INSTITUTE OF MATHEMATICS
of the
Polish Academy of Sciences

ul. Śniadeckich 8, P.O.B. 21, 00-956 Warszawa 10, Poland http://www.impan.gov.pl

Danuta Przeworska-Rolewicz

LINEAR SPACES and LINEAR OPERATORS

Warsaw 2007

LINEAR SPACES AND LINEAR OPERATORS

TABLE OF CONTENTS

PREFACE

Putting this book into hands of readers, I hope that it will serve mainly to mathematicians and engineers of different specialities, among others, those who work in automatic control theory, electronics, informatics and electrical engineering, as well as economists. It could be also used by students of the corresponding faculties as an auxiliary book, in order to make a "bridge" among the first year courses and more advanced parts of mathematics necessary in the theory and practice of engineering, economics and related topics.

In this book there are given basic notions and theorems concerning linear spaces of finite and infinite dimension and linear operators in these spaces in a possibly uniform way.

In order to understand the subject presented here, the reader should be acquainted with fundamental notions and theorems of logic, set theory and algebra.

Theoretical considerations are complemented by several examples and exercises.

Chapters 1-9 are a slightly extended version of a book edited in Polish in 1977 (cf. the author, PR[2]) related to my lectures for students of the 1st course at the Cybernetics Faculty of the Engineering Military Academy in Warsaw in years 1973-75. This was connected with a new programme of Mathematics based on ideas of Algebraic Analysis (cf. the author, PR[3]) then prepared by mathematicians and engineers from this school working in the Operations Research Department. Also we have prepared for this course new textbooks. Note that the name "Algebraic Analysis" has been introduced by J. L. LAGRANGE in the years 1797-1813 to emphasize that the analysis under consideration was more or less "different" from other concepts of analysis at that time. In the abovementioned book (PR[3]) this name is used for an algebraic approach to Calculus which is very efficient in teaching. For instance, with a well prepared programme one can start with solving of differential equations at the first semester of studies, which we did, indeed. However, it should be pointed out that this programme requires a particular team to be realized in a good manner.

Chapter 5 contains results of my works concerning algebraic operators. Theorems in Chapter 8 and 9 about the index and perturbations of linear operators were obtained together with Stefan ROLEWICZ (cf. PRR[1]).

A uniform treatment of Linear Algebra in infinite and finite dimensional spaces shows the unity of some ideas - independently of its traditional classifications. In addition, this approach permits to make shorter and simpler some proofs (for instance, Jordan theorem, Sylvester inertia law, and so on).

It should be also pointed out that Examples and Exercises make an essential part of the book for notions introduced and applied there. Note that Examples and Exercises end with the sign □, while proofs of theorems end with the sign ■.

The present book is not a literal translation of the first Polish edition (cf. PR[2]). There are some corrections, also some parts are either complemented or little changed. However, the main stream of the Polish version is preserved. I repeat, I hope that this book will serve readers well, as was the case with the first edition.

Finally, I should say that I am and I will be always very indebted to my husband, Professor Stefan Rolewicz, for his permanent assistance during 55 years of our common life.

Warszawa, September 2, 2006

Danuta Przeworska-Rolewicz

Chapter 1.

Linear spaces.

As usual, denote by $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ the sets of positive integers, integers, rational numbers, real numbers and complex numbers, respectively.

Denote by $\mathbb{F}$ a field of numbers, which we shall call *field of scalars*. Here $\mathbb{F}$ is either the field $\mathbb{R}$ of real numbers or the field $\mathbb{C}$ of complex numbers. Elements of the field $\mathbb{F}$ will be called *scalars*.

A commutative group X is said to be a *linear space* over a field $\mathbb{F}$ of scalars if in X besides the operation of addition of elements there is defined *multiplication of elements by scalars* satisfying the following conditions:

$$t(x+y) = tx + ty; \quad (t+s)x = tx + sx; \quad (ts)x = t(sx) \tag{1.1}$$

$$for\ arbitrary \quad x, y \in X,\ t, s \in \mathbb{F},$$

i.e. the multiplication by scalars is *distributive* with respect to the addition of elements and scalars and *associative*. These conditions imply that

$$1 \cdot x = x; \quad 0 \cdot x = 0 \quad for\ all\ x, y \in X \tag{1.2}$$

Indeed, if $t \neq 0$ is an arbitrarily fixed scalar, then

$$0 \cdot x = (t-t)x = tx - tx = 0; \quad t(1 \cdot x) = (t \cdot 1)x = tx.$$

Note that on the right hand side of the equality $0 \cdot x = 0$ we have the neutral element of the group X, which without any misunderstanding can be also denoted by "0".

Conditions (1.1) also imply that

$$if \quad tx = 0 \quad and \quad x \neq 0 \quad then \quad t = 0.$$

Indeed, suppose that $tx = 0$, $x \neq 0$ and $t \neq 0$. Then

$$1 \cdot x = (t^{-1}t)x = t^{-1}(tx) = t^{-1} \cdot 0 = 0,$$

which contradicts our assumption that x is different than zero.

Linear spaces either over the field $\mathbb{R}$ of reals or over the field $\mathbb{C}$ of complex numbers are the most often considered.

A linear space over the field $\mathbb{R}$ of reals (over the field $\mathbb{C}$ of complex numbers) is called sometimes briefly *real (complex) linear space*. Since most theorems for linear spaces over the fields $\mathbb{R}$ and $\mathbb{C}$ are the same, by the term "linear space" we shall understand both kinds of spaces, if it does not lead to any misunderstanding. Similarly, by words: "number" and "scalar" we shall understand elements of the field under consideration.

Suppose that X is a linear space over a field $\mathbb{F}$, $Y \subset X$ and the sum of two elements belonging to Y and the product of an arbitrary element $y \in Y$ by a scalar belonging to $\mathbb{F}$ again belong to Y. In other words, Y is a linear space with respect to the same operations as the space X. Then Y is said to be a *linear subset* or a *subspace* of the space X. Hence a subset Y of a linear space X over a field $\mathbb{F}$ is a linear subset (subspace) of X if for arbitrary $x, y \in X$, $t \in \mathbb{F}$ we have $x + y \in Y$, $tx \in Y$.

Suppose that Y is an arbitrary subset of a linear space X. The least linear subset containing Y is said to be a *linear space spanned by the set* Y (otherwise called *linear span*) of Y and it is denoted by lin Y.

Theorem 1.1. *If X is a linear space over a field $\mathbb{F}$ and $Y \subset X$ then*

$$\text{lin } Y = \big\{x \in X : x = \sum_{j=1}^{n} t_j x_j, \quad \textit{where } t_1, ..., t_n \in \mathbb{F};\ x_1, ..., x_n \in Y\big\}.$$

Proof. Write

$$Y_1 = \big\{x \in X : x = \sum_{j=1}^{n} t_j x_j, \quad \textit{where } t_1, ..., t_n \in \mathbb{F};\ x_1, ..., x_n \in Y\big\}.$$

By definition of Y_1, we have $Y \subset Y_1$. Clearly, Y_1 is a linear subset of the space X. If $Y_2 \supset Y$ is another linear subset of the space X then, by the definition of a linear subset, it should contains elements of the form $\sum_{j=1}^{n} t_j x_j$. We therefore conclude that $Y_1 = Y_2$. ■

Elements of the form

$$\sum_{j=1}^{n} t_j x_j, \quad \textit{where } t_1, ..., t_n \in \mathbb{F};\ x_1, ..., x_n \in X,$$

are said to be *linear combinations* of elements $x_1, ..., x_n$.

We say that an element $x \in X$ is *linearly dependent on a set* $Y \subset X$ (or : on elements of the set Y) if $x \in$ lin Y, i.e. if there exist $x_1, ..., x_n \in Y$ and $t_1, ..., t_n \in \mathbb{F}$ such that $x = \sum_{j=1}^{n} t_j x_j$. In other words: *an element $x \in X$ is linearly dependent on $x_1, ..., x_n$ if x is a linear combination of these elements.*

Elements of a set $Y \subset X$ are said to be *linearly independent* if there is no element $x \in Y$ linearly dependent on the set of the remaining elements, i.e. if $x \notin \text{lin}\,\big(Y \setminus \{x\}\big)$ for every $x \in Y$. By the definition of the set lin Y, elements $x_1, ..., x_n \in Y$ are linearly independent if

$$t_1 x_1 + ... + t_n x_n = 0 \quad \textit{implies} \quad t_1 = t_2 = ... = t_n = 0,$$

which means that the only vanishing linear combination of elements $x_1, ..., x_n$ is a combination with all coefficients equal to zero.

A linear space X is said to be *n-dimensional* if n is the least number of linearly independent elements $x_1, ..., x_n \in X$ such that

$$\text{lin}\ \{x_1, ..., x_n\} = X.$$

If this is the case, then the number n is said to be the *dimension* of the space X and we write dim $X = n$.

If there is a positive integer n such that dim $X = n$, then X is said to be *finite dimensional* and we write dim $X < +\infty$.

If X is not finite dimensional, then we say that X if *infinite dimensional* and we write dim $X = +\infty$.

Note that a finite dimensional space X considered simultaneously over the field $\mathbb{R}$ of reals and over the field $\mathbb{C}$ of complexes, has in the first case the dimension twice bigger than in the second one. It follows from the fact that the set of complex numbers considered as a real space (i.e. with the usual addition of complex numbers, but with the multiplication only by real numbers) has the dimension 2. Thus, whenever we shall consider a linear space over the field $\mathbb{C}$, we shall understand by its dimension the dimension of that space over the complex numbers.

A set (a system) $\mathcal{B}$ of elements of a linear space X is said to be a *basis* if every element $x \in X$ can be represented as a linear combination of elements belonging to $\mathcal{B}$ and this representation is uniquely determined. The uniqueness of that representation implies that the set $\mathcal{B}$ consists of linearly independent elements.

Indeed, suppose that elements $x_1, ..., x_n \in \mathcal{B}$ are linearly dependent. Then there are scalars $a_1, ..., a_n$ non vanishing simultaneously such that $a_1x_1 + ... + a_nx_n = 0$. Thus zero has two different representations, which contradicts to our assumption.

A consequence of Theorem 1.1 is the following

Corollary 1.1. *If X is a linear space and $Y \subset X$ is a set of linearly independent elements, then Y is a basis in* lin Y.

Proof. By Theorem 1.1, every element of the linear span lin Y can be represented as a linear combination of elements belonging to Y. The uniqueness of this representation follows from the linear independence of elements of the set Y. ■

Suppose now that we are given a system $X_1, ..., X_n$ of linear spaces, all of them over the same field $\mathbb{F}$ of scalars. The set X whose elements are all ordered n-tuples of n elements $x = (x_1, ..., x_n)$, where $x_j \in X_j$ $(j = 1, 2, ..., n)$, is said to be the *Cartesian product* of spaces $X_1, ..., X_n$.

The addition and the multiplication by scalars of elements of the set X are defined by formulae:

$$(x_1, ..., x_n) + (y_1, ..., y_n) = (x_1 + y_1, ..., x_n + y_n); \quad t(x_1, ..., x_n) = (tx_1, ..., tx_n),$$

$$\textit{where } x_j, y_j \in X_j \ (j = 1, ..., n); \quad t \in \mathbb{F}.$$

It is easy to check that the set X with these operations of addition and multiplication by scalars is a linear space over the field $\mathbb{F}$.

The Cartesian product X of spaces $X_1, ..., X_n$ usually is denoted as follows:

$$X = X_1 \times X_2 \times ... \times X_n.$$

If $X_1, ..., X_n$ are subspaces of a linear space X then the set

$$X_1 + ... + X_n = \{x_1 + ... + x_n : x_j \in X_j,\ j = 1, ..., n\}$$

is said to be *algebraic sum* of subspaces $X_1, ..., X_n$.

If $X_1, ..., X_n$ are subspaces of a linear space X such that

$$X_j \cap \Big[\bigcup_{k \neq j}^{n} X_k\Big] = \{0\} \quad for\ j \neq k \quad (j, k = 1, ..., n), \tag{1.4}$$

then the algebraic sum $X_1 + ... + X_n$ is called the *direct sum* of spaces $X_1, ..., X_n$ and is denoted by

$$X_1 \oplus ... \oplus X_n.$$

Note that from Condition (1.4) it follows that every element x of the direct sum $X_1 \oplus ... \oplus X_n$ can be written in a unique way in the form $x = x_1 + ... + x_n$, where $x_j \in X_j$ $(j = 1, ..., n)$.

If $X = X_1 \oplus ... \oplus X_n$ then we say that X is *decomposed onto the direct sum* of spaces $X_1, ..., X_n$.

Suppose that X is a linear space over the field $\mathbb{R}$ of reals. This space can be in a natural way "embedded" in a linear space over the field $\mathbb{C}$ of complexes.

Indeed, consider the set of all ordered pairs (x, y), where $x, y \in X$, with the addition and multiplication by scalars defined by formulae:

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2) \quad for\ x_1, x_2, y_1, y_2 \in X, \tag{1.5}$$

$$(a + ib)(x, y) = (ax - by, ay + bx) \quad for\ x, y \in X,\ a + ib \in \mathbb{C}. \tag{1.6}$$

It is easy to verify that these operations satisfy the distribution conditions assumed in the definition of a linear space. Thus it is enough to show that multiplication by scalars is associative. Indeed, if $a + ib,\ c + id \in \mathbb{C}$ and $x, y \in X$, then we have

$$[(a + ib)(c + id)](x, y) = [ac - bd + i(bc + ad)](x, y) =$$

$$= \big((ac - bd)x - (bc + ad)y, (ac - bd)y + (bc + ad)x\big),$$

$$(a + ib)[(c + id)(x, y)] = (a + ib)(cx - dy, cy + dx) =$$

$$= \big((ac - bd)x - (bc + ad)y, (ac - bd)y + (bc + ad)x\big).$$

We shall denote a linear space over the field $\mathbb{C}$ defined in this way by $X + iX$. By definition, $X \times \{0\} \subset X + iX$.

Let Y be a subspace of a linear space X over a field $\mathbb{F}$. To every element $x \in X$ there corresponds the set

$$[x] = \{x + y : y \in Y\} = x + Y. \tag{1.7}$$

The set $[x]$ is called a *coset* induced by an element $x \in X$ and the subspace Y. Note that without any misunderstanding we can write here $x + Y$ instead of $\{x\} + Y$.

The coset $[x]$ can be written also in another equivalent form:

$$[x] = \{z \in X : x - z \in Y\}. \tag{1.8}$$

We shall prove that two cosets are either disjoint or equal. Indeed, let x_1, x_2 be arbitrary elements of the space X such that $[x_1] \cap [x_2] \neq \emptyset$. Then there is an element $z \in X$ such that

$$z = x_1 + y_1 \quad \textit{for a } y_1 \in Y \textit{ and, simultaneously, } z = x_2 + y_2 \quad \textit{for a } y_2 \in Y.$$

This implies

$$x_2 - x_1 = z - y_2 - (z - y_1) = y_1 - y_2, \quad i.e. \quad x_1 \in [x_2].$$

Hence $[x_2] \subset [x_1]$. Similarly, we prove that $[x_1] \subset [x_2]$. Thus $[x_2] = [x_1]$.

Denote by

$$X/Y = \{[x] : x \in X\} \tag{1.9}$$

the set of all cosets induced by elements $x \in X$ and determine in the set X/Y the addition and the multiplication by scalars as follows:

$$[x] + [y] = [x + y]; \quad t[x] = [tx] \quad \textit{for } x, y \in X,\ t \in \mathbb{F}. \tag{1.10}$$

It is easy to verify that the set X/Y with the addition and multiplication by scalars defined by Formulae (1.10) is again a linear space over the field $\mathbb{F}$. The space X/Y is said to be a *quotient space*. By the definition of a coset, we conclude that $[0] = Y$. Hence the quotient space X/Y has the set Y as the neutral element.

By the *defect* or *codimension* of a subspace Y of a linear space X over a field $\mathbb{F}$, we mean the dimension of the quotient space X/Y. Hence, by definition,

$$\operatorname{codim} Y = \dim\ X/Y. \tag{1.11}$$

Suppose that X is a linear space over the field $\mathbb{R}$. An element $x \in X$ is said to be a *convex combination* of elements $x_1, ..., x_n \in X$ if

$$x = \sum_{j=1}^{n} a_j x_j; \quad a_j \in \mathbb{R},\ a_j \geq 0, \quad \sum_{j=1}^{n} a_j = 1.$$

The set $[x, y]$, $x, y \in X$, of all convex combinations of elements x and y is called a *closed interval.* Thus, by definition,

$$[x, y] = \{tx + (1 - t)y : t \in [0, 1]\}.$$

A subset Y of a linear space X over $\mathbb{R}$ is said to be a *convex set* if every convex combination of elements x and y belonging to Y again belongs to Y. One can prove that Y is a convex set if and only if $[x, y] \subset Y$ whenever $x, y \in Y$.

If Y is a subset of a linear space X over the field $\mathbb{R}$ then the least convex set containing the set Y is said to be the *convex hull* of Y and is denoted by conv Y.

From the definition of the convex hull it follows that

$$\text{(1.12)} \qquad \operatorname{conv} Y = \Big\{x \in X : x = \sum_{j=1}^{n} a_j x_j;\ x_j \in Y;\ a_j \geq 0;\ \sum_{j=1}^{n} a_j = 1\Big\}.$$

Indeed,

$$Y \subset Y_1 = \Big\{x \in X : x = \sum_{j=1}^{n} a_j x_j;\ x_j \in Y;\ a_j \geq 0;\ \sum_{j=1}^{n} a_j = 1\Big\}$$

and Y_1 is a convex subset of the space X. If $Y_2 \supset Y$ is another convex subset of X then, by the definition of a convex subset, it should contain also elements of the form

$$x = \sum_{j=1}^{n} a_j x_j;\ x_j \in Y;\ a_j \geq 0;\ \sum_{j=1}^{n} a_j = 1.$$

Thus $Y_2 \supset Y_1$. We therefore conclude that $Y_1 = \operatorname{conv} Y$.

A subset Y of a linear space over a field $\mathbb{F}$ is said to be a *linear manifold* if $tx + sy \in Y$ whenever $x, y \in Y$, $t, s \in \mathbb{F}$ and $t + s = 1$.

Theorem 1.2. *If X is a linear space over a field $\mathbb{F}$, Y is a linear manifold in X and x_0 is an arbitrary element of Y, then the set*

$$Y_0 = Y - x_0 = \{y - x_0 : y \in Y\}$$

is a linear subspace in Y independently of the choice of the element $x_0 \in Y$.

Proof. If $y_1, y_2 \in Y_0$ then $y_1 = y_1' - x_0$, $y_2 = y_2' - x_0$, where $y_1', y_2' \in Y$. Let $t, s \in \mathbb{F}$ be arbitrary. Consider the element

$$y = ty_1 + sy_2 + x_0 = ty_1' + sy_2' + (1 - t - s)x_0.$$

Since $t + s + (1 - t - s) = 1$, we get $y \in Y$. Hence $ty_1 + sy_2 = y - x_0 \in Y_0$, which means that Y_0 is a linear subspace of X. If x_1 is an arbitrary element of Y then $x_1 - x_0 \in Y_0$. Since Y_0 is a linear subspace of X, we get

$$Y - x_0 = Y_0 = Y_0 - (x_1 - x_0) = Y - x_0 + x_0 - x_1 = Y - x_1.$$

■

We say that a *linear manifold* $Y \subset X$ *is of codimension* n if the quotient space X/Y_0 (where the set $Y_0 = Y - x_0$ is defined as in Theorem 1.2) is of dimension n. Linear manifolds of codimension 1 are said to be *hyperplanes*.

If a linear space X over a field $\mathbb{F}$ is a ring (with respect to the same addition) and $a(xy) = (ax)y = x(ay)$ for all $a \in \mathbb{F}$ and $x, y \in X$ then X is said to be an *algebra* (otherwise: a *linear ring*). A non-empty subset Y of an algebra X is said to be a *subalgebra* (*linear subring*) if Y is an algebra with respect to the same addition, multiplication of elements and multiplication by scalars. By this definition, a subset Y of an algebra X over $\mathbb{F}$ is a subalgebra of the algebra X if

$$x - y \in Y; \quad xy \in Y; \quad tx \in Y \quad \textit{whenever } x, y \in Y,\ t \in \mathbb{F}.$$

If in an algebra X there exists an element e such that $ex = xe = x$ whenever $x \in X$, then e is said to be the *unit* of X. The unit (provided that it exists) is unique. Indeed, suppose that there is another element $e' \in X$ such that $e'x = xe' = x$ whenever $x \in X$. In particular, this implies $e' = ee' = e$. One can prove that any algebra X can be extended to an algebra with unit (cf. Example 1.15).

Let Y be a *left (right) ideal* in an algebra X over the field $\mathbb{F}$. It means that Y is a subalgebra of X such that $x - y \in Y$ for all $x, y \in Y$ and $xy \in Y$ ($yx \in Y$, respectively) whenever $x \in X$, $y \in Y$. If Y is simultaneously a left and right ideal, we say that Y is a *two-sided* ideal or briefly: an *ideal*, (if it does not lead to any misunderstanding). A left (right, two-sided) ideal Y is said to be *proper* if $Y \neq \{0\}$ and $Y \neq X$. If it is not the case, i.e. either $Y = \{0\}$ or $Y = X$, then we say that Y is a *trivial* ideal.

An algebra X is said to be *commutative* if each pair of its elements commute one with another, i.e. if $xy = yx$ whenever $x, y \in X$. Clearly, in a commutative algebra left and right ideals do coincide, i.e. there are only two-sided ideals.

A proper left (right, two-sided) ideal Y is said to be *maximal* if every proper left (right, two-sided) ideal $Y_1 \supset Y$ is equal to Y: $Y_1 = Y$. Every proper left (right, two-sided ideal) is contained in a maximal left (right, two-sided) ideal (cf. Jacobson [J]).

Let Y be a proper ideal in an algebra over a field $\mathbb{F}$. Consider the set

$$X/Y = \{[x] : x \in X\},$$

where, as before, we denote by $[x]$ the coset corresponding to the element x, i.e. $[x] = x + Y$. The addition and multiplication in the set X/Y are defined by Formulae (1.10). The multiplication of elements is defined as follows:

$$[x][y] = [xy] \quad for\ x, y \in X. \tag{1.13}$$

This operation is well defined, since always $[x][y] \subset [xy]$ for arbitrary $x, y \in X$. Indeed, since by our assumption, Y is a two-sided ideal, we have $xY + Yy + Y \cdot Y \subset Y$, where $Y_1 \cdot Y_2 = \{y_1 y_2 : y_1 \in Y_1,\ y_2 \in Y_2\}$ is the *algebraic product* of Y_1 and Y_2. Hence

$$[x][y] = (x + Y)(y + Y) = xy + xY + Yy + Y \cdot Y \subset xy + Y = [xy].$$

It is easy to verify that the set X/Y with the operations on cosets defined by Formulae (1.10), (1.13) is an algebra over the field $\mathbb{F}$ called the *quotient algebra*.

Examples and Exercises.

Example 1.1. The set $\mathbb{R}$ considered as a set of vectors on the real line with the usual addition and multiplication of vectors by reals is a linear space over the field $\mathbb{R}$. The basis in this space consists of the unit vector 1. Clearly, dim $\mathbb{R} = 1$. □

Example 1.2. The Cartesian product $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ is also a linear space over the field $\mathbb{R}$, since it is the set of vectors on the plane with the usual addition and multiplication by reals. The basis in that space consists of unit vectors $(1, 0)$ and $(0, 1)$. Clearly, dim $\mathbb{R}^2 = 2$. Observe that we can write also $\mathbb{R}^2 = \mathbb{R} \oplus \mathbb{R}$. □

Example 1.3. Recall that $\mathbb{R}^{n+1} = \mathbb{R} \times \mathbb{R}^n$ for $n \in \mathbb{N}$. The set $\mathbb{R}^n$ is a linear space over the field $\mathbb{R}$ with respect to the usual addition of vectors and multiplication of vectors by reals, i.e.

$$x + y = (x_1 + y_1, ..., x_n + y_n), \quad tx = (tx_1, ..., tx_n)$$

$$whenever \quad x = (x_1, ..., x_n),\ y = (y_1, ..., y_n) \in \mathbb{R}^n;\ t \in \mathbb{R}.$$

A basis in that space consists of *unit vectors*, i.e. vectors of the form

$$(1, 0, ..., 0)$$

$$(0, 1, ..., 0)$$

$$...........$$

$$(0, ..., 0, 1)$$

Clearly, dim $\mathbb{R}^n = n$. Observe that we can write also

$$\mathbb{R}^n = \underbrace{\mathbb{R} \oplus ... \oplus \mathbb{R}}_{n-fold}.$$

Note that in Examples 1.1, 1.2, 1.3 one can consider without any essential change the set $\mathbb{R}$ as the set of points on the real line instead of a set of vectors. $\square$

Example 1.4. The set $\mathbb{C}$ considered over the field $\mathbb{F}$ with the usual operations of addition and multiplication by complex numbers, i.e. the plane $\mathbb{C}$ of complex numbers is a linear space over the field $\mathbb{C}$ with the basis $\{1\}$. Clearly, dim $\mathbb{C} = 1$.

The same space over the field $\mathbb{R}$, written in the form $\mathbb{C} = \mathbb{R} + i\mathbb{R}$, has the basis $\{1, i\} = \{(1,0),(0,1)\}$ and dim $(\mathbb{R} + i\mathbb{R}) = 2$. $\square$

Example 1.5. The set $\mathbb{F}_n[t]$ of all polynomials in t of degree n with coefficients from the field $\mathbb{F}$:

$$p_n(t) = \sum_{k=0}^{n} a_k t^k, \quad a_0, ..., a_n \in \mathbb{F},$$

is a linear space over the field $\mathbb{F}$ if the addition of polynomials and multiplication of polynomials by scalars are defined as follows:

$$p_n(t) + q_n(t) = \sum_{k=0}^{n} (a_k + b_k) t^k, \quad \textit{where} \quad q_n(t) = \sum_{k=0}^{n} b_k t^k, \quad b_0, ..., b_n \in \mathbb{F}, \tag{1.14}$$

$$\alpha p_n(t) = \sum_{k=0}^{n} (\alpha a_k) t^k, \quad \textit{where } \alpha \in \mathbb{F}. \tag{1.15}$$

The set $\{1, t, ..., t^n\}$ is a basis in the space $\mathbb{F}_n[t]$. Clearly, $\dim \mathbb{F}_n[t] = n + 1$. $\square$

Example 1.6. The set $\mathbb{F}[t]$ of **all** polynomials in t with coefficients belonging to the field $\mathbb{F}$ is a linear space over $\mathbb{F}$ if we define the multiplication of polynomials by scalars by means of Formula (1.15) and the addition of polynomials of different degrees

$$p_n(t) = \sum_{k=0}^{n} a_k t^k \quad \textit{and} \quad q_m(t) = \sum_{j=0}^{m} b_k t^t, \quad \textit{where } a_k, b_j \in \mathbb{F},\ n \leq m,$$

in the following manner:

$$p_n + q_m = r_N, \quad \textit{where } N = \max(n, m) = m, \quad r_N(t) = \sum_{k=0}^{N} c_k t^k,$$

$$c_k = \begin{cases} a_k + b_k & \text{for } k = 0, 1, ..., \min(n, m); \\ b_k & \text{for } k = n + 1, ..., m. \end{cases}$$

The set $\{1, t, t^2, ...\}$ is a basis in the space $\mathbb{F}[t]$. Clearly, dim $\mathbb{F}[t] = +\infty$. $\square$

Example 1.7. The set of all functions defined on a set Ω and with values in a field $\mathbb{F}$ is a linear space over the field $\mathbb{F}$ if the operations of addition and multiplication by scalars of functions we define as follows:

$$(x + y)(t) = x(t) + y(t), \quad (\alpha x)(t) = a x(t) \quad \textit{for } x, y \in X,\ \alpha \in \mathbb{F}. \tag{1.16}$$

Clearly, dim $X = +\infty$ if the set Ω is infinite. If this is not the case, the dimension of X is equal to the number of elements of the set Ω. □

Example 1.8. The set of all real-valued functions defined and bounded on a set Ω with the addition and multiplication by reals defined by Formulae (1.16) (where we put $\mathbb{F} = \mathbb{R}$) is a linear space over the field $\mathbb{R}$. Indeed, a sum of two bounded functions and the product of a bounded function by a real number are again bounded functions. Clearly, dim $X = +\infty$ if the set Ω is infinite. □

Example 1.9. The set $C[a,b]$ of all real-valued functions defined and continuous on a closed interval $[a,b]$ with the addition and multiplication by scalars determined by Formulae (1.16) (where we admit $\mathbb{F} = \mathbb{R}$, $\Omega = [a,b]$) is a linear space over the field $\mathbb{R}$. Indeed, a sum of two continuous functions and the product of a continuous functions by a real number are again continuous functions.

□

Example 1.10. The set X of all real-valued functions defined and continuous on a closed interval $[a,b]$ and differentiable at each point $t \in [a,b]$, with the addition and multiplication by reals defined by Formulae (1.16), is a linear space over the field $\mathbb{R}$. Indeed, if $x, y \in X$ and there exist derivatives $x'(t)$, $y'(t)$ at each point $t \in [a,b]$ then

$$(x+y)' = x' + y', \quad (\alpha x)' = \alpha x' \quad \textit{for all } \alpha \in \mathbb{R}. \tag{1.17}$$

□

Example 1.11. The set $C^1[a,b]$ of all real-valued functions, defined on a closed interval and having a continuous derivative on this interval is a linear space over the field $\mathbb{R}$. Indeed, by Formulae (1.16), (1.17), since a sum of two continuous functions and a product of a continuous function by a real number are again continuous, we conclude that $x' + y' = (x+y)'$ and $(\alpha x)' = \alpha x'$ are again continuous functions. This implies that a sum of two continuously differentiable functions on the interval $[a,b]$ and a product of a continuously differentiable function on $[a,b]$ by a real number are again continuously differentiable functions on $[a,b]$. Clearly, $\dim C^1[a,b] = +\infty$. □

Exercise 1.1. Let $\mathbb{N}$ be the set of all positive integers. Let $n \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$ (i.e. $\mathbb{N}_0$ is the set of all nonnegative integers). Write $C^0[a,b] = C[a,b]$. Prove that

(i) the set $C^n[a,b]$ of all real-valued functions defined on a closed interval $[a,b]$ and having in this interval the continuous nth derivative is a linear space over the field $\mathbb{R}$;

(ii) the space $C^{n+1}[a,b]$ is a linear subspace of the space $C^n[a,b]$;

(iii) the set $C^\infty[a,b]$ of all real-valued functions defined on the closed interval $[a,b]$ and having in this interval continuous derivatives of an arbitrary order (i.e. the space of infinitely differentiable functions) is a linear space over $\mathbb{R}$;

(iv) the space $C^\infty[a,b]$ is a linear subspace of every space $C^n[a,b]$ $(n \in \mathbb{N}_0)$;

(v) $C^\infty[a,b] = \bigcap\limits_{n=0}^{\infty} C^n[a,b]$. □

Exercise 1.2. Suppose that sets Ω_1 and Ω_2 are disjoint. Write $\Omega = \Omega_1 \cup \Omega_2$. For an arbitrary set E denote by X_E the space of all functions defined on E and with values in a field $\mathbb{F}$. Prove that $X_\Omega = X_{\Omega_1} \times X_{\Omega_2}$. $\square$

Example 1.12. Suppose that $X = (s)_\mathbb{F}$ is the set of all sequences $a = \{a_n\}$, where $a_n \in \mathbb{F}$ for $n \in \mathbb{N}$. Traditionally, we write $(s)_\mathbb{R} = (s)$. Define the *coordinatewise* addition and multiplication by scalars of sequences:

$$(1.18) \qquad \{a_n\} + \{b_n\} = \{a_n + b_n\}; \quad \alpha\{a_n\} = \{\alpha a_n\} \quad \textit{for } \{a_n\}, \{b_n\} \in X;\ \alpha \in \mathbb{F}.$$

It is easy to verify that X is a linear space over the field $\mathbb{F}$. $\square$

Example 1.13. Let $\mathbb{Z}$ be the set of all integers. Suppose that X is the set of all two-sided sequences $a = \{a_n\}$, where $a_n \in \mathbb{F}$ for $n \in \mathbb{Z}$. Define the *coordinatewise* addition and multiplication by scalars of sequences by Formulae (1.18) (taking into account that here $n \in \mathbb{Z}$). It is easy to verify that X is a linear space over the field $\mathbb{F}$. $\square$

Example 1.14. The linear space $C[a, b]$ defined in Example 1.9 is a commutative algebra over the field $\mathbb{R}$ if we define the *pointwise* multiplication of two functions:

$$(1.19) \qquad (xy)(t) = x(t)y(t) \quad \textit{for all } x, y \in C[a, b],\ t \in [a, b].$$

For an arbitrarily fixed $t_0 \in [a, b]$ the set

$$X_0 = \{x \in C[a, b] : x(t_0) = 0\}$$

is a proper ideal in the algebra $C[a, b]$. Indeed, if $x, y \in X_0$ then $x(t_0) - y(t_0) = 0$. Hence $x - y \in X_0$. If $x \in C[a, b]$, $y \in X_0$ then $x(t_0)y(t_0) = y(t_0)x(t_0) = 0$. Hence $xy, yx \in X_0$. Clearly, X_0 is a proper ideal. Observe that the function $e(t) \equiv 1$, which is the unit of the algebra $C[a, b]$, does not belong to X_0, for $e(t_0) = 1 \neq 0$.

The quotient algebra $C[a, b]/X_0$ can be identify with the set of all functions *constant* on the interval $[a, b]$. This follows from the fact that $y \in [x]$ for an $x \in C[a, b]$ if and only if $x - y \in X_0$, i.e. if $x(t_0) - y(t_0) = 0$. This implies that codim $X_0 =$ dim $C[a, b]/X_0 = 1$. $\square$

Exercise 1.3. Prove that

(i) the set $X_1 = \{x \in C[a, b] : x(t_0) = 0,\ x(t_1) = 0\}$ is a proper ideal in the algebra $C[a, b]$ for arbitrarily fixed $t_0, t_1 \in [a, b]$ (cf. Example 1.14);

(ii) if $t_1 \neq t_0$ then codim $X_1 = 2$. $\square$

Example 1.15. Let X_1 be defined as in Exercise 1.3. Then the quotient algebra $C[a, b]/X_1$ can be identified with the set of all linear functions, i.e. the set

$$\{x : x(t) = \alpha t + \beta,\ \textit{where } \alpha, \beta \in \mathbb{R},\ t \in [a, b]\}.$$

$\square$

Exercise 1.4. Prove that the set $Y = \{x \in C[a, b] : x(t) = 0 \textit{ for } a \leq t_0 \leq t \leq t_1 \leq b\}$ is a proper ideal in the algebra $C[a, b]$ (over $\mathbb{R}$) if $t_1 \neq t_0$. $\square$

Exercise 1.5. Prove that

(i) the space $X = (s)_{\mathbb{F}}$ defined in Example 1.12 with the *coordinatewise multiplication* of sequences

$$\{a_n\}\{b_n\} = \{a_n b_n\} \quad for\ \{a_n\}, \{b_n\} \in X \tag{1.20}$$

is a commutative algebra over the field $\mathbb{F}$;

(ii) the set $X_0 = \{a = \{a_n\} \in X : a_m = 0\ for\ an\ m \in \mathbb{N}\}$ is a proper ideal in X. □

Exercise 1.6. Prove that the space $X = (s)_{\mathbb{F}}$ defined in Example 1.13. with the *convolution* of sequences, i.e. the operation define in the following manner:

$$\{a_n\} \star \{b_n\} = \{\sum_{k=1}^{n} a_k b_{n-k}\} \quad for\ \{a_n\}, \{b_n\} \in X, \tag{1.21}$$

as a multiplication of elements is a commutative algebra over the field $\mathbb{F}$. Has this ring *zero divisors*, i.e. such elements $x, y \in X$ that $x \neq 0$, $y \neq 0$, although $xy = 0$? □

Exercise 1.7. Prove that

(i) the space X defined in Example 1.13 with the *coordinatewise multiplication* of sequences (1.20) (with $n \in \mathbb{Z}$) is a commutative algebra over the field $\mathbb{F}$;

(ii) the set $X_0 = \{a = \{a_n\} \in X : a_m = 0\ for\ an\ m \in \mathbb{Z}\}$ is a proper ideal in X. □

Exercise 1.8. The multiplication in a ring is said to be *trivial* if $xy = 0$ for all $x, y \in X$. Prove that every linear space is an algebra with the trivial multiplication. □

Exercise 1.9. Are linear spaces appearing in Examples 1.6, 1.10 and Exercise 1.1, 1.2 algebras with the multiplication defined by Formula (1.19) ? Are some of these algebras commutative ? Have some of these algebras zero divisors ? □

Example 1.16. Suppose that X is a ring without unit. We shall prove that X can be extended to a ring X_1 with unit. Define the product of an arbitrary element $x \in X$ by an integer a as follows:

$$ax = \begin{cases} 0 & \text{if } a = 0; \\ nx = \underbrace{x + ... + x}_{n-fold}; & \text{if } a = n \in \mathbb{N}; \\ (-n)x = n(-x); & \text{if } a = -n;\ n \in \mathbb{N}, \end{cases}$$

where we denote by $-x$ the element $0 - x$. Consider the set $X_1 = \{(x, a) : x \in X;\ a \in \mathbb{Z}\}$. The operations of addition, multiplication of elements and multiplication by scalars can be defined in X_1 in the following manner:

$$b(x, a) = (bx, ba), \quad (x, a) + (y, b) = (x + y, a + b), \quad (x, a)(y, b) = (xy + bx + ay, ab)$$

$$for\ x, y \in X;\ a, b \in \mathbb{Z}.$$

It is easy to check that X_1 with these operations is a ring and that the pair $(0,1)$ is its unit. Moreover, $X = \{(x,0) : x \in X\} \subset X_1$. In order to prove that any algebra X over a field $\mathbb{F}$ without unit can be extended to an algebra with unit it is enough to consider the corresponding extension of X with scalars belonging to $\mathbb{F}$ instead of integers. $\square$

Exercise 1.10. Prove that

(i) a proper ideal $Y \subset X$ does not contain the unit of the algebra X;

(ii) a proper left (right) ideal does not contain any *left (right) invertible element*, i.e an element x such that there is $y \in X$ with the property $yx = e$ ($xy = e$, $yx = xy = e$, respectively);

(iii) a proper ideal does not contain any invertible element.

On the other hand, if $x \in X$ is not a left (right) invertible element then there is a proper left (right) ideal containing x, namely, xX (Xx, respectively). $\square$

Example 1.17. The convex hull of two points $x, y \in \mathbb{R}$, $y \neq x$ is the interval $[x,y]$. The convex hull of 3 linearly independent points in $\mathbb{R}^2$ is the triangle whose vertices are these points. The convex hull of four linearly independent points in $\mathbb{R}^3$ is the tetrahedron whose vertices are these points. $\square$

Example 1.18. The convex hull of $n+1$ linearly independent elements in $\mathbb{R}^n$ is said to be an *n-dimensional simplex* with vertices in these points ($n = 1, 2, 3, ...$). $\square$

Exercise 1.11. Determine the convex hull of n points ($n = 2, 3, ...$) in the spaces $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$. $\square$

Exercise 1.12. Does the convex hull of all functions determined on a set Ω and admitting only the values either $+1$ or -1 is the set of all functions determined on the set Ω and with the modulus less or equal 1 ? If this is not the case, determine that convex hull. $\square$

Exercise 1.13. Determine linear manifolds in $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$. $\square$

Exercise 1.14. Prove that

(i) hyperplanes in $\mathbb{R}$ are one-point sets;

(ii) hyperplanes in $\mathbb{R}^2$ are straight lines;

(iii) hyperplanes in $\mathbb{R}^3$ are planes. $\square$

Chapter 2.

Linear operators and linear functionals.

Let two linear spaces X and Y, both over the same field $\mathbb{F}$ of scalars, be given. A mapping A of a linear subset $\mathcal{D}_A$ of the space X into the space Y is said to be a *linear operator* if the following conditions are satisfied:

$$A(x+y) = Ax + Ay; \quad A(tx) = tAx \quad for\ all\ x, y \in \mathcal{D}_A,\ t \in \mathbb{F}. \tag{2.1}$$

The set $\mathcal{D}_A$ is called the *domain* of the operator A (sometimes denoted by dom A). More exactly, a linear operator is a pair $(\mathcal{D}_A, A)$, since A is determined by its domain and the form of the mapping. However, we shall use the shorter traditional notation A, because it does not lead to any misunderstanding.

Let $G \subset \mathcal{D}_A$. Write

$$AG = \{y \in Y : y = Ax\ for\ x \in G\}. \tag{2.2}$$

The set $\mathcal{E}_A = A\mathcal{D}_A$ of the *values* of the operator A is called either its *range* or the *image* of A.

The *graph* of the operator A is a subset of the Cartesian product $X \times Y$ defined as follows:

$$\text{graph } A = \{(x, y) : x \in \mathcal{D}_A,\ y = Ax\}.$$

By $L(X \to Y)$ we denote the set of all linear operators whose domains are contained in X and values belong to the space Y.

The *identity operator* (briefly: the *identity*) I_X in the space is an operator defined by means of the equality

$$I_X x = x \quad for\ all\ x \in X. \tag{2.3}$$

If it will not lead to any misunderstanding, we shall denote briefly the identity operator by "I" instead of "I_X".

If the operator $A \in L(X \to Y)$ is a one-to-one mapping then we can define the inverse operator A^{-1} in the following way: for all $y \in \mathcal{E}_A$

$$A^{-1}y = x; \quad x \in \mathcal{D}_A,\ y = Ax. \tag{2.4}$$

Observe that to every element $y \in \mathcal{E}_A$ there corresponds a unique element $x \in \mathcal{D}_A$. By definition,

$$\mathcal{D}_{A^{-1}} = \mathcal{E}_A \subset Y; \quad \mathcal{E}_{A^{-1}} = \mathcal{D}_A \subset X. \tag{2.5}$$

For arbitrary $x \in \mathcal{D}_A$, $y = Ax$ we have

$$(A^{-1}A)x = A^{-1}(Ax) = A^{-1}y = x, \quad (AA^{-1})y = A(A^{-1}y) = Ax = y,$$

which implies

$$A^{-1}A = I_{\mathcal{D}_A}, \quad AA^{-1} = I_{\mathcal{E}_A}. \tag{2.6}$$

This means that A^{-1} is, indeed, an inverse mapping for A.

Observe that this inverse mapping is uniquely determined. Indeed, suppose that there is another mapping B of the set $\mathcal{E}_A$ onto the set $\mathcal{D}_A$ satisfying Condition (2.6), i.e. such that

$$BA = I_{\mathcal{D}_A}, \quad AB = I_{\mathcal{E}_A}. \tag{2.7}$$

Then Formulae (2.6) and (2.7) together imply that

$$B = BI_{\mathcal{E}_A} = B(AA^{-1}) = (BA)A^{-1} = I_{\mathcal{D}_A}A^{-1} = A^{-1}.$$

We therefore conclude that the mapping A^{-1} is uniquely determined.

We shall show that the inverse mapping A^{-1} is a linear operator. Indeed, for arbitrary $y_1, y_2 \in \mathcal{D}_{A^{-1}} = \mathcal{E}_A$ there exist unique elements $x_1, x_2 \in \mathcal{E}_A^{-1} = \mathcal{D}_A$ such that $y_1 = Ax_1$, $y_2 = Ax_2$. By definition, $x_1 = A^{-1}y_1$, $x_2 = A^{-1}y_2$. Since A is a linear operator, for arbitrary scalars λ,μ we find

$$A^{-1}(\lambda y_1 + \mu y_2) = A^{-1}(\lambda Ax_1 + \mu Ax_2) = A^{-1}A(\lambda x_1 + \mu x_2)+$$

$$= \lambda x_1 + \mu x_2 = \lambda A^{-1}y_1 + \mu A^{-1}y_2.$$

Hence $A^{-1} \in L(Y \to X)$.

