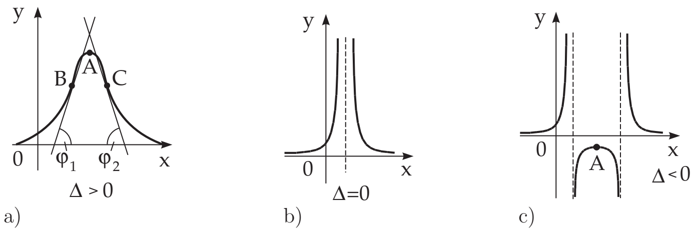

Figure 2.19

is a *curve of third degree* (type III) which goes through the origin, and has the x-axis **(Fig. 2.20)** as an asymptote. The behavior of the function depends on the signs of a and of $\Delta = 4ac - b^2$, and for $\Delta < 0$ also on the signs of the roots α and β of the equation $ax^2 + bx + c = 0$, and for $\Delta = 0$ also on the sign of b. From the two cases, $a > 0$ and $a < 0$, only the first one is considered because reflecting the curve of $y = \dfrac{x}{(-a)x^2 - bx - c}$ with respect to the x-axis yields the second one.

Case a) $\boldsymbol{\Delta > 0}$: The function is continuous everywhere, its value falls from 0 to the minimum, then rises to the maximum, then falls again to 0.
The extreme points of the curve, A and B, are at $\left(\pm\sqrt{\dfrac{c}{a}},\ \dfrac{-b \pm 2\sqrt{ac}}{\Delta}\right)$; there are three inflection points **(Fig. 2.20a)**.

Case b) $\boldsymbol{\Delta = 0}$: The behavior of the function depends on the sign of b, so there are two cases. In both cases there is a point of discontinuity at $x = -\dfrac{b}{2a}$; both curves have one inflection point.

• $b > 0$: The value of the function falls from 0 to $-\infty$, the function has a point of discontinuity, then the value of the function rises from $-\infty$ to the maximum, then decreases to 0 **(Fig. 2.20b$_1$)**. The extreme point A of the curve is at $A\left(+\sqrt{\dfrac{c}{a}}, \dfrac{1}{2\sqrt{ac}+b}\right)$.

• $b < 0$: The value of the function falls from 0 to the minimum, then rises to $+\infty$, running through the origin, then the function has a point of discontinuity, then the value of the function falls from $+\infty$ to 0 **(Fig. 2.20b$_2$)**. The extreme point A of the curve is at $A\left(-\sqrt{\dfrac{c}{a}}, -\dfrac{1}{2\sqrt{ac}-b}\right)$.

Case c) $\boldsymbol{\Delta < 0}$: The function has two points of discontinuity, at $x = \alpha$ and $x = \beta$; its behavior depends on the signs of α and β.

• The signs of α and β are different: The value of the function falls from 0 to $-\infty$, jumps up to $+\infty$, then falls again from $+\infty$ to $-\infty$, running through the origin, then jumps again up to $+\infty$, then it falls tending to 0 **(Fig. 2.20c$_1$)**. The function has no extremum.

• The signs of α and β are both negative: The value of the function falls from 0 to $-\infty$, jumps up to $+\infty$, from here it goes through a minimum up to $+\infty$ again, jumps down to $-\infty$, then rises to a maximum, then falls tending to 0 **(Fig. 2.20c$_2$)**.
The extremum points A and B can be calculated with the same formula as in case a) of 2.4.5.

• The signs of α and β are both positive: The value of the function falls from 0 until the minimum, then rises to $+\infty$, jumps down to $-\infty$, then it rises to the maximum, then it falls again to $-\infty$, then jumps up to $+\infty$ and then it tends to 0 **(Fig. 2.20c$_3$)**.
The extremum points A and B can be calculated by the same formula as in case a) of 2.4.5.

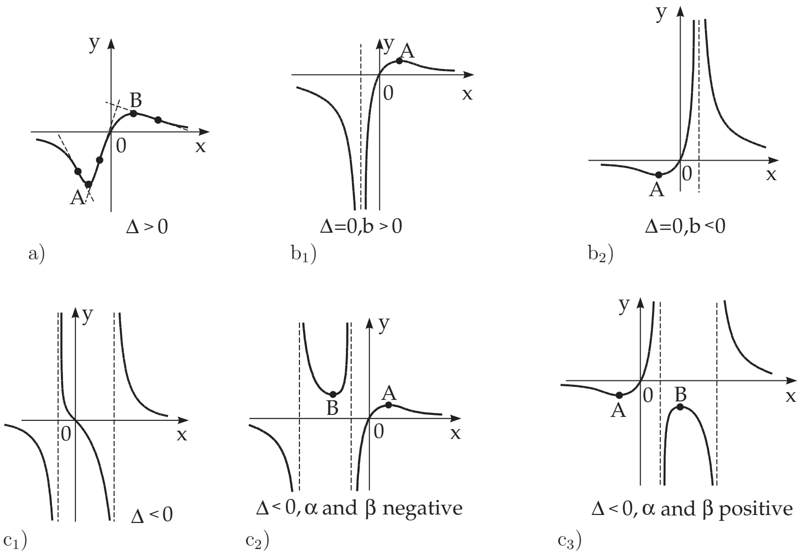

Figure 2.20

In all three cases the curve has one inflection point.

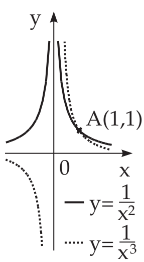

Figure 2.21

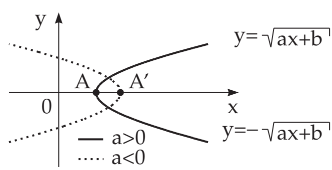

Figure 2.22

2.4.6 Reciprocal Powers

The graph of the function

$$y = \frac{a}{x^n} = ax^{-n} \quad (n > 0, \ \text{integer}; a \neq 0) \tag{2.50}$$

is a *curve of hyperbolic type* with the coordinate axes as asymptotes. The point of discontinuity is at $x = 0$ **(Fig. 2.21)**.

Case a) For $\boldsymbol{a > 0}$ and for even n the value of the function rises from 0 to $+\infty$, then it falls tending to 0, and it is always positive. For odd n it falls from 0 to $-\infty$, it jumps up to $+\infty$, then it falls tending to 0.

Case b) For $\boldsymbol{a < 0}$ and for even n the value of the function falls from 0 to $-\infty$, then it tends to 0, and it is always negative. For odd n it rises from 0 up to $+\infty$, jumps down to $-\infty$, then it tends to 0. The function does not have any extremum. The larger n is, the faster the curve approaches the x-axis, and the slower it approaches the y-axis. For even n the curve is symmetric with respect to the y-axis, for odd n it is center-symmetric and its center of symmetry is the origin. The **Fig. 2.21** shows the cases $n = 2$ and $n = 3$ for $a = 1$.

2.5 Irrational Functions

2.5.1 Square Root of a Linear Binomial

The union of the curve of the two functions

$$y = \pm\sqrt{ax + b} \quad (a \neq 0) \tag{2.51}$$

is a *parabola* with the x-axis as the symmetry axis. The vertex A is at $\left(-\frac{b}{a}, 0\right)$, the *semifocal chord* (see 3.5.2.10, p. 204) is $p = \frac{a}{2}$. The domain of the function and the shape of the curve depend on the sign of a **(Fig. 2.22)** (for more details about the parabola see 3.5.2.10, p. 204).

2.5.2 Square Root of a Quadratic Polynomial

The union of the graphs of the two functions

$$y = \pm\sqrt{ax^2 + bx + c} \quad (a \neq 0,\ \Delta = 4ac - b^2 \neq 0) \tag{2.52}$$

is for $a < 0$ an *ellipse*, for $a > 0$ a *hyperbola* **(Fig. 2.23)**. One of the two symmetry axes is the x-axis, the other one is the line $x = -\frac{b}{2a}$.

The vertices A, C and B, D are at $\left(-\frac{b \pm \sqrt{-\Delta}}{2a}, 0\right)$ and $\left(-\frac{b}{2a}, \pm\sqrt{\frac{\Delta}{4a}}\right)$, where $\Delta = 4ac - b^2$.

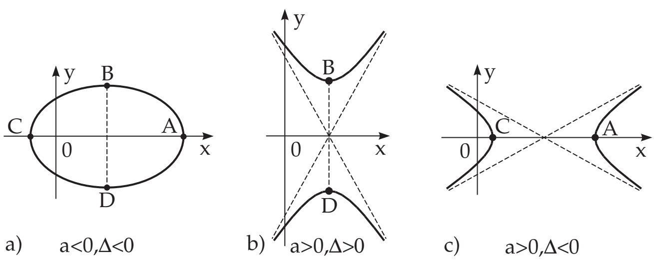

Figure 2.23

The domain of the function and the shape of the curve depend on the signs of a and Δ **(Fig. 2.23)**. For $a < 0$ and $\Delta > 0$ the function has only imaginary values, so no curve exists (for more details about the ellipse and hyperbola see 3.5.2.8, p. 199 and 3.5.2.9, p. 201).

2.5.3 Power Function

The power function

$$y = ax^k = ax^{\pm m/n} \quad (m,\ n \text{ integer, positive, coprime}) \tag{2.53}$$

is to be discussed for $k > 0$ and for $k < 0$ **(Fig. 2.24, Fig. 2.25)**. The investigation here can be restricted to the case $a = 1$, because for $a \neq 1$ the curve differs from the curve of $y = x^k$ only by a stretching in the direction of the y-axis by a factor $|a|$, and for a negative a by a reflection to the x-axis.

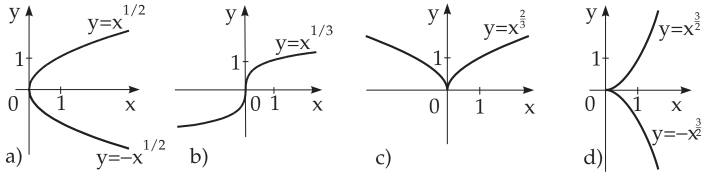

Figure 2.24

Case a) $k > 0$, $y = x^{m/n}$: The shape of the curve is represented in four characteristic cases depending on the numbers m and n in **Fig. 2.24**. The curve goes through the points $(0, 0)$ and $(1, 1)$. For $k > 1$ the x-axis is a tangent line of the curve at the origin **(Fig. 2.24d)**, for $k < 1$ the y-axis is a tangent line also at the origin**(Fig. 2.24a,b,c)**. For even n the union of the graph of functions $y = \pm x^k$ may be considered: it has two branches symmetric to the x-axis **(Fig. 2.24a,d)**, for even m the curve is symmetric to the y-axis **(Fig. 2.24c)**. If m and n are both odd, the curve is symmetric with respect to the origin **(Fig. 2.24b)**. So the curves can have a vertex, a cusp or an inflection point at the origin **(Fig. 2.24)**. None of them has any asymptote.

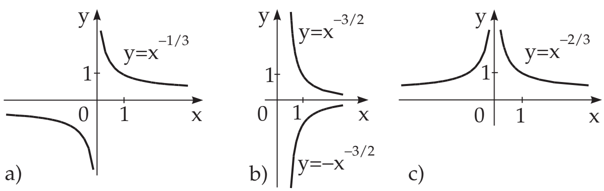

Figure 2.25

Case b) $k < 0$, $y = x^{-m/n}$: The shape of the curve is represented in three characteristic cases depending on m and n in **Fig. 2.25**. The curve is a hyperbolic type curve, where the asymptotes coincide with the coordinate axes **(Fig. 2.25)**. The point of discontinuity is at $x = 0$. The greater $|k|$ is the faster the curve approaches the x-axis, and the slower it approaches the y-axis. The symmetry properties of the curves are the same as above for $k > 0$; they depend on whether m and n are even or odd. There is no extreme value.

2.6 Exponential Functions and Logarithmic Functions

2.6.1 Exponential Functions

The function

$$y = a^x = e^{bx} \quad (a > 0,\ \ b = \ln a)\,, \tag{2.54}$$

is called the *exponential function* and its graphical representation the *exponential curve* **(Fig. 2.26)**. From (2.54) for $a = e$ follows the function of the *natural exponential curve*

$$y = e^x. \tag{2.55}$$

The function has only positive values. Its domain is the interval $(-\infty, +\infty)$. For $a > 1$, i.e., for $b > 0$, the function is strictly monotone increasing and takes its values from 0 until ∞. For $a < 1$, i.e., for $b < 0$, it is strictly monotone decreasing, its value falls from ∞ until 0. The larger $|b|$ is, the greater is the speed of growth and decay. The curve goes through the point $(0,1)$ and approaches asymptotically the x-axis, for $b > 0$ on the right and for $b < 0$ on the left, and faster for greater values of $|b|$. The function $y = a^{-x} = \left(\frac{1}{a}\right)^x$ increases for $a < 1$ and decreases for $a > 1$.

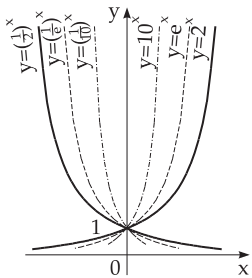

Figure 2.26

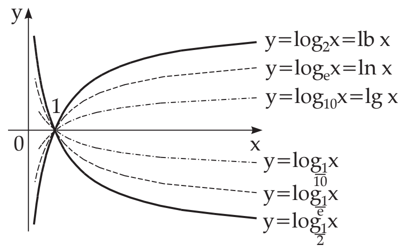

Figure 2.27

2.6.2 Logarithmic Functions

The function

$$y = \log_a x \quad (a > 0, \;\; a \neq 1) \tag{2.56}$$

gives the *logarithmic curve* **(Fig. 2.27)**; the curve is the reflection of the exponential curve with respect to the line $y = x$. From (2.56) for $a = e$ follows the curve of the *natural logarithm*

$$y = \ln x\,. \tag{2.57}$$

The real logarithmic function is defined only for $x > 0$. For $a > 1$ it is strictly monotone increasing and takes its values from $-\infty$ to $+\infty$, for $a < 1$ it is strictly monotone decreasing, and takes its values from $+\infty$ to $-\infty$, and the greater $|\ln a|$ is, the faster the growth and decay. The curve goes through the point $(1,0)$ and approaches asymptotically the y-axis, for $a > 1$ down, for $a < 1$ up, and again faster for larger values of $|\ln a|$.

2.6.3 Error Curve

The function

$$y = e^{-(ax)^2} \tag{2.58}$$

gives the *error curve (Gauss error distribution curve)* **(Fig. 2.28)**. Since the function is even, the y-axis is the symmetry axis of the curve and the larger $|a|$ is, the faster it approaches asymptotically the x-axis. It takes its maximum at zero, and it is equal to one, so the extreme point A of the curve is at $(0,1)$, the inflection points of the curve B, C are at $\left(\pm\frac{1}{a\sqrt{2}}, \frac{1}{\sqrt{e}}\right)$.

The slopes of the tangent lines are here $\tan\varphi = \mp a\sqrt{2/e}\,$.

A very important application of the *error curve* (2.58) is the description of the *normal distribution properties of the observational error* (see 16.2.4.1, p. 818.):

$$y = \varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right). \tag{2.59}$$

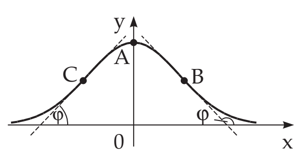

Figure 2.28

2.6.4 Exponential Sum

The function

$$y = ae^{bx} + ce^{dx} \tag{2.60}$$

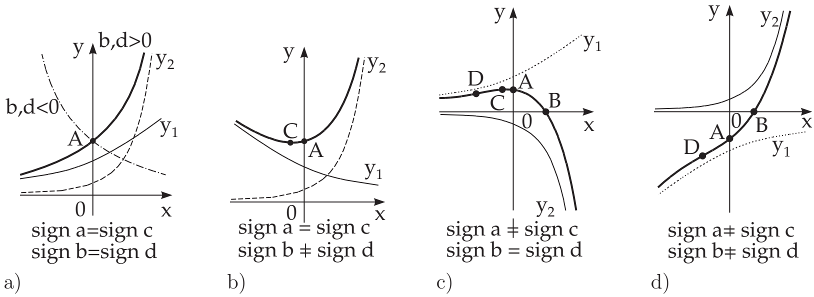

Figure 2.29

is represented in **Fig. 2.29** for the characteristic sign relations. The sum of the functions is got by adding the ordinates of the curves, i.e., the summands are $y_1 = ae^{bx}$ and $y_2 = ce^{dx}$. The function is continuous. If none of the numbers a, b, c, d is equal to 0, the curve has one of the four forms represented in **Fig. 2.29**. Depending on the signs of the parameters it is possible, that the graphs are reflected over a coordinate axis.

The intersection points A and B of the curve with the y-axis and with the x-axis are at $(0, a+c)$, and at $\left(\frac{\ln(-a/c)}{d-b}, 0\right)$ respectively, the extremum C is at $x = \frac{1}{d-b}\ln\left(-\frac{ab}{cd}\right)$, and the inflection point D is at $x = \frac{1}{d-b}\ln\left(-\frac{ab^2}{cd^2}\right)$, in the case when they exist.

Case a) The parameters a and c, and b and d have the same signs: The function does not change its sign, it is strictly monotone; its value is changing from 0 to $+\infty$ or to $-\infty$ or it is changing from $+\infty$ or from $-\infty$ to 0. There is no inflection point. The asymptote is the x-axis **(Fig. 2.29a)**.

Case b) The parameters a and c have the same sign, b and d have different signs: The function does not change its sign and either comes from $+\infty$ and arrives at $+\infty$ and has a minimum or comes from $-\infty$, goes to $-\infty$ and has a maximum. There is no inflection point **(Fig. 2.29b)**.

Case c) The parameters a and c have different signs, b and d have the same signs: The function has one extremum and it is strictly monotone before and after. It changes its sign once. Its value changes whether from 0 until the extremum, then goes to $+\infty$ or $-\infty$ or it comes first from $+\infty$ or $-\infty$, takes the extremum, then approaches 0. The x-axis is an asymptote, the extreme point of the curve is at C and the inflection point at D **(Fig. 2.29c)**.

Case d) The parameters a and c and also b and d have different signs: The function is strictly monotone, its value rises from $-\infty$ to $+\infty$ or it falls from $+\infty$ to $-\infty$. It has an inflection point D **(Fig. 2.29d)**.

2.6.5 Generalized Error Function

The curve of the function

$$y = a\exp(bx + cx^2) = a\exp\left(-\frac{b^2}{4c^2}\right)\exp\left(c\left(x + \frac{b}{2c}\right)^2\right) \quad (c \neq 0) \tag{2.61}$$

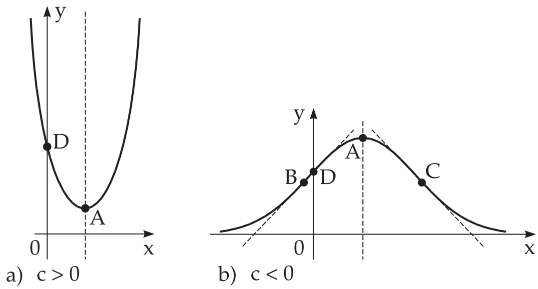

Figure 2.30

can be considered as the generalization of the error function (2.58); it results in a symmetric curve with respect to the vertical line $x = -\frac{b}{2c}$, it has no intersection point with the x-axis, and the intersection point D with the y-axis is at $(0, a)$ **(Fig. 2.30a,b)**.

The shape of the curve depends on the signs of a and c. Here only the case $a > 0$ is discussed, because the curve for $a < 0$ is got by reflecting it in the x-axis.

Case a) $c > 0$: The value of the function falls from $+\infty$ until the minimum, and then rises again to $+\infty$. It is always positive. The extreme point A of the curve is at $\left(-\frac{b}{2c}, a\exp\left(\frac{b^2}{4c}\right)\right)$ and it corresponds to the minimum of the function; there is no inflection point or asymptote **(Fig. 2.30a)**.

Case b) $c < 0$: The x-axis is the asymptote. The extreme point A of the curve is at $\left(-\frac{b}{2c}, a\exp\left(-\frac{b^2}{4c}\right)\right)$ and it corresponds to the maximum of the function. The inflection points B and C are at
$\left(\frac{-b \pm \sqrt{-2c}}{2c}, a\exp\left(\frac{-(b^2+2c)}{4c}\right)\right)$ **(Fig 2.30b)**.

2.6.6 Product of Power and Exponential Functions

The function

$$y = ax^b e^{cx} \tag{2.62}$$

is discussed here only in the case $a > 0$, because in the case $a < 0$ the curve is got by reflecting it in the x-axis. For a non-integer b the function is defined only for $x > 0$, and for an integer b the shape of the curve for negative x can be deduced also from the following cases **(Fig. 2.31)**.

Fig. 2.31 shows how the curve behaves for arbitrary parameters.

For $b > 0$ the curve passes through the origin. The tangent line at this point for $b > 1$ is the x-axis, for $b = 1$ the line $y = x$, for $0 < b < 1$ the y-axis. For $b < 0$ the y-axis is an asymptote. For $c > 0$ the function is increasing and exceeds any value, for $c < 0$ it tends asymptotically to 0. For different signs of b and c the function has an extremum at $x = -\frac{b}{c}$ (point A on the curve). The curve has either no or one or two inflection points at $x = -\frac{b \pm \sqrt{b}}{c}$ (points C and D see **Fig. 2.31c,e,f,g**).

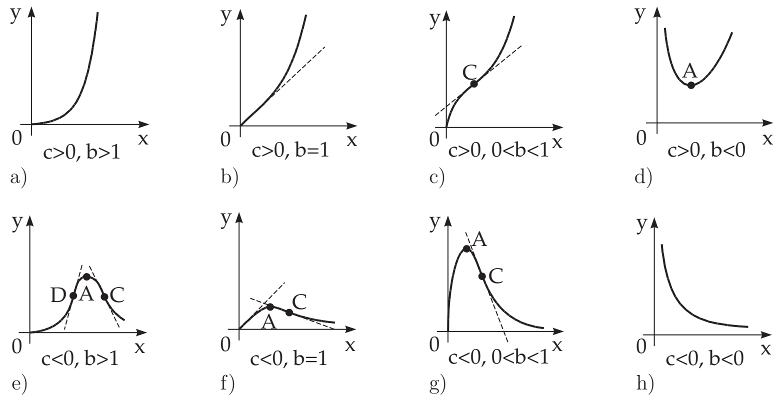

Figure 2.31

2.7 Trigonometric Functions (Functions of Angles)

2.7.1 Basic Notions

2.7.1.1 Definition and Representation

1. Definition

The trigonometric functions are introduced by geometric considerations. So in their definition and also in their arguments degree or radian measure is used (see 3.1.1.5, p. 131).

2. Sine

The *standard sine function*

$$y = \sin x \tag{2.63}$$

is a continuous curve with period $T = 2\pi$ (see **Fig. 2.32a**).

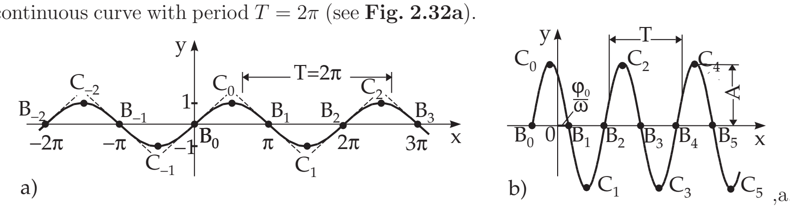

Figure 2.32

The intersection points B_0, B_1, B_{-1}, B_2, B_{-2},..., with $B_k = (k\pi, 0)$ $(k = 0, \pm 1, \pm 2, \ldots)$ of the standard sine curve and the x-axis are the inflection points of the curve. Here the angle of slope of the tangent line with the x axis is $\pm\dfrac{\pi}{4}$. The extreme points of the curve are at C_0, C_1, C_{-1}, C_2, $C_{-2}, \ldots$ with $C_k = \Big((k + \frac{1}{2})\pi, (-1)^k)\ \ (k = 0, \pm 1, \pm 2, \ldots\Big)$. For every value of the function y there is $-1 \le y \le 1$.

