12 Functional Analysis

1. Functional Analysis
Functional analysis arose after the recognition of a common structure in different disciplines such as the sciences, engineering and economics. General principles were discovered that resulted in a common and unified approach in calculus, linear algebra, geometry, and other mathematical fields, showing their interrelations.

2. Infinite Dimensional Spaces
There are many problems, the mathematical modeling of which requires the introduction of infinite systems of equations or inequalities. Differential or integral equations, approximation, variational or optimization problems could not be treated by using only finite dimensional spaces.

3. Linear and Non-Linear Operators
In the first phase of applying functional analysis – mainly in the first half of the twentieth century – linear or linearized problems were thoroughly examined, which resulted in the development of the theory of linear operators. More recently the application of functional analysis in practical problems required the development of the theory of non-linear operators, since more and more problems had to be solved that could be described only by non-linear methods. Functional analysis is increasingly used in solving differential equations, in numerical analysis and in optimization, and its principles and methods became a necessary tool in engineering and other applied sciences.

4. Basic Structures
In this chapter only the basic structures will be introduced, and only the most important types of abstract spaces and some special classes of operators in these spaces will be discussed. The abstract notion will be demonstrated by some examples, which are discussed in detail in other chapters of this book, and the existence and uniqueness theorems of the solutions of such problems are stated and proved there. Because of its abstract and general nature it is clear that functional analysis offers a large range of general relations in the form of mathematical theorems that can be directly used in solving a wide variety of practical problems.

12.1 Vector Spaces

12.1.1 Notion of a Vector Space

A non-empty set V is called a *vector space* or *linear space* over the field $\mathbb{F}$ of scalars if there exist two operations on V – addition of the elements and multiplication by scalars from $\mathbb{F}$ – such that they have the following properties:

1. for any two elements $x, y \in$ V, there exists an element $z = x + y \in$ V, which is called their *sum*.

2. For every $x \in$ V and every scalar (number) $\alpha \in \mathbb{F}$ there exists an element $\alpha x \in$ V, the *product* of x and the scalar α so that the following properties, the *axioms of vector spaces* (see also 5.3.8.1, p. 365), are satisfied for arbitrary elements $x, y, z \in$ V and scalars $\alpha, \beta \in \mathbb{F}$:

(V1) $x + (y + z) = (x + y) + z.$ (12.1)

(V2) There exists an element $0 \in$ V, the zero element, such that $x + 0 = x$. (12.2)

(V3) To every vector x there is a vector $-x$ such that $x + (-x) = 0$. (12.3)

(V4) $x + y = y + x.$ (12.4)

(V5) $1 \cdot x = x, \quad 0 \cdot x = 0.$ (12.5)

(V6) $\alpha(\beta x) = (\alpha\beta)x.$ (12.6)

(V7) $(\alpha + \beta)x = \alpha x + \beta x.$ (12.7)

(V8) $\alpha(x + y) = \alpha x + \alpha y.$ (12.8)

I.N. Bronshtein et al., *Handbook of Mathematics*,
DOI 10.1007/978-3-662-46221-8_12

V is called a real or complex vector space, depending on whether $\mathbb{F}$ is the field $\mathbb{R}$ of real numbers or the field $\mathbb{C}$ of complex numbers. The elements of V are also called either points or, according to linear algebra, *vectors*. The vector notation $\vec{x}$ or $\underline{x}$ is not used in functional analysis.

The difference $x - y$ of two arbitrary vectors $x, y \in \mathrm{V}$ also can be defined in V as $x - y = x + (-y)$. From the previous definition, it follows that the equation $x + y = z$ can be solved uniquely for arbitrary elements y and z. The solution is $x = z - y$. Further properties follow from axioms **(V1)**–**(V8)**:

- the zero element is uniquely defined,
- $\alpha x = \beta x$ and $x \neq 0$, imply $\alpha = \beta$,
- $\alpha x = \alpha y$ and $\alpha \neq 0$, imply $x = y$,
- $-(\alpha x) = \alpha \cdot (-x)$.

12.1.2 Linear and Affine Linear Subsets

1. Linear Subsets

A non-empty subset V_0 of a vector space V is called a *linear subspace* or a *linear manifold* of V if together with two arbitrary elements $x, y \in \mathrm{V}_0$ and two arbitrary scalars $\alpha, \beta \in \mathbb{F}$, their linear combination $\alpha x + \beta y$ is also in V_0. V_0 is a vector space in its own right, and therefore satisfies the axioms **(V1)**–**(V8)**. The subspace V_0 can be V itself or only the zero point. In these cases the subspace is called trivial.

2. Affine Subspaces

A subset of a vector space V is called an *affine linear subspace* or an *affine manifold* if it has the form

$$\{x_0 + y : y \in \mathrm{V}_0\}, \tag{12.9}$$

where $x_0 \in \mathrm{V}$ is a given element and V_0 is a linear subspace. It can be considered (in the case $x_0 \neq 0$) as the generalization of the lines or planes not passing through the origin in $\mathbb{R}^3$.

3. The Linear Hull

The intersection of an arbitrary number of subspaces in V is also a subspace. Consequently, for every non-empty subset $E \subset \mathrm{V}$, there exists a smallest linear subset $lin(E)$ or $[E]$ in V containing E, namely the intersection of all the linear subspaces, which contain E. The set $lin(E)$ is called the *linear hull of the set* E, or the *linear subspace generated by the set* E. It coincides with the set of all (finite) linear combinations

$$\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n, \tag{12.10}$$

comprised of elements $x_1, x_2, \ldots, x_n \in E$ and scalars $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}$.

4. Examples for Vector Spaces of Sequences

■ **A Vector Space $\mathbb{F}^n$:** Let n be a given natural number and V the set of all n-tuples, i.e., all finite sequences consisting of n scalar terms $\{(\xi_1, \ldots, \xi_n) : \ \xi_i \in \mathbb{F}, \ i = 1, \ldots, n\}$. The operations will be defined componentwise or termwise, i.e., if $x = (\xi_1, \ldots, \xi_n)$ and $y = (\eta_1, \ldots, \eta_n)$ are two arbitrary elements from V and α is an arbitrary scalar, $\alpha \in \mathbb{F}$, then

$$x + y = (\xi_1 + \eta_1, \ldots, \xi_n + \eta_n), \ (12.11\text{a}) \qquad \alpha \cdot x = (\alpha\xi_1, \ldots, \alpha\xi_n). \ (12.11\text{b})$$

In this way, the vector space $\mathbb{F}^n$ is defined. The linear spaces $\mathbb{R}$ or $\mathbb{C}$ are special cases for $n = 1$. This example can be generalized in two different ways (see examples **B** and **C**).

■ **B Vector Space s of all Sequences:** Considering the infinite sequences as elements $x = \{\xi_n\}_{n=1}^{\infty}$, $\xi_n \in \mathbb{F}$ and defining the operations componentwise, similar to (12.11a) and (12.11b), the vector space **s** of all sequences are obtained.

■ **C Vector Space φ(also $\mathbf{c_{00}}$) of all Finite Sequences:** Let V be the subset of all elements of **s** containing only a finite number of non-zero components, where the number of non-zero components depends on the element. This vector space – the operations are again introduced termwise – is denoted by $\boldsymbol{\varphi}$ or also by $\mathbf{c_{00}}$, and it is called the space of all *finite sequences of numbers*.

■ **D Vector Space m (also $\mathbf{l}^\infty$) of all Bounded Sequences:** A sequence $x = \{\xi_n\}_{n=1}^\infty$ belongs to $\mathbf{m}$ if and only if there exists $C_x > 0$ with $|\xi_n| \le C_x,\ \forall n = 1, 2, \ldots$. This vector space is also denoted by $\mathbf{l}^\infty$.

■ **E Vector Space c of all Convergent Sequences:** A sequence $x = \{\xi_n\}_{n=1}^\infty$ belongs to $\mathbf{c}$ if and only if there exists a number $\xi_0 \in \mathbb{F}$ such that for $\forall\, \varepsilon > 0$ there exists an index $n_0 = n_0(\varepsilon)$ such that for all $n > n_0$ one has $|\xi_n - \xi_0| < \varepsilon$ (see 7.1.2, p. 458).

■ **F Vector Space $\mathbf{c_0}$ of all Null Sequences:** The vector space $\mathbf{c_0}$ of all null sequences, i.e., the subspace of $\mathbf{c}$ consisting of all sequences converging to zero ($\xi_0 = 0$).

■ **G Vector Space $\mathbf{l}^p$:** The vector space of all sequences $x = \{\xi_n\}_{n=1}^\infty$ such that $\sum_{n=1}^\infty |\xi_n|^p$ is convergent, is denoted by $\mathbf{l}^p$ $(1 \le p < \infty)$.
It can be shown by the Minkowski inequality that the sum of two sequences from $\mathbf{l}^p$ also belongs to $\mathbf{l}^p$, (see 1.4.2.13, p. 32).

Remark: For the vector spaces introduced in examples **A**–**G**, the following inclusions hold:

$$\varphi \subset \mathbf{c_0} \subset \mathbf{c} \subset \mathbf{m} \subset \mathbf{s} \quad \text{and} \quad \varphi \subset \mathbf{l}^p \subset \mathbf{l}^q \subset \mathbf{c_0}, \quad \text{where} \quad 1 \le p < q < \infty. \tag{12.12}$$

5. Examples of Vector Spaces of Functions

■ **A Vector Space $\mathcal{F}(T)$:** Let V be the set of all real or complex valued functions defined on a given set T, where the operations are defined point-wise, i.e., if $x = x(t)$ and $y = y(t)$ are two arbitrary elements of V and $\alpha \in \mathbb{F}$ is an arbitrary scalar, then we define the elements (functions) $x + y$ and $\alpha \cdot x$ by the rules

$$(x + y)(t) = x(t) + y(t) \quad \forall\, t \in T, \tag{12.13a}$$

$$(\alpha x)(t) = \alpha \cdot x(t) \quad \forall\, t \in T. \tag{12.13b}$$

This vector space is denoted by $\mathcal{F}(T)$.
Some of the subspaces are introduced in the following examples.

■ **B Vector Space $\mathcal{B}(T)$ or $\mathcal{M}(T)$:** The space $\mathcal{B}(T)$ is the space of all functions bounded on T. This vector space is often denoted by $\mathcal{M}(T)$. In the case of $T = \mathbb{N}$, one gets the space $\mathcal{M}(\mathbb{N}) = \mathbf{m}$ from example **D** of the previous paragraph.

■ **C Vector Space $\mathcal{C}([a, b])$:** The set $\mathcal{C}([a, b])$ of all functions continuous on the interval $[a, b]$ (see 2.1.5.1, p. 58).

■ **D Vector Space $\mathcal{C}^{(k)}([a, b])$:** Let $k \in \mathbb{N},\ k \ge 1$. The set $\mathcal{C}^{(k)}([a, b])$ of all functions k-times continuously differentiable on $[a, b]$ (see 6.1, p. 432–437) is a vector space. At the endpoints a and b of the interval $[a, b]$, the derivatives have to be considered as right-hand and left-hand derivatives, respectively.

Remark: For the vector spaces of examples **A**–**D** of this paragraph, and $T = [a, b]$ the following subspace relations hold:

$$\mathcal{C}^{(k)}([a, b]) \subset \mathcal{C}([a, b]) \subset \mathcal{B}([a, b]) \subset \mathcal{F}([a, b]). \tag{12.14}$$

■ **E Vector Subspace of $\mathcal{C}([a, b])$:** For any given point $t_0 \in [a, b]$, the set $\{x \in \mathcal{C}([a, b]) \colon x(t_0) = 0\}$ forms a linear subspace of $\mathcal{C}([a, b])$.

12.1.3 Linearly Independent Elements

1. Linear Independence

A finite subset $\{x_1, \ldots, x_n\}$ of a vector space V is called *linearly independent* if

$$\alpha_1 x_1 + \cdots + \alpha_n x_n = 0 \quad \text{implies} \quad \alpha_1 = \cdots = \alpha_n = 0. \tag{12.15}$$

Otherwise, it is called *linearly dependent*. If $\alpha_1 = \cdots = \alpha_n = 0$, then for arbitrary vectors $x_1, \ldots, x_n$ from V, the vector $\alpha_1 x_1 + \cdots + \alpha_n x_n$ is trivially the zero element of V. Linear independence of the vectors $x_1, \ldots, x_n$ means that the only way to produce the zero element $0 = \alpha_1 x_1 + \cdots + \alpha_n x_n$ is when all coefficients are equal to zero $\alpha_1 = \cdots = \alpha_n = 0$. This important notion is well known from linear algebra (see 5.3.8.2, p. 366) and was used e.g. for the definition of a fundamental system of solutions of linear homogeneous differential equations (see 9.1.2.3, **2.**, p. 553). An infinite subset $E \subset \mathrm{V}$ is called *linearly independent* if every finite subset of E is linearly independent. Otherwise, E is called *linearly dependent*.

■ If the sequence whose k-th term is equal to 1 and all the others are 0 is denoted by e_k , then belongs e_k to the space $\boldsymbol{\varphi}$ and consequently to any space of sequences. The set $\{e_1, e_2, \ldots\}$ is linearly independent in every one of these spaces. In the space $\mathcal{C}([0,\pi])$, e.g., the system of functions

$$1,\ \sin nt,\ \cos nt \quad (n = 1, 2, 3, \ldots)$$

is linearly independent, but the functions $1, \cos 2t, \cos^2 t$ are linearly dependent (see (2.97), p. 81).

2. Basis and Dimension of a Vector Space

A linearly independent subset B from V, which generates the whole space V, i.e., $lin(B) = \mathrm{V}$ holds, is called an *algebraic basis* or a *Hamel basis* of the vector space V (see 5.3.8.2, p. 366). $B = \{x_\xi : \xi \in \Xi\}$ is a basis of V if and only if every vector $x \in \mathrm{V}$ can be written in the form $x = \sum_{\xi\in\Xi} \alpha_\xi x_\xi$, where the coefficients α_ξ are uniquely determined by x and only a finite number of them (depending on x) can be different from zero. Every non-trivial vector space V, i.e., $\mathrm{V} \neq \{0\}$, has at least one algebraic basis, and for every linearly independent subset E of V, there exists at least one algebraic basis of V, which contains this subset of E.

A vector space V is *m-dimensional* if it possesses a basis consisting of m vectors. That is, there exist m linearly independent vectors in V, and every system of $m+1$ vectors is linearly dependent.

A vector space is *infinite dimensional* if it has no finite basis, i.e., if for every natural number m there are m linearly independent vectors in V.

The space $\mathbb{F}^n$ is n-dimensional, and all the other spaces in examples **B**–**E** are infinite dimensional. The subspace $lin(\{1, t, t^2\}) \subset \mathcal{C}([a,b])$ is three-dimensional.

In the finite dimensional case, every two bases of the same vector space have the same number of elements. Also in an infinite dimensional vector space any two bases have the same cardinality, which is denoted by $dim(\mathrm{V})$. The dimension is an invariant quantity of the vector space, it does not depend on the particular choice of an algebraic basis.

12.1.4 Convex Subsets and the Convex Hull

12.1.4.1 Convex Sets

A subset C of a real vector space V is called *convex* if for every pair of vectors $x, y \in C$ all vectors of the form $\lambda x + (1-\lambda)y,\ \ 0 \le \lambda \le 1$, also belong to C. In other words, the set C is convex, if for any two elements x and y, the whole line segment

$$\{\lambda x + (1-\lambda)y :\ \ 0 \le \lambda \le 1\}, \tag{12.16}$$

(which is also called an interval), belongs to C. (For examples of convex sets in $\mathbb{R}^2$ see the sets denoted by A and B in **Fig. 12.5**, p. 684.)

The intersection of an arbitrary number of convex sets is also a convex set, where the empty set is agreed to be convex. Consequently, for every subset $E \subset \mathrm{V}$ there exists a smallest convex set which contains E, namely, the intersection of all convex subsets of V containing E. It is called the *convex hull* of the set E and it is denoted by $co\,(E)$. $co\,(E)$ is identical to the set of all finite *convex* linear combinations of elements from E, i.e., $co\,(E)$ consists of all elements of the form $\lambda_1 x_1 + \cdots + \lambda_n x_n$, where $x_1, \ldots, x_n$ are arbitrary elements from E and $\lambda_i \in [0,1]$ satisfy the equality $\lambda_1 + \cdots + \lambda_n = 1$. Linear and affine subspaces are always convex.

12.1.4.2 Cones

A non-empty subset C of a (real) vector space V is called a *convex cone* if it satisfies the following properties:

1. C is a convex set.
2. From $x \in C$ and $\lambda \geq 0$, it follows that $\lambda x \in C$.
3. From $x \in C$ and $-x \in C$, it follows that $x = 0$.

A cone can be characterized also by **3.** together with

$$x, y \in C \quad \text{and} \quad \lambda, \mu \geq 0 \quad \text{imply} \quad \lambda x + \mu y \in C. \tag{12.17}$$

■ **A:** The set $\mathbb{R}^n_+$ of all vectors $x = (\xi_1, \ldots, \xi_n)$ with non-negative components is a cone in $\mathbb{R}^n$.

■ **B:** The set $\mathcal{C}_+$ of all real continuous functions on $[a, b]$ with only non-negative values is a cone in the space $\mathcal{C}([a, b])$.

■ **C:** The set of all sequences of real numbers $\{\xi_n\}_{n=1}^{\infty}$ with only non-negative terms, i.e., $\xi_n \geq 0,\ \forall n$, is a cone in **s**. Analogously, cones are obtained in the spaces of examples **C**–**G** in 12.1.2, p. 655, if the sets of non-negative sequences are considered in these spaces.

■ **D:** The set $C \subset \mathbf{l}^p$ $(1 \leq p < \infty)$, consisting of all sequences $\{\xi_n\}_{n=1}^{\infty}$, such that for some $a > 0$

$$\sum_{n=1}^{\infty} |\xi_n|^p \leq a \tag{12.18}$$

is a convex set in $\mathbf{l}^p$, but obviously, not a cone.

■ **E:** Examples from $\mathbb{R}^2$ see **Fig. 12.1**: **a)** convex set, not a cone, **b)** not convex, **c)** convex hull.

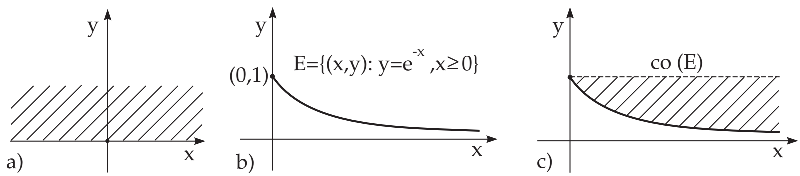

Figure 12.1

12.1.5 Linear Operators and Functionals

12.1.5.1 Mappings

A mapping T: $D \longrightarrow \mathbb{Y}$ from the set $D \subset \mathbb{X}$ into the set $\mathbb{Y}$ is called

- *injective*, if

$$T(x) = T(y) \Longrightarrow x = y, \tag{12.19}$$

- *surjective*, if for

$$\forall\, y \in \mathbb{Y} \quad \text{there exists an element} \quad x \in D \quad \text{such that} \quad T(x) = y, \tag{12.20}$$

- *bijective*, if T is both injective and surjective.

D is called the *domain* of the mapping T and is denoted by D_T or $D(T)$, while the subset $\{y \in \mathbb{Y} : \exists\, x \in D_T \text{ with } T(x) = y\}$ of $\mathbb{Y}$ is called the *range* of the mapping T and is denoted by $\mathcal{R}(T)$ or $Im(T)$.

12.1.5.2 Homomorphism and Endomorphism

Let $\mathbb{X}$ and $\mathbb{Y}$ be two vector spaces over the same field $\mathbb{F}$ and D a linear subset of $\mathbb{X}$. A mapping T: $D \longrightarrow \mathbb{Y}$ is called *linear* (or a *linear transformation, linear operator* or *homomorphism*), if for arbitrary $x, y \in D$ and $\alpha, \beta \in \mathbb{F}$,

$$T(\alpha x + \beta y) = \alpha Tx + \beta Ty. \tag{12.21}$$

For a linear operator T the notation Tx is preferred, which is similarly used for linear functions, while the notation $T(x)$ is used for general operators.

The range $\mathcal{R}(T)$ is the set of all $y \in \mathrm{Y}$ such that the equation $Tx = y$ has at least one solution. $N(T) = \{x \in \mathrm{X} : Tx = 0\}$ is the *null space* or *kernel* of the operator T and is also denoted by $\ker(T)$.

A mapping of the vector space X into itself is called an *endomorphism*. If T is an injective linear mapping, then the mapping defined on $\mathcal{R}(T)$ by

$$y \longmapsto x, \text{ such that } Tx = y,\ y \in \mathcal{R}(T) \tag{12.22}$$

is linear. It is denoted by T^{-1}: $\mathcal{R}(T) \longrightarrow \mathrm{X}$ and is called the *inverse* of T. If Y is the vector space $\mathbb{F}$, then a linear mapping f: $\mathrm{X} \longrightarrow \mathbb{F}$ is called a *linear functional* or a *linear form*.

12.1.5.3 Isomorphic Vector Spaces

A bijective linear mapping T: $\mathrm{X} \longrightarrow \mathrm{Y}$ is called an *isomorphism* of the vector spaces X and Y. Two vector spaces are called *isomorphic* provided an isomorphism exists.

12.1.6 Complexification of Real Vector Spaces

Every real vector space V can be extended to a complex vector space $\tilde{\mathrm{V}}$. The set $\tilde{\mathrm{V}}$ consists of all pairs (x, y) with $x, y \in \mathrm{V}$. The operations (addition and multiplication by a complex number $a + \mathrm{i}b \in \mathbb{C}$) are defined as follows:

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2), \quad (12.23\text{a}) \qquad (a+\mathrm{i}b)(x, y) = (ax - by, bx + ay). \quad (12.23\text{b})$$

Since the special relations

$$(x, y) = (x, 0) + (0, y) \text{ and } \mathrm{i}(y, 0) = (0 + \mathrm{i}1)(y, 0) = (0 \cdot y - 1 \cdot 0, 1y + 0 \cdot 0) = (0, y) \tag{12.24}$$

hold, the pair (x, y) can also be written as $x + \mathrm{i}y$. The set $\tilde{\mathrm{V}}$ is a complex vector space, where the set V is identified with the linear subspace $\tilde{\mathrm{V}}_0 = \{(x, 0)\colon x \in \mathrm{V}\}$, i.e., $x \in \mathrm{V}$ is considered as $(x, 0)$ or as $x + \mathrm{i}0$.

This procedure is called the *complexification* of the vector space V. A linearly independent subset in V is also linearly independent in $\tilde{\mathrm{V}}$. The same statement is valid for a basis in V, so $dim(\mathrm{V}) = dim(\tilde{\mathrm{V}})$.

12.1.7 Ordered Vector Spaces

12.1.7.1 Cone and Partial Ordering

If a cone C is fixed in a vector space V, then an *order* can be introduced for certain pairs of vectors in V. Namely, if $x - y \in C$ for some $x, y \in \mathrm{V}$ then one writes $x \geq y$ or $y \leq x$ and say x is *greater than or equal to* y or y is *smaller than or equal to* x. The pair (V, C) is called an *ordered vector space* or a vector space *partially ordered by the cone* C. An element x is called *positive*, if $x \geq 0$ or, which means the same, if $x \in C$ holds. Moreover

$$C = \{x \in \mathrm{V}\colon x \geq 0\}. \tag{12.25}$$

If the vector space $\mathbb{R}^2$ ordered by its first quadrant as the cone $C(= \mathbb{R}^2_+)$ is under consideration, then a typical phenomenon of ordered vector spaces will be seen. This is referred to as "partially ordered" or sometimes as "semi-ordered". Namely, only certain pairs of two vectors are comparable. Considering the vectors $x = (1, -1)$ and $y = (0, 2)$, neither the vector $x - y = (1, -3)$ nor $y - x = (-1, 3)$ is in C, so neither $x \geq y$ nor $x \leq y$ holds. An ordering in a vector space, generated by a cone, is always only a partial ordering.