If an operator $A \in L(X \to Y)$ has an inverse operator then we say that A is *invertible*.

A linear operator $A \in L(X \to Y)$ is said to be an *isomorphism* if $\mathcal{D}_A = X$, $\mathcal{E}_A = Y$ and A is a one-to-one mapping. If A is an isomorphism then A is invertible, the inverse operator A^{-1} is one-to-one and $\mathcal{D}_{A^{-1}} = Y$, $\mathcal{E}_{A^{-1}} = X$. Hence A^{-1} is also an isomorphism.

Linear spaces X and Y are *isomorphic* if there is an isomorphism mapping X onto Y.

Define the *sum* of two linear operators $A, B \in L(X \to Y)$ and the *product of an operator by a scalar* by following formulae:

$$(A + B)x = Ax + Bx \quad for \quad x \in \mathcal{D}_A \cap \mathcal{D}_B, \tag{2.8}$$

$$(tA)x = A(tx) \quad for \quad x \in \mathcal{D}_A, \ t \in \mathbb{F}. \tag{2.9}$$

It is easy to verify that the *addition* of operators just defined is *associative* and *commutative*, i.e.

$$(A + B) + C = A + (B + C), \quad B + A = A + B$$

for all $A, B, C \in L(X \to Y)$ such that the corresponding sums are well defined.

Clearly, an operator C such that $A + C = B$ does not exist always for $A, B \in L(X \to Y)$. It is so, for instance, when $\mathcal{D}_A \cap \mathcal{D}_B = \{0\}$. However, if such an operator C exists then we write $C = A - B$ and C is called the *difference* of A and B. The operation "-" is called the *subtraction* of operators.

If the operator $A - B$ is well defined then, by definition, $A - B = A + (-B)$ on $\mathcal{D}_A \cap \mathcal{D}_B$.

Write

$$L_0(X \to Y) = \{A \in L(X \to Y) : \mathcal{D}_A = X\}.$$

Since the addition of any two operators $A, B \in L_0(X \to Y)$ is well defined, associative, commutative and there exists the operator $C = A - B$, we conclude that $L_0(X \to Y)$ is an *Abelian group*. The *neutral element* of that group is an operator A such that $Ax = 0$ for every $x \in X$. In the sequel we shall denote this "zero operator" by 0, since it does not lead to any misunderstanding. Formula (2.9) implies that the Abelian group $L_0(X \to Y)$ is a *linear space over the field* $\mathbb{F}$.

Suppose that X, Y, Z are linear spaces over the field $\mathbb{F}$ of scalars, $A \in L(Y \to Z)$, $B \in L(X \to Y)$ and $\mathcal{E}_B = B\mathcal{D}_B \subset \mathcal{D}_A \subset Y$. A *superposition* of the operators B and A is defined as an operator AB satisfying the equality

$$(AB)x = A(Bx) \quad \textit{for all } x \in \mathcal{D}_A.$$

Clearly, by definition, $AB \in L(X \to Z)$. Moreover,

$$\mathcal{D}_{AB} = \mathcal{D}_A, \quad \mathcal{E}_{AB} = A\mathcal{E}_B.$$

The superposition of operators (provided that it exists) is *distributive* with respect to the addition of operators. Indeed, if $A,\ A_1,\ A_2 \in L(Y \to Z)$, $B, B_1, B_2 \in L(X \to Y)$ and the operators $AB_1,\ AB_2,\ B_1 + B_2,\ A(B_1 + B_2),\ A_1B,\ A_2B,\ A_1 + A_2,\ (A_1 + A_2)B$ are well defined then

$$A(B_1 + B_2) = AB_1 + AB_2, \quad (A_1 + A_2)B = A_1B + A_2B. \tag{2.10}$$

Two operators $A, B \in L(X \to Y)$ are said to be *commutative* if there exist superpositions AB, BA and $AB = BA$. In general, the last equality is not satisfied. However, it could be useful in several cases to consider the operator $AB - BA$ whenever it is well defined. This operator is called the *commutator* of A and B. Clearly, if A and B commute each with another then their commutator is equal to zero.

Write

$$L(X) = L(X \to X) \quad \textit{and} \quad L_0(X) = L_0(X \to X).$$

By Formulae (2.10), we conclude that $L(X)$ is not only a linear space, it is also an algebra with respect to the multiplication of operators $A, B \in L_0(X)$ defined as their superposition.

Indeed, if $A, B \in L_0(X)$ then $\mathcal{E}_B \subset \mathcal{D}_A = X$. Hence the superposition AB is well defined for all $A, B \in L_0(X)$. The algebra $L_0(X)$ has a *unit*, namely, the identity $I_X = I$).

A linear operator $P \in L_0(X)$ is said to be a *projector* (otherwise: a *projection operator*) if $P^2 = P$, where $P^2 = P \cdot P$. If P is a projector then $I - P$ is also a projector, since we have for $(I-P)^2 = I - 2P + P^2 = I - 2P + P = I - P$.

Theorem 2.1. *Every projector $P \in L_0(X)$ determines the decomposition of the space X onto a direct sum $X = Y \oplus Z$, where*

$$Y = \{x \in X : Px = x\}, \quad Z = \{x \in X : Px = 0\}.$$

Conversely, if $X = Y \oplus Z$ then there is a projector $P \in L_0(X)$ such that $PX = Y$, $(I-P)X = Z$.

Proof. Suppose that $P \in L_0(X)$ is a projector and x is an arbitrary element of the space X. Write $z = x - Px = (I-P)x$. Since $Pz = Px - P^2x = Px - Px = 0$, we conclude that $z \in Z$ and $Z = (I-P)X$. Thus $x = y + z$, where $y = Px$. Moreover, $Py = P(Px) = P^2x = Px = y$. Hence $y \in Y$ and $Y = PX$. The arbitrariness of $x \in X$ implies that X is an algebraic sum of Y and Z. We shall show that this sum is direct, i.e. $Y \cap Z = \{0\}$. Indeed, if $u \in Y \cap Z$ then $u \in Y$, hence $u = Pu$. On the other hand, since $u \in Z$, we have $Pu = 0$. This implies $u = 0$.

Conversely, suppose that $X = Y \oplus Z$. Then every element $x \in X$ can be written as a sum $x = y + z$, where $y \in Y$, $z \in Z$. Define the operator P by means of the equality

$$Px = y \quad for \quad x = y + z, \quad y \in Y,\ z \in Z.$$

The operator $P \in L_0(X)$ is a projector, for $P^2x = Py = y$. Hence $Y = PX$ and

$$Z = \{z \in X : z = x - y,\ x \in X,\ y \in Y\} = \{z \in X : z = x - Px,\ x \in X\} = (I-P)X.$$

■

Having already defined such a correspondence between projectors and decompositions of the space X onto direct sums, we can say that Y is a *projection* of the space X "in the direction Z" and Z is a projection of X "in the direction Y" or that P *projects* X onto Y "in the direction Z".

Let X_0 be a subspace of a linear space X. Every linear operator $A \in L_0(X \to Y)$ induces an operator $[A] \in L_0(X/X_0 \to Y/AX_0)$ defined by the following formula:

$$[A][x] = [Ax] \quad for \quad x \in [x],$$

where $[x]$ is a coset determined by an element $x \in X$, i.e. $[x] = x + X_0$, and $[Ax]$ is a coset in the quotient space Y/AX_0, i.e. $[Ax] = Ax + AX_0$. It is easy to verify that $[A]$ is a linear operator.

If X_0 is a subspace of a linear space X and $A \in L(X \to Y)$, then an operator $A_0 \in L_0(X_0 \to Y)$ defined by means of the equality

$$A_0 x = Ax \quad for \quad x \in \mathcal{D}_A \cap X_0$$

is said to be the *restriction of the operator* A *to the subspace* X_0. The operator A_0 is often denoted by the symbol

$$A_0 = A|_{X_0}.$$

An operator $A_1 \in L_0(X \to Y)$ is called an *extension* of the operator $A \in L_0(X_0 \to Y)$, $X_0 \subset X$, to the space X if

$$A_1 x = Ax \quad for \quad x \in X, \quad i.e. \quad A_1|_{X_0} = A.$$

Let $A \in L(X \to Y)$. Denote by

$$\text{ker } A = \{x \in \mathcal{D}_A : Ax = 0\}$$

the set of zeros of the operator A. This set is otherwise called the *kernel* of A or the *space of its zeros.*

The set ker A is a subspace of the space X. Indeed, if $x, y \in$ ker A then $Ax = 0$, $Ay = 0$. Hence $A(x + Y) = Ax + Ay = 0$, which implies $x + y \in$ ker A. Moreover, if $t \in \mathbb{F}$ then $A(tx) = tAx = 0$.

The dimension of the kernel of an operator $A \in L(X \to Y)$ is called its *nullity* and denoted by α_A. Hence, by definition,

$$\alpha_A = \text{dim ker } A. \tag{2.11}$$

The *cokernel* of an operator $A \in L(X \to Y)$ is the quotient space $Y/A\mathcal{D}_A$. The number dim $Y/A\mathcal{D}_A$ is said to be the *defect* of the range of the operator A.

The *deficiency* β_A of an operator $A \in L(X \to Y)$ is defined by means of the equality

$$\beta_A = \text{dim } Y/A\mathcal{D}_A = \text{codim } A\mathcal{D}_A. \tag{2.12}$$

In other words, *the deficiency of a linear operator is equal to the defect of its range* .

Theorem 2.2. *If in a linear space* X *the maximal number of linearly independent elements is* n *then* dim $X = n$.

Proof. Let n be the maximal number of linearly independent elements in the space X. Then there exist elements $x_1, ..., x_n \in X$ which are linearly independent. Suppose that dim $X = m < n$. Hence in X there is a basis $\{e_1, ..., e_m\}$. Define an operator A in the following way:

$$A\Big(\sum_{j=1}^{m} a_j e_j\Big) = \sum_{j=1}^{m} a_j x_j; \quad a_1, ..., a_m \in \mathbb{F}.$$

Since elements $x_1, ..., x_m$ are linearly independent, the operator A is one-to-one, i.e. it transforms arbitrary linearly independent elements into linearly independent elements. Since $X = \text{lin}\ \{e_1, ..., e_m\}$, we conclude that there are n linearly independent elements $y_1, ..., y_n$ belonging to the set lin $\{x_1, ..., x_m\}$. This implies that elements

$$y_1, ..., y_n, x_{m+1}, ..., x_n \in X$$

are linearly independent and their number is $2n - m > n$. This contradicts to our assumption that n is the maximal number of linearly independent elements in X. Then $\dim X = n$. ■

An immediate consequence of this theorem is

Corollary 2.1. *If X is an n-dimensional linear space then every system of n linearly independent elements $x_1, ..., x_n \in X$ is a basis in X.*

Theorem 2.3. *If Y is a subspace of a linear space X such that* codim $Y < +\infty$, *then there exists a subspace Z such that $X = Y \oplus Z$ and* $\dim Z = \text{codim}\ Y$.

Proof. Suppose that codim $Y = n$. Write $[X] = X/Y$. By our assumption, there exist n and only n linearly independent cosets $[x_1], ..., [x_n] \in [X]$ and each coset $[x] \in [X]$ can be written in a unique manner in the form

$$[x] = \sum_{j=1}^{n} t_j [x_j], \quad \textit{where } t_1, ..., t_n \textit{ are scalars.}$$

Let $y_1, ..., y_n$ be arbitrarily fixed elements such that $y_j \in [x_j]$. By definition, elements $y_1, ..., y_n$ are linearly independent. Hence every element $x \in X$ can be represented in a unique way in the form:

$$x = y + \sum_{j=1}^{n} t_j y_j, \quad \textit{where} \quad y \in Y.$$

Write $Z = \text{lin}\ \{y_1, ..., y_n\}$. Then we conclude that $X = Y \oplus Z$ and $\dim Z = \text{codim}\ Y$. ■

If the *axiom of choice* is admitted then Theorem 2.3 is satisfied without the assumption codim $Y < +\infty$ (cf. PRR[1]). Namely, we have the following

Theorem 2.4. *If Y is an arbitrary subspace of a linear space X, then there is a subspace $Z \subset X$ such that $X = Y \oplus Z$ (provided that the axiom of choice is admitted).*

The proof can be found in the book PRR[1].

Theorems 2.3 and 2.4 play an important role, because they decide on the decomposition of a given linear space onto a direct sum. For instance, by these theorems, we conclude that the algebra $L_0(X)$ has *zero divisors* *) whenever $\dim X > 1$. Indeed, suppose that X_0 is an arbitrary subspace of a linear space X such that $X_0 \neq X$. Let $Y \subset X$ be a subspace

*) Recall that $x \neq 0$ is a zero divisor in a ring X if there is a $y \in X \setminus \{0\}$ such that $xy = 0$.

such that $X = X_0 \oplus Y$. By Theorem 2.1, there is a projection operator $B \in L_0(X)$ such that $B \neq 0$ and $BX = X_0$. Write

$$Ax = \begin{cases} 0 & \text{for } x \in X_0, \\ x & \text{for } x \in Y. \end{cases}$$

Since $A = I - B$, we have $A \in L_0(X)$. Clearly, $A \neq 0$. However, for an arbitrary $x \in X$ we find $Bx = y \in X_0$. Then, by definition,

$$(AB)x = A(Bx) = Ay = 0.$$

We therefore conclude that $AB = 0$, i.e. the algebra $L_0(X)$ has zero divisors.

Another consequence of Theorem 2.4 (or Theorem 2.3 when codim $Y < +\infty$) is the following

Corollary 2.2. *Suppose that X_0 is a subspace of a linear space X. Then every operator $A \in L(X \to Y)$ with $\mathcal{D}_A = X_0$ has an extension to an operator A_1 such that $\mathcal{D}_{A_1} = X$, $\mathcal{E}_{A_1} = \mathcal{E}_A$.*

Proof. By our assumption, there is a subspace $Z \subset X$ such that $X = X_0 \oplus Z$. By Theorem 2.1, there is a projection operator $P \in L_0(X)$ such that $PX = X_0$. The operator $A_1 = AP$ is the operator, we are looking for. Indeed, the superposition AP is well defined and

$$\mathcal{D}_{A_1} = \mathcal{D}_{AP} = \mathcal{D}_P = X; \quad \mathcal{E}_{A_1} = A_1X = APX = AX_0 = \mathcal{E}_A.$$

■

Let X be a linear space over the field $\mathbb{F}$ and let either $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. A linear operator f is said to be a *linear functional* if $f \in L_0(X \to \mathbb{F})$.

Denote by X' the set of all linear functionals defined on the space X, i.e.

$$X' = L_0(X \to \mathbb{F}).$$

By definition, X' is a linear space over the field $\mathbb{F}$. The space X' is called the *conjugate space* to X.

The space $(X')'$ conjugate to the space X' is said to be the *second conjugate* and it is denoted by X''. Every element $x \in X$ induces a functional $x'' \in X''$ defined as follows:

$$x''(x') = x'(x) \quad \textit{for all } x' \in X'.$$

The correspondence $x'' = \kappa(x)$ between elements $x \in X$ and $x'' \in X''$ is called the *canonical mapping*. Whenever it does not lead to a misunderstanding, we identify an element x with its canonical image $\kappa(x)$.

If X is an n-dimensional linear space over the field $\mathbb{F}$ ($\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$) with the basis $\{x_1, ..., x_n\}$ then every linear functional f defined on X is of the form:

$$f(x) = \sum_{j=1}^{n} t_j a_j \quad \textit{for} \quad x = \sum_{j=1}^{n} t_j x_j; \quad t_j \in \mathbb{F};\ a_j = f(x_j)\ (j = 1, ..., n). \tag{2.13}$$

Indeed,

$$f(x) = f\big(\sum_{j=1}^{n} t_j x_j\big) = \sum_{j=1}^{n} f(t_j x_j) = \sum_{j=1}^{n} t_j f(x_j) = \sum_{j=1}^{n} t_j a_j.$$

Theorem 2.5. *Suppose that X is an n-dimensional linear space over the field $\mathbb{F}$ ($\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$) with the basis $\{x_1, ..., x_n\}$. Then there exist linearly independent linear functionals $f_1, ..., f_n \in X'$ such that*

$$f_j(x_m) = \delta_{jm} \quad (j, m = 1, ..., n), \tag{2.14}$$

where by δ_{jm} is denoted the so-called Kronecker symbol, *i.e.*

$$\delta_{jm} = \begin{cases} 1, & \text{if } m = j, \\ 0 & \text{if } m \neq j. \end{cases} \tag{2.15}$$

Proof. For an arbitrary element $x \in X$ of the form

$$x = \sum_{j=1}^{n} t_j x_j$$

define $f_j(x) = t_j$ $(j = 1, ..., n)$. By definition, $f_j(x_j) = 1$ and $f_j(x_m) = 0$ for $m \neq j$. We shall show that the functionals $f_1, ..., f_n$ are linearly independent. Suppose then that $f_1, ..., f_n$ are linearly dependent. Then there are scalars $a_1, ..., a_n$ non-vanishing simultaneously and such that $\sum_{j=1}^{n} a_j f_j = 0$. This implies that

$$a_m = \sum_{j=1}^{n} a_j \delta_{jm} = \sum_{j=1}^{n} a_j f_j(x_m) = \big(\sum_{j=1}^{n} a_j f_j\big)(x_m) = 0 \quad \textit{for } m = 1, ..., n.$$

This contradicts our assumption that $a_1, ..., a_n$ do not vanish simultaneously. Hence $f_1, ..., f_n$ are linearly independent. ■

Corollary 2.3. *If X is an n-dimensional linear space then dim X' = dim $X = n$ and the set$\{f_1, ..., f_n\}$ (where the functionals $f_1, ..., f_n \in X'$ are defined in Theorem 2.5) is a basis in X'.*

Indeed, by the general form (2.13) of linear functionals in an n-dimensional space X, we conclude that dim $X' = n$. Corollary 2.1 and Theorem 2.5 together imply that the set $\{f_1, ..., f_n\}$ is a basis in X'.

Theorem 2.6. *Suppose that X is a linear space over the field $\mathbb{F}$ ($\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$) and $g, f_1, ..., f_n \in X'$ satisfy the following condition:*

$$if \quad f_j(x) = 0 \quad for\ j = 1, 2, ..., n \quad then \quad g(x) = 0. \tag{2.16}$$

Then the functional g is linearly dependent on the functionals $f_1, ..., f_n$.

Proof. Without any loss of generality, we may assume that the functionals $f_1, ..., f_n$ are linearly independent. Write

$$X_0 = \{x \in X : f_j(x) = 0 \ (j = 1, ..., n)\}.$$

Clearly, codim $X_0 = n$. Condition (2.16) implies that $g(x) = 0$ for $x \in X_0$. Consider the quotient space X/X_0. Write

$$f'_j([x]) = f_j(x) \quad (j = 1, ..., n) \quad \textit{and} \quad g'([x]) = 0 \quad \textit{for} \quad x \in [x],$$

where $[x] = x + X_0$ are cosets induced by elements $x \in X$. Since $f_1(x) = 0$,...,$f_n(x) = 0$, $g(x) = 0$ for $x \in X_0$, the functionals $f'_1, ..., f'_n, g'$ are well defined. Moreover, $f'_1, ..., f'_n$ are linearly independent and dim $X/X_0 =$ codim $X_0 = n$. This implies that the functional g' is linearly dependent on the functionals $f_1, ..., f_n$. We therefore conclude that there are scalars $a_1, ..., a_n$ non-vanishing simultaneously and such that $g' = \sum_{j=1}^{n} a_j f'_j$. Hence, by definition, for every $x \in X$ we have

$$g(x) = g'([x]) = \sum_{j=1}^{n} a_j f'_j([x]) = \sum_{j=1}^{n} a_j f_j(x).$$

This implies that $g = \sum_{j=1}^{n} a_j f_j$, which was to be proved. ■

Theorem 2.7. *Suppose that X is an n-dimensional linear space over the field $\mathbb{F}$ ($\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$) and f is a linear functional defined on X. Then the set*

$$H_f = \{x \in X : f(x) = 1\} \tag{2.17}$$

is a hyperplane which does not contain the zero element. Conversely, if H is a hyperplane which does not contain the zero element, then there exist a linear functional f such that $H = \{x \in X : f(x) = 1\}$.

Proof. Let $x, y \in H_f$, $a, b \in \mathbb{F}$ and let $a + b = 1$. Then

$$f(ax + by) = af(x) + bf(y) = a + b = 1,$$

i.e. $ax+by \in H_f$. Write $H_f^0 = H_f - x_0$, where $x_0 \in H_f$. By Theorem 1.2, H_f^0 is a subspace of the linear space X and $H_f^0 = \{x \in X : f(x) = 0\}$.

Consider the quotient space X/H_f^0. Observe that the functional f is constant on each coset. Moreover, x and y belong to the same coset whenever $f(x) = f(y)$. Indeed,

$$[x] = \{y \in X; f(x - y) = 0\} = \{y \in X : f(x) = f(y)\}.$$

Hence there is a one-to-one linear correspondence between scalars and cosets. This implies that the quotient space X/H_f^0 is one-dimensional, i.e. H_f is a hyperplane. Since $f(0) = 0$, we conclude that $0 \notin H_f$.

Conversely, suppose that H is a hyperplane. By definition, H is a linear manifold such that the quotient space X/H^0, where H^0 is a subspace determined by H (cf. Theorem 1.2) is a one-dimensional subspace. This means that every element of the space X/H^0 can be written in the form $t[e]$, where $e \notin H^0$ is an arbitrary element, t is a scalar. We can choose e in such way that $[e] = H$. Indeed, $H^0 = H - x_0$ and H^0 does not depend on the choice of $x_0 \in H$. Therefore we can write

$$H = \{y \in X : y - x_0 \in H^0\} = [x_0].$$

If we define $e = x_0$ then we obtain the required equality $[e] = H$.

By our assumption, $0 \notin H$. Define on X a linear functional f by means of the equality

$$f(x) = t \quad for \quad x \in tH.$$

It is easy to verify that f is a functional such that $H_f = H$, Indeed,

$$H_f = \{xX : f(x) = 1\} = \{x \in X : x \in H\} = H, .$$

which was to be proved. ■

Let X and Y be linear spaces over a field $\mathbb{F}$. Every operator $A \in L_0(X \to Y)$ satisfies the following identity

$$(fA)x = f(Ax) \quad for\ all\ x \in X \quad and\ f \in Y', \tag{2.18}$$

where Y' is the space conjugate to Y. Then to the operator A there corresponds an operator A' defined by means of the equality

$$A'f = fA \quad for \quad f \in Y'. \tag{2.19}$$

The operator A' is said to be the *conjugate operator* to A. Clearly, $A' \in L(Y' \to X')$. We have

$$(A + B)' = A' + B', \quad (tA)' = tA' \quad whenever\ A, B \in L_0(X \to Y),\ t \in \mathbb{F}. \tag{2.20}$$

Indeed, by definition of conjugate operators, for arbitrary $x \in X$, $f' \in Y'$ we have

$$[(A + B)'f](x) = f[(A + B)x] = f(Ax + Bx) = f(Ax) + f(Bx) =$$

$$= (Af)(x) + (Bf)(x) = [(A' + B')f](x),$$

$$[(tA)'](x) = f(tAx) = tf(Ax) = (tA'f)(x),$$

which implies the required Formulae (2.20). In particular, it is easy to verify that

$$(I_X)' = I_{X'} \quad \textit{for an arbitrary linear space } X. \tag{2.21}$$

Consider three linear spaces X, Y, Z over the same field $\mathbb{F}$ of scalars. If $A \in L_0(Y \to Z)$, $B \in L_0(X \to Y)$, then

$$(AB)' = B'A'. \tag{2.22}$$

Indeed, by our assumption, the operator AB exists and belongs to $L_0(X \to Z)$. Hence $(AB)' \in L_0(Z' \to X')$. By definition, for arbitrary $x \in X$ and $f \in Z'$ we have

$$[(AB)'f](x) = f[(AB)x] = f[A(Bx)] = (A'f)(Bx) = (B'A'f)(x).$$

Formula (2.22) implies

$$(A')' = A \quad on \quad \kappa(X) \quad \textit{for an arbitrary } A \in L_0(X \to Y), \tag{2.23}$$

where κ is the canonical mapping of the space X into the space x''.

Indeed, for an arbitrary $f \in Y'$ we have

$$A'f' = (fA)' = (A'f)' = f'(A')' \quad on \quad \kappa(X).$$

Writing $g = f'$, we find $g(A')' = A'g = gA$. The arbitrariness of g implies the required Formula (2.23).

Suppose that X is an n-dimensional linear space with the basis $\{x_1, ..., x_n\}$ and Y is an m-dimensional space with the basis $\{y_1, ..., y_m\}$, both over the same field $\mathbb{F}$ of scalars. Let $A \in L_0(X \to Y)$. Then

$$Ax = \sum_{j=1}^{n} t_j A x_j \quad whenever \quad x = \sum_{j=1}^{n} t_j x_j, \quad t_1, ..., t_n \in \mathbb{F}.$$

On the other hand, since $Ax \in Y$, there are $c_1, ..., c_m \in \mathbb{F}$ such that $Ax = \sum_{k=1}^{m} c_k y_k$. We should determine the coefficients c_k. Since $Ax_j \in Y$, we have $Ax_j = \sum_{k=1}^{m} a_{jk} y_k$, where $a_{jk} \in \mathbb{F}$ $(j = 1, ..., n; k = 1, ..., m)$. The coefficients a_{jk} determine the way how the operator A transforms elements of the basis. Namely,

$$Ax = \sum_{j=1}^{n} t_j A x_j = \sum_{j=1}^{n} t_j \sum_{k=1}^{m} a_{jk} y_k = \sum_{j=1}^{n} \Big(\sum_{k=1}^{m} t_j a_{kj}\Big) y_k,$$

i.e.

$$c_k = \sum_{j=1}^{n} t_j a_{jk} \quad (k = 1, ..., m).$$

These last equalities show how to transform the coefficients of the expansion of a given element into the basis elements. Hence there is a one-to-one correspondence between the operator A and the system of $n \cdot m$ numbers a_{jk}:

$$A = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ \dots & \dots & \dots & \dots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{pmatrix} \tag{2.24}$$

The system of numbers

$$\begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ \dots & \dots & \dots & \dots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{pmatrix}$$

is said to be the *the matrix of the operator* A or, shortly, the *matrix*. Without any confusion, we can denote an operator and its matrix by the same letter. If j is fixed then the system $a_{j1}, ..., a_{jm}$ of numbers is said to be the jth-*column* of the matrix A. If k is fixed then the system $a_{1k}, ..., a_{nk}$ is said to be the kth-*row* of the matrix A. The numbers a_{jk} are called jk-*entries* of A. If $m = n$ then A is said to be a *square matrix* (of *dimension* n). For our convenience, sometimes we shall denote matrices in the form:

$$A = [a_{jk}]_{\substack{j=1,\dots,n\\k=1,\dots,m}}.$$

Theorem 2.8. *If* $\dim X = n$, $\dim Y = m$, $A, B \in L_0(X \to Y)$ *and*

$$A = [a_{jk}]_{\substack{j=1,\dots,n\\k=1,\dots,m}}, \quad B = [b_{jk}]_{\substack{j=1,\dots,n\\k=1,\dots,m}},$$

then

$$A + B = [a_{jk} + b_{jk}]_{\substack{j=1,\dots,n\\k=1,\dots,m}}, \quad \lambda A = [\lambda a_{jk}]_{\substack{j=1,\dots,n\\k=1,\dots,m}} \quad (\lambda \in \mathbb{F}). \tag{2.25}$$

The proof immediately follows from the fact that $(A + B)x = Ax + Bx$, $\lambda(Ax) = A(\lambda x)$ for $x \in X$.

Theorem 2.9. *If* $\dim X = n$, $\dim Y = m$, $\dim Z = p$, $A \in L_0(Y \to Z)$, $B \in L_0(X \to Y)$ *and*

$$A = [a_{jk}]_{\substack{j=1,\dots,n\\k=1,\dots,p}}, \quad B = [b_{lj}]_{\substack{l=1,\dots,n\\j=1,\dots,m}},$$

then there exists the superposition $AB \in L_0(X \to Z)$ *and*

$$A = [c_{lk}]_{\substack{l=1,\dots,n\\k=1,\dots,p}},$$

where

$$c_{lk} = \sum_{j=1}^{m} a_{jk} b_{lj} \qquad (l = 1, ..., n;\ k = 1, ..., p). \tag{2.26}$$

Proof. Let $\{x_1, ..., x_n\}$, $\{y_1, ..., y_n\}$, $\{z_1, ..., z_n\}$ denote the bases in the spaces X, Y, Z, respectively. By definition of matrices, we have

$$Bx_l = \sum_{j=1}^{n} b_{lj} y_j \quad (l = 1, ..., n), \quad Ay_j = \sum_{k=1}^{p} a_{jk} z_k \quad (j = 1, ..., m).$$

Hence for $l = 1, ..., n$

$$ABx_l = A\Big(\sum_{j=1}^{m} b_{lj} y_j\Big) = \sum_{j=1}^{m} b_{lj} Ay_j =$$

$$= \sum_{j=1}^{m} b_{lj} \sum_{k=1}^{p} a_{jk} z_k = \sum_{k=1}^{p} \Big(\sum_{j=1}^{m} a_{jk} b_{lj}\Big) z_k = \sum_{k=1}^{p} c_{lk} z_k,$$

which implies the required Formula (2.26). ■

Formula (2.26) implies that in order to multiply matrices A and B, we multiply *every row* of the matrix A by *every column* of the matrix B. Thus the multiplication of matrices is not commutative: in general, $AB \neq BA$. For instance,

$$if \quad A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad then \quad AB = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}, \quad BA = \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix} \neq AB.$$

The above example implies that the algebra $L_0(X)$ is *noncommutative* whenever dim $X > 1$. On the other hand, general properties of linear operators imply that the multiplication of matrices is distributive with respect to their addition and it is also associative, whenever it is well defined.

Let

$$A \in L_0(X \to Y) \quad and \quad A = [a_{jk}]_{\substack{j=1,\ldots,n \\ k=1,\ldots,m}}.$$

The matrix

$$A^T = [a_{kj}]_{\substack{k=1,\ldots,m \\ j=1,\ldots,n}} \tag{2.27}$$

is said to be the *transposed matrix* of the matrix A. It means the *transposition* of a matrix changes rows onto columns and columns onto rows.

Theorem 2.10. *Let $\{x_1, ..., x_n\}$, $\{y_1, ..., y_m\}$ be bases in linear spaces X and Y, respectively. Let $A \in L_0(X \to Y)$ and let $A = [a_{jk}]_{\substack{j=1,\ldots,n \\ k=1,\ldots,m}}$. Then $A' = A^T$, where the matrix A^T is defined by Formula (2.27).*

In other words, *in finite dimensional linear spaces conjugate operators are determined by transposed matrices.*

Proof. Let $f_i \in Y'$, $g_j \in X'$ be functionals such that

$$f_i(y_j) = \delta_{ij} \quad g_l(x_k) = \delta_{lk} \quad (i, j = 1, ..., m;\ l, k = 1, ..., n).$$

By Theorem 2.5, such functionals exist and $\{f_1, ..., f_m\}$, $\{g_1, ..., g_n\}$ are linearly independent systems. By Corollary 2.3, these systems are bases in the spaces Y' and X', respectively. Since $Ax_j \in Y$, the definition of matrices implies that

$$f_i(Ax_j) = f_i\Big(\sum_{k=1}^{m} a_{jk}y_k\Big) = \sum_{k=1}^{m} a_{jk}f_i(y_k) = \sum_{k=1}^{m} a_{jk}\delta_{ij} = a_{ji}$$

$$(j = 1, ..., n;\ i = 1, ..., m).$$

On the other hand, writing $A' = [b_{jk}]_{\substack{j=1,\dots,n\\ k=1,\dots,m}}$, by definition of the conjugate operator $A' \in L_0(Y' \to X')$, we find

$$f_i(Ax_j) = (A'f_i)x_j = \Big(\sum_{l=1}^{n} b_{il}g_l\Big)x_j = \sum_{l=1}^{n} b_{il}g_l(x_j) = \sum_{l=1}^{n} b_{il}\delta_{lj} = b_{ij}$$

$$(j = 1, ..., n;\ i = 1, ..., m).$$

Hence $b_{ij} = a_{ji}$ $(j = 1, ..., n;\ i = 1, ..., m)$, which implies $b_{jk} = a_{kj}$ $(j = 1, ..., n;\ k = 1, ..., m)$. ■

Observe that a transposed matrix of a transposed matrix is equal to the given matrix,

$$(A^T)^T = A$$

(cf. Formula (2.23)).

An operator $A \in L_0(X \to Y)$ is said to be *finite dimensional* if its range $\mathcal{E}_A$ is finite dimensional. If dim $\mathcal{E}_A = n$ then we say that A is an *n-dimensional operator*. Hence $A \in L_0(X \to Y)$ is a finite dimensional operator whenever dim $Y < +\infty$.

Theorem 2.11. *An operator $K \in L_0(X \to Y)$ is n-dimensional if and only if it is of the form:*

$$Kx = \sum_{j=1}^{n} f_j(x)y_j \quad \textit{for} \quad x \in X \tag{2.28}$$

for fixed $f_1, ..., f_n \in X'$ and fixed linearly independent $y_1, ..., y_n \in Y$.

Proof. Suppose that K is of the form (2.28). Write $a_j = f_j(x)$ for $j = 1, ..., n$. By definition, $a_1, ..., a_n$ are scalars. Then for an arbitrary $x \in X$ we have $Kx = \sum_{j=1}^{n} a_jy_j$. This means that K maps X onto the set lin $\{y_1, ..., y_n\} \subset Y$. Hence dim KX = dim lin $\{y_1, ..., y_n\} = n$. We therefore conclude that the operator K is n-dimensional.

Conversely, suppose that the operator K is n-dimensional, i.e. dim $KX = n$. Then there are n linearly independent elements $y_1, ..., y_n \in Y$ such that for every $x \in X$ we have $Kx = \sum_{j=1}^{n} a_j y_j$ for some scalars $a_1, ..., a_n$. Let $Y_0 =$ lin $\{y_1, ..., y_n\}$. Define linear functionals $\varphi_j \in Y'$ by means of the equality

$$\varphi_j\big(\sum_{m=1}^{n} a_m y_m\big) = a_j \quad (j = 1, ..., n).$$

Let $f_j(x) = \varphi_j(Kx)$ for $j = 1, ..., n$. Then $f_1, ..., f_n$ are linear functionals, since they are superpositions of linear functionals and a linear operator. Moreover, by definition,

$$f_j(x) = \varphi_j(Kx) = \varphi\big(\sum_{m=1}^{n} a_m y_m\big) = a_j,$$

which was to be proved. ■

Let be given a linear space X over the field $\mathbb{F}$ and its subspace Y. Let $\mathcal{S}$ be a subset of $L_0(X)$. Then Y is said to be an *$\mathcal{S}$-invariant subspace* if $AY \subset Y$ for all $A \in \mathcal{S}$. If $\mathcal{S} = \{A\}$ for an $A \in L_0(X)$, an $\mathcal{S}$-invariant subspace is called shortly: an *A-invariant* subspace.

Denote by $\mathbb{F}[t]$ the set of all polynomials in the variable t with coefficients in $\mathbb{F}$ (cf. Examples 1.5 and 1.6).

Theorem 2.12. *Let $A \in \mathcal{S} = L_0(X)$ and let $AB = BA$ for all $B \in \mathcal{S}$. Then*

(i) the range and the kernel of A are $\mathcal{S}$-invariant subspaces;

(ii) $p(A)B = Bp(A)$ for all $B \in \mathcal{S}$ whenever $p \in \mathbb{F}[t]$.

Proof. (i) Let $B \in \mathcal{S}$ be arbitrarily fixed. Let $y = Ax$ for an $x \in X$. Then $By = BAx = ABx$. This means that By also belongs to the range of A and this range is an $\mathcal{S}$-invariant subspace. Let $x \in$ ker A. Then $ABx = BAx = 0$. Hence also $Bx \in$ ker A, which implies that the kernel of A is an $\mathcal{S}$-invariant subspace. Point (ii) easily follows from our assumptions. ■

Other properties of linear operators in finite dimensional linear spaces and their matrices will be considered in the next chapter.

Examples and Exercises.

Example 2.1. Consider the space $\mathbb{C}$ of complex numbers with the usual addition and multiplication by *real* numbers. Define the mapping A of the space $\mathbb{C}$ into itself by means of the equality

$$Az = az + b\bar{z}, \quad where \quad a, b \in \mathbb{R},$$

where $\bar{z}$ denotes the complex number conjugate to z. It is easy to verify that A is a linear operator. Write

$$f(z) = \mathrm{Re}\ z, \quad g(z) = \mathrm{Im}\ z, \quad h(z) = |z| \quad for\ z \in \mathbb{C}.$$

It is easy to verify that f, g, h map $\mathbb{C}$ into $\mathbb{R}$, hence they are functionals. Moreover, f and g are linear, however, h is not a linear functional. Indeed,

$$h(1+i) = |1+i| = \sqrt{2}, \quad h(1) + h(i) = |1| + |i| = 2 \neq h(1+i).$$

$\square$

Example 2.2. Consider the space $X = C[a,b]$. Suppose, we are given a function $g \in C[a,b]$ such that $g(t) \neq 0$ for $a \leq t \leq b$. Write

$$(Ax)(t) = g(t)x(t) \quad for \quad x \in C[a,b],\ t \in [a,b].$$

It is easy to verify that $A \in L_0(X)$. Moreover, the operator A is invertible and

$$(A^{-1}x)(t) = \frac{1}{g(t)}x(t) \quad for \quad x \in C[a,b],\ t \in [a,b].$$

Suppose now that there is given a function $h \in C^1[a,b]$ such that $h(a) = a$, $h(b) = b$ and, moreover, $h'(t) > 0$ for $a \leq t \leq b$. Write

$$(Bx)(t) = x(h(t)) \quad for \quad x \in C[a,b],\ t \in [a,b].$$

It is easy to verify that $B \in L_0(X)$. Moreover, the operator B is invertible and

$$(B^{-1}x)(t) = x(h^{-1}(t)) \quad for \quad x \in C[a,b],\ t \in [a,b],$$

where by h^{-1} we denote the inverse function of h, which exists by our assumption that $h'(t) > 0$ for $a \leq t \leq b$.

Write now

$$f(x) = \sum_{j=1}^{n} a_j x(t_j) \quad for \quad x \in C[a,b],$$

where $a_1, ..., a_n \in \mathbb{R}$, $t_1, ..., t_n$ are arbitrarily fixed. It is easy to verify that f is a linear functional over the space $C[a,b]$. $\square$

If linear spaces $X_1, ..., X_n$ (over the same field $\mathbb{F}$ of scalars) are finite dimensional the following obvious equalities hold:

$$\dim\ X_1 \times X_2 \times ... \times X_n = \dim\ X_1 + \dim\ X_2 + ... + \dim\ X_n, \tag{2.29}$$

$$\dim\ X_1 \oplus X_2 \oplus ... \oplus X_n = \dim\ X_1 + \dim\ X_2 + ... + \dim\ X_n, \tag{2.30}$$

If at least one of the spaces $X_1, ..., X_n$ is infinite dimensional then both, the Cartesian product $X_1 \times X_2 \times ... \times X_n$ and the direct sum $X_1 \oplus X_2 \oplus ... \oplus X_n$ are infinite dimensional. We therefore conclude that Formulae (2.29) and (2.30) hold also in the case of infinite dimensional linear spaces. □

Exercise 2.1. Is a translation of points in the space $\mathbb{R}^n$ ($n = 1, 2....$) (i.e. a mapping of the form: $T_h x = x + h$, where $h \in \mathbb{R}^n$ is a fixed point) a linear operator ? □

Exercise 2.2. Give examples of projection operators in the spaces $\mathbb{R}^n$, $\mathbb{F}_n[t]$ ($n \in \mathbb{N}$) and $C[a, b]$ (cf. Examples 1.3, 1.5 and 1.9). □

Exercise 2.3. Prove that the rotation by an angle φ in the space $\mathbb{R}^2$ is a linear operator. Determine its matrix. □

Exercise 2.4. Give examples of linear functionals in the spaces $\mathbb{R}^n$ ($n \in \mathbb{N}$) and determine the corresponding hyperplanes (cf. Theorem 2.7). For instance, if $x \in (x_1, ..., x_n) \in \mathbb{R}^n$ and $f(x) = x_1 + ... + x_n$ then $H_f = \{x \in \mathbb{R}^n : x_1 + ... + x_n = 1\}$. Determine the general form of hyperplanes in $\mathbb{R}^n$ containing zero. □

Exercise 2.5. Give examples of linear functionals in the spaces $\mathbb{F}_n[t]$, $\mathbb{F}[t]$, $C[a, b]$ and determine the corresponding hyperplanes (cf. Examples 1.5, 1.6 and 1.9). □

Exercise 2.6. Mappings preserving all linear manifolds are called *affine transformations*. Describe all affine transformations in the spaces $\mathbb{R}^2$ and $\mathbb{R}^3$. □

Exercise 2.7. Suppose that

$$A = [a_{jk}]_{\substack{j=1,...,n \\ k=1,...,m}}, \quad B = [b_{jk}]_{\substack{j=1,...,n \\ k=1,...,m}}, \quad C = [c_{lj}]_{\substack{l=1,...,p \\ j=1,...,n}}.$$

Prove that

$$(A + B)^T = A^T + B^T, \quad (tA)^T = tA^T \quad for\ t \in \mathbb{F}, \quad (A^T)^T = A, \quad (CA)^T = A^T C^T$$

(cf. Formulae (2.20), (2.22), (2.23)). □

Exercise 2.8. A square matrix A is said to be *symmetric* if $A^T = A$ and *antisymmetric* if $A^T = -A$. Prove that every square matrix A is a sum of a symmetric matrix A^+ and an antisymmetric A^- and that this sum is uniquely determined, namely

$$A = A^+ + A^-, \quad where \quad A^+ = \frac{1}{2}(A + A^T), \quad A^- = \frac{1}{2}(A - A^T).$$

□

Exercise 2.9. Write $E = [\delta_{jk}]_{j,k=1,...,n}$, where δ_{jk} is the Kronecker symbol (cf. Formula (2.15)). Prove that for every square matrix A of dimension n we have $AE = EA = A$, i.e. E is the matrix of the identity operator I. □

Exercise 2.10. Prove that set of all square matrices of dimension n is an algebra with unit with the addition, multiplication by scalars determined and multiplication determined by Formulae (2.25), (2.26), respectively. Give examples of zero divisors in this algebra. □

Exercise 2.11. A square matrix A of dimension n is said to be a *diagonal matrix* if $a_{kk} = \lambda_k$ and $a_{jk} = 0$ for $j \neq k$, where $\lambda_k \in \mathbb{F}$ $(j, k = 1, ..., n)$. The system $\{a_{11}, ..., a_{nn}\}$ is called the *principal diagonal* (otherwise just the *diagonal*). Prove that every diagonal matrix of dimension n has the following properties:

(i) $A^T = A$;

(ii) A^N $(N \in \mathbb{N})$ is also a diagonal matrix with the principal diagonal $\{\lambda_1^N, ..., \lambda_n^N\}$, where we define by induction $A^{k+1} = A \cdot A^k$ $(k \in \mathbb{N})$;

(iii) the set of all diagonal matrices of dimension n is a subalgebra of the algebra described in Exercise 2.10. Has this subalgebra zero divisors ? □

Exercise 2.12. Suppose that either $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. Let $A = [a_{jk}]_{j,k=1,...,n}$, where $a_{n-k+1,k} = \lambda \in \mathbb{F}$, $a_{jk} = 0$ for $j \neq n - k + 1$ $(j, k = 1, ..., n)$. Prove that $A^T = A$. Does the equality $A^T = A$ hold also when $a_{n-k+1,k} = \lambda_k \in \mathbb{F}$ are different from each other ? □

Exercise 2.13. An operator $A \in L_0(X)$ is said to be *nilpotent of order* n if $A^n = 0$ but $A^{n-1} \neq 0$ $(n \in \mathbb{N}$. Let $A \in L_0(X)$ be nilpotent of order n. Prove that there exists an element $x_0 \in X$ such that $x_0 \neq 0$ and elements $x_0, Ax_0, ..., A^{n-1}x_0$ are linearly independent. □

Exercise 2.14. Let $X \neq \{0\}$ be a finite dimensional linear space over the field $\mathbb{F}$ and let $\mathcal{S} = L_0(X)$. Prove that the only $\mathcal{S}$-invariant subspaces of X are X itself and $\{0\}$. □

Exercise 2.15. Let $A \in L_0(X)$ be such that $AB = BA$ for all $B \in \mathcal{S} \subset L_0(X)$. Prove that the range and the kernel of A are $\mathcal{S}$-invariant subspaces of X. □

Exercise 2.16. Let X be a a linear space over the field $\mathbb{F}$ and let $\mathcal{S} = L_0(X)$. Suppose that U, V are $\mathcal{S}$-invariant subspaces of X. Prove that $U + V$ and $U \cap V$ are $\mathcal{S}$-invariant subspaces of X. □

Chapter 3.