The *general sine function*

$$y = A\sin(\omega x + \varphi_0) \tag{2.64}$$

with an amplitude $|A|$, frequency ω, and phase shift φ_0 is represented in **Fig. 2.32b**.
Comparing the standard and the general sine curve **(Fig. 2.32b)** it can be seen that in the general case the curve is stretched in the direction of y by a factor $|A|$, in the direction of x it is compressed by a factor $\frac{1}{\omega}$, and it is shifted to the left by a segment $\frac{\varphi_0}{\omega}$. The period is $T = \frac{2\pi}{\omega}$. The intersection points with the x-axis are $B_k = \left(\frac{k\pi - \varphi_0}{\omega}, 0\right)$ $(k = 0, \pm 1, \pm 2, \ldots)$. The extreme points are $C_k = \left(\frac{\left[\left(k+\frac{1}{2}\right)\pi - \varphi_0\right]}{\omega}, (-1)^k A\right)$ $(k = 0, \pm 1, \pm 2, \ldots)$.

3. Cosine

The *standard cosine function*

$$y = \cos x = \sin(x + \frac{\pi}{2}) \tag{2.65}$$

is represented in **Fig. 2.33**.

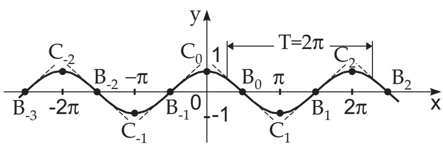

Figure 2.33

The intersection points with the x-axis $B_0, B_1, B_2, \ldots, B_k = \left(\left(k + \frac{1}{2}\right)\pi, 0\right)$ $(k = 0, \pm 1, \pm 2, \ldots)$ are also the inflection points. The angle of slope of the tangent line is $\pm\frac{\pi}{4}$.
The extreme points are $C_0, C_1, \ldots, C_k = (k\pi, (-1)^k)$ $(k = 0, \pm 1, \pm 2, \ldots)$.
The *general cosine function*

$$y = A\cos(\omega x + \varphi_0) \tag{2.66a}$$

can be transformed into the form

$$y = A\sin\left(\omega x + \varphi_0 + \frac{\pi}{2}\right), \tag{2.66b}$$

i.e., the general sine function shifted left by $\varphi = \frac{\pi}{2}$.

4. Tangent

The *tangent function*

$$y = \tan x = \frac{\sin x}{\cos x} \tag{2.67}$$

has period $T = \pi$ and the asymptotes are $x = \left(k + \frac{1}{2}\right)\pi$ $(k = 0, \pm 1, \pm 2, \ldots)$ **(Fig. 2.34)**. The function is monotone increasing in the intervals $\left(-\frac{\pi}{2} + k\pi, +\frac{\pi}{2} + k\pi\right)$ $(k = 0, \pm 1, \pm 2, \ldots)$ and takes values from $-\infty$ to $+\infty$. The curve has intersection points with the x-axis at A_0, A_1, A_{-1}, A_2, A_{-2}, …, $A_k = (k\pi, 0)$ $(k = 0, \pm 1, \pm 2, \ldots)$, these points are the inflection points and the angle of slope of the tangent line is $\frac{\pi}{4}$.

5. Cotangent

The *cotangent function*

$$y = \cot x = \frac{\cos x}{\sin x} = \frac{1}{\tan x} = -\tan\left(x + \frac{\pi}{2}\right) \tag{2.68}$$

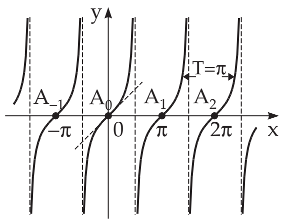

Figure 2.34

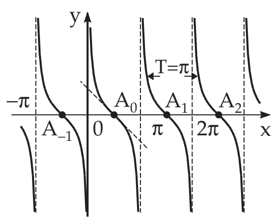

Figure 2.35

has a graph which is the tangent curve reflected with respect to the x-axis and shifted to the left by $\frac{\pi}{2}$ **(Fig. 2.35)**. The asymptotes are $x = k\pi$ $(k = 0, \pm 1, \pm 2, \ldots)$. Between 0 and π the function is monotone decreasing and takes its values from $+\infty$ until $-\infty$; the function has period $T = \pi$. The intersection points with the x-axis are at A_0, A_1, A_{-1}, A_2, $A_{-2}, \ldots$ with $A_k = \left(\left(k + \frac{1}{2}\right)\pi, 0\right)$ $(k = 0, \pm 1, \pm 2, \ldots)$, they are the inflection points of the curve and here the angle of the tangent line is $-\frac{\pi}{4}$.

6. Secant

The *secant function*

$$y = \sec x = \frac{1}{\cos x} \tag{2.69}$$

has period $T = 2\pi$, the asymptotes are $x = \left(k + \frac{1}{2}\right)\pi$ $(k = 0, \pm 1, \pm 2, \ldots)$; and obviously $|y| \geq 1$ holds. The extreme points corresponding to the maxima of the function are A_0, A_1, $A_{-1}, \ldots$

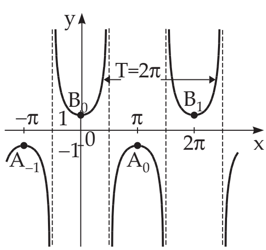

Figure 2.36

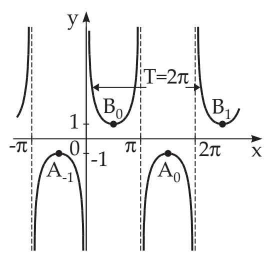

Figure 2.37

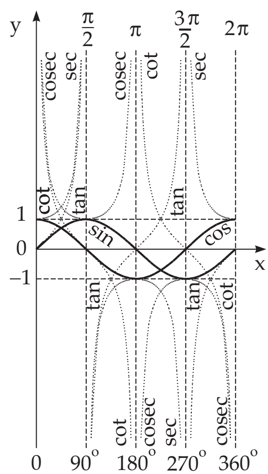

Figure 2.38

with $A_k = ((2k + 1)\pi, -1)$ $(k = 0, \pm 1, \pm 2, \ldots)$, the extreme points corresponding to the minima of the function are B_0, B_1, $B_{-1}, \ldots$ with $B_k = (2k\pi, +1)$ $(k = 0, \pm 1, \pm 2, \ldots)$ **(Fig. 2.36)**.

7. Cosecant

The *cosecant function*

$$y = \operatorname{cosec} x = \frac{1}{\sin x} \tag{2.70}$$

has a graph which is the graph of the secant shifted to the right by $x = \frac{\pi}{2}$. The asymptotes are $x = k\pi$ $(k = 0, \pm 1, \pm 2, \ldots)$. The extreme points corresponding to the maxima of the function are $A_0,\ A_1,\ A_{-1}, \ldots$ with $A_k = \left(\frac{4k+3}{2}\pi, -1\right)$ $(k = 0, \pm 1, \pm 2, \ldots)$ and the the points corresponding to the minima of the function are $B_0,\ B_1,\ B_{-1}, \ldots$ with $B_k = \left(\frac{4k+1}{2}\pi, +1\right)$ $(k = 0, \pm 1, \pm 2, \ldots)$ **(Fig. 2.37)**.

2.7.1.2 Range and Behavior of the Functions

1. Angle Domain $0 \le x \le 360°$

The six trigonometric functions are represented together in **Fig. 2.38** in all the four quadrants for a complete domain of angles from 0° to 360° or for a complete domain of radians from 0 to 2π.
In **Table 2.1** there is a review of the domain and the range of these functions. The signs of the functions depend on the quadrant where the argument is taken from, and these are reviewed in **Table 2.2**.

Table 2.1 Domain and range of trigonometric functions

Domain	Range	Domain	Range
$-\infty < x < \infty$	$\begin{cases} -1 \le \sin x \le 1 \\ -1 \le \cos x \le 1 \end{cases}$	$x \ne (2k+1)\frac{\pi}{2}$ $x \ne k\pi$ $(k = 0, \pm 1, \pm 2, \ldots)$	$-\infty < \tan x < \infty$ $-\infty < \cot x < \infty$

2. Function Values for Some Special Arguments (see Table 2.3)

3. Arbitrary Angle

Since the trigonometric functions are periodic (period 360° or 180°), the determination of their values for an arbitrary argument x can be reduced by the following rules.
Argument $x > 360°$ or $x > 180°$: If the angle is greater than 360° or greater than 180°, then it is to be reduced for a value α, for which $0 \le \alpha \le 360°$ or $0 \le \alpha \le 180°$ holds, in the following way (n integer):

$$\sin(360° \cdot n + \alpha) = \sin\alpha\,, \quad (2.71) \qquad \cos(360° \cdot n + \alpha) = \cos\alpha\,, \quad (2.72)$$

$$\tan(180° \cdot n + \alpha) = \tan\alpha\,, \quad (2.73) \qquad \cot(180° \cdot n + \alpha) = \cot\alpha\,. \quad (2.74)$$

Table 2.2 Signs of trigonometric functions

Quadrant	Angle	sin	cos	tan	cot	sec	csc
I	from 0° to 90°	+	+	+	+	+	+
II	from 90° to 180°	+	−	−	−	−	+
III	from 180° to 270°	−	−	+	+	−	−
IV	from 270° to 360°	−	+	−	−	+	−

Argument $x < 0$: If the argument is negative, then the following formulas reduce the calculations to functions for positive argument:

$$\sin(-\alpha) = -\sin\alpha\,, \quad (2.75) \qquad \cos(-\alpha) = \cos\alpha\,, \quad (2.76)$$

$$\tan(-\alpha) = -\tan\alpha\,, \quad (2.77) \qquad \cot(-\alpha) = -\cot\alpha\,. \quad (2.78)$$

Table 2.3 Values of trigonometric functions for $0°, 30°, 45°, 60°$ and $90°$.

Angle	Radian	sin	cos	tan	cot	sec	csc
$0°$	0	0	1	0	$\mp\infty$	1	$\mp\infty$
$30°$	$\frac{1}{6}\pi$	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{3}$	$\sqrt{3}$	$\frac{2\sqrt{3}}{3}$	2
$45°$	$\frac{1}{4}\pi$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}$	1	1	$\sqrt{2}$	$\sqrt{2}$
$60°$	$\frac{1}{3}\pi$	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$	$\sqrt{3}$	$\frac{\sqrt{3}}{3}$	2	$\frac{2\sqrt{3}}{3}$
$90°$	$\frac{1}{2}\pi$	1	0	$\pm\infty$	0	$\pm\infty$	1

Table 2.4 Reduction formulas and quadrant relations of trigonometric functions

Function	$x = 90° \pm \alpha$	$x = 180° \pm \alpha$	$x = 270° \pm \alpha$	$x = 360° - \alpha$
$\sin x$	$+\cos\alpha$	$\mp\sin\alpha$	$-\cos\alpha$	$-\sin\alpha$
$\cos x$	$\mp\sin\alpha$	$-\cos\alpha$	$\pm\sin\alpha$	$+\cos\alpha$
$\tan x$	$\mp\cot\alpha$	$\pm\tan\alpha$	$\mp\cot\alpha$	$-\tan\alpha$
$\cot x$	$\mp\tan\alpha$	$\pm\cot\alpha$	$\mp\tan\alpha$	$-\cot\alpha$

Argument x for $90° < x < 360°$: If $90° < x < 360°$ holds, then the arguments are to be reduced for an acute angle α by the *reduction formulas* given in **Table 2.4**. The relations between the values of the functions belonging to the arguments which differ from each other by $90°, 180°$ or $270°$ or which complete each other to $90°, 180°$ or $270°$ are called *quadrant relations.*
The first and second columns of **Table 2.4** give the *complementary angle formulas*, and the first and third ones give the *supplementary angle formulas.* Because $x = 90° - \alpha$ is the complementary angle (see 3.1.1.2, p. 130) of α, the following relations

$$\cos\alpha = \sin x = \sin(90° - \alpha)\,, \quad (2.79a) \qquad \sin\alpha = \cos x = \cos(90° - \alpha) \quad (2.79b)$$

are called the *complementary angle formulas.*

For $\alpha + x = 180°$ the relations between the trigonometric functions for supplementary angles (see 3.1.1.2, p. 130)

$$\sin\alpha = \sin x = \sin(180° - \alpha)\,, \quad (2.80a) \qquad -\cos\alpha = \cos x = \cos(180° - \alpha) \quad (2.80b)$$

are called *supplementary angle formulas.*
Argument x for $0° < x < 90°$: The values of trigonometric functions for acute angles ($0° < x < 90°$) have been taken formerly from tables, today calculators are used.

■ $\sin(-1000°) = -\sin 1000° = -\sin(360° \cdot 2 + 280°) = -\sin 280° = +\cos 10° = +0.9848.$

4. Angles in Radian Measure

The arguments given in radian measure, i.e., in units of radians, can be easily converted by formula (3.2) (see 3.1.1.5, p. 131).

2.7.2 Important Formulas for Trigonometric Functions

Remark: Trigonometric functions with complex argument z are discussed in 14.5.2, p. 759.

2.7.2.1 Relations Between the Trigonometric Functions

$$\sin^2\alpha + \cos^2\alpha = 1\,, \qquad (2.81)$$

$$\sec^2\alpha - \tan^2\alpha = 1, \qquad (2.82)$$

$$\operatorname{cosec}^2\alpha - \cot^2\alpha = 1\,, \qquad (2.83)$$

$$\sin\alpha \cdot \operatorname{cosec}\alpha = 1\,, \qquad (2.84)$$

$$\cos\alpha \cdot \sec\alpha = 1\,, \qquad (2.85)$$

$$\tan\alpha \cdot \cot\alpha = 1\,, \qquad (2.86)$$

$$\frac{\sin\alpha}{\cos\alpha} = \tan\alpha\,, \qquad (2.87)$$

$$\frac{\cos\alpha}{\sin\alpha} = \cot\alpha\,. \qquad (2.88)$$

Some important relations are summarized in **Table 2.5** for $0 < \alpha < \pi/2$ in order to create an easy survey. For other intervals in **Table 2.5** the square roots are always considered with the sign which corresponds to the quadrant where the argument is.

2.7.2.2 Trigonometric Functions of the Sum and Difference of Two Angles (Addition Theorems)

$$\sin(\alpha\pm\beta) = \sin\alpha\cos\beta \pm \cos\alpha\sin\beta, \qquad (2.89)$$

$$\cos(\alpha\pm\beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta, \qquad (2.90)$$

$$\tan(\alpha\pm\beta) = \frac{\tan\alpha \pm \tan\beta}{1 \mp \tan\alpha\tan\beta}, \qquad (2.91)$$

$$\cot(\alpha\pm\beta) = \frac{\cot\alpha\cot\beta \mp 1}{\cot\beta \pm \cot\alpha}, \qquad (2.92)$$

$$\begin{aligned}\sin(\alpha+\beta+\gamma) &= \sin\alpha\cos\beta\cos\gamma + \cos\alpha\sin\beta\cos\gamma \\ &\quad + \cos\alpha\cos\beta\sin\gamma - \sin\alpha\sin\beta\sin\gamma,\end{aligned} \qquad (2.93)$$

$$\begin{aligned}\cos(\alpha+\beta+\gamma) &= \cos\alpha\cos\beta\cos\gamma - \sin\alpha\sin\beta\cos\gamma \\ &\quad - \sin\alpha\cos\beta\sin\gamma - \cos\alpha\sin\beta\sin\gamma.\end{aligned} \qquad (2.94)$$

2.7.2.3 Trigonometric Functions of an Integer Multiple of an Angle

$$\sin 2\alpha = 2\sin\alpha\cos\alpha, \qquad (2.95)$$

$$\sin 3\alpha = 3\sin\alpha - 4\sin^3\alpha, \qquad (2.96)$$

$$\cos 2\alpha = \cos^2\alpha - \sin^2\alpha, \qquad (2.97)$$

$$\cos 3\alpha = 4\cos^3\alpha - 3\cos\alpha, \qquad (2.98)$$

$$\sin 4\alpha = 8\cos^3\alpha\sin\alpha - 4\cos\alpha\sin\alpha, \qquad (2.99)$$

$$\cos 4\alpha = 8\cos^4\alpha - 8\cos^2\alpha + 1, \qquad (2.100)$$

$$\tan 2\alpha = \frac{2\tan\alpha}{1-\tan^2\alpha}, \qquad (2.101)$$

$$\tan 3\alpha = \frac{3\tan\alpha - \tan^3\alpha}{1 - 3\tan^2\alpha}, \qquad (2.102)$$

$$\tan 4\alpha = \frac{4\tan\alpha - 4\tan^3\alpha}{1 - 6\tan^2\alpha + \tan^4\alpha,} \qquad (2.103)$$

$$\cot 2\alpha = \frac{\cot^2\alpha - 1}{2\cot\alpha}, \qquad (2.104)$$

$$\cot 3\alpha = \frac{\cot^3\alpha - 3\cot\alpha}{3\cot^2\alpha - 1}, \qquad (2.105)$$

$$\cot 4\alpha = \frac{\cot^4\alpha - 6\cot^2\alpha + 1}{4\cot^3\alpha - 4\cot\alpha}. \qquad (2.106)$$

For larger values of n in order to gain a formula for $\sin n\alpha$ and $\cos n\alpha$ the de Moivre formula is to be used (see 1.5.3.5, p. 38).

Using the binomial theorem (see 1.1.6.4, p. 12) gives:

$$\cos n\alpha + \mathrm{i}\sin n\alpha = \sum_{k=0}^{n} \binom{n}{k} \mathrm{i}^k \cos^{n-k}\alpha \sin^k \alpha = (\cos\alpha + \mathrm{i}\sin\alpha)^n$$

$$= \cos^n\alpha + \mathrm{i}n\cos^{n-1}\alpha\sin\alpha$$

$$-\binom{n}{2}\cos^{n-2}\alpha\sin^2\alpha - \mathrm{i}\binom{n}{3}\cos^{n-3}\alpha\sin^3\alpha + \binom{n}{4}\cos^{n-4}\alpha\sin^4\alpha + \dots . \tag{2.107}$$

With this it follows:

$$\cos n\alpha = \cos^n\alpha - \binom{n}{2}\cos^{n-2}\alpha\sin^2\alpha + \binom{n}{4}\cos^{n-4}\alpha\sin^4\alpha - \binom{n}{6}\cos^{n-6}\alpha\sin^6\alpha + \dots , \tag{2.108}$$

$$\sin n\alpha = n\cos^{n-1}\alpha\sin\alpha - \binom{n}{3}\cos^{n-3}\alpha\sin^3\alpha + \binom{n}{5}\cos^{n-5}\alpha\sin^5\alpha - \dots . \tag{2.109}$$

Table 2.5 Relations between the trigonometric functions of the same argument in the interval $0 < \alpha < \frac{\pi}{2}$

α	$\sin\alpha$	$\cos\alpha$	$\tan\alpha$	$\cot\alpha$
$\sin\alpha$	–	$\sqrt{1-\cos^2\alpha}$	$\frac{\tan\alpha}{\sqrt{1+\tan^2\alpha}}$	$\frac{1}{\sqrt{1+\cot^2\alpha}}$
$\cos\alpha$	$\sqrt{1-\sin^2\alpha}$	–	$\frac{1}{\sqrt{1+\tan^2\alpha}}$	$\frac{\cot\alpha}{\sqrt{1+\cot^2\alpha}}$
$\tan\alpha$	$\frac{\sin\alpha}{\sqrt{1-\sin^2\alpha}}$	$\frac{\sqrt{1-\cos^2\alpha}}{\cos\alpha}$	–	$\frac{1}{\cot\alpha}$
$\cot\alpha$	$\frac{\sqrt{1-\sin^2\alpha}}{\sin\alpha}$	$\frac{\cos\alpha}{\sqrt{1-\cos^2\alpha}}$	$\frac{1}{\tan\alpha}$	–

2.7.2.4 Trigonometric Functions of Half-Angles

In the following formulas the sign of the square root must be chosen positive or negative, according to the quadrant where the half-angle is.

$$\sin\frac{\alpha}{2} = \sqrt{\frac{1}{2}(1-\cos\alpha)}, \tag{2.110}$$

$$\cos\frac{\alpha}{2} = \sqrt{\frac{1}{2}(1+\cos\alpha)}, \tag{2.111}$$

$$\tan\frac{\alpha}{2} = \sqrt{\frac{1-\cos\alpha}{1+\cos\alpha}} = \frac{1-\cos\alpha}{\sin\alpha} = \frac{\sin\alpha}{1+\cos\alpha}, \tag{2.112}$$

$$\cot\frac{\alpha}{2} = \sqrt{\frac{1+\cos\alpha}{1-\cos\alpha}} = \frac{1+\cos\alpha}{\sin\alpha} = \frac{\sin\alpha}{1-\cos\alpha}. \tag{2.113}$$

2.7.2.5 Sum and Difference of Two Trigonometric Functions

$$\sin\alpha + \sin\beta = 2\sin\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}, \quad (2.114)$$

$$\sin\alpha - \sin\beta = 2\cos\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}, \quad (2.115)$$

$$\cos\alpha + \cos\beta = 2\cos\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}, \quad (2.116)$$

$$\cos\alpha - \cos\beta = -2\sin\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}, \quad (2.117)$$

$$\tan\alpha \pm \tan\beta = \frac{\sin(\alpha\pm\beta)}{\cos\alpha\cos\beta}, \quad (2.118)$$

$$\cot\alpha \pm \cot\beta = \pm\frac{\sin(\alpha\pm\beta)}{\sin\alpha\sin\beta}, \quad (2.119)$$

$$\tan\alpha + \cot\beta = \frac{\cos(\alpha-\beta)}{\cos\alpha\sin\beta}, \quad (2.120)$$

$$\cot\alpha - \tan\beta = \frac{\cos(\alpha+\beta)}{\sin\alpha\cos\beta}. \quad (2.121)$$

2.7.2.6 Products of Trigonometric Functions

$$\sin\alpha\sin\beta = \frac{1}{2}[\cos(\alpha-\beta) - \cos(\alpha+\beta)], \quad (2.122)$$

$$\cos\alpha\cos\beta = \frac{1}{2}[\cos(\alpha-\beta) + \cos(\alpha+\beta)], \quad (2.123)$$

$$\sin\alpha\cos\beta = \frac{1}{2}[\sin(\alpha-\beta) + \sin(\alpha+\beta)], \quad (2.124)$$

$$\begin{aligned}\sin\alpha\sin\beta\sin\gamma = \frac{1}{4}[&\sin(\alpha+\beta-\gamma) + \sin(\beta+\gamma-\alpha)\\ &+\sin(\gamma+\alpha-\beta) - \sin(\alpha+\beta+\gamma)],\end{aligned} \quad (2.125)$$

$$\begin{aligned}\sin\alpha\cos\beta\cos\gamma = \frac{1}{4}[&\sin(\alpha+\beta-\gamma) - \sin(\beta+\gamma-\alpha)\\ &+\sin(\gamma+\alpha-\beta) + \sin(\alpha+\beta+\gamma)],\end{aligned} \quad (2.126)$$

$$\begin{aligned}\sin\alpha\sin\beta\cos\gamma = \frac{1}{4}[&-\cos(\alpha+\beta-\gamma) + \cos(\beta+\gamma-\alpha)\\ &+\cos(\gamma+\alpha-\beta) - \cos(\alpha+\beta+\gamma)],\end{aligned} \quad (2.127)$$

$$\begin{aligned}\cos\alpha\cos\beta\cos\gamma = \frac{1}{4}[&\cos(\alpha+\beta-\gamma) + \cos(\beta+\gamma-\alpha)\\ &+\cos(\gamma+\alpha-\beta) + \cos(\alpha+\beta+\gamma)].\end{aligned} \quad (2.128)$$

2.7.2.7 Powers of Trigonometric Functions

$$\sin^2\alpha = \frac{1}{2}(1-\cos 2\alpha), \quad (2.129)$$

$$\cos^2\alpha = \frac{1}{2}(1+\cos 2\alpha), \quad (2.130)$$

$$\sin^3\alpha = \frac{1}{4}(3\sin\alpha - \sin 3\alpha), \quad (2.131)$$

$$\cos^3\alpha = \frac{1}{4}(\cos 3\alpha + 3\cos\alpha), \quad (2.132)$$

$$\sin^4\alpha = \frac{1}{8}(\cos 4\alpha - 4\cos 2\alpha + 3), \quad (2.133)$$

$$\cos^4\alpha = \frac{1}{8}(\cos 4\alpha + 4\cos 2\alpha + 3). \quad (2.134)$$

For large values of n $\sin^n \alpha$ and $\cos^n \alpha$, can be expressd by applying the formulas for $\cos n\alpha$ and $\sin n\alpha$ (see 2.7.2.3, p. 82).