It can be shown that the binary relation $\geq$ has the following properties:

(O1) $x \geq x\ \forall\, x \in \mathrm{V}$ (reflexivity). (12.26)

(O2) $x \geq y$ and $y \geq z$ imply $x \geq z$ (transitivity). (12.27)

(O3) $x \geq y$ and $\alpha \geq 0,\ \alpha \in \mathbb{R}$, imply $\alpha x \geq \alpha y$. (12.28)

(O4) $x_1 \geq y_1$ and $x_2 \geq y_2$ imply $x_1 + x_2 \geq y_1 + y_2$. (12.29)

Conversely, if in a vector space V there exists an ordering relation, i.e., a binary relation $\geq$ is defined for certain pairs of elements and satisfies axioms **(O1)**–**(O4)**, and if one puts

$$V_+ = \{x \in V\colon x \geq 0\}, \tag{12.30}$$

then it can be shown that V_+ is a cone. The order $\geq_{V_+}$ in V induced by V_+ is identical to the original order $\geq$; consequently, the two possibilities of introducing an order in a vector space are equivalent. A cone $C \subset V$ is called *generating* or *reproducing* if every element $x \in V$ can be represented as $x = u - v$ with $u, v \in C$. It can be written in the form $V = C – C$.

■ **A:** An obvious order in the space **s** (see example **B**, p. 655) is induced by means of the cone

$$C = \{x = \{\xi_n\}_{n=1}^{\infty} : \xi_n \geq 0 \quad \forall\, n\} \tag{12.31}$$

(see example **C**, p. 658).

In the spaces of sequences (see (12.12), p 656) usually the natural coordinate-wise order is considered. This is defined by the cone obtained as the intersection of the considered space with C (see (12.31), p. 660). The positive elements in these ordered vector spaces are then the sequences with non-negative terms. It is clear that other orders can be defined by other cones, as well. Then orderings different from the natural ordering can be obtained (see [12.17], [12.19]).

■ **B:** In the real spaces of functions $\mathcal{F}(T)$, $\mathcal{B}(T)$, $\mathcal{C}([a,b])$ and $\mathcal{C}^{(k)}([a,b])$ (see 12.1.2, **5.**, p. 656), the natural order $x \geq y$ for two functions x and y is defined by $x(t) \geq y(t)$, $\forall\, t \in T$, or $\forall\, t \in [a,b]$. Then $x \geq 0$ if and only if x is a non-negative function in T. The corresponding cones are denoted by $\mathcal{F}_+(T)$, $\mathcal{B}_+(T)$, etc. Also $C_+ = \mathcal{C}_+(T) = \mathcal{F}_+(T) \cap \mathcal{C}(T)$ can be obtained if $T = [a,b]$.

12.1.7.2 Order Bounded Sets

Let E be an arbitrary non-empty subset of an ordered vector space V. An element $z \in V$ is called an *upper bound* of the set E if for every $x \in E$, $x \leq z$. An element $u \in V$ is a *lower bound* of E if $u \leq x$, $\forall\, x \in E$. For any two elements $x, y \in V$ with $x \leq y$, the set

$$[x, y] = \{v \in V\colon x \leq v \leq y\} \tag{12.32}$$

is called an *order interval* or (o)-*interval*.

Obviously, the elements x and y are a lower bound and an upper bound of the set $[x,y]$, respectively, where they even belong to the set. A set $E \subset V$ is called *order bounded* or simply (o) *bounded*, if E is a subset of an order interval, i.e., if there exist two elements $u, z \in V$ such that $u \leq x \leq z$, $\forall\, x \in E$ or, equivalently, $E \subset [u, z]$. A set is called *bounded above* or *bounded below* if it has an upper bound, or a lower bound, respectively.

12.1.7.3 Positive Operators

A linear operator (see [12.2], [12.17]) $T\colon X \longrightarrow Y$ from an ordered vector space $X = (X, X_+)$ into an ordered vector space $Y = (Y, Y_+)$ is called *positive*, if

$$T(X_+) \subset Y_+, \quad \text{i.e.,} \quad Tx \geq 0 \quad \text{for all} \quad x \geq 0. \tag{12.33}$$

12.1.7.4 Vector Lattices

1. Vector Lattices

In the vector space $\mathbb{R}^1$ of the real numbers the notions of (o)-boundedness and boundedness (in the usual sense) are identical. It is known that every set of real numbers which is bounded from above has a supremum: the smallest of its upper bounds (or the least upper bound, sometimes denoted by lub). Analogously, if a set of reals is bounded from below, then it has an *infimum*, the greatest lower bound, sometimes denoted by glb. In a general ordered vector space, the existence of the supremum and infimum cannot be guaranteed even for finite sets. They must be given by axioms. An ordered vector space V is called a *vector lattice* or a *linear lattice* or a *Riesz space*, if for two arbitrary elements $x, y \in V$ there exists an element $z \in V$ with the following properties:

1. $x \leq z$ and $y \leq z$,

2. if $u \in V$ with $x \leq u$ and $y \leq u$, then $z \leq u$.

Such an element z is uniquely determined, it is denoted by $x \vee y$, and it is called the *supremum* of x and y (more precisely: supremum of the set consisting of the elements x and y). In a vector lattice, there also exists the infimum for any x and y, which is denoted by $x \wedge y$. For applications of positive operators in vector lattices see, e.g., [12.2], [12.3] [12.15].

A vector lattice is called *Dedekind complete* or a *K-space* (*Kantorovich space*) if every non-empty subset E that is order bounded from above has a supremum lub(E) (equivalently, if every non-empty subset that is order bounded from below has an infimum glb(E)).

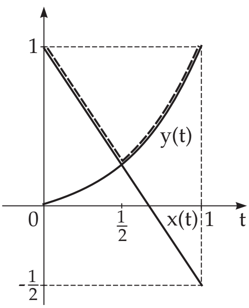

Figure 12.2

■ **A:** In the vector lattice $\mathcal{F}([a,b])$ (see 12.1.2, **5.**, p. 656), the supremum of two functions x, y is calculated pointwise by the formula

$$(x \vee y)(t) = \max\{x(t), y(t)\} \quad \forall\, t \in [a,b]. \tag{12.34}$$

In the case of $[a,b] = [0,1]$, $x(t) = 1 - \frac{3}{2}t$ and $y(t) = t^2$ **(Fig. 12.2)**,

$$(x \vee y)(t) = \begin{cases} 1 - \frac{3}{2}t, & \text{if } \; 0 \le t \le \frac{1}{2}, \\ t^2, & \text{if } \; \frac{1}{2} \le t \le 1 \end{cases} \tag{12.35}$$

is obtained.

■ **B:** The spaces $\mathcal{C}([a,b])$ and $\mathcal{B}([a,b])$ (see 12.1.2, **5.**, p. 656) are also vector lattices, while the ordered vector space $\mathcal{C}^{(1)}([a,b])$ is not a vector lattice, since the minimum or maximum of two differentiable functions may not be differentiable on $[a,b]$, in general.

A linear operator $T\colon$ X $\longrightarrow$ Y from a vector lattice X into a vector lattice Y is called a *vector lattice homomorphism* or *homomorphism of the vector lattice*, if for all $x, y \in$ X

$$T(x \vee y) = Tx \vee Ty \quad \text{and} \quad T(x \wedge y) = Tx \wedge Ty. \tag{12.36}$$

2. Positive and Negative Parts, Modulus of an Element

For an arbitrary element x of a vector lattice V, the elements

$$x_+ = x \vee 0, \quad x_- = (-x) \vee 0 \quad \text{and} \quad |x| = x_+ + x_- \tag{12.37}$$

are called the *positive part, negative part*, and *modulus* of the element x, respectively. For every element $x \in$ V, the three elements $x_+, x_-, |x|$ are positive, where for $x, y \in$ V the following relations are valid:

$$x \le x_+ \le |x|, \quad x = x_+ - x_-, \quad x_+ \wedge x_- = 0, \quad |x| = x \vee (-x), \tag{12.38a}$$

$$(x+y)_+ \le x_+ + y_+, \quad (x+y)_- \le x_- + y_-, \quad |x+y| \le |x| + |y|, \tag{12.38b}$$

$$x \le y \quad \text{implies} \quad x_+ \le y_+ \quad \text{and} \quad x_- \ge y_- \tag{12.38c}$$

and for arbitrary $\alpha \ge 0$

$$(\alpha\, x)_+ = \alpha\, x_+, \quad (\alpha\, x)_- = \alpha\, x_-, \quad |\alpha\, x| = \alpha |x|. \tag{12.38d}$$

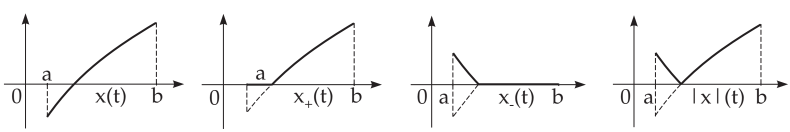

Figure 12.3

In the vector spaces $\mathcal{F}([a,b])$ and $\mathcal{C}([a,b])$, the positive part, the negative part, and the modulus of a function $x(t)$ can be got by means of the following formulas **(Fig. 12.3)**:

$$x_+(t) = \begin{cases} x(t), & \text{if } x(t) \ge 0, \\ 0, & \text{if } x(t) < 0, \end{cases} \qquad (12.39a)$$

$$x_-(t) = \begin{cases} 0, & \text{if } x(t) > 0, \\ -x(t), & \text{if } x(t) \le 0, \end{cases} \qquad (12.39b) \qquad |x|(t) = |x(t)| \quad \forall\, t \in [a,b]. \qquad (12.39c)$$

12.2 Metric Spaces

12.2.1 Notion of a Metric Space

Let X be a set, and suppose a real, non-negative function $\rho(x,y)$ $(x, y \in \mathrm{X})$ is defined on $\mathrm{X} \times \mathrm{X}$. If this function $\rho:\ \mathrm{X} \times \mathrm{X} \to \mathbb{R}^1_+$ satisfies the following properties **(M1)**–**(M3)** for arbitrary elements $x, y, z \in \mathrm{X}$, then it is called a *metric* or *distance* in the set X, and the pair $\mathrm{X} = (\mathrm{X}, \rho)$ is called a *metric space*. The axioms of *metric spaces* are:

(M1) $\rho(x,y) \ge 0$ and $\rho(x,y) = 0$ if and only if $x = y$ (non-negativity), (12.40)

(M2) $\rho(x,y) = \rho(y,x)$ (symmetry), (12.41)

(M3) $\rho(x,y) \le \rho(x,z) + \rho(z,y)$ (triangle inequality). (12.42)

A metric can be defined on every subset Y of a metric space $\mathrm{X} = (\mathrm{X}, \rho)$ in a natural way if the metric ρ of the space X is restricted to the set Y, i.e., if ρ is considered only on the subset $\mathrm{Y} \times \mathrm{Y}$. The space (Y, ρ) of $\mathrm{X} \times \mathrm{X}$ is called a *subspace* of the metric space X.

■ **A:** The sets $\mathbb{R}^n$ and $\mathbb{C}^n$ are metric spaces with the *Euclidean metric* defined for points $x = (\xi_1, \ldots, \xi_n)$ and $y = (\eta_1, \ldots, \eta_n)$ as

$$\rho(x,y) = \sqrt{\sum_{k=1}^{n} |\xi_k - \eta_k|^2}. \qquad (12.43)$$

■ **B:** The function

$$\rho(x,y) = \max_{1 \le k \le n} |\xi_k - \eta_k| \qquad (12.44)$$

for vectors $x = (\xi_1, \ldots, \xi_n)$ and $y = (\eta_1, \ldots, \eta_n)$ also defines a metric in $\mathbb{R}^n$ and $\mathbb{C}^n$, the so-called *maximum metric*. If $\tilde{x} = (\tilde{\xi}_1, \ldots, \tilde{\xi}_n)$ is an approximation of the vector x, then it is of interest to know how much is the maximal deviation between the coordinates: $\max\limits_{1 \le k \le n} |\xi_k - \tilde{\xi}_k|$.

The function

$$\rho(x,y) = \sum_{k=1}^{n} |\xi_k - \eta_k| \qquad (12.45)$$

for vectors $x, y \in \mathbb{R}^n$ (or $\mathbb{C}^n$) defines a metric in $\mathbb{R}^n$ and $\mathbb{C}^n$, the so-called *absolute value metric*. The metrics (12.43), (12.44) and (12.45) are reduced in the case of $n = 1$ to the absolute value $|x - y|$ in the spaces $\mathbb{R}$ and $\mathbb{C}$ (the sets of real and complex numbers).

■ **C:** Finite 0-1 sequences, e.g., 1110 and 010110, are called *words* in coding theory. If the number of positions is counted where two words of the same length n have different digits, i.e., for $x = (\xi_1, \ldots, \xi_n)$, $y = (\eta_1, \ldots, \eta_n)$, $\xi_k, \eta_k \in \{0, 1\}$, $\varrho(x,y)$ is defined as the number of the $k \in \{1, \ldots, n\}$ values such that $\xi_k \ne \eta_k$, then the set of words with a given length n is a metric space, and the metric is the so-called *Hamming distance*, e.g., $\varrho((1110), (0100)) = 2$.

■ **D:** In the set $\mathbf{m}$ and in its subsets $\mathbf{c}$ and $\mathbf{c_0}$ (see (12.12), p. 656) a metric is defined by

$$\rho(x,y) = \sup_k |\xi_k - \eta_k|, \quad (x = (\xi_1, \xi_2, \ldots), y = \eta_1, \eta_2, \ldots)). \qquad (12.46)$$

■ **E:** In the set $\mathbf{l}^p$ $(1 \le p < \infty)$ of sequences $x = (\xi_1, \xi_2, \ldots)$ with absolutely convergent series $\sum\limits_{n=1}^{\infty} |\xi_n|^p$ a metric is defined by

$$\rho(x, y) = \sqrt[p]{\sum_{n=1}^{\infty} |\xi_n - \eta_n|^p}, \quad (x, y \in l^p). \tag{12.47}$$

■ **F:** In the set $\mathcal{C}([a,b])$ a metric is defined by

$$\rho(x, y) = \max_{t \in [a,b]} |x(t) - y(t)|. \tag{12.48}$$

■ **G:** In the set $\mathcal{C}^{(k)}([a,b])$ a metric is defined by

$$\rho(x, y) = \sum_{l=0}^{k} \max_{t \in [a,b]} |x^{(l)}(t) - y^{(l)}(t)|, \tag{12.49}$$

where (see (12.14) $\mathcal{C}^{(0)}([a,b])$ is understood as $\mathcal{C}([a,b])$).

■ **H:** Consider the set $L^p(\Omega)$ $(1 \le p < \infty)$ of the equivalence classes of Lebesgue measurable functions which are defined almost everywhere on a bounded domain $\Omega \subset \mathbb{R}^n$ and $\int\limits_{\Omega} |x(t)|^p \, d\mu < \infty$ (see also 12.9, p. 693). A metric in this set is defined by

$$\rho(x, y) = \sqrt[p]{\int_{\Omega} |x(t) - y(t)|^p \, d\mu}. \tag{12.50}$$

12.2.1.1 Balls, Neighborhoods and Open Sets

In a metric space $\mathrm{X} = (\mathrm{X}, \rho)$, whose elements are also called points, the following sets

$$B(x_0; r) = \{x \in \mathrm{X} : \rho(x, x_0) < r\}, \ (12.51) \qquad \overline{B}(x_0; r) = \{x \in \mathrm{X} : \rho(x, x_0) \le r\} \tag{12.52}$$

defined by means of a real number $r > 0$ and a fixed point x_0, are called an *open* and *closed ball* with radius r and center at x_0, respectively.
The balls (circles) defined by the metrics (12.43) and (12.44) and (12.45) in the vector space $\mathbb{R}^2$ are represented in **Fig. 12.4a,b** with $x_0 = 0$ and $r = 1$.

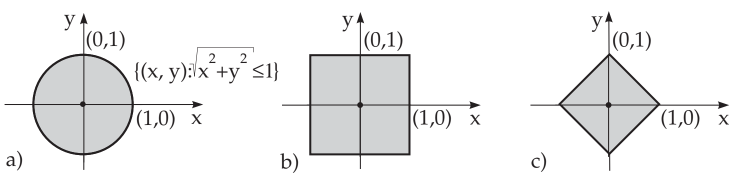

Figure 12.4

A subset U of a metric space $\mathrm{X} = (\mathrm{X}, \rho)$ is called a *neighborhood* of the point x_0 if U contains x_0 together with an open ball centered at x_0, in other words, if there exists an $r > 0$ such that $B(x_0; r) \subset U$. A neighborhood U of the point x is also denoted by $U(x)$. Obviously, every ball is a neighborhood of its center; an open ball is a neighborhood of all of its points. A point x_0 is called an *interior point* of a set $A \subset \mathrm{X}$ if x_0 belongs to A together with some of its neighborhood, i.e., there is a neighborhood U of x_0 such that $x_0 \in U \subset A$. A subset of a metric space is called *open* if all of its points are interior points. Obviously, X is an open set.
The open balls in every metric space, especially the open intervals in $\mathbb{R}$, are the prototypes of open sets.

The set of all open sets satisfies the following *axioms of open sets*:
- If G_α is open for $\forall\alpha \in I$, then the set $\bigcup\limits_{\alpha\in I} G_\alpha$ is also open.
- If $G_1, G_2, \ldots, G_n$ are finitely many arbitrary open sets, then the set $\bigcap\limits_{k=1}^{n} G_k$ is also open.
- The empty set $\emptyset$ is open by definition.

A subset A of a metric space is *bounded* if for a certain element x_0 (which does not necessarily belong to A) and a real number $R > 0$ the set A is in the ball $B(x_0; R)$, i.e., $\rho(x, x_0) < R$ for all $x \in A$.

12.2.1.2 Convergence of Sequences in Metric Spaces

Let $\mathrm{X} = (\mathrm{X}, \rho)$ be a metric space, $x_0 \in \mathrm{X}$ and $\{x_n\}_{n=1}^{\infty},\ x_n \in \mathrm{X}$ a sequence of elements of X.
The sequence $\{x_n\}_{n=1}^{\infty}$ is called *convergent* to the point x_0 if for every neighborhood $U(x_0)$ there is an index $n_0 = n_0(U(x_0))$ such that for all $n > n_0$, $x_n \in U(x_0)$. The usual notation

$$x_n \longrightarrow x_0 \quad (n \to \infty) \quad \text{or} \quad \lim_{n\to\infty} x_n = x_0 \tag{12.53}$$

is used and the point x_0 is called the *limit of the sequence* $\{x_n\}_{n=1}^{\infty}$. The limit of a sequence is uniquely determined. Instead of an arbitrary neighborhood of the point x_0, it is sufficient to consider only open balls with arbitrary radii, so (12.53) is equivalent to the following: $\forall\, \varepsilon > 0$ (now thinking about the open ball $B(x_0; \varepsilon)$), there is an index $n_0 = n_0(\varepsilon)$, such that if $n > n_0$, then $\rho(x_n, x_0) < \varepsilon$. Notice that (12.53) means $\rho(x_n, x_0) \longrightarrow 0$.

With these notions introduced in special metric spaces the distance between points can be calculated and the convergence of point sequences can be investigated. This has a great importance in numerical methods and in approximating functions by certain classes of functions (see, e.g., 19.6, p. 982).
In the space $\mathbb{R}^n$, equipped with one of the metrics given above, convergence always means coordinate-wise convergence.
In the spaces $\mathcal{B}([a, b])$ and $\mathcal{C}([a, b])$, the convergence introduced by (12.48) means uniform convergence of the function sequence on the set $[a, b]$ (see 7.3.2, p. 468).
In the space $L^2(\Omega)$ convergence with respect to the metric (12.50) means convergence in the (quadratic) mean, i.e., $x_n \to x_0$ if

$$\int_\Omega |x_n - x_0|^2\, d\mu \longrightarrow 0 \quad \text{for} \quad n \to \infty\,. \tag{12.54}$$

12.2.1.3 Closed Sets and Closure

1. Closed Sets

A subset F of a metric space X is called *closed* if $\mathrm{X} \setminus F$ is an open set. Every closed ball in a metric space, especially every interval of the form $[a, b],\ [a, \infty), (-\infty, a]$ in $\mathbb{R}$, is a closed set.
Corresponding to the axioms of open sets, the collection of all closed sets of a metric space has the following properties:
- If F_α are closed for $\forall\alpha \in I$, then the set $\bigcap\limits_{\alpha\in I} F_\alpha$ is closed.
- If $F_1, \ldots, F_n$ are finitely many closed sets, then the set $\bigcup\limits_{k=1}^{n} F_k$ is closed.
- The empty set $\emptyset$ is a closed set by definition.

The sets $\emptyset$ and X are open and closed at the same time.
A point x_0 of a metric space X is called a *limit point* of the subset $A \subset \mathrm{X}$ if for every neighborhood $U(x_0)$,

$$U(x_0) \cap A \neq \emptyset. \tag{12.55}$$

If this intersection always contains at least one point different from x_0, then x_0 is called an *accumulation point* of the set A. A limit point, which is not an accumulation point, is called an *isolated point*.
An accumulation point of A does not need to belong to the set A, e.g., the point a with respect to the set $A = (a, b]$, while an isolated point of A must belong to the set A.

A point x_0 is a limit point of the set A if there exists a sequence $\{x_n\}_{n=1}^{\infty}$ with elements x_n from A, which converges to x_0. If x_0 is an isolated point, then $x_n = x_0,\ \forall\, n \geq n_0$ for some index n_0.

2. The Closure of a Set

Every subset A of a metric space X obviously lies in the closed set X. Therefore, there always exists a smallest closed set containing A, namely the intersection of all closed sets of X, which contain A. This set is called the *closure* of the set A and it is usually denoted by $\overline{A}$. $\overline{A}$ is identical to the set of all limit points of A; $\overline{A}$ is obtained from the set A by adding all of its accumulation points to it. A is a closed set if and only if $A = \overline{A}$. Consequently, closed sets can be characterized by sequences in the following way: A is closed if and only if for every sequence $\{x_n\}_{n=1}^{\infty}$ of elements of A, which converges in X to an element $x_0 (\in \mathrm{X})$, the limit x_0 also belongs to A.

Boundary points of A are defined as follows: x_0 is a boundary point of A if for every neighborhood $U(x_0)$, $U(x_0) \cap A \neq \emptyset$ and also $U(x_0) \cap (X \setminus A) \neq \emptyset$. x_0 itself does not need to belong to A. Another characterization of a closed set is the following: A is closed if it contains all of its boundary points. (The set of boundary points of the metric space X is the empty set.)

12.2.1.4 Dense Subsets and Separable Metric Spaces

A subset A of a metric space X is called *everywhere dense* if $\overline{A} = \mathrm{X}$, i.e., each point $x \in \mathrm{X}$ is a limit point of the set A. That is, for each $x \in \mathrm{X}$, there is a sequence $\{x_n\}$ $x_n \in A$ such that $x_n \longrightarrow x$.

■ **A:** According to the Weierstrass approximation theorem, every continuous function on a bounded closed interval $[a, b]$ can be approximated arbitrarily well by polynomials in the metric space of the space $\mathcal{C}([a, b])$, i.e., uniformly. This theorem can now be formulated as follows: The set of polynomials on the interval $[a, b]$ is everywhere dense in $\mathcal{C}([a, b])$.

■ **B:** Further examples for everywhere dense subsets are the set of rational numbers $\mathbf{Q}$ and the set of irrational numbers in the space of the real numbers $\mathbb{R}$.

A metric space X is called *separable* if there exists a countable everywhere dense subset in X. A countable everywhere dense subset in $\mathbb{R}^n$ is, e.g., the set of all vectors with rational components. The space $\mathbf{l} = \mathbf{l}^1$ is also separable, since a countable everywhere dense subset is formed, for example, by the set of its elements of the form $x = (r_1, r_2, \ldots, r_N, 0, 0, \ldots)$, where r_i are rational numbers and $N = N(x)$ is an arbitrary natural number. The space $\mathbf{m}$ is not separable.