Matrices and determinants. Solutions to systems of linear equations.

A field of scalars $\mathbb{F}$ has the *characteristic zero* if it is commutative and the intersection of all its subfields (which is again a subfield of $\mathbb{F}$) is isomorphic with the field $\mathbb{Q}$ of all rational numbers, i.e. if there is a one-to-one mapping f of $\mathbb{F}$ which preserves the operations in $\mathbb{F}$:

$$f(x+y) = f(x) + f(y), \quad f(O) = 0, \quad f(-x) = -f(x), \quad f(xy) = f(x)f(y),$$

$$f(1) = 1 \quad and \quad f(x^{-1}) = [f(x)]^{-1} \quad for\ all\ x, y \in \mathbb{F}.$$

For instance, $\mathbb{Q}$ itself, the field $\mathbb{R}$ of reals and the field $\mathbb{C}$ of complex numbers are fields of characteristic zero. All calculations in a field of characteristic zero proceed in the same manner as in these fields. Thus all considerations of Chapters 1 and 2 are valid for fields of characteristic zero. In the sequel, in general, we shall mean by a "field" - a field of characteristic zero.

A *permutation* of numbers $1, ..., n \in \mathbb{N}$ is any one-to-one mapping p of the set $\{1, ..., n\}$ onto itself. We shall denote permutations by $\{p_1, ..., p_n\}$. By definition, the number of different permutations of the numbers $1, ..., n$ is $n!$.

Write

(3.1) $$\text{sign } a = \begin{cases} 1, & \text{if } a > 0, \\ 0 & \text{if } a = 0, \\ -1 & \text{if } a < 0, \end{cases} \qquad whenever \quad a \in \mathbb{R}.$$

It is easy to verify by induction that

$$\text{sign } a_1 ... a_n = (\text{sign } a_1) ... (\text{sign } a_n).$$

A *sign of a system* $\{a_1, ..., a_n\}$ of real numbers is said to be the number

(3.2) $$\text{sign } \{a_1, ..., a_n\} = \prod_{j,k=1,...,n; j>k} \text{sign } (a_j - a_k).$$

Suppose that a system $\{b_1, ..., b_n\} \subset \mathbb{R}$ is obtained from a system $\{a_1, ..., a_n\} \subset \mathbb{R}$ by a transposition of two terms: $b_i = a_i$ for $i \neq j$, $i \neq k$, $b_j = a_k$, $b_k = a_j$, (j, k are arbitrarily fixed). Then

(3.3) $$\text{sign } \{b_1, ..., b_n\} = - \text{ sign } \{a_1, ..., a_n\},$$

i.e. the system changes its sign.

Indeed.

$$\text{sign}\ \{a_1, ..., a_n\} =$$

$$= \text{sign}\ (a_j - a_k)\Big[\prod_{\substack{r,s=1,\dots,n \\ r>s;\ s\neq j,k}} \text{sign}\ (a_r - a_s)\Big]\ \Big[\prod_{\substack{r=1,\dots,n \\ r\neq j,k}} \text{sign}\ (a_j - a_r)\text{sign}\ (a_r - a_k)\Big].$$

Observe that the interchange of the term a_j and the term a_k does not change the sign of the last two products. This easily implies Formula (3.3).

The *determinant* of a square matrix $A = [a_{jk}]_{j,k=1,\dots,n}$ is said to be the number (belonging to the field $\mathbb{F}$ of scalars under consideration which, by our assumption, should have the characteristic zero)

$$\text{(3.4)} \qquad \det\ A = \sum_{\{p_1,\dots,p_n\}} \text{sign}\ \{p_1, ..., p_n\}\ a_{1p_1}...a_{np_n},$$

where the summation is extended over all permutations $\{p_1, ..., p_n\}$ of the numbers $1, ..., n$. The determinant of the matrix $A = [a_{jk}]_{j,k=1,\dots,n}$ is denoted also in the following way:

$$\det\ A = \begin{vmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{vmatrix}.$$

Columns and *rows* of a determinant det A are, by definition, columns and rows of the matrix A.

Example 3.1. If $n = 1$ then det $A = |a_{11}|$. If $n = 2$ then

$$\det\ A = \begin{vmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

If $n = 3$ then

$$\det\ A = \begin{vmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{vmatrix} =$$

$$= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31} - a_{13}a_{22}a_{31}.$$

Observe that in each product $a_{1p_1}...a_{np_n}$ a term from every column and every row appears only once. □

We shall give now the fundamental properties of determinants.

Property 3.1. *The determinant of a matrix such that either one column or one row consists of zeros only is equal to zero.*

Proof. Suppose that $a_{jk} = 0$ for a j and $k = 1, ..., n$. Then any component of the sum (3.4) contains a term of the form a_{jp_j}. Hence det $A = 0$. A similar result holds if $a_{jk} = 0$ for $j = 1, ..., n$ and a k. ■

Property 3.2. *If* $\det B$ *is formed from a determinant* $\det\ A$ *by multiplication of one column (one row) by a scalar* c*, then* $\det\ B = c \det\ A$.

Proof. If we multiply the jth column by c, then we find

$$\det\ B = \sum_{\{p_1,\ldots,p_n\}} \text{sign}\ \{p_1,...,p_n\} a_{1p_1} ....a_{j-1,p_{j-1}} c a_{j+1,p_{j+1}} ...a_{np_n} =$$

$$= \sum_{\{p_1,\ldots,p_n\}} c\ \text{sign}\ \{p_1,...,p_n\} a_{1p_1} ....a_{j-1,p_{j-1}} a_{j+1,p_{j+1}} ...a_{np_n} =$$

$$= c \sum_{\{p_1,\ldots,p_n\}} \text{sign}\ \{p_1,...,p_n\} a_{1p_1} ....a_{np_n} = c\ \det\ A.$$

A similar proof holds for rows. ■

Property 3.3. *If the terms of the* j*th column (row) of a determinant* $\det\ A$ *are sums of two components*

$$a_{jk} = a'_{jk} + a''_{jk} \quad (k = 1,...,n) \quad (j = 1,...,n)$$

then this determinant is a sum of two determinants such that the j*th column (row) of the first one consists of terms* a'_{jk} *and the* j*th column (row) of the second one consists of terms* a''_{jk}*, i.e.* $A = B + C$*, where* $B = \det[b_{im}]_{i,m=1,\ldots,n}$, $C = \det[c_{im}]_{i,m=1,\ldots,m}$, $b_{im} = c_{im} = a_{im}$ *for* $i \neq j$, $b_{jm} = a'_{jm}$, $c_{jm} = a''_{jm}$ *for* $i = j$ $(m = 1,...,n)$*. Similarly, for rows* $b_{im} = c_{im} = a_{im}$ *for* $m \neq k$, $b_{ik} = a'_{ik}$, $c_{ik} = a''_{ik}$ *for* $m = k$ $(i = 1,...,n)$.

Proof. By our assumptions,

$$\det\ A = \sum_{\{p_1,\ldots,p_n\}} \text{sign}\ \{p_1,...,p_n\} a_{1p_1} ....a_{j-1,p_{j-1}} a_{jp_j} a_{j+1,p_{j+1}} ...a_{np_n} =$$

$$= \sum_{\{p_1,\ldots,p_n\}} \text{sign}\ \{p_1,...,p_n\} a_{1p_1} ....a_{j-1,p_{j-1}} (a'_{jp_j} + a''_{jp_j}) a_{j+1,p_{j+1}} ...a_{np_n} =$$

$$= \sum_{\{p_1,\ldots,p_n\}} \text{sign}\ \{p_1,...,p_n\} a_{1p_1} ...a_{j-1,p_{j-1}} a'_{jp_j} a_{j+1,p_{j+1}} ...a_{np_n} +$$

$$+ \sum_{\{p_1,\ldots,p_n\}} \text{sign}\ \{p_1,...,p_n\} a_{1p_1} ...a_{j-1,p_{j-1}} a''_{jp_j} a_{j+1,p_{j+1}} ...a_{np_n}.$$

A similar proof holds for rows. ■

Formula (3.3) immediately implies

Property 3.4. *If* $n \geq 2$ *and in a determinant* $\det\ A$ *two columns (two rows) change their place, then the determinant* $\det\ B$ *obtained in this way changes its sign:* $\det\ B = -\det\ A$.

Property 3.5. *If a determinant* $\det\ A$ *contains two identical columns (rows), then* $\det\ A$ *= 0.*

Proof. If we exchange the place of the identical columns, then we obtain a new determinant $\det\ B = -\det\ A$. By Property 3.4,

$$\det\ A = \det\ B = -\det\ A \quad \textit{which implies} \quad 2\det\ A = 0 \quad \textit{and} \quad \det\ A = 0.$$

A similar proof holds for rows. ■

Property 3.6. *A determinant does not change its value if to elements of its column (row) there are added elements of another column (row) multiplied by an arbitrary number c.*

Proof. By our assumptions,

$$\sum_{\{p_1,\dots,p_n\}} \text{sign } \{p_1,...,p_n\} a_{1p_1}...a_{j-1,p_{j-1}}(a_{jp_j} + c a_{kp_k}) a_{j+1,p_{j+1}}...a_{np_n} =$$

$$\sum_{\{p_1,\dots,p_n\}} \text{sign } \{p_1,...,p_n\} a_{1p_1}...a_{j-1,p_{j-1}} a_{jp_j} a_{j+1,p_{j+1}}...a_{np_n} +$$

$$+c \sum_{\{p_1,\dots,p_n\}} \text{sign } \{p_1,...,p_n\} a_{1p_1}...a_{j-1,p_{j-1}} a_{kp_k} a_{j+1,p_{j+1}}...a_{np_n} = \det\ A + c\det\ A'.$$

But $\det\ A'$ has two identical columns: the jth column and the kth column. This, and Property 3.5 together imply that $\det A' = 0$. A similar proof holds for rows. ■

A *minor determinant* of a square matrix A of dimension n is said to be a determinant of A obtained by canceling the same number of columns and rows in A. Denote by M_{jk} minor determinants obtained by canceling the jth column and the kth row. Minor determinants M_{jk} are determinants of a matrix of dimension $n-1$. By definition,

$$M_{jk} = \tag{3.5}$$

$$= \sum_{\{p_1,\dots p_{j-1},p_{j+1},\dots,p_n\}} \text{sign } \{p_1,...,p_{j-1},p_{j+1},...,p_n\} a_{1p_1}...a_{j-1,p_{j-1}} a_{j+1,p_{j+1}}...a_{np_n}.$$

Theorem 3.1 (Laplace Theorem). *Let* $\det\ A = \det\ [a_{jk}]_{j,k=1,\dots,n}$. *Then*

$$\det\ A = \sum_{j=1}^{n} (-1)^{j+k} a_{jk} M_{jk}, \tag{3.6}$$

$$\det\ A = \sum_{k=1}^{n} (-1)^{j+k} a_{jk} M_{jk}. \tag{3.7}$$

Proof. We shall prove only Formula (3.6). A proof of Formula (3.7) is similar. Observe that if we cancel in the permutation $\{p_1,...,p_n\}$ the term $p_j = k$ then

$$\text{sign } \{p_1,...,p_n\} = (-1)^{j+k} \text{sign } \{p_1,...,p_{j-1},p_{j+1},...,p_n\}. \tag{3.8}$$

Indeed, let

q_1 be the number of those $p_1, .., p_{j-1}$ which are less than k,
q_2 be the number of those $p_1, .., p_{j-1}$ which are greater than k,
r_1 be the number of those $p_{j+1}, .., p_n$ which are less than k,
r_2 be the number of those $p_{j+1}, .., p_n$ which are greater than k.

If we cancel the term $p_j = k$ then we have to multiply sign $\{p_1, ..., p_n\}$ by the number $(-1)^{q_2+r_1}$. On the other hand,

$$q_1 + q_2 = j - 1, \qquad q_1 + r_1 = k - 1,$$

which implies

$$2q_1 + q_2 + r_1 = j + k - 1.$$

Hence

$$(-1)^{q_2+r_1} = (-1)^{j+k-2-2q_1} = (-1)^{j+k}.$$

Then, by Formula (3.5), we find

$$(-1)^{j+k} a_{jk} M_{jk} =$$

$$\sum_{\substack{\{p_1, \dots p_{j-1}, p_{j+1}, \dots, p_n\} \\ p_j = k}} \text{sign } \{p_1, ..., p_{j-1}, p_{j+1}, ..., p_n\} (-1)^{j+k} a_{1p_1} ... a_{j-1,p_{j-1}} a_{jk} a_{j+1,p_{j+1}} ... a_{np_n}$$

$$= \sum_{\substack{\{p_1, \dots p_{j-1}, p_{j+1}, \dots, p_n\} \\ p_j = k}} \text{sign } \{p_1, ..., p_n\} a_{1p_1} ... a_{j-1,p_{j-1}} a_{jp_j} a_{j+1,p_{j+1}} ... a_{np_n} =$$

$$= \sum_{\substack{\{p_1, \dots p_{j-1}, p_{j+1}, \dots, p_n\} \\ p_j = k}} \text{sign } \{p_1, ..., p_n\} a_{1p_1} ... a_{np_n}.$$

We therefore conclude that

$$(-1)^{j+k} a_{jk} M_{jk} = \sum_{j=1}^{N} \sum_{\substack{\{p_1, \dots p_{j-1}, p_{j+1}, \dots, p_n\} \\ p_j = k}} \text{sign } \{p_1, ..., p_n\} a_{1p_1} ... a_{np_n} =$$

$$= \sum_{\{p_1, \dots, p_n\}} \text{sign } \{p_1, ..., p_n\} a_{1p_1} ... a_{np_n} = \det \ A.$$

$\blacksquare$

Theorem 3.2 (Cauchy Theorem). *If A and B are square matrices of the same dimension then*

$$\det \ (AB) = (\det \ A)(\det \ B).$$

Proof (by induction). Clearly, Theorem 3.2 is true for $n = 1$. Indeed, for $A = [a_{11}]$, $B = [b_{11}]$ we have

$$(\det\ A)(\det\ B) = a_{11}b_{11} = \det\ [a_{11}b_{11}] = \det\ (AB).$$

Suppose now that Theorem 3.2 holds for an arbitrarily fixed positive integer $n \geq 1$. By Theorem 2.9, for arbitrary matrices P and Q of dimension n

$$P = [p_{jk}]_{j,k=1,\dots,n}, \quad Q = [q_{jk}]_{j,k=1,\dots,n}$$

the following formula is satisfied:

$$(3.9). \qquad \det\ \Big[\sum_{j=1}^{n} p_{jk}q_{lj}\Big]_{k,l=1,\dots,n} = \big(\det\ [p_{jk}]_{j,k=1,\dots,n}\big)\big(\det\ [q_{lj}]_{l,j=1,\dots,n}\big)$$

Let A and B be arbitrary square matrices of dimension $n + 1$,

$$A = [a_{jk}]_{j,k=1,\dots,n+1}, \quad B = [b_{jk}]_{j,k=1,\dots,n+1}.$$

Write

$$C = AB, \quad i.e. \quad C = [c_{lk}]_{l,k=1,\dots,n+1} \quad where \quad c_{lk} = \sum_{j=1}^{n+1} a_{jk}b_{lj}.$$

Denote by M'_{jk}, M''_{jk}, M''_{jk} minor determinants of matrices A, B, C, respectively, which are obtained by canceling their jth column and kth row. Denote also by A_{jk}, B_{jk}, C_{jk} matrices of those minor determinants, respectively. Then

$$M'_{jk} = \det\ A, \quad M''_{jk} = \det\ B_{jk}, \quad M_{jk} = \det\ C_{jk} \quad (j, k = 1, ..., n).$$

By our induction assumption,

$$(3.10) \qquad M'_{jk}M''_{km} = M_{jm} \quad (j, k, m = 1, ..., n).$$

Indeed,

$$M'_{jk}M''_{km} = (\det\ A_{jk})(\det\ B_{km}) = \det\ (A_{jk}B_{km}) = \det\ C_{jm} = M_{jm}.$$

This, the Laplace theorem (Theorem 3.1) and Theorem 2.9 together imply

$$(\det\ A)(\det\ B) = \Big[\sum_{j=1}^{n+1} (-1)^{j+k}a_{jk}M'_{jk}\Big]\Big[\sum_{k=1}^{n+1} (-1)^{k+m}b_{km}M''_{km}\Big] =$$

$$= \sum_{j,k=1}^{n+1} (-1)^{m+j+2k}a_{jk}b_{km}M'_{jk}M''_{km} =$$

$$= \sum_{j,k=1}^{n+1} (-1)^{m+j} a_{jk} b_{km} M_{jk} = \sum_{j=1}^{n+1} (-1)^{m+j} \Big(\sum_{k=1}^{n+1} a_{jk} b_{km} \Big) M_{jk} =$$

$$= \sum_{j=1}^{n+1} (-1)^{m+j} c_{jm} M_{jm} = \det\ C = \det\ (AB).$$

We therefore have verified that the formula $\det\ (AB) = (\det\ A)(\det\ B)$ is true for square matrices of dimension $n+1$. Hence Theorem 3.2 holds for square matrices of an arbitrary dimension. ■

Example 3.2. Let

$$A = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{pmatrix}.$$

Then

$$\det\ (AB) = \det\ \begin{pmatrix} a_{11}b_{11} + a_{21}b_{21} & a_{11}b_{21} + a_{21}b_{22} \\ a_{12}b_{11} + a_{22}b_{12} & a_{12}b_{21} + a_{22}b_{22} \end{pmatrix} =$$

$$= (a_{11}b_{11} + a_{21}b_{21})(a_{12}b_{21} + a_{22}b_{22}) - (a_{11}b_{21} + a_{21}b_{22})(a_{12}b_{11} + a_{22}b_{12}) =$$

$$= a_{11}a_{12}b_{11}b_{21} + a_{21}a_{12}b_{12}b_{21} + a_{11}a_{22}b_{11}b_{22} + a_{21}a_{22}b_{12}b_{22} -$$

$$- a_{11}a_{12}b_{11}b_{21} - a_{21}a_{12}b_{11}b_{22} - a_{11}a_{22}b_{21}b_{12} - a_{21}a_{22}b_{12}b_{22} =$$

$$= (a_{21}a_{12} - a_{11}a_{22})b_{12}b_{21} + (a_{11}a_{22} - a_{21}a_{22})b_{11}b_{22} =$$

$$= (a_{11}a_{22} - a_{12}a_{22})(b_{11}b_{22} - b_{12}b_{22}) =$$

$$= \begin{vmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{vmatrix} \cdot \begin{vmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{vmatrix} = (\det\ A)(\det\ B).$$

■

Theorem 3.3. *If A is a square matrix then* $\det\ A^T = \det\ A$.

Proof by induction with respect to the dimension of the matrix A. For $n = 1$ this theorem is obvious, since $A^T = [a_{11}] = A$. Suppose that Theorem 3.3 is true for an arbitrarily fixed $k = n-1$, where $n > 1$. Similarly, as in Formula (3.5), denote by M_{kj}^T the minor determinants of the matrix A^T. Then, by our induction assumption, $M_{kj}^T = M_{jk}$, since these determinants are determinants of matrices of dimension $n-1$. By Theorem 3.1 applied to the matrix A^T with entries $b_{jk} = a_{kj}$, we find for $k = n$

$$\det\ A = \sum_{j=1}^{n} (-1)^{j+k} b_{jk} M_{jk}^T = \sum_{j=1}^{n} (-1)^{j+k} a_{kj} M_{kj} = \det\ A.$$

■

Corollary 3.1. *If A and B are square matrices of the same dimension, then*

$$\det\ (A^T B^T) = \det\ (A^T B) = \det\ (AB^T) = \det\ (AB) = (\det\ A)(\det\ B).$$

Proof. By the Cauchy theorem and Theorem 3.2, we have

$$\det\ (A^T B^T) = (\det\ A^T)(\det\ B^T) = (\det\ A)(\det\ B),$$

$$\det\ (A^T B) = (\det\ A^T)(\det\ B) = (\det\ A)(\det\ B),$$

$$\det\ (AB^T) = (\det\ A)(\det\ B^T) = (\det\ A)(\det\ B).$$

■

In other words: *the determinant of the product of two square matrices has its value independent of multiplication either rows by columns or columns by columns or columns by rows or rows by rows.*

A square matrix A is said to be a *non-singular matrix* if $\det\ A \neq 0$. If $\det\ A = 0$ then A is called a *singular matrix*.

Theorem 3.4. *Every non-singular square matrix A is invertible and its inverse A^{-1} is determined by the formula*

$$A^{-1} = \frac{1}{\det\ A}[A_{jk}]_{j,k=1,\ldots,n}, \quad \textit{where} \quad A_{jk} = (-1)^{j+k} M_{jk}$$

and the minor determinants M_{jk} are determined by Formula (3.5).

Proof. Write $a = \det\ A$. Recall (cf. Exercise 2.9) that to the identity operator there corresponds the matrix $E = [\delta_{jk}]_{j,k=1,\ldots,n}$. We are looking for a matrix $B = [b_{jk}]_{j,k=1,\ldots,n}$ such that $AB = BA = E$. By the Cauchy Theorem,

$$aE = AB = [a_{jk}]_{j,k=1,\ldots n}[b_{jk}]_{j,k=1,\ldots,n} = [c_{lk}]_{l,k=1,\ldots,n},$$

$$\textit{where} \quad c_{lk} = \sum_{j=1}^{n} a_{jk} b_{lj}.$$

Then the following equalities should be satisfied:

$$c_{kk} = \sum_{j=1}^{n} a_{jk} b_{jk} = a, \quad c_{lk} = 0 \quad \textit{for } l \neq k \quad (l, k = 1, \ldots, n).$$

Let $b_{jk} = A_{jk}$. Then, by the Laplace theorem, we obtain

$$c_{kk} = \sum_{j=1}^{n} a_{jk} A_{jk} = a = \det\ A.$$

Furthermore, if $l \neq k$ then

$$c_{lk} = \sum_{j=1}^{n} (-1)^{j+l} a_{jl} M_{jl},$$

where the minor determinant M_{jl} determined by formula (3.5) is obtained by canceling the jth column and the lth row. Hence C_{lk} is a development of a determinant such that instead of the lth row there is the kth row. Hence this determinant has two identical rows, which implies that it is equal zero. We therefore conclude that $c_{lk} = 0$ for $l \neq k$. A similar proof shows that $BA = (\det\ A)E$. Therefore we have shown that $A^{-1} = \frac{1}{\det\ A} B$. ■

Let either $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. Let $X = \mathbb{F}^n$ and let $A = [a_{jk}]_{j,k=1,\ldots,n} \in L_0(X)$. Consider a system of linear equations

$$\sum_{j=1}^{n} a_{jk} x_j = y_k, \quad \textit{where} \quad y_k \in \mathbb{F} \quad (k = 1, ..., n). \tag{3.11}$$

This system can be rewritten in an equivalent form:

$$Ax = y, \quad \textit{where} \quad \begin{cases} x = (x_1, ..., x_n) \in X, \\ y = (y_1, ..., y_n) \in X. \end{cases} \tag{3.11$'$}$$

The system (3.11′) with $\det\ A \neq 0$ is said to be *Cramer system of linear equations.*

Theorem 3.5 (Cramer Formulae). *A Cramer system of linear equations* (3.11′) *has a unique solution of the form:*

$$x = (x_1, ..., x_n), \quad \textit{where}$$

$$x_k = \frac{1}{\det\ A} \sum_{j=1}^{n} A_{jk} y_j = \frac{1}{\det\ A} \det\ A_k = \tag{3.12}$$

$$= \frac{1}{\det\ A} \sum_{j=1}^{n} (-1)^{j+k} y_j M_{jk},$$

where the matrix A_j is obtained from the matrix A by putting instead of its jth column the column $y = (y_1, ..., y_n)$ and A_{jk} are determined in Theorem 3.4.

Proof. Since $\det\ A \neq 0$, Theorems 3.4 and 2.4 together imply that there exists the inverse operator A^{-1}. Hence

$$(x_1, ..., x_n) = x = A^{-1}Ax = A^{-1}y = \frac{1}{\det\ A}[a_{jk}]_{j,k=1,\ldots,n}(y_1, ..., y_n) =$$

$$= \frac{1}{\det\ A}\Big(\sum_{j=1}^{n} A_{j1} y_j, ..., \sum_{j=1}^{n} A_{jn} y_j\Big) = \frac{1}{\det\ A}(\det\ A_1, ..., \det\ A_n).$$

By a comparison of the corresponding components, we find Formula (3.12). ■

A system (3.11′) of linear equations is said to be *homogeneous* if $y_1 = ... = y_n = 0$.

Corollary 3.2. *A homogeneous Cramer system of linear equations*

$$Ax = 0, \quad where \quad X = \mathbb{F}^n, \quad A \in L_0(X), \tag{3.13}$$

has only zero as a solution: $x = 0$, *i.e.* $x_1 = ... = x_n = 0$.

Corollary 3.3. *If a homogeneous system (3.13) of linear equations has a solution* $x \neq 0$ *then* det $A = 0$, *i.e. the system (3.13) is not a Cramer system.*

Proof by a reduction to a contradiction. Indeed, suppose that det $A \neq 0$. Then there exists the inverse operator A^{-1}. Acting on both sides of the system (3.13) by this operator, by Corollary 3.2, we obtain $x = 0$. This contradicts our assumption that $x \neq 0$. Hence det $A = 0$. ■

Cramer Formulae and Corollaries 3.2 and 3.3 concern the case when the operator A maps an n-dimensional space onto itself, i.e. a case such that the number of unknowns is equal to the number of equations. Now we shall examine the case when these numbers are not equal. In order to do that, we have to consider non-square matrices, i.e. so called *rectangular matrices.*

Minor determinants of a rectangular matrix A are said to be determinants of square matrices obtained from A by canceling of some numbers of its columns and rows. The *rank* of a rectangular matrix A is said to be the maximal dimension of its minor determinants different than zero. The rank of a rectangular matrix A will be denoted by $r(A)$. By definition and Theorem 3.3, we conclude that

$$r(A^T) = r(A). \tag{3.14}$$

Formula (3.14) and properties of determinants together imply

Corollary 3.4. *The rank of a rectangular matrix does not change if*

(i) we multiply rows (columns) of this matrix by numbers different than zero;

(ii) we exchange the place of rows (columns);

(iii) we add to one row (column) linear combinations of other rows (columns).

Theorem 3.6. *Let* $A = [a_{jk}]_{\substack{j=1,\dots,n \\ k=1,\dots,m}}$. *Denote columns and rows of A as follow:*

$$a_j = (a_{j1}, ..., a_{jm}), \quad (a^k = (a_{1k}, ..., a_{nk})) \quad (j = 1, ..., n;\ k = 1, ..., m).$$

Then the rank of the matrix A is equal to the dimension of the linear span of vectors $a_1, ..., a_n$ $(a^1, ..., a^m$, *respectively), i.e.*

$$r(A) = \dim \operatorname{lin} \{a_1, ..., a_n\} = \dim \operatorname{lin} \{a^1, ..., a^m\}. \tag{3.15}$$

Proof. If some rows (columns) of the matrix under consideration are linearly dependent then one of these rows (columns) should be a linear combination of other rows (columns). Hence the corresponding minor determinant is equal to zero. Conversely, if some rows (columns) are linearly independent then the corresponding minor determinant is different than zero. ■

Let, as before, either $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. Let $X = \mathbb{F}^n$, $Y = \mathbb{F}^m$ and let $A = [a_{jk}]_{\substack{j=1,\dots,n \\ k=1,\dots,m}} \in L_0(X \to Y)$. Consider a system of m linear equations with n unknowns:

$$\sum_{j=1}^{n} a_{jk}x_j = y_k, \quad \textit{where} \quad y_k \in \mathbb{F} \quad (k = 1, ..., m), \tag{3.16}$$

which can be written in an equivalent form

$$Ax = y, \quad \textit{where} \quad x = (x_1, ..., x_n) \in X, \; y = (y_1, ..., y_m) \in Y. \tag{3.17}$$

Every system $x = (x_1, ..., x_n)$ satisfying the system (3.16) of linear equations is said to be a *solution* either of the system (3.16) of linear equations or Equation (3.17). A system (3.16) of linear equations (Equation (3.17)) may have no solutions. If this is the case, then it is said to be *contradictory*. In particular, if $m > n$ then the operator A maps the space X into an n-dimensional space $Y_n \subset Y$. Hence Equation (3.17) has no solutions whenever $y \notin Y_n$, i.e. it is contradictory.

Denote by B the so called *extended matrix* obtained from the matrix A if we join to A a column $(y_1, ..., y_m)$, i.e. the matrix

$$B = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} & y_1 \\ a_{12} & a_{22} & \dots & a_{n2} & y_2 \\ \dots & \dots & \dots & \dots & \dots \\ a_{1m} & a_{2m} & \dots & a_{nm} & y_m \end{pmatrix}. \tag{3.18}$$

Theorem 3.7 (Kronecker-Capella Theorem). *The system (3.16) of linear equations (Equation (3.17)) has a solution $x = (x_1, ..., x_n)$ if and only if the rank of the extended matrix B* (cf. Formula (3.18)) *is equal to the rank of the matrix A, i.e. if $r(B) = r(A)$.*

Proof. Suppose that $x = (x_1, ..., x_n)$ is a solution of the system (3.16) of linear equations (Equation (3.17)), i.e.

$$y_k = \sum_{j=1}^{n} a_{jk}x_j, \qquad (k = 1, ..., m).$$

Then the last column in the extended matrix defined by Formula (3.18) is a linear combination of columns of the matrix A with known coefficients $x_1, ..., x_n$. Hence this column cannot increase the rank of the matrix, i.e. $r(B) = r(A)$.

Conversely, suppose that $r(B) = r(A)$. This means that by joining of the column $y = (y_1, ..., y_m)$ of free terms to the matrix A the maximal dimension of a determinant different than zero is not changed. In other words, if $r(A) = r$ then, even with the joined column of free terms, every determinant of dimension greater than r is equal to zero. Hence the column $y = (y_1, ..., y_m)$ is a linear combination of columns $a_j = (a_{j1}, ..., a_{jm})$, $(j = 1, ..., n)$ of the matrix A, i.e. there exist numbers $x_1, ..., x_n$ such that

$$y_k = \sum_{j=1}^{n} a_{jk}x_j, \qquad (k = 1, ..., m).$$

By definition of numbers $x_1, ..., x_n$, we conclude that $x = (x_1, ..., x_n)$ is a solution of the system (3.16) of linear equations (Equation (3.17)). ■

Corollary 3.5. *Suppose that the system (3.16) of linear equations (Equation (3.17)) has a solution and $r(A) = r$. Denote by M minor determinant of the matrix A of dimension r different than zero. If we cancel in the system (3.16) these equations whose coefficients do not appear in the minor determinant M, then we obtain an equivalent system of linear equations (i.e. every solution of the obtained system of linear equations is a solution to the system (3.16), and conversely).*

Proof. By Theorem 3.6, if $r(A) = r$ then all columns (rows) of the matrix A, which do not appear in the minor determinant M, are linearly dependent on the remains ones. By our assumption, the system (3.16) of linear equations (Equation (3.17)) has solutions. This, and Kronecker-Capella Theorem (Theorem 3.7) together imply that the rank of the extended matrix $r(B) = r(A) = r$. Then free terms y_k corresponding to the canceled equations are also linearly dependent on the remained free terms. Even more, the corresponding linear combinations have the same coefficients. Hence every solution of the system (3.16) of linear equations is a solution of the system reduced to r linear equations corresponding to the minor determinant M, and conversely, every solution of the reduced system is a solution to the system (3.16). ■

Immediate consequences of Corollary 3.5 are

Corollary 3.6. *If $r(A) = m \leq n$ (i.e. the rank of the matrix A is less or equal to the number of equations) then all solutions to the system (3.16) of linear equations (Equation (3.17)) are obtained by canceling in the system (3.16) those equations whose coefficients do not appear in the minor determinant M, where M is any minor determinant of dimension $r(A) = m$ different than zero. We admit for $n-m$ unknowns corresponding to the canceled columns arbitrary numbers, and we calculate the remained unknowns by means of Cramer Formulae applied to the minor determinant M.*

Corollary 3.7. *If $r(A) = m \leq n$ (i.e. the rank of the matrix A is less or equal to the number of equations), then solutions to the system (3.16) of linear equations (Equation (3.17)) depend on $n - m = n - r(A)$ arbitrary numbers called* parameters.

Corollary 3.8. *A system (3.16) of linear equations (Equation (3.17)) has a unique solution if and only if $r(A) = r(B) = n$.*

Proof. By Kronecker-Capella Theorem, the condition $r(A) = r(B)$ is a necessary and sufficient condition of the existence of solutions to the system (3.16). By our assumption that $r(A) = r(B) = n$, it follows that $m = n = r(A)$. Hence, by Corollary 3.6, we find $n - m = 0$ and the system (3.16) is a Cramer system of linear equations. We therefore conclude that this system has a unique solution.

Suppose that $r(A) = n - p$, where $p \geq 1$. Since $r(A) < n$, Corollary 3.7 implies that all solutions to the system (3.16) of linear equations depend on $n-(n-p) = p \geq 1$ parameters. Hence they are not unique. ■

Consider now a homogeneous system corresponding to the system (3.16) of linear equations, i.e. the system

$$\sum_{j=1}^{n} a_{jk}x_j = 0 \qquad (k = 1, ..., m). \tag{3.19}$$

Theorem 3.8. *A necessary and sufficient condition for a homogeneous system (3.19) of linear equations to have a unique solution is that $r(A) = n$.*

Proof. Since the column $y = (y_1, ..., y_n)$ of free terms consists of zeros only, we have $r(A) = r(B)$. By our assumption, $r(B) = r(A) = n$. By Corollary 3.8, the condition $r(A) = n$ is a necessary and sufficient condition for the system (3.19) to have a unique solution. ■

Theorem 3.9. *If $x = (x_1, ..., x_n)$ is a non-zero solution to a system (3.19) of linear equations and $r(A) = m = n - 1$ then*

$$\frac{x_1}{M_1} = -\frac{x_2}{M_2} = ... = \frac{x_n}{(-1)^{n+1}M_n}, \tag{3.20}$$

where M_k is the minor determinant of the matrix A obtained by canceling the kth column (if $M_k = 0$ for a k then we let $x_k = 0$).

Proof. By the assumption that $r(A) = m = n - 1$, there is at least one minor determinant $M_k \neq 0$ of dimension $n - 1 = m$. This assumption and Corollary 3.6 together imply that solutions to the system (3.19) depend on one parameter for $n - m - (n - 1) = 1$ and that solutions to the system (3.19) are of the form:

$$x_k = a(-1)^{k+1}M_k \quad (k = 1, ..., m = n - 1), \quad \textit{with } a \neq 0.$$

Hence Formula (3.20) holds.

These solutions are obtained in the following way: instead of the unknown x_j corresponding to a canceled column of the matrix A we admit an arbitrary number $a \neq 0$, and we applied to the obtained system of linear equations (already non-homogeneous) Cramer Formulae. ■

Theorem 3.10. *If $A = [a_{jk}]_{\substack{j=1,...,n \\ k=1,...,m}}$ and*

$$\ker A = \{x : Ax = 0\} = \{(x_1, ..., x_n) : \sum_{j=1}^{n} a_{jk}x_j = 0 \quad \textit{for } k = 1, ..., m\}$$

then

$$\text{dim ker } A = n - r(A). \tag{3.21}$$

Proof. The nullity $\alpha_A = \text{dim ker } A$ is the number of linearly independent non-zero solutions to the system (3.19) of linear equations. Let $r(A) = r$. Denote by M a minor determinant of the matrix A of dimension r which is different than zero. Cancelling in the system (3.19) those equations whose coefficients do not appear in the minor determinant M, we obtain by Theorem 3.5, a system equivalent to the system (3.19). However, the system of r linear equations just obtained has $n \geq r$ unknowns. On the other hand, $n - r$ columns do not appear in M. Admitting arbitrary numbers for the corresponding unknowns, we obtain a new non-homogeneous system of r linear equations with r unknowns. Solving this system by means of Cramer Formulae we obtain a non-zero solution of the system (3.19) which depends on $n - r = n - r(A)$ parameters. In other words, the number of linearly independent solutions of the system (3.19) is $\alpha_A = n - r(A)$. ■

At the end of this chapter we shall give other conditions of solvability of Equation (3.17) (cf. also Chapter 9).

Theorem 3.11. *Suppose that $A \in L_0(X \to Y)$, dim $X = n$, dim $Y = m$. Then $A^T \in L_0(Y \to X)$, hence $A^T A \in L_0(X \to X)$, i.e. $A^T A$ is a square matrix. Then the following conditions are satisfied:*

(i) If det $(A^T A) \neq 0$*, then the equation $Ax = y$, $y \in Y$, has a unique solution $x = (A^T A)^{-1} A^T y$, i.e. the homogeneous equation $Ax = 0$ has only zero as a solution.*

(ii) If there is an $x \neq 0$ such that $Ax = 0$, then det $(A^T A) = 0$.

(iii) If $y \neq 0$, $A^T y = 0$ and det $(A^T A) \neq 0$*, then the equation $Ax = y$ has not solutions.*

Proof. Acting on both sides of the equation $Ax = y$ by means of the operator A^T, we obtain the equation

$$A^T A x = A^T y. \tag{3.22}$$

(i) Since det $(A^T A) \neq 0$ by our assumption, the square matrix $A^T A$ is invertible and $x = (A^T A)^{-1} A^T y$. If $y = 0$ then $x = 0$.

(ii) If there an $x \neq 0$ such that $Ax = 0$, then $A^T A x = 0$. This, and Corollary 3.3, together imply that det $(A^T A) = 0$.

(iii) If $y \neq 0$, then a solution x of the equation $Ax = y$ (provided that it exists) is different than zero. Suppose that $A^T y = 0$. By Equation (3.22), we conclude that $A^T A x = 0$. But det $(A^T A) \neq 0$. This, and Point (i) of this Theorem together imply $x = 0$. This contradicts our assumption that $x \neq 0$. Hence the equation $Ax = y$ has no solutions.

■

Exercises.

Exercise 3.1. Prove that the determinant of a diagonal matrix (cf. Example 2.11) is equal to the product of terms on the principal diagonal. $\square$

Exercise 3.2. Prove that the determinant of a square matrix $[a_{jk}]_{j,k}$, where $a_{jk} = 0$ for $j > k$, is equal to the product of terms on the principal diagonal. $\square$

Exercise 3.3. Prove that the determinant of a square matrix $[a_{jk}]_{j,k}$, where $a_{jk} = 0$ for $j < k$, is equal to the product of terms on the principal diagonal. $\square$

Exercise 3.4. Determine the value of the *Vandermonde determinant*

$$V_n = \det\, [t_j^{k-1}]_{j,k=1,\dots,n} \qquad (n \in \mathbb{N};\ t \in \mathbb{R}).$$

Prove that $V_n \neq 0$ whenever $t_i \neq t_j$ for $i \neq j$. $\square$

Exercise 3.5. Prove that $B^T = B$ whenever $B = AA^T$. $\square$

Exercise 3.6. Prove that a *cyclic determinant*

$$D_n = \begin{vmatrix} a_0 & a_1 & a_2 & \dots & a_{n-1} \\ a_{n-1} & a_0 & a_1 & \dots & a_{n-2} \\ \dots & \dots & \dots & \dots & \dots \\ a_1 & a_2 & a_3 & \dots & a_0 \end{vmatrix} = \phi(\epsilon)\phi(\epsilon^2)\dots\phi(\epsilon^n) \qquad (n \in \mathbb{N}),$$

where $\phi(t) = a_0 + a_1 t + \dots + a_n t^n$, $\epsilon = \mathrm{e}^{2\pi i/n}$ (Hint: multiply rows of the determinant D_n by columns of the Vandermonde determinant considered in Exercise 3.4). $\square$

Exercise 3.7. Let $\dim X = n$ and let $A \in L_0(X)$. By a *change of the coordinates* we mean a one-to-one mapping B of the basis in the space X into another basis. This definition implies that a linear extension $\tilde{B}$ of the mapping B onto the whole space X is invertible and that $\tilde{B}$, $\tilde{B}^{-1} \in L_0(X)$. Prove that a change B of the coordinates transforms the operator A into the operator $\tilde{B}^{-1}A\tilde{B}$. Give examples. $\square$

Exercise 3.8. Prove that homogeneous polynomials $1, t, \dots, t^n$ $n \in \mathbb{N})$ are linearly independent (Hint: Use Exercise 3.4). $\square$

Exercise 3.9. Prove that in the space $\mathbb{F}_n[t]$ determined in Example 1.5, homogeneous polynomials $1, t, \dots, t^n$ form a basis, and that $\dim \mathbb{F}_n[t] = n + 1$ $(n \in \mathbb{N})$. $\square$

Exercise 3.10. Prove that the space $\mathbb{F}[t]$ determined in Example 1.6 is infinite dimensional. $\square$

Exercise 3.11. Prove that vectors $(\delta_{k1}, \dots, \delta_{kn})$ $(k = 1, \dots, n \in \mathbb{N})$ form a basis in the space $\mathbb{R}^n$ and that $\dim \mathbb{R}^n = n$. $\square$

Exercise 3.12. There are given k hyperplanes in $\mathbb{R}^n$. When do these hyperplanes cut each another? Consider the cases: $k < n$, $k = n$, $k > n$. Give a geometric interpretation for $n = 2$ and $n = 3$. $\square$

Exercise 3.13. Determine kernels of matrices

$$\begin{pmatrix} 1 & 0 & -1 \\ -1 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 1 \\ 2 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix}.$$

□

Exercise 3.14. Give conditions of solvability of the equation $Ax = y$, where

$$(i) \quad A = \begin{pmatrix} a & b & c \\ c & a & b \end{pmatrix}; \qquad (ii) \quad A = \begin{pmatrix} a & 0 \\ 0 & t \\ a & b \end{pmatrix}.$$

(Hint: apply Theorem 3.7). □

Exercise 3.15. Suppose that all conditions of Theorem 3.11 are satisfied. Prove that the equation $Ax = y$ has a unique solution $x = A^T(AA^T)^{-1}y$ whenever det $(AA^T) \neq 0$ (Hint: substitute $x = A^T y$). Give a condition of solvability in the case det $(AA^T) = 0$. Explain why Theorem 3.6 is more useful when $n < m$, while results obtained in this exercise are nonuseful in the case $m < n$. □

Exercise 3.16. Prove that

(i) the set of all finite dimensional linear operators acting in a linear space X (either over the field $\mathbb{R}$ or over the field $\mathbb{C}$) is an ideal in the algebra $L_0(X)$;

(ii) this ideal is proper if and only if dim $X = +\infty$. □

Exercise 3.17. Prove that all considerations of this chapter hold for matrices

$$A = [a_{jk}]_{\substack{j=1,\ldots,n \\ k=1,\ldots,m}},$$

where a_{jk} belong to a commutative algebra (either over $\mathbb{R}$ or over $\mathbb{C}$). □

Chapter 4.

Inner product and vector product.