2.7.3 Description of Oscillations

2.7.3.1 Formulation of the Problem

In engineering and physics one often meets quantities depending on time and given in the form

$$u(t) = A\sin(\omega t + \varphi). \tag{2.135}$$

They are called also *sinusoidal quantities*. Their dependence on time results in a *harmonic oscillation*. The graphical representation of (2.135) results in a *general sine curve*, as shown in **Fig. 2.39**.

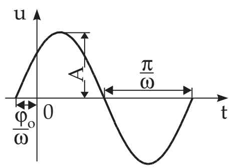

Figure 2.39

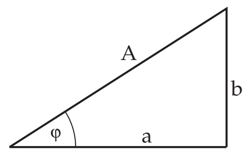

Figure 2.40

The general sine curve differs from the simple sine curve $y = \sin x$:

a) by the *amplitude* A, i.e., the greatest distance between its points and the time axis t,

b) by the *period* $T = \dfrac{2\pi}{\omega}$, which corresponds to the *wavelength* (with ω as the *frequency of the oscillation*, which is called the *angular* or *radial frequency* in wave theory),

c) by the *initial phase* or *phase shift* by the initial angle $\varphi \neq 0$.

The quantity $u(t)$ can also be written in the form

$$u(t) = a\sin\omega t + b\cos\omega t. \tag{2.136}$$

Here for a and b $A = \sqrt{a^2 + b^2}$ and $\tan\varphi = \dfrac{b}{a}$ holds. The quantities a, b, A and φ can be represented as sides and angle of a right triangle **(Fig. 2.40)**.

2.7.3.2 Superposition of Oscillations

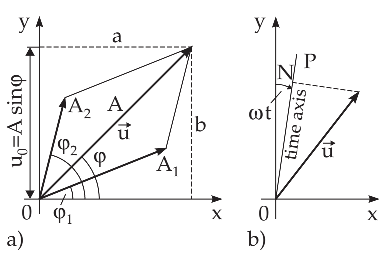

Figure 2.41

In the simplest case the *superposition of oscillations* is the *addition of two oscillations* with the same frequency. It results again in a harmonic oscillation with the same frequency:

$$A_1\sin(\omega t+\varphi_1)+A_2\sin(\omega t+\varphi_2) = A\sin(\omega t+\varphi) \tag{2.137a}$$

with

$$A = \sqrt{{A_1}^2 + {A_2}^2 + 2A_1A_2\cos(\varphi_2 - \varphi_1)}, \tag{2.137b}$$

$$\tan\varphi = \frac{A_1\sin\varphi_1 + A_2\sin\varphi_2}{A_1\cos\varphi_1 + A_2\cos\varphi_2}, \tag{2.137c}$$

where the quantities A and φ can be determined by a vector diagram **(Fig. 2.41a)**.

A *linear combination* of several sine functions with the same frequency is also possible and yields a general sine function (harmonic oscillation) with the same frequency:

$$\sum_i c_iA_i\sin(\omega t + \varphi_i) = A\sin(\omega t + \varphi). \tag{2.138}$$

2.7.3.3 Vector Diagram for Oscillations

The general sine function (2.135, 2.136) can be represented easily by the polar coordinates $\rho = A, \varphi$ and by the Cartesian coordinates $x = a$, $y = b$ (see 3.5.2.1, p. 190) in a plane. The sum of two such quantities then behaves as the sum of two summand vectors **(Fig. 2.41a)**. Similarly the sum of several vectors results in a *linear combination* of several general sine functions. This representation is called a *vector diagram*.

The quantity u for a given time t can be determined from the vector diagram with the help of **Fig. 2.41b**: First the time axis $OP(t)$ has to be put through the origin O, which rotates clockwise around O by a constant angular velocity ω. At start $t = 0$ the axes y and t coincide. Then at any time t the projection ON of the vector $\vec{\mathbf{u}}$ onto the time axis is equal to the absolute value of the general sine function $u = A\sin(\omega t + \varphi)$. For time $t = 0$ the value $u_0 = A\sin\varphi$ is the projection onto the y-axis **(Fig. 2.41b)**.

2.7.3.4 Damping of Oscillations

The function $u(t) = Ae^{-at}\sin(\omega t + \varphi_0) \quad (a,\, t > 0)$ (2.139)

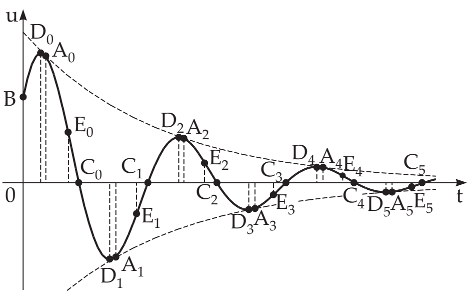

Figure 2.42

yields the *curve of a damped oscillation* **(Fig. 2.42)**.

The oscillation proceeds along the t-axis, while the curve asymptotically approaches the t-axis. The sine curve is enclosed by the exponential curves $u(t) = \pm Ae^{-at}$, and it contacts them in the points

$$A_0, A_1, A_2, \ldots, A_k = \left(\frac{\left(k + \frac{1}{2}\right)\pi - \varphi_0}{\omega}, \right.$$

$$\left. (-1)^k A \exp\left(-a\frac{\left(k + \frac{1}{2}\right)\pi - \varphi_0}{\omega} \right) \right).$$

The intersection points with the coordinate axes are $B = (0, A\sin\varphi_0), C_0, C_1, C_2, \ldots, C_k = \left(\frac{k\pi - \varphi_0}{\omega}, 0\right)$. The extrema $D_0, D_1, D_2, \ldots$ are at $t_k = \frac{k\pi - \varphi_0 + \alpha}{\omega}$; and the inflection points $E_0, E_1, E_2, \ldots$ are at $t_k = \frac{k\pi - \varphi_0 + 2\alpha}{\omega}$ with $\tan\alpha = \frac{\omega}{a}$.

The *logarithmic decrement* of the damping is $\delta = \ln\left|\frac{y_i}{y_{i+1}}\right| = a\frac{\pi}{\omega}$, where y_i and y_{i+1} are the ordinates of two consecutive extrema.

2.8 Cyclometric or Inverse Trigonometric Functions

The *cyclometric functions* or *arcus functions* are the inverses of the trigonometric functions. For a unequivocal definition the domain of the trigonometric functions is to be decomposed into monotony intervals, to get an inverse function for every monotony interval. So, there are infinitely many such intervals, and for each its inverse is to be defined. In order to distinguish them an index k is to be assigned according to the corresponding interval. Obviously the trigonometric inverse functions are monotony in these intervals.

2.8.1 Definition of the Inverse Trigonometric Functions

How to define the inverse trigonometric functions will be shown here for the inverse of the sin function **(Fig. 2.43)**. The usual notation for it is $\arcsin x$. The domain of $y = \sin x$ will be split into monotony

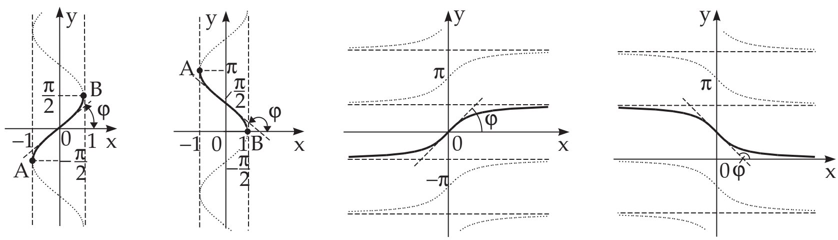

Figure 2.43 Figure 2.44 Figure 2.45 Figure 2.46

intervals $k\pi - \frac{\pi}{2} \le x \le k\pi + \frac{\pi}{2}$ with $k = 0, \pm 1, \pm 2, \ldots$. Reflecting the curve of $y = \sin x$ in the line $y = x$ yields the curve of the inverse function

$$y = \operatorname{arc}_k \sin x \qquad (2.140a)$$

with the domains and ranges

$$-1 \le x \le +1 \quad \text{and} \quad k\pi - \frac{\pi}{2} \le y \le k\pi + \frac{\pi}{2}, \quad \text{where} \quad k = 0, \pm 1, \pm 2, \ldots . \qquad (2.140b)$$

The form $y = \operatorname{arc}_k \sin x$ has the same meaning as $x = \sin y$.

Similarly, one can get the other inverse trigonometric functions which are represented in **Fig. 2.44–2.46**. The domains and ranges of the inverse functions can be found in **Table 2.6**.

2.8.2 Reduction to the Principal Value

In their domain the arcus functions have the so-called *principal values* for $k = 0$, written usually without an index, g.e., as $\arcsin x \equiv \operatorname{arc}_0 \sin x$. In **Fig. 2.47** the principal values of the inverse functions are presented.

The values of the different inverses can be calculated from the principal values by the following formulas:

$$\operatorname{arc}_k \sin x = k\pi + (-1)^k \arcsin x. \qquad (2.141)$$

$$\operatorname{arc}_k \cos x = \begin{cases} (k+1)\pi - \arccos x & (k \text{ odd}), \\ k\pi + \arccos x & (k \text{ even}). \end{cases} \qquad (2.142)$$

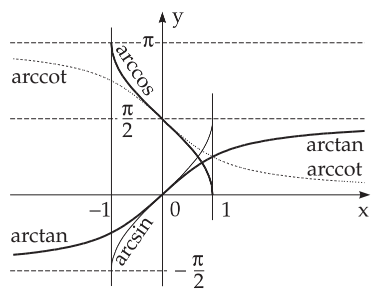

Figure 2.47

$$\operatorname{arc}_k \tan x = k\pi + \arctan x. \qquad (2.143)$$

$$\operatorname{arc}_k \cot x = k\pi + \operatorname{arccot} x. \qquad (2.144)$$

■ **A:** $\arcsin 0 = 0\,,\ \operatorname{arc}_k \sin 0 = k\pi$.

■ **B:** $\operatorname{arccot} 1 = \frac{\pi}{4}\,,\ \operatorname{arc}_k \cot 1 = \frac{\pi}{4} + k\pi$.

■ **C:** $\arccos \frac{1}{2} = \frac{\pi}{3},\ \operatorname{arc}_k \cos \frac{1}{2} = -\frac{\pi}{3} + (k+1)\pi$ for odd k,
$= \frac{\pi}{3} + k\pi$ for even k.

Remark: Calculators give the principal values of the trigonometric inverses.

Table 2.6 Domains and ranges of the inverses of trigonometric functions

Inverse function	Domain	Range	Trigonometric function with the same meaning
arc sine $y = \mathrm{arc}_k \sin x$	$-1 \le x \le 1$	$k\pi - \dfrac{\pi}{2} \le y \le k\pi + \dfrac{\pi}{2}$	$x = \sin y$
arc cosine $y = \mathrm{arc}_k \cos x$	$-1 \le x \le 1$	$k\pi \le y \le (k+1)\pi$	$x = \cos y$
arc tangent $y = \mathrm{arc}_k \tan x$	$-\infty < x < \infty$	$k\pi - \dfrac{\pi}{2} < y < k\pi + \dfrac{\pi}{2}$	$x = \tan y$
arc cotangent $y = \mathrm{arc}_k \cot x$	$-\infty < x < \infty$	$k\pi < y < (k+1)\pi$	$x = \cot y$
$k = 0, \pm 1, \pm 2, \ldots$. For $k = 0$ one gets the principal value of the inverse functions, which is usually written without an index, e.g., $\arcsin x \equiv \mathrm{arc}_0 \sin x$.			

2.8.3 Relations Between the Principal Values

$$\arcsin x = \frac{\pi}{2} - \arccos x = \arctan \frac{x}{\sqrt{1-x^2}} = \begin{cases} -\arccos\sqrt{1-x^2} & (-1 \le x \le 0), \\ \arccos\sqrt{1-x^2} & (0 \le x \le 1). \end{cases} \tag{2.145}$$

$$\arccos x = \frac{\pi}{2} - \arcsin x = \operatorname{arccot} \frac{x}{\sqrt{1-x^2}} = \begin{cases} \pi - \arcsin\sqrt{1-x^2} & (\pi - 1 \le x \le 0), \\ \arcsin\sqrt{1-x^2} & (0 \le x \le 1). \end{cases} \tag{2.146}$$

$$\arctan x = \frac{\pi}{2} - \operatorname{arccot} x = \arcsin \frac{x}{\sqrt{1+x^2}}\,. \tag{2.147}$$

$$\arctan x = \begin{cases} \operatorname{arccot}\dfrac{1}{x} - \pi & (x < 0) \\ \operatorname{arccot}\dfrac{1}{x} & (x > 0) \end{cases} = \begin{cases} -\arccos\dfrac{1}{\sqrt{1+x^2}} & (x \le 0)\,, \\ \arccos\dfrac{1}{\sqrt{1+x^2}} & (x \ge 0). \end{cases} \tag{2.148}$$

$$\operatorname{arccot} x = \frac{\pi}{2} - \arctan x = \arccos \frac{x}{\sqrt{1+x^2}}\,. \tag{2.149}$$

$$\operatorname{arccot} x = \begin{cases} \arctan\dfrac{1}{x} + \pi & (x < 0) \\ \arctan\dfrac{1}{x} & (x > 0) \end{cases} = \begin{cases} \pi - \arcsin\dfrac{1}{\sqrt{1+x^2}} & (x \le 0), \\ \arcsin\dfrac{1}{\sqrt{1+x^2}} & (x \ge 0). \end{cases} \tag{2.150}$$

2.8.4 Formulas for Negative Arguments

$$\arcsin(-x) = -\arcsin x. \tag{2.151}$$

$$\arctan(-x) = -\arctan x. \tag{2.152}$$

$$\arccos(-x) = \pi - \arccos x. \tag{2.153}$$

$$\operatorname{arccot}(-x) = \pi - \operatorname{arccot} x. \tag{2.154}$$

2.8.5 Sum and Difference of $\arcsin x$ and $\arcsin y$

$$\arcsin x + \arcsin y = \arcsin\left(x\sqrt{1-y^2} + y\sqrt{1-x^2}\right) \quad (xy \le 0 \text{ or } x^2+y^2 \le 1), \tag{2.155a}$$

$$= \pi - \arcsin\left(x\sqrt{1-y^2} + y\sqrt{1-x^2}\right) \quad (x>0, y>0,\ x^2+y^2>1), \tag{2.155b}$$

$$= -\pi - \arcsin\left(x\sqrt{1-y^2} + y\sqrt{1-x^2}\right) \quad (x<0, y<0,\ x^2+y^2>1). \tag{2.155c}$$

$$\arcsin x - \arcsin y = \arcsin\left(x\sqrt{1-y^2} - y\sqrt{1-x^2}\right) \quad (xy \ge 0 \text{ or } x^2+y^2 \le 1), \tag{2.156a}$$

$$= \pi - \arcsin\left(x\sqrt{1-y^2} - y\sqrt{1-x^2}\right) \quad (x>0,\ y<0,\ x^2+y^2>1), \tag{2.156b}$$

$$= -\pi - \arcsin\left(x\sqrt{1-y^2} - y\sqrt{1-x^2}\right) \quad (x<0,\ y>0,\ x^2+y^2>1). \tag{2.156c}$$

2.8.6 Sum and Difference of $\arccos x$ and $\arccos y$

$$\arccos x + \arccos y = \arccos\left(xy - \sqrt{1-x^2}\sqrt{1-y^2}\right) \quad (x+y \ge 0), \tag{2.157a}$$

$$= 2\pi - \arccos\left(xy - \sqrt{1-x^2}\sqrt{1-y^2}\right) \quad (x+y<0). \tag{2.157b}$$

$$\arccos x - \arccos y = -\arccos\left(xy + \sqrt{1-x^2}\sqrt{1-y^2}\right) \quad (x \ge y), \tag{2.158a}$$

$$= \arccos\left(xy + \sqrt{1-x^2}\sqrt{1-y^2}\right) \quad (x<y). \tag{2.158b}$$

2.8.7 Sum and Difference of $\arctan x$ and $\arctan y$

$$\arctan x + \arctan y = \arctan\frac{x+y}{1-xy} \quad (xy<1), \tag{2.159a}$$

$$= \pi + \arctan\frac{x+y}{1-xy} \quad (x>0,\ xy>1), \tag{2.159b}$$

$$= -\pi + \arctan\frac{x+y}{1-xy} \quad (x<0,\ xy>1). \tag{2.159c}$$

$$\arctan x - \arctan y = \arctan\frac{x-y}{1+xy} \quad (xy>-1), \tag{2.160a}$$

$$= \pi + \arctan\frac{x-y}{1+xy} \quad (x>0,\ xy<-1), \tag{2.160b}$$

$$= -\pi + \arctan\frac{x-y}{1+xy} \quad (x<0,\ xy<-1). \tag{2.160c}$$

2.8.8 Special Relations for $\arcsin x$, $\arccos x$, $\arctan x$

$$2\arcsin x = \arcsin\left(2x\sqrt{1-x^2}\right) \quad \left(|x| \le \frac{1}{\sqrt{2}}\right), \tag{2.161a}$$

$$= \pi - \arcsin\left(2x\sqrt{1-x^2}\right) \quad \left(\frac{1}{\sqrt{2}} < x \le 1\right), \tag{2.161b}$$

$$= -\pi - \arcsin\left(2x\sqrt{1-x^2}\right) \quad \left(-1 \le x < -\frac{1}{\sqrt{2}}\right). \tag{2.161c}$$

$$2\arccos x = \arccos(2x^2-1) \quad (0 \le x \le 1), \tag{2.162a}$$

$$= 2\pi - \arccos(2x^2-1) \quad (-1 \le x < 0). \tag{2.162b}$$

$$2\arctan x = \arctan\frac{2x}{1-x^2} \quad (|x| < 1), \tag{2.163a}$$

$$= \pi + \arctan\frac{2x}{1-x^2} \quad (x > 1), \tag{2.163b}$$

$$= -\pi + \arctan\frac{2x}{1-x^2} \quad (x < -1). \tag{2.163c}$$

$$\cos(n\arccos x) = T_n(x) \quad (n \ge 1), \tag{2.164}$$

where $n \ge 1$ can also be a fractional number and $T_n(x)$ is given by the equation

$$T_n(x) = \frac{\left(x+\sqrt{x^2-1}\right)^n + \left(x-\sqrt{x^2-1}\right)^n}{2}. \tag{2.165}$$

For any integer n, $T_n(x)$ is a polynomial of x (a *Chebyshev polynomial*). To study the properties of the Chebyshev polynomials see 19.6.3, p. 988.

2.9 Hyperbolic Functions

2.9.1 Definition of Hyperbolic Functions

Hyperbolic sine, *hyperbolic cosine* and *hyperbolic tangent* are defined by the following formulas:

$$\sinh x = \frac{e^x - e^{-x}}{2}, \tag{2.166}$$

$$\cosh x = \frac{e^x + e^{-x}}{2}, \tag{2.167}$$

$$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \tag{2.168}$$

The geometric definition (see 3.1.2.2, p. 132), is an analogy to the trigonometric functions.
Hyperbolic cotangent, *hyperbolic secant* and *hyperbolic cosecant* are defined as reciprocal values of the above hyperbolic functions:

$$\coth x = \frac{1}{\tanh x} = \frac{e^x + e^{-x}}{e^x - e^{-x}}, \tag{2.169}$$

$$\operatorname{sech} x = \frac{1}{\cosh x} = \frac{2}{e^x + e^{-x}}, \tag{2.170}$$

$$\operatorname{cosech} x = \frac{1}{\sinh x} = \frac{2}{e^x - e^{-x}}. \tag{2.171}$$

The shapes of curves of hyperbolic functions are shown in **Fig. 2.48–2.52**.

2.9.2 Graphical Representation of the Hyperbolic Functions

2.9.2.1 Hyperbolic Sine

$y = \sinh x$ (2.166) is an odd strictly monotone increasing function between $-\infty$ and $+\infty$ **(Fig. 2.49)**. The origin is its symmetry center, the inflection point, and here the angle of slope of the tangent line is $\varphi = \frac{\pi}{4}$. There is no asymptote.

2.9.2.2 Hyperbolic Cosine

$y = \cosh x$ (2.167) is an even function, it is strictly monotone decreasing for $x < 0$ from $+\infty$ to 1, and for $x > 0$ it is strictly monotone increasing from 1 until $+\infty$ **(Fig. 2.50)**. The minimum is at $x = 0$

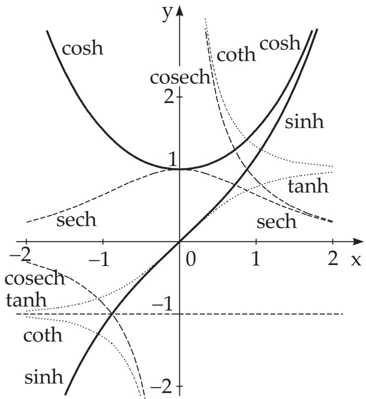

Figure 2.48

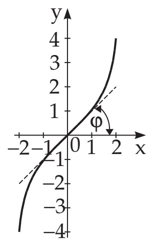

Figure 2.49

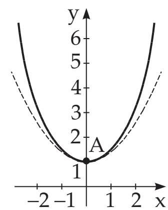

Figure 2.50

and it is equal to 1 (point $A(0,1)$); it has no asymptote. The curve is symmetric with respect to the y-axis and it always stays above the curve of the quadratic parabola $y = 1 + \frac{x^2}{2}$ (the broken-line curve). Because the function demonstrates a *catenary curve*, the curve is called the *catenoid* (see 2.15.1, p. 107).

2.9.2.3 Hyperbolic Tangent

$y = \tanh x$ (2.168) is an odd function, for $-\infty < x < +\infty$ strictly monotone increasing from -1 to $+1$ **(Fig. 2.51)**. The origin is the center of symmetry, and the inflection point, and here the angle of slope of the tangent line is $\varphi = \frac{\pi}{4}$. The asymptotes are the lines $y = \pm 1$.

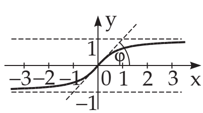

Figure 2.51

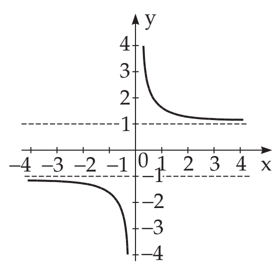

Figure 2.52

2.9.2.4 Hyperbolic Cotangent

$y = \coth x$ (2.169) is an odd function which is not continuous at $x = 0$ **(Fig. 2.52)**. It is strictly monotone decreasing in the interval $-\infty < x < 0$ and it takes its values from -1 until $-\infty$; in the interval $0 < x < +\infty$ it is also strictly monotone decreasing with values from $+\infty$ to $+1$. It has no inflection point, no extreme value. The asymptotes are the lines $x = 0$ and $y = \pm 1$.