12.2.2 Complete Metric Spaces

12.2.2.1 Cauchy Sequences

Let $\mathrm{X} = (\mathrm{X}, \rho)$ be a metric space. A sequence $\{x_n\}_{n=1}^{\infty}$ with $x_n \in \mathrm{X}$ is called a *Cauchy sequence* if for $\forall\, \varepsilon > 0$ there is an index $n_0 = n_0(\varepsilon)$ such that for $\forall\, n, m > n_0$ there holds the inequality

$$\rho(x_n, x_m) < \varepsilon. \tag{12.56}$$

Every Cauchy sequence is a bounded set. Furthermore, every convergent sequence is a Cauchy sequence. In general, the converse statement is not true, as is shown in the following example.

■ Consider the space $\mathbf{l}^1$ with the metric (12.46) of the space $\mathbf{m}$. Obviously, the elements $x^{(n)} = \left(1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, 0, 0, \ldots\right)$ belong to $\mathbf{l}^1$ for every $n = 1, 2, \ldots$ and the sequence $\{x^{(n)}\}_{n=1}^{\infty}$ is a Cauchy sequence in this space. If the sequence (of sequences) $\{x^{(n)}\}_{n=1}^{\infty}$ converges, then it has to be convergent also coordinate-wise to the element $x^{(0)} = \left(1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \frac{1}{n+1}, \ldots\right)$. However, $x^{(0)}$ does not belong to $\mathbf{l}^1$, since $\sum\limits_{n=1}^{\infty} \frac{1}{n} = +\infty$ (see 7.2.1.1, **2.**, p. 459, harmonic series).

12.2.2.2 Complete Metric Spaces

A metric space X is called *complete* if every Cauchy sequence converges in X. Hence, complete metric spaces are the spaces for which the *Cauchy principle*, known from real calculus, is valid: A sequence is convergent if and only if it is a Cauchy sequence. Every closed subspace of a complete metric space (considered as a metric space on its own) is complete. The converse statement is valid in a certain way: If a subspace Y of a (not necessary complete) metric space X is complete, then the set Y is closed in X.

■ Complete metric spaces are, e.g., the spaces: $\mathbf{m}$, $\mathbf{l}^p$ $(1 \le p < \infty)$, $\mathbf{c}$, $\mathcal{B}(T)$, $\mathcal{C}([a,b])$, $\mathcal{C}^{(k)}([a,b])$, $L^p(a,b)$ $(1 \le p < \infty)$.

12.2.2.3 Some Fundamental Theorems in Complete Metric Spaces

The importance of complete metric spaces can be illustrated by a series of theorems and principles, which are known and used in real calculus, and which are to be applied even in the case of infinite dimensional spaces.

1. Theorem on Nested Balls

Let X be a complete metric space. If

$$\overline{B}(x_1;r_1) \supset \overline{B}(x_2;r_2) \supset \cdots \supset \overline{B}(x_n;r_n) \supset \cdots \tag{12.57}$$

is a sequence of nested closed balls with $r_n \longrightarrow 0$, then the intersection of all of those balls is non-empty and consists of only a single point. If this property is valid in some metric space for any sequence satisfying the assumptions, then the metric space is complete.

2. Baire Category Theorem

Let X be a complete metric space and $\{F_k\}_{k=1}^{\infty}$ a sequence of closed sets in X with $\bigcup\limits_{k=1}^{\infty} F_k = \mathrm{X}$. Then there exists at least one index k_0 such that the set F_{k_0} has an interior point.

3. Banach Fixed-Point Theorem

Let F be a non-empty closed subset of a complete metric space (X, ρ). Let T: $X \longrightarrow \mathrm{X}$ be a contracting operator on F, i.e., there exists a constant $q \in [0,1)$ such that

$$\rho(Tx,Ty) \le q\,\rho(x,y) \quad \text{for all} \quad x,y \in F. \tag{12.58}$$

Suppose, if $x \in F$, then $Tx \in F$. Then the following statements are valid:

a) For an arbitrary initial point $x_0 \in F$ the iteration

$$x_{n+1} := Tx_n \quad (n = 0,1,2,\ldots) \tag{12.59}$$

is well defined, i.e., $x_n \in F$ for every n.

b) The iteration sequence $\{x_n\}_{n=0}^{\infty}$ converges to an element $x^* \in F$.

c) $Tx^* = x^*$, i.e., x^* is a fixed point of the operator T. (12.60)

d) The only fixed point of T in F is x^*.

e) The following error estimation is valid:

$$\rho(x^*, x_n) \le \frac{q^n}{1-q}\rho(x_1, x_0). \tag{12.61}$$

The *Banach fixed-point theorem* is sometimes called the *contraction mapping principle*.

12.2.2.4 Some Applications of the Contraction Mapping Principle

1. Iteration Method for Solving a System of Linear Equations

The given linear (n,n) system of equations

$$\begin{matrix} a_{11}x_1 & +a_{12}x_2 & +\ldots+ & a_{1n}x_1 & = b_1, \\ a_{21}x_1 & +a_{22}x_2 & +\ldots+ & a_{2n}x_n & = b_2, \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{n1}x_1 & +a_{n2}x_2 & +\ldots+ & a_{nn}x_n & = b_n \end{matrix} \tag{12.62a}$$

can be transformed according to 19.2.1, p. 955, into the equivalent system

$$\begin{array}{llllll} x_1 & -(1-a_{11})x_1 & +a_{12}x_2 & +\cdots & +a_{1n}x_n & = b_1, \\ x_2 & +a_{21}x_1 & -(1-a_{22})x_2 & +\cdots & +a_{2n}x_n & = b_2, \\ \multicolumn{6}{c}{\dotfill} \\ x_n & +a_{n1}x_1 & +a_{n2}x_2 & +\cdots & -(1-a_{nn})x_n & = b_n. \end{array} \tag{12.62b}$$

If the operator $T\colon\ \mathbb{F}^n \to \mathbb{F}^n$ is defined by

$$Tx = \left(x_1 - \sum_{k=1}^{n} a_{1k}x_k + b_1, \ldots, x_n - \sum_{k=1}^{n} a_{nk}x_k + b_n \right)^{\mathrm{T}}, \tag{12.63}$$

then the last system is transformed into the fixed-point problem

$$x = Tx \tag{12.64}$$

in the metric space $\mathbb{F}^n$, where an appropriate metric is considered: The Euclidean (12.43), the maximum (12.44) or the absolute value metric $\rho(x,y) = \sum_{k=1}^{n} |x_k - y_k|$ (compare with (12.45)). If one of the numbers

$$\sqrt{\sum_{j,k=1}^{n} |a_{jk}|^2}, \quad \max_{1\le j\le n} \sum_{k=1}^{n} |a_{jk}|, \quad \max_{1\le k\le n} \sum_{j=1}^{n} |a_{jk}| \tag{12.65}$$

is smaller than one, then T turns out to be a contracting operator. It has exactly one fixed point according to the Banach fixed-point theorem, which is the componentwise limit of the iteration sequence started from an arbitrary point of $\mathbb{F}^n$.

2. Fredholm Integral Equations

The Fredholm integral equation of second kind (see also 11.2, p. 622)

$$\varphi(x) - \int_a^b K(x,y)\varphi(y)\,dy = f(x), \quad x \in [a,b] \tag{12.66}$$

with a continuous kernel $K(x,y)$ and continuous right-hand side $f(x)$ can be solved by iteration. By means of the operator $T\colon\ \mathcal{C}([a,b]) \longrightarrow \mathcal{C}([a,b])$ defined as

$$T\varphi(x) = \int_a^b K(x,y)\varphi(y)\,dy + f(x) \quad \forall\, \varphi \in \mathcal{C}([a,b]), \tag{12.67}$$

it is transformed into a fixed-point problem $T\varphi = \varphi$ in the metric space $\mathcal{C}([a,b])$ (see example **A** in 12.1.2, **4.**, p. 655). If $\max_{a\le x\le b} \int_a^b |K(x,y)|\,dy < 1$, then T is a contracting operator and the fixed-point theorem can be applied. The unique solution is now obtained as the uniform limit of the iteration sequence $\{\varphi_n\}_{n=1}^{\infty}$, where $\varphi_n = T\varphi_{n-1}$, starting with an arbitrary function $\varphi_0(x) \in \mathcal{C}([a,b])$. It is clear that $\varphi_n = T^n\varphi_0$ and the iteration sequence is $\{T^n\varphi_0\}_{n=1}^{\infty}$.

3. Volterra Integral Equations

The Volterra integral equation of second kind (see 11.4, p. 643)

$$\varphi(x) - \int_a^x K(x,y)\varphi(y)\,dy = f(x), \quad x \in [a,b] \tag{12.68}$$

with a continuous kernel and a continuous right-hand side can be solved by means of the Volterra integral operator

$$(V\varphi)(x) := \int_a^x K(x,y)\varphi(y)\,dy \quad \forall\, \varphi \in \mathcal{C}([a,b]) \tag{12.69}$$

and $T\varphi = f + V\varphi$ as the fixed-point problem $T\varphi = \varphi$ in the space $\mathcal{C}([a,b])$.

4. Picard-Lindelöf Theorem

Consider the differential equation

$$\dot{x} = f(t,x) \tag{12.70}$$

with a continuous mapping $f\colon I \times G \longrightarrow \mathbb{R}^n$, where I is an open interval of $\mathbb{R}$ and G is an open domain of $\mathbb{R}^n$. Suppose the function f satisfies a Lipschitz condition with respect to x (see 9.1.1.1, **2.** p. 541), i.e., there is a positive constant L such that

$$\varrho(f(t,x_1), f(t,x_2)) \le L\varrho(x_1,x_2) \quad \forall\,(t,x_1),(t,x_2) \in I \times G, \tag{12.71}$$

where ϱ is the Euclidean metric in $\mathbb{R}^n$. (Using the norm (see 12.3.1, p. 669) and the formula (12.81) $\varrho(x,y) = \|x-y\|$ (12.71) can be written as $\|f(t,x_1) - f(t,x_2)\| \le L \cdot \|x_1 - x_2\|$.) Let $(t_0,x_0) \in I \times G$. Then there are numbers $\beta > 0$ and $r > 0$ such that the set $\Omega = \{(t,x) \in \mathbb{R}\times\mathbb{R}^n\colon |t-t_0| \le \beta,\ \varrho(x,x_0) \le r\}$ lies in $I \times G$. Let $M = \max_\Omega \varrho(f(t,x),0)$ and $\alpha = \min\{\beta, \frac{r}{M}\}$. Then there is a number $b > 0$ such that for each $\tilde{x} \in B = \{x \in \mathbb{R}^n\colon \varrho(x,x_0) \le b\}$, the initial value problem

$$\dot{x} = f(t,x), \quad x(t_0) = \tilde{x} \tag{12.72}$$

has exactly one solution $\varphi(t,\tilde{x})$, i.e., $\dot{\varphi}(t,\tilde{x}) = f(t,\varphi(t,\tilde{x}))$ for $\forall t$ satisfying $|t-t_0| \le \alpha$ and $\varphi(t_0,\tilde{x}) = \tilde{x}$. The solution of this initial value problem is equivalent to the solution of the integral equation

$$\varphi(t,\tilde{x}) = \tilde{x} + \int_{t_0}^{t} f(s,\varphi(s,\tilde{x}))\,ds, \quad t \in [t_0 - \alpha, t_0 + \alpha]. \tag{12.73}$$

If X denotes the closed ball $\{\varphi(t,x)\colon d(\varphi(t,x),x_0) \le r\}$ in the complete metric space $\mathcal{C}([t_0 - \alpha, t_0 + \alpha] \times B;\ \mathbb{R}^n)$ with metric

$$d(\varphi,\psi) = \max_{(t,x)\in\{|t-t_0|\le\alpha\}\times B} \varrho(\varphi(t,x),\psi(t,x)), \tag{12.74}$$

then X is a complete metric space with the induced metric. If the operator $T\colon \mathrm{X} \longrightarrow \mathrm{X}$ is defined by

$$T\varphi(t,x) = \tilde{x} + \int_{t_0}^{t} f(s,\varphi(s,\tilde{x}))\,ds \tag{12.75}$$

then T is a contracting operator and the solution of the integral equation (12.73) is the unique fixed point of T which can be calculated by iteration.

12.2.2.5 Completion of a Metric Space

Every (non-complete) metric space X can be completed; more precisely, there exists a metric space $\tilde{\mathrm{X}}$ with the following properties:

a) $\tilde{\mathrm{X}}$ contains a subspace Y isometric to X (see 12.2.3, **2.**, p. 669).

b) Y is everywhere dense in $\tilde{\mathrm{X}}$.

c) $\tilde{\mathrm{X}}$ is a complete metric space.

d) If Z is any metric space with the properties **a)**–**c)**, then Z and $\tilde{\mathrm{X}}$ are isometric.

The complete metric space, defined uniquely in this way up to isometry, is called the *completion* of the space X.

12.2.3 Continuous Operators

1. Continuous Operators

Let $T\colon \mathrm{X} \longrightarrow \mathrm{Y}$ be a mapping of the metric space $\mathrm{X} = (\mathrm{X},\rho)$ into the metric space $\mathrm{Y} = (\mathrm{Y},\varrho)$. T is called *continuous at the point* $x_0 \in \mathrm{X}$ if for every neighborhood $V = V(y_0)$ of the point $y_0 = T(x_0)$

there is a neighborhood $U = U(x_0)$ such that:

$$T(x) \in V \quad \text{for all} \quad x \in U. \tag{12.76}$$

T is called *continuous on the set* $A \subset \mathrm{X}$ if T is continuous at every point of A. Equivalent properties for T to be continuous on X are:

a) For any point $x \in \mathrm{X}$ and any arbitrary sequence $\{x_n\}_{n=1}^{\infty}$, $x_n \in \mathrm{X}$ with $x_n \longrightarrow x$ there always holds $T(x_n) \longrightarrow T(x)$. Hence $\rho(x_n, x) \to 0$ implies $\varrho(T(x_n), T(x)) \to 0$.

b) For any open subset $G \subset \mathrm{Y}$ the inverse image $T^{-1}(G)$ is an open subset in X.

c) For any closed subset $F \subset \mathrm{Y}$ the inverse image $T^{-1}(F)$ is a closed subset in X.

d) For any subset $A \subset \mathrm{X}$ one has $T(\overline{A}) \subset \overline{T(A)}$.

2. Isometric Spaces

If there is a bijective mapping $T\colon \mathrm{X} \longrightarrow \mathrm{Y}$ for two metric spaces $\mathrm{X} = (\mathrm{X}, \rho)$ and $\mathrm{Y} = (\mathrm{Y}, \varrho)$ such that

$$\rho(x, y) = \varrho(T(x), T(y)) \quad \forall\, x, y \in \mathrm{X}, \tag{12.77}$$

then the spaces X and Y are called *isometric*, and T is called an *isometry*.

12.3 Normed Spaces

12.3.1 Notion of a Normed Space

12.3.1.1 Axioms of a Normed Space

Let X be a vector space over the field $\mathbb{F}$. A function $\|\cdot\|\colon \mathrm{X} \longrightarrow \mathbb{R}^1_+$ is called a *norm* on the vector space X and the pair $\mathrm{X} = (\mathrm{X}, \|\cdot\|)$ is called a *normed space* over the field $\mathbb{F}$, if for arbitrary elements $x, y \in \mathrm{X}$ and for any scalar $\alpha \in \mathbb{F}$ the following properties, the so-called *axioms of a normed space*, are fulfilled:

(N1) $\|x\| \geq 0, \quad \text{and} \quad \|x\| = 0 \quad \text{if and only if } x = 0,$ (12.78)

(N2) $\|\alpha x\| = |\alpha| \cdot \|x\|$ (homogenity), (12.79)

(N3) $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality). (12.80)

A metric can be introduced by means of

$$\rho(x, y) = \|x - y\|, \quad x, y \in \mathrm{X}, \tag{12.81}$$

in any normed space. The metric (12.81) has the following additional properties which are compatible with the structure of the vector space:

$$\rho(x + z, y + z) = \rho(x, y), \qquad z \in X \tag{12.82a}$$

$$\rho(\alpha x, \alpha y) = |\alpha|\rho(x, y), \quad \alpha \in \mathbb{F}. \tag{12.82b}$$

So, in a normed space there are available both the properties of a vector space and the properties of a metric space. These properties are compatible in the sense of (12.82a) and (12.82b). The advantage is that most of the local investigations can be restricted to the *unit ball*

$$B(0; 1) = \{x \in \mathrm{X} : \|x\| < 1\} \quad \text{or} \quad \overline{B}(0; 1) = \{x \in \mathrm{X} : \|x\| \leq 1\} \tag{12.83}$$

since

$$B(x; r) = \{y \in \mathrm{X} : \|y - x\| < r\} = x + rB(0; 1), \quad \forall\, x \in X \quad \text{and} \quad \forall\, r > 0. \tag{12.84}$$

Moreover, the algebraic operations in a vector space are continuous, i.e.,

$$x_n \to x, \quad y_n \to y, \quad \alpha_n \to \alpha \quad \text{imply}$$
$$x_n + y_n \to x + y, \quad \alpha_n x_n \to \alpha x, \quad \|x_n\| \to \|x\|. \tag{12.85}$$

In normed spaces instead of (12.53) one may write for convergent sequences

$$\|x_n - x_0\| \longrightarrow 0 \quad (n \to \infty). \tag{12.86}$$

12.3.1.2 Some Properties of Normed Spaces

Among the linear metric spaces, those spaces are *normable* (i.e., a norm can be introduced by means of the metric, if one defines $\|x\| = \rho(x, 0)$) whose metric satisfies the conditions (12.82a) and (12.82b).

Two normed spaces X and Y are called *norm isomorphic* if there is a bijective linear mapping $T\colon \mathrm{X} \longrightarrow \mathrm{Y}$ with $\|Tx\| = \|x\|$ for all $x \in \mathrm{X}$. Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be two norms on the vector space X, and denote the corresponding normed spaces by X_1 and X_2, i.e., $\mathrm{X}_1 = (\mathrm{X}, \|\cdot\|_1)$ and $\mathrm{X}_2 = (X, \|\cdot\|_2)$.

The norm $\|\cdot\|_1$ is *stronger* than the norm $\|\cdot\|_2$, if there is a number $\gamma > 0$ such that $\|x\|_2 \le \gamma \|x\|_1$, for all $x \in \mathrm{X}$. In this case, the convergence of a sequence $\{x_n\}_{n=1}^{\infty}$ to x with respect to the stronger norm $\|\cdot\|_1$, i.e., $\|x_n - x\|_1 \longrightarrow 0$, implies the convergence to x with respect to the norm $\|\cdot\|_2$, i.e., $\|x_n - x\|_2 \longrightarrow 0$.

Two norms $\|\cdot\|$ and $\|\cdot\|_1$ are called *equivalent* if there are two numbers $\gamma_1 > 0,\ \gamma_2 > 0$ such that $\forall x \in \mathrm{X}$ there holds $\gamma_1\|x\| \le \|x\|_1 \le \gamma_2\|x\|$. In a finite dimensional vector space all norms are equivalent to each other.

A *subspace of a normed space* is a closed linear subspace of the space.

12.3.2 Banach Spaces

A complete normed space is called a *Banach space*. Every normed space X can be completed into a Banach space $\tilde{\mathrm{X}}$ by the completion procedure given in 12.2.2.5, p. 668, and by the natural extension of its algebraic operations and the norm to $\tilde{\mathrm{X}}$.

12.3.2.1 Series in Normed Spaces

In a normed space X *infinite series* can be considered. That means for a given sequence $\{x_n\}_{n=1}^{\infty}$ of elements $x_n \in \mathrm{X}$ a new sequence $\{s_k\}_{k=1}^{\infty}$ is constructed by

$$s_1 = x_1, s_2 = x_1 + x_2, \ldots, s_k = x_1 + \cdots + x_k = s_{k-1} + x_k, \ \ldots . \qquad (12.87)$$

If the sequence $\{s_k\}_{k=1}^{\infty}$ is convergent, i.e., $\|s_k - s\| \longrightarrow 0 \ (k \to \infty)$ for some $s \in X$, then a convergent series is defined. The elements $s_1, s_2, \ldots, s_k, \ldots$ are called the partial sums of the series. The limit

$$s = \lim_{k\to\infty} \sum_{n=1}^{k} x_n \qquad (12.88)$$

is the *sum* of the series, and it is denoted by $s = \sum\limits_{n=1}^{\infty} x_n$. A series $\sum\limits_{n=1}^{\infty} x_n$ is called *absolutely convergent* if the number series $\sum\limits_{n=1}^{\infty} \|x_n\|$ is convergent. In a Banach space every absolutely convergent series is convergent, and $\|s\| \le \sum\limits_{n=1}^{\infty} \|x_n\|$ holds for its sum s.

12.3.2.2 Examples of Banach Spaces

■ **A**: $\mathbb{F}^n$ with $\|x\| = \left(\sum\limits_{k=1}^{n} |\xi_k|^p\right)^{\frac{1}{p}}$, if $1 \le p < \infty$; $\|x\| = \max\limits_{1\le k\le n} |\xi_k|$, if $p = \infty$. (12.89a)

These normed spaces over the same vector space $\mathbb{F}^n$ are often denoted by $\mathbf{l}^p(n)$ $(1 \le p \le \infty)$. For $1 \le p < \infty$, they are called Euclidean spaces in the case of $\mathbb{F} = \mathbb{R}$, and *unitary spaces* in the case of $\mathbb{F} = \mathbb{C}$.

■ **B**: $\mathbf{m}$ with $\|x\| = \sup\limits_{k} |\xi_k|$. (12.89b)

■ **C**: $\mathbf{c}$ and $\mathbf{c_0}$ with the norm from $\mathbf{m}$. (12.89c)

■ **D**: $\mathbf{l}^p$ with $\|x\| = \|x\|_p = \left(\sum\limits_{n=1}^{\infty} |\xi_n|^p\right)^{\frac{1}{p}}$ $(1 \le p < \infty)$. (12.89d)

■ **E** : $\mathcal{C}([a,b])$ with $\|x\| = \max_{t\in[a,b]} |x(t)|$. (12.89e)

■ **F** : $L^p((a,b))$ $(1 \le p < \infty)$ with $\|x\| = \|x\|_p = \left(\int_a^b |x(t)|^p\,dt\right)^{\frac{1}{p}}$. (12.89f)

■ **G** : $\mathcal{C}^{(k)}([a,b])$ with $\|x\| = \sum_{l=0}^{k} \max_{t\in[a,b]} |x^{(l)}(t)|$, where $x^{(0)}(t)$ stands for $x(t)$. (12.89g)

12.3.2.3 Sobolev Spaces

Let $\Omega \subset \mathbb{R}^n$ be a bounded domain, i.e., an open connected set, with a sufficiently smooth boundary $\partial\Omega$. For $n = 1$ or $n = 2, 3$ one can imagine Ω being something similar to an interval (a, b) or a bounded convex set.
A function $f\colon \overline{\Omega} \longrightarrow \mathbb{R}$ is k-times continuously differentiable on the closed domain $\overline{\Omega}$ if f is k-times continuously differentiable on Ω and each of its partial derivatives has a finite limit on the boundary, i.e., if x approaches an arbitrary point of $\partial\Omega$. In other words, all partial derivatives can be continuously extended on the boundary of Ω, i.e., each partial derivative is a continuous function on $\overline{\Omega}$. In this vector space (for $p \in [1, \infty)$) and with the Lebesgue measure λ in $\mathbb{R}^n$ (see example **C** in 12.9.1, **2.**, p. 695) the following norm is defined:

$$\|f\|_{k,p} = \|f\| = \left(\int_{\Omega} |f(x)|^p\,d\lambda + \sum_{1\le|\alpha|\le k} \int_{\Omega} |D^\alpha f|^p\,d\lambda\right)^{\frac{1}{p}}. \tag{12.90}$$

The resulting normed space is denoted by $\tilde{W}^{k,p}(\Omega)$ or also by $\tilde{W}^k_p(\Omega)$ (in contrast to the space $\mathcal{C}^{(k)}([a,b])$ which has a quite different norm). Here α means a *multi-index*, i.e., an ordered n-tuple $(\alpha_1, \ldots, \alpha_n)$ of non-negative integers, where the sum of the components of α is denoted by $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_n$. For a function $f(x) = f(\xi_1, \ldots, \xi_n)$ with $x = (\xi_1, \ldots, \xi_n) \in \overline{\Omega}$ the brief notation is used as in (12.90):

$$D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial\xi_1^{\alpha_1} \cdots \partial\xi_n^{\alpha_n}}. \tag{12.91}$$

The normed space $\tilde{W}^{k,p}(\Omega)$ is not complete. Its completion is denoted by $W^{k,p}(\Omega)$ or in the case of $p = 2$ by $\mathbb{H}^k(\Omega)$ and it is called a *Sobolev space*.