We say that in a linear space X over the field $\mathbb{C}$ of complex numbers a *inner product* is defined if there is a function $\langle x, y\rangle$ defined on the set $X \times X$, with values $\in \mathbb{C}$ and such that the following conditions are satisfied for all $a \in \mathbb{C}$, $x, y, x_1, x_2, y_1, y_2 \in X$:

(i) $\langle x_1 + x_2, y\rangle = \langle x_1, y\rangle + \langle x_2, y\rangle$,

(ii) $\langle x, y\rangle = \overline{\langle y, x\rangle}$ *),

(iii) $\langle ax, y\rangle = a\langle x, y\rangle$,

(iv) $\langle x, x\rangle > 0$ for $x \neq 0$.

Conditions (i) and (ii) together imply the additivity of an inner product with respect to its second variable:

(i') $\langle x, y_1 + y_2\rangle = \langle x, y_1\rangle + \langle x, y_2\rangle$.

Indeed,

$$\langle x, y_1 + y_2\rangle = \overline{\langle y_1 + y_2, x\rangle} = \overline{\langle y_1, x\rangle} + \overline{\langle y_2, x\rangle} = \langle x, y_1\rangle + \langle x, y_2\rangle.$$

Conditions (ii) and (iii) together imply that

(iii') $\langle x, ay\rangle = \overline{a}\langle x, y\rangle$.

Indeed, $\langle x, ay\rangle = \overline{\langle ay, x\rangle} = \overline{a < y, x >} = \overline{a}\ \overline{\langle y, x\rangle} = \overline{a}\ \langle x, y\rangle$.

Conditions (iii) and (iv) together imply that

(iv') $\langle x, x\rangle = 0$ if and only if $x = 0$.

Indeed, suppose that $x = 0$. Let a be an arbitrary complex number different than zero and than one. Then, by Condition (iii), we find

$$r = \langle 0, 0\rangle = \langle a0, 0\rangle = a\langle 0, 0\rangle = ar.$$

Hence $r = 0$, i.e. $\langle 0, 0\rangle = 0$. Conversely, suppose that $\langle x, x\rangle = 0$. By Condition (iv), if $x \neq 0$ then $\langle x, x\rangle > 0$, a contradiction. Then $x = 0$.

If X is a linear space over the field $\mathbb{R}$ then Condition (ii) in the definition of the inner product admits the form

(ii') $\langle x, y\rangle = \langle y, x\rangle$.

---

*) By $\overline{z}$ we denote the number conjugate with $z \in \mathbb{C}$.

Other conditions remain the same as before.

An infinite dimensional space X with an inner product defined in is said to be a *pre-Hilbert space*. An n-dimensional pre-Hilbert space over the field $\mathbb{R}$ is said to be an *n-dimensional Euclidean space* and it is denoted by $\mathbb{E}^n$.

A *norm* of an element x belonging to a pre-Hilbert space X is the function

$$\|x\| = \sqrt{\langle x, x\rangle}.$$

Observe that, by definition, $\|x\| \geq 0$.

By the *distance* of two points belonging to a pre-Hilbert space is meant the norm of their difference:

$$\|x-y\| = \sqrt{\langle x-y, x-y\rangle}.$$

In other words, *the norm of a point in a pre-Hilbert space is equal to its distance from zero.*

In order to show properties of the distance and the norm in a pre-Hilbert space, first we shall prove the following

Lemma 4.1 (Schwarz inequality). *If X is a pre-Hilbert space then*

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\| \quad for \quad x, y \in X. \tag{4.1}$$

Proof. For an arbitrary $a \in \mathbb{R}$ we have

$$0 \leq \langle x+ay, x+ay\rangle = \langle x, y\rangle + a[\langle x, y\rangle + \langle y, x\rangle] + a^2\langle y, y\rangle +$$

$$+\|x\|^2 + a[\langle x, y\rangle + \langle y, x\rangle] + a^2\|y\|^2.$$

Hence the discriminant of the last trinomial satisfies the inequality

$$\frac{1}{4}[\langle x, y\rangle + \langle y, x\rangle]^2 - \|x\|^2\|y\|^2 \leq 0 \tag{4.1$'$}$$

for arbitrary $x, y \in X$.

It is easy to choose such a number b that $|b| = 1$ and that the inner product $\langle x, by\rangle$ is a real number. Indeed, if $\langle x, y\rangle = r(\cos\alpha + i\sin\alpha)$ then we let $b = \cos\alpha + i\sin\alpha$. Then $|b| = 1$ and $\langle x, by\rangle = \bar{b}\langle x, y\rangle = \bar{b}rb = \bar{b}br = |b|^2 r = r$, where r is a real number. We therefore conclude that b is the number, we were looking for. Let now $y_0 = by$. Inequality (4.1$'$) (satisfied also by elements x, y_0) implies that

$$|\langle x, y\rangle| = \frac{1}{|b|}\ |\langle x, y_0\rangle| = \frac{1}{2}|\langle x, y_0\rangle + \langle y_0, x\rangle| \leq \|x\| \cdot \|y\| =$$

$$= \|x\|\ \|by\| = |b|\ \|x\|\ \|y\| = \|x\|\ \|y\|,$$

which we had to prove. ■

Theorem 4.1. *If X is a pre-Hilbert space then the norm of its elements determined by the inner product has the following properties ($x, y \in X$):*

(a) $\|x\| = 0$ *if and only if* $x = 0$;

(b) $\|\lambda x\| = |\lambda|\ \|x\|$ *for* $\lambda \in \mathbb{F}$ (homogeneity)*);

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

Proof. Equality (a) is an immediate consequence of Condition (iv). Conditions (ii) and (iii) together imply that $\|\lambda x\| = \langle \lambda x, \lambda x\rangle = \lambda\overline{\lambda}\langle x, x\rangle = |\lambda|^2\|x\|^2$. By the Schwarz inequality (Lemma 4.1), we get

$$\|x + y\|^2 = |\langle x + y, x + y\rangle| = |\langle x, x\rangle| + \langle y, y\rangle + \langle x, y\rangle + \langle y, x\rangle| \leq$$

$$\leq \|x\|^2 + 2\|x\| \cdot \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2,$$

i.e. the triangle inequality is satisfied. ∎

An immediate consequence of Theorem 4.1 is

Corollary 4.1. If X is a pre-Hilbert space, then the distance of two points determined by the inner product has the following properties ($x, y, z \in X$):

(α) $\|x - y\| = 0$ *if and only if* $x = 0$;

(β) $\|y - x\| = \|x - y\|$ (symmetry);

(γ) $\|x - y\| \leq \|x - z\| + \|y - z\|$ (triangle inequality).

We say that two elements x and y of a pre-Hilbert space X are *orthogonal* if $\langle x, y\rangle = 0$. Then we write $x \perp y$.

Two subspaces Y and Z of a pre-Hilbert space X are said to be *orthogonal* if $\langle y, z\rangle = 0$ for all $y \in Y$, $z \in Z$. Then we write $Y \perp Z$. If $X = Y \oplus Z$ and $Y \perp Z$ then X is said to be *decomposed onto the orthogonal direct sum* and Z is said to be an *orthogonal complement* of the subspace Y, what we write $Z = Y^\perp$. Respectively, Y is an orthogonal complement of Z. By the definition of an orthogonal complement, we have

$$Y^\perp = \{z \in X : \langle y, z\rangle = 0 \quad for \quad y \in Y\} \quad and \quad (Y^\perp)^\perp = Y.$$

Let Y be a subspace of a pre-Hilbert space X. If $Y^\perp$ is an orthogonal complement of Y then there are two projection operators P_Y and $P_Y^\perp$ such that $Y = P_Y X$, $Y^\perp = P_Y^\perp X$ and $P_Y + P_Y^\perp = I$, $P_Y P_Y^\perp = P_Y^\perp P_Y = 0$ (cf. Theorems 2.1, 2.3 and 2.4).

If Y is a subspace of a pre-Hilbert space X, and $Y^\perp$ is its orthogonal complement, then the *distance of an element $x \in X$ from Y* is the number

$$\mathrm{d}(x, Y) = \|x - P_Y x\|, \tag{4.2}$$

---

*) where, as before, either $\mathbb{F} = \mathbb{C}$ or $\mathbb{F} = \mathbb{R}$.

where P_Y is the projection operator defined as before: $P_Y X = Y$.

In the sequel we shall restrict ourselves to n-dimensional Euclidean spaces $\mathbb{E}^n$. It is easy to verify that a function $\langle x, y\rangle$ defined by means of the formula

$$(4.3) \qquad \langle x, y\rangle = \sum_{j=1}^{n} x_j y_j \quad for \quad x = (x_1, ..., x_n),\ y = (y_1, ..., y_n) \in \mathbb{R}^n$$

is an inner product of elements x and y, i.e. it satisfies Conditions (i), (ii'), (iii), (iv). Then $\mathbb{R}^n$ with the inner product defined by Formula (4.3) is the n-dimensional Euclidean space $\mathbb{E}^n$. We denote the norm of an element $x \in \mathbb{E}^n$ by single lines: $|x|$, i.e.

$$|x| = \Big(\sum_{j=1}^{n} x_j^2\Big)^{\frac{1}{2}} \quad for \quad x \in \mathbb{E}^n.$$

The number $|x|$, where $x \in \mathbb{E}^n$, can be interpreted in two manners, either as the *length of a vector* or the distance of the point x from zero. Hence the number $|x - y|$ denotes either the lenght of a vector which is a difference of vectors x and y or the *distance of points* x and y. By the definition (4.3) of the inner product in $\mathbb{E}^n$, we have

Property 4.1. *Two vectors x and y in $\mathbb{E}^n$ are orthogonal if and only if*

$$(4.4) \qquad \langle x, y\rangle = \sum_{j=1}^{n} x_j y_j = 0 \qquad (x, y \in \mathbb{E}^n).$$

A vector $x \in \mathbb{E}^n$ is said to be *normed* if $|x| = 1$. If $x = (x_1, ..., x_n)$ is a normed vector then, by definition,

$$|x|^2 = \sum_{j=1}^{n} x_j^2 = 1 \quad for \quad x \in \mathbb{E}^n.$$

Coordinates $x_1, ..., x_n$ of a normed vector are called *directional cosines* of that vector. The sum of their squares is then equal to one.

A basis $\{e_1, ..., e_n\}$ of the space $\mathbb{E}^n$ is said to be *orthonormal* if $\langle e_i, e_k\rangle = \delta_{ik}$ $(i, k = 1, ..., n)$, where δ_{ik} is the Kronecker symbol *). Clearly, if this is the case, then $|e_i| = 1$. Hence an orthonormal basis consists of normed orthogonal vectors.

Property 4.2. *The Euclidean space $\mathbb{E}^n$ has an orthonormal basis which consists of vectors $e_k = (\delta_{k1}, ..., \delta_{kn})$ $(k = 1, ..., n)$.*

Proof. The system $\{e_1, ..., e_n\}$ constitutes a basis in the space $\mathbb{R}^n$ (cf. Exercise (3.11), hence also in the space $\mathbb{E}^n$. By definition, for $i, k = 1, ..., n$

$$\langle e_i, e_k\rangle = \sum_{j=1}^{N} \delta_{ij}\delta_{ki} = \delta_{ki} = \begin{cases} 1 & \text{for } k = i, \\ 0 & \text{for } k \neq i, \end{cases}$$

*) cf. Exercise 3.11.

which we had to prove. ■

Directional cosines of a normed vector in the space $\mathbb{E}^n$ are coefficients of the expansion of that vector with respect to the basis $\{e_1, ..., e_n\}$ determined in Property 4.2.

Property 4.3. *Every subspace Y of the Euclidean space $\mathbb{E}^n$ has an orthogonal complement $Y^\perp$. Moreover, if $\{e_{k_1}, ..., e_{k_m}\}$, $m < n$ is an orthonormal basis in Y then*

$$Y^\perp = \text{lin}\ \{e_k : e_k \notin \text{lin}\ \{e_{k_1}, ..., e_{k_m}\}\}. \tag{4.5}$$

Proof. Let $y \in Y$ and $e_k \notin \{e_{k_1}, ..., e_{k_m}\}$ be arbitrary (as before, e_j denote elements of an orthonormal basis in $\mathbb{E}^n$). Then $y = \sum_{j=1}^{n} a_j e_{k_j}$, where $a_j \in \mathbb{R}$. This, and Corollary 4.2 together imply that

$$\langle y, e_k \rangle = \Big\langle \sum_{j=1}^{n} a_j e_{k_j}, e_k \Big\rangle = \sum_{j=1}^{n} a_j \langle e_{k_j}, e_k \rangle = 0,$$

which proves that $e_k \perp Y$. Hence the subspace $Y^\perp$ defined by Formula (4.5) is the orthogonal complement of the subspace Y. ■

Property 4.4. *Every linear functional defined on the Euclidean space $\mathbb{E}^n$ is of the form $f(x) = \langle x, a \rangle$, where $a \in \mathbb{E}^n$ is a fixed element. Conversely, if $f(x) = \langle x, a \rangle$ then f is a linear functional over the space $\mathbb{E}^n$.*

Proof. By Property 4.2, if x is an arbitrary element of the space $\mathbb{E}^n$ then $x = \sum_{k=1}^{n} t_k e_k = (t_1, ..., t_n)$, where $t_k \in \mathbb{R}$. By Formula (3,3), every linear functional over the space $\mathbb{E}^n$ can be written in the form $f(x) = \sum_{j=1}^{n} t_j a_j$, where $a_j \in \mathbb{R}$. Write $a = (a_1, ..., a_n)$. Then we have $a \in \mathbb{E}^n$ and

$$f(x) = \sum_{j=1}^{n} t_j a_j = \langle (t_1, ..., t_n), (a_1, ..., a_n) \rangle = \langle x, a \rangle,$$

which we had to prove. On the other hand, if $f(x) = \langle x, a \rangle$ then f is a linear mapping of $\mathbb{E}^n$ into $\mathbb{R}$, hence it is a linear functional. ■

An immediate consequence of Theorem 2.7 and Property 4.4 is

Property 4.5. *Hyperplanes in the Euclidean space $\mathbb{E}^n$ not passing through zero are of the form*

$$H_a = \{x \in \mathbb{E}^n : \langle x, a \rangle = 1;\ a\ is\ a\ fixed\ element\ of\ \mathbb{E}^n\}. \tag{4.6}$$

If a hyperplane H_a passing through a point $x_0 \in \mathbb{E}^n$ satisfies the equation $\langle x, a\rangle = 1$ then it also satisfies an equation of the form:

$$\langle x - x_0, a\rangle = 0. \tag{4.7}$$

Indeed, if this is the case then $\langle x, a\rangle = 1$, $\langle x_0, a\rangle = 1$. Subtracting these equations by sides, we obtain Formula (4.7).

Note that every hyperplane in $\mathbb{E}^n$ passing through the point $x_0 = 0$ has the equation of the form:

$$\langle x, a\rangle = 0. \tag{4.8}$$

Hence an equation of an arbitrary hyperplane can be written in the form:

$$\langle x, a\rangle = c. \tag{4.9}$$

Property 4.6. *The distance of a point $y \in \mathbb{E}^n$ from a hyperplane $H_a \subset \mathbb{E}^n$ defined by means of the equation $\langle x, a\rangle = c$ is given by the following formula:*

$$\mathrm{d}\ (y, H_a) = \frac{1}{|a|}\ |\langle y, a\rangle - c|. \tag{4.10}$$

Proof. Formula (4.2) implies that $\mathrm{d}(y, H_a) = |y - Py|$, where by P is denoted the projection operator onto the hyperplane H_a. Write $y' = Py$. By the definition of a projection, it follows that $y - y' = \lambda a$ for a $\lambda \neq 0$. Hence $\mathrm{d}(y, H_a) = |y - y'| = |\lambda| \cdot |a|$. On the other hand, since $y' \in H_a$, we get $\langle y', a\rangle = c$ and

$$\langle y, a\rangle = \langle y, a\rangle + \langle y', a\rangle - \langle y', a\rangle = \langle y - y', a\rangle + c =$$

$$= \langle \lambda a, a\rangle + c = \lambda\langle a, a\rangle = \lambda\ |a|^2 + c.$$

Then

$$\lambda = \frac{\langle y, a\rangle - c}{|a|^2}, \qquad |\lambda| = \frac{|\langle y, a\rangle - c|}{|a|^2}.$$

Finally,

$$\mathrm{d}\ (y, H_a) = |\lambda|\ |a| = \frac{|\langle y, a\rangle - c|}{|a|}.$$

■

Property 4.1. *If points $y_j = (y_{j1}, ..., y_{jn}) \in \mathbb{E}^n$ $(j = 1, ...n)$ are linearly independent then they uniquely determine a hyperplane H defined by means of the equation*

$$H = \{x \in \mathbb{E}^n : \det\ \Delta_n = 0\} \quad \textit{where} \quad \Delta_n = \begin{pmatrix} x_1 & ... & x_n & 1 \\ y_{11} & ... & y_{1n} & 1 \\ y_{21} & ... & y_{2n} & 1 \\ ... & ... & ... & 1 \\ y_{n1} & ... & y_{nn} & 1 \end{pmatrix}. \tag{4.11}$$

Proof. All points of a hyperplane H satisfy the equation

$$\langle x, a\rangle = 1. \tag{4.12}$$

Since H passes through the points $y_1, ..., y_n$, we conclude that the points $y_1, ..., y_n$ satisfy the equations

$$\langle y_j, a\rangle = 1, \quad (j = 1, ..., n), \tag{4.12$'$}$$

where $x = (x_1, ..., x_n)$ and the vector $a = (a_1, ..., a_n) \in \mathbb{E}^n$ is fixed. Equations (4.12′) can be rewritten as a system of linear equations

$$\sum_{k=1}^{n} x_k a_k = 1, \tag{4.13}$$

$$\sum_{k=1}^{n} y_{jk} = 1 \qquad (j = 1, ..., n). \tag{4.14}$$

This is a system of $n+1$ non-homogeneous linear equations with n unknowns $a_1, ..., a_n$. We therefore conclude that the system (4.13)-(4.14) has a unique solution if $r(B) = r(\Delta_n) = n$, where

$$B = \begin{pmatrix} x_1 & ... & x_n \\ y_{11} & ... & y_{1n} \\ ... & ... & ... \\ y_{n1} & ... & y_{nn} \end{pmatrix}.$$

However, the linear independence of vectors y_j implies that the determinant of the square matrix $A = [y_{jk}]_{j,k=1,...,n}$ is different than zero (since it has linearly independent columns). Hence the system (4.14) is a Cramer system and, by Theorem 3.5, has a unique solution of the form

$$a = (a_1, ..., a_n) = \frac{1}{\det\ A}\ (\det\ A_1, ..., \det\ A_n), \tag{4.15}$$

where the matrix A_k is obtained from the matrix A by putting instead of its kth column the column $(1, 1, ..., 1)$. Since $\det\ A \neq 0$, we conclude that $r(\Delta_n) \geq r(B) = r(A) = n$. Still we have to prove that $\det\ \Delta_n = 0$. Develop $\det\ \Delta_n$ with respect to its first row. Then we obtain

$$\det\ \Delta_n = \sum_{k=1}^{n} (-1)^{k+1} x_k (-1)^{n-1} \det\ A_k + (-1)^{n+2} \det\ A =$$

$$= (-1)^{n+1} \sum_{k=1}^{n} x_k \det\ A_k + (-1)^{n+2} \det\ A = (-1)^{n+1} \det\ A + (-1)^{n+2} \det\ A = 0.$$

For, by Formulae (4.13) and (4.15), we find

$$\sum_{k=1}^{n} x_k \det\ A_k = \sum_{k=1}^{n} x_k a_k \det\ A = \det\ A \sum_{k=1}^{n} x_k a_k = \det\ A.$$

Hence $r(A) = n = r(B)$ and the system (4.13)-(4.14) has a unique solution (4.15). ■

Suppose that we are given two non-zero vectors $x, y \in \mathbb{E}^n$. Then the modulus of the number

$$c = \frac{\langle x, y\rangle}{|x| \cdot |y|}$$

is not greater than 1 (by the Schwarz inequality (4.1)). Hence there is an angle θ such that $0 \geq \theta \geq \pi$ and $c = \cos\ \theta$. The angle θ is said to be the *angle* between vectors x and y. An angle θ between non-zero vectors x and y belonging to $\mathbb{E}^n$ is then defined by the formula

$$\cos\ \theta = \frac{\langle x, y\rangle}{|x| \cdot |y|}. \tag{4.16}$$

Two non-zero vectors x, y in the Euclidean space $\mathbb{E}^n$ are said to be *orthogonal* (otherwise: *perpendicular* one to each another) if $\langle x, y\rangle = 0$. If this is the case, then $\cos\ \theta = 0$, i.e. $\theta = \frac{\pi}{2}$.

Two non-zero vectors $x, y \in \mathbb{E}^n$ are said to be *parallel* if there is a real number $\lambda \neq 0$ such that $y = \lambda x$, since in that case

$$\cos\ \theta = \frac{\langle x, y\rangle}{|x| \cdot |y|} = \frac{\langle x, \lambda x\rangle}{|x| \cdot |\lambda x|} = \frac{\lambda\langle x, x\rangle}{|\lambda|\ |x|^2} = \frac{\lambda}{|\lambda|} = \pm\ 1,$$

i.e. either $\theta = 0$ or $\theta = \pi$. If this is the case, then we write $x \parallel y$.

Similarly, a *cosine of an angle* between two hyperplanes H_a and H_b in the Euclidean space $\mathbb{E}^n$ is defined by means of the formula

$$\cos\ \theta = \frac{\langle a, b\rangle}{|a| \cdot |b|}. \tag{4.17}$$

Two hyperplanes H_a and H_b in the Euclidean space $\mathbb{E}^n$ are *orthogonal* (otherwise: *perpendicular*) if

$$\langle a, b\rangle = 0. \tag{4.18}$$

Indeed,if this is the case, then $\cos\ \theta = 0$, hence $\theta = \frac{\pi}{2}$.

Two hyperplanes H_a and H_b in the Euclidean space $\mathbb{E}^n$ are *parallel* if there is a real number $\lambda \neq 0$ such that $b = \lambda a$.

Indeed, if it the case, then $\cos\ \theta = \pm\ 1$, i.e. either $\theta = 0$ or $\theta = \pi$.

Let $\{e_1, ..., e_n\}$ be an orthonormal basis in the Euclidean space $\mathbb{E}^n$. An ordered system $\{x_1, ..., x_n\}$ of linearly independent vectors $x_j = (x_{j1}, ..., x_{jn})$ has an *orientation compatible with the basis* $\{e_1, ..., e_n\}$, in other words, a *positive orientation* with respect to that basis if

$$\Theta(x_1, ..., x_n) = \det\, [x_{jk}]_{j,k=1,...,n} \ > 0.$$

In the opposite case, i.e. when $\Theta(x_1, ..., x_n) < 0$, this system has a *negative orientation*. Otherwise, these system are called *dextrally oriented* and *sinistrally oriented* systems. (or, in the space $\mathbb{E}^2$, *clockwise oriented* and *anticlockwise oriented*).

Suppose that we are given two linearly independent vectors (i.e. vectors which are non-parallel) $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$ in the Euclidean space $\mathbb{E}^3$. We are looking for a vector $z = (z_1, z_2, z_3)$ orthogonal to x and y and such that the triple $\{x, y, z\}$ of vectors is positively oriented. The vector in question is constructed in the following manner: if $x, y \in \mathbb{E}^3$ then $z = (z_1, z_2, z_3)$, where

$$z_1 = \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix}, \quad z_2 = \begin{vmatrix} x_3 & x_1 \\ y_3 & y_1 \end{vmatrix}, \quad z_3 = \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix}. \tag{4.19}$$

This vector will be denoted by $x \times y$ and called the *vector product* of x and y.

Observe that coordinates z_1, z_2, z_3 of a vector product are minor determinants of the determinant

$$\begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}$$

obtained by its development with respect to the first row. We will show that the vector z is orthogonal to x and y. Indeed,

$$\langle x, z\rangle = x_1z_1 + x_2z_2 + x_3z_3 = x_1 \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} + x_2 \begin{vmatrix} x_3 & x_1 \\ y_3 & y_1 \end{vmatrix} + x_3 \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} =$$

$$= x_1 \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} - x_2 \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} + x_3 \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} = \begin{vmatrix} x_1 & x_2 & x_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} = 0,$$

for the last determinant has two identical rows. A similar proof shows that $\langle y, z\rangle = 0$.

We still have to prove that the triple $\{x, y, z\}$ is positively oriented. Develop the determinant $\Theta(x, y, z)$ with respect to the last row. Then we get

$$\Theta(x, y, z) = \begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} & \begin{vmatrix} x_3 & x_1 \\ y_3 & y_1 \end{vmatrix} & \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \end{vmatrix} =$$

$$= \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix}^2 - \begin{vmatrix} x_3 & x_1 \\ y_3 & y_1 \end{vmatrix} \cdot \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} + \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix}^2 =$$

$$= (x_2y_3 - y_2x_3)^2 + (x_3y_1 - y_3x_1)^2 + (x_1y_2 - y_1x_2)^2 > 0.$$

Hence the vector product has all required properties.

It is not difficult to generalize the above construction in order to form in the Euclidean space $\mathbb{E}^n$ $(n > 3)$ a vector product of $n-1$ linearly independent vectors. Namely, suppose that we are given vectors $x_j = (x_{j1}, ..., x_{jn}) \in \mathbb{E}^n$ $(j = 1, ..., n-1)$ which are linearly independent. Define a vector product in $\mathbb{E}^n$ in the following way:

$$x_n = (x_{n1}, ..., x_{nn}),$$

where

$$x_{nj} = (-1)^{j+1} \det \left[x_{km}\right]_{\substack{k=1,\dots,n-1\\ m=1,\dots,n;\ m\neq j}} \quad for \quad j = 1, ..., n.$$

It is easy to verify that the vector x_n is orthogonal to each of vectors $x_1, ..., x_{n-1}$ and that the n-tuple $\{x_1, ..., x_n\}$ of vectors is positively oriented, i.e. $\det\, [x_{k,m}]_{k,m=1,\dots,n} > 0$.

Property 4.8. *If θ is an angle between two vectors x, $y \in \mathbb{E}^3$ then*

$$|x \times y| = |x|\ |y| \sin\ \theta, \tag{4.20}$$

i.e.

$$\sin\ \theta = \frac{|x \times y|}{|x| \cdot |y|}.$$

Proof. By definition, we have

$$|x \times y|^2 = \det \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix}^2 + \det \begin{vmatrix} x_3 & x_1 \\ y_3 & y_1 \end{vmatrix}^2 + \det \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix}^2 =$$

$$= (x_2y_3 - x_3y_2)^2 + (x_3y_1 - x_1y_3)^2)^2 + (x_1y_2 - x_2y_1)^2 =$$
$$= x_2^2y_3^2 + x_3^2y_2^2 + x_3^2y_1^2 + x_1^2y_3^2 + x_1^2y_2^2 + x_2^2y_1^2 -$$
$$-2(x_2x_3y_2y_3 + x_1x_3y_1y_3 + x_1x_2y_1y_2).$$

On the other hand,

$$|x|^2|y|^2 \sin^2\theta =$$
$$= |x|^2|y|(1 - \cos^2\theta) = \big(1 - \frac{\langle x, y\rangle^2}{|x|^2|y|^2}\big)|x|^2|y|^2 = |x|^2|y|^2 - \langle x, y\rangle^2 =$$
$$= x_2^2y_3^2 + x_3^2y_2^2 + x_3^2y_1^2 + x_1^2y_3^2 + x_1^2y_2^2 + x_2^2y_1^2 -$$
$$-2(x_2x_3y_2y_3 + x_1x_3y_1y_3 + x_1x_2y_1y_2) = |x \times y|^2.$$

$\blacksquare$

Property 4.8 immediately implies

Property 4.9. *Let $x, y \in \mathbb{E}^3$. Then*

$$x \times y = 0 \quad \textit{if and only if either } x = 0 \textit{ or } y = 0 \textit{ or } x \parallel y; \tag{4.21}$$

$$|x \times y| = |x|\ |y| \quad \textit{if and only if } x \perp y. \tag{4.22}$$

Property 4.10. *The lenght of a vector product $|x \times y|$ in the Euclidean space $\mathbb{E}^2$ is equal to the area of a parallelogram spanned by the vectors x and y.*

Indded, the area of a paralleogram spanned by vectors x and y is equal to $|x| \cdot |y| \sin\ \theta$, where θ is the angle between vectors x and y.

By direct calculations, we get

Property 4.11. *The vector product is an associative, distributive with respect to the addition and* anti*commutative operation, i.e. for arbitrary $x, y, z \in \mathbb{E}^3$ the following equalities hold:*

$$x \times (y \times z) = (x \times y) \times z; \tag{4.23}$$

$$(x + y) \times z = x \times z + y \times z; \quad x \times (y + z) = x \times y + x \times z; \tag{4.24}$$

$$y \times x = -x \times y. \tag{4.25}$$

Property 4.12. *The volume of a parallelepiped spanned on three linearly independent vectors $x = (x_1, x_2, x_3)$, $y = (y_1, y_2, y_3)$, $z = (z_1, z_2, z_3)$ in the Euclidean space $\mathbb{E}^3$ is equal to the determinant*

$$V = \begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix}.$$

Proof. The volume of a parallelepided under consideration is $V = P \cdot h$, where h is its height and P is the area of its base. By Property 4.10, we have $P = |x \times y|$. Denote by ω the angle between vectors $u = x \times y$ and z. Then

$$h = |z| \sin\ \Big(\frac{\pi}{2} - \omega\Big) = |z| \cos\ \omega = |z|\ \frac{\langle u, z\rangle}{|u| \cdot |z|} = \frac{\langle u, z\rangle}{|u|}.$$

Hence

$$V = P \cdot h = |x \times y|\ \frac{\langle u, z\rangle}{|u|} = |x \times y|\ \frac{\langle x \times y, z\rangle}{|x \times y|} = \langle x \times y, z\rangle =$$

$$= z_1 \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} - z_2 \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} + z_3 \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} = \begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix},$$

which we had to prove. ■

With appropriate definitions of n-dimensional volumes the last property holds for the vector product in the Euclidean space $\mathbb{E}^n$ for $n > 3$ without any essential change in the proofs.

Exercises.

Exercise 4.1. Give a geometric interpretation of the triangle condition (for the lenght of a vector and the distance of two points in the Euclidean space $\mathbb{E}^2$). □

Exercise 4.2. Prove the Schwarz inequality (4.1) for the Euclidean space $\mathbb{E}^n$. □

Exercise 4.3. Determine an orthogonal complement of the sets $\{(r, 0, 0) : r \in \mathbb{E}\}$ and $\{(r_1, r_2, 0) : (r_1, r_2) \in \mathbb{E}^2\}$ in the Euclidean space $\mathbb{E}^3$. □

Exercise 4.4. Give formulae for

(i) the distance of a point from a straight line in $\mathbb{E}^2$;

(ii) the distance of a point from a plane in $\mathbb{E}^3$. □

Exercise 4.5. Give conditions for

(i) two straight lines in $\mathbb{E}^2$,

(ii) two planes in $\mathbb{E}^3$,

(iii) two hyperplanes in $\mathbb{E}^n$

to intersect each another. Determine their common parts, provided that they exist. □

Exercise 4.6. Show that one can draw in the Euclidean space $\mathbb{E}^3$ a uniquely determined hyperplane through

(i) a point and a straight line;

(ii) two parallel straight lines;

(iii) two intersecting straight lines.

Write equations of these hyperplanes. □

Exercise 4.7. Give examples of two straight lines in the Euclidean space $\mathbb{E}^3$ such that no hyperplane can be drawn through them. If this is the case, such straight lines are said to be *skew*. □

Exercise 4.8. Determine in the Euclidean space $\mathbb{E}^3$

(i) the angle between two straight lines;

(ii) the angle between a straight line and a hyperplane;

(iii) the distance of two straight lines;

(iv) the distance of a point from a straight line;

(v) the point of intersection of two intersecting straight lines;

(vi) a straight line orthogonal to two skew straight lines (cf. Exercise 4.7). □

Exercise 4.9. Give formulae for

(i) the area of a triangle with given vertices in $\mathbb{E}^2$;

(ii) the volume of a tetrahedron with given vertices in $\mathbb{E}^n$. □

Exercise 4.10. Explain why the so-called *clockwise orientation* and *anticlockwise orientations* in the space $\mathbb{E}^2$ correspond to the negative and positive orientation, respectively. □

Exercise 4.11. Prove the *Lagrange identity*:

$$\Big(\sum_{j=1}^{n} a_j^2\Big)\Big(\sum_{j=1}^{n} b_j^2\Big) - \Big(\sum_{j=1}^{n} a_j b_j\Big)^2 = \sum_{\substack{j,k=1\\ j>k}}^{n} (a_j b_k - a_k b_j)^2.$$

□

Exercise 4.12. Prove that the vector product of $n-1$ linearly independent vectors in the Euclidean space $\mathbb{E}^n$ is associative, distributive with respect to the addition and anticommutative (cf. Property 4.11 for $\mathbb{E}^3$). □

Exercise 4.13. Prove that in the space $\mathbb{C}^n$ the expression

$$\langle x, y\rangle = \sum_{j=1}^{N} x_j \overline{y}_j, \quad \textit{where} \quad x_1, ..., x_n, y_1, ..., y_n \in \mathbb{C}$$

is an inner product. A linear operator A^* conjugate to $A \in L_0(\mathbb{C}^n)$ is defined by means of that inner product in the following way: $\langle Ax, y\rangle = \langle x, A^* y\rangle$ for all $x, y \in \mathbb{C}^n$. Determine the matrix corresponding to A^*. □

Chapter 5.

Eigenspaces and principal spaces. Algebraic operators.

Let X be a linear space over the field $\mathbb{F}$ (of characteristic zero) and let $A \in L_0(X)$. We shall consider linear operators of the form $T = A - \lambda I$, where $\lambda \in \mathbb{F}$.

A number $\lambda \in \mathbb{F}$ is said to be a *regular value* of the operator A if the operator $A - \lambda I$ is invertible. The set of all those λ, which are not regular values of A, is called the *spectrum* of the operator A and it is denoted by spectr A. Clearly, spectr $A \subset \mathbb{F}$. If $\lambda \in$ spectr A and there exists an $x \in X$ different than zero and such that $(A - \lambda I)x = 0$, i.e. if there is a non-zero solution of the equation $Ax = \lambda x$, then λ and x are said to be an *eigenvalue of the operator* A and its *eigenvector corresponding to the eigenvalue* λ, respectively. The linear span of all eigenvectors corresponding to an eigenvalue λ is said to be an *eigenspace of the operator* A *corresponding to its eigenvalue* λ. By definition, an eigenspace is the set

$$\{x \in X : Ax = \lambda x\} = \ker\,(A - \lambda I), \tag{5.1}$$

hence it is a linear subspace of the space X.

An element $x \in X$ such that $(A - \lambda_0 I)^n x = 0$ for a positive integer n is said to be a *principal vector* corresponding to the eigenvalue λ_0. The linear span of all principal vectors corresponding to an eigenvalue λ_0 is said to be a *principal space* of the operator A corresponding to the eigenvalue λ_0. By this definition, it follows that a principal space corresponding to an eigenvalue λ_0 is the set

$$\bigcup_{n=1}^{\infty} \{x \in X : (A - \lambda_0 I)^n x = 0\} = \bigcup_{n=1}^{\infty} \ker\,(A - \lambda_0 I)^n. \tag{5.1$'$}$$

The dimension of a principal space corresponding to an eigenvalue λ_0 of the operator A is said to be its *multiplicity*.

If there exist principal vectors corresponding to an eigenvalue λ_0 then also there exist eigenvectors corresponding to this eigenvalue. Indeed, if n is the smallest number such that $(A - \lambda_0 I)^n x = 0$, then $x_0 = (A - \lambda_0 I)^{n-1} x$ is an eigenvector corresponding to the eigenvalue λ_0 for $(A - \lambda_0 I)x_0 = (A - \lambda_0 I)^n x = 0$. On the other hand, every eigenvector corresponding to the eigenvalue λ_0 is a principal vector, since $Ax = \lambda_0 x$ implies $(A - \lambda_0 I)^n x = (A - \lambda_0 I)^{n-1}(A - \lambda_0 I)x = 0$. We therefore conclude that the dimension of a principal space is not less than the dimension of the eigenspace corresponding to the same eigenvalue λ_0.

Property 5.1. *Eigenvalues of a square matrix* $A = [a_{jk}]_{j,k=1,\ldots,n}$ *are roots of the equation*

$$\det\,(A - \lambda I) = 0. \tag{5.2}$$

Indeed, if λ_0 is an eigenvalue of the linear operator A, then there is an $x \in X$ different than zero and such that $(A - \lambda_0 I)x = 0$. This, and Corollary 3.3 together imply that $\det\,(A - \lambda_0 I) = 0$.

The determinant $\det\,(A - \lambda I)$ is said to be the *characteristic determinant* of the matrix A. By definition, this determinant is a polynomial in λ (we denote here the matrix E corresponding to the identity operator I also by I).

We shall study now a class of linear operators whose principal spaces and eigenspaces are easy to determine. In particular, we shall see that to the class in question belong all linear operators mapping finite dimensional spaces into themselves.

A field $\mathbb{F}$ is said to be *algebraically closed* if every polynomial

$$a(t) = \sum_{k=0}^{n} a_k t^k$$

with coefficients $a_0, ..., a_n \in \mathbb{F}$ has n and only n roots $t_1, ..., t_n$, i.e. if

$$a(t) = c \prod_{m=1}^{n} (t - t_m), \quad \textit{where} \quad c, t_1, ..., t_n \in \mathbb{F}.$$

For instance, the field $\mathbb{R}$ of reals is not algebraically closed, the field $\mathbb{C}$ of complex numbers is algebraically closed.

Let X be a linear space over an algebraically closed field $\mathbb{F}$ of scalars. An operator $A \in L_0(X)$ is said to be *algebraic* on X if there is a polynomial $P(t) \in \mathbb{F}[t]$ such that $P(A)x = 0$ for all $x \in X$, i.e. $P(A) = 0$ on X. Without any loss of generality we may assume here and in the sequel that any polynomial under consideration is *normalized*, i.e. its coefficient of the term of the highest degree is 1. The operator A is algebraic of the *order* N if $\deg P(t) = N$ and there is no polynomial $Q(t) \in \mathbb{F}[t]$ of degree $M < N$ such that $Q(A) = 0$ on X . If this is the case, then $P(t)$ is said to be a *characteristic polynomial* of A and its roots are called *characteristic roots* of A. We shall see that in order to determine principal spaces and eigenspaces of an algebraic operator it is enough to determine its characteristic polynomial and characteristic roots. In the most of cases it is enough to consider instead of an arbitrary algebraically closed field $\mathbb{F}$ the field $\mathbb{C}$ of complexes.

Lemma 5.1. Hermite formula for interpolation with multiple knots.$^{*)}$ *There exists a unique polynomial $Q(t)$ of degree $N-1$ which together with its derivatives admits given values y_{ki} at different points t_i $(i = 1, ..., n;\ k = 0, ..., r_i - 1;\ r_1 + ... + r_n = N)$. More precisely,*

$$Q^{(k)}(t_i) = y_{ki}, \quad \textit{where} \quad Q^{(k)} = \frac{\mathrm{d}^k Q}{\mathrm{d}t^k},\ Q^{(0)} = Q,$$

$$(i = 1, ..., n;\ k = 0, ..., r_i - 1;\ r_1 + ... + r_n = N).$$

$^{*)}$ For the proof in the case when $\mathbb{F} = \mathbb{C}$ cf. Ch. Hermite H[1].

The polynomial $Q(t)$ is of the form

$$Q(t) = \sum_{i=1}^{n} \frac{P(t)}{(t-t_i)^{r_i}} \sum_{k=0}^{r_i-1} \left\{ \frac{(t-t_i)^{r_1}}{P(t)} \right\}_{(r_i-1-k;\ t_i)} \frac{(t-t_i)^k}{k!}, \tag{5.3}$$

where we write *)

$$P(t) = \prod_{m=1}^{n} (t-t_m)^{r_m}, \quad \{f(t)\}_{(k;s)} = \sum_{m=0}^{k} \frac{(t-s)^m}{m!} \cdot \left[\frac{\mathrm{d}^m f(t)}{\mathrm{d}t^m} \right]_{t=s} \tag{5.4}$$

for any function f k-times differentiable in a neighbourhood of the point s.

Proof. This proof can be omitted in the first reading. Consider a polynomial $\tilde{Q}(t)$ of degree $N-1$ which is of the form

$$\tilde{Q}(t) = P(t) \sum_{i=1}^{n} \sum_{j=0}^{r_i-1} \frac{a_{i,j}}{(t-t_i)^{r_i-j}},$$

where the scalar coefficients are not yet determined. If we let

$$F_i(t) = \frac{P(t)}{(t-t_i)^{r_i}} \quad (i = 1, ..., n),$$

then

$$\frac{\tilde{Q}(t)}{F_i(t)} = \frac{\tilde{Q}(t)}{P(t)}(t-t_i)^{r_i} = \sum_{j=0}^{r_i-1} a_{i,j}(t-t_i)^j + (t-t_i)^{r_i} R_i(t),$$

where the function $R_i(t)$ is regular at the point t_i $(i = 1, ..., n)$. If we decompose the rational function $\tilde{Q}(t)/P(t)$ into vulgar fractions, then we can calculate scalar coefficients

$$a_{i,0} = \left[\frac{\tilde{Q}(t)}{P(t)} \right]_{t=t_i},$$

....

$$a_{i,j} = \frac{1}{j!} \left[\frac{\mathrm{d}}{\mathrm{d}t^j} \frac{\tilde{Q}(t)}{F_i(t)} \right]_{t=t_i} = \frac{1}{j!} \sum_{k=0}^{j} \binom{j}{k} \tilde{Q}^{(k)}(t_i) \left[\frac{1}{F_i(t)} \right]_{t=t_i}^{(j-k)} +$$

$$= \frac{1}{j!} \sum_{k=0}^{j} \binom{j}{k} y_{ki} \left[\frac{1}{F_i(t)} \right]_{t=t_i}^{(j-k)},$$

where we let $\tilde{Q}^{(k)}(t_i) = y_{kj} = Q^{(k)}(t_i)$, $(i = 1, ..., n;\ k = 0, ..., r_i - 1)$. Hence

$$\tilde{Q}(t) = P(t) \sum_{i=1}^{n} \sum_{j=0}^{r_i-1} \frac{a_{i,j}}{(t-t_i)^{r_i-j}} = \sum_{i=1}^{n} \sum_{j=0}^{r_i-1} \frac{P(t)}{(t=t_i)^{r_i}} (t-t_i)^j a_{i,j} =$$

*) Roots t_i of the polynomial $Q(t)$ are called its *knots*.

$$= \sum_{i=1}^{n} \sum_{j=0}^{r_i-1} \frac{P(t)}{(t-t_i)^{r_i}} (t-t_i)^j \frac{1}{j!} \sum_{k=0}^{j} \binom{j}{k} y_{ki} \left[\frac{1}{F_i(t)}\right]_{t=t_i}^{(j-k)} =$$

$$= \sum_{i=1}^{n} \frac{P(t)}{(t-t_i)^{r_i}} \sum_{j=0}^{r_i-1} \sum_{k=0}^{j} a_{k,i} \frac{(t-t_i)^{j-k}}{(j-k)!} \left[\frac{1}{F_i(t)}\right]_{t=t_i}^{(j-k)} \frac{(t-t_i)^k}{k!} =$$

$$= \sum_{i=1}^{n} \frac{P(t)}{(t-t_i)^{r_i}} \sum_{k=0}^{r_i-1} y_{ki} \left(\sum_{j=k}^{r_i-1} \frac{(t-t_i)^{j-k}}{(j-k)!} \left[\frac{1}{F_i(t)}\right]_{t=t_i}^{(j-k)} \right) \frac{(t-t_i)^k}{k!} =$$

$$= \sum_{i=1}^{n} \frac{P(t)}{(t-t_i)^{r_i}} \sum_{k=0}^{r_i-1} y_{ki} \left(\sum_{m=0}^{r_i-1} \frac{(t-t_i)^{m}}{m!} \left[\frac{1}{F_i(t)}\right]_{t=t_i}^{(m)} \right) \frac{(t-t_i)^k}{k!} =$$

$$= \sum_{i=1}^{n} \frac{P(t)}{(t-t_i)^{r_i}} \sum_{k=0}^{r_i-1} y_{ki} \left\{ \frac{1}{F_i(t)} \right\}_{(r_i-1-k;\ t_i)} \frac{(t-t_i)^k}{k!} =$$

$$= \sum_{i=1}^{n} \frac{P(t)}{(t-t_i)^{r_i}} \sum_{k=0}^{r_i-1} y_{ki} \left\{ \frac{(t-t_i)^{r_i}}{P(t)} \right\}_{(r_i-1-k;\ t_i)} \frac{(t-t_i)^k}{k!} = Q(t),$$

where the polynomial $Q(t)$ is defined by Formula (5.3). We therefore conclude that $Q(t)$ is the unique polynomial satisfying the required conditions. ■

If t_i are *single* knots, i.e. if $r_i = 1$ for $i = 1, ..., n$), then the Hermite interpolation formula (5.4) implies the *Lagrange interpolation formula*

(5.5)
$$Q(t) = \sum_{i=1}^{n} \prod_{\substack{m=1 \\ m \neq i}}^{n} \frac{t-t_m}{t_i-t_m}.$$

Indeed, if $r_i = 1$ for $i = 1, ..., n$ then

$$\left\{ \frac{(t-t_i)^{r_i}}{P(t)} \right\} = \left\{ \frac{t-t_i}{P(t)} \right\}_{(0,t_i)} = \frac{t-t_i}{P(t)}|_{t=t_i} = \prod_{\substack{m=1 \\ m \neq i}}^{n} (t-t_m)^{-1}\Big|_{t=t_i} = \prod_{\substack{m=1 \\ m \neq i}}^{n} (t_i-t_m)^{-1}.$$

Writing $y_i = y_{0i}$ for $i = 1, ..., n$ in Formula (5.4), we obtain Formula (5.5).