2.9.3 Important Formulas for the Hyperbolic Functions

There are similar relations between the hyperbolic functions as between trigonometric functions. The validity of the following formulas can be shown directly from the definitions of hyperbolic functions, or considering the definitions and relations of these functions also for complex arguments, from (2.199)–(2.206), they can be calculated from the formulas known for trigonometric functions.

2.9.3.1 Hyperbolic Functions of One Variable

$$\cosh^2 x - \sinh^2 x = 1, \quad (2.172) \qquad \coth^2 x - \operatorname{cosech}^2 x = 1, \quad (2.173)$$

$$\operatorname{sech}^2 x + \tanh^2 x = 1, \quad (2.174) \qquad \tanh x \cdot \coth x = 1, \quad (2.175)$$

$$\frac{\sinh x}{\cosh x} = \tanh x, \quad (2.176) \qquad \frac{\cosh x}{\sinh x} = \coth x. \quad (2.177)$$

2.9.3.2 Expressing a Hyperbolic Function by Another One with the Same Argument

The corresponding formulas are collected in **Table 2.7**.

2.9.3.3 Formulas for Negative Arguments

$$\sinh(-x) = -\sinh x, \quad (2.178) \qquad \cosh(-x) = \cosh x, \quad (2.180)$$

$$\tanh(-x) = -\tanh x, \quad (2.179) \qquad \coth(-x) = -\coth x. \quad (2.181)$$

Table 2.7 Relations between two hyperbolic functions with the same arguments for $x > 0$

	$\sinh x$	$\cosh x$	$\tanh x$	$\coth x$
$\sinh x$	–	$\sqrt{\cosh^2 x - 1}$	$\frac{\tanh x}{\sqrt{1 - \tanh^2 x}}$	$\frac{1}{\sqrt{\coth^2 x - 1}}$
$\cosh x$	$\sqrt{\sinh^2 x + 1}$	–	$\frac{1}{\sqrt{1 - \tanh^2 x}}$	$\frac{\coth x}{\sqrt{\coth^2 x - 1}}$
$\tanh x$	$\frac{\sinh x}{\sqrt{\sinh^2 x + 1}}$	$\frac{\sqrt{\cosh^2 x - 1}}{\cosh x}$	–	$\frac{1}{\coth x}$
$\coth x$	$\frac{\sqrt{\sinh^2 x + 1}}{\sinh x}$	$\frac{\cosh x}{\sqrt{\cosh^2 x - 1}}$	$\frac{1}{\tanh x}$	–

2.9.3.4 Hyperbolic Functions of the Sum and Difference of Two Arguments (Addition Theorems)

$$\sinh(x \pm y) = \sinh x \cosh y \pm \cosh x \sinh y, \quad (2.182)$$

$$\cosh(x \pm y) = \cosh x \cosh y \pm \sinh x \sinh y, \quad (2.183)$$

$$\tanh(x \pm y) = \frac{\tanh x \pm \tanh y}{1 \pm \tanh x \tanh y}, \quad (2.184) \qquad \coth(x \pm y) = \frac{1 \pm \coth x \coth y}{\coth x \pm \coth y}. \quad (2.185)$$

2.9.3.5 Hyperbolic Functions of Double Arguments

$$\sinh 2x = 2 \sinh x \cosh x, \quad (2.186)$$

$$\cosh 2x = \sinh^2 x + \cosh^2 x, \quad (2.187)$$

$$\tanh 2x = \frac{2 \tanh x}{1 + \tanh^2 x}, \quad (2.188)$$

$$\coth 2x = \frac{1 + \coth^2 x}{2 \coth x}. \quad (2.189)$$

2.9.3.6 De Moivre Formula for Hyperbolic Functions

$$(\cosh x \pm \sinh x)^n = \left(e^{\pm x}\right)^n = e^{\pm nx} = \cosh nx \pm \sinh nx. \quad (2.190)$$

2.9.3.7 Hyperbolic Functions of Half-Argument

$$\sinh \frac{x}{2} = \pm\sqrt{\frac{1}{2}(\cosh x - 1)}, \quad (2.191) \qquad \cosh \frac{x}{2} = \sqrt{\frac{1}{2}(\cosh x + 1)}, \quad (2.192)$$

The sign of the square root in (2.191) is positive for $x > 0$ and negative for $x < 0$.

$$\tanh \frac{x}{2} = \frac{\cosh x - 1}{\sinh x} = \frac{\sinh x}{\cosh x + 1}, \quad (2.193) \qquad \coth \frac{x}{2} = \frac{\sinh x}{\cosh x - 1} = \frac{\cosh x + 1}{\sinh x}. \quad (2.194)$$

2.9.3.8 Sum and Difference of Hyperbolic Functions

$$\sinh x \pm \sinh y = 2 \sinh \frac{x \pm y}{2} \cosh \frac{x \mp y}{2}, \quad (2.195)$$

$$\cosh x + \cosh y = 2 \cosh \frac{x + y}{2} \cosh \frac{x - y}{2}, \quad (2.196)$$

$$\cosh x - \cosh y = 2 \sinh \frac{x + y}{2} \sinh \frac{x - y}{2}, \quad (2.197)$$

$$\tanh x \pm \tanh y = \frac{\sinh(x \pm y)}{\cosh x \cosh y}. \quad (2.198)$$

2.9.3.9 Relation Between Hyperbolic and Trigonometric Functions with Complex Arguments z

$$\sin z = -\mathrm{i} \sinh \mathrm{i}z, \quad (2.199) \qquad \sinh z = -\mathrm{i} \sin \mathrm{i}z, \quad (2.203)$$

$$\cos z = \cosh \mathrm{i}z, \quad (2.200) \qquad \cosh z = \cos \mathrm{i}z, \quad (2.204)$$

$$\tan z = -\mathrm{i} \tanh \mathrm{i}z, \quad (2.201) \qquad \tanh z = -\mathrm{i} \tan \mathrm{i}z, \quad (2.205)$$

$$\cot z = \mathrm{i} \coth \mathrm{i}z, \quad (2.202) \qquad \coth z = \mathrm{i} \cot \mathrm{i}z. \quad (2.206)$$

Every relation between hyperbolic functions, which contains x or ax but not $ax+b$, can be derived from the corresponding trigonometric relation with the substitution $\mathrm{i} \sinh x$ for $\sin \alpha$ and $\cosh x$ for $\cos \alpha$.

■ **A:** $\cos^2 \alpha + \sin^2 \alpha = 1$, $\cosh^2 x + \mathrm{i}^2 \sinh^2 x = 1$ or $\cosh^2 x - \sinh^2 x = 1$.

■ **B:** $\sin 2\alpha = 2\sin\alpha\cos\alpha$, $\mathrm{i}\sinh 2x = 2\mathrm{i}\sinh x\cosh x$ or $\sinh 2x = 2\sinh x\cosh x$.

2.10 Area Functions

2.10.1 Definitions

The *area functions* are the inverse functions of the hyperbolic functions, i.e., the *inverse hyperbolic functions*. The functions $\sinh x$, $\tanh x$, and $\coth x$ are strictly monotone, so they have unique inverses without any restriction; the function $\cosh x$ has two monotonic intervals so there are to consider two inverse functions. The name *area* refers to the fact that the geometric definition of the functions is the area of certain hyperbolic sectors (see 3.1.2.2, p. 132).

2.10.1.1 Area Sine

The function $y = \operatorname{Arsinh} x$ (2.207)

(Fig. 2.53) is an odd, strictly monotone increasing function, with domain and range given in **Table 2.8**. (2.53) is equivalent to the expression $x = \sinh y$. The origin is the center of symmetry and the inflection point of the curve, where the angle of slope of the tangent line is $\varphi = \dfrac{\pi}{4}$.

2.10.1.2 Area Cosine

The functions $y = \operatorname{Arcosh} x$ and $y = -\operatorname{Arcosh} x$ (2.208)

(Fig. 2.54) or $x = \cosh y$ have the domain and range given in **Table 2.8**; they are defined only for $x \geq 1$. The function curve starts at the point $A(1, 0)$ with a vertical tangent line and the function increases or decreases strictly monotonically respectively.

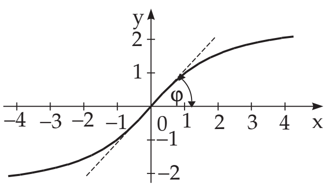

Figure 2.53

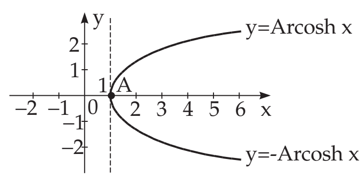

Figure 2.54

Table 2.8 Domains and ranges of the area functions

Area function	Domain	Range	Hyperbolic function with same meaning
area sine $y = \operatorname{Arsinh} x$	$-\infty < x < \infty$	$-\infty < y < \infty$	$x = \sinh y$
area cosine $y = \operatorname{Arcosh} x$ $y = -\operatorname{Arcosh} x$	$1 \leq x < \infty$	$0 \leq y < \infty$ $-\infty < y \leq 0$	$x = \cosh y$
area tangent $y = \operatorname{Artanh} x$	$\lvert x\rvert < 1$	$-\infty < y < \infty$	$x = \tanh y$
area cotangent $y = \operatorname{Arcoth} x$	$\lvert x\rvert > 1$	$-\infty < y < 0$ $0 < y < \infty$	$x = \coth y$

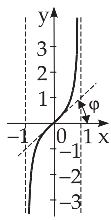

Figure 2.55

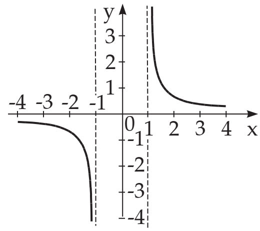

Figure 2.56

2.10.1.3 Area Tangent

The function $\quad y = \operatorname{Artanh} x$ (2.209)

(Fig. 2.55) or $x = \tanh y$ is an odd function, defined only for $|x| < 1$, with domain and range given in **Table 2.8**. The origin is the center of symmetry and also the inflection point of the curve, and here the angle of slope of the tangent line is $\varphi = \dfrac{\pi}{4}$. The asymptotes are vertical, their equations are $x = \pm 1$.

2.10.1.4 Area Cotangent

The function $\quad y = \operatorname{Arcoth} x$ (2.210)

(Fig. 2.56) or $x = \coth y$ is an odd function, defined only for $|x| > 1$, with domain and range given in **Table 2.8**. In the interval $-\infty < x < -1$ the function is strictly monotone decreasing from 0 until $-\infty$, in the interval $1 < x < +\infty$ it is strictly monotone decreasing from $+\infty$ to 0. It has three asymptotes: their equations are $y = 0$ and $x = \pm 1$.

2.10.2 Determination of Area Functions Using Natural Logarithm

From the definition of hyperbolic functions ((2.166)–(2.171), see 2.9.1, p. 89) follows that the area functions can be expressed with the logarithm function:

$$\operatorname{Arsinh} x = \ln\left(x + \sqrt{x^2+1}\right), \tag{2.211}$$

$$\operatorname{Arcosh} x = \ln\left(x + \sqrt{x^2-1}\right) = \ln\left(\frac{1}{x - \sqrt{x^2-1}}\right) \quad (x \ge 1), \tag{2.212}$$

$$\operatorname{Artanh} x = \frac{1}{2}\ln\frac{1+x}{1-x} \quad (|x| < 1), \tag{2.213}$$

$$\operatorname{Arcoth} x = \frac{1}{2}\ln\frac{x+1}{x-1} \quad (|x| > 1). \tag{2.214}$$

2.10.3 Relations Between Different Area Functions

$$\operatorname{Arsinh} x = (\operatorname{sign} x)\operatorname{Arcosh}\sqrt{x^2+1} = \operatorname{Artanh}\frac{x}{\sqrt{x^2+1}} = \operatorname{Arcoth}\frac{\sqrt{x^2+1}}{x} \quad (|x| < \infty), \tag{2.215}$$

$$\operatorname{Arcosh} x = \operatorname{Arsinh}\sqrt{x^2-1} = \operatorname{Artanh}\frac{\sqrt{x^2-1}}{x} = \operatorname{Arcoth}\frac{x}{\sqrt{x^2-1}} \quad (x \ge 1), \tag{2.216}$$

$$\operatorname{Artanh} x = \operatorname{Arsinh}\frac{x}{\sqrt{1-x^2}} = \operatorname{Arcoth}\frac{1}{x} = (\operatorname{sign} x)\operatorname{Arcosh}\frac{1}{\sqrt{1-x^2}} \quad (|x| < 1), \tag{2.217}$$

$$\operatorname{Arcoth} x = \operatorname{Artanh}\frac{1}{x} = (\operatorname{sign} x)\operatorname{Arsinh}\frac{1}{\sqrt{x^2-1}}$$
$$= (\operatorname{sign} x)\operatorname{Arcosh}\frac{|x|}{\sqrt{x^2-1}} \qquad (|x|>1). \tag{2.218}$$

2.10.4 Sum and Difference of Area Functions

$$\operatorname{Arsinh} x \pm \operatorname{Arsinh} y = \operatorname{Arsinh}\left(x\sqrt{1+y^2} \pm y\sqrt{1+x^2}\right), \tag{2.219}$$

$$\operatorname{Arcosh} x \pm \operatorname{Arcosh} y = \operatorname{Arcosh}\left(xy \pm \sqrt{(x^2-1)(y^2-1)}\right), \tag{2.220}$$

$$\operatorname{Artanh} x \pm \operatorname{Artanh} y = \operatorname{Artanh}\frac{x \pm y}{1 \pm xy}. \tag{2.221}$$

2.10.5 Formulas for Negative Arguments

$$\operatorname{Arsinh}(-x) = -\operatorname{Arsinh} x, \tag{2.222}$$

$$\operatorname{Artanh}(-x) = -\operatorname{Artanh} x, \tag{2.223}$$ $$\operatorname{Arcoth}(-x) = -\operatorname{Arcoth} x. \tag{2.224}$$

The functions Arsinh, Artanh and Arcoth are odd functions, and Arcosh (2.212) is not defined for arguments $x < 1$.

2.11 Curves of Order Three (Cubic Curves)

A curve is called an algebraic curve of order n if it can be written in the form of a polynomical equation $F(x, y) = 0$ of two variables where the left-hand side is a poynomial expression of degree n.

■ The cardioid with equation $(x^2+y^2)(x^2+y^2-2ax)-a^2y^2 = 0$ $(a > 0)$ (see 2.12.2, p. 98) is a curve of order four. The well-known conic sections (see 3.5.2.11, p. 206) result in curves of order two.

2.11.1 Semicubic Parabola

The equation $\quad y = ax^{3/2} \quad (a > 0,\ x \ge 0)$ (2.225a)

or in parametric form $\quad x = t^2, \quad y = at^3 \quad (a > 0,\ -\infty < t < \infty)$ (2.225b)

gives the *semicubic parabola* **(Fig. 2.57)**. It has a cuspidal point at the origin, it has no asymptote. The curvature $K = \dfrac{6a}{\sqrt{x}(4+9a^2x)^{3/2}}$ takes all the values between ∞ and 0. The arclength of the curve between the origin and the point $P(x, y)$ is $L = \dfrac{1}{27a^2}[(4+9a^2x)^{3/2} - 8]$.

2.11.2 Witch of Agnesi

The equation $$y = \frac{a^3}{a^2+x^2} \quad (a > 0,\ -\infty < x < \infty) \tag{2.226a}$$

determines the curve represented in **Fig. 2.58**, the *witch of Agnesi*. It has an asymptote with the equation $y = 0$, it has an extreme point at $A(0, a)$, where the radius of curvature is $r = \dfrac{a}{2}$. The inflection points B and C are at $\left(\pm\dfrac{a}{\sqrt{3}}, \dfrac{3a}{4}\right)$, where the angles of slope of the tangent lines are $\tan\varphi = \mp\dfrac{3\sqrt{3}}{8}$.

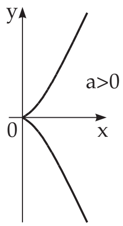

Figure 2.57

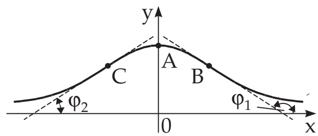

Figure 2.58

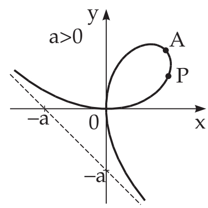

Figure 2.59

The area of the region between the curve and its asymptote is equal to $S = \pi a^2$. The witch of Agnesi (2.226a) is a special case of the Lorentz or Breit–Wigner curve

$$y = \frac{a}{b^2 + (x-c)^2} \quad (a > 0\,, b \neq 0)\,. \tag{2.226b}$$

■ The Fourier transform of the damped oscillation is the Lorentz or Breit–Wigner curve (see 15.3.1.4, p. 791).

2.11.3 Cartesian Folium (Folium of Descartes)

The equation $\quad x^3 + y^3 = 3axy \quad (a > 0) \quad$ or (2.227a)

in parametric form $\quad x = \dfrac{3at}{1+t^3}\,, \; y = \dfrac{3at^2}{1+t^3} \quad$ with

$$t = \tan \sphericalangle P0x \quad (a > 0, -\infty < t < -1 \text{ and } -1 < t < \infty) \tag{2.227b}$$

gives the *Cartesian folium curve* represented in **Fig. 2.59**. The origin is a double point because the curve passes through it twice, and here both coordinate axes are tangent lines. At the origin the radius of curvature for both branches of the curve is $r = \dfrac{3a}{2}$. The equation of the asymptote is $x + y + a = 0$. The vertex A has the coordinates $A\left(\dfrac{3}{2}a, \dfrac{3}{2}a\right)$. The area of the loop is $S_1 = \dfrac{3a^2}{2}$. The area S_2 between the curve and the asymptote has the same value.

2.11.4 Cissoid

The equation $\quad y^2 = \dfrac{x^3}{a-x} \quad (a > 0),$ (2.228a)

or in parametric form $\quad x = \dfrac{at^2}{1+t^2}\,, \; y = \dfrac{at^3}{1+t^2} \quad$ with

$$t = \tan \sphericalangle P0x \quad (a > 0, -\infty < t < \infty) \tag{2.228b}$$

or with polar coordinates $\rho = \dfrac{a \sin^2 \varphi}{\cos \varphi} \quad (a > 0)$ (2.228c)

(Fig. 2.60) describes the locus of the points P for which

$$\overline{0P} = \overline{MQ} \tag{2.229}$$

is valid. Here M is the second intersection point of the line $0P$ with the drawn circle of radius $\frac{a}{2}$, and Q is the intersection point of the line $0P$ with the asymptote $x = a$. The area between the curve and the asymptote is equal to $S = \frac{3}{4}\pi a^2$.

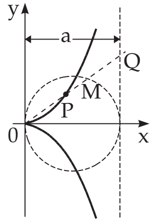

Figure 2.60

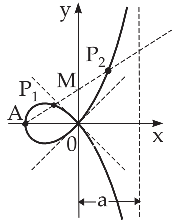

Figure 2.61

2.11.5 Strophoide

Strophoide is the locus of the points P_1 and P_2, which are on an arbitrary half-line starting at A (A is on the negative x-axis) and for which the equalities

$$\overline{MP_1} = \overline{MP_2} = \overline{0M} \tag{2.230}$$

are valid. Here M is the intersection point with the y-axis **(Fig. 2.61)**. The equation of the strophoide in Cartesian, and in polar coordinates, and in parametric form is:

$$y^2 = x^2\left(\frac{a+x}{a-x}\right) \quad (a > 0), \tag{2.231a}$$

$$\rho = -a\frac{\cos 2\varphi}{\cos\varphi} \quad (a > 0), \tag{2.231b}$$

$$x = a\frac{t^2-1}{t^2+1}, \quad y = at\frac{t^2-1}{t^2+1} \text{ with } t = \tan \sphericalangle P0x \quad (a > 0, -\infty < t < \infty). \tag{2.231c}$$

The origin is a double point with tangent lines $y = \pm x$. The asymptote has the equation $x = a$. The vertex is $A(-a, 0)$. The area of the loop is $S_1 = 2a^2 - \frac{1}{2}\pi a^2$, and the area between the curve and the asymptote is $S_2 = 2a^2 + \frac{1}{2}\pi a^2$.

2.12 Curves of Order Four (Quartics)

2.12.1 Conchoid of Nicomedes

The *Conchoid of Nicomedes* **(Fig. 2.62)** is the locus of the points P, for which

$$\overline{0P} = \overline{0M} \pm l \tag{2.232}$$

holds, where M is the intersection point of the line $\overline{0P_10P_2}$ with the asymptote $x = a$. The "+" sign belongs to the right branch of the curve, the "−" sign belongs to the left one in relation to the asymptote. The equations for the *conchoid of Nicomedes* are the following in Cartesian coordinates, in parametric form and in polar coordinates:

$$(x-a)^2(x^2+y^2) - l^2x^2 = 0 \quad (a > 0, l > 0), \tag{2.233a}$$

$$x = a + l\cos\varphi, \quad y = a\tan\varphi + l\sin\varphi$$
$$\left(a > 0, \text{ right branch: } -\frac{\pi}{2} < \varphi < \frac{\pi}{2}, \text{ left branch: } \frac{\pi}{2} < \varphi < \frac{3\pi}{2}\right), \tag{2.233b}$$

$$\rho = \frac{a}{\cos\varphi} \pm l \quad (\text{“+” sign: right branch, “−” sign: left branch,}) \tag{2.233c}$$

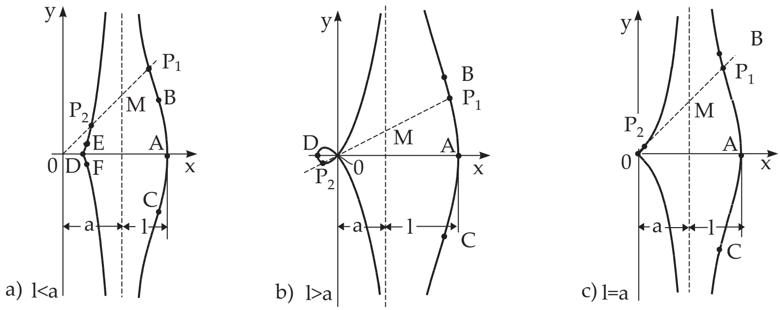

Figure 2.62

1. Right Branch: The asymptote is $x = a$. The vertex A is at $(a + l, 0)$, the inflection points B, C have as x-coordinate the greatest root of the equation $x^3 - 3a^2x + 2a(a^2 - l^2) = 0$. The area between the right branch and the asymptote is $S = \infty$.
2. Left Branch: The asymptote is $x = a$. The vertex D is at $(a - l, 0)$. The origin is a singular point, whose type depends on a and l:

Case a) For $l < a$ it is an isolated point **(Fig. 2.62a)**. The curve has two further inflection points E and F, whose abscissa is the second greatest root of the equation $x^3 - 3a^2x + 2a(a^2 - l^2) = 0$.

Case b) For $l > a$ the origin is a double point **(Fig. 2.62b)**. The curve has a maximum and a minimum value at $x = a - \sqrt[3]{al^2}$. At the origin the slopes of the tangent lines are $\tan\alpha = \dfrac{\pm\sqrt{l^2 - a^2}}{a}$. Here the radius of curvature is $r_0 = \dfrac{l\sqrt{l^2 - a^2}}{2a}$.

Case c) For $l = a$ the origin is a cuspidal point **(Fig. 2.62c)**.

2.12.2 General Conchoid

The *conchoid of Nicomedes* is a special case of the *general conchoid*. One gets the conchoid of a given curve by elongating the length of the position vector of every point by a given constant segment $\pm l$. Considering a curve in a polar coordinate system with an equation $\rho = f(\varphi)$, then the equation of its conchoid is

$$\rho = f(\varphi) \pm l. \tag{2.234}$$

So, the conchoid of Nicomedes is the *conchoid of the line*.