12.3.3 Ordered Normed Spaces

1. Cones in a Normed Space

Let X be a real normed space with the norm $\|\cdot\|$. A cone $\mathrm{X}_+ \subset \mathrm{X}$ (see 12.1.4.2, p. 657) is called *solid*, if X_+ contains a ball (with positive radius), or equivalently, X_+ contains at least one interior point.

■ The usual cones are solid in the spaces $\mathbb{R}$, $\mathcal{C}([a,b])$, $\mathbf{c}$, but in the spaces $L^p((a,b))$ and $\mathbf{l}^p$ $(1 \le p < \infty)$ they are not solid.

A cone X_+ is called *normal* if the norm in X is *semi-monotonic*, i.e., there exists a constant $M > 0$ such that

$$0 \le x \le y \implies \|x\| \le M\|y\|. \tag{12.92}$$

If X is a Banach space ordered by a cone X_+, then every (o)-interval is bounded with respect to the norm if and only if the cone X_+ is normal.

■ The cones of the vectors with non-negative components and of the non-negative functions in the spaces $\mathbb{R}^n$, $\mathbf{m}$, $\mathbf{c}$, $\mathbf{c_0}$, $\mathcal{C}$, $\mathbf{l}^p$ and L^p, respectively, are normal.

A cone is called *regular* if every monotonically increasing sequence which is bounded above,

$$x_1 \le x_2 \le \cdots \le x_n \le \cdots \le z \tag{12.93}$$

is a Cauchy sequence in X. In a Banach space every closed regular cone is normal.

■ The cones in $\mathbb{R}^n$, $\mathbf{l}^p$ and L^p for $1 \le p < \infty$ are regular, but in $\mathcal{C}$ and $\mathbf{m}$ they are not.

2. Normed Vector Lattices and Banach Lattices

Let X be a vector lattice, which is a normed space at the same time. X is called a *normed lattice* or *normed vector lattice* (see [12.15], [12.19], [12.22], [12.23]), if the norm satisfies the condition

$$|x| \le |y| \quad \text{implies} \quad \|x\| \le \|y\| \quad \forall\, x, y \in \mathrm{X} \quad \text{(monotonicity of the norm).} \qquad (12.94)$$

A complete (with respect to the norm) normed lattice is called a *Banach lattice*.

■ The spaces $\mathcal{C}([a,b])$, L^p, $\mathbf{l}^p$, $\mathcal{B}([a,b])$ are Banach lattices.

12.3.4 Normed Algebras

A vector space X over $\mathbb{F}$ is called an *algebra*, if in addition to the operations defined in the vector space X and satisfying the axioms **(V1)**–**(V8)** (see 12.1.1, p. 654), a product $x \cdot y \in \mathrm{X}$ is defined for every two elements $x, y \in \mathrm{X}$ (or with a simplified notation by a product xy), such that for arbitrary $x, y, z \in \mathrm{X}$ and $\alpha \in \mathbb{F}$ the following conditions are satisfied:

(**A1**) $x(yz) = (xy)z,$ (12.95)

(**A2**) $x(y+z) = xy + xz,$ (12.96)

(**A3**) $(x+y)z = xz + yz,$ (12.97)

(**A4**) $\alpha(xy) = (\alpha x)y = x(\alpha y).$ (12.98)

An algebra is *commutative* if $xy = yx$ holds for two arbitrary elements x, y. A linear operator (see (12.21), p. 658) $T\colon \mathrm{X} \longrightarrow \mathrm{Y}$ of the algebra X into the algebra Y is called an *algebra homomorphism* if for any $x_1, x_2 \in \mathrm{X}$:

$$T(x_1 \cdot x_2) = Tx_1 \cdot Tx_2. \qquad (12.99)$$

An algebra X is called a *normed algebra* or a *Banach algebra* if it is a normed vector space or a Banach space and the norm has the additional property

$$\|x \cdot y\| \le \|x\| \cdot \|y\|. \qquad (12.100)$$

In a normed algebra all the operations are continuous, i.e., additionally to (12.85), if $x_n \longrightarrow x$ and $y_n \longrightarrow y$, then also $x_n y_n \longrightarrow xy$ (see [12.20]).

Every normed algebra can be completed to a Banach algebra, where the product is extended to the norm completion with respect to (12.100).

■ **A:** $\mathcal{C}([a,b])$ with the norm (12.89e) and the usual (pointwise) product of continuous functions.

■ **B:** The vector space $W([0,2\pi])$ of all complex-valued functions $x(t)$ continuous on $[0,2\pi]$ and having an absolutely convergent Fourier series expansion, i.e.,

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{int}, \qquad (12.101)$$

with the norm $\|x\| = \sum\limits_{n=-\infty}^{\infty} |c_n|$ and the usual multiplication.

■ **C:** The space $L(\mathrm{X})$ of all bounded linear operators on the normed space X with the operator norm and the usual algebraic operations (see 12.5.1.2, p. 677), where the product TS of two operators is defined as the sequential application, i.e., $TS(x) = T(S(x)),\ x \in \mathrm{X}$.

■ **D:** The space $L^1(-\infty, \infty)$ of all measurable and absolutely integrable functions on the real axis (see 12.9, p. 693) with the norm

$$\|x\| = \int_{-\infty}^{\infty} |x(t)|\, dt \qquad (12.102)$$

is a Banach algebra if the multiplication is defined as the convolution $(x * y)(t) = \int_{-\infty}^{\infty} x(t-s)y(s)\, ds$.

12.4 Hilbert Spaces

12.4.1 Notion of a Hilbert Space

12.4.1.1 Scalar Product

A vector space V over a field $\mathbb{F}$ (mostly $\mathbb{F} = \mathbb{C}$) is called a *space with scalar product* or an *inner product space* or *pre-Hilbert space* if to every pair of elements $x, y \in \mathrm{V}$ there is assigned a number $(x, y) \in \mathbb{F}$ (the scalar product of x and y), such that the *axioms of the scalar product* are satisfied, i.e., for arbitrary $x, y, z \in \mathrm{V}$ and $\alpha \in \mathbb{F}$:

(H1) $(x, x) \geq 0$, (i.e., (x, x) is real), and $(x, x) = 0$ if and only if $x = 0$, (12.103)

(H2) $(\alpha x, y) = \alpha(x, y)$, (12.104)

(H3) $(x + y, z) = (x, z) + (y, z)$, (12.105)

(H4) $(x, y) = \overline{(y, x)}$. (12.106)

(Here $\overline{\omega}$ denotes the conjugate of the complex number ω, which is denoted by ω^* in (1.133c). Sometimes the notation of a scalar product is $\langle x, y\rangle$.)
In the case of $\mathbb{F} = \mathbb{R}$, i.e., in a real vector space, **(H4)** means the commutativity of the scalar product. Some further properties follow from the axioms:

$$(x, \alpha y) = \bar{\alpha}(x, y) \quad \text{and} \quad (x, y + z) = (x, y) + (x, z). \tag{12.107}$$

12.4.1.2 Unitary Spaces and Some of their Properties

In a pre-Hilbert space $\mathbb{H}$ a norm can be introduced by means of the scalar product as follows:

$$\|x\| = \sqrt{(x, x)} \quad (x \in \mathbb{H}). \tag{12.108}$$

A normed space $\mathbb{H} = (\mathbb{H}, \|\cdot\|)$ is called *unitary* if there is a scalar product satisfying (12.108). Based on the previous properties of scalar products and (12.108) in unitary spaces the following facts are valid:

a) Triangle Inequality:

$$\|x + y\| \leq \|x\| + \|y\|. \tag{12.109}$$

b) Cauchy-Schwarz Inequality or Schwarz-Buniakowski Inequality (see also 1.4.2.9, p. 31):

$$|(x, y)| \leq \sqrt{(x, x)}\sqrt{(y, y)}\,. \tag{12.110}$$

c) Parallelogram Identity: This characterizes the unitary spaces among the normed spaces:

$$\|x + y\|^2 + \|x - y\|^2 = 2\left(\|x\|^2 + \|y\|^2\right). \tag{12.111}$$

d) Continuity of the Scalar Product:

$$x_n \to x,\ y_n \to y \quad \text{imply} \quad (x_n, y_n) \to (x, y). \tag{12.112}$$

12.4.1.3 Hilbert Space

A complete unitary space is called a *Hilbert space*. Since Hilbert spaces are also Banach spaces, they possess their properties (see 12.3.1, p. 669; 12.3.1.2, p. 670; 12.3.2, p. 670). In addition they have the properties of unitary spaces 12.4.1.2, p. 673. A subspace of a Hilbert space is a closed linear subspace.

■ **A:** $\mathbf{l}^2(n)$, $\mathbf{l}^2$ and $L^2((a, b))$ with the scalar products

$$(x, y) = \sum_{k=1}^{n} \xi_k\overline{\eta_k},\ (x, y) = \sum_{k=1}^{\infty} \xi_k\overline{\eta_k} \quad \text{and} \quad (x, y) = \int_a^b x(t)\overline{y(t)}\,dt. \tag{12.113}$$

■ **B:** The space $\mathbb{H}^2(\Omega)$ with the scalar product

$$(f, g) = \int\limits_{\Omega} f(x)\overline{g(x)}\,dx + \sum_{1\leq|\alpha|\leq k}\int\limits_{\Omega} D^\alpha f(x)\overline{D^\alpha g(x)}\,dx. \tag{12.114}$$

■ **C:** Let $\varphi(t)$ be a measurable positive function on $[a,b]$. The complex space $L^2((a,b),\varphi)$ of all measurable functions, which are quadratically integrable with the *weight* function φ on (a,b), is a Hilbert space if the scalar product is defined as

$$(x,y) = \int_a^b x(t)\overline{y(t)}\varphi(t)\,dt. \tag{12.115}$$

12.4.2 Orthogonality

Two elements x, y of a Hilbert space $\mathbb{H}$ are called *orthogonal* (denoted by $x \perp y$) if $(x,y) = 0$ (the notions of this paragraph also make sense in pre-Hilbert spaces and in unitary spaces). For an arbitrary subset $A \subset \mathbb{H}$, the set

$$A^{\perp} = \{x \in \mathbb{H}\colon (x,y) = 0 \quad \forall\, y \in A\} \tag{12.116}$$

of all vectors which are orthogonal to each vector in A is a (closed linear) subspace of $\mathbb{H}$ and it is called the *orthogonal space* to A or the *orthogonal complement* of A. The notation $A \perp B$ means that $(x,y) = 0$ for all $x \in A$ and $y \in B$. If A consists of a single element x, then the notation $x \perp B$ is used.

12.4.2.1 Properties of Orthogonality

The zero vector is orthogonal to every vector of $\mathbb{H}$. The following statements hold:

a) $x \perp y$ and $x \perp z$ imply $x \perp (\alpha y + \beta z)$ for any $\alpha, \beta \in \mathbb{C}$.

b) From $x \perp y_n$ and $y_n \to y$ it follows that $x \perp y$.

c) $x \perp A$ if and only if $x \perp \overline{lin(A)}$, where $\overline{lin(A)}$ denotes the *closed linear hull* of the set A.

d) If $x \perp A$ and A is a *fundamental set*, i.e., $lin(A)$ is everywhere dense in $\mathbb{H}$, then $x = 0$.

e) Pythagoras Theorem: If the elements $x_1, \ldots, x_n$ are pairwise orthogonal, that is $x_k \perp x_l$ for all $k \neq l$, then

$$\|\sum_{k=1}^{n} x_k\|^2 = \sum_{k=1}^{n} \|x_k\|^2. \tag{12.117}$$

f) Projection Theorem: If $\mathbb{H}_0$ is a subspace of $\mathbb{H}$, then each vector $x \in \mathbb{H}$ can be written uniquely as

$$x = x' + x'', \quad x' \in \mathbb{H}_0,\ x'' \perp \mathbb{H}_0. \tag{12.118}$$

g) Approximation Problem: Furthermore, the equation $\|x'\| = \rho(x, \mathbb{H}_0) = \inf_{y\in \mathbb{H}_0}\{\|x-y\|\}$ holds, and so the problem

$$\|x - y\| \to \inf, \quad y \in \mathbb{H}_0 \tag{12.119}$$

has the unique solution x' in $\mathbb{H}_0$. In this statement $\mathbb{H}_0$ can be replaced by a convex closed non-empty subset of $\mathbb{H}$.

The element x' is called the *projection* of the element x on $\mathbb{H}_0$. It has the smallest distance from x (to $\mathbb{H}_0$), and the space $\mathbb{H}$ can be decomposed: $\mathbb{H} = \mathbb{H}_0 \oplus \mathbb{H}_0^{\perp}$.

12.4.2.2 Orthogonal Systems

A set $\{x_\xi\colon \xi \in \Xi\}$ of vectors from $\mathbb{H}$ is called an *orthogonal* system if it does not contain the zero vector and $x_\xi \perp x_\eta$, $\xi \neq \eta$, hence $(x_\xi, x_\eta) = \delta_{\xi\eta}$ holds, where

$$\delta_{\xi\eta} = \begin{cases} 1 & \text{for } \xi = \eta, \\ 0 & \text{for } \xi \neq \eta \end{cases} \tag{12.120}$$

denotes the Kronecker symbol (see 4.1.2, **10.**, p. 271). An orthogonal system is called *orthonormal* if in addition $\|x_\xi\| = 1\ \forall\, \xi$.

In a separable Hilbert space an orthogonal system may contain at most countably many elements. Therefore $\Xi = \mathbb{N}$ is assumed from now on.

■ **A:** The system

$$\frac{1}{\sqrt{2\pi}},\ \frac{1}{\sqrt{\pi}}\cos t,\ \frac{1}{\sqrt{\pi}}\sin t,\ \frac{1}{\sqrt{\pi}}\cos 2t,\ \frac{1}{\sqrt{\pi}}\sin 2t,\ \ldots \tag{12.121}$$

in the real space $L^2((-\pi,\pi))$ and the system

$$\frac{1}{\sqrt{2\pi}}e^{int} \quad (n = 0, \pm 1, \pm 2, \ldots) \tag{12.122}$$

in the complex space $L^2((-\pi,\pi))$ are orthonormal systems. Both of these systems are called *trigonometric*.

■ **B:** The Legendre polynomials of the first kind (see 9.1.2.6, **2.**, p. 566)

$$P_n(t) = \frac{d^n}{d\,t^n}[(t^2-1)^n] \quad (n = 0, 1, \ldots) \tag{12.123}$$

form an orthogonal system of elements in the space $L^2((-1,1))$. The corresponding orthonormal system is

$$\tilde{P}_n(t) = \sqrt{n + \frac{1}{2}}\,\frac{1}{(2n)!!}\,P_n(t). \tag{12.124}$$

■ **C:** The Hermite polynomials (see 9.1.2.6, **6.**, p. 568 and 9.2.4, **3.**, 602) according to the second definition of the Hermite differential equation (9.66b)

$$H_n(t) = e^{t^2}\frac{d^n}{dt^n}e^{-t^2} \quad (n = 0, 1, \ldots) \tag{12.125}$$

form an orthogonal system in the space $L^2((-\infty,\infty))$.

■ **D:** The Laguerre polynomials form an orthogonal system (see 9.1.2.6, **5.**, p. 568) in the space $L^2((0,\infty))$.

Every orthogonal system is linearly independent, since the zero vector was excluded. Conversely, if $x_1, x_2, \ldots x_n, \ldots$ is a system of linearly independent elements in a Hilbert space $\mathbb{H}$, then there exist vectors $e_1, e_2, \ldots, e_n, \ldots$, obtained by the *Gram–Schmidt orthogonalization method* (see 4.6.2.2, **1.**, p. 316) which form an orthonormal system. They span the same subspace, and by the method they are determined up to a scalar factor with modulus 1.

12.4.3 Fourier Series in Hilbert Spaces

12.4.3.1 Best Approximation

Let $\mathbb{H}$ be a separable Hilbert space and

$$\{e_n\colon\ n = 1, 2, \ldots\} \tag{12.126}$$

a fixed orthonormal system in $\mathbb{H}$. For an element $x \in \mathbb{H}$ the numbers $c_n = (x, e_n)$ are called the *Fourier coefficients* of x with respect to the system (12.126). The (formal) series

$$\sum_{n=1}^{\infty} c_n e_n \tag{12.127}$$

is called the *Fourier series* of the element x with respect to the system (12.126) (see 7.4.1.1, **1.**, p. 474). The n-th partial sum of the Fourier series of an element x has the property of the *best approximation*, i.e., for fixed n, the n-th partial sum of the Fourier series

$$\sigma_n = \sum_{k=1}^{n}(x, e_k)e_k \tag{12.128}$$

gives the smallest value of $\|x - \sum_{k=1}^{n} \alpha_k e_k\|$ among all vectors of $\mathbb{H}_n = lin(\{e_1, \ldots, e_n\})$. Furthermore, $x - \sigma_n$ is orthogonal to $\mathbb{H}_n$, and there holds the *Bessel inequality*:

$$\sum_{n=1}^{\infty} |c_n|^2 \le \|x\|^2, \quad c_n = (x, e_n) \quad (n = 1, 2, \ldots). \tag{12.129}$$

12.4.3.2 Parseval Equation, Riesz-Fischer Theorem

The Fourier series of an arbitrary element $x \in \mathbb{H}$ is always convergent. Its sum is the projection of the element x onto the subspace $\mathbb{H}_0 = \overline{lin(\{e_n\}_{n=1}^{\infty})}$. If an element $x \in \mathbb{H}$ has the representation $x = \sum_{n=1}^{\infty} \alpha_n e_n$, then α_n are the Fourier coefficients of x $(n = 1, 2, \ldots)$. If $\{\alpha_n\}_{n=1}^{\infty}$ is an arbitrary sequence of numbers with the property $\sum_{n=1}^{\infty} |\alpha_n|^2 < \infty$, then there is a unique element x in $\mathbb{H}$, whose Fourier coefficients are equal to α_n and for which the Parseval equation holds:

$$\sum_{n=1}^{\infty} |(x, e_n)|^2 = \sum_{n=1}^{\infty} |\alpha_n|^2 = \|x\|^2 \qquad (\textit{Riesz–Fischer theorem}). \tag{12.130}$$

An orthonormal system $\{e_n\}$ in $\mathbb{H}$ is called *complete* if there is no non-zero vector y orthogonal to every e_n; it is called a *basis* if every vector $x \in \mathbb{H}$ has the representation $x = \sum_{n=1}^{\infty} \alpha_n e_n$, i.e., $\alpha_n = (x, e_n)$ and x is equal to the sum of its Fourier series. In this case, one also says that x has a Fourier expansion. The following statements are equivalent:

a) $\{e_n\}$ is a fundamental set in $\mathbb{H}$.

b) $\{e_n\}$ is complete in $\mathbb{H}$.

c) $\{e_n\}$ is a basis in $\mathbb{H}$.

d) For $\forall\, x, y \in \mathbb{H}$ with the corresponding Fourier coefficients c_n and d_n $(n = 1, 2, \ldots)$ there holds

$$(x, y) = \sum_{n=1}^{\infty} c_n \overline{d_n}. \tag{12.131}$$

e) For every vector $x \in \mathbb{H}$, the Parseval equation (12.130) holds.

■ **A:** The trigonometric system (12.121) is a basis in the space $L^2((-\pi, \pi))$.

■ **B:** The system of the normalized Legendre polynomials (12.124) $\tilde{P}_n(t)$ $(n = 0, 1, \ldots)$ is complete and consequently a basis in the space $L^2((-1, 1))$.

12.4.4 Existence of a Basis, Isomorphic Hilbert Spaces

In every separable Hilbert space there exits a basis. From this fact it follows that every orthonormal system can be completed to a basis.

Two Hilbert spaces $\mathbb{H}_1$ and $\mathbb{H}_2$ are called *isometric* or *isomorphic* as Hilbert spaces if there is a linear bijective mapping T: $\mathbb{H}_1 \longrightarrow \mathbb{H}_2$ with the property $(Tx, Ty)_{\mathbb{H}_2} = (x, y)_{\mathbb{H}_1}$ (that is, it preserves the scalar product and because of (12.108) also the norm). Any two arbitrary infinite dimensional separable Hilbert spaces are isometric, in particular every such space is isometric to the separable space $\mathbf{l}^2$.

12.5 Continuous Linear Operators and Functionals

12.5.1 Boundedness, Norm and Continuity of Linear Operators

12.5.1.1 Boundedness and the Norm of Linear Operators

Let $\mathrm{X} = (\mathrm{X}, \|\cdot\|)$ and $\mathrm{Y} = (\mathrm{Y}, \|\cdot\|)$ be normed spaces. In the following discussion the index X in the notation $\|\cdot\|_{\mathrm{X}}$, which emphasizes being in the space X, is omitted, because from the text it will be always clear, which norms and spaces are considered. An arbitrary operator $T\colon\ \mathrm{X} \longrightarrow \mathrm{Y}$ is called *bounded* if there is a real number $\lambda > 0$ such that

$$\|T(x)\| \le \lambda \|x\| \quad \forall\, x \in \mathrm{X}. \tag{12.132}$$

A bounded operator with a constant λ "stretches" every vector at most λ times and it transforms every bounded set of X into a bounded set of Y, in particular the image of the unit ball of X is bounded in Y. This last property is characteristic of bounded linear operators. A linear operator is continuous (see 12.2.3, p. 668) if and only if it is bounded.

The smallest constant λ, for which (12.132) still holds, is called the *norm of the operator* T and it is denoted by $\|T\|$, i.e.,

$$\|T\| := \inf\{\lambda > 0 :\ \|Tx\| \le \lambda\|x\|,\ x \in \mathrm{X}\}. \tag{12.133}$$

For a continuous linear operator the following equalities hold:

$$\|T\| = \sup_{\|x\|\le 1} \|Tx\| = \sup_{\|x\|<1} \|Tx\| = \sup_{\|x\|=1} \|Tx\| \tag{12.134}$$

and, furthermore, the following estimation holds

$$\|Tx\| \le \|T\| \cdot \|x\| \quad \forall\, x \in \mathrm{X}. \tag{12.135}$$

■ Let T be the operator in the space $\mathcal{C}([a,b])$ with the norm (12.89e), defined by the integral

$$(Tx)(s) = y(s) = \int_a^b K(s,t)x(t)\,dt \quad (s \in [a,b]), \tag{12.136}$$

where $K(s,t)$ is a (complex-valued) continuous function on the rectangle $\{a \le s, t \le b\}$. Then T is a bounded linear operator, which maps $\mathcal{C}([a,b])$ into $\mathcal{C}([a,b])$. Its norm is

$$\|T\| = \max_{s\in[a,b]} \int_a^b |K(s,t)|\,dt. \tag{12.137}$$

12.5.1.2 The Space of Linear Continuous Operators

The sum $U = S + T$ and the multiple αT of two linear (continuous) operators $S, T\colon\ \mathrm{X} \longrightarrow \mathrm{Y}$ are defined point-wise:

$$U(x) = S(x) + T(x), \quad (\alpha T)(x) = \alpha \cdot T(x), \quad \forall\, x \in \mathrm{X} \text{ and } \forall\, \alpha \in \mathbb{F}. \tag{12.138}$$

The set $L(\mathrm{X},\mathrm{Y})$, often denoted by $B(\mathrm{X},\mathrm{Y})$, of all linear continuous operators T from X into Y equipped with the operations (12.138) is a vector space, where $\|T\|$ (12.133) turns out to be a norm on it. So, $L(\mathrm{X},\mathrm{Y})$ is a normed space and even a Banach space if Y is a Banach space. So the axioms **(V1)**–**(V8)** and **(N1)**–**(N3)** are satisfied (see 12.1.1, p. 654, 12.3.1, p. 669).