Lemma 5.2. (Partition of unity). *Write*

(5.6)
$$\mathbf{p}_i(t) = q_i(t) \prod_{\substack{m=1 \\ m \neq i}}^{n} (t-t_m)^{r_m},$$

where

$$q_i(t) = \left\{ \frac{(t-t_i)^{r_i}}{P(t)} \right\}_{(r_i-1;t_i)} \qquad (i = 1, ..., n).$$

Then

$$\sum_{i=1}^{n} \mathbf{p}_i(t) \equiv 1 \tag{5.7}$$

and this representation is unique for fixed t_i and r_i.

Proof. Let $Q(t) \equiv 1$. Then $q^{(0)}(t_i) = 1$ and $Q^{(k)}(t_i) = 0$ for $k \geq 1$, $i = 1, ..., n$. The Hermite interpolation formula (5.4) implies that

$$1 = \sum_{i=1}^{n} \left[\prod_{\substack{m=1 \\ m \neq i}}^{n} (t - t_m)^{r_m} \right] \left\{ \frac{(t - t_i)^{r_i}}{P(t)} \right\}_{(r_i - 1; t_i)} = \sum_{i=1}^{n} \mathbf{p}_i(t).$$

■

In the case of single roots, by the Lagrange interpolation formula (5.5), we get instead of Formula (5.7) the following formula for the partition of unity:

$$\sum_{i=1}^{n} \prod_{\substack{m=1 \\ m \neq i}}^{n} \frac{t - t_m}{t_i - t_m} \equiv 1. \tag{5.8}$$

Theorem 5.1. *Let $A \in L_0(X)$. Then the following conditions are equivalent:*

(i) the operator A is algebraic on X with the characteristic polynomial

$$P(t) = \prod_{j=1}^{n} (t - t_j)^{r_j}, \quad t_j \neq t_k \; if \; j \neq k$$

with the order $N = r_1 + ... + r_n$;

(ii) there exist n disjoint projectors $P_1, ..., P_n \in L_0(X)$ giving the partition of unity, i.e. such that

$$X_j = P_j X, \quad P_j = \mathbf{p}_j(A),$$

$$P_j P_k = \begin{cases} P_k & \text{for } j = k, \\ 0 & \text{for } j \neq k, \end{cases}, \qquad \sum_{j=1}^{N} P_j = I \quad and \quad (A - t_j I)^{r_j} P_j = 0 \qquad (j, k = 1, ..., n),$$

where polynomials $\mathbf{p}_j(t)$ are defined by Formulae (5.6);

(iii) X is the direct sum of principal spaces of the operator A corresponding to the eigenvalues $t_1, ..., t_n$:

$$X = X_1 \oplus ... \oplus X_n, \quad where \; (A - t_j I)^{r_j} x_j = 0 \; for \; x_j \in X_j,$$

i.e. $X_j = \ker\,(A - t_j I)^{r_j}$ $(j = 1, ..., n)$.

Proof. We will prove the implications (i) $\to$ (ii) $\to$ (iii) $\to$ (i).

Proof of the implication (i) $\to$ (ii). Admit $P_j = \mathbf{p}_j(A)$ $(j = 1, ..., n)$, where the polynomials $\mathbf{p}_j(t)$ are defined by Formula (5.6). By Lemma 5.2 on the partition of unity, it follows that

$$\sum_{j=1}^{n} P_j = \sum_{j=1}^{n} \mathbf{p}_j(A) = I.$$

Since all polynomials with scalar coefficients commute each with another, we conclude that for $j \neq k$

$$P_j P_k = \mathbf{p}_j(A)\mathbf{p}_k(A) = q_j(A)q_k(A)\Big[\prod_{\substack{m=1\\ m\neq j}}^{n} (A - t_m I)^{r_m}\Big] \Big[\prod_{\substack{i=1\\ i\neq j}}^{n} (A - t_i I)^{r_i}\Big] =$$

$$= q_j(A)q_k(A)\Big[\prod_{\substack{m=1\\ m\neq j; m\neq k}}^{n} (A - t_m)^{r_m}\Big]P(A) = 0.$$

If $j = k$ we have

$$P_j = P_j \sum_{k=1}^{n} P_k = \sum_{k=1}^{n} P_j P_k = P_j^2 \qquad (j = 1, ..., n).$$

Moreover, for a fixed number m $(m = 1, ..., n)$ we get

$$(A - t_m I)^{r_m} P_m = (A - t_m I)^{r_m} \mathbf{p}_m(A) =$$

$$= (A - t_m I)^{r_m} q_m(A) \prod_{\substack{k=1\\ k\neq m}}^{n} (A - t_k I)^{r_k} = q_m(A)P(A) = 0.$$

Hence the operators $P_1, ..., P_n$ have all required properties.

Proof of the implication (ii) $\to$ (iii). By our assumptions, there are n operators $P_1, .., P_n \in L_0(X)$ such that

$$\sum_{j=1}^{n} P_j = I, \quad P_j P_k = 0 \quad for \quad j \neq k, \quad P_j^2 = P_j$$

$$and \quad (A - t_j I)^{r_j} P_j = 0 \quad for \quad j = 1, ..., n.$$

Hence the operators $P_1, ..., P_n$ are disjoint projectors such that their sum is the identity operator I, i.e. the space X is a direct sum of n spaces $X_1, ..., X_n$ defined by means of the equalities $X_j = P_j X$ for $j = 1, ..., n$. But the equalities $(A - t_j I)^{r_j} P_j = 0$ imply that $(A - t_j I)^{r_j} x = 0$ for $x \in X_j = P_j X$. Hence each of spaces X_j is a principal space of the operator A corresponding to the root t_j, i.e. $X_j = \ker\,(A - t_j I)^{r_j}$.

Proof of the implication (iii) $\to$ (i). Suppose that X is a direct sum of principal spaces $X_j = \ker\,(A - t_j I)^{r_j}$. Write

$$P(A) = \prod_{m=1}^{n} (A - t_m I)^{r_m}.$$

Let x be an arbitrary element of the space X. By our assumptions,

$$x = \sum_{j=1}^{n} x_j, \quad where \quad x_j \in X_j$$

for every $x \in X$. Hence

$$P(A)x = P(A)\sum_{j=1}^{n} x_j = \sum_{j=1}^{n} P(A)x_j = \sum_{j=1}^{n} \left[\prod_{\substack{m=1\\ m\neq j}}^{n} (A - t_m I)^{r_m}\right](A - t_j I)^{r_j} x_j = 0$$

for every $x \in X$. Then $P(A) = 0$ on the space X, which implies that A is an algebraic operator of the order $N = r_1 + ... + r_n$. ■

If $r_1 = ... = r_n = 1$, i.e. if the characteristic roots are single, then $N = r_1 + ... + r_n = n$. This, and Theorem 5.1, together immediately imply

Corollary 5.1. *Let $A \in L_0(X)$. Then the following conditions are equivalent:*

(i) A is an algebraic operator of the order n with the characteristic polynomial

$$P(t) = \prod_{j=1}^{n} (t - t_j) \qquad (t_j \neq t_k \ for\ j \neq k);$$

(ii) there exists n projectors $P_j \in L_0(X)$ such that

$$P_j P_k = \begin{cases} P_k & \text{for } j = k, \\ 0 & \text{for } j \neq k, \end{cases}, \qquad \sum_{j=1}^{n} P_j = I \quad AP_j = t_j P_j \quad (j = 1, ..., n),$$

namely,

$$P_j(t) = \prod_{\substack{k=1\\ k\neq j}}^{n} \frac{t - t_k}{t_j - t_k} \quad (j = 1, ..., n);$$

(iii) the space X is a direct sum of n eigenspaces of the operator A corresponding to its eigenvalues $t_1, ..., t_n$, respectively, i.e.

$$X = X_1 \oplus ... \oplus X_n, \quad where \quad X_j = P_j X = \ker\,(A - t_j I).$$

Corollary 5.2. *If $A \in L_0(X)$ is an algebraic operator with the characteristic polynomial $P(t)$, then its conjugate operator A' is also an algebraic operator with the same characteristic polynomial.*

Proof. Indeed, if

$$P(t) = \sum_{k=1}^{n} p_k A^k = 0$$

then

$$P(A') = \sum_{k=1}^{n} p_k (A')^k = \left[\sum_{k=1}^{n} p_k A^k \right]' = [P(t)]' = 0.$$

■

Corollary 5.3. (Cayley-Hamilton Theorem). *If* dim $X < +\infty$ *then every operator $A \in L_0(X)$ is an algebraic operator such that its characteristic polynomial is a divisor of the polynomial $Q(\lambda) =$* det $(A - \lambda I)$.

Proof. Property 5.1 and Point (iii) of Theorem 5.1 immediately imply that to every operator $A \in L_0(X)$ there corresponds a square matrix $A = [a_{jk}]_{j,k=1,\ldots,\dim X}$. ■

Note that in the case when dim $X < +\infty$ and $A \in L_0(X)$ the characteristic polynomial of the algebraic operator A is said to be the *minimal polynomial* of the matrix A. Hence, by definition, a minimal polynomial of a matrix is a divisor of its characteristic polynomial.

Suppose that dim $X < +\infty$ and $A \in L_0(X)$ is an algebraic operator with the characteristic polynomial

$$P(t) = \prod_{m=1}^{n} (t - t_m)^{r_m}.$$

We define a *Jordan matrix* corresponding to a characteristic root t_m as a square matrix of dimension $k \le r_m$ of the form

$$J_{m,k} = \begin{cases} (t_m) & \text{if } r_m = 1; \\ \begin{pmatrix} t_m & 1 & 0 & \ldots & 0 & 0 \\ 0 & t_m & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \ldots & t_m & 1 \\ 0 & 0 & \ldots & \ldots & \ldots & t_m \end{pmatrix} & \text{if } r_m \ge 2. \end{cases} \tag{5.9}$$

where $k = 1, \ldots, r_m$; $m = 1, \ldots, n$.

For a given a system of square matrices $A_1, \ldots, A_M$ of dimensions $n_1, \ldots, n_M$, respectively, denote a square matrix

$$[[A_i]]_{i=1,\ldots,M} = \begin{pmatrix} A_1 & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & \ldots & A_M \end{pmatrix}. \tag{5.9'}$$

of dimension $n_1 + ... + n_m$.

Theorem 5.2. (Jordan Theorem). *Let* dim $X < +\infty$. *Let be given a square matrix* $A = [a_{jk}]_{j,k=1,...,\dim X}$ *with the characteristic polynomial*

$$P(t) = \prod_{m=1}^{n} (t - t_m)^{r_m}.$$

Then by an appropriate change of the basis in the space X the matrix A can be represented in the form

$$A = [[J_{m,k_m}]]_{\substack{m=1,...,n; \\ k_m=1,...,r_m}},$$

where J_{m,k_m} are Jordan matrices corresponding to the characteristic root t_m; some of them may appear several times, while some may not appear.

Proof. By the Cayley-Hamilton theorem (Corollary 5.3), A is an algebraic operator acting in a finite dimensional space X. By Theorem 5.1, the space X is a direct sum of spaces $X_m = \ker\,(A - t_m I)^{r_m}$ $(m = 1, ..., n)$. Hence it is enough to prove that, by an appropriate change of the basis in a fixed space X_m, we obtain a representation of the matrix A by Jordan matrices J_{m,k_m}.

First, we shall consider the case when dim $X_m = r_m$. If $r_m = 1$ then for all $x \neq 0$ such that $X_m = \ker\,(A - t_m I)$ we have $Ax = t_m x$. Hence $A = t_m I = J_{m,1}$ on the space X_m. Consider now the case $r_m \geq 2$. Write $B = A - t_m I$. By definition, $B^{r_m} = (A - t_m I)^{r_m}$ $= 0$ on the space X_m. But $B^{r_m - 1} \neq 0$ (i.e. B is a nilpotent operator of the order r_m; cf. Exercise 2.13). Then there is an element $x_0 \neq 0$ such that $x_0 \in X_m$ and $y = B^{r_m-1}x_0 \neq 0$. It is easy to verify that all elements $x_0, Bx_0, ..., B^{r_m-1}x_0$ are linearly independent. Hence the system $\mathcal{B}_m = \{B^{r_m-1}x_0, ..., Bx_0, x_0\}$ is a basis in the space X_m. Let $x \in X_m$ be arbitrary. Then x can be represented in a unique way (with respect to the basis $\mathcal{B}_m$) in the form

$$x = \sum_{j=0}^{r_m-1} b_j B^{r_m-1-j} x_0, \quad \textit{where} \quad b_j \in \mathbb{C}.$$

Since $B^{r_m} = 0$, we conclude that

$$Bx = \sum_{j=0}^{r_m-1} b_j B^{r_m-1-j+1} x_0 = \sum_{j=0}^{r_m-1} b_j B^{r_m-j} x_0 = \sum_{j=1}^{r_m-1} b_j B^{r_m-j} x_0.$$

Then, acting by the operator B on the vector $(b_1, ..., b_n)$ written with respect to the basis $\mathcal{B}_m$, we obtain the vector $(b_1, ..., b_{r_m-1}, 0)$ in the same basis. On the other hand.,

$$\begin{pmatrix} 0 & 1 & 0 & ... & 0 & 0 \\ 0 & 0 & 1 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & ... & 0 & 1 \\ 0 & 0 & ... & ... & ... & 0 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_{r_m-1} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ 0 \end{pmatrix}.$$

Hence to the operator B in the basis $\mathcal{B}_m$ there corresponds the matrix

$$\begin{pmatrix} 0 & 1 & 0 & ... & 0 & 0 \\ 0 & 0 & 1 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & ... & 0 & 1 \\ 0 & 0 & ... & ... & ... & 0 \end{pmatrix}.$$

This implies that to the operator $A = B + t_m I$ in the basis $\mathcal{B}_m$ there correspond the expression

$$\begin{pmatrix} 0 & 1 & 0 & ... & 0 & 0 \\ 0 & 0 & 1 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & ... & 0 & 1 \\ 0 & 0 & ... & ... & ... & 0 \end{pmatrix} + t_m \begin{pmatrix} 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 1 \end{pmatrix} =$$

$$= \begin{pmatrix} t_m & 1 & 0 & ... & 0 & 0 \\ 0 & t_m & 1 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & ... & t_m & 1 \\ 0 & 0 & ... & ... & ... & t_m \end{pmatrix} = J_{m,k_m},$$

which we wanted to prove.

Suppose now that $r_m < \dim X_m < +\infty$. Observe that

$$\ker\ B \supset ... \supset \ker\ B^{r_m} = X_m.$$

Moreover, the operator B maps $\ker\ B^k$ into $\ker\ B^{k-1}$ for $k = 2, ..., r_m$. By our assumption that $r_m < \dim\ X_m$, there are elements $x_1, ..., x_M \in X_m$ such that $x_j \neq 0$, $Bx_j = 0$ for $j = 1, ..., M$ and the system $\{x_1, ..., x_M\}$ is a basis in the space $\ker\ B^{r_m - 1}$ (in the previous step of that proof we have admitted $M = 1$). As before, we can show that all elements

$$x_1, ..., x_M, ..., B^{r_m} x_1, ..., B^{r_m} x_M$$

are linearly independent. In order to do so, consider the system

$$\mathcal{B}_m^{(M)} = \{B^{r_m - 1} x_1, ..., B^{r_m - 1} x_M, ..., x_1, ..., x_M\}.$$

If $\dim X_m = M r_m$ then the system $\mathcal{B}_m^{(M)}$ is a basis in the space X_m. In this case, as in the first step of our proof, we prove that the matrix A can be represented in the basis $\mathcal{B}_m^{(M)}$ by M Jordan matrices corresponding to the root t_m, each of them of dimension r_m, i.e.

$$A = \begin{pmatrix} A_1 & ... & 0 \\ ... & ... & ... \\ 0 & ... & A_M \end{pmatrix} \quad \textit{where} \quad A_1 = ... = A_M = J_{m,r_m}.$$

Consider the last case: the system $\mathcal{B}_m^{(M)}$ is not a basis in the space X_m. Then lin $\mathcal{B}_m^{(M)} \neq X_m$. Denote by k the greatest positive integer such that there is an element $x \in X_m$ which does not belong to lin $\mathcal{B}_m^{(M)}$. Clearly, $k < r_m$, since for $k = r_m$ we obtain the previous case. Let elements $y_1, ..., y_p \in \ker\ B^k$ be such that the set $\{y_1, ..., y_p, \mathcal{B}^{(M)}\}$ consists of linearly independent elements. Consider the set

$$\mathcal{B}_m^{(M+1)} = \{B^{r_m - 1}y_1, ..., B^{r_m-1}y_p, \mathcal{B}_m^{(M)}\}.$$

If this set is a basis in the space X_m, i.e. if dim $X_m = (M+1)r_m$, then in a similar way, as before, we show that the matrix X can be represented in the basis $\mathcal{B}_m^{(M+1)}$ by $M+1$ Jordan matrices. If this is not the case, i.e. if $\mathcal{B}_m^{(M+1)}$ is not a basis in the space X_m, then we repeat the previous arguments. Since the space X_m is finite dimensional, after a finite number of steps we shall obtain a finite system which is a basis in the space X_m. We therefore conclude that in the already constructed basis the matrix A can be represented by Jordan matrices J_{m,r_m} corresponding to the root t_m. This finishes the proof of the Jordan theorem. ■

Property 5.2. *Suppose that $A \in L_0(X)$ is an algebraic operator with the characteristic polynomial $\prod_{m=1}^{n} (t - t_m)^{r_m}$ and*

$$Q_k = (A - t_k I)P_k \qquad (k = 1, ..., n), \tag{5.10}$$

where the projectors $P_1, ..., P_n$ are defined in Theorem 5.1(ii), i.e. $P_j = \mathbf{p}_j(A)$ with $\mathbf{p}_j(t)$ determined by Formula (5.6) $(j = 1, ..., n)$. Then

$$Q_j Q_k = 0 \quad for \quad j \neq k, \tag{5.11}$$

$$P_k Q_j = Q_j P_k = \begin{cases} 0 & \text{for } j \neq k; \\ Q_j & \text{for } j = k \end{cases} \quad (j, k = 1, ..., n), \tag{5.12}$$

$$Q_k^m = (A - t_k I)^m P_k \quad for \quad k = 1, ..., n;\ m \in \mathbb{N}. \tag{5.13}$$

In particular,

$$Q_k^{r_k} = 0 \qquad (k = 1, ..., n). \tag{5.14}$$

Proof. By definition of the projectors $P_1, ..., P_n$, we have $P_j P_k = 0$ for $j \neq k$ and $P_j^2 = P_j$ $(j, k = 1, ..., n)$. Moreover, the operators $A - t_k I$ commute with the operators $P_1, .., P_n$. Hence for $j \neq k$ we have

$$Q_j Q_k = (A - t_j I)P_j(A - t_k I)P_k = (A - t_j I)(A - t_k I)P_j P_k = 0,$$

$$Q_j P_k = (a - t_j I)P_j P_k = 0,$$

$$P_kQ_j = P_k(A - t_jI)P_j = (A - t_jI)P_kP_j = 0.$$

We have also

$$Q_jP_j = (a - t_jI)P_j^2 = (A - t_jI)P_j = Q_j,$$

$$P_jQ_j = P_j(A - t_jI)P_j = (A - t_jI)P_j^2 = (A - t_jI)P_j = Q_j.$$

Then Formulae (5.11) and (5.12) are already proved.

We shall prove now Formula (5.13) by induction with respect to m for an arbitrarily fixed positive integer $k = 1, ..., m$. For $m = 1$, Formula (5.13) is true by the definition of the operator Q_k. Suppose that Formula (5.13) holds for an arbitrarily fixed positive integer $m \geq 1$. Then, by our induction assumption,

$$Q_k^{m+1} = Q_kQ_k^m = Q_k(A - t_kI)^mP_k = (A - t_kI)P_k(A - t_kI)^mP_k =$$

$$= (A - t_kI)(A - t_kI)^mP_k^2 = (A - t_kI)^{m+1}P_k,$$

i.e. Formula (5.13) holds for $m+1$. We therefore conclude that Formula (5.13) holds for an arbitrary positive integer m, what was to be proved. In particular (cf. Theorem 5.1(ii)),

$$Q_{r_k}^m = Q_m^{r_k} = (A - t_kI)^{r_k}P_k = 0.$$

■

Theorem 5.3. *Suppose that $A \in L_0(X)$ is an algebraic operator with the characteristic polynomial $\prod_{m=1}^{n} (t - t_m)^{r_m}$. Then*

$$A^m = \sum_{j=1}^{n} \left[\binom{m}{k} t_j^{m-j}(A - t_jI)^k\right] \qquad (m \in \mathbb{N}). \tag{5.15}$$

Proof (by induction). Let $m = 1$. Since $\sum_{j=1}^{n} P_j = I$, we have

$$A = A\sum_{j=1}^{n} P_j = \sum_{j=0}^{n} AP_j = \sum_{j=1}^{n} [t_jP_j + (A - t_jI)P_j] = \sum_{j=1}^{n} [t_jI + (A - t_jI)]P_j,$$

i.e. Formula (5.15) holds for $m = 1$.

Suppose now that Formula (5.15) is true for an arbitrary positive integer $m \geq 1$. Formulae (5.10) and (5.11) together imply that

$$A^m = \sum_{j=0}^{n} \left[t_j^mI + \sum_{k=1}^{n} \binom{m}{k} t_j^{m-k}(A - t_jI)^k\right]P_j =$$

$$= \sum_{j=0}^{n} \left[t_j^m P_j + \sum_{k=1}^{n} \binom{m}{k} t^{m-k}(A - t_j I)^k P_j\right] = \sum_{j=0}^{n} \left[t_j^m P_j + \sum_{k=1}^{n} \binom{m}{k} t_j^{m-k} Q_j^k\right].$$

Since $P_i P_j = 0$ for $i \neq j$ and

$$A = A \cdot I = A \sum_{j=1}^{n} P_j = \sum_{j=1}^{n} AP_j,$$

by Formulae (5.11) and (5.12), we find

$$A^{m+1} = A \cdot A^m = \sum_{i=1}^{n} (t_i P_i + Q_i) \sum_{j=1}^{n} \left[t_j^m P_j + \sum_{k=1}^{n} \binom{m}{k} t_j^{m-k} Q_j^k k\right] =$$

$$= \sum_{i,j=1}^{n} \left[t_i t_j^m P_i P_j + t_j^m Q_i P_j + t_j + \sum_{k=1}^{n} \binom{m}{k} t_j^{m-k} Q_k^j + \sum_{k=1}^{n} \binom{m}{k} t_j^{m-k} Q_i Q_k^k\right] =$$

$$= \sum_{j=1}^{n} \left[t_j^{m+1} P_j^2 + t_j^m Q_j P_j + \sum_{k=1}^{n} \binom{m}{k} t_j^{m+1-k} P_j Q_j^k \sum_{k=1}^{n} \binom{m}{k} t_j^{m-k} Q_j^{k+1}\right] =$$

$$= \sum_{j=1}^{n} \left[t_j^{m+1} P_j + t_j^m Q_j + \sum_{k=1}^{n} \binom{m}{k} t_j^{m+1-k} Q_j^k + \sum_{k=1}^{n} \binom{m}{k} t_j^{m-k} Q_j^{k+1}\right] =$$

$$= \sum_{j=1}^{n} \left[t_j^{m+1} P_j + t_j^m Q_j + \sum_{\mu=0}^{m-1} \binom{m}{\mu+1} t_j^{m-\mu} Q_j^{\mu+1} + \sum_{k=1}^{m} \binom{m}{k} t_j^{m-k} Q_j^{k+1}\right] =$$

$$= \sum_{j=1}^{n} \left\{t_j^{m+1} P_j + t_j^m Q_j \binom{m}{1} t_j^m Q_j + \sum_{\mu=1}^{m-1} \left[\binom{m}{\mu+1} + \binom{m}{\mu}\right] t_j^{m-\mu} Q_j^{\mu+1} + Q_j^{m+1}\right\} =$$

$$= \sum_{j=1}^{n} \left[t_j^{m+1} P_j + (m+1) t_j^m Q_j + \sum_{\mu=1}^{m-1} \binom{m+1}{\mu+1} t_j^{m-\mu} Q_j^{\mu+1} + Q_j^{m+1}\right] =$$

$$= \sum_{j=1}^{n} \left[t_j^{m+1} P_j + (m+1) t_j^m Q_j + \sum_{k=2}^{m-1} \binom{m+1}{k} t_j^{m+1-k} Q_j^k + Q_j^{m+1}\right] =$$

$$= \sum_{j=1}^{n} \left[t_k^{m+1} P_j + \sum_{k=1}^{m+1} \binom{m+1}{k} t_j^{m+1-k} Q_j^k\right],$$

i.e. Formula (5.15) holds for $m + 1$. We therefore conclude that Formula (5.15) holds for an arbitrary positive integer m, what was to be proved. ■

Corollary 5.4. *If an operator A satisfies all conditions of Theorem 5.3 then*

$$(5.16) \qquad A^m P_i = t_i^m P_i + \sum_{k=1}^{m} \binom{m}{k} t_i^{m-k} (A - t_i I)^k P_i \quad \textit{for} \quad i = 1, ..., n;\ m \in \mathbb{N}.$$

Proof. Since $P_iP_j = 0$ for $i \neq j$ $(i, j = 1, ..., n)$, Formula (5.15) implies that

$$A^m P_i = \sum_{j=1}^{n} \left[t_j^m I + \sum_{k=1}^{n} \binom{m}{k} t_j^{m-k} (A - t_j I)^k \right] P_j P_i =$$

$$= \left[t_i^m + \sum_{k=1}^{m} \binom{m}{k} t_j^{m-k} (A - t_i I)^k \right] P_i = t_i^m P_i + \sum_{k=1}^{n} \binom{m}{k} t_i^{m-k} (A - t_i I)^k P_i.$$

∎

Corollary 5.5. *If $A \in L_0(X)$ is an algebraic operator with single characteristic roots, i.e. its characteristic polynomial is of the form $\prod_{j=1}^{n} (t - t_j)$, then*

$$A^m = \sum_{j=1}^{n} t_j^m P_j \qquad (m \in \mathbb{N}), \tag{5.17}$$

$$A^m P_i = t_j^m P_i \qquad (i = 1, ..., m;\ m \in \mathbb{N}) \tag{5.18}$$

Moreover, for an arbitrary polynomial $Q(t)$ with coefficients in $\mathbb{F}$ we have

$$Q(A) = \sum_{j=0}^{n} Q(t_j) P_j. \tag{5.19}$$

Proof. If all characteristic roots of the operator A under consideration are single then, by Corollary 5.1, we find $AP_j = t_jP_j$ $(j = 1, ..., n)$. Hence $(A - t_jI)P_j = 0$ $(j = 1, ..., n)$. This implies that all components of the form $(A - t_jI)^kP_j$ $(j = 1, ..., n;\ k \in \mathbb{N}$ in Formula (5.15) vanish. We therefore obtain Formula (5.17). This, and Corollary 5.4 together immediately imply Formula (5.15). Suppose now that we are given a polynomial

$$Q(t) = \sum_{k=0}^{M} q_k t^k, \quad \textit{where} \quad q_0, ..., q_M \in \mathbb{F}.$$

Formula(5.18) implies that

$$Q(A) = \sum_{k=0}^{M} q_k A^k = \sum_{k=0}^{M} q_k \sum_{j=0}^{n} t_j^k P_j = \sum_{k=0}^{M} \left[\sum_{j=0}^{n} q_k t_j^k \right] P_j = \sum_{j=1}^{n} Q(t_j) P_j.$$

∎

Theorem 5.4. *If $A \in L_0(X)$ is an algebraic operator with the characteristic polynomial $P(t) = \prod_{j=1}^{n} (t-t_j)^{r_j}$ then for every $\lambda \neq t_1, ..., t_n$ the operator $A - \lambda I$ is invertible and*

$$(A-\lambda I)^{-1} = \sum_{j=1}^{n} \left[\frac{1}{t_j-\lambda} I + \sum_{m=1}^{r_j-1} \frac{(-1)^{m+1}}{(t_j-\lambda)^{m+1}} (A-t_jI)^m \right] P_j. \tag{5.20}$$

Proof. Write for $\lambda \neq t_1, ..., t_n$

$$B = \sum_{j=1}^{n} \left[\frac{1}{t_j-\lambda} P_j + \sum_{m=1}^{r_j-1} \frac{(-1)^{m+1}}{(t_j-\lambda)^{m+1}} (A-t_jI)^m P_j \right] = \tag{5.21}$$

$$= \sum_{j=1}^{n} \frac{1}{t_j-\lambda} \left[P_j + \sum_{m=1}^{r_j-1} \frac{(-1)^{m+1}}{(t_j-\lambda)^{m}} Q_j^m \right].$$

Theorem 5.3 and Property 5.2 together imply that

$$(A-\lambda I)B = \sum_{i=1}^{n} \Big[(t_i-\lambda)P_i + Q_i \Big] \sum_{j=1}^{n} \frac{1}{t_j-\lambda} \left[P_j + \sum_{m=1}^{r_j-1} \frac{(-1)^{m+1}}{(t_j-\lambda)^{m}} Q_j^m \right] =$$

$$= \sum_{i,j=1}^{n} \left\{ (t_i-\lambda) \frac{1}{t_j-\lambda} \left[P_iP_j + \sum_{m=1}^{r_j-1} \frac{(-1)^{m+1}}{(t_j-\lambda)^{m}} P_iQ_j^m \right] + \right.$$

$$\left. + \frac{1}{t_j-\lambda} \left[Q_iP_j + \sum_{m=1}^{r_j-1} \frac{(-1)^{m+1}}{(t_j-\lambda)^{m}} Q_iQ_j^m \right] \right\} =$$

$$= \sum_{j=1}^{n} \left\{ (t_i-\lambda) \frac{1}{t_j-\lambda} \left[P_j^2 + \sum_{m=1}^{r_j-1} \frac{(-1)^{m+1}}{(t_j-\lambda)^{m}} P_jQ_j^m \right] + \right.$$

$$\left. + \frac{1}{t_j-\lambda} \left[Q_jP_j + \sum_{m=1}^{r_j-1} \frac{(-1)^{m+1}}{(t_j-\lambda)^{m}} Q_j^{m+1} \right] \right\} =$$

$$= \sum_{j=1}^{n} \left[P_j + \sum_{m=1}^{r_j-1} \frac{(-1)^{m+1}}{(t_j-\lambda)^{m}} Q_j^m + \frac{1}{t_j-\lambda} Q_j + \sum_{m=1}^{r_j-1} \frac{(-1)^{m+1}}{(t_j-\lambda)^{m+1}} Q_j^{m+1} \right] =$$

$$= \sum_{j=1}^{n} \left[P_j + \sum_{m=1}^{r_j-1} \frac{(-1)^{m+1}}{(t_j-\lambda)^{m}} Q_j^m + \frac{1}{t_j-\lambda} Q_j + \sum_{k=2}^{r_j} \frac{(-1)^{k}}{(t_j-\lambda)^{k}} Q_j^k \right] =$$

$$= \sum_{j=1}^{n} P_j + \sum_{j=1}^{n} \left[-\sum_{k=1}^{r_j} \frac{(-1)^{k}}{(t_j-\lambda)^{k}} Q_j^k + \sum_{k=1}^{r_j} \frac{(-1)^{k}}{(t_j-\lambda)^{k}} Q_j^k \right] =$$

$$= I + \sum_{j=1}^{n} \frac{(-1)^{r_j}}{(t_j - \lambda)^{r_j}} Q_j^{r_j} = I$$

because $Q_j^{r_j} = 0$ for $j = 1, ..., n$ (cf. Formula(5.14)). In similar way we prove that $B(A - \lambda I) = I$ for $\lambda \neq t_1, ..., t_n$. We therefore conclude that the operator $A - \lambda I$ is invertible for $\lambda \neq t_1, ..., t_n$ and that $(A - \lambda I)^{-1} = B$, where the operator B is defined by Formula (5.21). ■

Theorem 5.5. *Suppose that $A \in L_0(X)$ is an algebraic operator with the characteristic polynomial $P(t) = \prod_{j=1}^{n} (t - t_j)^{r_j}$. Decompose the rational function $1/P(t)$ onto vulgar fractions:*

$$\frac{1}{P(t)} = \sum_{j=1}^{n} \sum_{k=0}^{r_j - 1} \frac{d_{jk}}{k!} \frac{1}{(t - t_j)^{r_j - k}}, \tag{5.22}$$

where

$$d_{jk} = \left\{ \frac{\mathrm{d}^k}{\mathrm{d}^k} [p_j(t)]^{-1} \right\}_{t = t_j}; \quad p_j(t) = (t - t_j)^{-r_j} P(t) \tag{5.23}$$

$$(j = 1, ..., n; \ k = 0, ..., r_j - 1).$$

Write

$$\mathbf{p}_j'(t(= p_j(t) \sum_{k=0}^{r_j - 1} \frac{d_{jk}}{k!} (t - t_j)^k \quad (j = 1, ..., n). \tag{5.24}$$

Then

$$P_j = \mathbf{p}_j'(A) \qquad (j = 1, ..., n), \tag{5.25}$$

where $P_1, ..., P_n$ are projectors defined in Theorem 5.1.

Proof. The definition of the polynomials $\mathbf{p}_j'(t)$ and Formulae (5.22) together imply that

$$\sum_{j=1}^{n} \mathbf{p}_j'(t) = \sum_{j=1}^{n} p_j(t) \sum_{k=0}^{r_j - 1} \frac{d_{jk}}{k!} (t - t_j)^k =$$

$$= \sum_{j=1}^{n} (t - t_j)^{-r_j} P(t) \sum_{k=0}^{r_j - 1} \frac{d_{jk}}{k!} (t - t_j)^k =$$

$$= P(t) \sum_{k=0}^{r_j - 1} \frac{d_{jk}}{k!} \frac{1}{(t - t_j)^{r_j - k}} = P(t) \frac{1}{P(t)} = 1.$$

Hence

$$\sum_{j=1}^{n} \mathbf{p}_j'(A) = I. \tag{5.26}$$

Moreover,

$$\mathbf{p}_j'(t)\mathbf{p}_m'(t) = (t-t_j)^{-r_j}P(t)(t-t_m)^{-r_m}P(t) = \Big[\prod_{\substack{\mu=1\\ \mu\neq j,\ \mu\neq m}}^{n} (t-t_\mu)^{r_\mu}\Big]P(t).$$

Then for $m \neq j$ we find

$$\mathbf{p}_j'(A)\mathbf{p}_m'(A) = \Big[\sum_{k=0}^{r_j-1} \frac{d_{jk}}{k!}(A-t_jI)^k\Big]\Big[\sum_{m=0}^{r_j-1} \frac{d_{m\mu}}{\mu!}(A-t_mI)^\mu\Big]\Big[\prod_{\substack{\mu=1\\ \mu\neq j,\ \mu\neq m}}^{n} (A-t_\mu)^{r_\mu}\Big]P(A) = 0.$$

This, and Formula (5.26) together imply that for $m = 1, ..., n$ we get

$$\mathbf{p}_m'(A) = \mathbf{p}_m'(A)\sum_{j=1}^{n} \mathbf{p}_j'(A) = \sum_{j=1}^{n} \mathbf{p}_m'\mathbf{p}_j'(A) = [\mathbf{p}_m'(A)]^2.$$

Since

$$(t-t_j)^{r_j}\mathbf{p}_j(t) = (t-t_j)^{r_j}p_j(t)\sum_{k=0}^{r_j-1} \frac{d_{jk}}{k!}(t-t_j)^k =$$

$$= (t-t_j)^{r_j}(t-t_j)^{-r_j}P(t)\sum_{k=0}^{r_j-1} \frac{d_{jk}}{k!}(t-t_j)^k =$$

$$= P(t)\sum_{k=0}^{r_j-1} \frac{d_{jk}}{k!}(t-t_j)^k,$$

we have for $j = 1, ..., n$

$$(A-t_jI)^{r_j}\mathbf{p}_j'(A) = P(A)\sum_{k=0}^{r_j-1} \frac{d_{jk}}{k!}(A-t_jI)^k = 0.$$

We therefore have proved that the operators $\mathbf{p}_j'(A)$ have the following properties:

$$\sum_{j=1}^{n} \mathbf{p}_j'(A) = I, \quad [\mathbf{p}_j'(A)]^2 = \mathbf{p}_j'(A), \quad (j, m = 1, 2, ..., m)$$

$$\mathbf{p}_j'(A)\mathbf{p}_m'(A) = 0 \quad for \quad j \neq 0, \quad (A-t_jI)^{r_j}\mathbf{p}_j'(A) = 0.$$

This, and Theorem 5.1 together imply that $P_j = \mathbf{p}'_j(A)$, where $P_j = p(A)$ $(j = 1, ..., n)$ and the polynomials $\mathbf{p}_j(t)$ are defined by Formulae (5.6). ■

Note that Theorems 5.3, 5.4 and Property 5.2 for multiple characteristic roots of an algebraic operator, also Theorem 5.5, have been proved by M. Tasche (cf. T[1]) in Banach space by means of analytic methods. Other properties of algebraic operators, their applications and references can be found in the author's books (cf. PR[3], PR[4]), also in the monograph of the author and S. Rolewicz (cf. PRR[1]) and the recent book of Nguyen Van Mao (cf. N[1]).

Let X be a pre-Hilbert space with an inner product $\langle x, y\rangle$. An operator A' is said to be an *adjoint operator* for an operator $A \in L_0(X)$ in the sense of that inner product if it satisfies the identity

$$(5.27) \qquad \langle Ax, y\rangle = \langle x, A'y\rangle \quad \textit{for all} \quad x, y \in X.$$

The operator A' is *antilinear*, i.e. $A'(\lambda x) = \overline{\lambda}(Ax)$ for all scalars λ and $x \in X$ (cf.Exercise 4.12). It means that in a real space X both notions, a conjugate operator in the sense of functionals (Formula (2.19)) and an adjoint operator in a sense of an inner product (Formula (5.27)), are identical.

An operator $A \in L_0(X)$ is said to be *self-adjoint* if $A' = A$. In the Euclidean space $\mathbb{E}^n$ to a self-adjoint linear operator A there corresponds a *symmetric* matrix, i.e. a matrix A such that $A^T = A$. (cf. Exercise 2.8).

Again, assume that X is a pre-Hilbert space with an inner product $\langle x, y\rangle$. Then an operator $A \in L_0(X)$ is said to be *unitary* if it is invertible and

$$(5.28) \qquad A'A = AA' = I.$$

In other words: *an operator $A \in L_0(X)$ is unitary if it is invertible and*

$$(5.29) \qquad A' = A^{-1}.$$

If X is an n-dimensional linear space over the field $\mathbb{R}$ (over the field $\mathbb{C}$, respectively) and for an $A \in L_0(X)$ the matrix A is non-singular then this matrix is said to be *orthogonal* (otherwise: *unitary*) if

$$(5.30) \qquad A^T = A^{-1} \qquad \overline{A}^T = A^{-1}, \ \textit{respectively},$$

where $A = [a_{j,k}]_{j,k=1,...,n}$, $\overline{A} = [\overline{a}_{j,k}]_{j,k=1,...,n}$.

Consequently, in the Euclidean space $\mathbb{E}^n$ (in the space $\mathbb{C}^n$, respectively) to a unitary operator there corresponds an orthogonal (unitary) matrix.

Theorem 5.6. *Suppose that X is a pre-Hilbert space and an operator $A \in L_0(X)$ is unitary. Then for all $x, y \in X$*

(i) $\langle Ax, Ay\rangle = \langle x, y\rangle$;

(ii) if $x \perp y$ *then* $Ax \perp Ay$;

(iii) $\|Ax\| = \|x\|$;

(iv) if λ *is an eigenvalue of* A *then* $|\lambda| = 1$.

Proof. Let x, y be arbitrary elements of the space X. Formulae (5.27) and (5.28) together imply that

$$\langle Ax, Ay\rangle = \langle x, A'Ay\rangle = \langle x, y\rangle.$$

If $x \perp y$ then $\langle x, y\rangle = 0$. This implies that $\langle Ax, Ay\rangle = \langle x, y\rangle = 0$, i.e. $Ax \perp Ay$. By Point (i) of this theorem,

$$\|Ax\|^2 = \langle Ax, Ay\rangle = \langle x, y\rangle = \|x\|^2.$$

If λ is an eigenvalue of the operator A and x is an eigenvector corresponding to this eigenvalue then, by definition, $x \neq 0$ and $Ax = \lambda x$. This, and Point (iii) together imply that

$$\|x\|^2 = \|Ax\|^2 = \langle Ax, Ax\rangle = \langle \lambda x, \lambda x\rangle = \lambda\overline{\lambda}\langle x, x\rangle = |\lambda|^2\|x\|^2.$$

Hence $|\lambda| = 1$. ■

Corollary 5.6. *If A is a unitary matrix in the space $\mathbb{C}^n$ with the inner product* *)

$$\langle x, y\rangle = \sum_{j=1}^{n} x_j\overline{y}_j,$$

then Points (i), (ii), (iii), (iv) of Theorem 5.6 hold for arbitrary $x, y \in \mathbb{C}^n$ with that

$$\|x\|^2 = |x|^2 = \sum_{j=1}^{n} x_j^2 \quad for \quad x = (x_1, ..., x_n),\ y = (y_1, ..., y_n) \in \mathbb{C}^n.$$

Moreover,

(v) $|\det\ A| = 1$;

(vi) if $a^j = (a_{j1}, ..., a_{jn})$, $a_k = (a_{1k}, ..., a_{nk})$ $(j, k = 1, ..., n)$ denote columns and rows of the matrix A, respectively, then

$$\langle a^i, \overline{a}^j\rangle = \delta_{ij}, \qquad \langle a_i, \overline{a}_k\rangle = \delta_{ik},$$

where $\overline{a}^j = (\overline{a}_{j1}, ..., \overline{a}_{jn})$, $\overline{a}_k = (\overline{a}_{1k}, ..., \overline{a}_{nk})$ $(j, k = 1, ..., n)$.

Proof. (v) By our assumption A, is a unitary matrix, hence it corresponds to a unitary operator in the space $\mathbb{C}^n$. This immediately implies the conclusions of Theorem 5.6,

*) cf. Exercise 4.12.