2.12.3 Pascal's Limaçon

The *conchoid of a circle* is called the *Pascal limaçon* **(Fig. 2.63)** if in (2.232) the origin is on the perimeter of the circle, which is a further special case of the general conchoid (see 2.12.2, p. 98). The equations in the Cartesian and in the polar coordinate systems and in parametric form are the following (see also (2.246c), p. 105):

$$(x^2 + y^2 - ax)^2 = l^2(x^2 + y^2) \quad (a > 0,\ l > 0), \tag{2.235a}$$

$$\rho = a\cos\varphi + l \quad (a > 0,\ l > 0), \tag{2.235b}$$

$$x = a\cos^2\varphi + l\cos\varphi, \quad y = a\cos\varphi\sin\varphi + l\sin\varphi \quad (a > 0,\ l > 0,\ 0 \le \varphi < 2\pi) \tag{2.235c}$$

with a as the diameter of the circle. The vertices A, B are at $(a \pm l, 0)$. The shape of the curve depends on the quantities a and l, as can be seen in **Fig. 2.63** and **Fig. 2.64**.

a) Extreme Points and Inflection Points: For $a > l$ the curve has four extreme points C, D, E, F; for $a \le l$ it has two; they are at $\left(\cos\varphi = \dfrac{-l \pm \sqrt{l^2 + 8a^2}}{4a}\right)$. For $a < l < 2a$ there exist two inflection points G and H at $\left(\cos\varphi = -\dfrac{2a^2 + l^2}{3al}\right)$.

b) Double Tangent: For $l < 2a$, at the points I and K at $\left(-\dfrac{l^2}{4a}, \pm\dfrac{l\sqrt{4a^2 - l^2}}{4a}\right)$ there is a double tangent.

c) Singular Points: The origin is a singular point: For $a < l$ it is an isolated point, for $a > l$ it is a

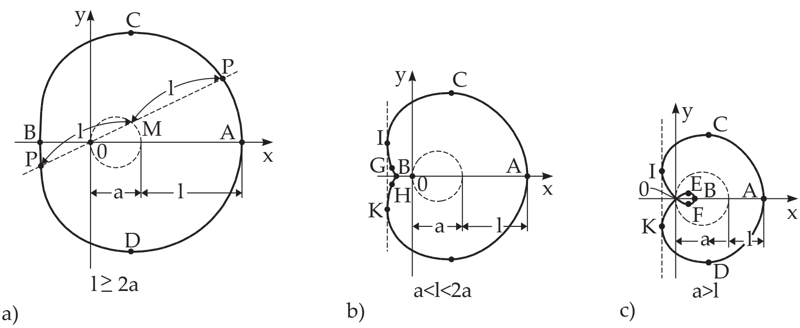

Figure 2.63

double point and the slopes of the tangent lines are $\tan\alpha = \pm\dfrac{\sqrt{a^2 - l^2}}{l}$, here the radius of curvature is $r_0 = \dfrac{1}{2}\sqrt{a^2 - l^2}$.

For $a = l$ the origin is a cuspidal point; then the curve is called a *cardioid* (see also 2.13.3, p. 103).

The area of the limaçon is $S = \dfrac{\pi a^2}{2} + \pi l^2$, where in the case $a > l$ **(Fig. 2.63c)** the area of the inside loop is counted twice.

2.12.4 Cardioid

The *cardioid* **(Fig. 2.64)** can be defined in two different ways, as:

1. Special case of the *Pascal limaçon* with

$$\overline{0P} = \overline{0M} \pm a, \tag{2.236}$$

where a is the diameter of the circle.

2. Special case of the *epicycloid* with the same diameter a for the fixed and for the moving circle (see 2.13.3, p. 103). The equation is

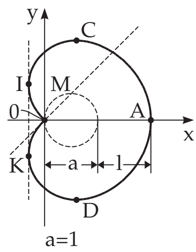

Figure 2.64

$$(x^2+y^2)^2 - 2ax(x^2+y^2) = a^2y^2 \quad (a>0), \tag{2.237a}$$

and the parametric form, and the equation in polar coordinates are:

$$x = a\cos\varphi(1+\cos\varphi), \quad y = a\sin\varphi(1+\cos\varphi)$$
$$(a>0, 0\le\varphi<2\pi), \tag{2.237b}$$
$$\rho = a(1+\cos\varphi) \quad (a>0). \tag{2.237c}$$

The origin is a cuspidal point. The vertex A is at $(2a,0)$; extreme points C and D are at $\cos\varphi = \frac{1}{2}$ with coordinates $\left(\frac{3}{4}a\,, \pm\frac{3\sqrt{3}}{4}a\right)$. The area is $S = \frac{3}{2}\pi a^2$, i.e., six times the area of a circle with diameter a. The length of the curve is $L = 8a$.

2.12.5 Cassinian Curve

The locus of the points P, for which the product of the distances from two fixed points F_1 and F_2 with coordinates $(c,0)$ and $(-c,0)$ resp., is equal to a constant $a^2 \ne 0$, is called a *Cassinian curve* **(Fig. 2.65)**:

$$\overline{F_1P}\cdot\overline{F_2P} = a^2. \tag{2.238}$$

The equations in Cartesian and polar coordinates are:

$$(x^2+y^2)^2 - 2c^2(x^2-y^2) = a^4 - c^4, \quad (a>0, c>0), \tag{2.239a}$$

$$\rho^2 = c^2\cos 2\varphi \pm \sqrt{c^4\cos^2 2\varphi + (a^4-c^4)} \quad (a>0, c>0). \tag{2.239b}$$

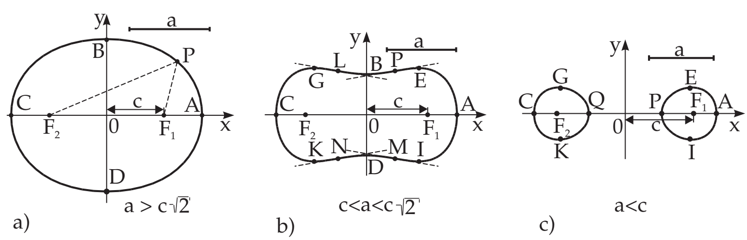

Figure 2.65

The shape of the curve depends on the quantities a and c :

Case $a > c\sqrt{2}$: For $a > c\sqrt{2}$ the curve is an oval whose shape resembles an ellipse **(Fig. 2.65a)**. The intersection points A, C with the x-axis are $(\pm\sqrt{a^2+c^2}\,, 0)$, the intersection points B, D with the y-axis are $(0,\ \pm\sqrt{a^2-c^2})$.

Case $a = c\sqrt{2}$: For $a = c\sqrt{2}$ the curve is of the same type with A, C $(\pm c\sqrt{3}\,,\ 0)$ and B, D $(0, \pm c)$, where the curvature at the points B and D is equal to 0, i.e., there is a narrow contact with the lines $y = \pm c$.

Case $c < a < c\sqrt{2}$: For $c < a < c\sqrt{2}$ the curve is a pressed oval **(Fig. 2.65b)**. The intersection points with the axes are the same as in the case $a > c\sqrt{2}$, also the extreme points B, D, while there are further extreme points E, G, K, I at $\left(\pm\frac{\sqrt{4c^4-a^4}}{2c}\,,\ \pm\frac{a^2}{2c}\right)$ and there are four inflection points P, L,

M, N at $\left(\pm\sqrt{\frac{1}{2}(m-n)},\ \pm\sqrt{\frac{1}{2}(m+n)}\right)$ with $n = \frac{a^4 - c^4}{3c^2}$ and $m = \sqrt{\frac{a^4 - c^4}{3}}$.

Case $a = c$: For $a = c$ there is the *lemniscate*.

Case $a < c$: For $a < c$ there are two ovals **(Fig. 2.65c)**. The intersection points A, C and P, Q with the x-axis are at $(\pm\sqrt{a^2+c^2},\ 0)$ and $(\pm\sqrt{c^2-a^2}, 0)$. The extreme points E, G, K, I are at $\left(\pm\frac{\sqrt{4c^4-a^4}}{2c},\ \pm\frac{a^2}{2c}\right)$. The radius of curvature is $r = \frac{2a^2\rho^3}{c^4 - a^4 + 3\rho^4}$, where ρ satisfies the polar coordinate representation.

2.12.6 Lemniscate

The *lemniscate* **(Fig. 2.66)** is the special case $a = c$ of the *Cassinian curve* satisfying the condition

$$\overline{F_1P} \cdot \overline{F_2P} = \left(\frac{\overline{F_1F_2}}{2}\right)^2, \tag{2.240}$$

where the fixed points F_1, F_2 are at $(\pm a, 0)$. The equation in Cartesian coordinates is

$$(x^2+y^2)^2 - 2a^2(x^2-y^2) = 0 \quad (a > 0) \tag{2.241a}$$

and in polar coordinates

$$\rho = a\sqrt{2\cos 2\varphi} \quad (a > 0). \tag{2.241b}$$

The origin is a double point and an inflection point at the same time, where the tangents are $y = \pm x$.

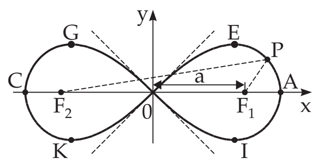

Figure 2.66

The intersection points A and C with the x-axis are at $(\pm a\sqrt{2}, 0)$, the extreme points of the curve E, G, K, I are at $\left(\pm\frac{a\sqrt{3}}{2},\ \pm\frac{a}{2}\right)$. The polar angle at these points is $\varphi = \pm\frac{\pi}{6}$. The radius of curvature is $r = \frac{2a^2}{3\rho}$ and the area of every loop is $S = a^2$.

2.13 Cycloids

2.13.1 Common (Standard) Cycloid

The *cycloid* is a curve which is described by a point of the perimeter of a circle while the circle rolls along a line without sliding **(Fig. 2.67)**. The equation of the *usual cycloid* written in parametric form is the following:

$$x = a(t - \sin t),\ \ y = a(1 - \cos t) \quad (a > 0,\ -\infty < t < \infty), \tag{2.242a}$$

where a is the radius of the circle and t is the angle $\measuredangle PC_1B$ in radian measure. In Cartesian coordinates

$$x + \sqrt{y(2a-y)} = a\operatorname{arc}_k\cos\frac{a-y}{a} \quad (a > 0,\ k = 0, \pm1, \pm2, \ldots) \tag{2.242b}$$

holds.

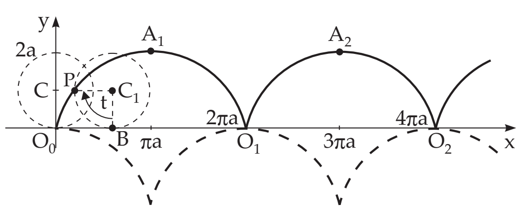

Figure 2.67

The curve is periodic with period $\overline{O_0O_1} = 2\pi a$. At $O_0, O_1, O_2, \ldots, O_k = (2k\pi a, 0)$ there are cusps, the vertices are at $A_{k+1} = ((2k+1)\pi a, 2a \ (k = 0, \pm 1, \pm 2, \ldots))$. The arc length of O_0P is $L = 8a\sin^2(t/4)$, the length of one arch is $L_{O_0A_1O_1} = 8a$. The area of one arch is $S = 3\pi a^2$. The radius of curvature is $r = 4a\sin\frac{1}{2}t$, at the vertices $r_A = 4a$. The evolute of a cycloid (see 3.6.1.6, p. 254) is a *congruent cycloid*, which is denoted in **Fig. 2.67** by the broken line.

2.13.2 Prolate and Curtate Cycloids or Trochoids

Prolate and *curtate cycloids* or *trochoids* are curves described by a point, which is inside or outside of a circle, fixed on a half-line starting from the center of the circle, while the circle rolls along a line without sliding **(Fig. 2.68).**

The equation of the trochoid in parametric form is

$$x = a(t - \lambda\sin t), \tag{2.243a}$$

$$y = a(1 - \lambda\cos t), \tag{2.243b}$$

where a is the radius of the circle, t is the angle $\sphericalangle PC_1M$, and $\lambda a = \overline{C_1P}$.
The case $\lambda > 1$ gives the prolate cycloid and $\lambda < 1$ the curtate one.
The period of the curve is $\overline{O_0O_1} = 2\pi a$, the maximum points are at $A_1, A_2, \ldots, A_{k+1} = ((2k+1)\pi a, (1+\lambda)a)$, the minimum points are at $B_0, B_1, B_2, \ldots, B_k = (2k\pi a, (1-\lambda)a)$ $(k = 0, \pm 1, \pm 2, \ldots)$. The prolate cycloid has double points at $D_0, D_1, D_2, \ldots, D_k = \left[2k\pi a\,,\, a\left(1 - \sqrt{\lambda^2 - {t_0}^2}\right)\right]$, where t_0 is the smallest positive root of the equation $t = \lambda\sin t$.

The curtate cycloid has inflection points at $E_1, E_2, \ldots, E_{k+1} = \left[a\left(\mathrm{arc}_k\cos\lambda - \lambda\sqrt{1-\lambda^2}\right), a\,(1-\lambda^2)\right]$.
The calculation of the length of one cycle can be done by the integral $L = a\int_0^{2\pi}\sqrt{1+\lambda^2-2\lambda\cos t}\,dt$. The shaded area in **Fig. 2.68** is $S = \pi a^2(2+\lambda^2)$. The radius of curvature is $r = a\dfrac{(1+\lambda^2-2\lambda\cos t)^{3/2}}{\lambda(\cos t - \lambda)}$, which has the value $r_A = -a\dfrac{(1+\lambda)^2}{\lambda}$ at the maxima and the value $r_B = a\dfrac{(1-\lambda)^2}{\lambda}$ at the minima.

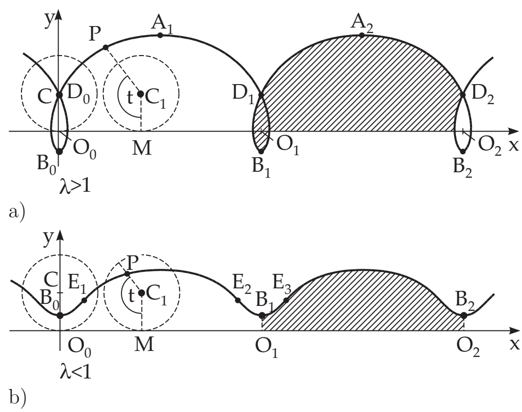

Figure 2.68

2.13.3 Epicycloid

A curve is called an *epicycloid*, if it is described by a point of the perimeter of a circle while this circle rolls along the outside of another circle without sliding **(Fig. 2.69)**. The equation of the epicycloid in parametric form is

$$x = (A+a)\cos\varphi - a\cos\frac{A+a}{a}\varphi, \quad y = (A+a)\sin\varphi - a\sin\frac{A+a}{a}\varphi \quad (-\infty < \varphi < \infty), \tag{2.244}$$

where A is the radius of the fixed circle, a is the radius of the rolling one, and φ is the angle $\sphericalangle C0x$. The shape of the curve depends on the quotient $m = \dfrac{A}{a}$.

For $m = 1$ one gets the *cardioid*.

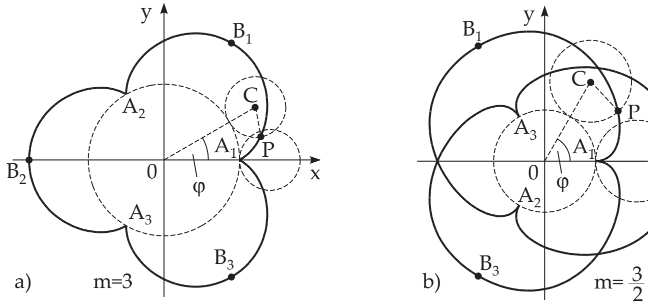

Figure 2.69

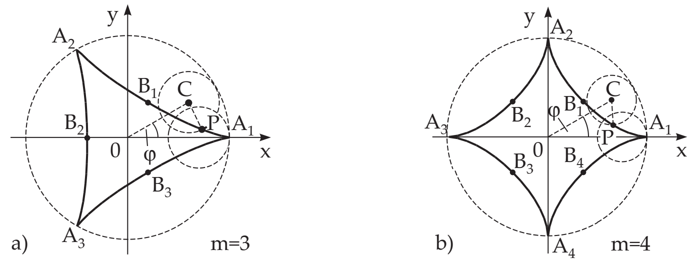

Figure 2.70

Case *m* integer: For an integer m the curve consists of m identically shaped branches surrounding the fixed curve **(Fig. 2.69a)**. The cusps $A_1, A_2, \ldots, A_m$ are at $\left(\rho = A, \;\; \varphi = \dfrac{2k\pi}{m} \;\; (k = 0, 1, \ldots, m-1)\right)$, the vertices $B_1, B_2, \ldots, B_m$ are at $\left(\rho = A + 2a, \quad \varphi = \dfrac{2\pi}{m}\left(k + \dfrac{1}{2}\right)\right)$.

Case *m* a rational fraction: If m is a non-integer rational number, the identically shaped branches follow each other around the fixed circle, overlapping the previous ones, until the moving point P returns back to the starting-point after a finite number of circuits **(Fig. 2.69b)**.

Case *m* an irrational: For an irrational m the number of round trips is infinite; the point P never returns to the starting-point.

The length of one branch is $L_{A_1B_1A_2} = \dfrac{8(A+a)}{m}$. For an integer m the total length of the closed curve is $L_{\text{total}} = 8(A+a)$. The area of the sector $A_1B_1A_2A_1$ (without the sector of the fixed circle) is $S = \pi a^2 \left(\dfrac{3A+2a}{A}\right)$. The radius of curvature is $r = \dfrac{4a(A+a)}{2a+A} \sin \dfrac{A\varphi}{2a}$, at the vertices $r_B = \dfrac{4a(A+a)}{2a+A}$.

2.13.4 Hypocycloid and Astroid

A curve is called a *hypocycloid* if it is described by a point of the perimeter of a circle, while this circle rolls along the inside of another circle without sliding **(Fig. 2.70)**. The equation of the hypocycloid, the coordinates of the vertices, the cusps, the formulas for the arc-length, the area, and the radius of curvature are similar to the corresponding formulas for the epicycloid, only "$+a$" is to be replaced by "$-a$". The number of cusps for integer, rational or irrational m is the same as for the epicycloid (now $m > 1$ holds).

Case $m = 2$: For $m = 2$ the curve is actually the diameter of the fixed circle.

Case $m = 3$: For $m = 3$ the hypocycloid has three branches **(Fig. 2.70a)** with the equation

$$x = a(2\cos\varphi + \cos 2\varphi), \qquad y = a(2\sin\varphi - \sin 2\varphi) \tag{2.245a}$$

and $L_{\text{total}} = 16a$, $S_{\text{total}} = 2\pi a^2$.

Case $m = 4$: For $m = 4$ **(Fig. 2.70b)** the hypocycloid has four branches, and it is called an *astroid* (or *asteroid*). Its equation in Cartesian coordinates and in parametric form is

$$x^{2/3} + y^{2/3} = A^{2/3}, \quad (2.245b) \qquad x = A\cos^3\varphi, \quad y = A\sin^3\varphi \quad (0 \le \varphi < \pi) \quad (2.245c)$$

and $L_{\text{total}} = 24a = 6A$, $S_{\text{total}} = \frac{3}{8}\pi A^2$.

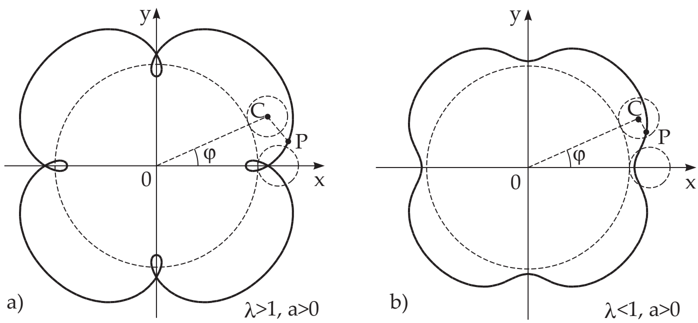

Figure 2.71

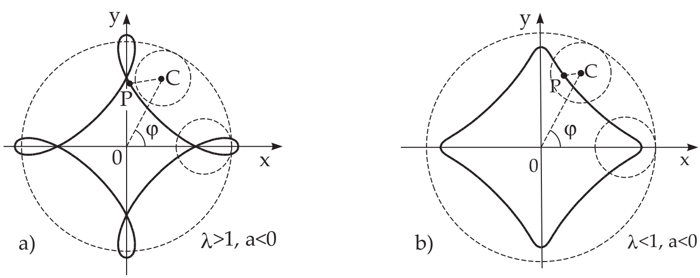

Figure 2.72

2.13.5 Prolate and Curtate Epicycloid and Hypocycloid

The *prolate* and *curtate epicycloid* and the *prolate* and *curtate hypocycloid*, which are also called the *epitrochoid* and *hypotrochoid*, are curves **(Fig. 2.71** and **Fig. 2.72)** described by a point, which is inside or outside of a circle, fixed on a half-line starting at the center of the circle, while the circle rolls around the outside (epitrochoid) or the inside (hypotrochoid) of another circle, without sliding. The equation of the epitrochoid in parametric form is

$$x = (A+a)\cos\varphi - \lambda a \cos\left(\frac{A+a}{a}\varphi\right), \qquad y = (A+a)\sin\varphi - \lambda a \sin\left(\frac{A+a}{a}\varphi\right), \tag{2.246a}$$

where A is the radius of the fixed circle and a is the radius of the rolling one. For the hypocycloid "$+a$" is to be replaced by "$-a$". For $\lambda a = \overline{CP}$ one of the inequalities $\lambda > 1$ or $\lambda < 1$ is valid, depending on whether the prolate or the curtate curve is considered.

For $A = 2a$, and for arbitrary $\lambda \neq 1$ the hypocycloid with equation

$$x = a(1+\lambda)\cos\varphi\,, \quad y = a(1-\lambda)\sin\varphi \quad (0 \le \varphi < 2\pi) \tag{2.246b}$$

describes an ellipse with semi-axes $a(1+\lambda)$ and $a(1-\lambda)$. For $A = a$ it results in the Pascal limaçon (see also 2.12.3, p. 98):

$$x = a(2\cos\varphi - \lambda\cos 2\varphi), \quad y = a(2\sin\varphi - \lambda\sin 2\varphi). \tag{2.246c}$$

Remark: For the Pascal limaçon on 2.12.3, p. 98 the quantity denoted by a there is denoted by $2\lambda a$ here, and the l there is the diameter $2a$ here. Furthermore the coordinate system is different.

2.14 Spirals

2.14.1 Archimedean Spiral

An *Archimedean spiral* is a curve **(Fig. 2.73)** described by a point which is moving with constant speed v on a ray, while this ray rotates around the origin at a constant angular velocity ω. The equation of the Archimedean spiral in polar coordinates is

$$\rho = a|\varphi|, \quad a = \frac{v}{\omega} \quad (a > 0, -\infty < \varphi < \infty)\,. \tag{2.247}$$

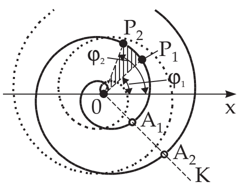

Figure 2.73

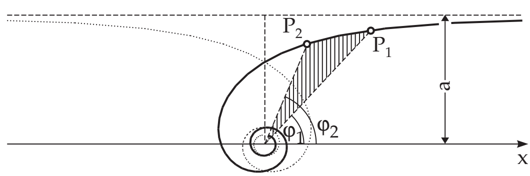

Figure 2.74

The curve has two branches ($\varphi < 0\,, \varphi > 0$) in a symmetric position with respect to the y-axis. Every ray $0K$ intersects the curve at the points 0, A_1, $A_2, \ldots, A_n, \ldots$; their distance is $\overline{A_iA_{i+1}} = 2\pi a$. The arclength $\overset{\frown}{0P}$ is $L = \frac{a}{2}\left(\varphi\sqrt{\varphi^2+1} + \operatorname{Arsinh}\varphi\right)$, where for large φ the expression $\frac{2L}{a\varphi^2}$ tends to 1. The area of the sector $P_1 0 P_2$ is $S = \frac{a^2}{6}({\varphi_2}^3 - {\varphi_1}^3)$. The radius of curvature is $r = a\frac{(\varphi^2+1)^{3/2}}{\varphi^2+2}$ and at the origin $r_0 = \frac{a}{2}$.