If $\mathrm{Y} = \mathrm{X}$, then a product can be defined for two arbitrary elements $S, T \in L(\mathrm{X},\mathrm{X}) = L(\mathrm{X}) = B(\mathrm{X})$ as

$$(ST)(x) = S(Tx) \quad (\forall\, x \in \mathrm{X}), \tag{12.139}$$

which satisfies the axioms **(A1)**–**(A4)** from 12.3.4, p. 672, and also the compatibility condition (12.100) with the norm. $L(\mathrm{X})$ is in general a non-commutative normed algebra, and if X is a Banach space, then it is a Banach algebra. Then for every operator $T \in L(\mathrm{X})$ its powers are defined by

$$T^0 = I,\ T^n = T^{n-1}T \quad (n = 1, 2, \ldots), \tag{12.140}$$

where I is the identity operator $Ix = x, \quad \forall\, x \in \mathrm{X}$. Then

$$\|T^n\| \le \|T\|^n \quad (n = 0, 1, \ldots), \tag{12.141}$$

and furthermore there always exists the (finite) limit

$$r(T) = \lim_{n\to\infty} \sqrt[n]{\|T^n\|}\,, \tag{12.142}$$

which is called the *spectral radius* of the operator T and satisfies the relations

$$r(T) \le \|T\|, \quad r(T^n) = [r(T)]^n, \quad r(\alpha T) = |\alpha| r(T), \quad r(T) = r(T^*), \tag{12.143}$$

where T^* is the adjoint operator to T (see 12.6, p. 684, and (12.159)). If $L(\mathrm{X})$ is complete, then for $|\lambda| > r(T)$, the operator $(\lambda I - T)^{-1}$ has the representation in the form of a Neumann series

$$(\lambda I - T)^{-1} = \lambda^{-1} I + \lambda^{-2} T + \ldots + \lambda^{-n} T^{n-1} + \ldots\,, \tag{12.144}$$

which is convergent for $|\lambda| > r(T)$ in the operator norm on $L(\mathrm{X})$.

12.5.1.3 Convergence of Operator Sequences

1. Point-wise Convergence

of a sequence of linear continuous operators $T_n\colon \mathrm{X} \longrightarrow \mathrm{Y}$ to an operator $T\colon \mathrm{X} \longrightarrow \mathrm{Y}$ means that:

$$T_n x \longrightarrow Tx \quad \text{in Y for each } x \in \mathrm{X}. \tag{12.145}$$

2. Uniform Convergence

The usual norm-convergence of a sequence of operators $\{T_n\}_{n=1}^{\infty}$ in a space $L(\mathrm{X}, \mathrm{Y})$ to T, i.e.,

$$\|T_n - T\| = \sup_{\|x\|\le 1} \|T_n x - Tx\| \longrightarrow 0 \quad (n \to \infty) \tag{12.146}$$

is the *uniform* convergence on the unit ball of X. It implies point-wise convergence, while the converse statement is not true in general.

3. Applications

The convergence of quadrature formulas when the number n of interpolation nodes tends to ∞, the performance principle of summation, limiting methods, etc.

12.5.2 Linear Continuous Operators in Banach Spaces

Now X and Y are supposed to be Banach spaces.

1. Banach-Steinhaus Theorem (Uniform Boundedness Principle)

The theorem characterizes the point-wise convergence of a sequence $\{T_n\}$ of linear continuous operators T_n to some linear continuous operator by the conditions:

a) For every element from an everywhere dense subset $D \subset \mathrm{X}$, the sequence $\{T_n x\}$ has a limit in Y,

b) there is a constant C such that $\|T_n\| \le C,\ \forall\, n$.

2. Open Mappings Theorem

The theorem tells us that a linear continuous operator mapping from X onto Y is *open*, i.e., the image $T(G)$ of every open set G from X is an open set in Y.

3. Closed Graph Theorem

An operator $T\colon D_T \longrightarrow \mathrm{Y}$ with $D_T \subset \mathrm{X}$ is called *closed* if $x_n \in D_T$, $x_n \to x_0$ in X and $Tx_n \to y_0$ in Y imply $x_0 \in D_T$ and $y_0 = Tx_0$. A necessary and sufficient condition is that the graph of the operator T in the space $\mathrm{X} \times \mathrm{Y}$, i.e., the set

$$\Gamma_T = \{(x, Tx)\colon\ x \in D_T\} \tag{12.147}$$

is closed, where here (x, y) denotes an element of the set $\mathrm{X} \times \mathrm{Y}$.
If T is a closed operator with a closed domain D_T, then T is continuous.

4. Hellinger-Toeplitz Theorem

Let T be a linear operator in a Hilbert space $\mathbb{H}$. If $(x, Ty) = (Tx, y)$ for every $x, y \in \mathbb{H}$, then T is continuous (here (x, Ty) denotes the scalar product in $\mathbb{H}$).

5. Krein-Losanovskij Theorem on the Continuity of Positive Linear Operators

If $\mathbb{X} = (\mathbb{X}, \mathbb{X}_+, \|\cdot\|)$ and $\mathbb{Y} = (\mathbb{Y}, \mathbb{Y}_+, \|\cdot\|)$ are ordered normed spaces, where $\mathbb{X}_+$ is a generating cone, then the set $L_+(\mathbb{X}, \mathbb{Y})$ of all positive linear and continuous operators T, i.e., $T(\mathbb{X}_+) \subset \mathbb{Y}_+$, is a cone in $L(\mathbb{X}, \mathbb{Y})$. The *theorem of Krein and Losanovskij* asserts (see [12.17]): If $\mathbb{X}$ and $\mathbb{Y}$ are ordered Banach spaces with closed cones $\mathbb{X}_+$ and $\mathbb{Y}_+$, and $\mathbb{X}_+$ is a generating cone, then the positivity of a linear operator implies its continuity.

6. Inverse Operator

Let $\mathbb{X}$ and $\mathbb{Y}$ be arbitrary normed spaces and let $T\colon \mathbb{X} \longrightarrow \mathbb{Y}$ be a linear, not necessarily continuous operator. T has a continuous inverse $T^{-1}\colon \mathbb{Y} \longrightarrow \mathbb{X}$, if $T(\mathbb{X}) = \mathbb{Y}$ and there exists a constant $m > 0$ such that $\|Tx\| \geq m\|x\|$ for each $x \in X$. Then $\|T^{-1}\| \leq \dfrac{1}{m}$. The situation considered here is less general than that in (12.22) (see 12.1.5.2, p. 659), since there may be $D \neq \mathbb{X}$ and $\mathcal{R}(T) \neq \mathbb{Y}$.

In the case of Banach spaces $\mathbb{X}, \mathbb{Y}$ the following theorem is valid:

7. Banach Theorem on the Continuity of the Inverse Operator

If T is a linear continuous bijective operator from $\mathbb{X}$ onto $\mathbb{Y}$, then the inverse operator T^{-1} is also continuous.

An important application is, e.g., the continuity of $(\lambda I - T)^{-1}$ given the injectivity and surjectivity of $\lambda I - T$. This fact has importance in investigating the spectrum of an operator (see 12.5.3.2, p. 680). It also applies to the

8. Continuous Dependence of the Solution

on the right-hand side and also on the initial data of initial value problems for linear differential equations. This fact is demonstrated by the following example.

■ The initial value problem

$$\ddot{x}(t) + p_1(t)\dot{x}(t) + p_2(t)x(t) = q(t), \quad t \in [a,b], \quad x(t_0) = \xi,\ \dot{x}(t_0) = \dot{\xi}, \quad t_0 \in [a,b] \tag{12.148a}$$

with coefficients $p_1(t), p_2(t) \in \mathcal{C}([a,b])$ has exactly one solution x from $\mathcal{C}^2([a,b])$ for every right-hand side $q(t) \in \mathcal{C}([a,b])$ and for every pair of numbers $\xi,\ \dot{\xi}$. The solution x depends continuously on $q(t),\ \xi$ and $\dot{\xi}$ in the following sense. If $q_n(t) \in \mathcal{C}([a,b]), \xi_n,\ \dot{\xi}_n \in \mathbb{R}^1$ are given and $x_n \in \mathcal{C}([a,b])$ denotes the solution of

$$\ddot{x}_n(t) + p_1(t)\dot{x}_n(t) + p_2(t)x_n(t) = q_n(t), \quad x_n(a) = \xi_n,\ \dot{x}_n(a) = \dot{\xi}_n, \tag{12.148b}$$

for $n = 1, 2, \ldots$, then:

$$\left.\begin{array}{r} q_n(t) \to q(t) \ \text{ in } \ \mathcal{C}([a,b]), \\ \xi_n \to \xi, \\ \dot{\xi}_n \to \dot{\xi}, \end{array}\right\} \quad \text{implies } x_n \to x \text{ in the space } \mathcal{C}^2([a,b]). \tag{12.148c}$$

9. Method of Successive Approximation

to solve an equation of the form

$$x - Tx = y \tag{12.149}$$

with a continuous linear operator T in a Banach space $\mathbb{X}$ for a given y. This method starts with an arbitrary initial element x_0, and constructs a sequence $\{x_n\}$ of approximating solutions by the formula

$$x_{n+1} = y + Tx_n \quad (n = 0, 1, \ldots). \tag{12.150}$$

This sequence converges to the solution x^* in $\mathbb{X}$ of (12.149). The convergence of the method, i.e., $x_n \to x^*$, is based on the convergence of the series (12.144) with $\lambda = 1$.

Let $\|T\| \leq q < 1$. Then the following statements are valid:

a) The operator $I - T$ has a continuous inverse with $\|(I - T)^{-1}\| \leq \dfrac{1}{1-q}$, and (12.149) has exactly one solution for each y.

b) The series (12.144) converges and its sum is the operator $(I - T)^{-1}$.

c) The method (12.150) converges to the unique solution x^* of (12.149) for any initial element x_0, if the series (12.144) converges. Then the following estimation holds:

$$\|x_n - x^*\| \le \frac{q^n}{1-q}\|Tx_0 - x_0\| \quad (n = 1, 2, \ldots). \tag{12.151}$$

Equations of the type

$$x - \mu Tx = y, \quad \lambda x - Tx = y, \quad \mu, \lambda \in \mathbb{F} \tag{12.152}$$

can be handled in an analogous way (see 11.2.2, p. 625, and [12.8]).

12.5.3 Elements of the Spectral Theory of Linear Operators

12.5.3.1 Resolvent Set and the Resolvent of an Operator

For an investigation of the solvability of equations one tries to rewrite the problem in the form

$$(I - T)x = y \tag{12.153}$$

with some operator T having a possible small norm. This is especially convenient for using a functional analytic method because of (12.143) and (12.144). In order to handle large values of $\|T\|$ as well, it is necessary to investigate the whole family of equations

$$(\lambda I - T)x = y \quad x \in \mathrm{X}, \ \text{with } \lambda \in \mathbb{C} \tag{12.154}$$

in a complex Banach space X. Let T be a linear, but in general not a bounded operator in a Banach space X. The set $\varrho(T)$ of all complex numbers λ such that $(\lambda I - T)^{-1} \in B(\mathrm{X}) = L(\mathrm{X})$ is called the *resolvent set* and the operator $R_\lambda = R_\lambda(T) = (\lambda I - T)^{-1}$ is called the *resolvent*. Let T now be a bounded linear operator in a complex Banach space X. Then the following statements are valid:

a) The set $\varrho(T)$ is open. More precisely, if $\lambda_0 \in \varrho(T)$ and $\lambda \in \mathbb{C}$ satisfy the inequality

$$|\lambda - \lambda_0| < \frac{1}{\|R_{\lambda_0}\|}, \tag{12.155}$$

then R_λ exists and

$$R_\lambda = R_{\lambda_0} + (\lambda - \lambda_0)R^2_{\lambda_0} + (\lambda - \lambda_0)^2 R^3_{\lambda_0} + \ldots = \sum_{k=1}^{\infty} (\lambda - \lambda_0)^{k-1} R^k_{\lambda_0}. \tag{12.156}$$

b) $\{\lambda \in \mathbb{C} : |\lambda| > \|T\|\} \subset \varrho(T)$. More precisely, $\forall\, \lambda \in \mathbb{C}$ with $|\lambda| > \|T\|$, the operator R_λ exists and

$$R_\lambda = -\frac{I}{\lambda} - \frac{T}{\lambda^2} - \frac{T^2}{\lambda^3} - \ldots. \tag{12.157}$$

c) $\|R_\lambda - R_{\lambda_0}\| \to 0$, if $\lambda \to \lambda_0 \quad (\lambda, \lambda_0 \in \varrho(T))$, and $\|R_\lambda\| \to 0$, if $\lambda \to \infty \quad (\lambda \in \varrho(T))$.

d) $\left\| \dfrac{R_\lambda - R_{\lambda_0}}{\lambda - \lambda_0} - R^2_{\lambda_0} \right\| \longrightarrow 0$, if $\lambda \to \lambda_0$.

e) For an arbitrary functional $f \in \mathrm{X}^*$ (see 12.5.4.1, p. 681) and arbitrary $x \in \mathrm{X}$ the function $F(\lambda) = f(R_\lambda(x))$ is holomorphic on $\varrho(T)$.

f) For arbitrary λ, $\mu \in \varrho(T)$, and $\lambda \ne \mu$ one has:

$$R_\lambda R_\mu = R_\mu R_\lambda = \frac{R_\lambda - R_\mu}{\lambda - \mu}. \tag{12.158}$$

12.5.3.2 Spectrum of an Operator

1. Definition of the Spectrum

The set $\sigma(T) = \mathbb{C} \setminus \varrho(T)$ is called the *spectrum* of the operator T. Since $I - T$ has a continuous inverse (and consequently (12.153) has a solution, which continuously depends on the right-hand side) if and only if $1 \in \varrho(T)$, the spectrum $\sigma(T)$ must be known as well as possible. From the properties of the resolvent set it follows immediately that the spectrum $\sigma(T)$ is a closed set of $\mathbb{C}$ which lies in the disk $\{\lambda \in \mathbb{C}\colon |\lambda| \le \|T\|\}$, however, in many cases $\sigma(T)$ is much smaller than this disk. The spectrum of any linear continuous operator on a complex Banach space is never empty and

$$r(T) = \sup_{\lambda\in\sigma(T)} |\lambda|. \tag{12.159}$$

It is possible to say more about the spectrum in the cases of different special classes of operators. If T is an operator in a finite dimensional space X and if the equation $(\lambda I - T)x = 0$ has only the trivial solution (i.e., $\lambda I - T$ is injective), then $\lambda \in \varrho(T)$ (i.e., $\lambda I - T$ is surjective). If this equation has a non-trivial solution in some Banach space, then the operator $\lambda I - T$ is not injective and $(\lambda I - T)^{-1}$ is in general not defined.

The number $\lambda \in \mathbb{C}$ is called an *eigenvalue* of the linear operator T, if the equation $\lambda x = Tx$ has a non-trivial solution. All those solutions are called *eigenvectors*, or in the case when X is a function space (which occurs very often in applications), they are called *eigenfunctions* of the operator T associated to λ. The subspace spanned by them is called the *eigenspace* (or *characteristic space*) associated to λ. The set $\sigma_p(T)$ of all eigenvalues of T is called the *point spectrum* of the operator T.

2. Comparison to Linear Algebra, Residual Spectrum

An essential difference between the finite dimensional case which is considered in linear algebra and the infinite dimensional case discussed in functional analysis is that in the first case $\sigma(T) = \sigma_p(T)$ always holds, while in the second case the spectrum usually also contains points which are not eigenvalues of T. If $\lambda I - T$ is injective and surjective as well, then $\lambda \in \varrho(T)$ due to the theorem on the continuity of the inverse (see 12.5.2, **7.**, p. 679). In contrast to the finite dimensional case where the surjectivity follows automatically from the injectivity, the infinite dimensional case has to be dealt with in a very different way.

The set $\sigma_c(T)$ of all $\lambda \in \sigma(T)$, for which $\lambda I - T$ is injective and $Im(\lambda I - T)$ is dense in X, is called the *continuous* spectrum and the set $\sigma_r(T)$ of all λ with an injective $\lambda I - T$ and a non-dense image, is called the *residual spectrum* of operator T.

For a bounded linear operator T in a complex Banach space X

$$\sigma(T) = \sigma_p(T) \cup \sigma_c(T) \cup \sigma_r(T), \tag{12.160}$$

where the terms of the right-hand side are mutually disjoint.

12.5.4 Continuous Linear Functionals

12.5.4.1 Definition

For $\mathrm{Y} = \mathbb{F}$ a linear mapping is called a *linear functional* or a *linear form*. In the following discussions, for a Hilbert space the complex case is considered; in other situations almost every times the real case is considered. The Banach space $L(\mathrm{X}, \mathbb{F})$ of all continuous linear functionals is called the *adjoint space* or the *dual space* of X and it is denoted by X^* (sometimes also by X'). The value (in $\mathbb{F}$) of a linear continuous functional $f \in \mathrm{X}^*$ on an element $x \in \mathrm{X}$ is denoted by $f(x)$, often also by (x, f) – emphasizing the bilinear relation of X and X^* – (compare also with the Riesz theorem (see 12.5.4.2, p. 682).

■ **A:** Let $t_1, t_2, \ldots, t_n$ be fixed points of the interval $[a, b]$ and $c_1, c_2, \ldots, c_n$ real numbers. By the formula

$$f(x) = \sum_{k=1}^{n} c_k x(t_k) \tag{12.161}$$

a linear continuous functional is defined on the space $\mathcal{C}([a, b])$; the norm of f is $\|f\| = \sum_{k=1}^{n} |c_k|$. A special case of (12.161) for a fixed $t \in [a, b]$ is the δ functional

$$\delta_t(x) = x(t) \quad (x \in \mathcal{C}([a, b])). \tag{12.162}$$

■ **B:** With an integrable function $\varphi(t)$ (see 12.9.3.1, p. 696) on $[a, b]$

$$f(x) = \int_a^b \varphi(t)x(t)\,dt \tag{12.163}$$

is a linear continuous functional on $\mathcal{C}([a, b])$ and also on $\mathcal{B}([a, b])$ in each case with the norm $\|f\| = \int_a^b |\varphi(t)|\,dt$.

12.5.4.2 Continuous Linear Functionals in Hilbert Spaces. Riesz Representation Theorem

In a Hilbert space $\mathbb{H}$ equipped with the scalar product $(\cdot,\cdot)$ every element $y \in \mathbb{H}$ defines a linear continuous functional by the formula $f(x) = (x, y)$, where its norm is $\|f\| = \|y\|$. Conversely, if f is a linear continuous functional on $\mathbb{H}$, then there exists a unique element $y \in \mathbb{H}$ such that

$$f(x) = (x, y) \quad \forall\, x \in \mathbb{H}, \tag{12.164}$$

where $\|f\| = \|y\|$. According to this theorem the spaces $\mathbb{H}$ and $\mathbb{H}^*$ are isomorphic and might be identified.

The Riesz representation theorem contains a hint on how to introduce the notion of *orthogonality in an arbitrary normed space.* Let $A \subset \mathrm{X}$ and $A^* \subset \mathrm{X}^*$. The sets

$$A^{\perp} = \{f \in \mathrm{X}\colon f(x) = 0 \quad \forall\, x \in A\} \quad \text{and} \quad A^{*\perp} = \{x \in \mathrm{X}\colon f(x) = 0 \quad \forall\, f \in A^*\} \tag{12.165}$$

are called the *orthogonal complement* or the *annulator* of A and A^*, respectively.

12.5.4.3 Continuous Linear Functionals in L^p

Let $p \geq 1$. The number q is called the *conjugate exponent* to p if $\frac{1}{p} + \frac{1}{q} = 1$, where it is assumed that $q = \infty$ in the case of $p = 1$.

■ Based on the Hölder integral inequality (see 1.4.2.12, p. 32) the functional (12.163) can be considered also in the spaces $L^p([a, b])$ $(1 \leq p \leq \infty)$ (see 12.9.4, p. 697) if $\varphi \in L^q([a, b])$ and $\frac{1}{p} + \frac{1}{q} = 1$. Its norm is then

$$\|f\| = \|\varphi\| = \begin{cases} \left(\int_a^b |\varphi(t)|^q dt\right)^{\frac{1}{q}}, & \text{if} \quad 1 < p \leq \infty, \\ \operatorname*{ess.\,sup}_{t\in[a,b]} |\varphi(t)|, & \text{if} \quad p = 1 \end{cases} \tag{12.166}$$

(with respect to the definition of ess. sup $|\varphi|$ see (12.221), p. 698). To every linear continuous functional f in the space $L^p([a, b])$ there is a uniquely (up to its equivalence class) defined element $y \in L^q([a, b])$ such that

$$f(x) = (x, y) = \int_a^b x(t)\overline{y(t)}dt, \quad x \in L^p \quad \text{and} \quad \|f\| = \|y\|_q = \left(\int_a^b |y(t)|^q dt\right)^{\frac{1}{q}}. \tag{12.167}$$

For the case of $p = \infty$ see [12.15].

12.5.5 Extension of a Linear Functional

1. Semi-Norm

A mapping p: $\mathrm{X} \longrightarrow \mathbb{R}$ of a vector space X is called a *semi-norm* or *pseudonorm*, if it has the following properties:

(HN1) $p(x) \geq 0$, (12.168)

(HN2) $p(\alpha x) = |\alpha| p(x)$, (12.169)

(HN3) $p(x + y) \leq p(x) + p(y)$. (12.170)

Comparison with 12.3.1, p. 669, shows that a semi-norm is a norm if and only if $p(x) = 0$ holds only for $x = 0$.

Both for theoretical mathematical questions and for practical reasons in applications of mathematics, the problem of the extension of a linear functional given on a linear subspace $\mathrm{X}_0 \subset \mathrm{X}$ to the entire space (and, in order to avoid trivial and uninteresting cases) with preserving certain " good" properties became a fundamental question. The solution of this problem is guaranteed by

2. Analytic Form of the Hahn-Banach Extension Theorem

Let X be a vector space over $\mathbb{F}$ and p a pseudonorm on X. Let X_0 be a linear (complex in the case of $\mathbb{F} = \mathbb{C}$ and real in the case of $\mathbb{F} = \mathbb{R}$) subspace of X, and let f_0 be a (complex-valued in the case of $\mathbb{F} = \mathbb{C}$ and real-valued in the case of $\mathbb{F} = \mathbb{R}$) linear functional on X_0 satisfying the relation

$$|f_0(x)| \le p(x) \quad \forall\, x \in \mathrm{X}_0. \tag{12.171}$$

Then there exists a linear functional f on X with the following properties:

$$f(x) = f_0(x) \quad \forall\, x \in \mathrm{X}_0, \quad |f(x)| \le p(x) \quad \forall\, x \in \mathrm{X}. \tag{12.172}$$

So, f is an extension of the functional f_0 onto the whole space X preserving the relation (12.171).

If X_0 is a linear subspace of a normed space X and f_0 is a continuous linear functional on X_0, then $p(x) = \|f_0\| \cdot \|x\|$ is a pseudonorm on X satisfying (12.171), so the Hahn-Banach extension theorem for continuous linear functionals is obtained.