Points (i), (ii), (iii), (iv). Moreover, Formula (5.30) and the Cauchy Theorem (Theorem 3.2) together imply that

$$|\det\ A|^2 = (\det\ A)(\overline{\det\ A}) = (\det\ A)(\det\ \overline{A}) =$$

$$= (\det\ A)[\det\ (A^{-1})^T] = (\det\ A)(\det\ A^{-1}) = \det\ (AA^{-1}) = \det\ I = 1.$$

(vi) Suppose now that a^i and a^j are columns of the matrix A $(i,j = 1,...,n)$. By our assumption, $A\overline{A^T} = I$. Hence entries of the matrix $A\overline{A^T}$ are of the form

$$c_{ji} = \sum_{k=1}^{n} a_{jk}\overline{a}_{ik} = \begin{cases} 1 & \text{for } i = j; \\ 0 & \text{for } i \neq j \end{cases} \qquad (i,j = 1,...,n)$$

for $\overline{A^T} = [\overline{a}_{j,k}]^T_{j,k=1,...,n} = [\overline{a}_{ik}]_{k,i=1,...,n}$. Then

$$\langle a^i, \overline{a^j}\rangle = \sum_{k=1}^{n} a_{jk}\overline{a}_{ik} = c_{ji} = \delta_{ij} \quad (i,j = 1,...,n).$$

A similar proof holds for rows of the matrix A. ■

Corollary 5.7. *If A is an orthogonal matrix in the Euclidean space $\mathbb{E}^n$ with an inner product*

$$\langle x, y\rangle = \sum_{j=1}^{n} x_j y_j \quad \textit{for} \quad x = (x_1,...,x_n),\ y = (y_1,...,y_n) \in \mathbb{E}^n$$

then Points (i), (ii), (iii) of Theorem 5.6 are satisfied for arbitrary $x, y \in \mathbb{E}^n$. Moreover,

(iv') if λ is an eigenvalue of the matrix A then either $\lambda = 1$ or $\lambda = -1$;

(v') either $\det\ A = 1$ *or* $\det\ A = -1$*;*

(vi') if $a^j = (a_{j1},...,a_{jn})$, $a_k = (a_{1k},...,a_{nk})$ $(j,k = 1,...,n)$ are columns and rows of the matrix A, respectively, then

$$\langle a^i, a^j\rangle = \delta_{ij}, \quad \langle a_i, a_k\rangle = \delta_{ik} \quad (i,j,k = 1,...,n), \tag{5.30}$$

i.e. two different columns (two different rows) of an orthogonal matrix are orthogonal each to other.

Proof. In order to prove Point (iv'), it is enough to observe that now $\lambda \in \mathbb{R}$. Then $\overline{\lambda} = \lambda$ and $\lambda\overline{\lambda} = \lambda^2$. Hence the equality $\lambda^2 = 1$ implies that $\lambda = \pm 1$.

In order to prove Point (v'), observe that every orthogonal matrix satisfies the first condition of (5.30) and that $\overline{\det\ A} = \det\ A$. Hence

$$|\det\ A|^2 = (\det\ A)(\det\ A) = (\det\ A)(\det\ A^T) = \det\ (AA^T) = \det\ I = 1.$$

This implies that $\det\ A = \pm 1$.

In order to prove Point (vi'), it is enough to observe that now $\overline{a^j} = a^j$, $\overline{a_k} = a_k$ $(j, k = 1, ..., n)$. ■

Corollary 5.8. *If a matrix $A \in L_0(\mathbb{C}^n)$ is symmetric (i.e. $A^T = A$) then all its eigenvalues are real.*

Proof. Indeed, if $x \neq 0$ is an eigenvector of the matrix A corresponding to an eigenvalue λ, $y \neq 0$ is an eigenvector of the matrix A^T corresponding to an eigenvalue μ and $\langle x, y\rangle \neq 0$, then

$$\langle A^T y, x\rangle = \langle \mu y, x\rangle = \mu\langle y, x\rangle = \mu\overline{\langle x, y\rangle} \qquad \textit{and}$$

$$\langle A^T y, x\rangle = \overline{\langle x, A^T y\rangle} = \overline{\langle Ax, y\rangle} = \overline{\langle \lambda x, y\rangle} = \overline{\lambda}\ \overline{\langle x, y\rangle}.$$

This implies that $\mu = \overline{\lambda}$. But $A^T = A$, hence eigenvalues and eigenvectors of these matrices are the same. Then, admitting $y = x$, $\mu = \lambda$, we get $\lambda = \mu = \overline{\lambda}$, i.e. $\lambda \in \mathbb{R}$. ■

Corollary 5.9. *If the matrix $A \in L_0(\mathbb{E}^n)$ is symmetric then all its eigenvalues are real (in the extension of $\mathbb{E}^n$ in $\mathbb{C}^n$).*

Proof. We proceed as in the proof of Corollary 5.8, so that the following equalities are obtained:

$$\langle A^T y, x\rangle = \langle \mu y, x\rangle = \mu\langle x, y\rangle,$$

$$\langle A^T x, y\rangle = \langle x, A^T y\rangle = \langle Ax, y\rangle = \langle \lambda x, y\rangle = \lambda\langle x, y\rangle.$$

Hence $\mu = \lambda$. ■

Corollary 5.10. *If the matrix $A \in L_0(\mathbb{E}^n)$ is symmetric and x, y are eigenvectors of A corresponding to its eigenvalues λ, μ respectively, with $\mu \neq \lambda$, then $\langle x, y\rangle = 0$.*

Proof. By our assumptions, $Ax = \lambda z$, $Ay = \mu y$ and $A^T = A$, $\lambda - \mu \neq 0$. This implies that

$$\langle x, y\rangle = \frac{1}{\lambda - \mu}(\lambda - \mu)\langle x, y\rangle = \frac{1}{\lambda - \mu}\big[\lambda\langle x, y\rangle - \mu\langle x, y\rangle\big] =$$

$$= \frac{1}{\lambda - \mu}\big[\langle \lambda x, y\rangle - \langle x, \mu y\rangle\big] = \frac{1}{\lambda - \mu}\big[\langle Ax, y\rangle - \langle x, A^T y\rangle\big] =$$

$$= \frac{1}{\lambda - \mu}\big[\langle Ax, y\rangle - \langle Ax, y\rangle\big] = 0.$$

■

Examples and Exercises.

Example 5.1. Every finite dimensional operator $A \in L_0(X)$, where $\dim X = +\infty$, is an algebraic operator.

Indeed, by definition, the space $Y = AX$ is finite dimensional. Denote by A_0 a matrix corresponding to the restriction of the operator A to the space $Y = AX \subset X$. Then, by the Cayley-Hamilton theorem (Corollary 5.3), the operator A_0 corresponding to that matrix is algebraic and its characteristic polynomial $P(\lambda)$ is a divisor of the polynomial det $(A_0 - \lambda I)$. Hence the characteristic polynomial of the operator A is of the form $\lambda P(\lambda)$ for $AP(A)X = P(A)AX = P(A_0)Y = 0$.

Observe that $\lambda = 0$ is a characteristic root of the operator A. This operator is not invertible, because the equation $Ax = 0$ has a non-trivial solution of the form $x = P(A)y$, where $y \in X \setminus AX \neq \{0\}$. □

Example 5.2. An operator $A \in L_0(X)$ is said to be an *involution of order* n if $A^n = I$ and $0 \neq A^k \neq I$ for $k = 1, ..., n-1$ $(n \in \mathbb{N})$. If $n = 2$ then A is said to be an *involution*. By definition, every involution of order n is an algebraic operator with the characteristic polynomial $P(t) = t^n - 1$ and with characteristic roots $1, \varepsilon, ..., \varepsilon^{n-1}$, where $\varepsilon = \mathrm{e}^{2\pi i/n} = \cos\, \frac{2\pi}{n} + i \sin\, \frac{2\pi i}{n}$. □

Exercise 5.1. Suppose that $A \in L_0(X)$ is an involution of order n. Prove that projectors P_k defined in Theorem 5.1 are of the form

$$P_k = \frac{1}{n} \sum_{j=1}^{n} \varepsilon^{-jk} A^j, \quad \textit{where} \quad \varepsilon = \mathrm{e}^{2\pi i/n} \quad (k = 1, ..., n). \tag{5.32}$$

If dim $X < +\infty$ then determine Jordan matrices of the matrix A. In particular, show that for an involution we have

$$P_1 = \frac{1}{2}(I - A), \quad P_2 = \frac{1}{2}(I + A)$$

(cf. PR[1], PRR[1]). □

Example 5.3. Let X be a linear space of functions determined for $t \in \mathbb{R}$. The *reflection* S of a function $x \in X$ is defined by means of the equality: $(Sx)(t) = x(-t)$ for $t \in \mathbb{R}$. By definition, $S^2 = I$, i.e. S is an involution. By Theorem 5.1, $X = X^+ \oplus X^-$, where X^+ is the space of all *even* functions, i.e. functions $x \in X$ which satisfy the condition $x(-t) = x(t)$, and X^- is the space of all *odd* functions , i.e. functions $x \in X$ which satisfy the condition $x(-t) = -x(t)$. Then every function $x \in X$ can be written in a unique way as the sum of an even function and an odd function. Namely, $x = x^+ + x^-$, where

$$x^+(t) = \frac{1}{2}[x(t) + x(-t)], \quad x^-(t) = \frac{1}{2}[x(t) - x(-t)] \quad (t \in \mathbb{R}).$$

□

Exercise 5.2. Let X be the space of all square matrices of dimension n with real entries. Prove that

(i) the transposition of matrices is an involution;

(ii) every matrix can be written in a unique way as the sum of a symmetric and antisymmetric matrix (cf. Exercise 2.8). Recall that a square matrix $A = [a_{jk}]_{j.k=1,...,n}$ is symmetric if $A^T = A$, i.e. $a_{kj} = a_{jk}$ $(j, k = 1, ..., n)$, and antisymmetric if $A^T = -A$, i.e. $a_{kj} = -a_{jk}$ $(j, k = 1, ..., n)$. □

Exercise 5.3. Let X be either real or complex linear space of functions determined on the plane of a complex variable. Let $S \in L_0(X)$ be the operator of a *rotation* through the angle $2\pi/N$, where N is a positive integer greater than 1, i.e. $(Sx)(t) = x(\mathrm{e}^{\frac{2\pi i}{N}} t)$ for $t \in \mathbb{C}$, $x \in X$. Prove that S is an involution of order N and determine components given by the decomposition of X onto the direct sum determined in Theorem 5.1. □

Exercise 5.4. Let X_ω be a linear space of all *periodic* functions with the period ω determined for $t \in \mathbb{R}$. Prove that the *shift operator* defined by means of the equality: $(Sx)(t) = x(t - \frac{\omega}{N})$, where N is an arbitrarily fixed positive integer greater than one, is an involution of order N. Determine components given by the decomposition of X onto the direct sum determined in Theorem 5.1. □

Exercise 5.5. Suppose that X is an arbitrary linear space over the field $\mathbb{C}$, operators $A, B \in L_0(X)$ and A is an algebraic operator. Prove that a complex number $\lambda \neq 0$ such that $AB - BA = \lambda I$ does not exist (cf. PRR[1], p. 71).

Corollary. Square matrices A and B (of the same dimension) such that $AB - BA = \lambda I$ for $\lambda \in \mathbb{C} \setminus \{0\}$ do not exist. □

Exercise 5.6. Consider linear spaces $\mathbb{F}_n[t]$ of all polynomials of order n with coefficients in $\mathbb{F}$ $(n \in \mathbb{N})$ (cf. Example 1.5). For an arbitrary polynomial $p(t) = \sum_{k=0}^{n} p_k t^k$, where $p_k \in \mathbb{F}$, define an operator D by means of the equality

$$(Dp)(t) = \sum_{k=1}^{n} k p_k t^{k-1}.$$

Prove that D is a nilpotent *) operator of order $n+1$ on the space $\mathbb{F}^n[t]$: $D^{n+1} = 0$, $D^n \neq 0$, i.e. D is an algebraic operator with a unique characteristic root zero of multiplicity $n+1$. □

Exercise 5.7. Prove that $\lambda \in \mathbb{F}$ is a regular value of the operator A^n $(n \in \mathbb{N})$ if and only if the nth roots of λ are regular values of the operator $A \in L_0(X)$. □

Exercise 5.8. Suppose that $P \in L_0(X)$ is a projector, i.e. $P^2 = P$.

(i) Determine characteristic roots of the operator P;

(ii) determine Jordan matrices of P when $\dim X < +\infty$. □

Exercise 5.9. Suppose that $B \in L_0(X)$, where $\dim X < +\infty$, is a nilpotent operator of order n (cf. Exercise 5.6). Prove that

(i) if $\mathbb{F} = \mathbb{C}$ then $B\overline{B}^T \neq I$;

(ii) if $\mathbb{F} = \mathbb{R}$ then $BB^T \neq I$, i.e. a nilpotent matrix is neither unitary nor orthogonal.

Can a nilpotent matrix be either symmetric or antisymmetric? □

*) (cf. Exercise 2.13.)

Exercise 5.10. Prove that orthogonal matrices of dimension n in $\mathbb{E}^n$ form a group. $\square$

Exercise 5.11. Prove that unitary matrices of dimension n form in $\mathbb{C}^n$ a group. $\square$

Exercise 5.12. Let $A \in L_0(\mathbb{C}^n)$ be unitary. Prove that the matrix A^T is also unitary. $\square$

Exercise 5.13. Let det $A \neq 0$. Prove that the matrix A^{-1} (which exists by our assumption) has as eigenvalues the inverses of eigenvalues of the matrix A. $\square$

Exercise 5.14. Let det $B \neq 0$. Prove that the eigenvalues of the matrix BAB^{-1} are equal to the eigenvalues of the matrix $A \in L_0(X)$. $\square$

Exercise 5.15. Prove that for an arbitrary polynomial $Q(t)$ with coefficients in $\mathbb{F}$ the operator $Q(A)$ has eigenvalues $Q(\lambda)$, where λ is an arbitrary eigenvalue of A. $\square$

Exercise 5.16. Prove that properties proved in Exercises 5.13, 5.14 and 5.15 are true for characteristic roots of algebraic operators. $\square$

Exercise 5.17. Formulate and prove Corollary 5.10 for matrices $A \in L_0(\mathbb{C}^n)$. (Hint: A matrix is said to be *Hermitian* (otherwise: *Hermite matrix*) if $a_{kj} = \overline{a}_{jk}$, i.e. if $A^T = \overline{A}$, where $\overline{A} = [\overline{a}_{jk}]_{j,k=1,\dots,n}$. Hence in $\mathbb{E}^n$ Hermite matrices are symmetric). Prove that to a selfadjoint operator $A \in L_0(\mathbb{C}^n)$ (i.e. such that $A' = A$) there corresponds a Hermite matrix. $\square$

Exercise 5.18. Prove that

(i) orthogonal operators in $\mathbb{E}^n$,

(ii) unitary operators in $\mathbb{C}^n$

form a group. $\square$

Exercise 5.18. Prove that rotations in $\mathbb{E}^2$ are orthogonal operators. $\square$

Chapter 6.

Quadratic forms.

Let X be an arbitrary linear space over the field $\mathbb{R}$. A function $f(x, y)$ of variables $x, y \in X$ and with values in the field $\mathbb{R}$ is said to be *bilinear* if $f(x, y)$ with a fixed x is a linear functional of the variable y and $f(x, y)$ with a fixed y is a linear functional of the variable x. A bilinear functional f is said to be *symmetric* if $f(y, x) = f(x, y)$ for all $x, y \in X$.

A symmetric functional with identified variables, i.e. with $y = x$, is said to be a *quadratic functional*. This definition implies that every quadratic functional is of the form $f(x, x)$ where $f(x, y)$ is a symmetric bilinear functional. In other words: Any quadratic functional is obtained from a symmetric bilinear functional by the identification of variables x and y.

If dim $X < +\infty$ then traditionally linear functionals are called *linear forms*, bilinear functionals are called *bilinear forms* and quadratic functionals are called *quadratic forms*.

In this Chapter we shall consider above all Euclidean spaces $\mathbb{E}^n$. We already have proved that every linear functional in the space $\mathbb{E}^n$ can be represented in the form $f(x) = \langle x, a\rangle$, where a is a fixed element of the space $\mathbb{E}^n$ (cf. Property 4.4). Hence every linear form in the Euclidean space $\mathbb{E}^n$ can be represented as follow:

$$f(x) = \langle x, a\rangle = \sum_{j=1}^{n} a_j x_j,$$

where $a = (a_1, ..., a_n) \in \mathbb{E}^n$ is fixed.

It is not difficult to verify that every bilinear form in the Euclidean space $\mathbb{E}^n$ can be represented in the following way:

$$f_A(x, y) = \langle x, Ay\rangle = \sum_{k=1}^{n} \sum_{j=1}^{n} a_{jk} x_k y_j, \tag{6.1}$$

where $A = [a_{jk}]_{j,k=1,...,n}$. Hence every quadratic form can be represented in the following way:

$$f_A(x, x) = \langle x, Ax\rangle = \sum_{k=1}^{n} \sum_{j=1}^{n} a_{jk} x_k x_j, \tag{6.2}$$

where $A = [a_{jk}]_{j,k=1,...,n}$ and $A^T = A$, i.e. $a_{kj} = a_{jk}$ for $j, k = 1, ..., n$.

By the *rank* of a quadratic form f_A we shall mean the rank of the matrix A, i.e. if we denote by $r(f_A)$ the rank of the form f_a then

$$r(f_A) = r(A). \tag{6.3}$$

Let f_A be a quadratic form. Then the determinant of the matrix A is called the *discriminant* of f_A. If we denote the discriminant by $\Delta(f_A)$ then, by definition,

$$\Delta(f_A) = \det\ A. \tag{6.4}$$

Discriminants of quadratic forms and discriminants of quadratic equations should be distinguished. Namely, if we consider a quadratic equation $ax^2+bx+c=0$ then its discriminant is, by definition, the number $\Delta = b^2 - 4ac$. A bilinear form corresponding to a trinomial $ax^2 + bx + c$ is the form $f_A(x,y) = ax^2 + bxy + cy^2$. Hence

$$A = \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}$$

and the discriminant of the form f_A is

$$\Delta(f_A) = \det\ A = ac - \frac{b^2}{4} = -4(b^2 - 4ac) = -4\Delta.$$

Then the discriminants $\Delta(f_A)$ and Δ are equal if and only if $b^2 - 4ac = 0$, i.e. if

$$f_A(x,y) = \left(\sqrt{a}\ x + \sqrt{c}\ y\right)^2.$$

A quadratic form is *positive* (*negative*) *definite* if

$$f_A(x,x) > 0 \qquad (f_A(x,x) < 0,\ \textit{respectively}) \quad \textit{for every}\ x \neq 0. \tag{6.5}$$

If Condition (6.5) is satisfied then the matrix A is said to be *positive* (*negative*) *definite.*

Positive definite quadratic forms are said to be *Hermite forms* (otherwise: *Hermitian forms*).

A quadratic form is said to be *canonical* if

$$f_A(x,x) = \sum_{j=1}^{n} a_j x_j^2. \tag{6.6}$$

By this definition, it follows that a quadratic form f_A is canonical if and only if its matrix A is diagonal with the principal diagonal $(a_1, ..., a_n)$.

Theorem 6.1. *If $A \in L_0(\mathbb{E}^n)$ is a symmetric matrix and $B \in L_0(\mathbb{E}^n)$ is an orthogonal matrix then the mapping defined by the equality*

$$y = Bx \qquad for \quad x \in \mathbb{E}^n \tag{6.7}$$

carries the quadratic form f_A into the quadratic form f_{BAB^T}.

Proof. By our assumptions, $A^T = A$ and $B^T B = BB^T = I$. Observe that the matrix BAB^T is also symmetric. Indeed, $(BAB^T)^T = (B^T)^T A^T B^T = BAB^T$. Moreover, $x = B^T Bx = B^T y$. Then

$$f_A(x,x) = \langle x, Ax\rangle = \langle B^T y, AB^T y\rangle = \langle BB^T y, BAB^T y\rangle =$$

$$= \langle y, BAB^T y\rangle = f_{BAB^T}(y,y).$$

■

Theorem 6.2. *The characteristic polynomial of a symmetric matrix is an* invariant of orthogonal mappings, *i.e. if $A \in L_0(\mathbb{E}^n)$ is a symmetric matrix and $B \in L_0(\mathbb{E}^n)$ is an orthogonal matrix then*

$$\det\,(BAB^T - \lambda I) = \det\,(A - \lambda I).$$

Proof. By our assumption, $BB^T = I$, hence $\det\ BB^T = 1$. This, and Corollary 3.1 together imply that

$$\det\,(BAB^T - \lambda I) = \det\,(BAB^T - \lambda BB^T) = \det\ B(A - \lambda I)B^T =$$

$$= (\det\ B)[\det\,(A - \lambda I)](\det\ B^T) = (\det\ B)(\det\ B^T)\det\,(A - \lambda I) =$$

$$= \det\,(BB^T)\det\,(A - \lambda I) = \det\,(A - \lambda I).$$

■

The following corollaries are immediate consequences of Theorem 6.2.

Corollary 6.1. *Characteristic roots of a symmetric matrix in the Euclidean space $\mathbb{E}^n$ are invariants of orthogonal mappings.*

Corollary 6.2. *If f_A is a quadratic form in the Euclidean space $\mathbb{E}^n$ then characteristic roots of the matrix A are invariants of orthogonal mappings.*

Corollary 6.3. *If f_A is a quadratic form in the Euclidean space $\mathbb{E}^n$ then the discriminant of the matrix A is an invariant of orthogonal mappings, i.e.*

$$\Delta(f_{BAB^T}) = \Delta(f_A) \tag{6.8}$$

for an arbitrary orthogonal matrix B.

Proof. Indeed, by Theorem 6.2, for $\lambda = 0$ we find

$$\Delta(f_{BAB^T}) = \det\ BAB^T = \det\ A = \Delta(f_A).$$

■

Theorem 6.3. *Every quadratic form f_A in the Euclidean space $\mathbb{E}^n$ can be reduced to the canonical quadratic form*

$$f_A(x,x) = f_{\tilde{A}}(y,y) = \sum_{j=1}^{n} \lambda_j y_j^2,$$

where $\lambda_1, .., \lambda_n$ are characteristic roots of the symmetric matrix $A \in L_0(\mathbb{E}^n)$; each multiple root appears here so many times as it is its multiplicity, by an orthogonal transformation $y = Bx$ of the basis in $\mathbb{E}^n$, with the linear operator B whose matrix satisfies the equality: $BAB^T = \tilde{A} = [\lambda_j \delta_{jk}]_{j,k=1,...,n}$.

Proof *). Let $\mathcal{B}$ be an orthonormal basis in $\mathbb{E}^n$. By the Cayley-Hamilton theorem (Corollary 5.3), A is an algebraic operator. Hence A has a finite number of eigenvalues $\lambda_1, ..., \lambda_n$ (each one counted as many times as it its multiplicity), which are characteristic roots of the operator A. Since the matrix A is symmetric, these roots are real.
Let X_j be the eigenspace of the operator A corresponding to the eigenvalue λ_j $(j = 1, ..., n)$ and let $k_j = \dim\ X_j \le n$. Then there are k_j linearly independent eigenvectors $e_m^{(j)}$ of the operator A which constitute a basis $\mathcal{B}^{(j)}$ in the space X_j. Without any loss of generality, we may assume that this basis is orthonormal. This , and Corollary 5.10 together imply that

$$\langle e_m^{(i)}, e_l^{(i)} \rangle = 0 \quad if\ either \quad j \ne i \quad or \quad l \ne m,$$

$$(k_1 + ... + k_n = n;\ m = 1, ..., k_j;\ l = 1, ..., i,\ i, j = 1, ..., n).$$

We therefore conclude that the spaces X_k and X_j are orthogonal to each other if $k \ne j$ $(j, k = 1, ..., n)$. Hence the set

$$\mathcal{B}_1 = \{e_1^{(1)}, ..., e_{k_1}^{(1)}, ..., e_1^{(n)}, ..., e_{k_n}^{(n)}\} \quad (k_1 + ... + k_n = n)$$

is an orthonormal basis in the space X. But in the basis $\mathcal{B}_1$ to the operator A there corresponds the diagonal matrix $\tilde{A} = [\lambda_j \delta_{jk}]_{j,k=1,...,n}$. Indeed, the definition of $\mathcal{B}_1$ implies

$$\tilde{A} e_m^{(j)} = \lambda_j e_m^{(j)} \qquad (m = 1, ..., k_j;\ j = 1, ..., n).$$

Denote by $B = [b_{jk}]_{j,k=1,...,n}$ the matrix of a one-to-one linear operator transforming the basis $\mathcal{B}$ into the basis $\mathcal{B}_1$. We shall show that the matrix B is orthogonal. Indeed, by our assumptions, for $g_j, g_k \in \mathcal{B}_1$ $(j, k = 1, ..., n)$ we have

$$\delta_{jk} = \langle g_j, g_k \rangle = \langle Be_j, Be_k \rangle = \langle \sum_{l=1}^{n} b_{jl} e_l, \sum_{m=1}^{n} b_{km} e_m \rangle =$$

*) This proof was never published in English (cf. PR[2]).

$$= \sum_{l,m=1}^{n} b_{jl} b_{km} \langle e_j, e_m \rangle = \sum_{m=1}^{n} b_{jm} b_{km} \langle e_m, e_m \rangle = \sum_{m=1}^{n} b_{jm} b_{km}.$$

Then $B^T B = BB^T = I$ (where I is the identity matrix), which proves that B is an orthogonal matrix (cf. Corollary 5.7, Point (vi').)

Let $x \in \mathbb{E}^n$ be arbitrarily fixed. Let $y = Bx$. Then $x = B^T Bx = B^T y$. Since $BAB^T = \tilde{A}$, we obtain

$$f_A(x, x) = \langle x, Ax \rangle = \langle B^T y, AB^T y \rangle = \langle y, BAB^T y \rangle = \langle y, \tilde{A} y \rangle =$$

$$= f_{\tilde{A}}(y, y) = \sum_{j=1}^{n} \lambda_j y_j^2.$$

■

Corollary 6.4. *A quadratic form f_A is Hermitian in the Euclidean space $\mathbb{E}^n$ if and only if all characteristic roots of the matrix A are positive.*

Proof. Indeed, if λ is a characteristic root of the matrix A then it is an eigenvalue of the operator A with a corresponding eigenvector $x \neq 0$. This implies that $|x| > 0$ and $\langle x, Ax \rangle = \langle x, \lambda x \rangle = \lambda \langle x, x \rangle = \lambda |x|$. Then $\langle x, Ax \rangle > 0$ whenever $\lambda > 0$. Conversely, if $\langle x, Ax \rangle > 0$ then $\lambda > 0$. ■

An immediate consequence of Corollaries 6.2 and 6.3 is

Corollary 6.5 (Sylvester inertia law). *The number of signs "+" and "−" of coefficients of a canonical quadratic form in the Euclidean space $\mathbb{E}^n$ is constant.*

Let $a \in \mathbb{E}^n$, $A \in L_0(\mathbb{E}^n)$, be arbitrarily fixed and let $A^T = A$. Sets of the form

(6.10) $$K_{A,a} =$$

$$= \{x \in \mathbb{E}^n : f_A(x, x) + 2\langle x\} =$$

$$= \{x \in \mathbb{E}^n : \langle x, Ax + 2a \rangle = 1\}$$

are said to be *hyperquadrics*. A hyperquadric is said to be *improper* if it is either an empty set or a finite set of linear manifolds of dimension n. Otherwise, hyperquadrics are called *proper*. Clearly, a hyperquadric $K_{A,a}$ *passing through a point* x_0 has the equation

(6.11) $$\langle x - x_0, A(x - x_0) + 2a \rangle = 0.$$

Let $H_b = \{x \in \mathbb{E}^n : \langle x, b \rangle = 1,\ b \in \mathbb{E}^n\}$ be a hyperplane in the Euclidan space $\mathbb{E}^n$. We say that the hyperplane H_b *intersects a hyperquadric* $K_{A,a}$ if the set $H_b \cap K_{A,a}$ is non-empty.

A straight line $P \subset \mathbb{E}^n$ is said to be *tangent* to a hyperquadric $K_{A,a} = \{x \in \mathbb{E}^n : \langle x, Ax + 2a \rangle = 1\}$ at a point $x_0 \in K_{A,a}$ if

(i) $x_0 \in P$;

(ii) the set $P_0 = \{\langle y, Ay + 2a\rangle : y \in P\}$ of real numbers has the following property: either $p \geq 1$ for all $p \in P_0$ or $p \leq 1$ for all $p \in P_0$.

A linear manifold $Y \subset \mathbb{E}^n$ is said to be *tangent to a hyperquadric* $K_{A,a}$ at a point $x_0 \in K_{A,a}$ if every straight line $P \subset Y$ is tangent to this hyperquadric at the point x_0.

A linear manifold $Y \subset \mathbb{E}^n$ is said to be *normal to a hyperquadric* $K_{A,a}$ at a point $x_0 \in K_{A,a}$ if $Y \cap K_{A,a} = \{x_0\}$ and $\langle x, y\rangle = 0$ for all x belonging to a hyperplane tangent at the point x_0 and for all $y \in Y$.

A *unit ball* in the Euclidean space $\mathbb{E}^n$ with the center at zero is the set

$$\{x \in \mathbb{E}^n : \sum_{j=1}^{n} x_j^2 \leq 1\}.$$

A ball with the center at the point $a = (a_1, ..., a_n)$ and with the radius r is the set

$$K_{r,a} = \{x \in \mathbb{E}^n : \sum_{j=1}^{n} (x_j - a_j)^2 \leq r^2\} = \{x \in \mathbb{E}^n : |x - a| \leq r\}.$$

Sets $K_{r,a}$ in $\mathbb{E}^2$ are called *discs*.

A boundary of a ball in the Euclidean space, i.e. the set

$$S_{r,a} = \{x \in \mathbb{E}^n : |x - a| = r\},$$

is said to be a *sphere*.

A sphere in the space $\mathbb{E}^2$ is called a *circle*. By definition, a sphere $S_{1,0}$ is a hyperquadric $K_{I,0}$.

Hyperquadrics in the Euclidean space $\mathbb{E}^2$ are called either *curves of the 2nd degree* or *conics*. Hyperquadrics in the Euclidean spaces $\mathbb{E}^3$ are called either *surfaces of the 2nd degree* or *quadrics*. We shall now consider with their classification.

As follows from the general form (6.11) of equations of hyperquadrics passing through a point, curves of order two have the general equation of the form

$$a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + 2a_{13}x_1 + 2a_{23}x_2 + a_{33} = 0, \tag{6.12}$$

where $a_{21} = a_{12}$ since the matrix $A = [a_{jk}]_{j,k=1,2}$ is symmetric. Taking into account this symmetry, we also let $a_{31} = a_{13}$, $a_{32} = a_{23}$.

By Equation (6.11), the general equation of surfaces of the 2nd degree is of the form

$$a_{11}x_1^2 + +a_{22}x_2^2 + a_{33}x_3^2+ \tag{6.13}$$

$$+2a_{12}x_1x_2 + 2a_{13}x_1x_3 + 2a_{23}x_2x_3 + 2a_{14}x_1 + 2a_{24}x_2 + 2a_{34}x_3 + a_{44} = 0.$$

A surface is said to be *rotary* if it is formed by a rotation of a plane curve around an axis which is perpendicular to that plane.

A surface made by a movement of a straight line along a curve which lies on a plane non-containing this straight line, is said to be a *rectilinear surface*. Each of such straight lines is said to be a *generatrix* of the surface under consideration.

A *cone* is a surface formed in the following way: we use straight lines passing through a curve lying on a plane $P \subset \mathbb{E}^3$ and a point $x_0 \in \mathbb{E}^3 \setminus P$ (i.e. x_0 does not lie on this plane). By definition, every cone is a rectilinear surface. If the curve under consideration is a circle then a cone obtained in this way is said to be *circular*. Clearly, every cone is a rectilinear surface. If a cone is circular then it is also a rotary surface.

Every surface formed by all parallel straight lines passing through a curve which lies in a plane $P \subset \mathbb{E}^3$ non-lying on this plane (in particular, straight lines orthogonal to the plane P) is said to be a *cylinder* (otherwise: a *cylindrical surface*). By definition, every cylinder is a rectilinear surface. If the curve under consideration is a circle then a cylinder is said to be *circular*. If it is the case then a cylinder under consideration is also a rotary surface.

Curves of the 2nd degree are called *conics*, since every conic can be obtained by means of a section of a cylindrical cone by a plane.

Note that a cylindrical cone in the Euclidean space $\mathbb{E}^n$ is a rotary rectilinear surface. Its characterization will be given later.

We already know one quantity characterizing a quadratic form f_A, namely, its discriminant $\Delta(f_A) = \det\ A$. We shall introduce now new notations concerning curves of the 2nd degree. Write

$$w = \Delta(f_A) = \det\ A = \begin{vmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{vmatrix}, \tag{6.14}$$

$$W = \begin{vmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{vmatrix}, \qquad V = \begin{vmatrix} a_{22} & a_{32} \\ a_{23} & a_{33} \end{vmatrix}. \tag{6.15}$$

We shall see that the determinants w and W called *small* and *big discriminants*, respectively, fully characterize conics, and only in few particular cases of improper conics it is necessary to known the determinant V. Observe that the small discriminant w is, by Corollary 6.3, an invariant of orthogonal mappings. The same is true concerning the big discriminant W.

Traditionally, for an unification of classification, in the case when a conic is an empty set, we will call it an *imaginary conic*. Let us study now possible particular cases. Suppose that we are given a cylindrical cone S in the Euclidean space $\mathbb{E}^3$. Denote by α an angle between a generatrix and the axis of the cone S. Then we have the following cases.

1. A section of the cone S by a plane inclined to the axis of S on an angle β, $\alpha < \beta \leq \frac{\pi}{2}$ and not passing through its vertex gives an **ellipse**. Here we have $W \neq 0$, $w > 0$ and $a_{11}w > 0$. The canonical equation is

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1.$$

If $\alpha = \frac{\pi}{2}$ then $b = a$, i.e. we have a section of the cone S by a hyperplane normal to the axis of the cone and not passing through its vertex, which gives a **circle** with radius a (a particular case of ellipses).

2. A section of the cone S by a plane parallel to the axis (or inclined to the axis of S on an angle β, $\alpha < \beta \leq \frac{\pi}{2}$) S and not passing through this axis gives a **hyperbola**. Here we have $W \neq 0$ and $w < 0$. The canonical equation is

$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1.$$

A hyperbola is said to be *equiaxial* if $a = b$.

3. A section of the cone S by a plane parallel to a generatrix of S and non-passing through this generatrix gives a **parabola**. Here we have $W \neq 0$ and $w = 0$. The canonical equation is $x_2 = ax_1^2$.

These three curves: ellipse, hyperbola and parabola are the only **proper** conics. Note that only in these cases $W \neq 0$. The following conics are **improper**.

4. A section of the cone S by a plane and passing through the axis of S gives **two intersecting straight lines** (generatrices). Here we have $W = 0$ and $w < 0$. The canonical equation is

$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 0.$$

5. A section of the cone S by a plane inclined to the axis of S on an angle $\alpha < \beta \leq \frac{\pi}{2}$ and passing through its vertex gives **two imaginary straight lines crossing at a real point** $(0, 0)$. Here we have $W = 0$ and $w > 0$. The canonical equation is

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 0.$$

6. A section of the cone S by a plane tangent to a generatrix of S gives one (double) **straight line**. Here we have $W = 0$, $w = 0$ and $V = 0$. The canonical equation is

$$\left(\frac{x_1}{a} - \frac{x_2}{b}\right)^2 = 0.$$

7. A section of the cone S by a plane giving an **empty set** - an **imaginary ellipse**. Here we have $W \neq 0$, $w > 0$ and $V > 0$. The canonical equation is

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = -1.$$

8. A cone S is improper if its generatrices are parallel to its axis. An improper circular cone is a circular cylinder. If this is the case, then a section of S by a plane parallel to the axis and not passing by this axis gives **two parallel straight lines**. Here we have $W = 0$, $w = 0$ and $V < 0$. The canonical equation is

$$(\frac{x_1}{a} - \frac{x_2}{b})^2 = 1.$$

9. If $W = 0$, $w = 0$ and $V > 0$ then we have **two parallel imaginary straight lines**, i.e. an *empty set*, for its canonical equation is

$$(\frac{x_1}{a} - \frac{x_2}{b})^2 = -1.$$

Now we shall consider quadrics, i.e. surfaces of the 2nd degree in the Euclidean space $\mathbb{E}^3$. Recall that every quadric is described by Equation (6.13). Let the determinant W be defined by the Formula (6.15), i.e.

$$W = \begin{vmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{vmatrix}.$$

Write

$$\widetilde{V} = \begin{vmatrix} a_{11} & a_{21} & a_{31} & a_{41} \\ a_{12} & a_{22} & a_{32} & a_{42} \\ a_{13} & a_{23} & a_{33} & a_{43} \\ a_{14} & a_{24} & a_{34} & a_{44} \end{vmatrix}, \qquad W_1 = a_{11} + a_{22} + a_{33}, \tag{6.16}$$

$$V_1 = \begin{vmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{vmatrix}, \quad V_2 = \begin{vmatrix} a_{22} & a_{32} \\ a_{23} & a_{33} \end{vmatrix}, \quad V_3 = \begin{vmatrix} a_{11} & a_{31} \\ a_{31} & a_{33} \end{vmatrix}. \tag{6.17}$$

As for conics, for an unification of classification, we introduce *imaginary* surfaces: imaginary planes, an imaginary ellipsoid and an imaginary cone. to whose there correspond either an empty set or a linear manifold of higher dimension (cf. Points 6, 8 and 13 of the following classification). Then we have the following cases.

Let $\widetilde{V} < 0$, the rank of $\widetilde{V}$, $r(\tilde{V}) = 4$ *), $W \neq 0$, either $W_2 > 0$ or $WW_1 > 0$. In this case we have an **ellipsoid**. Its canonical equation is

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} + \frac{x_3^2}{c^2} = 1.$$

An ellipsoid is rotary if either $a = b$ or $b = c$ or $a = c$. If $a = b = c$ then it is a **ball** with radius a.

*) the rank of a determinant is, by definition, equal to the rank of its matrix.

2. Let $\widetilde{V} < 0$, $r(\widetilde{V}) = 4$, $W \neq 0$, either $W_2 \leq 0$ or $WW_1 \leq 0$. Then we have a **hyperboloid of two sheets**. Its canonical equation is

$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} - \frac{x_3^2}{c^2} = 1.$$

A hyperboloid of two sheets is rotary if $b = c$.

3. Let $\widetilde{V} > 0$, $r(\widetilde{V}) = 4$, $W \neq 0$, either $W_2 \leq 0$ or $WW_1 \leq 0$. Then we have a **hyperboloid of one sheet**. Its canonical equation is

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} - \frac{x_3^2}{c^2} = 1.$$

A hyperboloid of one sheet is a rectilinear surface. If $a = b$ then it is a rotary surface.

4. Let $\widetilde{V} \neq 0$, $r(\widetilde{V}) = 4$, $W = 0$ and let $W_2 > 0$. Then we have an **elliptic paraboloid** . Its canonical equation is

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 2x_3.$$

If $a = b$ then an elliptic paraboloid is a rotary surface.

5. Let $\widetilde{V} \neq 0$, $r(\widetilde{V}) = 4$, $W = 0$ and let $W_2 < 0$. Then we have a **hyperbolic paraboloid** . Its canonical equation is

$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 2x_3.$$

Any hyperbolic paraboloid is a rectilinear surface.

The just mentioned five surfaces: ellipsoid, hyperboloid of two sheets, hyperboloid of one sheet, elliptic hyperboloid and hyperbolic paraboloid, are the only **proper** quadrics. The following quadrics are **improper**.

6. Let $\widetilde{V} > 0$, $r(\widetilde{V}) = 4$, $W \neq 0$, $W_2 > 0$ and let $WW_1 > 0$. Then we have an **imaginary ellipsoid**, i.e. an **empty set**. Its canonical equation is

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} + \frac{x_3^2}{c^2} = -1.$$

7. Let $\widetilde{V} = 0$, $r(\widetilde{V}) = 3$, $W \neq 0$, either $W_2 \leq 0$ or $WW_1 \leq 0$. Then we have a **cone**. Its canonical equation is

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} - \frac{x_3^2}{c^2} = 0.$$

If $a = b$ then a cone is circular, i.e. it is a rotary surface.

8. Let $\widetilde{V} = 0$, $r(\widetilde{V}) = 3$, $W \neq 0$, $W_2 > 0$ and let $WW_1 > 0$. Then we have an **imaginary cone**, i.e. a **real point** , namely, $(0, 0, 0)$. Its canonical equation is

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} + \frac{x_3^2}{c^2} = 0.$$

9. Let $\widetilde{V} = 0$, $r(\widetilde{V}) = 3$, $W = 0$, $r(W) = 2$ and $W_2 > 0$. Then we have an **elliptic cylinder**. Its canonical equation is

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1.$$

An elliptic cylinder is a rectilinear surface. If $a = b$ then an elliptic cylinder is circular, i.e. it is a rotary surface.

10. Let $\widetilde{V} = 0$, $r(\widetilde{V}) = 3$, $W = 0$, $r(W) = 2$ and let $W_2 < 0$. Then we have a **hyperbolic cylinder**. Its canonical equation is

$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1.$$

A hyperbolic cylinder is a rectilinear surface. If $a = b$ then a hyperbolic cylinder is equiaxial.

11. Let $\widetilde{V} = 0$, $r(\widetilde{V}) = 3$, $W = 0$ and let $r(W) = 1$. Then we have a **parabolic cylinder**. Its canonical equation is

$$x_1^2 = 2ax_2.$$

A parabolic cylinder is a rectilinear surface.

12. Let $\widetilde{V} = 0$, $r(\widetilde{V}) = 2$, $W = 0$, $r(W) = 2$ and let $W_2 > 0$. Then we have **two intersecting planes**. Their canonical equation is

$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 0.$$

13. Let $\widetilde{V} = 0$, $r(\widetilde{V}) = 2$, $W = 0$, $r(W) = 2$ and let $W_2 < 0$. Then we have **two imaginary planes intersecting along a real straight line**. Their canonical equation is

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 0.$$

The mentioned real straight line has equations $x_1 = 0$, $x_2 = 0$.

14. Let $\widetilde{V} = 0$, $r(\widetilde{V}) = 2$, $W = 0$ and let $r(W) = 1$. Then we have **two parallel planes**. Their canonical equation is

$$x_1^2 - a^2 = 0.$$

15. Let $\widetilde{V} = 0$, $r(\widetilde{V}) = 2$, $W = 0$, $r(W) = 1$ and let $V_1 > 0$, $V_2 > 0$, $V_3 > 0$. Then we have **two parallel imaginary planes**, i.e. an **empty set**. Their canonical equation is

$$x_1^2 + a^2 = 0.$$

16. Let $\widetilde{V} = 0$, $r(\widetilde{V}) = 1$, $W = 0$, $r(W) = 1$ and let $V_1 = V_2 = V_3 = 0$. Then we have a (double) **one plane**. Its canonical equation is

$$x_3^2 = 0.$$

To summarize, we see that the following quadrics are rotary surfaces: balls, equiaxial ellipsoids, equiaxial hyperboloids of one sheet and of two sheets, circular cones and circular cylinders. The following quadrics are rectilinear: hyperboloids of one sheets, hyperbolic paraboloids, cones and elliptic, hyperbolic and parabolic cylinders.

Finally, note that every intersection of a surface of the 2nd degree by a plane is a conic either proper or improper. For instance, if we substitute either $x_1 = 0$ or $x_2 = 0$ or $x_3 = 0$ in canonical equations of quadrics, then we obtain some conics. The knowledge about these sections is important for a description of surfaces under consideration.

Exercises.