2.14.2 Hyperbolic Spiral

The equation of the *hyperbolic spiral* in polar coordinates is

$$\rho = \frac{a}{|\varphi|} \quad (a > 0,\ -\infty < \varphi < 0,\ 0 < \varphi < \infty)\,. \tag{2.248}$$

The curve of the hyperbolic spiral **(Fig. 2.74)** has two branches ($\varphi < 0\,,\varphi > 0$) in a symmetric position with respect to the y-axis. The line $y = a$ is the asymptote for both branches, and the origin is an asymptotic point. The area of the sector $P_1 0 P_2$ is $S = \dfrac{a^2}{2}\left(\dfrac{1}{\varphi_1} - \dfrac{1}{\varphi_2}\right)$, and $\lim\limits_{\varphi_2\to\infty} S = \dfrac{a^2}{2\varphi_1}$ is valid.

The radius of curvature is $r = \dfrac{a}{\varphi}\left(\dfrac{\sqrt{1+\varphi^2}}{\varphi}\right)^3$.

2.14.3 Logarithmic Spiral

The *logarithmic spiral* is a curve **(Fig. 2.75)** which intersects all the rays starting at the origin 0 at the same angle α. The equation of the logarithmic spiral in polar coordinates is

$$\rho = ae^{k\varphi} \quad (a > 0,\ -\infty < \varphi < \infty)\,, \tag{2.249}$$

where $k = \cot\alpha\,$. The origin is the asymptotic point of the curve. The arclength $\overset{\frown}{P_1P_2}$ is $L = \dfrac{\sqrt{1+k^2}}{k}(\rho_2 - \rho_1)\,$, the limit of the arclength $\overset{\frown}{0P}$ calculated from the origin is $L_0 = \dfrac{\sqrt{1+k^2}}{k}\rho$. The radius of curvature is $r = \sqrt{1+k^2}\rho = L_0 k$.

Special case of a circle: For $\alpha = \dfrac{\pi}{2}$ holds $k = 0$, and the curve is a circle.

2.14.4 Evolvent of the Circle

The *evolvent of the circle* is a curve **(Fig. 2.76)** which is described by the endpoint of a string while rolling it off a circle, and always keeping it tight, so that $\overset{\frown}{AB} = \overline{BP}$. The equation of the *evolvent of the circle* is in parametric form

$$x = a\cos\varphi + a\varphi\sin\varphi, \quad y = a\sin\varphi - a\varphi\cos\varphi, \tag{2.250}$$

where a is the radius of the circle, and $\varphi = \sphericalangle B0x$. The curve has two branches in symmetric position with respect to the x-axis. It has a cusp at $A(a, 0)$, and the intersection points with the x-axis are at $x = \dfrac{a}{\cos\varphi_0}$, where φ_0 are the roots of the equation $\tan\varphi = \varphi$. The arclength of $\overset{\frown}{AP}$ is $L = \dfrac{1}{2}a\varphi^2$. The radius of curvature is $r = a\varphi = \sqrt{2aL}\,$; the centre of curvature B is on the circle, i.e., the circle is the evolute of the curve.

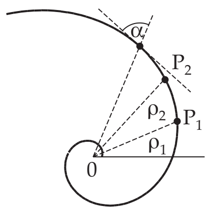

Figure 2.75

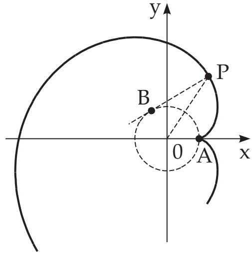

Figure 2.76

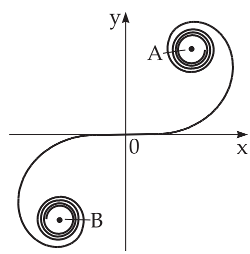

Figure 2.77

2.14.5 Clothoid

The *clothoid* is a curve **(Fig. 2.77)** such that at every point the radius of curvature is inversely proportional to the arclength between the origin and the considered point:

$$r = \frac{a^2}{s} \quad (a > 0)\,. \tag{2.251a}$$

The equation of the clothoid in parametric form is

$$x = a\sqrt{\pi}\int\limits_0^t \cos\frac{\pi t^2}{2}\,dt, \quad y = a\sqrt{\pi}\int\limits_0^t \sin\frac{\pi t^2}{2}\,dt \quad \text{with} \quad t = \frac{s}{a\sqrt{\pi}}, \quad s = \overset{\frown}{0P}\,. \tag{2.251b}$$

The integrals cannot be expressed in terms of elementary functions; but for any given value of the parameter $t = t_0, t_1, \ldots$ it is possible to calculate them by numerical integration (see 19.3, p. 963), so the clothoid can be drawn pointwise. About calculations with a computer see the literature.

The curve is centrosymmetric with respect to the origin, which is also the inflection point. At the inflection point the x-axis is the tangent line. At A and B the curve has asymptotic points with coordinates $\left(+\frac{a\sqrt{\pi}}{2}, +\frac{a\sqrt{\pi}}{2}\right)$ and $\left(-\frac{a\sqrt{\pi}}{2}, -\frac{a\sqrt{\pi}}{2}\right)$. The clothoid is applied, for instance in road construction, where the transition between a line and a circular arc is made by a clothoid segment.

2.15 Various Other Curves

2.15.1 Catenary Curve

The *catenary curve* is a curve wich has the shape of a homogeneous, flexible but inextensible heavy chain hung at both ends **(Fig. 2.78)** represented by a continuous line. The equation of the *catenary curve* is

$$y = a\cosh\frac{x}{a} = a\frac{e^{x/a} + e^{-x/a}}{2} \quad (a > 0). \tag{2.252}$$

The parameter a determines the vertex A at $(0, a)$. The curve is symmetric to the y-axis, and is always higher than the parabola $y = a + \frac{x^2}{2a}$, which is represented by the broken line in **Fig. 2.78**. The arclength of $\overset{\frown}{AP}$ is $L = a\sinh\frac{x}{a} = a\frac{e^{x/a} - e^{-x/a}}{2}$. The area of the region $0APM$ has the value $S = a\,L = a^2\sinh\frac{x}{a}$. The radius of curvature is $r = \frac{y^2}{a} = a\cosh^2\frac{x}{a} = a + \frac{L^2}{a}$.

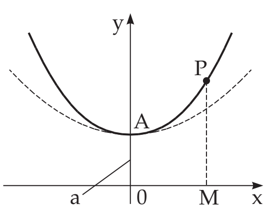

Figure 2.78

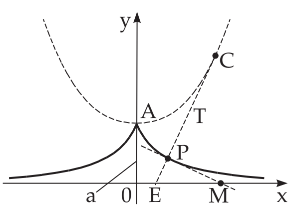

Figure 2.79

The catenary curve is the evolute of the tractrix (see 3.6.1.6, p. 254), so the tractrix is the evolvent (see 3.6.1.6, p. 255) of the catenary curve with vertex A at $(0, a)$.

2.15.2 Tractrix

The *tractrix* (the thick line in **Fig. 2.79**) is a curve such that the length of the segment $\overline{PM}$ of the tangent line between the point of contact P and the intersection point with a given straight line, here the x-axis, is a constant a. Fasting one end of an inextensible string of length a to a material point P, and dragging the other end along a straight line, here the x-axis, then P draws a tractrix. The equation of the tractrix is

$$x = a\,\mathrm{Arcosh}\,\frac{a}{y} \pm \sqrt{a^2 - y^2} = a \ln \frac{a \pm \sqrt{a^2 - y^2}}{y} \mp \sqrt{a^2 - y^2} \quad (a > 0\,, 0 < y \le a). \tag{2.253}$$

The x-axis is the asymptote. The point A at $(0, a)$ is a cusp. The curve is symmetric with respect to the y-axis. The arclength of $\overset{\frown}{AP}$ is $L = a \ln \frac{a}{y}$. For increasing arclength L the difference $L - x$ tends to the value $a(1 - \ln 2) \approx 0.307a$, where x is the abscissa of the point P. The radius of curvature is $r = a \cot \frac{x}{y}$. The radius of curvature $\overline{PC}$ and the segment $\overline{PE} = b$ are inversely proportional: $rb = a^2$. The evolute (see 3.6.1.6, p. 254) of the tractrix, i.e., the geometric locus of the centers of circles of curvature C, is the catenary curve (2.252), represented by the dotted line in **Fig. 2.79**.

2.16 Determination of Empirical Curves

2.16.1 Procedure

2.16.1.1 Curve-Shape Comparison

If there are only empirical data for a function $y = f(x)$, it is possible to get an approximate formula in two steps. First choose a formula for an approximation which contains free parameters. Then calculate the values of the parameters. If there is no theoretical description for the type of formula, then first choose the approximate formula which is the simplest among the possible functions, comparing their curves with the curve of empirical data. Estimation of similarity by eye can be deceptive. Therefore, after the choice of an approximate formula, and before the determination of the parameters, it is to check whether it is appropriate.

2.16.1.2 Rectification

Supposing there is a definite relation between x and y and in the chosen approximate formula two functions $X = \varphi(x, y)$ and $Y = \psi(x, y)$ are introduced such that a linear relation of the form

$$Y = AX + B \tag{2.254}$$

holds, where A and B are constant. Calculating the corresponding X and Y values for the given x and y values, and considering their graphical representation, it is easy to check if they are approximately on a straight line, or not. Then it can be decided whether the chosen formula is appropriate.

■ **A:** If the approximate formula is $y = \frac{x}{ax + b}$, then substituting $X = x$, $Y = \frac{x}{y}$, one gets $Y = aX + b$. Another possible substitution is $X = \frac{1}{x}$, $Y = \frac{1}{y}$. Then $Y = a + bX$ follows.

■ **B:** Using semi-logarithmic paper, 2.17.2.1, p. 116.

■ **C:** Using double logarithmic paper, 2.17.2.2, p. 117.

In order to decide whether empirical data satisfy a linear relation $Y = AX + B$ or not, one can use *linear regression* or *correlation* (see 16.3.4, p. 839). The reduction of a functional relationship to a linear relation is called *rectification*. Examples of rectification of some formulas are given in 2.16.2, p. 109, and for an example discussed in detail, see in 2.16, p. 114.

2.16.1.3 Determination of Parameters

The most important and most accurate method of determining the parameters is the *least squares method* (see 16.3.4.2, p. 841). In several cases, however, even simpler methods can be used with success, for instance the *mean value method*.

1. Mean Value Method

The *mean value method* uses the linear dependence of the "rectified" variables X and Y, i.e., $Y = AX + B$ as follows: The conditional equations $Y_i = AX_i + B$ for the given values Y_i, X_i are to be divided into two groups, which have the same size, or approximately the same size. By adding the equations in the groups one gets two equations, from which A and B can be determined. Then replacing X and Y by the original variables x and y again, one gets the connection between x and y, which is what one was looking for.
If not all the parameters are to be determined, one can apply the mean value method again with a rectification by other amounts $\overline{X}$ and $\overline{Y}$ (see, e.g., 2.16.2.11, p. 113).
Rectification and the mean value method are used above all when certain parameters occur in non-linear relations in an approximate formula, as for instance in (2.267b), (2.267c).

2. Least Squares Method

When certain parameters occur in non-linear relations in the approximation formula, the *least squares method* usually leads to a *non-linear fitting problem*. Their solution needs a lot of numerical calculations and also a good initial approximation. These approximations can be determined by the rectification and mean value method.

2.16.2 Useful Empirical Formulas

In this paragraph some of the simplest cases of empirical functional dependence are discussed, and the corresponding graphs are presented. Each figure shows several curves corresponding to different parameter values involved in the formula. The influence of the parameters upon the forms of the curves is discussed in the following sections.
For the choice of the appropriate function, usually only that part of the corresponding graph is to be considered, which is used for the reproduction of the empirical data. Therefore, e.g., one should not think that the formula $y = ax^2 + bx + c$ is suitable only in the case when the curve of the empirical data have a maximum or minimum.

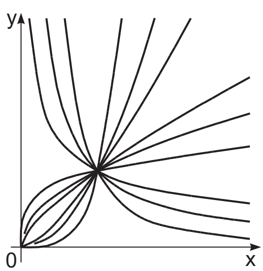

Figure 2.80

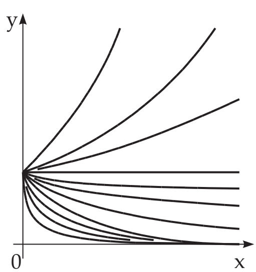

Figure 2.81

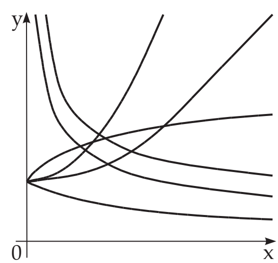

Figure 2.82

2.16.2.1 Power Functions

1. Type $y = ax^b$:

Typical shapes of curve for different values of the exponent b

$$y = ax^b \tag{2.255a}$$

are shown in **Fig. 2.80**. The curves for different values of the exponent are also represented in **Figs. 2.15, 2.21, 2.24, 2.25** and **Fig. 2.26**. The functions are discussed on pages 66, 70 and 71 for the formula (2.44) as a *parabola of order n*, formula (2.45) as a *reciprocal proportionality* and formula (2.50) as a *reciprocal power function*. The rectification is made by taking the logarithm

$$X = \log x, \quad Y = \log y\colon \quad Y = \log a + bX. \tag{2.255b}$$

2. Type $y = ax^b + c$:
The formula

$$y = ax^b + c \tag{2.256a}$$

produces the same curve as in (2.255a), but it is shifted by c in the direction of y **(Fig. 2.82)**. If b is given, the following rectification can be used:

$$X = x^b, \quad Y = y\colon \quad Y = aX + c. \tag{2.256b}$$

If b is not known first c is to be determined, then the rectification can be done in accordance with

$$X = \log x\,, \quad Y = \log(y - c)\colon \quad Y = \log a + bx\,. \tag{2.256c}$$

In order to determine a first approach of c, one can choose two arbitrary abscissae x_1, x_2 and a third one, $x_3 = \sqrt{x_1 x_2}$ as well as the corresponding ordinates y_1, y_2, y_3, so that now

$$c = \frac{y_1 y_2 - {y_3}^2}{y_1 + y_2 - 2y_3} \tag{2.256d}$$

holds. After having determined a and b, one can get a corrected c, namely as the average of the amounts $y - ax^b$.

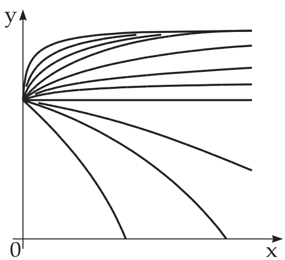

Figure 2.83

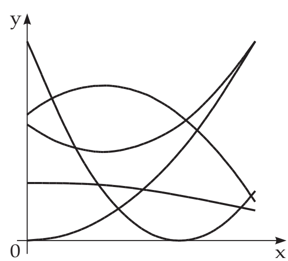

Figure 2.84

2.16.2.2 Exponential Functions

1. Type $y = ae^{bx}$:
The characteristic shapes of the curves of the function

$$y = ae^{bx} \tag{2.257a}$$

are shown in **Fig. 2.81**. The discussion of the *exponential function* (2.54) and its graph **(Fig. 2.26)** is presented in 2.6.1, p. 72. For the rectification one takes

$$X = x, \quad Y = \log y\colon \quad Y = \log a + b \log e \cdot X. \tag{2.257b}$$

2. Type $y = ae^{bx} + c$:
The formula

$$y = ae^{bx} + c \tag{2.258a}$$

produce the same curve as (2.257a), but it is shifted by c in the direction of y **(Fig. 2.83)**. Begin with the determination of a first approach c_1 for c then the rectification by logarithm:

$$Y = \log(y - c_1), \quad X = x\colon \quad Y = \log a + b \log e \cdot X. \tag{2.258b}$$

In order to determine c as for (2.256d) two arbitrary abscissae x_1, x_2 are to be chosen and $x_3 = \dfrac{x_1 + x_2}{2}$ as well as the corresponding ordinates y_1, y_2, y_3 to get $c = \dfrac{y_1 y_2 - {y_3}^2}{y_1 + y_2 - 2y_3}$. After having determined a and b one can get a corrected c, namely as the average of the amounts $y - ax^b$.

2.16.2.3 Quadratic Polynomial

Possible shapes of curves of the *quadratic polynomial*

$$y = ax^2 + bx + c \tag{2.259a}$$

are shown in **Fig. 2.84**. For the discussion of quadratic polynomials (2.41) and their curves **Fig. 2.12** (see 2.3.2, p. 64). Usually the coefficients a, b and c are determined by the least squares method; but in this case also rectification is possible. Choosing an arbitrary point of data (x_1, y_1) one rectifies

$$X = x, \quad Y = \frac{y - y_1}{x - x_1}: \quad Y = (b + ax_1) + aX. \tag{2.259b}$$

If the given x values form an arithmetical sequence with a difference h, one rectifies

$$Y = \Delta y, \quad X = x: \quad Y = (bh + ah^2) + 2ahX. \tag{2.259c}$$

In both cases after the determination of a and b from the equation

$$\sum y = a \sum x^2 + b \sum x + nc \tag{2.259d}$$

c is to be calculated; n is the number of the given x values, for which the sum is calculated.

2.16.2.4 Rational Linear Functions

The *rational linear function*

$$y = \frac{ax + b}{cx + d} \tag{2.260a}$$

is discussed in 2.4 with (2.46) and graphical representation **Fig. 2.17** (see p. 66).
Choosing an arbitrary data point (x_1, y_1) the rectification is done by the formulas

$$Y = \frac{x - x_1}{y - y_1}, \quad X = x: \quad Y = A + BX. \tag{2.260b}$$

After determining the values A and B the relation can be written in the form (2.260c). Sometimes instead of (2.260a) the form (2.260d)is sufficient:

$$y = y_1 + \frac{x - x_1}{A + Bx}, \quad (2.260c) \qquad\qquad y = \frac{x}{cx + d} \quad \text{or} \quad y = \frac{1}{cx + d}. \tag{2.260d}$$

Then in the first case the rectification can be made by $X = \dfrac{1}{x}$ and $Y = \dfrac{1}{y}$ or $X = x$ and $Y = \dfrac{x}{y}$ and by $X = x$ and $Y = \dfrac{1}{y}$ in the second case.

2.16.2.5 Square Root of a Quadratic Polynomial

Several possible shapes of curves of the equation

$$y^2 = ax^2 + bx + c \tag{2.261}$$

are shown in **Fig. 2.85**. The discussion of the function (2.52) and its graph **Fig. 2.23** (see p. 71).
If introducing the new variable $Y = y^2$, the problem can be reduced to the case of the quadratic polynomial in 2.16.2.3, p. 111.

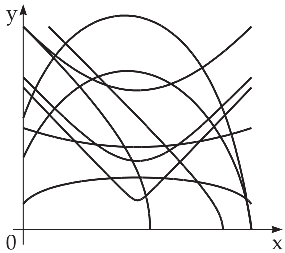

Figure 2.85

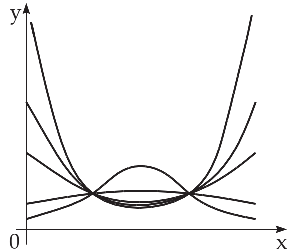

Figure 2.86

2.16.2.6 General Error Curve

The typical shapes of curves of the functions

$$y = ae^{bx+cx^2} \quad \text{or} \quad \log y = \log a + bx \log e + cx^2 \log e \tag{2.262}$$

are shown in **Fig. 2.86**. The discussion of the function with equation (2.61) and its graph **Fig. 2.31** (see p. 75).
Introducing the new variable $Y = \log y$, the problem is reduced to the case of the quadratic polynomial in 2.16.2.3, p. 111.

2.16.2.7 Curve of Order Three, Type II

The possible shapes of graphs of the function

$$y = \frac{1}{ax^2 + bx + c}. \tag{2.263}$$

are represented in **Fig. 2.87**. The discussion of the function with equation (2.48) and with graphs **Fig. 2.19** (see p. 67).
By introducing the new variable $Y = \dfrac{1}{y}$, the problem is reduced to the case of the quadratic polynomial in 2.16.2.3, p. 111.

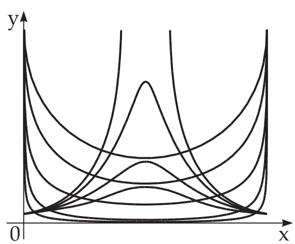

Figure 2.87

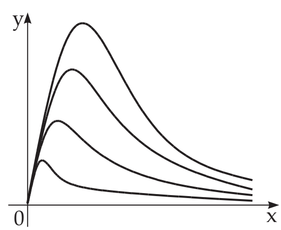

Figure 2.88

2.16.2.8 Curve of Order Three, Type III

Typical shapes of curves of functions of the type

$$y = \frac{x}{ax^2 + bx + c} \tag{2.264}$$

are represented in **Fig. 2.88**. The discussion of the function with equation (2.49) and with graphs **Fig. 2.20** (see p. 68).

Introducing the new variable $Y = \frac{x}{y}$ the problem can be reduced to the case of the quadratic polynomial in 2.16.2.3, p. 111.

2.16.2.9 Curve of Order Three, Type I

Typical shapes of curves of functions of the type

$$y = a + \frac{b}{x} + \frac{c}{x^2} \tag{2.265}$$

are represented in **Fig. 2.89**. The discussion of the function with equation (2.47) and with graphs **Fig. 2.18** (see p. 67).

Introducing the new variable $X = \frac{1}{x}$ the problem can be reduced to the case of the quadratic polynomial in 2.16.2.3, p. 111.

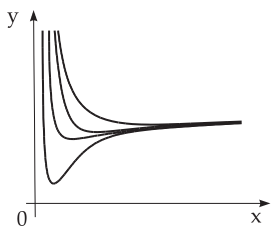

Figure 2.89

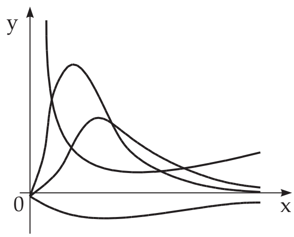

Figure 2.90

2.16.2.10 Product of Power and Exponential Functions

Typical shapes of curves of functions of the type

$$y = ax^b e^{cx} \tag{2.266a}$$

are represented in **Fig. 2.90**. The discussion of the function with equation (2.62) and with graphs **Fig. 2.31** (see p. 75).

If the empirical values of x form an arithmetical sequence with difference h, the rectification follows in accordance with

$$Y = \Delta \log y, \quad X = \Delta \log x: \quad Y = hc \log e + bX. \tag{2.266b}$$

Here $\Delta \log y$ and $\Delta \log x$ denote the difference of two subsequent values of $\log y$ and $\log x$ respectively. If the x values form a geometric sequence with quotient q, then the rectification follows by

$$X = x, \quad Y = \Delta \log y: \quad Y = b \log q + c(q-1)X \log e. \tag{2.266c}$$

After b and c are determined the logarithm of the given equation is taken, and the value of $\log a$ is calculated like in (2.259d).