Two important consequences are:

1. For every element $x \neq 0$ there is a functional $f \in \mathrm{X}^*$ with $f(x) = \|x\|$ and $\|f\| = 1$.

2. For every linear subspace $\mathrm{X}_0 \subset \mathrm{X}$ and $x_0 \notin \mathrm{X}_0$ with the positive distance $d = \inf_{x \in X_0} \|x - x_0\| > 0$ there is an $f \in \mathrm{X}^*$ such that

$$f(x) = 0 \ \forall\, x \in \mathrm{X}_0, \quad f(x_0) = 1 \quad \text{and} \quad \|f\| = \frac{1}{d}. \tag{12.173}$$

12.5.6 Separation of Convex Sets

1. Hyperplanes

A linear subset L of the real vector space X, $L \neq \mathrm{X}$, is called a *hypersubspace* or *hyperplane through* 0 if there exists an $x_0 \in \mathrm{X}$ such that $\mathrm{X} = lin(x_0, L)$. Sets of the form $x + L$ (L a linear subset) are affine-linear manifolds (see 12.1.2, p. 655). If L is a hypersubspace, these manifolds are called *hyperplanes*.

There exist the following close relations between hypersubspaces, hyperplanes and linear functionals:

a) The kernel $f^{-1}(0) = \{x \in \mathrm{X}\colon f(x) = 0\}$ of a linear functional f on X is a hypersubspace in X, and for each number $\lambda \in \mathbb{R}$ there exists an element $x_\lambda \in \mathrm{X}$ with $f(x_\lambda) = \lambda$ and $f^{-1}(\lambda) = x_\lambda + f^{-1}(0)$.

b) For any given hypersubspace $L \subset X$ and each $x_0 \notin L$ and $\lambda \neq 0$ ($\lambda \in \mathbb{R}$) there always exists a uniquely determined linear functional f on X with $f^{-1}(0) = L$ and $f(x_0) = \lambda$.

The closedness of $f^{-1}(0)$ in the case of a normed space X is equivalent to the continuity of the functional f.

2. Geometric Form of the Hahn–Banach Extension Theorem

Let X be a normed space, $x_0 \in \mathrm{X}$ and L a linear subspace of X. Then for every non-empty convex open set K which does not intersect the affine-linear manifold $x_0 + L$, there exists a closed hypersubspace H such that $x_0 + L \subset H$ and $H \cap K = \emptyset$.

3. Separation of Convex Sets

Two subsets A, B of a real normed space X are called *separated* by a hyperplane if there is a functional $f \in \mathrm{X}^*$ such that:

$$\sup_{x \in A} f(x) \le \inf_{y \in B} f(y). \tag{12.174}$$

The separating hyperplane is then given by $f^{-1}(\alpha)$ with $\alpha = \sup_{x \in A} f(x)$, which means that the two sets are contained in the different half-spaces

$$A \subset \{x \in \mathrm{X}\colon f(x) \leq \alpha\} \quad \text{and} \quad B \subset \{x \in \mathrm{X}\colon f(x) \geq \alpha\}. \tag{12.175}$$

In **Fig. 12.5b,c** two cases of the separation by a hyperplane are shown.
Their disjointness is less decisive for the separation of two sets. In fact, **Fig. 12.5a** shows two sets E and B, which are not separated although E and B are disjoint and B is convex. The convexity of both sets is the intrinsic property for separating them. In this case it is possible that the sets have common points which are contained in the hyperplane.

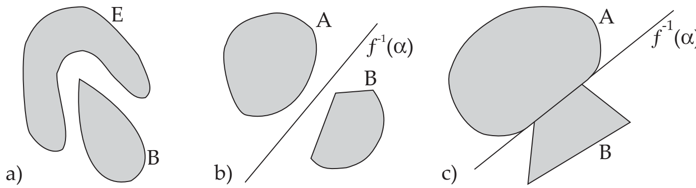

Figure 12.5

If A is a convex set of a normed space X with a non-empty interior $Int(A)$ and $B \subset \mathrm{X}$ is a non-empty convex set with $Int(A) \cap B = \emptyset$, then A and B can be separated. The hypothesis $Int(A) \neq \emptyset$ in that statement cannot be dropped (see [12.3], example 4.47). A (real linear) functional $f \in \mathrm{X}^*$ is called a *supporting functional* of the set A at the point $x_0 \in A$, if there is a real number $\lambda \in \mathbb{R}$ such that $f(x_0) = \lambda$, and $A \subset \{x \in \mathrm{X} : f(x) \leq \lambda\}$. $f^{-1}(\lambda)$ is called the *supporting hyperplane* at the point x_0.
For a convex set K with a non-empty interior, there exists a supporting functional at each of its boundary points.

Remark: The famous Kuhn-Tucker theorem (see 18.2, p. 925) which yields practical methods to determine the minimum of convex optimization problems (see [12.5]), is also based on the separation of convex sets.

12.5.7 Second Adjoint Space and Reflexive Spaces

The adjoint space X^* of a normed space X is also a normed space if it is equipped with the norm $\|f\| = \sup\limits_{\|x\|\leq 1} |f(x)|$, so $(\mathrm{X}^*)^* = \mathrm{X}^{**}$ – the *second adjoint space* to X can also be considered. The *canonical embedding*

$$J\colon \mathrm{X} \longrightarrow \mathrm{X}^{**} \quad \text{with} \quad Jx = F_x, \text{ where } \quad F_x(f) = f(x) \quad \forall\, f \in \mathrm{X}^* \tag{12.176}$$

is a norm isomorphism (see 12.3.1, p. 669), hence X is identified with the subset $J(\mathrm{X}) \subset \mathrm{X}^{**}$. A Banach space X is called *reflexive* if $J(\mathrm{X}) = \mathrm{X}^{**}$. Hence the canonical embedding is then a surjective norm isomorphism.

■ Every finite dimensional Banach space and every Hilbert space is reflexive, as well as the spaces L^p $(1 \leq p < \infty)$, however $\mathcal{C}([a,b])$, $L^1([0,1])$, $\mathbf{c_0}$ are examples of non-reflexive spaces.

12.6 Adjoint Operators in Normed Spaces

12.6.1 Adjoint of a Bounded Operator

For a given linear continuous operator $T\colon \mathrm{X} \longrightarrow \mathrm{Y}$ (X, Y are normed spaces) to every $g \in \mathrm{Y}^*$ there is assigned a functional $f \in \mathrm{X}^*$ by $f(x) = g(Tx)$, $\forall\, x \in \mathrm{X}$. In this way, a linear continuous operator

$$T^*\colon \mathrm{Y}^* \longrightarrow \mathrm{X}^*, \quad (T^*g)(x) = g(Tx), \quad \forall\, g \in \mathrm{Y}^* \text{ and } \forall\, x \in \mathrm{X} \tag{12.177}$$

is obtained which is called the *adjoint operator* of T and has the following properties:

$(T+S)^* = T^* + S^*$, $(ST)^* = S^*T^*$, $\|T^*\| = \|T\|$, where for the linear continuous operators $T\colon \mathrm{X} \to \mathrm{Y}$ and $S : \mathrm{Y} \to \mathrm{Z}$ (X, Y, Z normed spaces), the operator $ST : \mathrm{X} \to \mathrm{Z}$ is defined in the natural way as $ST(x) = S(T(x))$ (see 12.3.4, ■ C, p. 672). With the notation introduced in 12.1.5, p. 658, and 12.5.4.2, p. 682, the following identities are valid for an operator $T \in B(\mathrm{X}, \mathrm{Y})$:

$$\overline{Im(T)} = \ker(T^*)^\perp, \quad \overline{Im(T^*)} = \ker(T)^\perp, \tag{12.178}$$

where the closedness of $Im(T)$ implies the closedness of $Im(T^*)$.

The operator $T^{**}: \mathrm{X}^{**} \to \mathrm{Y}^{**}$, obtained as $(T^*)^*$ from T^*, is called the *second adjoint* of T. Due to $(T^{**}(F_x))g = F_x(T^*g) = (T^*g)(x) = g(Tx) = F_{Tx}(g)$ the operator T^{**} has the following property: If $F_x \in \mathrm{X}^{**}$, then $T^{**}F_x = F_{Tx} \in \mathrm{Y}^{**}$. Hence, the operator $T^{**}: \mathrm{X}^{**} \to \mathrm{Y}^{**}$ is an extension of T.

In a Hilbert space $\mathbb{H}$ the adjoint operator can also be introduced by means of the scalar product $(Tx, y) = (x, T^*y),\ x, y \in \mathbb{H}$. This is based on the Riesz representation theorem, where the identification of $\mathbb{H}$ and $\mathbb{H}^{**}$ implies $(\lambda T)^* = \overline{\lambda} T^*$, $I^* = I$ and even $T^{**} = T$. If T is bijective, then the same holds for T^*, and also $(T^*)^{-1} = (T^{-1})^*$. For the resolvents of T and T^* there holds

$$[R_\lambda(T)]^* = R_{\overline{\lambda}}(T^*), \tag{12.179}$$

from which $\sigma(T^*) = \{\overline{\lambda}\colon \lambda \in \sigma(T)\}$ follows for the spectrum of the adjoint operator.

■ **A:** Let T be an integral operator in the space $L^p([a, b])$ $(1 < p < \infty)$

$$(Tx)(s) = \int_a^b K(s,t)x(t)\,dt \tag{12.180}$$

with a continuous kernel $K(s,t)$. The adjoint operator of T is also an integral operator, namely

$$(T^*g)(t) = \int_a^b K^*(t,s)y_g(s)\,ds \tag{12.181}$$

with the kernel $K^*(s,t) = K(t,s)$, where y_g is the element from L^q associated to $g \in (L^p)^*$ according to (12.167).

■ **B:** In a finite dimensional complex vector space the adjoint of an operator represented by the matrix $\mathbf{A} = (a_{ij})$ is defined by the matrix $\mathbf{A}^*$ with $a^*_{ij} = \overline{a_{ji}}$.

12.6.2 Adjoint Operator of an Unbounded Operator

Let X and Y be real normed spaces and T a (not necessarily bounded) linear operator with a (linear) domain $D(T) \subset \mathrm{X}$ and values in Y. For a given $g \in \mathrm{Y}^*$, the expression $g(Tx)$, depending obviously linearly on x, is meaningful. Now the question is: Does there exist a well-defined functional $f \in \mathrm{X}^*$ such that

$$f(x) = g(Tx) \quad \forall\, x \in D(T). \tag{12.182}$$

Let $D^* \subset \mathrm{Y}^*$ be the set of all those $g \in \mathrm{Y}^*$ for which the representation (12.182) holds for a certain $f \in \mathrm{X}^*$. If $\overline{D(T)} = \mathrm{X}$, then for given g the functional f is uniquely defined. So a linear operator T^* is defined by $f = T^*g$ with $D(T^*) = D^*$. Then for arbitrary $x \in D(T)$ and $g \in D(T^*)$

$$g(Tx) = (T^*g)(x). \tag{12.183}$$

The operator T^* turns out to be closed and is called the *adjoint* of T. The naturalness of this general procedure stems from the fact that $D(T^*) = \mathrm{Y}^*$ holds if and only if T is bounded on $D(T)$. In this case $T^* \in B(\mathrm{Y}^*, \mathrm{X}^*)$ and $\|T^*\| = \|T\|$ hold.

12.6.3 Self-Adjoint Operators

An operator $T \in B(\mathbb{H})$ ($\mathbb{H}$ is a Hilbert space) is called *self-adjoint* if $T^* = T$. In this case

$$(Tx, y) = (x, Ty), \ \ x, y \in \mathbb{H} \tag{12.184a}$$

is valid and the number (Tx, x) is real for each $x \in \mathbb{H}$. Then the equality

$$\|T\| = \sup_{\|x\|=1} |(Tx, x)| \tag{12.184b}$$

holds and with $m = m(T) = \inf_{\|x\|=1} (Tx, x)$ and $M = M(T) = \sup_{\|x\|=1} (Tx, x)$ also the relations

$$m(T)\|x\|^2 \le (Tx, x) \le M(T)\|x\|^2 \quad \text{and} \quad \|T\| = r(T) = \max\{|m|, M\} \tag{12.185}$$

are valid. The equality (12.184a) characterizes the self-adjoint operators. The spectrum of a self-adjoint (bounded) operator lies in the interval $[m, M]$ and $m, M \in \sigma(T)$ holds.

12.6.3.1 Positive Definite Operators

A partial ordering can be introduced in the set of all self-adjoint operators of $B(\mathbb{H})$ by defining

$$T \ge 0 \quad \text{if and only if} \quad (Tx, x) \ge 0 \quad \forall\, x \in \mathbb{H}. \tag{12.186}$$

An operator T with $T \ge 0$ is called *positive* (or, more exactly *positive definite*). For any self-adjoint operator T (with **(H1)** from 12.4.1.1, p. 673), $(T^2x, x) = (Tx, Tx) \ge 0$, so T^2 is positive definite. Every positive definite operator T possesses a square root, i.e., there exists a unique positive definite operator W such that $W^2 = T$. Moreover, the vector space of all self-adjoint operators is a K-space (Kantorovich space, see 12.1.7.4, p. 660), where the operators

$$|T| = \sqrt{T^2}, \quad T^+ = \frac{1}{2}(|T| + T), \quad T^- = \frac{1}{2}(|T| - T) \tag{12.187}$$

are the corresponding elements with respect to (12.37). They are of particular importance for the spectral decomposition and spectral and integral representations of self-adjoint operators by means of some Stieltjes integral (see 8.2.3.1, **2.**, p. 506, and [12.1], [12.11], [12.12], [12.15], [12.18]).

12.6.3.2 Projectors in a Hilbert Space

Let $\mathbb{H}_0$ be a subspace of a Hilbert space $\mathbb{H}$. Then every element $x \in \mathbb{H}$ has its projection x' onto $\mathbb{H}_0$ according to the projection theorem (see 12.4.2, p. 674), and therefore, an operator P with $Px = x'$ is defined on $\mathbb{H}$ with values in $\mathbb{H}_0$. P is called a *projector* onto $\mathbb{H}_0$. Obviously, P is linear, continuous, and $\|P\| = 1$. A continuous linear operator P in $\mathbb{H}$ is a projector (onto a certain subspace) if and only if:

a) $P = P^*$, i.e., P is self-adjoint, and

b) $P^2 = P$, i.e., P is *idempotent*.

12.7 Compact Sets and Compact Operators

12.7.1 Compact Subsets of a Normed Space

A subset A of a normed space[†] X is called

• *compact,* if every sequence of elements from A contains a convergent subsequence whose limit lies in A,

• *relatively compact* or *precompact* if its closure (see 12.2.1.3, p. 664) is compact, i.e., every sequence of elements from A contains a convergent subsequence (whose limit does not necessarily belong to A).

This is the Bolzano–Weierstrass theorem in real calculus for bounded sequences in $\mathbb{R}^n$, and one says that such a set has the *Bolzano–Weierstrass property.*

Every compact set is closed and bounded. Conversely, if the space X is finite dimensional, then every such set is compact. The closed unit ball in a normed space X is compact if and only if X is finite dimensional.

For some characterizations of relatively compact subsets in metric spaces (the Hausdorff theorem on the existence of a finite ε-net) and in the spaces **s**, $\mathcal{C}$ (Arzela–Ascoli theorem) and in the spaces $\boldsymbol{L^p}(1 < p < \infty)$ see [12.15].

12.7.2 Compact Operators

12.7.2.1 Definition of Compact Operator

An arbitrary operator T: X $\longrightarrow$ Y of a normed space X into a normed space Y is called *compact* if the

[†]It is enough that X is a metric (or an even more general) space. This generality is not used in what follows.

image $T(A)$ of every bounded set $A \subset \mathrm{X}$ is a relatively compact set in Y. If, in addition the operator T is also continuous, then it is called *completely continuous*. Every *compact linear* operator is bounded and consequently completely continuous. For a linear operator to be compact it is sufficient to require that it transforms the unit ball of X into a relatively compact set in Y.

12.7.2.2 Properties of Linear Compact Operators

A characterization by sequences of the compactness of an operator from $B(\mathrm{X},\mathrm{Y})$ is the following: For every bounded sequence $\{x_n\}_{n=1}^{\infty}$ from X the sequence $\{Tx_n\}_{n=1}^{\infty}$ contains a convergent subsequence. A linear combination of compact operators is also compact. If one of the operators $U \in B(\mathrm{W},\mathrm{X})$, $T \in B(\mathrm{X},\mathrm{Y})$, $S \in B(\mathrm{Y},\mathrm{Z})$ in each of the following products is compact, then the operators TU and ST are also compact. If Y is a Banach space, then the following important statements are valid.

a) Convergence: If a sequence of compact operators $\{T_n\}_{n=1}^{\infty}$ is convergent in the space $B(\mathrm{X},\mathrm{Y})$, then its limit is a compact operator, too.

b) Schauder Theorem: If T is a linear continuous operator, then either both T and T^* are compact or both are not.

c) Spectral Properties of a Compact Operator T in an (Infinite Dimensional) Banach Space X: The zero belongs to the spectrum. Every non-zero point of the spectrum $\sigma(T)$ is an eigenvalue with a finite dimensional eigenspace $\mathrm{X}_\lambda = \{x \in \mathrm{X}\colon (\lambda I - T)x = 0\}$, and $\forall\, \varepsilon > 0$ there is always only a finite number of eigenvalues of T outside the circle $\{|\lambda| \le \varepsilon\}$, where only the zero can be an accumulation point of the set of eigenvalues. If $\lambda = 0$ is not an eigenvalue of T, then T^{-1} is unbounded if it exists.

12.7.2.3 Weak Convergence of Elements

A sequence $\{x_n\}_{n=1}^{\infty}$ of elements of a normed space X is called *weakly convergent* to an element x_0 if for each $f \in \mathrm{X}^*$ the relation $f(x_n) \to f(x_0)$ holds (written as: $x_n \rightharpoonup x_0$ or as $x_n \xrightarrow{w} x_0$).
Obviously: $x_n \to x_0$ implies $x_n \rightharpoonup x_0$. If Y is another normed space and $T\colon \mathrm{X} \longrightarrow \mathrm{Y}$ is a continuous linear operator, then:

a) $x_n \rightharpoonup x_0$ implies $Tx_n \rightharpoonup Tx_0$,

b) if T is compact, then $x_n \rightharpoonup x_0$ implies $Tx_n \to Tx_0$.

■ **A:** Every finite dimensional operator is compact. From this fact it follows that the identity operator in an infinite dimensional space cannot be compact (see 12.7.1, p. 686).

■ **B:** Suppose $\mathrm{X} = \mathbf{l}^2$, and let T be the operator in $\mathbf{l}^2$ given by the infinite matrix

$$\begin{pmatrix} t_{11} & t_{12} & t_{13} & \cdots \\ t_{21} & t_{22} & t_{23} & \cdots \\ t_{31} & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdots \end{pmatrix} \quad \text{with} \quad Tx = \left(\sum_{k=1}^{\infty} t_{1k}x_k, \ldots, \sum_{k=1}^{\infty} t_{nk}x_k, \ldots \right). \tag{12.188}$$

If $\sum\limits_{k,n=1}^{\infty} |t_{nk}|^2 = M < \infty$, then T is a compact operator from $\mathbf{l}^2$ into $\mathbf{l}^2$ with $\|T\| \le M$.

■ **C:** The integral operator (12.136) is a compact operator in the spaces $\mathcal{C}([a,b])$ and $L^p((a,b))$ $(1 < p < \infty)$.

12.7.3 Fredholm Alternative

Let T be a compact linear operator in a Banach space X. The following equations (of the second kind) are considered with a parameter $\lambda \ne 0$:

$$\begin{aligned} \lambda x - Tx &= y, \quad & \lambda x - Tx &= 0, \\ \lambda f - T^* f &= g, \quad & \lambda f - T^* f &= 0. \end{aligned} \tag{12.189}$$

The following statements are valid:

a) $\dim(\ker(\lambda I - T)) = \dim(\ker(\lambda I - T^*)) < +\infty$, i.e., both homogeneous equations always have the same number of linearly independent solutions.

b) $\mathrm{Im}(\lambda I - T) = \ker(\lambda I - T^*)^{\perp}$ and [‡] $\mathrm{Im}(\lambda I - T^*) = \ker(\lambda I - T)^{\perp}$.

c) $\mathrm{Im}(\lambda I - T) = \mathrm{X}$ if and only if $\ker(\lambda I - T) = 0$.

d) The *Fredholm alternative* (also called the Riesz–Schauder theorem):

α) Either the homogeneous equation has only the trivial solution. In this case $\lambda \in \varrho(T)$, the operator $(\lambda I - T)^{-1}$ is bounded, and the inhomogeneous equation has exactly one solution $x = (\lambda I - T)^{-1} y$ for arbitrary $y \in \mathrm{X}$.

β) Or the homogeneous equation has at least one non-trivial solution. In this case λ is an eigenvalue of T, i.e., $\lambda \in \sigma(T)$, and the inhomogeneous equation has a (non-unique) solution if and only if the right-hand side y satisfies the condition $f(y) = 0$ for every solution f of the adjoint equation $T^* f = \lambda f$. In this last case every solution x of the inhomogeneous equation has the form $x = x_0 + h$, where x_0 is a fixed solution of the inhomogeneous equation and $h \in \ker(\lambda I - T)$.

Linear equations of the form $Tx = y$ with a compact operator T are called equations of the first kind. Their mathematical investigation is in general more difficult (see [12.11],[12.18]).

12.7.4 Compact Operators in Hilbert Space

Let $T : \mathbb{H} \longrightarrow \mathbb{H}$ be a compact operator. Then T is the limit (in $B(\mathbb{H})$) of a sequence of finite dimensional operators. The similarity to the finite dimensional case can be seen from the statements:

If C is a finite dimensional operator and $T = I - C$, then the injectivity of T implies the existence of T^{-1} and $T^{-1} \in B(\mathbb{H})$.

If C is a compact operator, then the following statements are equivalent:

a) $\exists\, T^{-1}$ and it is continuous,

b) $x \neq 0 \Rightarrow Tx \neq 0$, i.e., T is injective,

c) $T(\mathbb{H}) = \mathbb{H}$, i.e., T is surjective.

12.7.5 Compact Self-Adjoint Operators

1. Eigenvalues

A compact self-adjoint operator $T \neq 0$ in a Hilbert space $\mathbb{H}$ possesses at least one non-zero eigenvalue. More precisely, T always has an eigenvalue λ with $|\lambda| = \|T\|$. The set of eigenvalues of T is at most countable.

Any compact self-adjoint operator T has the representation $T = \sum\limits_k \lambda_k P_{\lambda_k}$ (in $B(\mathbb{H})$), where λ_k are the different eigenvalues of T and P_λ denotes the projector onto the eigenspace $\mathbb{H}_\lambda$. In this case the operator T is diagonalizable. From this fact it follows that $Tx = \sum\limits_k \lambda_k (x, e_k) e_k$ for every $x \in \mathbb{H}$, where $\{e_k\}$ is the orthonormal system of the eigenvectors of T. If $\lambda \notin \sigma(T)$ and $y \in \mathbb{H}$, then the solution of the equation $(\lambda I - T)x = y$ can be represented as $x = R_\lambda(T) y = \sum\limits_k \dfrac{1}{\lambda - \lambda_k} (y, e_k) e_k$.

2. Hilbert-Schmidt Theorem

If T is a compact self-adjoint operator in a separable Hilbert space $\mathbb{H}$, then there is a basis in $\mathbb{H}$ consisting of the eigenvectors of T.

The so-called spectral (mapping) theorems (see [12.8], [12.10], [12.12], [12.13], [12.18]) can be considered as the generalization of the Hilbert–Schmidt theorem for the non-compact case of self-adjoint (bounded or unbounded) operators.

[‡] Here the orthogonality is considered in Banach spaces (see 12.5.4.2, p. 682).

12.8 Non-Linear Operators

In the theory of non-linear operator equations the most important methods are based on the following principles:

1. Principle of the Contracting Mapping, Banach Fixed-Point Theorem (see 12.2.2.3, p. 666, and 12.2.2.4, p. 666). For further modifications of this principle see [12.8],[12.11], [12.12], [12.18].