Exercise 6.1. Prove that in the Euclidean space $\mathbb{E}^n$ every bilinear form induces a quadratic form and conversely, i.e. every quadratic form induces a symmetric bilinear form. Is this theorem true for

(i) an arbitrary pre-Hilbert space over the field $\mathbb{R}$?

(ii) an arbitrary linear space over the field $\mathbb{R}$? □

Exercise 6.2. In any pre-Hilbert space X over the field $\mathbb{C}$ the inner product is *antilinear* with respect to the second variable. Indeed, by definition, $\langle y, x\rangle = \overline{\langle x, y\rangle}$ which implies

$$\langle x, \lambda y\rangle = \overline{\lambda\langle y, x\rangle} = \overline{\lambda}\langle x, y\rangle.$$

Hence one cannot determine quadratic forms in the same manner as in the real case. However, it is possible to admit the following definition: a functional is said to be *quadratic* if it results from a function $f(x, y)$ determined for $x, y \in X$ and with values in the field $\mathbb{C}$ by an identification of variables, i.e. if $y = x$. In the space $\mathbb{C}^n$ with this definition, similar theorems to Theorems 6.1, 6.2 and Corollaries 6.1, 6.2, 6.3, 6.4, 6.5 hold? □

Exercise 6.3. Write equations of a hyperplane tangent to a hyperquadric $K_{A,a}$ at a point $x_0 \in K_{A,a}$ in the Euclidean spaces $\mathbb{E}^2$, $\mathbb{E}^3$, $\mathbb{E}^n$. □

Exercise 6.4. Write equations of a hyperplane passing through a point x_0 and tangent to a hyperquadric $K_{A,a}$ in the Euclidean spaces $\mathbb{E}^2$, $\mathbb{E}^3$, $\mathbb{E}^n$. □

Exercise 6.5. Prove that a set obtained as a section by a plane of a surface of the second degree is a curve of the second degree. □

Exercise 6.6. Prove that a hyperboloid of one sheet and a hyperbolic paraboloid can be obtained by means of two families of generating straight lines skew each to another. □

Exercise 6.7. Write an equation of the geometric locus of points in the Euclidean space $\mathbb{E}^n$ whose

(i) distance from a given point is constant;

(ii) sum of distances from two given points is constant;

(iii) sum of distances from a given straight line and a point not lying on that line is constant.

Reduce the obtained equations to their canonical forms. Determine curves represented by these equations. □

Chapter 7.

Transformation groups. Representations of finite groups.

In the previous chapters we have considered transformations (otherwise called: mappings of the Euclidean spaces $\mathbb{E}^n$ of various typesI. Now we shall review systematically these transformations.

A property W of subsets of the Euclidian space $\mathbb{E}^n$ is said to be an *invariant* of a group $G(\mathbb{E}^n)$ of transformations of $\mathbb{E}^n$ onto itself if

$$(7.1) \qquad AZ \textit{ has the property } W \textit{ for every transformation } A \in G(\mathbb{E}^n)$$

$$\textit{whenever } Z \subset \mathbb{E}^n \textit{ has the property } W.$$

A transformation A in this definition does not need to be linear.

A transformation $A \in L_0(\mathbb{E}^n)$ *preserves the orientation* of n vectors in the Euclidian space $\mathbb{E}^n$ if det $A > 0$. If det $A < 0$ then A *changes the orientation* to an opposite one.

A transformation A of the Euclidean space $\mathbb{E}^n$ onto itself is said to be an *isometry* if

$$(7.2) \qquad |Ax - Ay| = |x - y| \quad \textit{for all } x, y \in \mathbb{E}^n.$$

Theorem 7.1. *A transformation $A \in L_0(\mathbb{E}^n)$ is orthogonal if and only if it is an isometry.*

Proof. Suppose that $A \in L_0(\mathbb{E}^n)$ is an orthogonal transformation, i.e. it is determined by an orthogonal matrix. Then Corollary 5.6 implies that $|Ax| = |x|$ for all $x \in \mathbb{E}^n$. Hence A is an isometry.

Conversely, suppose that $A \in L_0(\mathbb{E}^n)$ is an isometry. Then A preserves the inner product. Indeed, for all $x, y \in \mathbb{E}^n$ we have

$$\langle Ax, Ay\rangle = \frac{1}{4}\big[\langle Ax + Ay, Ax + Ay\rangle - \langle Ax - Ay, Ax - Ay\rangle\big] =$$

$$= \frac{1}{4}\big[|A(x+y)|^2 - |A(x-y)|^2\big] = \frac{1}{4}\big[|x+y|^2 - |x-y|^2\big] =$$

$$= \frac{1}{4}\big[\langle x+y, x+y\rangle - \langle x-y, x-y\rangle\big] = \langle x, y\rangle.$$

This implies that

$$\langle A^T Ax, y\rangle - \langle Ax, Ay\rangle = \langle x, y\rangle.$$

Then

$$\langle A^T Ax - x, y\rangle = \langle A^T Ax, y\rangle - \langle x, y\rangle = 0.$$

The arbitrariness of $y \in \mathbb{E}^n$ implies that $A^T Ax - x = 0$ for every $x \in \mathbb{E}^n$. We therefore conclude that A is an orthogonal transformation. ■

Theorem 7.2. *The set $\mathcal{O}(\mathbb{E}^n)$ of all orthogonal transformations *) of the Euclidean space $\mathbb{E}^n$ is a group with respect to superposition as the group operation.*

Proof. Suppose that $A, B \in \mathcal{O}(\mathbb{E}^n)$. Then $AA^T = A^T A = I$, $BB^T = B^T B = I$. Hence

$$(AB)(AB)^T = ABB^T A^T = AA^T = I,$$

i.e. $AB \in \mathcal{O}(E^n)$. It is easy to verify that the operation of superposition of orthogonal transformations is associative. A unit of this operation is the identity operator. The definition of orthogonal transformations implies that for every $A \in \mathcal{O}(\mathbb{E}^n)$ there exists an inverse transformation A^{-1} and $A^{-1} = A^T \in \mathcal{O}(E^n)$. Then $\mathcal{O}(\mathbb{E}^n)$ is a group. ■

In the previous chapter we have considered three invariants of the group of orthogonal transformations in the Euclidean space $\mathbb{E}^n$: the characteristic polynomials of a matrix, its characteristic roots and the discriminant of a quadratic form (cf. Theorem 6.2 and Corollaries 6.2, 6.3).

Theorem 7.1 implies the following

Corollary 7.1. *The length of an interval is an invariant of the group $\mathcal{O}(\mathbb{E}^n)$ of orthogonal transformations.*

Proof. If $A \in \mathcal{O}(\mathbb{E}^n)$ then A is an isometry. Hence $|Ax - Ay| = |A(x - y)| = |x - y|$ for arbitrary $x, y \in \mathbb{E}^n$. ■

Corollary 7.1 immediately implies

Corollary 7.2. *The property of triangles (or other figures and solids) to be congruent is an invariant of the group $\mathcal{O}(\mathbb{E}^n)$ of orthogonal transformations.*

A transformation $A \in L_0(\mathbb{E}^n)$ is said to be a *reflection* if

(7.3) $$Ax = -x \quad for\ every\ x \in \mathbb{E}^n.$$

A reflection is an involution for $A^2 = I$. Since $A^T = A$, the matrix of this transformation is symmetric. Hence $A^T A = A^2 = I$, which implies that A is an orthogonal transformation. Moreover, since det $A = (-1)^n$ when n is odd, we conclude that any reflection changes the orientation of systems of n vectors in $\mathbb{E}^n$ on an opposite one whenever the number n is odd.

A *translation* by an element $h \in \mathbb{E}^n$ (otherwise called a *shift* by h) is a transformation defined for an arbitrary $x \in \mathbb{E}^n$ by means of the formula

(7.4) $$T_h x = x + h.$$

---

*) An orthogonal transformation is, by definition, a linear mapping.

Theorem 7.3. *The set $\mathcal{T}(\mathbb{E}^n)$ of all translations in the Euclidean space $\mathbb{E}^n$ is a commutative group with respect to the superposition as the group operation.*

Proof. We find

$$T_g T_h = T_{g+h} \quad for\ arbitrary\ g, h \in \mathbb{E}^n. \tag{7.5}$$

Indeed, for every $x \in \mathbb{E}^n$ we have

$$(T_g T_h)x = T_g(T_h x) = T_g(x+h) = x + g + h = T_{g+h}x.$$

Clearly, the operation defined by Formula (7.4) is commutative and associative, because the summation of vectors (points) in the Euclidean space $\mathbb{E}^n$ is commutative and associative. Furthermore, since $T_0 x = x + 0 = 0$ for every $x \in \mathbb{E}^n$, we conclude that the identity transformation $T_0 = I$ is a unit with respect to the operation (7.5). Again, by Formula (7.5), for an arbitrary $h \in \mathbb{E}^n$ we have $T_h T_{-h} = T_{h-h} = T_0 = I$. The operation of superposition of translations is commutative. Then $T_h T_{-h} = T_{-h} T_h = I$ which implies that every transformation $T_h \in \mathcal{T}(\mathbb{E}^n)$ is invertible and $T_h^{-1} = T_{-h}$. We therefore conclude that $\mathcal{T}(\mathbb{E}^n)$ is a commutative group. ■

Observe that a translation on $h \neq 0$ is not a linear operator *) for $T_h(0) = h \neq 0$. However, it is an isometry, since we have

Corollary 7.3. *The length of an interval is an invariant of the group $\mathcal{T}(\mathbb{E}^n)$ of translations in the Euclidean space $\mathbb{E}^n$.*

Proof. For an arbitrary $T_h \in \mathcal{T}(\mathbb{E}^n)$ and for arbitrary $x, y \in \mathbb{E}^n$ we have

$$|T_h x - T_h y| = |(x+h) - (y+h)| = |x - y|.$$

■

Corollary 7.4. *The property of triangles to be congruent is an invariant of the group $\mathcal{T}(\mathbb{E}^n)$ of translations in the Euclidean space $\mathbb{E}^n$.*

Consider now transformations which are superpositions of translations and orthogonal transformations, all in the Euclidean space $\mathbb{E}^n$. They also form a group $\mathcal{E}(\mathbb{E}^n)$ of transformations of $\mathbb{E}^n$ onto itself such that its invariant are: the longitude of an interval and the property of the congruence of triangles (or other figures or solids).

A *symmetry with respect to a point* $a \in \mathbb{E}^n$ is a transformation A defined by means of the formula

$$Ax - a = a - x \quad for\ every\ x \in \mathbb{E}^n. \tag{7.6}$$

The transformation A defined by Formula (7.6) is a superposition of the reflection and the translation T_{2a} for $Ax = 2a - x$. Then $A \in \mathcal{E}(\mathbb{E}^n)$.

---

*) Cf. Exercise 2.1.

The point a is said to be the *symmetry center*. We say that a set $Y \subset \mathbb{E}^n$ has a *center of symmetry* $a \in \mathbb{E}^n$ if for every $x \in Y$ there is a (unique) point $y \in Y$ such that

$$y - a = a - x.$$

Observe that the reflection is a symmetry with respect to the point $a = 0$.

A *symmetry with respect to a linear manifold* $Y \subset \mathbb{E}^n$ is a transformation A defined by means of the formula

$$Ax - P_Y x = P_Y x - x \quad for\ every\ x \in \mathbb{E}^n, \tag{7.7}$$

where P_Y is an orthogonal projector onto the manifold Y, i.e. $P_Y x$ is a projection of the point x onto Y.

Since $Ax = 2P_Y x - x$, the transformation A defined by Formula (7.7) is a superposition of the projection, the operator $2I$ and the reflection.

If the linear manifold Y is a straight line then it is called the *axis of symmetry*. We say that a set $Z \subset \mathbb{E}^n$ *has an axis (a plane) of symmetry* Y if for every $x \in Z$ there is a unique point $y \in Z$ such that

$$y - P_Y x = P_Y x - x,$$

where P_Y denotes an orthogonal projector of the point x onto Y.

A *rotation* by an angle φ with respect to a point $a \in \mathbb{E}^2$, called the rotation center is a transformation defined by the matrix

$$A = \begin{pmatrix} \cos\ \varphi & \sin\ \varphi \\ -\sin\ \varphi & \cos\ \varphi \end{pmatrix}. \tag{7.8}$$

Since $A \in L_0(\mathbb{E}^2)$ and

$$A^T = \begin{pmatrix} \cos\ \varphi & -\sin\ \varphi \\ \sin\ \varphi & \cos\ \varphi \end{pmatrix} \quad and \quad A^T A = I,$$

we conclude that rotations are orthogonal transformations.

Similarly, we define rotations with respect to a straight line in the Euclidean space $\mathbb{E}^3$. This straight line is called the *axis of the rotation*. Note that a rotation by the angle $\frac{2\pi}{n}$ $(n \in \mathbb{N})$ is an involution of order n, i.e. $A^n = I$.

It is possible to prove that every transformation belonging to the group $\mathcal{E}(\mathbb{E}^n)$ can be represented as a superposition of a translation, a rotation and a reflection.

A *similarity* is a transformation A of the Euclidean space $\mathbb{E}^n$ which does not move the zero point and such that for a positive number κ we have

$$|Ax - Ay| = \kappa|x - y| \quad for\ x, y \in \mathbb{E}^n. \tag{7.9}$$

Observe that for $\kappa = 1$ we have an isometry, hence an orthogonal transformation provided that $A \in L_0(\mathbb{E}^n)$. One can prove that $|\det\ A| = \kappa^n$.

Theorem 7.4. *The set $\mathcal{S}(\mathbb{E}^n)$ of all similarities in the Euclidean space $\mathbb{E}^n$ is a group with respect to superposition as the group operation.*

Proof. Suppose that $A, B \in \mathcal{S}(\mathbb{E}^n)$. Then there are positive numbers κ, μ such that

$$|Ax - Ay| = \kappa|x - y|, \quad |Ax - Ay| = \mu|x - y| \quad for\ x, y \in \mathbb{E}^n.$$

This implies that

$$|ABx - ABy| = \kappa|Bx - By| = \kappa\mu|x - y| \quad for\ x, y \in \mathbb{E}^n.$$

Hence $AB \in \mathcal{S}(\mathbb{E}^n)$. Moreover, the operation of superposition is associative. Since $\det\ A = \pm\kappa^n \neq 0$, we conclude that there is a transformation A^{-1} which also is a similarity and its similarity coefficient is $\frac{1}{\kappa}$. Then $\mathcal{S}(\mathbb{E}^n)$ is a group. ■

An immediate consequence of this theorem is

Corollary 7.5. *The group $\mathcal{O}(\mathbb{E}^n)$ of orthogonal transformations is a subgroup of the group $\mathcal{S}(\mathbb{E}^n)$ of similarities.*

Corollary 7.6. *The angle between two vectors is an invariant of the group $\mathcal{S}(\mathbb{E}^n)$ of similarities.*

Proof. Indeed, suppose that $A \in \mathcal{S}(\mathbb{E}^n)$ and that α denotes the angle between vectors x and y in $\mathbb{E}^n$. Let β be the angle between the vectors Ax and Ay. Then we have

$$\cos\ \beta = \frac{\langle Ax, Ay\rangle}{|Ax|\ |Ay|} =$$

$$= \frac{\langle A(x+y), A(x+y)\rangle - \langle A(x-y), A(x-y)\rangle}{4|Ax|\ |Ay|} =$$

$$= \frac{|A(x+y)|^2 - |A(x-y)|^2}{4|Ax|\ |Ay|} = \frac{\kappa^2|x+y|^2 - \kappa^2|x-y|^2}{4\kappa^2|Ax|\ |Ay|} =$$

$$= \frac{|x+y|^2 - |x-y|^2}{4|x|\ |y|} = \frac{\langle x, y\rangle}{|x|\ |y|} = \cos\ \alpha.$$

■

Corollary 7.7. *The equality of vectors is an invariant of the group $\mathcal{S}(\mathbb{E}^n)$ of similarities.*

Proof. Indeed, suppose that $A \in \mathcal{S}(\mathbb{E}^n)$, $x, y \in \mathbb{E}^n$ and $y = x$. Then $|Ax - Ay| = \kappa|x - y| = 0$. This implies $Ax = Ay$. ■

Corollary 7.7 implies that the congruence of triangles is also an invariant of the group $\mathcal{S}(\mathbb{E}^n)$ of similarities.

An *affine* transformation is a transformation defined by means of the formula

$$Ax = a + Bx \quad for\ x \in \mathbb{E}^n, \tag{7.10}$$

where $a \in \mathbb{E}^n$, $B \in L_0(\mathbb{E}^n)$ are fixed and det $B \neq 0$.

Theorem 7.5. *The set $\mathcal{A}(\mathbb{E}^n)$ of all affine transformations in the Euclidean space $\mathbb{E}^n$ is a group with respect to superposition as the group operation.*

Proof. Suppose that $a, \tilde{A} \in \mathcal{A}(\mathbb{E}^n)$. It means that for every $x \in \mathbb{E}^n$ we have $Ax = a + Bx$, $\tilde{A}x = \tilde{a} + \tilde{B}x$, where $a, \tilde{a} \in \mathbb{E}^n$, $B, \tilde{B} \in L_0(\mathbb{E}^n)$ are fixed and det $B \neq 0$, det $\tilde{B} \neq 0$. Then

$$A\tilde{A}x = a + B\tilde{A}x = a + B(\tilde{a} + \tilde{B}x) = a + B\tilde{a} + B\tilde{B}x = d + Dx,$$

where $d = a + B\tilde{a} \in \mathbb{E}^n$, $D = B\tilde{B} \in L_0(\mathbb{E}^n)$ and

$$\det\ D = \det\ (B\tilde{B}) = (\det\ B)(\det\ \tilde{B}) \neq 0.$$

This implies that $\tilde{A} \in \mathcal{A}(\mathbb{E}^n)$. The identity operator I is the unit of the set $\mathcal{A}(\mathbb{E}^n)$. Since det $B \neq 0$, we conclude that there exists an inverse transformation defined by means of the formula $A^{-1}x = B^{-1}a + B^{-1}x$. ∎

Theorem 7.5 immediately implies

Corollary 7.8. *The group $\mathcal{S}(\mathbb{E}^n)$ of similarities is a subgroup of the group $\mathcal{A}(\mathbb{E}^n)$ of affine transformations.*

It is easy to verify that an angle between vectors is not an invariant of the group $\mathcal{A}(\mathbb{E}^n)$ of affine transformations. However, one can prove the following

Corollary 7.9. *Straight lines (and, in general, linear manifolds of dimension $k \leq n$) are invariants of the group $\mathcal{A}(\mathbb{E}^n)$ of affine transformations.*

Suppose that we are given points $a, b \in \mathbb{E}^n$ and $p = ta + (1 - t)b$ for a $t \in \mathbb{R}$. Then the number $\frac{t-1}{t}$ is also an invariant of the group $\mathcal{A}(\mathbb{E}^n)$ of affine transformations.

The theory of invariants of the group of orthogonal transformations, i.e. the group of isometries is called *metric geometry*. The theory of invariants of the group of similarities is called *similarities geometry*. The theory of invariants of the group of affine transformations is called *affine geometry* (cf. Borsuk B[2], Borsuk and Szmielew B[3]). It follows from the above considerations that a larger group of transformations has even less invariants.

An *inversion* with respect to a point $a \in \mathbb{E}^n$ is a transformation A of the Euclidean space $\mathbb{E}^n 2$ into itself such that for every $x \in \mathbb{E}^n$ the point Ax lies on a half-line connecting the points x and a and satisfying the condition

$$|Ax - a| \cdot |x - a| = c^2, \quad where\ c \in \mathbb{R}\ is\ fixed. \tag{7.11}$$

The point a is called the *center of inversion*.

If $y = Ax$ then, by definition of inversions, $Ay = x$. Hence $A^2 = I$, i.e. any inversion is an involution.

A straight line completed by the point at infinity can be treated as a circle with the radius $r = \infty$. Straight lines completed by the point at infinity are called *improper circles*. Using this notion, we obtain the following

Theorem 7.6. *Circles are invariants of inversions.*

Proof. We prove this theorem for $\mathbb{E}^2$. In the general case of $\mathbb{E}^n$ the proof is similar. Suppose then that A is an inversion with the center $a \in \mathbb{E}^2$. Condition (7.11) implies that $|Ax - a| > c$ whenever $|x - a| < c$. Then the inversion A transforms points lying inside the circle

$$K_{a,c} = \{x \in \mathbb{E}^2 : |x - a| = c\}$$

onto points lying outside this circle, and conversely. To simplify calculations, let $a = 0$. Write $x = (x_1, x_2)$, $y = Ax = (y_1, y_2)$. Then, by Condition (7.11), we find

$$\frac{x_1}{|x|^2} = \frac{y_1}{|y|^2}, \qquad \frac{x_2}{|x|^2} = \frac{y_2}{|y|^2}. \tag{7.12}$$

Formulae (7.12) imply that the inversion A transforms straight lines passing through the origin of the coordinate system into circles. Indeed, consider a straight line determined by the equation $\alpha x_1 + \beta x_2 + \gamma = 0$. Then, by Formulae (7.12), we find

$$0 = \alpha x_1 + \beta x_2 + \gamma = \alpha \frac{y_1}{|y|^2}|x|^2 + \beta \frac{y_2}{|y|^2}|x|^2 + \gamma =$$

$$= (\alpha y_1 + \beta y_2 + \gamma |y|^2)\frac{|x|^2}{|y|^2} = [\alpha y_1 + \beta y_2 + \gamma(y_1^2 + y_2^2)]\frac{c^2}{|y|^4},$$

which yields to the equation

$$\alpha y_1 + \beta y_2 + \gamma(y_1^2 + y_2^2) = 0. \tag{7.13}$$

Hence for $\gamma = 0$ the inversion A transforms the straight line under consideration onto itself.

If $\gamma \neq 0$ then the straight line does not pass through the origin of the coordinate system and it is transformed by the inversion into a circle passing through that origin (since the point (0,0) satisfies Equation (7.13)). Clearly, if it is the case then the center of this inversion does not lie in the origin of the coordinate system. Any circle with the center at the origin of the coordinate system is transformed into itself. Indeed, if we have a circle with the equation $x_1^2 + x_2^2 = r^2$ then $|x| = r$ and, after an obvious transformation, we obtain the equation

$$|y|^2 = y_1^2 + y_2^2 = \frac{x_1^2}{|x|^4}|y|^4 + \frac{x_2^2}{|x|^4}|y|^4 = \frac{x_1^2 + x_2^2}{|x|^4}|y|^4 =$$

$$= \frac{r^2}{r^4}|y|^4 = \frac{1}{r^2}|y|^4,$$

i.e. the equation $|y| = r$. On the other hand, since an inversion, as involution, transforms circles with centers not lying at the origin of the coordinate systems onto straight lines not passing by this origin. ■

Often it is convenient to determine an inversion as a transformation of the form

$$Ax = \frac{1}{|x|^2}x.$$

A superposition of this last transformation with a translation and a similarity leads to a transformation defined by means of Formula (7.11).

If we have a transformation $y = Q(x)$, where $Q(x)$ is an arbitrary rational function, then this transformation is a superposition of a number of affine transformations and inversions. Transformations of that form generate a group, since superposition of two rational functions and their inverses are again rational functions.

When considering arbitrary groups, it is sometimes convenient to consider instead of a given group its homomorphic mappings into groups of linear operators. We shall show a scheme of such a procedure.

A set $T_G = \{T_g\}_{g\in G}$ is said to be a *representation of a finite group* G in a linear space X if to every element $g \in G$ there corresponds an operator $T_g \in L_0(X)$ with the property

$$T_g T_h = T_{gh}, \quad whenever \quad g, h \in G. \tag{7.14}$$

The *dimension of a representation* is, by definition, the dimension of the space X. The same group G may have finite and infinite dimensional representations. Here we shall consider only finite dimensional representations.

Representations T_G and $\tilde{T}_G$ in linear spaces X and Y, respectively, are said to be *equivalent* if there is an operator $A \in L_0(X \to Y)$ such that

$$\tilde{T}_g = AT_gA^{-1} \quad for\ every\ g \in G.$$

One can prove that every equivalence class of representations of a finite group contains a *unitary representation*, i.e. a representation T_G such that T_g is a unitary operator for every $g \in G$. Then, in order to determine all non-equivalent representations of a finite group, it is enough to determine all its unitary representations.

A representation T_G of a group G in a linear space X is said to be *reducible* if there is a proper subspace $Y \subset X$ *) which is *invariant* for all operators T_g $(g \in G)$, i.e. a representation T_g such that

$$T_gY \subset Y \quad for\ all\ g \in G.$$

*) i.e. $Y \neq \{0\}$ and $Y \neq X$.

Representations, which are not reducible, are said to be *irreducible*. Therefore we conclude that operators from an irreducible representation have no common proper invariant subspace.

Every finite group G has the so-called *natural representation*. Namely, if the number of elements in G is n then there is a one-to-one mapping of the group G onto the set $\{1, ..., n\}$. Write $a_g = \{\delta_{jg}\}_{g,j=1,\dots,n}$ for $g \in G$. Then a_g is a vector with 1 on the gth place and with zeros on the remained places. Let $\mathbb{G}$ be an arbitrary field of scalars and let

$$X = \big\{x : x = \sum_{g \in G} t_g a_g, \quad where\ t_g \in \mathbb{G}\big\}.$$

Clearly, X is a linear space over the field $\mathbb{G}$. For an arbitrary $h \in G$ define a mapping of the space X into itself by means of the formula

$$A_h x = \sum_{g \in G} t_g a_{gh}, \quad where \quad x = \sum_{g \in G} t_g a_g.$$

Since $A_h \in L_0(X)$ and $A_{h_1} A_{h_2} = A_{h_1 h_2}$ for arbitrary $h, h_1, h_2 \in G$, we conclude that A_h is a representation of the group G.

If $X = Y \oplus Z$, then any representation T_G can be decomposed onto representations T'_G and T''_G uniquely determined by means of the equality

$$T_G x = T'_G y + T''_G z, \quad where \quad x = y + z,\ y \in Y,\ z \in Z. \tag{7.15}$$

A representation T_G defined by Formula (7.15) is said to be a *sum of representations* T'_G and T''_G.

Every reducible finite dimensional unitary representation is a sum of irreducible unitary representations.

Theorem 7.7 (Schur Lemma). *If T'_G and T''_G are two equivalent irreducible representations of a finite group G in linear spaces X' and X'', respectively, and an operator $A \in L_0(X'' \to X')$ satisfies the condition*

$$T'_g A = A T''_g \quad for\ every\ g \in G \tag{7.16}$$

then $A = 0$.

Examples and other properties of group representations can be found, for instance, in the books of Lax (cf. L[2]) and Prostakov (cf. P[1]).

Exercises.

Exercise 7.1. Determine transformations groups in $\mathbb{C}^n$. □

Exercise 7.2. Let G_n be the set of nth roots of unity. Prove that

(i) G_n is a group consisting of n elements with respect to the multiplication in $\mathbb{C}$;

(ii) a representation of the group G_n in an arbitrary finite dimensional linear space X is the collection of operators I, S,...,S^{n-1}, where $S \in L_0(X)$ is an involution of order n, i.e. $S^n = I$. Is this representation reducible? If this is the case, then determine its invariant subspace and describe a decomposition into irreducible representations. □

Chapter 8.

Index and perturbations of linear operators.

Let X and Y be linear spaces over the same field $\mathbb{F}$ of scalars. Recall (cf. Formulae (2.11) and (2.12)) that the nullity of a linear operator $A \in L(X \to Y)$ is the number $\alpha_A = \dim\ker\, A$ and the deficiency of A is the number $\beta_A = \operatorname{codim} A\mathcal{D}_A = \dim Y/A\mathcal{D}_A$.

The ordered pair (α_A, β_A) is said to be the *dimensional characteristic* of the operator A (shortly: its *d-characteristic*). We say that the d-characteristic is *finite* if both numbers, α_A and β_A, are finite. If at least one of the numbers α_A, β_A is finite then we say that the d-characteristic of the operator A is *semi-finite*.

Write

$$\begin{aligned} D(X \to Y) &= \{A \in L(X \to Y) : \alpha_A < +\infty,\ \beta_A < +\infty\}, \\ D^-(X \to Y) &= \{A \in L(X \to Y) : \alpha_A = +\infty,\ \beta_A < +\infty\}, \\ D^+(X \to Y) &= \{A \in L(X \to Y) : \alpha_A < +\infty,\ \beta_A = +\infty\}, \end{aligned}$$

For linear operators A belonging to one of these three sets, i.e. for operators with either finite or semi-finite d-characteristic, we define the *index* κ_A in the following manner:

$$\kappa_A = \begin{cases} \beta_A - \alpha_A & \text{if } A \in D(X \to Y), \\ +\infty & \text{if } A \in D^+(X \to Y), \\ -\infty & \text{if } A \in D^-(X \to Y). \end{cases}$$

Notions already introduced are very useful when solving linear equations. For instance, consider the equation

$$Ax = y. \tag{8.1}$$

Clearly, a solution of this equation exists if and only if $y \in \mathcal{E}_A = A\mathcal{D}_A$. On the other hand, if we know a solution x_1 of Equation (8.1), i.e. an element $x_1 \in \mathcal{D}_A$ such that $Ax_1 = y$, then a general solution of Equation (8.1) is of the form $x = x_0 + x_1$, where x_0 is an arbitrary element of the kernel of A. Then, in order to solve this equation, the knowledge of the sets $\mathcal{D}_A$ and $\mathcal{E}_A = A\mathcal{D}_A$ is, indeed, essential. The nullity α_A and the deficiency β_A characterize in a sense these sets, although they do not describe them exactly.

Very often a given linear equation can be reduced to another equation in such a manner that the nullity and the deficiency of the operator obtainedare easy to determine. This is a reason, why the following two theorems about superpositions of linear operators will play a fundamental role in our subsequent considerations. Here and in the sequel we shall assume that all linear spaces under question are considered over the same field $\mathbb{F}$ of scalars (of characteristic zero).

Theorem 8.1. *Let X, Y, Z be linear spaces over a field $\mathbb{F}$ of scalars. Let*

$$B \in \begin{cases} D(X \to Y) \\ D^-(X \to Y) \\ D^+(X \to Y) \end{cases} \qquad A \in \begin{cases} D(Y \to Z) \\ D^-(Y \to Z) \\ D^+(Y \to Z) \end{cases}$$

and let $\mathcal{D}_A = Y \supset \mathcal{E}_B = B\mathcal{D}_B$. Then the superposition AB exists,

$$AB \in \begin{cases} D(X \to Z) \\ D^-(X \to Z) \\ D^+(X \to Z) \end{cases}$$

respectively, and

$$\kappa_{AB} = \kappa_A + \kappa_B. \tag{8.2}$$

Proof. To begin with, we shall prove our theorem for linear operators with a finite d-characteristic, i.e. in the case when $A \in D(Y \to Z)$, $B \in D(X \to Y)$. Let $V_1 = \mathcal{E}_B \cap \ker A$. Write $n_1 = \dim V_1$. The subspace ker A can be written as the direct sum

$$\ker A = V_1 \oplus V_2, \quad \textit{where} \quad \dim V_2 = \alpha_A - n_1, \tag{8.3}$$

and the space Y as the direct sum

$$Y = \mathcal{E}_A \oplus V_2 \oplus V_3, \quad \textit{where} \quad \dim V_3 = n_3. \tag{8.4}$$

The definition of the deficiency and Theorem 2.3 together imply that

$$\dim (V_2 \oplus V_3) = \operatorname{codim} Y/\mathcal{E}_B = \beta_B.$$

Then $\alpha_A - n_1 + n_3 = \beta_B$ and $\alpha_A - \beta_B = n_1 - n_3$. Formula (8.3) implies that $AV_2 = 0$. Hence

$$\mathcal{E}_A = A\mathcal{D}_A = AY = A\mathcal{E}_B \oplus AV_1 \oplus AV_3 = AB\mathcal{D}_B \oplus AV_3 = \mathcal{E}_{AB} \oplus AV_3. \tag{8.4}$$

But $\dim AV_3 = \dim V_3 = n_3$, since the operator A maps in a one-to-one way the subspace V_1 onto AV_3. Finally, we get

$$\kappa_{AB} = \beta_{AB} - \alpha_{AB} = \beta_A + n_3 - (\alpha B + n_1) = \beta_A - \alpha_A + \beta_B - \alpha_B = \kappa_A + \kappa_B.$$

Suppose now that $\alpha_A, \alpha_B < +\infty$. Then $n_1 < +\infty$, which implies $\alpha_{AB} = \alpha_B + n_1 < +\infty$. If $\beta_A, \beta_B < +\infty$ then we have $n_3 < \beta_B < +\infty$. Hence $\beta_{AB} = \beta_B - n_3 < +\infty$. ■

Theorem 8.2. *Let X, Y, Z be linear spaces over a field $\mathbb{F}$ of scalars. Suppose that $A \in L_0(Y \to Z)$ and $B \in L(X \to Y)$. Then the superposition AB exists, and*

(a) $\alpha_{AB} < +\infty$ *implies* $\alpha_B < +\infty$;

(b) $\beta_{AB} < +\infty$ *implies* $\beta_A < +\infty$.

Proof. Since

$$\ker\ B = \{x \in \mathcal{D}_B : Bx = 0\} \subset \{x \in \mathcal{D}_B : ABx = 0\} = \ker\ AB,$$

we find

$$\alpha_B \le \alpha_{AB}, \tag{8.5}$$

i.e. $\alpha_B < +\infty$. Since $A\mathcal{D}_A = AY \supset AB\mathcal{D}_B$, we have

$$\beta_A \le \beta_{AB}, \tag{8.6}$$

i.e. $\beta_A < +\infty$. ■

Theorem 8.2 immediately implies

Corollary 8.1 *Suppose that $A \in L_0(X \to Y)$, $B \in L_0(Y \to X)$, hence $AB \in L_0(Y)$, $BA \in L_0(X)$. If $AB \in D(Y \to Y)$, $BA \in D(X \to X)$ then $A \in D(X \to Y)$ and $B \in D(Y \to X)$.*

Corollary 8.2. *Suppose that $A \in L_0(X)$ and there is a positive integer m such that $I - T^m \in D(X \to X)$. Then $I - T \in D(X \to X)$.*

Proof. For the proof it is enough to assume in Corollary 8.1 that $A = I - T$, $B = I + T + ... + T^{m-1}$. ■

Corollary 8.3. *A linear operator $A \in L_0(X \to Y)$ is invertible (is an isomorphism) if and only if $\alpha_A = \beta_A = 0$.*

A linear operator $A \in L_0(X \to Y)$ is said to be *right invertible* (*left invertible*) if there is a linear operator $B \in L_0(Y \to X)$ such that $AB = I_Y$ ($BA = I_X$, respectively). The operator B is said to be a *right* (*left*) *inverse* of A.

Theorem 8.3. *A linear operator $A \in L_0(X \to Y)$ is right (left) invertible if and only if $\beta_A = 0$ ($\alpha_A = 0$, respectively).*

Proof. Suppose that $\alpha_A = 0$. Then $\ker\ A = \{0\}$ and the operator A is a one-to-one mapping of the space X onto the set $\mathcal{E}_A = AX$. Decompose the space Y onto the direct sum: $Y = AX \oplus Z$. Define a linear operator $B \in L_0(Y \to X)$ in the following way:

$$By = \begin{cases} 0 & \text{for } y \in Z, \\ x & \text{for } y = Ax,\ x \in X. \end{cases}$$

Clearly, $B(Ax) = x$ for $x \in X$. Then $BA = I_X$ and the operator A is left invertible.

Suppose now that $\beta_A = 0$. This means that the operator A maps the space X onto the space Y. Decompose the space X onto the direct sum: $X = \ker\ A \oplus X_1$ and denote by A_0 the restriction of the operator A to the space X_1. Clearly, the operator A_0 is a one-to-one mapping of X_1 onto Y. Then there exists its inverse operator $A^{-1} \in L_0(Y \to X_1)$. Let

$B = A_0^{-1}$. Then $A(By) = Ax = y$ for $y \in Y$, i.e. $AB = I_Y$. Hence the operator A is right invertible.

Next, suppose that a linear operator $A \in L_0(X \to Y)$ is left invertible, i.e. there is a linear operator $B \in L_0(Y \to X)$ such that $BA = I_X$. Suppose, moreover, that there is an element $x \in \ker\ A$ such that $x \neq 0$. Then $(BA)x = B(Ax) = B(0) = 0$, which contradicts to our assumption that $BAx = x$. Hence $\ker\ A = \{0\}$ which implies $\alpha_A = 0$.

If a linear operator $A \in L_0(X \to Y)$ is right invertible then there is a linear operator $B \in L_0(Y \to X)$ such that $AB = I_Y$. This, and Inequality (8.6) together imply that $\beta_A \leq \beta_{AB} = \beta_{I_Y} = 0$. Then $\beta_A = 0$. ■

Example 8.1. The operator A_p of multiplication by a function $p \in C[a,b]$ such that $p(t) \neq 0$ for $a \leq t \leq b$ is in the space $C[a,b]$ invertible and $A_p^{-1} = A_{1/p}$. Then $\alpha_{A_p} = \beta_{A_p} = 0$ which implies $\kappa_{A_p} = \beta_{A_p} - \alpha_{A_p} = 0$. □

Example 8.2. A function $y \in C^1[a,b]$ is said to be *primitive* for a function $x \in C[a,b]$ if $\frac{\mathrm{d}}{\mathrm{d}t}y(t) = x(t)$ for $a \leq t \leq b$. It is well known that every continuous function has a primitive function and that for a fixed $t_0 \in [a,b]$ there exists a unique primitive function y such that $y(t_0) = 0$. If it is the case, then this primitive function is traditionally written as

$$y(t) = \int_{t_0}^{t} x(s)ds$$

and it is called an *integral* of x with lower and upper limits t_0n and t, respectively.

Let now $Dx = \frac{\mathrm{d}x}{\mathrm{d}t}$ for $x \in C^1[a,b]$ and $(Rx)(t) = \int_{t_0}^{t} x(s)ds$ for $x \in C[a,b]$. Write $X = C[a,b]$, $Y = C^1[a,b]$. Then $Y \subset X$ (cf. Example 1.11 and Exercise 1.1). Moreover, $D \in L_0(Y \to X)$, $R \in L_0(X \to Y)$ and $DR = I_X$. Hence the operator D is right invertible and the operator R is left invertible. This implies $\beta_D = 0$ and $\alpha_R = 0$. Next, since $\frac{\mathrm{d}}{\mathrm{d}t}x(t) = 0$ if and only if the function $x(t)$ is constant, i.e. $x(t) \equiv a \in \mathbb{R}$, we conclude that $\alpha_D = \dim \ker`D = \dim \mathbb{R} = 1$. Then the operator D is not left invertible, hence D is not invertible. Moreover, $\kappa_D = \beta_D - \alpha_D = -1$, $\kappa_R = \beta_R - \alpha_R = 1$. We conclude that also the operator R is not invertible, which implies $\beta_R \neq 0$. □

Denote by $K(X \to Y)$ the set of all finite dimensional operators defined on a linear space X and with values in a linear space Y. Theorem 2.11 implies that every operator $K \in K(X \to Y)$ is of the form

$$Kx = \sum_{j=1}^{n} f_j(x)y_j \quad for\ x \in X,$$

where $f_1, ..., f_n \in X'$ and $y_1, ..., y_n \in Y$ are given.

Theorem 8.4. *If $K \in K(X \to Y)$ then $I + K \in D(X \to Y)$ and $\kappa_{I+K} = 0$.*

Proof. Consider the equation

$$(I + K)x = y, \quad where \quad Kx = \sum_{j=1}^{n} f_j(x)y_j, \quad y \in X. \tag{8.7}$$

Since f_j are linear functional, writing $C_i = f_i(x)$ for $i = 1, ..., n$, we obtain Equation (8.7) in the form

$$x + \sum_{i=1}^{n} C_j y_j = y.$$

Hence its solution should be of the form

$$x = y - \sum_{i=1}^{n} C_j y_j. \tag{8.8}$$

Acting by functionals $f_1, ..., f_n$ on both sides of Equation (8.7), we obtain the following system of n linear equations with n unknowns $C_1, ..., C_n$:

$$C_j + \sum_{i=1}^{n} K_{ji} C_i = L_j \qquad (j = 1, ..., n), \tag{8.9}$$

where $K_{ji} = f_j(y_i)$ and $L_j = f_j(y)$ $(i, j = 1, ..., n)$ are known. If the determinant of this system

$$\Delta_K = \begin{vmatrix} K_{11}+1 & K_{12} & ... & K_{1n} \\ K_{21} & K_{22}+1 & ... & K_{2n} \\ ... & ... & .... & .... \\ K_{n1} & K_{n2} & ... & K_{nn}+1 \end{vmatrix} \tag{8.10}$$

is different than zero then to every system $(L_1, ..., L_n)$ of numbers there corresponds a unique solution of the system (8.9) given by the Cramer formulae. Then Equation (8.7) has for every $y \in Y$ a unique solution of the form (8.8), where the coefficients $C_1, ..., C_n$ are a unique solution of the system (8.9).

If $y = 0$ then $L_1 = ... = L_n = 0$ and the homogeneous system (8.9) with $\Delta_K not = 0$ has only zeros as a solution: $C_1 = ... = C_n = 0$. Hence the equality (8.8) implies that the only solution of Equation (8.7) is $x = 0$. We therefore conclude that the operator $I + K$ is invertible in the space X. This implies that $\alpha_{I+K} = \beta_{I+K} = 0$ (cf. Corollary 8.3). Then $\kappa_{I+K} = \beta_{I+K} - \alpha_{I+K} = 0$.

If the determinant $\Delta_K = 0$ then a solution of the system (8.9) does not exist for every system $(L_1, ..., L_n)$ of numbers. Write the matrix of coefficients of the system (8.9) in the form

$$I + K = [K_{ji} + \delta_{ji}]_{j,i=1,...,n}$$

and denote its rank by k, i.e. $r(I + K) = k$.

A necessary and sufficient condition for the existence of a solution to the system (8.9) is that the vector $(L_1, ..., L_n)$ has to belong to a k-dimensional subspace (cf. the Kronecker-Capella theorem (Theorem 3.8) and Corollary 3.7). Then

$$\beta_{I+K} = n - k = n - r(I + K).$$

By Theorem 3.11, we find

$$\alpha_{I+K} = n - r(I+K).$$

Then also in the case $\Delta_K = 0$ we find

$$\kappa_{I+K} = \beta_{I+K} - \alpha_{I+K} = n - r(I+K) - [n - r(I+K)] = 0.$$

■

Example 8.3. Consider the following equation, traditionally called an *integral equation with a degenerate kernel*:

$$x(t) + \int_a^b \left[\sum_{j=1}^n g_j(t)h_j(s)\right] x(s)ds = y(t), \tag{8.11}$$

where functions $y,\, g_1, ..., g_n \in C[a,b]$ are given and we write

$$v(t) = \int_a^t u(s)ds \quad for\ a \le t \le b\ and\ for\ an\ arbitrary\ function\ u \in C[a,b]$$

(cf. Example 8.2). By definition of the function $v(t)$, it follows that the number

$$v(b) = \int_a^b u(s)ds \qquad for\ u \in C[a,b]$$

is a real number called the *definite integral* of the function u. It is easy to verify that the mapping such that to every function $u \in C[a,b]$ there corresponds a real number $\int_a^b u(s)ds$ is linear operator, hence a linear functional on the space $C[a,b]$.