If the given x values do not form a geometric sequence, but one can choose pairs of two values of x such that their quotient q is the same constant, then the rectification is the same as in the case of a geometric sequence of x values with the substitution $Y = \Delta_1 \log y$. Here $\Delta_1 \log y$ denotes the difference of the two values of $\log y$ whose corresponding x values result in the constant quotient q (see 2.16.2.12, p. 114).

2.16.2.11 Exponential Sum

Typical shapes of curves of the *exponential sum*

$$y = ae^{bx} + ce^{dx} \tag{2.267a}$$

are represented in **Fig. 2.91**. The discussion of the function with equation (2.60) and with graphs **Fig. 2.29** (see p. 74).
If the values of x form an arithmetical sequence with difference h, and y, y_1, y_2 are any three consecutive values of the given function, then the rectification is made by

$$Y = \frac{y_2}{y}, \quad X = \frac{y_1}{y}: \quad Y = (e^{bh} + e^{dh})X - e^{bh} \cdot e^{dh}. \tag{2.267b}$$

After b and d are determined, follows again a rectification by

$$\overline{Y} = ye^{-dx}, \quad \overline{X} = e^{(b-d)x}: \quad \overline{Y} = a\overline{X} + c. \tag{2.267c}$$

2.16.2.12 Numerical Example

Find an empirical formula to describe the relation between x and y for given values in **Table 2.9**.
Choice of the Approximation Function: Comparing the graph prepared from the given data (**Fig. 2.92**) with the curves discussed before, one can see that formulas (2.264) or (2.266a) with curves in **Fig. 2.88** and **Fig. 2.90** can fit the considered case.
Determination of Parameters: Using the formula (2.264) to rectify are $\Delta\frac{x}{y}$ and x. The calculation shows, however, the relationship between x and $\Delta\frac{x}{y}$ is far from linear. To verify whether the formula (2.266a) is suitable one plots the graph of the relation between $\Delta \log x$ and $\Delta \log y$ for $h = 0,1$ in **Fig. 2.93**, and also between $\Delta_1 \log y$ and x for $q = 2$ in **Fig. 2.94**. In both cases the points fit a straight line well enough, so the formula $y = ax^b e^{cx}$ can be used.

Table 2.9 For the approximate determination of an empirically given function relation

x	y	$\frac{x}{y}$	$\Delta\frac{x}{y}$	$\lg x$	$\lg y$	$\Delta \lg x$	$\Delta \lg y$	$\Delta_1 \lg y$	y_{err}
0.1	1.78	0.056	0.007	−1.000	0.250	0.301	0.252	0.252	1.78
0.2	3.18	0.063	0.031	−0.699	0.502	0.176	+0.002	−0.097	3.15
0.3	3.19	0.094	0.063	−0.523	0.504	0.125	−0.099	−0,447	3.16
0.4	2.54	0.157	0.125	−0.398	0.405	0.097	−0.157	−0.803	2.52
0.5	1.77	0.282	0.244	−0.301	0.248	0.079	−0.191	−1.134	1.76
0.6	1.14	0.526	0.488	−0.222	0.057	0.067	−0.218	−1.455	1.14
0.7	0.69	1.014	0.986	−0.155	−0.161	0.058	−0.237	–	0.70
0.8	0.40	2.000	1.913	−0.097	−0.398	0.051	−0.240	–	0.41
0.9	0.23	3.913	3.78	−0.046	−0.638	0.046	−0.248	–	0.23
1.0	0.13	7.69	8.02	0.000	−0.886	0.041	−0.269	–	0.13
1.1	0.07	15.71	14.29	0.041	−1.155	0.038	−0.243	–	0.07
1.2	0.04	30.0	–	0.079	−1.398	–	–	–	0.04

In order to determine the constants a, b and c, a linear relation between x and $\Delta_1 \log y$ is to be searched by the method of mean values. Adding the conditional equations $\Delta_1 \log y = b \log 2 + cx \log e$ in groups of three equations each, yields

$$-0.292 = 0.903b + 0.2606c\,, \quad -3.392 = 0.903b + 0.6514c,$$

and from here $b = 1.966$ and $c = -7.932$ holds. To determine a, the equations of the form $\log y = \log a + b \log x + c \log e \cdot x$ are to be added, which yields $-2.670 = 12 \log a - 6.529 - 26.87$, so from $\log a = 2.561$, $a = 364$ follows. The values of y calculated from the formula $y = 364x^{1{,}966}e^{-7.032x}$ are given in the last column of **Table 2.9**; they are denoted by y_{err} as an approximation of y. The error sum of squares is 0.0024.
Using the parameters determined by rectification as initial values for the iterative solution of the non-

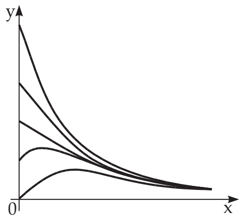

Figure 2.91

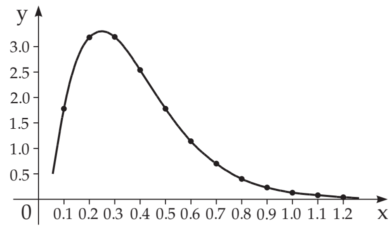

Figure 2.92

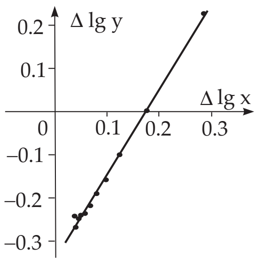

Figure 2.93

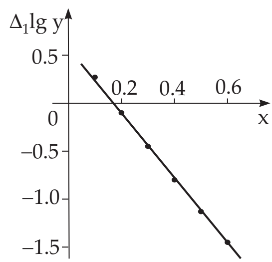

Figure 2.94

linear least squares problem (see 19.6.2.4, p. 987)

$$\sum_{i=1}^{12}[y_i - ax_i^b e^{cx_i}]^2 = \text{min!}$$

yields $a = 396.601\,986$, $b = 1.998\,098$, $c = -8.000\,0916$ with the small error sum of squares $0.000\,0916$.

2.17 Scales and Graph Paper

2.17.1 Scales

The base of a scale is a function $y = f(x)$. The task is to construct a *scale* from this function so that on a curve, e.g. on a line, the function values of y are to be inserted as the values of the argument x. A scale can be considered as a one-dimensional representation of a table of values.
The *scale equation* for the function $y = f(x)$ is:

$$y = l[f(x) - f(x_0)]. \tag{2.268}$$

The starting point of the scale is fixed at x_0. The *scale factor* l takes into consideration that for a concrete scale there it is only one given scale length.

■ **A Logarithmic Scale:** For $l = 10$ cm and $x_0 = 1$ the scale equation is
$y = 10[\lg x - \lg 1] = 10 \lg x$ (in cm). For the table of values

x	1	2	3	4	5	6	7	8	9	10
$y = \lg x$	0	0.30	0.48	0.60	0.70	0.78	0.85	0.90	0.95	1

one gets the scale shown in **Fig. 2.95**.

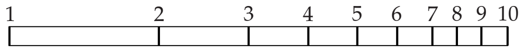

Figure 2.95

■ **B Slide Rule:** The most important application of the logarithmic scale, from a historical viewpoint, was the *slide rule*. Here, for instance, multiplication and division were performed with the help of two identically calibrated logarithmic scales, which can be shifted along each other.
From **Fig. 2.96** one can read: $y_3 = y_1 + y_2$, i.e., $\lg x_3 = \lg x_1 + \lg x_2 = \lg x_1 x_2$, hence $x_3 = x_1 \cdot x_2$; $y_1 = y_3 - y_2$, i.e., $\lg x_1 = \lg x_3 - \lg x_2 = \lg \dfrac{x_3}{x_2}$, so $x_1 = \dfrac{x_3}{x_2}$.

■ **C Volume Scale** on the lateral surface of a conical shaped funnel: A scale is to be marked on the funnel, so that the volume inside could be read from it. The data of the funnel are: Height $H = 15$ cm, upper diameter $D = 10$ cm.

Taking in mind **Fig. 2.97a** gives the scale equation as follows: Volume $V = \dfrac{1}{3} r^2 \pi h$, apothem $s = \sqrt{h^2 + r^2}$, $\tan\alpha = \dfrac{r}{h} = \dfrac{D/2}{H} = \dfrac{1}{3}$. From these $h = 3r$, $s = r\sqrt{10}$, $V = \dfrac{\pi}{(\sqrt{10})^3}$ follows, so the scale equation is $s = \dfrac{\sqrt{10}}{\sqrt[3]{\pi}} \sqrt[3]{V} \approx 2.16 \sqrt[3]{V}$. The following table of values contains the calibration marks on the funnel as in the figure:

V	0	50	100	150	200	250	300	350
s	0	7.96	10.03	11.48	12.63	13.61	14.46	15.22

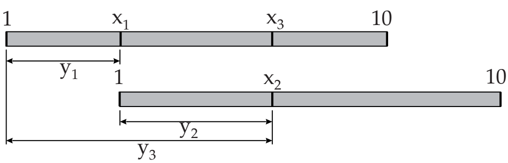

Figure 2.96

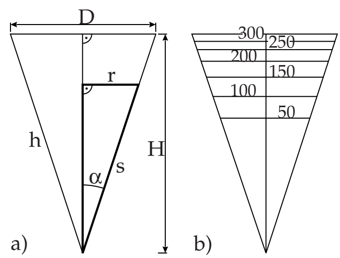

Figure 2.97

2.17.2 Graph Paper

The most useful graph paper is prepared so that the axes of a right-angle coordinate system are calibrated by the scale equations

$$x = l_1[g(u) - g(u_0)], \quad y = l_2[f(v) - f(v_0)]. \tag{2.269}$$

Here l_1 and l_2 are the scale factors; u_0 and v_0 are the initial points of the scale.

2.17.2.1 Semilogarithmic Paper

If the x-axis has an equidistant subdivision, and the y-axis has a logarithmic one, then one talks about *semilogarithmic paper* or about a *semilogarithmic coordinate system*.

Scale Equations:

$$x = l_1[u - u_0] \quad \text{(linear scale)}, \qquad y = l_2[\lg v - \lg v_0] \quad \text{(logarithmic scale)}. \tag{2.270}$$

The **Fig. 2.98** shows an example of semilogarithmic paper.

Representation of Exponential Functions: On semilogarithmic paper the graph of the exponential function

$$y = \alpha e^{\beta x} \quad (\alpha, \beta \text{ const}) \tag{2.271a}$$

is a straight line (see rectification in 2.16.2.2, p. 110). This property can be used in the following way: If the measuring points, introduced on semilogarithmic paper, lie approximately on a line, it can be supposed a relation between the variables as in (2.271a). With this line, estimated by eye, one can determine the approximate values of α and β: Considering two points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ from this line one gets

$$\beta = \frac{\ln y_2 - \ln y_1}{x_2 - x_1} \quad \text{and, e.g.,} \quad \alpha = y_1 e^{\beta x_1}. \tag{2.271b}$$

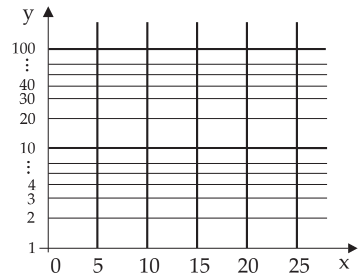

Figure 2.98

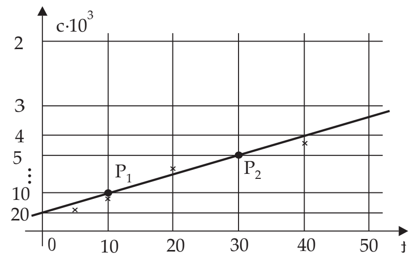

Figure 2.99

2.17.2.2 Double Logarithmic Paper

If both axes of a right-angle x, y coordinate system are calibrated with respect to the logarithm function, then one talks about *double logarithmic paper* or *log–log paper* or a *double logarithmic coordinate system*.

Scale Equations: The scale equations are

$$x = l_1[\lg u - \lg u_0], \qquad y = l_2[\lg v - \lg v_0], \tag{2.272}$$

where l_1, l_2 are the scale factors and u_0, v_0 are the initial points.

Representation of Power Functions (see 2.5.3, p. 71): Log–log paper has a similar arrangement to semilogarithmic paper, but the x-axis also has a logarithmic subdivision. In this coordinate system the graph of the power function

$$y = \alpha x^{\beta} \quad (\alpha, \beta \text{ const}) \tag{2.273}$$

is a straight line (see rectification of a power function in 2.16.2.1, p. 109). This property can be used in the same way as in the case of semilogarithmic paper.

2.17.2.3 Graph Paper with a Reciprocal Scale

The subdivisions of the scales on the coordinate axes follow from (2.45) for the function of inverse proportionality (see 2.4.1, p. 66).

Scale Equations: $x = l_1[u - u_0], \quad y = l_2\left[\frac{a}{v} - \frac{a}{v_0}\right] \quad (a \text{ const}),$ (2.274)

where l_1 and l_2 are the scale factors, and u_0, v_0 are the starting points.

■ **Concentration in a Chemical Reaction:** For a chemical reaction, the concentration denoted by $c = c(t)$, has been measured as a function of the time t, giving the following results for c:

t/min	5	10	20	40
$c \cdot 10^3/\text{mol/l}$	15.53	11.26	7.27	4.25

It can be supposed that the reaction is of second order, i.e., the relation should be

$$c(t) = \frac{c_0}{1 + c_0 kt} \quad (c_0, k \text{ const}). \tag{2.275}$$

Taking the reciprocal value of both sides, one gets $\frac{1}{c} = \frac{1}{c_0} + kt$, i.e., (2.275) can be represented as a line, if the corresponding graph paper has a reciprocal subdivision on the y-axis and a linear one on the x-axis. The scale equation for the y-axis is, e.g., $y = 10 \cdot \frac{1}{v}$ cm.

It is obvious from the corresponding **Fig. 2.99** that the measuring points lie approximately on a line, i.e., the supposed relation (2.275) is acceptable.
From these points the approximate values of both parameters k (reaction rate) and c_0 (initial concentration) can be determined. Choosing two points, e.g., $P_1(10, 10)$ and $P_2(30, 5)$, one gets:

$$k = \frac{1/c_1 - 1/c_2}{t_2 - t_1} \approx 0.005, \quad c_0 \approx 20 \cdot 10^{-3}.$$

2.17.2.4 Remark

There are several other possibilities for constructing and using graph papers. Although today in most cases there are high-capacity computers to analyze empirical data and measurement results, in everyday laboratory practice, when getting only a few data, graph papers are used quite often to show the functional relations and approximate parameter values needed as initial data for applied numerical methods (see the non-linear least squares method in 19.6.2.4, p. 987).

2.18 Functions of Several Variables

2.18.1 Definition and Representation

2.18.1.1 Representation of Functions of Several Variables

A variable value u is called a function of n *independent variables* $x_1, x_2, \ldots, x_n$, if for given values of the independent variables, u is a uniquely defined value. Depending on how many variables there are, two, three, or n, one writes

$$u = f(x, y), \quad u = f(x, y, z), \quad u = f(x_1, x_2, \ldots, x_n). \tag{2.276}$$

Substituting given numbers for the n independent variables yields a value system of the variables, which can be considered as a *point of the n-dimensional space*. The single independent variables are called arguments; sometimes the entire n *tuple* together is called the argument of the function.

Examples of Values of Functions:

■ **A:** $u = f(x, y) = xy^2$ has for the value system $x = 2$, $y = 3$ the value $f(2, 3) = 2 \cdot 3^2 = 18$.
■ **B:** $u = f(x, y, z, t) = x \ln(y - zt)$ takes for the value system $x = 3$, $y = 4$, $z = 3$, $t = 1$ the value $f(3, 4, 3, 1) = 3 \ln(4 - 3 \cdot 1) = 0$.

2.18.1.2 Geometric Representation of Functions of Several Variables

1. Representation of the Value System of the Variables

The value system of an argument of two variables x and y can be represented as a point of the plane given by Cartesian coordinates x and y. A value system of three variables x, y, z corresponds to a point given by the coordinates x, y, z in a three-dimensional Cartesian coordinate system. Systems of four or more coordinates cannot be represented obviously in our three-dimensional imagination.

Similarly to the three-dimensional case the system of n variables $x_1, x_2, \ldots, x_n$ is to be considered as a point of the n-dimensional space given by Cartesian coordinates x_1, x_2,. . . ,x_n. In the above example **B**, the four variables define a point in four-dimensional space, with coordinates $x = 3$, $y = 4$, $z = 3$ and $t = 1$.

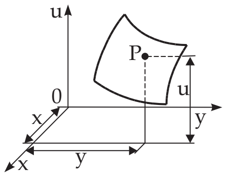

Figure 2.100

2. Representation of the Function $u = f(x, y)$ of Two Variables

a) A function of two independent variables can be represented by a surface in three-dimensional space, similarly to the graph representation of functions of one variable (**Fig. 2.100**, see also 3.6.3, p. 261). Considering the values of the independent variables of the domain as the first two coordinates, and the value of the function $u = f(x, y)$ as the third coordinate of a point in a Cartesian coordinate system, these points form a surface in three-dimensional space.

Examples of Surfaces of Functions:

■ **A:** $u = 1 - \dfrac{x}{2} - \dfrac{y}{3}$ represents a plane (**Fig. 2.101a**, see also 3.5.3.10, p. 218).

■ **B:** $u = \dfrac{x^2}{2} + \dfrac{y^2}{4}$ represents an elliptic paraboloid (**Fig. 2.101b**, see also 3.5.3.13, **5.**, p. 226).

■ **C:** $u = \sqrt{16 - x^2 - y^2}$ represents a hemisphere with $r = 4$ **(Fig. 2.101c)**.

b) The shape of the surface of the function $u = f(x, y)$ can be pictured with the help of intersection curves, which can be get by intersecting the surface parallel to the coordinate planes. The intersection curves $u =$ const are called *level curves*.

■ In **Fig. 2.101b,c** the level curves are ellipses or concentric circles (not denoted in the figure).

Remark: A function with an argument of three or more variables cannot be represented in three-dimensional space. Similarly to surfaces in three-dimensional space also the notion of a *hyper surface* in n-dimensional space is in use.

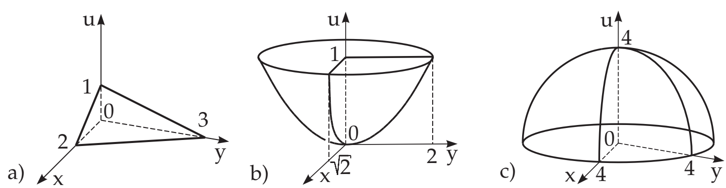

Figure 2.101

2.18.2 Different Domains in the Plane

2.18.2.1 Domain of a Function

The *domain of definition of a function* (or domain of a function) is the *set of the system of values* or *points* which can be taken by the variables of the argument of the function. The domains defined in this way can be very different. Mostly they are bounded or unbounded connected *sets of points*. Depending on whether the boundary belongs to the domain or not, the domain is closed or open. An open, connected set of points is called a *domain*. If the boundary belongs to the domain, it is called a *closed domain*, if it does not, sometimes it is called an *open domain*.

2.18.2.2 Two-Dimensional Domains

Fig. 2.102 shows the simplest cases of connected sets of points of two variables and their notation. Domains are represented here as the shaded part; closed domains, i.e., domains whose boundary belongs to them, are bounded by thick curves in the figures; open domains are bounded by dotted curves. Including the entire plane there are only *simply connected domains* or *simply connected regions* in **Fig. 2.102**.

2.18.2.3 Three or Multidimensional Domains

These are handled similarly to the two-dimensional case. It concerns also the distinction between simply and multiply connected domains. Functions of more than three variables will be geometrically represented in the corresponding n-dimensional space.

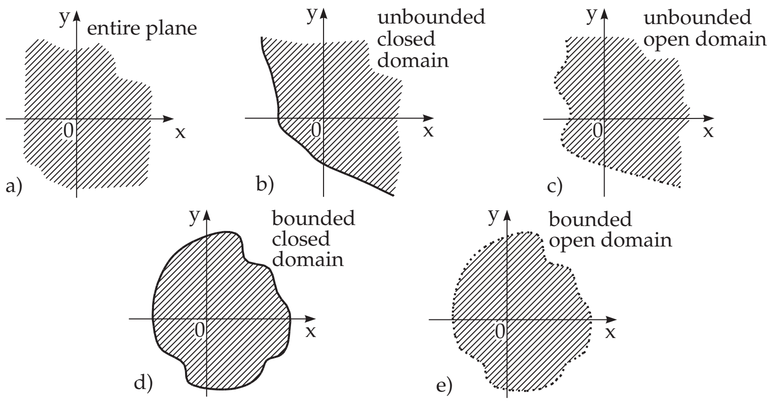

Figure 2.102

2.18.2.4 Methods to Determine a Function

1. Definition by Table of Values Functions of several variables can be defined by a table of values. An example of functions of two independent variables are the tables of values of elliptic integrals (see 21.9, p. 1103). The values of the independent variables are denoted on the top and on the left-hand side of the table. The required value of the function is in the intersection of the corresponding row and column. It is called a *table with double entry*.

2. Definition by Formulas Functions of several variables can be defined by one or more formulas.

■ **A:** $u = xy^2$.
■ **B:** $u = x \ln(y - zt)$.
■ **C:** $u = \begin{cases} x + y \ for \ x \geq 0, \ y \geq 0, \\ x - y \ for \ x \geq 0, \ y < 0, \\ -x + y \ for \ x < 0, \ y \geq 0, \\ -x - y \ for \ x < 0, \ y < 0. \end{cases}$

3. Domain of a Function Given by One Formula In the analysis mostly such functions are considered which are defined by formulas. Here the union of all value systems for which the analytical expression has a meaning is considered to be the domain, i.e., for which the expression has a unique, finite, real value.

Examples for Domains:

■ **A:** $u = x^2 + y^2$: The domain is the entire plane.

■ **B:** $u = \dfrac{1}{\sqrt{16 - x^2 - y^2}}$: The domain consists of all value systems x, y, satisfying the inequality $x^2 + y^2 < 16$. Geometrically this domain is the interior of the circle in **Fig. 2.103a**, an open domain.

■ **C:** $u = \arcsin(x + y)$: The domain consists of all value systems x, y, satisfying the inequality $-1 \leq$

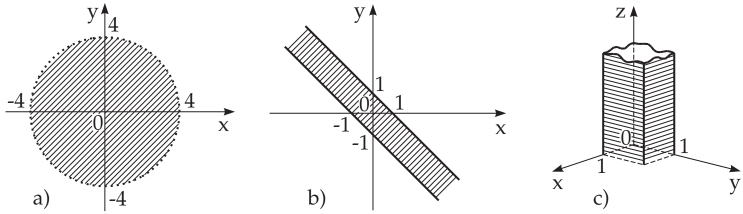

Figure 2.103

$x + y \leq +1$, i.e., the domain of the function is a closed domain, the stripe between the two parallel lines in **Fig. 2.103b**.

■ **D:** $u = \arcsin(2x-1) + \sqrt{1-y^2} + \sqrt{y} + \ln z$: The domain consists of the value system x, y, z, satisfying the inequalities $0 \leq x \leq 1$, $0 \leq y \leq 1$, $z > 0$, i.e., it consists of all points of the three dimensional x, y, z space lying above a square with side-length 1 shown in **Fig. 2.103c**.