2. Generalization of the Newton Method (see 18.2.5.2, p. 931 and 19.1.1.2, p. 950) for the infinite dimensional case.

3. Schauder Fixed-Point Principle (see 12.8.4, p. 691)

4. Leray-Schauder Theory (see 12.8.5, p. 692)

Methods based on principles **1** and **2** yield information on the existence, uniqueness, constructivity etc. of the solution, while methods based on principles **3** and **4**, in general, allow "only" the qualitative statement of the existence of a solution. If further properties of operators are known then see also 12.8.6, p. 692, and 12.8.7, p. 693.

12.8.1 Examples of Non-Linear Operators

For non-linear operators the relation between continuity and boundedness discussed for linear operators in 12.5.1, p. 677 is no longer valid in general. In studying non-linear operator equations, e.g., non-linear boundary value problems or integral equations, the following non-linear operators occur most often. Iteration methods described in 12.2.2.4, p. 666, can be successfully applied for solving non-linear integral equations.

1. Nemytskij Operator

Let Ω be an open measurable subset from $\mathbb{R}^n$ (12.9.1, p. 693) and $f\colon\ \Omega\times\mathbb{R}\longrightarrow\mathbb{R}$ a function of two variables $f(x,s)$, which is continuous with respect to x for almost every s and measurable with respect to s for every x (*Caratheodory conditions*). The non-linear operator $\mathcal{N}$ to $\mathcal{F}(\Omega)$ defined as

$$(\mathcal{N}u)(x) = f[x, u(x)] \quad (x\in\Omega) \tag{12.190}$$

is called the *Nemytskij operator*. It is continuous and bounded if it maps $L^p(\Omega)$ into $L^q(\Omega)$, where $\frac{1}{p}+\frac{1}{q}=1$. This is the case, e.g., if

$$|f(x,s)| \le a(x) + b|s|^{\frac{p}{q}} \quad \text{with} \quad a(x)\in L^q(\Omega) \quad (b>0) \tag{12.191}$$

or if $f\colon\ \Omega\times\mathbb{R}\longrightarrow\mathbb{R}$ is continuous. The operator $\mathcal{N}$ is compact only in special cases.

2. Hammerstein Operator

Let Ω be a relatively compact subset of $\mathbb{R}^n$, f a function satisfying the Caratheodory conditions and $K(x,y)$ a continuous function on $\overline{\Omega}\times\overline{\Omega}$. The non-linear operator $\mathcal{H}$ on $\mathcal{F}(\Omega)$

$$(\mathcal{H}u)(x) = \int_\Omega K(x,y) f[y, u(y)]\, dy \quad (x\in\Omega) \tag{12.192}$$

is called the *Hammerstein operator*. $\mathcal{H}$ can be written in the form $\mathcal{H}=\mathcal{K}\cdot\mathcal{N}$ with the Nemytskij operator $\mathcal{N}$ and the integral operator $\mathcal{K}$ determined by the kernel K

$$(\mathcal{K}u)(x) = \int_\Omega K(x,y) u(y)\, dy \quad (x\in\Omega). \tag{12.193}$$

If the kernel $K(x,y)$ satisfies the additional condition

$$\int_{\Omega\times\Omega} |K(x,y)|^q\, dx\, dy < \infty \tag{12.194}$$

and the function f satisfies the condition (12.191), then $\mathcal{H}$ is a continuous and compact operator on $L^p(\Omega)$.

3. Urysohn Operator

Let $\Omega \subset \mathbb{R}^n$ be an open measurable subset and $K(x, y, s)\colon\ \Omega \times \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ a function of three variables. Then the non-linear operator $\mathcal{U}$ on $\mathcal{F}(\Omega)$

$$(\mathcal{U}u)(x) = \int\limits_{\Omega} K[x, y, u(y)]\, dy \quad (x \in \Omega) \tag{12.195}$$

is called the *Urysohn operator*. If the kernel K satisfies the appropriate conditions, then $\mathcal{U}$ is a continuous and compact operator in $\mathcal{C}(\Omega)$ or in $L^p(\Omega)$, respectively.

12.8.2 Differentiability of Non-Linear Operators

Let X, Y be Banach spaces, $D \subset \mathrm{X}$ be an open set and $T\colon D \longrightarrow \mathrm{Y}$. The operator T is called *Fréchet differentiable* (or, briefly, differentiable) at the point $x \in D$ if there exists a linear operator $L \in B(\mathrm{X}, \mathrm{Y})$ (in general depending on the point x) such that

$$T(x + h) - T(x) = Lh + \omega(h) \quad \text{with} \quad \|\omega(h)\| = o(\|h\|) \tag{12.196}$$

or in an equivalent form

$$\lim_{\|h\| \to 0} \frac{\|T(x + h) - T(x) - Lh\|}{\|h\|} = 0, \tag{12.197}$$

i.e., $\forall\, \varepsilon > 0,\ \exists\, \delta > 0$, such that $\|h\| < \delta$ implies $\|T(x + h) - T(x) - Lh\| \le \varepsilon\|h\|$. The operator L, which is usually denoted by $T'(x)$, $T'(x, \cdot)$ or $T'(x)(\cdot)$, is called the *Fréchet derivative* of the operator T at the point x. The value $dT(x; h) = T'(x)h$ is called the *Fréchet differential* of the operator T at the point x (for the increment h).

The differentiability of an operator at a point implies its continuity at that point. If $T \in B(\mathrm{X}, \mathrm{Y})$, i,e., T itself is linear and continuous, then T is differentiable at every point, and its derivative is equal to T.

12.8.3 Newton's Method

Let X, D be as in the previous paragraph and $T\colon D \longrightarrow \mathrm{Y}$. Under the assumption of the differentiability of T at every point of the set D an operator $T'\colon D \longrightarrow B(\mathrm{X}, \mathrm{Y})$ is defined by assigning the element $T'(x) \in B(\mathrm{X}, \mathrm{Y})$ to every point $x \in D$. Suppose the operator T' is continuous on D (in the operator norm); in this case T is called *continuously differentiable* on D.

Suppose $\mathrm{Y} = \mathrm{X}$ and also that the set D contains a solution x^* of the equation

$$T(x) = 0. \tag{12.198}$$

Furthermore, it is assumed that the operator $T'(x)$ is continuously invertible for each $x \in D$, hence $[T'(x)]^{-1}$ is in $B(\mathrm{X})$. Because of (12.196) for an arbitrary $x_0 \in D$ one conjectures that the elements $T(x_0) = T(x_0) - T(x^*)$ and $T'(x_0)(x_0 - x^*)$ are "not far" from each other and therefore the element x_1 defined as

$$x_1 = x_0 - [T'(x_0)]^{-1} T(x_0) \tag{12.199}$$

is an approximation of x^* (under the given assumptions). Starting with an arbitrary x_0 the so-called *Newton approximation sequence*

$$x_{n+1} = x_n - [T'(x_n)]^{-1} T(x_n) \quad (n = 0, 1, \ldots) \tag{12.200}$$

can be constructed. There are many theorems known from the literature discussing the behavior and the convergence properties of this method. Here only the following most important result is mentioned which demonstrates the main properties and advantages of Newton's method:

$\forall\, \varepsilon \in (0, 1)$ there exists a ball $B = B(x_0; \delta)$, $\delta = \delta(\varepsilon)$ in X, such that all points x_n lie in B and the Newton sequence converges to the solution x^* of (12.198). Moreover, $\|x_n - x_0\| \le \varepsilon^n \|x_0 - x^*\|$ which yields a practical error estimation.

The *modified Newton's method* is obtained if the operator $[T'(x_0)]^{-1}$ is used instead of $[T'(x_n)]^{-1}$ $\forall n = 0, 1, \ldots$ in formula (12.200). For further estimations of the speed of convergence and for the (in general sensitive) dependence of the method on the choice of the starting point x_0 see [12.7], [12.12], [12.18].

■ **Jacobian or Functional Matrix** Given a non-linear operator $T = F\colon D \longrightarrow \mathbb{R}^m$ on an open set $D \subset \mathbb{R}^n$ with m non-linear coordinate functions $F_1, F_2, \ldots, F_m$ and n independent variables $x_1, x_2, \ldots, x_n$. Then

$$F(x) = \begin{pmatrix} F_1(x) \\ F_2(x) \\ \vdots \\ F_m(x) \end{pmatrix} \in \mathbb{R}^m \quad \forall\, x = (x_1, x_2, \ldots, x_n) \in D \tag{12.201}$$

holds. If the partial derivatives $\frac{\partial F_i}{\partial x_k}$ $(k = 1, 2, \ldots, n)$ of the coordinate functions F_i $(i = 1, 2, \ldots, m)$ on D exist and are continuous, then the mapping (the operator) F in every point of D is differentiable and its derivative at the point $x = (x_1, x_2, \ldots, x_n) \in D$ is the linear operator $F'(x) : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ with the matrix representation

$$F'(x) = \begin{pmatrix} \frac{\partial F_1(x)}{\partial x_1} & \frac{\partial F_1(x)}{\partial x_2} & \cdots & \frac{\partial F_1(x)}{\partial x_n} \\ \frac{\partial F_2(x)}{\partial x_1} & \frac{\partial F_2(x)}{\partial x_2} & \cdots & \frac{\partial F_2(x)}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_m(x)}{\partial x_1} & \frac{\partial F_m(x)}{\partial x_2} & \cdots & \frac{\partial F_m(x)}{\partial x_n} \end{pmatrix}. \tag{12.202}$$

The derivative $F'(x)$ is a matrix of the type (m, n). It is called the *Jacobian* or *functional matrix* of F. Special cases of occurrence are, e.g., the iterative solution of systems of non-linear equations by the Newton method (see 19.2.2.2, p. 962) or describing the independence of functions (see 2.18.2.6, **3.**, p. 123).
For $m = n$ the so-called *functional determinant* or *Jacobian determinant* can be formed, which is denoted shortly by

$$\frac{D(F_1, F_2, \ldots, F_n)}{D(x_1, x_2, \ldots, x_n)}. \tag{12.203}$$

This determinant is used for the solution of (mostly inner-mathematical) problems (see also, e.g., 8.5.3.2, p. 539).

12.8.4 Schauder's Fixed-Point Theorem

Let $T : D \longrightarrow X$ be a non-linear operator defined on a subset D of a Banach space X. The non-trivial question of whether the equation $x = T(x)$ has at least one solution, can be answered as follows: If $X = \mathbb{R}$ and $D = [-1, 1]$, then every continuous function, mapping D into D, has a fixed point in D. If X is an arbitrary *finite dimensional* normed space (dimX ≥ 2), then *Brouwer's fixed-point theorem* holds.

1. Brouwer's Fixed-Point Theorem Let D be a non-empty closed bounded and convex subset of a finite dimensional normed space. If T is a continuous operator, which maps D into itself, then T has at least one fixed point in D.
The answer in the case of an arbitrary infinite dimensional Banach space X is given by *Schauder's fixed-point theorem*.

2. Schauder's Fixed-Point Theorem Let D be a non-empty closed bounded and convex subset of a Banach space X. If the operator $T : D \longrightarrow$ X is continuous and compact (hence completely continuous) and it maps D into itself, then T has at least one fixed point in D.
By using this theorem, it is proved, e.g., that the initial value problem (12.70), p. 668, always has a

local solution for $t \geq 0$, if the right-hand side is assumed only to be continuous.

12.8.5 Leray-Schauder Theory

For the existence of solutions of the equations $x = T(x)$ and $(I+T)(x) = y$ with a completely continuous operator T, a further principle is found which is based on deep properties of the mapping degree. It can be successfully applied to prove the existence of a solution of non-linear boundary value problems. Here only those results of this theory are mentioned which are the most useful ones in practical problems, and for simplicity a formulation is chosen which avoids the notion of the mapping degree.

Leray–Schauder Theorem: Let D be an open bounded set in a real Banach space X and let T : $\overline{D}$:$\longrightarrow$ X be a completely continuous operator. Let $y \in D$ be a point such that $x + \lambda T(x) \neq y$ for each $x \in \partial D$ and $\lambda \in [0,1]$, where ∂D denotes the boundary of the set D. Then the equation $(I+T)(x) = y$ has at least one solution.

The following version of this theorem is very useful in applications:

Let T be a completely continuous operator in the Banach space X. If all solutions of the family of equations

$$x = \lambda T(x) \quad (\lambda \in [0,1]) \tag{12.204}$$

are uniformly bounded, i.e., $\exists\, c > 0$ such that $\forall\, \lambda$ and $\forall\, x$ satisfying (12.204) the a priori estimation $\|x\| \leq c$ holds, then the equation $x = T(x)$ has a solution.

12.8.6 Positive Non-Linear Operators

The successful application of Schauder's fixed-point theorem requires the choice of a set with appropriate properties, which is mapped into itself by the considered operator. In applications, especially in the theory of non-linear boundary value problems, ordered normed function spaces and positive operators are often considered, i.e., which leave the corresponding cone invariant, or *isotone increasing* operators, i.e., if $x \leq y \Rightarrow T(x) \leq T(y)$. If confusions (see, e.g., 12.8.7, p. 693) are excluded, these operators are also called *monotone*.

Let X $=$ (X, X$_+$, $\|\cdot\|$) be an ordered Banach space, X$_+$ a closed cone and $[a,b]$ an order interval of X. If X$_+$ is normal and T is a completely continuous (not necessarily isotone) operator that satisfies $T([a,b]) \subset [a,b]$, then T has at least one fixed point in $[a,b]$ **(Fig. 12.6b)**.

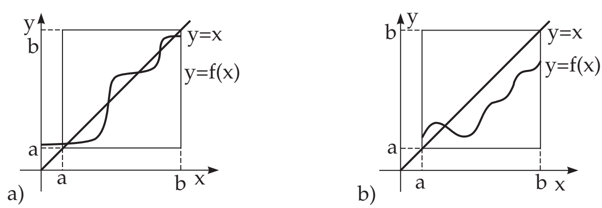

Figure 12.6

Notice that the condition $T([a,b]) \subset [a,b]$ automatically holds for any isotone increasing operator T, which is defined on an (o)-interval (order interval) $[a,b]$ of the space X if it maps only the endpoints a, b into $[a,b]$, i.e., when the two conditions $T(a) \geq a$ and $T(b) \leq b$ are satisfied. Then both sequences

$$x_0 = a \text{ and } x_{n+1} = T(x_n) \;\; (n \geq 0) \quad \text{and} \quad y_0 = b \text{ and } y_{n+1} = T(y_n) \;\; (n \geq 0) \tag{12.205}$$

are well defined, i.e., $x_n, y_n \in [a,b],\ n \geq 0$. They are monotone increasing and decreasing, respectively, i.e., $a = x_0 \leq x_1 \leq \ldots \leq x_n \leq \ldots$ and $b = y_0 \geq y_1 \geq \ldots y_n \geq \ldots$. A fixed point x_*, x^* of the operator T is called *minimal*, *maximal*, respectively, if for every fixed point z of T the inequalities $x_* \leq z$, $z \leq x^*$ hold, respectively.

Now,the following statement is valid (**Fig. 12.6a)**): Let X be an ordered Banach space with a closed cone X_+ and $T\colon D \longrightarrow X$, $D \subset X$ a continuous isotone increasing operator. Let $[a,b] \subset D$ be such that $T(a) \geq a$ and $T(b) \leq b$. Then $T([a,b]) \subset [a,b]$, and the operator T has a fixed point in $[a,b]$ if one of the following conditions is fulfilled:

a) X_+ is normal and T is compact;

b) X_+ is regular.

Then the sequences $\{x_n\}_{n=0}^{\infty}$ and $\{y_n\}_{n=0}^{\infty}$, defined in (12.205), converge to the minimal and to the maximal fixed points of T in $[a,b]$, respectively.

The notion of the *super- and sub-solutions* is based on these results (see [12.14]).

12.8.7 Monotone Operators in Banach Spaces

1. Special Properties

An arbitrary operator $T\colon D \subset X \longrightarrow Y$ (X, Y normed spaces) is called *demi-continuous* at the point $x_0 \in D$ if for each sequence $\{x_n\}_{n=1}^{\infty} \subset D$ converging to x_0 (in the norm of X) the sequence $\{T(x_n)\}_{n=1}^{\infty}$ converges weakly to $T(x_0)$ in Y. T is called *demi-continuous* on the set D if T is *demi-continuous* at every point of D.

In this paragraph another generalization of the notion of monotonicity known from real analysis are introduced. Let X now be a real Banach space, X^* its dual, $D \subset X$ and $T\colon D \longrightarrow X^*$ a non-linear operator. T is called *monotone* if $\forall\, x, y \in D$ the inequality $(T(x) - T(y), x - y) \geq 0$ holds. If $X = \mathbb{H}$ is a Hilbert space, then $(\cdot,\cdot)$ means the scalar product, while in the case of an arbitrary Banach space one refers to the notation introduced in 12.5.4.1, p. 681. The operator T is called *strongly monotone* if there is a constant $c > 0$ such that $(T(x) - T(y), x - y) \geq c\|x - y\|^2$ for $\forall\, x, y \in D$. An operator $T\colon X \longrightarrow X^*$ is called *coercive* if $\displaystyle\lim_{\|x\|\to\infty} \frac{(T(x), x)}{\|x\|} = \infty$.

2. Existence Theorems

for solutions of operator equations with monotone operators are given here only exemplarily: If the operator T, mapping the real separable Banach space X into X^*, ($D_T = X$), is monotone demi-continuous and coercive, then the equation $T(x) = f$ has a solution for arbitrary $f \in X^*$.

If in addition the operator T is strongly monotone, then the solution is unique. In this case the inverse operator T^{-1} also exists.

For a monotone, demi-continuous operator $T\colon \mathbb{H} \longrightarrow \mathbb{H}$ in a Hilbert space $\mathbb{H}$ with $D_T = \mathbb{H}$, there holds $Im(I+T) = \mathbb{H}$, where $(I+T)^{-1}$ is continuous. If T is supposed to be strongly monotone, then T^{-1} is bijective with a continuous T^{-1}.

Constructive approximation methods for the solution of the equation $T(x) = 0$ with a monotone operator T in a Hilbert space are based on the idea of Galerkin's method (see 19.4.2.2, p. 974, or [12.10], [12.18]). By means of this theory set-valued operators $T\colon X \longrightarrow 2^{X^*}$ can also be handled. The notion of monotonicity is then generalized by $(f - g, x - y) \geq 0$, $\forall\, x, y \in D_T$ and $f \in T(x), g \in T(y)$.

12.9 Measure and Lebesgue Integral

12.9.1 Set Algebras and Measures

The initial point for introducing measures is a generalization of the notion of the length of an interval in $\mathbb{R}$, of the area, and of the volume of subsets of $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. This generalization is necessary in order to "measure" as many sets as possible and to "make integrable" as many functions as possible. For instance, the volume of an n-dimensional rectangular parallelepiped

$$Q = \{x \in \mathbb{R}^n\colon\ a_k \leq x_k \leq b_k \quad (k = 1, 2, \ldots, n)\} \quad \text{has the value} \quad \prod_{k=1}^{n} (b_k - a_k). \tag{12.206}$$

1. σ Algebra or Set Algebra

Let X be an arbitrary set. A non-empty system $\mathcal{A}$ of subsets from X is called a *σ algebra* if:

a) $A \in \mathcal{A}$ implies $X \setminus A \in \mathcal{A}$ and (12.207a)

b) $A_1, A_2, \dots, A_n, \dots \in \mathcal{A}$ implies $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{A}$. (12.207b)

Every σ algebra contains the sets ∅ and X, the intersection of countably many of its sets and also the difference sets of any two of its sets.

In the following $\overline{\mathbb{R}}$ denotes the set $\mathbb{R}$ of real numbers extended by the elements $-\infty$ and $+\infty$ (extended real line), where the algebraic operations and the order properties from $\mathbb{R}$ are extended to $\overline{\mathbb{R}}$ in the natural way. The expressions $(\pm\infty) + (\mp\infty)$ and $\frac{\infty}{\infty}$ are meaningless, while $0 \cdot (+\infty)$ and $0 \cdot (-\infty)$ are assigned the value 0.

2. Measure

A function μ: $\mathcal{A} \longrightarrow \overline{\mathbb{R}}_+ = \mathbb{R}_+ \cup +\infty$, defined on a σ algebra $\mathcal{A}$, is called a *measure* if

a) $\mu(A) \ge 0 \quad \forall A \in \mathcal{A}$, (12.208a)

b) $\mu(\emptyset) = 0$, (12.208b)

c) $A_1, A_2, \dots A_n, \dots \in \mathcal{A},\ A_k \cap A_l = \emptyset \ (k \ne l)$ implies $\mu\left(\bigcup\limits_{n=1}^{\infty} A_n\right) = \sum\limits_{n=1}^{\infty} \mu(A_n)$. (12.208c)

The property c) is called *σ additivity* of the measure. If μ is a measure on $\mathcal{A}$, and for the sets $A, B \in \mathcal{A}$, $A \subset B$ holds, then $\mu(A) \le \mu(B)$ (*monotonicity*). If $A_n \in \mathcal{A}$ $(n = 1, 2, \dots)$ and $A_1 \subset A_2 \subset \cdots$, then $\mu\left(\bigcup\limits_{n=1}^{\infty} A_n\right) = \lim\limits_{n\to\infty} \mu(A_n)$ (*continuity from below*).

Let $\mathcal{A}$ be a σ algebra of subsets of X and μ a measure on $\mathcal{A}$. The triplet $X = (X, \mathcal{A}, \mu)$ is called a *measure space*, and the sets belonging to $\mathcal{A}$ are called *measurable* or *$\mathcal{A}$-measurable*.

■ **A: Counting Measure:** Let X be a finite set $\{x_1, x_2, \dots, x_N\}$, $\mathcal{A}$ the σ algebra of all subsets of X, and let assign a non-negative number p_k to each x_k $(k = 1, \dots, N)$. Then the function μ defined on $\mathcal{A}$ for every set $A \in \mathcal{A}$, $A = \{x_{n_1}, x_{n_2}, \dots, x_{n_k}\}$ by $\mu(A) = p_{n_1} + p_{n_2} + \cdots + p_{n_k}$ is a measure which takes on only finite values since $\mu(X) = p_1 + \cdots + p_N < \infty$. This measure is called the *counting measure*.

■ **B: Dirac Measure:** Let $\mathcal{A}$ be a σ algebra of subsets of a set X and a an arbitrary given point from X. Then a measure (called *Dirac Measure*) is defined on $\mathcal{A}$ by

$$\delta_a(A) = \begin{cases} 1, & \text{if } a \in A, \\ 0, & \text{if } a \notin A. \end{cases} \tag{12.209a}$$

It is called the *δ function* (concentrated on a). The *characteristic function* or *indicator function* of a subset $A \subseteq X$ denotes the function χ_A: $X \longrightarrow \{0, 1\}$ of X on $\{0, 1\}$, which has the value 1 for $x \in A$ and for all other x the value 0:

$$\chi_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{otherwise.} \end{cases} \tag{12.209b}$$

Obviously $\delta_a(A) = \delta_a(\chi_A) = \chi_A(a)$ (see 12.5.4, p. 681), where χ_A denotes the characteristic function of the set A.

■ **C: Lebesgue Measure:** Let X be a metric space and $\mathcal{B}(X)$ the smallest σ algebra of subsets of X which contains all the open sets from X. $\mathcal{B}(X)$ exists as the intersection of all the σ algebras containing all the open sets, and is called the *Borel σ algebra* of X. Every element from $\mathcal{B}(X)$ is called a *Borel set* (see [12.6]).

Suppose now, $X = \mathbb{R}^n$ $(n \ge 1)$. Using an extension procedure a σ algebra and a measure on it can be constructed, which coincides with the volume on the set of all rectangular parallelepipeds in $\mathbb{R}^n$. More precisely: There exists a uniquely defined σ algebra $\mathcal{A}$ of subsets of $\mathbb{R}^n$ and a uniquely defined measure λ on $\mathcal{A}$ with the following properties:

a) Each open set from $\mathbb{R}^n$ belongs to $\mathcal{A}$, in other words: $\mathcal{B}(\mathbb{R}^n) \subset \mathcal{A}$.

b) If $A \in \mathcal{A}$, $\lambda(A) = 0$ and $B \subset A$ then $B \in \mathcal{A}$ and $\lambda(B) = 0$.

c) If Q is a rectangular parallelepiped, then $Q \in \mathcal{A}$, and $\lambda(Q) = \prod_{k=1}^{n} (b_k - a_k)$.

d) λ is translation invariant, i.e., for every vector $x \in \mathbb{R}^n$ and every set $A \in \mathcal{A}$ one has $x + A = \{x + y\colon y \in A\} \in \mathcal{A}$ and $\lambda(x + A) = \lambda(A)$.