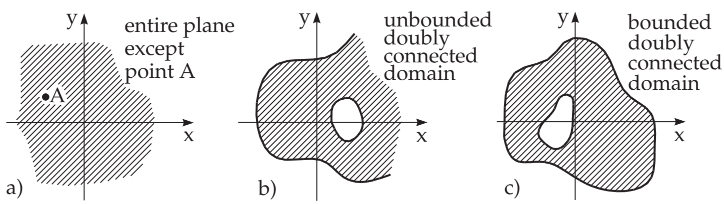

Figure 2.104

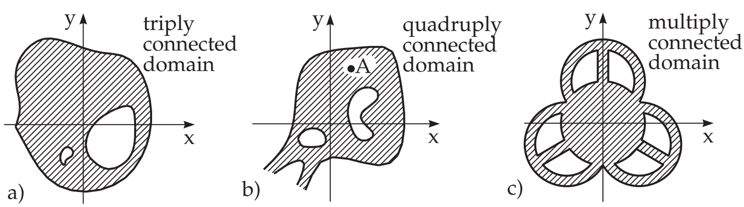

Figure 2.105

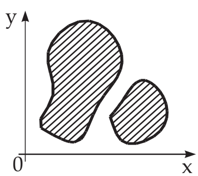

Figure 2.106

If from the interior of the considered part of the plane a point or a bounded, simply connected point set is missing, as shown in **Fig. 2.104**, it is called a *doubly-connected domain* or *doubly-connected region. Multiply connected domains* are represented in **Fig. 2.105**. A *non-connected region* is shown in **Fig. 2.106**.

2.18.2.5 Various Forms for the Analytical Representation of a Function

Functions of several variables can be defined in different ways, just as functions of one variable.

1. Explicit Representation

A function is given or defined in an explicit way if its value (the dependent variable) can be expressed by the independent variables:

$$u = f(x_1, x_2, \ldots, x_n)\,. \tag{2.277}$$

2. Implicit Representation

A function is given or defined in an implicit way if the relation between its value and the independent variables is given in the form:

$$F(x_1, x_2, \ldots, x_n, u) = 0\,, \tag{2.278}$$

if there is a unique value of u satisfying this equality.

3. Parametric Representation

A function is given in parametric form if the n arguments and the function are defined by n new variables, the parameters, in an explicit way, supposing there is a one-to-one correspondence between the parameters and the arguments. For a two-variable function, for instance

$$x = \varphi(r,s)\,, \quad y = \psi(r,s)\,, \quad u = \chi(r,s)\,, \tag{2.279a}$$

and for a three-variable function

$$x = \varphi(r,s,t)\,, \quad y = \psi(r,s,t)\,, \quad z = \chi(r,s,t)\,, \quad u = \kappa(r,s,t) \tag{2.279b}$$

etc.

4. Homogeneous Functions

A function $f(x_1, x_2, \ldots, x_n)$ of several variables is called a homogeneous function if the relation

$$f(\lambda x_1, \lambda x_2, \ldots, \lambda x_n) = \lambda^m f(x_1, x_2, \ldots, x_n) \tag{2.280}$$

holds for arbitrary λ. The number m is the *degree of homogeneity*.

■ **A:** For $u(x,y) = x^2 - 3xy + y^2 + x\sqrt{xy + \dfrac{x^3}{y}}$, the degree of homogeneity is $m = 2$.

■ **B:** For $u(x,y) = \dfrac{x+z}{2x-3y}$, the degree of homogeneity is $m = 0$.

2.18.2.6 Dependence of Functions

1. Special Case of Two Functions

Two functions of two variables $u = f(x,y)$ and $v = \varphi(x,y)$, with the same domain G, are called *dependent functions* if one of them can be expressed as a function of the other one $u = F(v)$. For every point of the domain G of the functions the identity

$$f(x,y) = F(\varphi(x,y)) \quad \text{or} \quad \Phi(f,\varphi) = 0 \tag{2.281}$$

holds. If there is no such function $F(\varphi)$ or $\Phi(f,\varphi)$, one speaks about *independent functions*.

■ $u(x,y) = (x^2+y^2)^2$, $v = \sqrt{x^2+y^2}$ are defined in the domain $x^2 + y^2 \geq 0$, i.e., in the whole x,y plain, and they are dependent, because $u = v^4$ holds.

2. General Case of Several Functions

Similarly to the case of two functions, the m functions $u_1, u_2, \ldots, u_m$ of n variables $x_1, x_2, \ldots, x_n$ in their common domain G are called dependent if one of them can be expressed as a function of the others, i.e., if for every point of the domain G the identity

$$u_i = f(u_1, u_2, \ldots, u_{i-1}, u_{i+1}, \ldots, u_m) \quad \text{or} \quad \Phi(u_1, u_2, \ldots, u_m) = 0 \tag{2.282}$$

is valid. If there is no such functional relationship, they are independent functions.

■ The functions $u = x_1 + x_2 + \cdots + x_n$, $v = {x_1}^2 + {x_2}^2 + \cdots + {x_n}^2$ and $w = x_1x_2 + x_1x_3 + \cdots + x_1x_n + x_2x_3 + \cdots + x_{n-1}x_n$ are dependent because $v = u^2 - 2w$ holds.

3. Analytical Conditions for Independence

Suppose every partial derivative mentioned below exists. Two functions $u = f(x, y)$ and $v = \varphi(x, y)$ are independent on a domain if their *functional determinant* or *Jacobian determinant* is not identically zero here. Analogously, in the case of n functions of n variables $u_1 = f_1(x_1, \ldots, x_n), \ldots, u_n = f_n(x_1, \ldots, x_n)$:

$$\begin{vmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial \varphi}{\partial x} & \frac{\partial \varphi}{\partial y} \end{vmatrix}, \text{ short } \frac{D(f, \varphi)}{D(x, y)} \text{ or } \frac{D(u, v)}{D(x, y)}, \qquad (2.283a)$$

$$\begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{vmatrix} \equiv \frac{D(f_1, f_2, \ldots, f_n)}{D(x_1, x_2, \ldots, x_n)} \neq 0 \,. \qquad (2.283b)$$

If the number m of the functions $u_1, u_2, \ldots, u_m$ is smaller than the number of variables $x_1, x_2, \ldots, x_n$, these functions are independent if at least one subdeterminant of order m of the matrix (2.283c) is not identically zero:

$$\begin{pmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} & \cdots & \frac{\partial u_1}{\partial x_n} \\ \frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} & \cdots & \frac{\partial u_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial u_m}{\partial x_1} & \frac{\partial u_m}{\partial x_2} & \cdots & \frac{\partial u_m}{\partial x_n} \end{pmatrix}. \qquad (2.283c)$$

The number of independent functions is equal to the rank r of the matrix (2.283c) (see 4.1.4, **7.**, p. 274). Here these functions are independent, whose derivatives are the elements of the non-vanishing determinant of order r. If $m > n$ holds, then among the given m functions at most n can be independent.

2.18.3 Limits

2.18.3.1 Definition

A function of two variables $u = f(x, y)$ has a limit A at $x = a$, $y = b$ if when x and y are arbitrarily close to a and b, respectively, then the value of the function $f(x, y)$ approaches arbitrarily closely the value A. Then one writes:

$$\lim_{\substack{x \to a \\ y \to b}} f(x, y) = A \,. \qquad (2.284)$$

The function may not be defined at (a, b), or if it is defined here, may not have the value A.

2.18.3.2 Exact Definition

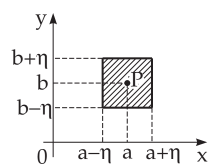

Figure 2.107

A function of two variables $u = f(x, y)$ has at the point (a, b) a limit

$$A = \lim_{\substack{x \to a \\ y \to b}} f(x, y) \qquad (2.285a)$$

if for arbitrary positive ε there is a positive η such that (see **Fig. 2.107**)

$$|f(x, y) - A| < \varepsilon \qquad (2.285b)$$

holds for every point (x, y) of the square

$$|x - a| < \eta, \quad |y - b| < \eta \,. \qquad (2.285c)$$

2.18.3.3 Generalization for Several Variables

a) The notion of limit of a function of several variables can be defined analogously to the case of two variables.

b) The criteria for the existence of a limit of a function of several variables can be obtained by generalization of the criterion for functions of one variable, i.e., by reducing to the limit of a sequence or by applying the Cauchy condition for convergence (see 2.1.4.3, p. 53).

2.18.3.4 Iterated Limit

If for a function of two variables $f(x,y)$ first the limit for $x \to a$ has been determined, i.e., $\lim\limits_{x\to a} f(x,y)$ for constant y, then for the function obtained, which is now a function only of y, one determines the limit for $y \to b$, then the resulting number

$$B = \lim_{y\to b}\left(\lim_{x\to a} f(x,y)\right) \tag{2.286a}$$

is called an *iterated limit*. Changing the order of calculations generally yields an other limit:

$$C = \lim_{x\to a}\left(\lim_{y\to b} f(x,y)\right). \tag{2.286b}$$

In general $B \neq C$ holds, even if both limits exist.

■ For the function $f(x,y) = \dfrac{x^2 - y^2 + x^3 + y^3}{x^2 + y^2}$ for $x \to 0$, $y \to 0$ one gets the iterated limits $B = -1$ and $C = +1$.

Remark: If the function $f(x,y)$ has a limit $A = \lim\limits_{\substack{x\to a\\ y\to b}} f(x,y)$, and both B and C exist, then $B = C = A$ is valid. The existence of B and C does not follow from the existence of A. From the equality of the limits $B = C$ the existence of the limit A does not follow.

2.18.4 Continuity

A function of two variables $f(x,y)$ is continuous at $x = a$, $y = b$, i.e., at the point (a,b), if 1. the point (a,b) belongs to the domain of the function and 2. the limit for $x \to a$, $y \to b$ exists and is equal to the value, i.e.,

$$\lim_{\substack{x\to a\\ y\to b}} f(x,y) = f(a,b). \tag{2.287}$$

Otherwise the function has a discontinuity at $x = a$, $y = b$. If a function is defined and continuous at every point of a connected domain, it is called continuous on this domain. Similarly can be defined the continuity of functions of more than two variables.

2.18.5 Properties of Continuous Functions

2.18.5.1 Theorem on Zeros of Bolzano

If a function $f(x,y)$ is defined and continuous in a connected domain, and at two points (x_1,y_1) and (x_2,y_2) of this domain the values have different signs, then there exists at least one point (x_3,y_3) in this domain such that $f(x,y)$ is equal to zero there:

$$f(x_3,y_3) = 0, \quad \text{if} \quad f(x_1,y_1) > 0 \quad \text{and} \quad f(x_2,y_2) < 0\,. \tag{2.288}$$

2.18.5.2 Intermediate Value Theorem

If a function $f(x,y)$ is defined and continuous in a connected domain, and at two points (x_1,y_1) and (x_2,y_2) it has different values $A = f(x_1,y_1)$ and $B = f(x_2,y_2)$, then for an arbitrary value C between A and B there is at least one point (x_3,y_3) such that:

$$f(x_3,y_3) = C, \quad A < C < B \quad \text{or} \quad B < C < A\,. \tag{2.289}$$

2.18.5.3 Theorem About the Boundedness of a Function

If a function $f(x,y)$ is continuous on a bounded and closed domain, it is bounded in this domain, i.e., there are two numbers m and M such that for every point (x,y) in this domain:

$$m \le f(x,y) \le M. \tag{2.290}$$

2.18.5.4 Weierstrass Theorem (About the Existence of Maximum and Minimum)

If a function $f(x,y)$ is continuous on a bounded and closed domain, then it takes its maximum and minimum here, i.e., there is at least one point (x',y') such that all the values $f(x,y)$ in this domain are less than or equal to the value $f(x',y')$, and there is at least one point (x'',y'') such that all the values $f(x,y)$ in this domain are greater than or equal to $f(x'',y'')$: For any point (x,y) of this domain

$$f(x',y') \geq f(x,y) \geq f(x'',y'') \tag{2.291}$$

is valid.

2.19 Nomography

2.19.1 Nomograms

Nomograms are graphical representations of a functional correspondence between three or more variables. From the nomogram, the corresponding values of the variables of a given formula – the *key formula* – in a given domain of the variables can be immediately read directly. Important examples of nomograms are *net charts* and *alignment charts*.
Nomograms are still used in laboratories, even in the computer age, for instance to get approximate values or starting guesses for iterations.

2.19.2 Net Charts

To represent a correspondence between the variables x, y, z given by the equation

$$F(x,y,z) = 0 \tag{2.292}$$

(or in many cases explicitly by $z = f(x,y)$), the variables can be considered as coordinates in space. The equation (2.292) defines a surface which can be visualized on two-dimensional paper by its level curves (see 2.18.1.2, p. 118). Here, a family of curves is assigned to each variable. These curves form a net: The variables x and y are represented by lines parallel to the axis, the variable z is represented by the family of level curves.

■ Ohm's law is $U = R \cdot I$. The voltage U can be represented by its level curves depending on two variables. If R and I are chosen as Cartesian coordinates, then the equation $U =$ const for every constant corresponds to a hyperbola **(Fig. 2.108)**. By looking at the figure one can tell the corresponding value of U for every pair of values R and I, and also I corresponding to every R, U, and also R corresponding to every I and U. Of course, the investigation is always to be restricted to the domain which is interpreted: In **Fig. 2.108** holds $0 < R < 10$, $0 < I < 10$ and $0 < U < 100$.

Remarks:

1. By changing the calibration, the nomogram can be used for other domains. If, e.g., in **(Fig. 2.108)** the domain $0 < I < 1$ should be represented but R should remain the same, then the hyperbolas of U are to be marked by $U/10$.

2. By application of scales (see 2.17.1, p. 115) it is possible to transform nomograms with complicated curves into straight-line nomograms. Using uniform scales on the x and y axis, every equation of the form

$$x\varphi(z) + y\psi(z) + \chi(z) = 0 \tag{2.293}$$

can be represented by a nomogram consisting of straight lines. If function scales $x = f(z_2)$ and $y = g(z_2)$ are used, the equation of the form

$$f(z_2)\varphi(z_1) + g(z_2)\psi(z_1) + \chi(z_1) = 0 \tag{2.294}$$

has a representation for the variables z_1, z_2 and z_3 as two families of curves parallel to the axis and an arbitrary family of straight lines.

■ By applying a logarithmic scale (see 2.17.1, p. 115), Ohm's law can be represented by a straight-line nomogram. Taking the logarithm of $R \cdot I = U$ gives $\log R + \log I = \log U$. Substituting $x = \log R$

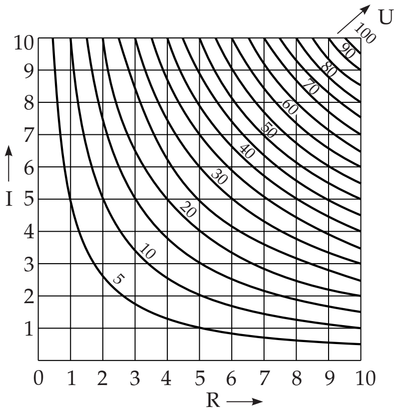

Figure 2.108

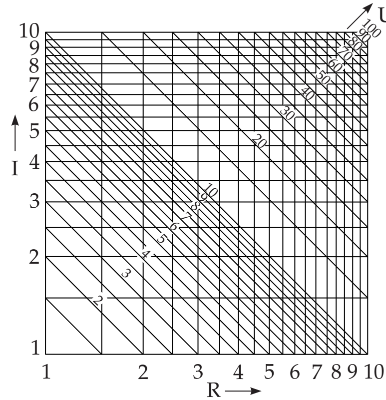

Figure 2.109

and $y = \log I$ results in $x + y = \log U$, i.e., a special form of (2.294). The corresponding nomogram is shown in **Fig. 2.109**.

2.19.3 Alignment Charts

A graphical representation of a relation between three variables z_1, z_2 and z_3 can be given by assigning a scale (see 2.17.1, p. 115) to each variable. The z_i scale has the equation

$$x_i = \varphi_i(z_i), y_i = \psi_i(z_i) \quad (i = 1, 2, 3). \tag{2.295}$$

The functions φ_i and ψ_i are chosen in such a manner that the values of the three variables z_1, z_2 and z_3 satisfying the nomogram equation should lie on a straight line. To satisfy this condition, the area of the triangle, given by the points $(x_1, y_1), (x_2, y_2)$ and (x_3, y_3) , must be zero (see (3.301), p. 195), i.e.,

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = \begin{vmatrix} \varphi_1(z_1) & \psi_1(z_1) & 1 \\ \varphi_2(z_2) & \psi_2(z_2) & 1 \\ \varphi_3(z_3) & \psi_3(z_3) & 1 \end{vmatrix} = 0 \tag{2.296}$$

must hold. Every relation between three variables z_1, z_2 and z_3 , which can be transformed into the form (2.296), can be represented by an *alignment nomogram* .
Next, follows the description of some important special cases of (2.296).

2.19.3.1 Alignment Charts with Three Straight-Line Scales Through a Point

If the zero point is chosen for the common point of the lines having the three scales z_1, z_2 or z_3 , then (2.296) has the form

$$\begin{vmatrix} \varphi_1(z_1) & m_1\varphi_1(z_1) & 1 \\ \varphi_2(z_2) & m_2\varphi_2(z_2) & 1 \\ \varphi_3(z_3) & m_3\varphi_3(z_3) & 1 \end{vmatrix} = 0\,, \tag{2.297}$$

since the equation of a line passing through the origin has the equation $y = mx$. Evaluating the determinant (2.297), yields

$$\frac{m_2 - m_3}{\varphi_1(z_1)} + \frac{m_3 - m_1}{\varphi_2(z_2)} + \frac{m_1 - m_2}{\varphi_3(z_3)} = 0 \tag{2.298a}$$

or

$$\frac{C_1}{\varphi_1(z_1)} + \frac{C_2}{\varphi_2(z_2)} + \frac{C_3}{\varphi_3(z_3)} = 0 \quad \text{with } C_1 + C_2 + C_3 = 0\,. \tag{2.298b}$$

Here C_1, C_2 and C_3 are constants.

■ The equation $\frac{1}{a} + \frac{1}{b} = \frac{2}{f}$ is a special case of (2.298b) and it is an important relation, for instance in optics or for the parallel connection of resistances. The corresponding alignment nomogram consists of three uniformly scaled lines.

2.19.3.2 Alignment Charts with Two Parallel Inclined Straight-Line Scales and One Inclined Straight-Line Scale

One of the scales is put on the y-axis, the other one on another line parallel to it at a distance d. The third scale is put on a line $y = mx$. In this case (2.296) has the form

$$\begin{vmatrix} 0 & \psi_1(z_1) & 1 \\ d & \psi_2(z_2) & 1 \\ \varphi_3(z_3) & m\varphi_3(z_3) & 1 \end{vmatrix} = 0. \tag{2.299}$$

Evaluation of the determinant by the first column yields

$$d\,(m\varphi_3(z_3) - \psi_1(z_1)) + \varphi_3(z_3)\,(\psi_1(z_1) - \psi_2(z_2)) = 0. \tag{2.300a}$$

Consequently:

$$\psi_1(z_1)\,\frac{\varphi_3(z_3) - d}{\varphi_3(z_3)} - (\psi_2(z_2) - md) = 0 \;\text{ oder }\; f(z_1)\cdot g(z_3) - h(z_2) = 0. \tag{2.300b}$$

It is often useful to introduce measure scales E_1 and E_2 of the form

$$E_1 f(z_1)\frac{E_2}{E_1} g(z_3) - E_2 h(z_2) = 0. \tag{2.300c}$$

Then, $\varphi_3(z_3) = \dfrac{d}{1 - \dfrac{E_2}{E_1} g(z_3)}$ holds. The relation $E_2 : E_1$ can be chosen so that the third scale is pulled near a certain point or it is gathered. Substituting $m = 0$, then $E_2 h(z_2) = \psi_2(z_2)$ and in this case, the line of the third scale passes through both the starting points of the first and of the second scale. Consequently, these two scales must be placed with a scale division in opposite directions, while the third one will be between them.

■ The relation between the Cartesian coordinates x and y of a point in the x, y plane and the corresponding angle φ in polar coordinates is:

$$y^2 = x^2 \tan^2 \varphi\,. \tag{2.301}$$

The corresponding nomogram is shown in **Fig. 2.110**. The scale division is the same for the scales of x and y but they are oriented in opposite directions. In order to get a better intersection with the third scale between them, their initial points are shifted by a suitable amount. The intersection points of the third scale with the first or with the second one are marked by $\varphi = 0$ or $\varphi = 90°$ respectively.

■ For instance: $x = 3$ $y = 3.5$, gives $\varphi \approx 49.5°$.

2.19.3.3 Alignment Charts with Two Parallel Straight Lines and a Curved Scale

If one of the straight-line scales is placed on the y-axis and the other one is placed at a distance d from it, then equation (2.296) has the form

$$\begin{vmatrix} 0 & \psi_1(z_1) & 1 \\ d & \psi_2(z_2) & 1 \\ \varphi_3(z_3) & \psi_3(z_3) & 1 \end{vmatrix} = 0\,. \tag{2.302}$$

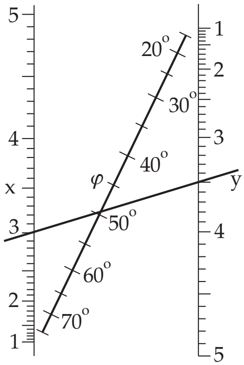

Figure 2.110

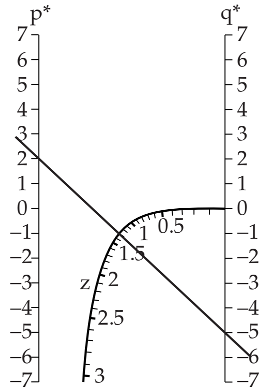

Figure 2.111

Consequently:

$$\psi_1(z_1) + \psi_2(z_2)\frac{\varphi_3(z_3)}{d - \varphi_3(z_3)} - d\frac{\psi_3(z_3)}{d - \varphi_3(z_3)} = 0. \tag{2.303a}$$

Choosing the scale E_1 for the first scale and E_2 for the second one, then (2.303a) is transformed into

$$E_1 f(z_1) + E_2 g(z_2)\frac{E_1}{E_2}h(z_3) + E_1 k(z_3) = 0, \tag{2.303b}$$

where $\psi_1(z_1) = E_1 f(z_1)$, $\psi_2(z_2) = E_2 g(z_2)$ and

$$\varphi_3(z_3) = \frac{dE_1 h(z_3)}{E_2 + E_1 h(z_3)} \quad \text{and} \quad \psi_3(z_3) = -\frac{E_1 E_2 k(z_3)}{E_2 + E_1 h(z_3)} \tag{2.303c}$$

holds.

■ The reduced third-degree equation $z^3 + p^*z + q^* = 0$ (see 1.6.2.3, p. 40) is of the form (2.303b). After the substitutions $E_1 = E_2 = 1$ and $f(z_1) = q^*$, $g(z_2) = p^*$, $h(z_3) = z$, the formulas to calculate the coordinates of the curved scale are $x = \varphi_3(z) = \dfrac{d \cdot z}{1 + z}$ and $y = \psi_3(z) = -\dfrac{z^3}{1+z}$.

In **Fig. 2.111** the curved scale is shown only for positive values of z . The negative values one gets by replacing z by $-z$ and by the determination of the positive roots from the equation $z^3 + p^*z - q^* = 0$. The complex roots $u + iv$ can also be determined by nomograms. Denoting the real root, which always exists, by z_1 , then the real part of the complex root is $u = -z_1/2$, and the imaginary part v can be determined from the equation $3u^2 - v^2 + p^* = \dfrac{3}{4}z_1^2 - v^2 + p^* = 0$.

■ For instance: $y^3 + 2y - 5 = 0$, i.e., $p^* = 2, q^* = -5$. One reads $z_1 \approx 1.3$.

2.19.4 Net Charts for More Than Three Variables

To construct a chart for formulas containing more than three variables, the expression is to decompose by the help of auxiliary variables into several formulas, each containing only three variables. Here, every auxiliary variable must be contained in exactly two of the new equations. Each of these equations is to be represented by an alignment chart so that the common auxiliary variable has the same scale.