The elements of $\mathcal{A}$ are called *Lebesgue measurable* subsets of $\mathbb{R}^n$. λ is the (n-dimensional) *Lebesgue measure* in $\mathbb{R}^n$.

Remark: In measure theory and integration theory one says that a certain statement (property, or condition) with respect to the measure μ is valid *almost everywhere* or *μ-almost everywhere* on a set X, if the set, where the statement is not valid, has measure zero. It is denoted by a.e. or μ-a.e.[§] For instance, if λ is the Lebesgue measure on $\mathbb{R}$ and A, B are two disjoint sets with $\mathbb{R} = A \cup B$ and f is a function on $\mathbb{R}$ with $f(x) = 1,\ \forall x \in A$ and $f(x) = 0,\ \forall\, x \in B$, then $f = 1,\ \lambda$-a.e. on $\mathbb{R}$ if and only if $\lambda(B) = 0$.

12.9.2 Measurable Functions

12.9.2.1 Measurable Function

Let $\mathcal{A}$ be a σ algebra of subsets of a set X. A function $f\colon \mathrm{X} \longrightarrow \overline{\mathbb{R}}$ is called *measurable* if for an arbitrary $\alpha \in \mathbb{R}$ the set $f^{-1}((\alpha, +\infty]) = \{x : x \in \mathrm{X},\ f(x) > \alpha\}$ is in $\mathcal{A}$.
A complex-valued function $g + \mathrm{i}h$ is called measurable if both functions g and h are measurable. The characteristic function χ_A of every set $A \in \mathcal{A}$ is measurable, because

$$\chi_A^{-1}((\alpha, +\infty]) = \begin{cases} A, & \text{if } \alpha \in (-\infty, 1), \\ \emptyset, & \text{if } \alpha \geq 1 \end{cases} \qquad (12.210)$$

is valid (see Dirac measure, p. 694). If $\mathcal{A}$ is the σ algebra of the Lebesgue measurable sets of $\mathbb{R}^n$ and $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ is a continuous function, then the set $f^{-1}((\alpha, +\infty]) = f^{-1}((\alpha, +\infty))$, according to 12.2.3, p. 668, is open for every $\alpha \in \mathbb{R}$, hence f is measurable.

12.9.2.2 Properties of the Class of Measurable Functions

The notion of measurable functions requires no measure but a σ algebra. Let $\mathcal{A}$ be a σ algebra of subsets of the set X and let $f, g, f_n : \mathrm{X} \longrightarrow \overline{\mathbb{R}}$ be measurable functions. Then the following functions (see 12.1.7.4, p. 660) are also measurable:

a) αf for every $\alpha \in \mathbb{R}$; $f \cdot g$;

b) $f_+,\ f_-,\ |f|,\ f \vee g$ and $f \wedge g$;

c) $f + g$, if there is no point from X where the expression $(\pm\infty) + (\mp\infty)$ occurs;

d) $\sup f_n,\ \inf f_n,\ \limsup f_n\ (= \lim\limits_{n\to\infty} \sup\limits_{k\geq n} f_k),\ \liminf f_n$;

e) the point-wise limit $\lim f_n$, in case it exists;

f) if $f \geq 0$ and $p \in \mathbb{R}, p > 0$, the f^p is measurable.

A function $f\colon \mathrm{X} \longrightarrow \mathbb{R}$ is called *elementary* or *simple* if there is a finite number of pairwise disjoint sets $A_1, \ldots, A_n \in \mathcal{A}$ and real numbers $\alpha_1, \ldots, \alpha_n$ such that $f = \sum_{k=1}^{n} \alpha_k \chi_k$, where χ_k denotes the characteristic function of the set A_k. Since each characteristic function of a measurable set is measurable (see (12.210)), so every elementary function is measurable. It is interesting that each measurable function can be approximated arbitrarily well by elementary functions: For each measurable function $f \geq 0$ there exists a monotone increasing sequence of non-negative elementary functions, which converges point-wise to f.

[§]Here and in the following parts "a.e." is an abbreviation for "almost everywhere".

12.9.3 Integration

12.9.3.1 Definition of the Integral

Let $(\mathrm{X}, \mathcal{A}, \mu)$ be a measure space. The integral $\int_X f\, d\mu$ (also denoted by $\int f\, d\mu$) for a measurable function f is defined by means of the following steps:

1. If f is an elementary function $f = \sum\limits_{k=1}^{n} \alpha_k \chi_k$, then

$$\int f\, d\mu = \sum_{k=1}^{n} \alpha_k \mu(A_k). \tag{12.211}$$

2. If $f\colon \mathrm{X} \longrightarrow \overline{\mathbb{R}}$ $(f \geq 0)$, then

$$\int f\, d\mu = \sup\left\{\int g\, d\mu : \; g \text{ is an elementary function with } 0 \leq g(x) \leq f(x),\ \forall\, x \in \mathrm{X}\right\}. \tag{12.212}$$

3. If $f\colon \mathrm{X} \longrightarrow \overline{\mathbb{R}}$ and f_+, f_- are the positive and the negative parts of f, then

$$\int f\, d\mu = \int f_+\, d\mu - \int f_-\, d\mu \tag{12.213}$$

under the condition that at least one of the integrals on the right side is finite (in order to avoid the meaningless expression $\infty - \infty$).

4. For a complex-valued function $f = g + \mathrm{i}h$, if the integrals (12.213) of the functions g, h are finite, put

$$\int f\, d\mu = \int g\, d\mu + \mathrm{i} \int h\, d\mu. \tag{12.214}$$

5. If for any measurable set A and a function f there exists the integral of the function $f\chi_A$ then put

$$\int\limits_A f\, d\mu := \int f\chi_A\, d\mu. \tag{12.215}$$

The integral of a measurable function is in general a number from $\overline{\mathbb{R}}$. A function $f\colon \mathrm{X} \longrightarrow \overline{\mathbb{R}}$ is called *integrable* or *summable* over X with respect to μ if it is measurable and $\int |f|\, d\mu < \infty$.

12.9.3.2 Some Properties of the Integral

Let $(\mathrm{X}, \mathcal{A}, \mu)$ be a measure space, $f, g : \mathrm{X} \longrightarrow \overline{\mathbb{R}}$ be measurable functions and $\alpha, \beta \in \mathbb{R}$.

1. If f is integrable, then f is finite a.e., i.e., $\mu\{x \in \mathrm{X}\colon |f(x)| = +\infty\} = 0$.

2. If f is integrable, then $\left|\int f\, d\mu\right| \leq \int |f|\, d\mu$.

3. If f is integrable and $f \geq 0$, then $\int f\, d\mu \geq 0$.

4. If $0 \leq g(x) \leq f(x)$ on X and f is integrable, then g is also integrable, and $\int g\, d\mu \leq \int f\, d\mu$.

5. If f, g are integrable, then $\alpha f + \beta g$ is integrable, and $\int (\alpha f + \beta g)\, d\mu = \alpha \int f\, d\mu + \beta \int g\, d\mu$.

6. If f, g are integrable on $A \in \mathcal{A}$, i.e., there exist the integrals $\int_A f\, d\mu$ and $\int_A g\, d\mu$ according to (12.215) and $f = g$ μ-a.e. on A, then $\int_A f\, d\mu = \int_A g\, d\mu$.

If $\mathrm{X} = \mathbb{R}^n$ and λ is the Lebesgue measure, then the introduced integral is the (n-dimensional) *Lebesgue integral* (see also 8.2.3.1, **3.**, p. 507). In the case $n = 1$ and $A = [a, b]$, for every continuous function

f on $[a,b]$ both the Riemann integral $\int_a^b f(x)\,dx$ (see 8.2.1.1, 2., p. 494) and the Lebesgue integral $\int_{[a,b]} f\,d\lambda$ are defined. Both values are finite and equal to each other. Furthermore, if f is a bounded Riemann integrable function on $[a,b]$, then it is also Lebesgue integrable and the values of the two integrals coincide.

The set of Lebesgue integrable functions is considerably larger than the set of the Riemann integrable functions and it has several advantages, e.g., when passing to the limit under the integral sign and f, $|f|$ are Lebesgue integrable simultaneously.

12.9.3.3 Convergence Theorems

Now Lebesgue measurable functions will be considered throughout.

1. B. Levi's Theorem on Monotone Convergence

Let $\{f_n\}_{n=1}^{\infty}$ be an a.e. monotone increasing sequence of non-negative integrable functions with values in $\overline{\mathbb{R}}$. Then

$$\lim_{n\to\infty}\int f_n\,d\mu = \int \lim_{n\to\infty} f_n\,d\mu. \tag{12.216}$$

2. Fatou's Theorem

Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of non-negative $\overline{\mathbb{R}}$-valued measurable functions. Then

$$\int \liminf f_n\,d\mu \le \liminf \int f_n\,d\mu. \tag{12.217}$$

3. Lebesgue's Dominated Convergence Theorem

Let $\{f_n\}$ be a sequence of measurable functions convergent on X a.e. to some function f. If there exists an integrable function g such that $|f_n| \le g$ a.e., then $f = \lim\limits_{n\to\infty} f_n$ is integrable and there holds

$$\lim_{n\to\infty}\int f_n\,d\mu = \int \lim_{n\to\infty} f_n\,d\mu\,. \tag{12.218}$$

4. Radon-Nikodym Theorem

a) Assumptions: Let $(\mathrm{X},\mathcal{A},\mu)$ be a σ-finite measure space, i.e., there exists a sequence $\{A_n\}$, $A_n \in \mathcal{A}$ such that $\mathrm{X} = \bigcup\limits_{n=1}^{\infty} A_n$ and $\mu(A_n) < \infty$ for $\forall\, n$. In this case the measure is called σ *finite*. It is called *finite* if $\mu(\mathrm{X}) < \infty$, and it is called a *probability measure* if $\mu(\mathrm{X}) = 1$. A real function φ defined on $\mathcal{A}$ is called *absolutely continuous* with respect to μ if $\mu(A) = 0$ implies $\varphi(A) = 0$. This property is denoted by $\varphi \prec \mu$.
For an integrable function f, the function φ defined on $\mathcal{A}$ by $\varphi(A) = \int_A f\,d\mu$ is σ additive and absolutely continuous with respect to the measure μ. The converse of this property plays a fundamental role in many theoretical investigations and practical applications:

b) Radon–Nikodym Theorem: Suppose a σ-additive function φ and a measure μ are given on a σ algebra $\mathcal{A}$, and let $\varphi \prec \mu$. Then there exists a μ-integrable function f such that for each set $A \in \mathcal{A}$,

$$\varphi(A) = \int\limits_A f\,d\mu. \tag{12.219}$$

The function f is uniquely determined up to its equivalence class, and φ is non-negative if and only if $f \ge 0$ μ-a.e.

12.9.4 L^p Spaces

Let $(\mathrm{X},\mathcal{A},\mu)$ be a measure space and p a real number $1 \le p < \infty$. For a measurable function f, according to 12.9.2.2, p. 695, the function $|f|^p$ is measurable as well, so the expression

$$N_p(f) = \left(\int |f|^p \, d\mu\right)^{\frac{1}{p}} \tag{12.220}$$

is defined (and may be equal to $+\infty$). A measurable function $f\colon \mathrm{X} \longrightarrow \overline{\mathbb{R}}$ is called *p-th power integrable*, or an *L^p-function* if $N_p(f) < +\infty$ holds or, equivalent to this, if $|f|^p$ is integrable.

For every p with $1 \le p < +\infty$, the set of all L^p-functions, i.e., all functions p-th power integrable with respect to μ on X, is denoted by $\mathcal{L}^p(\mu)$ or by $\mathcal{L}^p(\mathrm{X})$ or in full detail $\mathcal{L}^p(\mathrm{X}, \mathcal{A}, \mu)$. For $p = 1$ the simple notation $\mathcal{L}(\mathrm{X})$ is used. For $p = 2$ the functions are called *quadratically integrable*.

The set of all measurable μ-a.e. bounded functions on X is denoted by $\mathcal{L}^\infty(\mu)$ and the essential supremum of a function f is defined as

$$N_\infty(f) = \text{ess. sup}\, f = \inf\{a \in \mathbb{R} : |f(x)| \le a \ \ \mu\text{–a.e.}\}. \tag{12.221}$$

$\mathcal{L}^p(\mu)$ $(1 \le p \le \infty)$ equipped with the usual operations for measurable functions and taking into consideration Minkowski inequality for integrals (see 1.4.2.13, p. 32), is a vector space and $N_p(\cdot)$ is a semi-norm on $\mathcal{L}^p(\mu)$. If $f \le g$ means that $f(x) \le g(x)$ holds μ-a.e., then $\mathcal{L}^p(\mu)$ is also a vector lattice and even a K-space (see 12.1.7.4, p. 660). Two functions $f, g \in \mathcal{L}^p(\mu)$ are called *equivalent* (or they are declared as equal) if $f = g$ μ-a.e. on X. In this way, functions are considered to be identical if they are equal μ-a.e. The factorization of the set $\mathcal{L}^p(\mathrm{X})$ modulo the linear subspace $N_p^{-1}(0)$ leads to a set of equivalence classes on which the algebraic operations and the order can be transferred naturally. So a vector lattice (K-space) is obtained again, which is denoted now by $L^p(\mathrm{X}, \mu)$ or $L^p(\mu)$. Its elements are called functions, as before, but actually they are classes of equivalent functions.

It is very important that $\|\hat{f}\|_p = N_p(f)$ is now a norm on $L^p(\mu)$ ($\hat{f}$ stands here for the equivalence class of f, which will later be denoted simply by f), and $(L^p(\mu), \|f\|_p)$ for every p with $1 \le p \le +\infty$ is a Banach lattice with several good compatibility conditions between norm and order. For $p = 2$ with $(f, g) = \displaystyle\int f\overline{g}\, d\mu$ as a scalar product, $L^2(\mu)$ is also a Hilbert space (see [12.12]).

Very often the space $L^p(\Omega)$ is considered for a measurable subset $\Omega \subset \mathbb{R}^n$. Its definition is not a problem because of step **5** in (12.9.3.1, p. 696).

The spaces $L^p(\Omega, \lambda)$, where λ is the n-dimensional Lebesgue measure, can also be introduced as the completions (see 12.2.2.5, p. 668 and 12.3.2, p. 670) of the non-complete normed spaces $\mathcal{C}(\Omega)$ of all continuous functions on the set $\Omega \subset \mathbb{R}^n$ equipped with the integral norm $\|x\|_p = \left(\displaystyle\int |x|^p \, d\lambda\right)^{\frac{1}{p}}$ $(1 \le p < \infty)$ (see [12.18]).

Let X be a set with a finite measure, i.e., $\mu(\mathrm{X}) < +\infty$, and suppose for the real numbers p_1, p_2, $1 \le p_1 < p_2 \le +\infty$. Then $L^{p_2}(\mathrm{X}, \mu) \subset L^{p_1}(\mathrm{X}, \mu)$, and with a constant $C = C(p_1, p_2, \mu(\mathrm{X})) > 0$ (independent of x), there holds the estimation $\|x\|_1 \le C\|x\|_2$ for $x \in L^{p_2}$ (here $\|x\|_k$ denotes the norm of the space $L^{p_k}(\mathrm{X}, \mu)$ $(k = 1, 2)$).

12.9.5 Distributions

12.9.5.1 Formula of Partial Integration

For an arbitrary (open) domain $\Omega \subseteq \mathbb{R}^n$, $\mathcal{C}_0^\infty(\Omega)$ denotes the set of all arbitrary many times in Ω differentiable functions φ with compact support, i.e., the set $\text{supp}(\varphi) = \overline{\{x \in \Omega : \varphi(x) \ne 0\}}$ is compact in $\mathbb{R}^n$ and lies in Ω. The set of all *locally summable* functions with respect to the Lebesgue measure in $\mathbb{R}^n$ is doted by $L^1_{loc}(\Omega)$, i.e., all the measurable functions f (equivalent classes) on Ω such that $\int_\omega |f|\, d\lambda < +\infty$ for every bounded domain $\omega \subset \Omega$.

Both sets are vector spaces (with the natural algebraic operations).

There hold $L^p(\Omega) \subset L^1_{loc}(\Omega)$ for $1 \le p \le \infty$, and $L^1_{loc}(\Omega) = L^1(\Omega)$ for a bounded Ω. If the elements of $\mathcal{C}^k(\overline{\Omega})$ are considered as the classes generated by them in $L^p(\Omega)$, then the inclusion $\mathcal{C}^k(\overline{\Omega}) \subset L^p(\Omega)$ holds for bounded Ω, where $\mathcal{C}^k(\overline{\Omega})$ is at once dense. If Ω is unbounded, then the set $\mathcal{C}_0^\infty(\Omega)$ is dense (in this sense) in $L^p(\Omega)$.

For a given function $f \in \mathcal{C}^k(\overline{\Omega})$ and an arbitrary function $\varphi \in \mathcal{C}_0^\infty(\Omega)$ the formula of partial integration has the form

$$\int\limits_\Omega f(x)D^\alpha\varphi(x)\,d\lambda = (-1)^{|\alpha|}\int\limits_\Omega \varphi(x)D^\alpha f(x)\,d\lambda \tag{12.222}$$

$\forall\ \alpha$ with $|\alpha| \le k$ (the fact that $D^\alpha\varphi|_{\partial\Omega} = 0$ is used), and will be taken as the starting point for the definition of the generalized derivative of a function $f \in L^1_{loc}(\Omega)$.

12.9.5.2 Generalized Derivative

Suppose $f \in L^1_{loc}(\Omega)$. If there exists a function $g \in L^1_{loc}(\Omega)$ such that $\forall\ \varphi \in \mathcal{C}_0^\infty(\Omega)$ with respect to some multi-index α the equation

$$\int\limits_\Omega f(x)D^\alpha\varphi(x)\,d\lambda = (-1)^{|\alpha|}\int\limits_\Omega g(x)\varphi(x)\,d\lambda \tag{12.223}$$

holds, then g is called the *generalized derivative* (*derivative in the Sobolev sense* or *distributional derivative*) of order α of f. It is denoted by $g = D^\alpha f$ as in the classical case.

The convergence of a sequence $\{\varphi_k\}_{k=1}^\infty$ in the vector space $\mathcal{C}_0^\infty(\Omega)$ to $\varphi \in \mathcal{C}_0^\infty(\Omega)$ is defined as

$$\varphi_k \longrightarrow \varphi \ \text{ if and only if } \begin{cases} \text{a) } \exists \text{ a compact set } K \subset \Omega \text{ with } \operatorname{supp}(\varphi_k) \subset K \text{ for any } k\,, \\ \text{b) } D^\alpha\varphi_k \to D^\alpha\varphi \text{ uniformly on } K \text{ for each multi-index } \alpha\,. \end{cases} \tag{12.224}$$

The set $\mathcal{C}_0^\infty(\Omega)$, equipped with this convergence of sequences, is called the *fundamental space*, and is denoted by $\mathcal{D}(\Omega)$. Its elements are often called test functions.

12.9.5.3 Distributions

A linear functional ℓ on $\mathcal{D}(\Omega)$ continuous in the following sense (see 12.2.3, p. 668):

$$\varphi_k,\ \varphi \in \mathcal{D}(\Omega) \ \text{ and } \ \varphi_k \longrightarrow \varphi \ \text{ imply } \ \ell(\varphi_k) \longrightarrow \ell(\varphi) \tag{12.225}$$

is called a *generalized function* or a *distribution*.

■ **A:** If $f \in L^1_{loc}(\Omega)$, then

$$\ell_f(\varphi) = (f, \varphi) = \int_\Omega f(x)\varphi(x)\,d\lambda, \quad \varphi \in \mathcal{D}(\Omega) \tag{12.226}$$

is a distribution. A distribution, defined by a locally summable function as in (12.226), is called *regular*. Two regular distributions are equal, i.e., $\ell_f(\varphi) = \ell_g(\varphi)\ \forall\varphi \in \mathcal{D}(\Omega)$, if and only if $f = g$ a.e. with respect to λ.

■ **B:** Let $a \in \Omega$ be an arbitrary fixed point. Then $\ell_{\delta_a}(\varphi) = \varphi(a),\ \varphi \in \mathcal{D}(\Omega)$ is a linear continuous functional on $\mathcal{D}(\Omega)$, hence a distribution, which is called the Dirac distribution, δ distribution or δ function.

Since ℓ_{δ_a} cannot be generated by any locally summable function (see [12.11], [12.24]), it is an example for a non-regular distribution.

The set of all distributions is denoted by $\mathcal{D}'(\Omega)$. From a more general duality theory than that discussed in 12.5.4, p. 681, $\mathcal{D}'(\Omega)$ can be obtained as the dual space of $\mathcal{D}(\Omega)$. Consequently, one should write $\mathcal{D}^*(\Omega)$ instead. In the space $\mathcal{D}'(\Omega)$, it is possible to define several operations with its elements and with functions from $\mathcal{C}^\infty(\Omega)$, e.g., the derivative of a distribution or the convolution of two distributions, which make $\mathcal{D}'(\Omega)$ important not only in theoretical investigations but also in practical applications in electrical engineering, mechanics, etc.

For a review and for simple examples in applications of generalized functions see, e.g., [12.11],[12.24].

Here, only the notion of the derivative of a generalized function is discussed.

12.9.5.4 Derivative of a Distribution

If ℓ is a given distribution, then the distribution $D^\alpha \ell$ defined by

$$(D^\alpha \ell)(\varphi) = (-1)^{|\alpha|}\ell(D^\alpha \varphi), \quad \varphi \in D(\Omega), \tag{12.227}$$

is called the *distributional derivative* of order α of ℓ.
Let f be a continuously differentiable function, say on $\mathbb{R}$ (so f is locally summable on $\mathbb{R}$, and f can be considered as a distribution), let f' be its classical derivative and $D^1 f$ its distributional derivative of order 1. Then:

$$(D^1 f, \varphi) = \int_{\mathrm{R}} f'(x)\varphi(x)\,dx, \tag{12.228a}$$

from which by partial integration there follows

$$(D^1 f, \varphi) = -\int_{\mathrm{R}} f(x)\varphi'(x)\,dx = -(f, \varphi'). \tag{12.228b}$$

In the case of a regular distribution ℓ_f with $f \in L^1_{loc}(\Omega)$ by using (12.226)

$$(D^\alpha \ell_f)(\varphi) = (-1)^{|\alpha|}\ell_f(D^\alpha \varphi) = (-1)^{|\alpha|}\int_\Omega f(x) D^\alpha \varphi\, d\lambda \tag{12.229}$$

is obtained, which is the generalized derivative of the function f in the Sobolev sense (see (12.223)).

■ **A:** For the regular distribution generated by the (obviously locally summable) Heaviside function

$$\Theta(x) = \begin{cases} 1 \text{ for } x \geq 0, \\ 0 \text{ for } x < 0 \end{cases} \tag{12.230}$$

the non-regular δ distribution is obtained as the derivative.

■ **B:** In mathematical modeling of technical and physical problems one is faced with (in a certain sense idealized) influences concentrated at one point, such as a "point-like" force, needle-deflection, collision, etc., which can be expressed mathematically by using the δ or Heaviside function. For example, $m\delta_a$ is the mass density of a point-like mass m concentrated at one point a $(0 \leq a \leq l)$ of a beam of length l.
The motion of a spring-mass system on which at time t_0 there acts a momentary external force F is described by the equation $\ddot{x} + \omega^2 x = F\delta_{t_0}$. With the initial conditions $x(0) = \dot{x}(0) = 0$ its solution is $x(t) = \dfrac{F}{\omega}\sin(\omega(t - t_0))\Theta(t - t_0)$.