Write now for every $x \in X$

$$Kx = \sum_{j=1}^n f_j(x)g_j \quad where \quad f_j(x) = \int_a^b h_j(s)x(s)ds \quad (j = 1, ..., n). \tag{8.12}$$

It is easy to verify that $f_1, ..., f_n$ are linear functionals on the space $C[a,b]$ and that K is a finite dimensional linear operator mapping the space $C[a,b]$ into itself. Then we can apply Theorem 8.4 to the operator $I + K$. By Formulae (8.12), Equation (8.11) can written in the form

$$(I + K)x = y,$$

for

$$(Kx)(t) = \left[\sum_{j=1}^n f_j(x)g_j\right](t) = \sum_{j=1}^n g_j(t)\int_a^b h_j(s)x(s)ds =$$

$$= \int_a^b \left[\sum_{j=1}^n g_j(t)h_j(s)\right] x(s)ds.$$

Then every solution of Equation (8.11) is of the form

$$x = y - \sum_{j=1}^{n} C_j g_j, \tag{8.13}$$

where the numbers $C_1, ..., C_n$ are to be determined in that way as in the proof of Theorem 8.4. In particular, if

$$\Delta_K = \det\,[f_j(g_i) + \delta_{ji}]_{j,i=1,\ldots,n} \neq 0$$

then Equation (8.11) has a unique solution of the form (8.13), where the numbers $C_1, ..., C_n$ are determined by the Cramer formulae. In this manner an integral equations with an unknown function x was reduced to a system of n linear equations with scalar coefficients and with n unknown numbers $C_1, ..., C_n$. □

Suppose that we are given a class $\mathfrak{A}$ (non-necessarily linear) of linear operators. A linear operator B is said to be an *$\mathfrak{A}$-perturbation of a linear operator A* if $A + B \in \mathfrak{A}$. A linear operator B is said to be a *perturbation of the class* $\mathfrak{A}$ of linear operators if $A + B \in \mathfrak{A}$ for every $A \in \mathfrak{A}$. We denote by $\Pi(\mathfrak{A})$ the set of all perturbations of the class $\mathfrak{A}$.

Theorem 8.5. *The set $\Pi(\mathfrak{A})$ of all perturbations of the class $\mathfrak{A}$ is additive, i.e. if T_1 and T_2 are perturbations of the class $\mathfrak{A}$ then $T_1 + T_2$ is also a perturbation of the class $\mathfrak{A}$.*

Proof. Let $A \in \mathfrak{A}$. By our assumption, $A + T_1 \in \mathfrak{A}$, because T_1 is a perturbation of the class $\mathfrak{A}$. But T_2 is also a perturbation of the class $\mathfrak{A}$, which implies that $A + (T_1 + T_2) = (A + T_1) + T_2 \in \mathfrak{A}$. We therefore conclude that the operator $T_1 + T_2$ is a perturbation of the class $\mathfrak{A}$. ■

Corollary 8.4. *If the class $\mathfrak{A}$ is* homogeneous, *i.e. $\alpha A \in \mathfrak{A}$ for every scalar α and for every $A \in \mathfrak{A}$, then the set $\Pi(\mathfrak{A})$ of all perturbations of the class $\mathfrak{A}$ is linear.*

Proof. Let $A \in \mathfrak{A}$ be arbitrary. Then, by our assumption, $\frac{1}{\alpha}A \in \mathfrak{A}$ for every scalar $\alpha \neq 0$. If a linear operator T is a perturbation of the class $\mathfrak{A}$ then $\frac{1}{\alpha}(A + \alpha T) = \frac{1}{\alpha}A + T \in \mathfrak{A}$. Hence $A + \alpha T \in \mathfrak{A}$. This, and the arbitrariness of $A \in \mathfrak{A}$ together imply that αT is a perturbation of the class $\mathfrak{A}$. Hence the set $\Pi(\mathfrak{a})$ is homogeneous. This, and Theorem 8.5 together imply that the set $\Pi(\mathfrak{A})$ is linear. ■

Theorem 8.6. *Every finite dimensional linear operator is a perturbation of the class of all linear operators with a finite d-characteristic (deficiency, nullity) and this perturbation* preserves the index, *i.e.*

$$\kappa_{A+K} = \kappa_A \qquad \textit{for all } K \in K(X \to Y) \quad \textit{and}$$

$$A \in \begin{cases} D(X \to Y), \\ D^+(X \to Y), \\ D^-(X \to Y), \end{cases} \quad \textit{respectively.}$$

Proof. Suppose that $A \in D(X \to Y)$. Decompose the space X onto the direct sum: $X = \ker\ A \oplus V$. The operator A_1, defined as the restriction of the operator A to the subspace V is, by Theorem 8.3, left invertible (for $\alpha_{A_1} = 0$). Let K_1 be the restriction of the operator K to the subspace V and let B_1 be a left inverse of A_1. Then $A_1 + K_1 = (I + K_1B_1)A_1$ on V, since $B_1A_1 = I$ on V.

Observe that the operator K_1B_1 defined on the space Y and with values in Y is finite dimensional for $K_1B_1Y < +\infty$, since the operator K, hence also the operator K_1 are finite dimensional. This, and Theorems 8.1, 8.2 and 8.4 together imply that

$$\kappa_{A_1+K_1} = \kappa_{(I+K_1B_1)A_1} = \kappa_{I+K_1B_1} + \kappa_{A_1} = \kappa_{A_1} + \beta_{A_1} = \beta_A, \tag{8.14}$$

for $\kappa_{I+K_1B_1} = 0$ and $\alpha_{A_1} = 0$.

Observe that the operator $A+K$ is an extension of the operator A_1+K_1. We shall prove that $\kappa_{A+K} = \kappa_{A_1+K_1} - \alpha_A$. Consider three cases:

(i) $K\ker\ A \subset \mathcal{E}_{A_1+K_1} = (A_1+K_1)X$. Then

$$\mathcal{E}_{A+K} = (A+K)X = (A_1+K_1)X = \mathcal{E}_{A_1+K_1}.$$

This implies $\beta_{A+K} + \beta_{A_1+K_1}$. But

$$\alpha_{A+K} = \dim\ \ker\ (A+K) = \dim\ \ker\ A + \dim\ \ker\ (A_1+K_1) = \alpha_A + \alpha_{A_1+K_1}.$$

Consequently,

$$\kappa_{A+K} = \beta_{A+K} - \alpha_{A+K} = \beta_{A_1+K_1} - \alpha_{A_1+K_1} - \alpha_A = \kappa_{A_1+K_1} - \alpha_A.$$

(ii) $Kx \notin \mathcal{E}_{A_1+K_1}$ for every $x \in \ker\ A$, $x \neq 0$. Write $r = \dim K\ker\ A$. Clearly, $\beta_{A+K} = \beta_{A_1+K_1} - r$. On the other hand, writing $K_3 = K|_{\ker\ A}$, we get $\alpha_{A+K} = \alpha_{A_1+K_1} + (\alpha_A - r)$, since $\alpha_A - r = \alpha_{K_3}$ for $(A+K)x = Kx$ whenever $x \in \ker\ A$. Then

$$\kappa_(A+K) = \beta_{A+K} - \alpha_{A+K} = \beta_{A_1+K_1} - r - (\alpha_{A_1+K_1} + \alpha_A - r) =$$

$$= \beta_{A_1+K_1} - \alpha_{A_1+K_1} - \alpha_A = \kappa_{A_1+K_1} - \alpha_A.$$

(iii) If (i) and (ii) do not hold then decompose the space $\ker\ A$ onto the direct sum $\ker\ A = V_1 \oplus V_2$, where $KV_1 \subset \mathcal{E}_{A_1+K_1}$ and $Kx \notin \mathcal{E}_{A_1+K_1}$ for every $x \in V_2$, $x \neq 0$. Let $A_2 + K_2$ be the restriction of the operator $A+K$ to the subspace $V \oplus V_1$. Then Point (ii) of our proof implies that $\kappa_{A_2+K_2} = \kappa_{A_1+K_1} - \dim\ V_2$. Moreover, Point (i) of this proof implies that $\kappa_{A+K} = \kappa_{A_2+K_2}$ - dim V_1. Since dim V_1 + dim V_2 = dim ker $A = \alpha_A$, we conclude that

$$\kappa_{A+K} + \kappa_{A_2+K_2} - \dim\ V_1 = \kappa_{A_1+K_1} - \dim\ V_2 - \dim\ V_1 = \kappa_{A_1+K_1} - \alpha_A.$$

This, and Formula (8.14) together imply that

$$\kappa_{A+K} = \kappa_{A_1+K_1} - \alpha_A = \beta_A - \alpha_A = \kappa_A$$

for every $A \in D(X \to Y)$ and $K \in K(X \to Y)$.

Suppose now that $A \in D^+(X \to Y)$, i.e. $\alpha_A < +\infty$, $\beta_A = +\infty$. Then, by Formula (8.14), $\kappa_{A_1+K_1} = +\infty$. Since $\alpha_A < +\infty$, we find $\kappa_{A+K} = +\infty = \kappa_A$.

Finally, suppose that $A \in D^-(X \to Y)$, i.e. $\alpha_A = +\infty$, $\beta_A < +\infty$. Then, by Formula (8.14), we have $\kappa_{A_1+K_1}$ ¡ $\beta_A < +\infty$. But $\alpha_A = +\infty$, hence $\kappa_{A+K} = \kappa_{A_1+K_1} - \alpha_A =$ - ∞ $= \kappa_A$¿ ■

Theorem 8.7. *Suppose that the set*

$$D_0(X \to Y) = L_0(X \to Y) \cap D(X \to Y)$$

is non-empty [*)]. *If a linear operator $K \in L_0(X \to Y)$ is a perturbation of the class $D_0(X \to Y)$ then it is finite dimensional.*

Proof. Suppose that an operator $K \in L_0(X \to Y)$ is a perturbation of the class $D_0(x \to Y)$ and it is not finite dimensional. Then There exists a sequence $\{y_n\}$ of linearly independent elements belonging to the set $\mathcal{E}_A = KX$. By the definition of the sequence $\{y_n\}$, there is a sequence $\{x_n\} \subset X$ such that $Kx_n = y_n$ $(n = 1, 2, ...)$. Let

$$X_0 = \text{lin}\ \{x_n\}, \qquad Y_0 = \text{lin}\ \{y_n\}.$$

Decompose the space X onto the direct sum $X = X_0 \oplus V$. The subspace V has the infinite codimension.

Let $A \in D_0(X \to Y)$ be arbitrarily fixed. The set $V_1 = AV$ has also the infinite codimension and only a finite number of elements y_n belong to this set. Indeed, suppose that there is a subsequence $\{y_{n_k}\}$ such that $y_{n_k} \in AV$. This means that there is a sequence $\{x'_{n_k}\}$ such that $Ax'_{n_k} = y_{n_k}$. Write $x''_{n_k} = x'_{n_k} - x_{n_k}$. Then $Ax''_{n_k} = 0$, i.e. $x''_{n_k} \in \ker\ A$. Since elements x''_{n_k} are linearly independent, we conclude that $\alpha_A = +\infty$, a contradiction with our assumption that A has a finite d-characteristic.

Define a linear operator $B \in L_0(X \to Y)$ by means of the equalities:

$$Bx = \begin{cases} Ax & \text{if } x \in V, \\ y_n & \text{if } x = x_n \text{ and } y_n \notin V_1, \\ 0 & \text{if } x = x_n \text{ and } y_n \in V_1. \end{cases}$$

Our previous considerations lead to the conclusion that $\alpha_B < +\infty$. On the other hand,

$$\mathcal{E}_B = BX = V_1 \oplus Y_0 = AV \oplus AX_0 = AX = \mathcal{E}_A.$$

Hence $\beta_B = \beta_A < +\infty$, i.e. $B \in D_0(x \to Y)$. However, the operator $B - K$ has not a finite d-characteristic, since $(B - K)x_n = 0$ for x_n such that $y_n \notin V_1$, i.e. for infinitely many x_n. This implies that $\alpha_{B-K} = +\infty$.Then the operator $-K$ is not a perturbation of

[*)] If the bases in linear spaces X and Y are not equipotent then linear operators with a finite dimensional characteristic mapping X into Y do not exist.

the operator $B \in D_0(X \to Y)$. This contradicts to our assumption. Then K is a finite dimensional operator. ■

Observe that the inequality $\beta_A < +\infty$ has been applied in the proof of Theorem 8.7 only in order to show that $\beta_B < +\infty$. This, and Theorems 8.6 and 8.7 together imply

Theorem 8.8. *Suppose that the class*

$$D_0^+(X \to Y) = L_0(X \to Y) \cap D^+(X \to Y),$$

$$(D_0(X \to Y) = L_0(X \to Y) \cap D(X \to Y), \quad \textit{respectively}),$$

is non-empty. Then a necessary and sufficient condition for an operator $K \in L_0(X \to Y)$ to be a perturbation of all operators with a finite nullity (with a finite dimensional characteristic) belonging to $L_0(X \to Y)$ is that K is a finite dimensional operator.

We shall consider now arbitrary algebras of linear operators mapping a linear space X into itself. We shall assume here and in the sequel that algebras under consideration contain the identity operator I. Recall that, in particular, the set $L_0(X)$ is an algebra.

Theorem 8.9. *Every algebra $\mathcal{X}$ with the unit e can be represented as an algebra of linear operators over a linear space X.*

Proof. Define the space X as a space equal to the given algebra $\mathcal{X}$. To every element $x \in X$ there corresponds a linear operator A_x defined by means of the equality

$$A_x y = xy \qquad for \quad y \in X.$$

By definition, $A_x \in L_0(X)$. It is easy to verify that $A_x + A_y = A_{x+y}$ and $A_x A_y = A_{xy}$ for arbitrary $x, y \in X$. Moreover, $A_0 = I$. Hence the set $\{A_x : x \in X\}$ is an algebra of linear operators over the space X. ■

Theorem 8.10. *If $\mathcal{X}(X)$ is an algebra of linear operators over a linear space $X \neq \{0\}$ then the set $K_{\mathcal{X}}(X)$ of all finite dimensional linear operators belonging to this algebra is an ideal in the algebra $\mathcal{X}(X)$. This ideal is proper if dim $X = +\infty$.*

Proof. Suppose that $K_1, K_2 \in K_{\mathcal{X}}(X)$, i.e. dim $K_1X = n_1 < +\infty$ and dim $K_2X = n_2 < +\infty$. Since $(K_1+K_2)X \subset K_1X+K_2X$, we have dim $(K_1+K_2)X =$ dim $K_1X +$ dim $K_2X = n_1 + n_2 < +\infty$. Then the operator $K_1 + K_2$ is also finite dimensional.

Suppose now that the operators $A \in \mathcal{X}(X)$ and $K_{\mathcal{X}}(X)$ are arbitrary. By our assumptions, dim $KX < +\infty$ and $AX \subset X$. Then $AKX \subset KX$ and

$$\dim AKX \leq \dim KX < +\infty,$$

i.e. the operator AK is finite dimensional. Similarly, the operator KA is finite dimensional. We therefore conclude that $K_{\mathcal{X}}(X)$ is an ideal in the algebra $\mathcal{X}(X)$. Observe that $I \notin K_{\mathcal{X}}(X)$ whenever dim $X = +\infty$, because dim $IX =$ dim $X = +\infty$. Then, in the case when dim $X = +\infty$, the ideal $K_{\mathcal{X}}(X)$ is proper.

■

Let be given an algebra $\mathcal{X}(X) \subset L_0(X)$ of linear operators and a proper ideal $\mathcal{J} \subset \mathcal{X}(X)$. As before, assume that $I \in \mathcal{X}(X)$. If for a given operator $A \in \mathcal{X}(X)$ there is a linear operator $R_A \in \mathcal{X}(X)$ such that

$$R_A A = I + T_1 \qquad (AR_A = I + T_2,\ \textit{respectively}), \quad \textit{where } T_1, T_2 \in \mathcal{J},$$

then R_A is said to be a *left* (*right regularizer* of the operator A to the ideal $\mathcal{J}$. If it will not lead to a misunderstanding then we shall omit words "to the ideal $\mathcal{J}$". If an operator R_A is simultaneously a left and a right regularizer then it is called a *simple regularizer* of the operator A to the ideal $\mathcal{J}$. Observe that in the case when $\mathcal{J} = \{0\}$ a left (right) regularizer is a left (right) inverse of the operator A. Hence the assumption that the ideal $\mathcal{J}$ is proper is, indeed, essential.

Theorem 8.11. *Suppose that $\mathcal{J}$ is a proper ideal in an algebra $\mathcal{X}(X) \subset L_0(X)$ of linear operators and that $A \in \mathcal{X}(X)$. Then regularizers of the operator A (provided that they exist) have the following properties:*

(i) If there is a left (right, simple) regularizer R_A to the ideal $\mathcal{J}$ then $R_A \notin \mathcal{J}$.

(ii) If there is a left (right, simple) regularizer R_A to the ideal $\mathcal{J}$ then the coset $[A]$ in the quotient algebra $\mathcal{X}(X)/\mathcal{J}$ is left invertible (right invertible, invertible).

(iii) If the operator A has a left regularizer R_1 and a right regularizer R_2 to the ideal $\mathcal{J}$ then both, R_1 and R_2, are simple regularizers and $R_2 - R_1 \in \mathcal{J}$. Then a simple regularizer is uniquely determined up to a component belonging to the ideal $\mathcal{J}$.

(iv) If $A = B + T$, where B has a left (right) inverse $B_1 \in \mathcal{X}(X)$ and $T \in \mathcal{J}$ then the operator B_1 is a left (right) regularizer of the operator A to the ideal $\mathcal{J}$. Conversely, if a left (right) regularizer R_A to the ideal $\mathcal{J}$ has a left (right) inverse $B \in \mathcal{X}(X)$ then $A = B + T$, where $T \in \mathcal{J}$.

(v) The operator A has an invertible simple regularizer to the ideal $\mathcal{J}$ if and only if $A = B + T$, where B is an invertible linear operator and $T \in \mathcal{J}$. If it is the case, then $R_A = B^{-1}$.

(vi) If the operator A has a left (right, simple) regularizer to the ideal $\mathcal{J}$ then for every $T \in \mathcal{J}$ the operator $A + T$ has a left (right, simple) regularizer R_{A+T} to the ideal $\mathcal{J}$ and $R_{A+T} = R_A$¿

(vii) If the operators $A, B \in \mathcal{X}(X)$ have left (right, simple) regularizers R_A, R_B to the ideal $\mathcal{J}$ and there exists a superposition AB then the operator AB has a left (right, simple) regularizer R_{AB} to the ideal $\mathcal{J}$ and $R_{AB} = R_B R_A$.

Proof. (i) Suppose that a left regularizer $R_A \in \mathcal{J}$. Since, by our assumption, $R_A A = I + T$, where $T \in \mathcal{J}$, we find $I = R_A A_T \in \mathcal{J}$. This contradicts to our assumption that the ideal $\mathcal{J}$ is proper. Similar proofs for right and simple regularizers.

(ii) If the operator A has a left regularizer R_A to the ideal $\mathcal{J}$ then $R_A A = I + T$, where $T \in \mathcal{J}$. Then in the quotient algebra $\mathcal{X}(X)/\mathcal{J}$ for the corresponding cosets the following equality hold: $[R_A][A] = [I]$, i.e. the coset $[A]$ is left invertible. A similar proof for a right regularizer. If a regularizer R_A is simple then $[R_A][A] = [A][R_A] = [I]$. Hence the coset $[A]$ is invertible and $[A]^{-1} = [R_A]$ *).

(iii) Suppose that R_1 and R_2 are left and right regularizers of the operator A to the ideal $\mathcal{J}$. The there are linear operators $T_1, T_2 \in \mathcal{J}$ such that

$$R_1 A = I + T_1, \qquad AR_2 = I + T_2.$$

These equalities imply that in the quotient algebra $\mathcal{X}(X)/\mathcal{J}$ the following equalities hold:

$$[R_1][A] = [I], \qquad [A][R_2] = [I].$$

Then the coset $[A]$ is simultaneously left and right invertible, i.e. this coset is invertible. Since an inverse (if it exists) is unique, we conclude that $[R_1] = [R_2]$, i.e.

$$[R_1 - R_2] = [R_1] - [R_2] = 0,$$

which implies that $R_1 - R_2 = T \in \mathcal{J}$. Hence

$$R_2 A = (R_1 + T)A = R_1 A + TA = I + T_1 + TA = I + T_3, \quad \textit{where } T_3 = T_1 + TA \in \mathcal{J}.$$

Then the operator R_2 is a simple regularizer. In a similar way we prove that the operator R_1 is also a simple regularizer. But $R_2 = R_1 + T$, where $T \in \mathcal{J}$. We therefore conclude that a simple regularizer is uniquely determined up to a component belonging to the ideal $\mathcal{J}$.

(iv) Suppose that $A = B + T$, the operator B has a left inverse B_1 and $T \in \mathcal{J}$. Then $B_1 A = B_1 B + B_1 T = I + B_1 T$, where $B_1 T \in \mathcal{J}$. Then B_1 is a left regularizer of the operator A to the ideal $\mathcal{J}$. A similar proof for the case when the operator B has a right inverse.

Conversely, if a left regularizer R_A of the operator A to the ideal $\mathcal{J}$ has a left inverse $B \in \mathcal{X}(X)$ then the following conditions are satisfied:

$$BR_A = I, \quad \textit{and} \quad R_A A = I + T_1, \quad \textit{where } T_1 \in \mathcal{J}.$$

Then

$$A = (BR_A)A = B(R_A A) = B(I + T_1) = B + T, \quad \textit{where } T = BT_1 \in \mathcal{J}.$$

A similar proof for the case when the operator B is a right inverse.

(v) follows immediately from Point (iv).

(vi) Suppose that the operator A has a left regularizer R_A to the ideal $\mathcal{J}$ and that $T \in \mathcal{J}$. By our assumption, there is an operator $T_1 \in \mathcal{J}$ such that $R_A = I + T_1$. Then

$$R_A(A + T) = R_A A + R_A T = I + T_1 + R_A T = I + T_2, \quad \textit{where } T_2 = T_1 + R_A T \in \mathcal{J}.$$

---

*) Observe that this last equality does not imply that the operator A is invertible.

Then the operator $A+T$ has a left regularizer R_{A+T} to the ideal $\mathcal{J}$ and $R_{A+T}=R_A$. Similar considerations for right and simple regularizers.

(vii) Suppose that linear operators $A, B \in \mathcal{X}(X)$ have left regularizers R_A, R_B (respectively) to the ideal $\mathcal{J}$, i.e.

$$R_A A = I + T_1, \quad R_B B = I + T_2 \qquad \textit{where } T_1, T_2 \in \mathcal{J}.$$

Suppose also that the superposition AB exists. Then

$$R_A(AB) = (R_A A)B = (I+T_1)B = B + T_1 B, \quad \textit{where } T_1 B \in \mathcal{J}.$$

Hence

$$(R_B R_A)(AB) = R_B(B+T_1B) = R_B + R_B T_1 + R_B T_1 B = I + T_2 + R_B T_1 B = I + T_3,$$

$$\textit{where} \quad T_3 = T_2 + R_B T_1 B \in \mathcal{J}.$$

Then the operator $R_{AB} = R_A R_B$ is a left regularizer of the operator AB to the ideal $\mathcal{J}$. Similar considerations for right and simple regularizers. ■

The notion of a regularizer is very useful in studies of properties of linear operators. For instance, if $A, B \in L_0(X)$ then the superpositions AB and BA exist and

$$\ker\,(BA) \subset \ker\, A, \tag{8.15}$$

$$\mathcal{E}_{AB} \subset \mathcal{E}_A \tag{8.15$'$}$$

(cf. Formulae (8.5) and (8.6)). Suppose that the operator A has a simple regularizer to a proper ideal $\mathcal{J} \subset L_0(X)$. Then

$$R_A A = I + T_1, \quad A R_A = I + T_2, \quad \textit{where } T_1, T_2 \in \mathcal{J}.$$

Formula (8.15) implies that in order to study the kernel of the operator A it is enough to examine the kernel of its restriction to the subspace $\ker\,(I+T_1) = \ker\, R_A A \supset \ker\, A$.

Similarly, in order to study the cokernel of the operator A it is enough to consider the operator $\tilde{A}$ induced by A in the quotient space $X/\mathcal{E}_{I+T_2}$, since $\mathcal{E}_{I+T_2} = \mathcal{E}_{AR_A} \subset \mathcal{E}_A$. In the case when $\alpha_{I+T_1} < +\infty$, $\beta_{I+T_2} < +\infty$ it is an essential simplification, since it reduces a problem in infinite dimensional spaces to a problem in finite dimensional spaces. Therefore in the sequel we shall examine linear operators for which a regularization leads to operators $I+T$ with either finite nullity or finite deficiency for all $T \in \mathcal{J}$.

Moreover, if an operator A has a left invertible left regularizer R_A then, by Formula (8.15), it follows that $\ker\,(R_A A) = \ker\, A$. Then $\alpha_A = \alpha_{R_A A} = \alpha_{I+T_1}$. Indeed, if R_1 is a left inverse of the operator R_A then

$$\ker\, A = \ker\,(R_1 R_A A) \supset \ker\,(R_A A) \supset \ker\, A.$$

SImilarly, if A has a right invertible right regularizer R_A then $\mathcal{E}_A = \mathcal{E}_{AR_A}$. Then $\beta_A = \beta_{AR_A} = \beta_{I+T_2}$.

Suppose that $\mathcal{J}$ is a proper ideal in an algebra $\mathcal{X}(X)$ of linear operators over a linear space X. Then $\mathcal{J}$ is said to be a *quasi-Fredholm ideal* if the operator $I+T$ has a finite dimensional characteristic whenever $T \in \mathcal{J}$. The ideal $\mathcal{J}$ is said to be a *Fredholm ideal* if it is a quasi-Fredholm ideal and, moreover, $\kappa_{I+T} = 0$ for every $T \in \mathcal{J}$. Theorem 8.4 implies that the ideal $K_{\mathcal{X}}(X)$ of all finite dimensional operators belonging to the algebra $\mathcal{X}(X)$ is a Fredholm ideal. There exist quasi-Fredholm ideals which are not Fredholm, as it will be shown by the following example.

Example 8.4 (G. Neubauer *)). Let X be the linear space (s) of all complex sequences. Define a linear operator R mapping the space X into itself in the following way:

$$Rx = y, \quad where \quad x = \{x_1, x_2, ...\}, \quad y = \{0, x_1, x_2, ...\}.$$

Write $B = R - I$. Let $\mathcal{X}(X)$ be the algebra of all polynomials in B with complex coefficients. Every operator belonging to $\mathcal{X}(X)$ and different than zero has a finite d-characteristic. Indeed, let $q(B) \in \mathcal{X}(X)$. Clearly, if $q(B) = I$ then $\alpha_{q(B)} = \beta_{q(B)} = 0$. If $q(B) \neq I$ then we can write this polynomial in the form:

$$q(B) = a_0 \prod_{j=1}^{n} (B - a_j I) = a_0 \prod_{j=1}^{n} (R - b_j I), \quad where \ b_j = a_j + 1.$$

Each of operators $R - b_j I$ has a finite d-characteristic. Then, by Theorem 8.1, the operator $q(B)$ has a finite d-characteristic. Hence every proper ideal in $\mathcal{X}(X)$ is a quasi-Fredholm ideal. Consider an ideal $\mathcal{J}$ which consists of operators of the form $A = (I+B)q(B)$. The ideal $\mathcal{J}$ is not a Fredholm ideal, since the equality $q(B) = I$ implies that $A = I + B = R$. Hence $\beta_{I+B} = \beta_R = 1$ and $\alpha_{I+B} = \alpha_R = 0$. Then $\kappa_{I+B} = 1 \neq 0$. □

An immediate consequence of Theorem 8.2 and Corollary 8.1 is

Theorem 8.12. *Suppose that $\mathcal{J}$ is a quasi-Fredholm ideal in an algebra $\mathcal{X}(X)$ of linear operators. If a linear operator $A \subset \mathcal{X}(X)$ has a simple regularizer R_A to the ideal $\mathcal{J}$ then A has a finite d-characteristic. Moreover, if $\mathcal{J}$ is a Fredholm ideal then $\kappa_A = -\kappa_{R_A}$.*

An algebra $\mathcal{X}(X) \subset L_0(X)$ of linear operators is said to be *regularizable to an ideal* $\mathcal{J} \subset \mathcal{X}(X)$ if every operator with a finite d-characteristic has a simple regularizer to the ideal $\mathcal{J}$. An algebra $\mathcal{X}(X)$ regularizable to the ideal $K_{\mathcal{X}}(X)$ of all finite dimensional operators will be called shortly a *regularizable algebra*.

Theorem 8.13. *If an algebra $\mathcal{X}(X) \subset L_0(X)$ of linear operators is regularizable to a quasi-Fredholm ideal $\mathcal{J} \subset \mathcal{X}(X)$ then every operator $T \in \mathcal{J}$ is a perturbation of the class $D(X \to X) \cap \mathcal{X}(X)$. Moreover, if $\mathcal{J}$ is a Fredholm ideal then perturbations $T \in \mathcal{J}$ preserve the index, i.e.*

$$\kappa_{A+T} = \kappa_A \quad for\ every\ A \in D(X \to X) \cap \mathcal{X}(X)\ and\ T \in \mathcal{J}.$$

*) cf. PRR[1], p.42.

Proof. Let $A \in D(X \to X) \cap \mathcal{X}(X)$. By our assumption, there is a simple regularizer R_A of the operator A to the ideal $\mathcal{J}$. Point (vi) of Theorem 8.11 implies that R_A is also a simple regularizer of operators $A + T$ for every $T \in \mathcal{J}$. By Theorem 8.12, the operator $A + T$ has a finite d-characteristic for every $T \in \mathcal{J}$. The arbitrariness of the operator A implies that all operators $T \in \mathcal{J}$ are perturbations of the class $D(X \to X) \cap \mathcal{X}(X)$. If, moreover, $\mathcal{J}$ is a Fredholm ideal then

$$\kappa_{A+T} = -\kappa_{R_{A+T}} = -\kappa_{R_A} = \kappa_A.$$

■

Theorem 8.14. *Suppose that $\mathcal{J}$ and $\mathcal{J}_1$ are quasi-Fredholm ideals in an algebra $\mathcal{X}(X) \subset L_0(X)$ of linear operators and that the algebra $\mathcal{X}(X)$ is regularizable to the ideal $\mathcal{J}_1$. Then every operator $T \in \mathcal{J}$ is a perturbation of the class $D(X \to X) \cap \mathcal{X}(X)$. Moreover, if $\mathcal{J}$ and $\mathcal{J}_1$ are Fredholm ideals then these perturbations preserve the index, i.e.*

$$\kappa_{A+T} = \kappa_A \quad \textit{for every } A \in D(X \to X) \cap \mathcal{X}(X) \textit{ and } T \in \mathcal{J}.$$

Proof. Write $\tilde{\mathcal{J}} = \mathcal{J} + \mathcal{J}_1$. Clearly, $\tilde{\mathcal{J}}$ is a linear set. We shall show that $\tilde{\mathcal{J}}$ is an ideal in the algebra $\mathcal{X}(X)$. Indeed, if $T \in \mathcal{J}$, $T_1 \in \mathcal{J}_1$, $A \in \mathcal{X}(X)$ then

$$A(T + T_1) = AT + AT_1 \in \mathcal{J} + \mathcal{J}_\infty = \tilde{\mathcal{J}},$$

$$(T + T_1)A = TA + T_1A \in \mathcal{J} + \mathcal{J}_\infty = \tilde{\mathcal{J}}.$$

The ideal $\tilde{\mathcal{J}}$ is quasi-Fredholm. Indeed, by Theorem 8.13, if the operator $I + T$ has a finite d-characteristic then also the operator $I + T + T_1$ has a finite d-characteristic. The algebra $\mathcal{X}(X)$ is regularizable to the ideal $\mathcal{J}_1$. So that there is a simple regularizer R_A of any operator $A \in \mathcal{X}(X)$ to the ideal $\mathcal{J}_1$, hence also to the ideal $\tilde{\mathcal{J}}$. By Point (vi) of Theorem 8.11, for every $T \in \tilde{\mathcal{J}}$, in particular, for every $T \in \mathcal{J}$, the operator R_A is a simple regularizer of the operator $A + T$ to the ideal $\tilde{\mathcal{J}}$. Theorem 8.12 implies that $A + T$ has a finite d-characteristic. If $\mathcal{J}$ and $\mathcal{J}_1$ are Fredholm ideals then also $\tilde{\mathcal{J}}$ is a Fredholm ideal. Indeed, by Theorem 8.13, we have $\kappa_{I+T+T_1} = \kappa_{I+T} = 0$ for all $T \in \mathcal{J}$, $T_1 \in \mathcal{J}_1$. Again by Theorem 8.13, we conclude that $\kappa_{A+T} = \kappa_A$ for all $A \in D(X \to X) \cap \mathcal{X}(X)$. ■

An immediate consequence of the last theorem is

Corollary 8.5. *Suppose that $\mathcal{J}$ is a quasi-Fredholm ideal in a regularizable algebra $\mathcal{X}(X) \subset L_0(X)$ of linear operators. Then every operator $T \in \mathcal{J}$ is a perturbation of the class $D(X \to X) \cap \mathcal{X}(X)$. Moreover, if $\mathcal{J}$ is a Fredholm ideal then these perturbations preserve the index, i.e.*

$$\kappa_{A+T} = \kappa_A \quad \textit{for every } A \in D(X \to X) \cap \mathcal{X}(X) \textit{ and } T \in \mathcal{J}.$$

Suppose that X is a ring with unit e. Then the set

$$R[X] = \{x \in X : \textit{elements } e + axb \textit{ are invertible for all } a, b \in X\}$$

is said to be the *radical* of the ring X (cf. JACOBSON, J[1]). It is easy to verify that the radical $R[X]$ is an ideal in X.

Theorem 8.15. *Suppose that an algebra $\mathcal{X}(X) \subset L_0(X)$ is regularizable to a quasi-Fredholm ideal $\mathcal{J} \subset \mathcal{X}(X)$. Denote by $R[\mathcal{X}(X)/\mathcal{J}]$ the radical of the quotient algebra $\mathcal{X}(X)/\mathcal{J}$ and write*

$$\mathcal{J}_0 = \{U \in \mathcal{X}(X) : [U] \in R[\mathcal{X}(X)/\mathcal{J}]\},$$

where $[U]$ is a coset determined by an operator $U \in \mathcal{X}(X)/\mathcal{J}$. Then (i) The set $\mathcal{J}_0$ is a quasi-Fredholm ideal in the algebra $\mathcal{X}(X)$.

(ii) Every operator $T \in \mathcal{J}_0$ is a perturbation of the class $D(X \to X) \cap \mathcal{X}(X)$.

(iii) If $\mathcal{J}_1$ is a quasi-Fredholm ideal in the algebra $\mathcal{X}(X)$ then $\mathcal{J}_1 \subset \mathcal{J}_0$, i.e. $\mathcal{J}_0$ is a maximal quasi-Fredholm ideal *in $\mathcal{X}(X)$.*

(iv) If, in addition, the algebra $\mathcal{X}(X)$ is regularizable then

$$\mathcal{J}_0 = \mathcal{K}_0 = \{U \in \mathcal{X}(X) : [U] \in R[\mathcal{X}(X)/K_{\mathcal{X}}(X)\}.$$

Proof. (i) Since the radical is an ideal, the set $\mathcal{J}_0$ is also an ideal in the algebra $\mathcal{X}(X)$. If $U \in \mathcal{J}_0$ is arbitrarily chosen then, by the definition of the radical, it follows that the coset $[I]+[U]$ is invertible in the quotient algebra $\mathcal{X}(X)/\mathcal{J}$. Then there is a coset $[V] \in \mathcal{X}(X)/\mathcal{J}$ such that

$$[V]([I] + [U]) = ([I] + [U])[V] = [I].$$

This implies that for every operator $V \in [V]$

$$V(I + U) = I + T_1, \quad (I + U)V = I + T_2, \quad \textit{where} \quad T_1, T_2 \in \mathcal{J}.$$

Hence the operator $I + U$ has a simple regularizer to the ideal $\mathcal{J}$. This, and Theorem 8.12 together imply that the operator $I + U$ has a finite d-characteristic. The arbitrariness of U implies that $\mathcal{J}_0$ is a quasi-Fredholm ideal.

(ii) The proof of this point is an immediate consequence of Point (i) of this theorem and Theorem 8.14.

(iii) Suppose that $U \in \mathcal{J}_1$. Then $AUB \in \mathcal{J}_1$ for arbitrary $A, B \in \mathcal{X}(X)$. Since the algebra $\mathcal{X}(X)$ is regularizable to the ideal $\mathcal{J}$, the operator $I + AUB$ has a simple regularizer to the ideal $\mathcal{J}$. Then the coset$[I] + [A][U][B]$ is invertible in the quotient algebra $\mathcal{X}(X)/\mathcal{J}$. The arbitrariness of operators A and B implies that the coset $[U]$ belongs to the radical $R[\mathcal{X}(X)/\mathcal{J}]$. Hence $U \in \mathcal{J}_0$ and $\mathcal{J}_1 \subset \mathcal{J}_0$. We therefore conclude that $\mathcal{J}_0$ is a maximal quasi-Fredholm ideal in the algebra $\mathcal{X}(X)$.

(iv) Observe that the construction of the ideal $\mathcal{J}_0$ does not depend of the ideal $\mathcal{J}$. Indeed, if it is not the case, then the ideal $\mathcal{J}_0$ is not maximal. This, and our assumption that the algebra $\mathcal{X}(X)$ is regularizable together imply that $\mathcal{J}_0 = \mathcal{K}_0$. ■

A characterization of the index is given by the following

Theorem 8.16. *Suppose that $\mathcal{J}$ is a quasi-Fredholm ideal in an algebra $\mathcal{X}(X) \subset L_0(X)$. Suppose, moreover, that the set $W \subset D(X \to X) \cap \mathcal{X}(X)$ satisfies the following conditions:*

(w_1) if $A, B \in W$ then $AB \in W$;

(w_2) every operator $T \in \mathcal{J}$ is a perturbation of the class W;

(w_3) every operator $A \in W$ has a simple regularizer $R_A \in W$ to the ideal $\mathcal{J}$.

If an integer-valued function $\nu(A)$ defined on the set W satisfies the following conditions:

(i) $\nu(A+T) = \nu(A)$ for every $A \in W$ and $T \in \mathcal{J}$,

(ii) $\nu(AB) = \nu(A) + \nu(B)$ for all $A, B \in W$,

(iii) if $\kappa_A = 0$ then $\nu(A) = 0$,

then there is a number $p \in \mathbb{R}$ such that

$$\nu(A) = p\kappa_A \qquad for\ all\ A \in W. \tag{8.16}$$

In particular, if there is an operator $S \in W$ such that $\kappa_S = 1$ then the number p in Formula (8.16) is an integer.

Proof. To begin with, suppose that there is an operator $S \in W$ with the index $\kappa_S = 1$. By Condition (w_3), there is a simple regularizer R_S of the operator S to the ideal $\mathcal{J}$. This, and Theorem 8.12 together imply that $\kappa_{R_S} = -\kappa_S = -1$.

Suppose now that $A \in W$ and that A has the positive index: $\kappa_A = n > 0$. Theorem 8.11 and Condition (w_1) together imply that $(R_S)^n A \in W$ and $\kappa_{(R_S)^n A} =$ - $n\kappa_S + n = -n + n = 0$. Then, by Point (iii), it follows that $\nu(A) = -\nu(R_S^n) = -n\nu(R_S) = p_1\kappa_A$, where $p_1 = -\nu(R_S)$.

If $A \in W$ and $\kappa_A = n \leq 0$ then in a similar way we prove that $\nu(A) + P\kappa_A$, where $p = \nu(S)$. However, there is an operator $T \in \mathcal{J}$ such that $R_S S = I + T$. Then $\nu(R_S S) = \nu(I+T) = \nu(I) = 0$. Hence $\nu(R_S) =$ - $\nu(S)$, which implies that $p_1 = -\nu(R_S) = \nu(S) = p$. Finally, we get $\nu(A) = p\kappa_A$ for all $A \in W$, where p is an integer.

Suppose now that an operator S belonging to W with the index equal to 1 does not exist. Let q be the least positive index of linear operators belonging to the set W. The number q is a divisor of the index of every operator from the set W. Indeed, let operators $A, B \in W$ have the indices q and s, respectively, where $s = nq + r$, $0 < r < q$. By our assumptions, the operator A has a simple regularizer R_A to the ideal $\mathcal{J}$ and $\kappa_{R_a} = -\kappa_A = -q$. Again by our assumptions, $(B(R_A)^n \in W$ and

$$\kappa_{B(R_A)^n} = \kappa_B - n\kappa_A = s - nq = r < q,$$

which is a contradiction with the definition of the number q. Then $r = 0$ and the index of the operator B is divisible by q. Furthermore the proof is going on the same lines as in the first part (where we have assumed that there exists an operator with the index 1), i.e.

we consider an arbitrary operator $A \in W$ with the index nq and an operator $S \in W$ such that $\kappa_S = q$. Since the index of the operator $A(R_S)^n$ is equal zero, we find

$$\nu(A) = n\nu(S) = \frac{np}{q}\kappa_S = \frac{p}{q}\kappa_A,$$

where the number $p + \nu(S)$ is an integer. ■

Theorem 8.17. *Suppose that $A \in L_0(X \to Y)$. Then*

$$A \in \begin{cases} D^+(X \to Y) \\ D^-(X \to Y) \end{cases} \quad \textit{if and only if} \quad A = S + K,$$

where the operators $S, K \in L_0(X \to Y)$, S is left (right) invertible and K is finite dimensional.

Proof of the necessary condition. Suppose that $A \in D^+(X \to Y)$. Then $\alpha_A < +\infty$, $\beta_A = +\infty$. Decompose the space X onto the direct sum $X = \ker\ A \oplus V$ and the space Y onto the direct sum $Y = \mathcal{E}_A \oplus V_1$. Then dim ker $A = \alpha_A \leq \beta_A =$ dim V_1. This implies that there is a finite dimensional operator K which is a one-to-one mapping of the subspace ker A into the subspace V_1. Hence the operator $S = A - K$ defined on the space X maps X into Y in a one-to-one way. Then the operator S is invertible and its inverse S^{-1} is defined on the set $\mathcal{E}_S = SX \subset Y$. Write $S^{(-1)}$ for an arbitrary extension of the operator S^{-1} onto the whole space Y. Clearly, $S^{(-1)}Sx = x$ for all $x \in X$. We therefore conclude that the operator S is left invertible and $A = S + K$.

Suppose now that $A \in D^-(X \to Y)$. Then $\alpha_A = +\infty$ and $\beta_A < +\infty$. Similarly, as in the previous case, decompose spaces X and Y onto direct sums. Then we conclude that dim $V_1 = \beta_A \leq \alpha_A =$ dim ker A. The operator A maps in a one-to-one way the subspace V onto the set $\mathcal{E}_A = AX$. Hence there exists a finite dimensional operator K mapping a subspace of the set ker A onto the whole subspace V_1 and this mapping is one-to-one. Then the operator $S = A - K$ maps a subspace X_1 of the space X onto the whole space Y. The operator S is right invertible on the subspace X_1 and $A = S + K$.

Proof of the sufficient condition. Suppose that $A = S + K$, where $K \in K(X \to Y)$ and S is left (right) invertible. By Theorem 8,3, we have $\alpha_S = 0$ ($\beta_S = 0$, respectively). Hence $S \in D^+(X \to Y)$ ($S \in D^-(X \to Y)$, respectively). Then Theorem 8.6 implies that $A = S + K \in D^+(X \to Y)$ ($A \in D^-(X \to Y)$). ■

Theorem 8.18. *Suppose that $A \in D(X \to Y) \cap L_0(X)$. Then $\kappa_A = 0$ if and only if $A = S + K$, where S is an invertible operator and K is a finite dimensional operator.*

Proof. The proof of this theorem is the same as the proof of Theorem 8.1, but in the proof of the necessary condition we are applying the equalities $\alpha_A = 0$, $\beta_A = 0$. The first part of the proof is like this for the case $\alpha_A \leq \beta_A$, the second one - like for the case $\beta_A \leq \alpha_A$. In the proof of the sufficient condition we are using the fact that $\alpha_S = \beta_S =$), hence $S \in D(X \to Y)$. ■

Corollary 8.6. *Suppose that $A \in L_0(X \to Y)$. If*

$$A \in \begin{cases} D^-(X \to Y) \\ D^+(X \to Y) \end{cases} \quad \textit{then there is an operator } R_A \textit{ such that } \begin{cases} AR_A - I \in K(X \to Y), \\ R_A A - I \in K(X \to Y), \end{cases}$$

respectively.

Proof. Indeed, if $\alpha_A < +\infty$ then $R_A = S^{(-1)}$, if $\beta_A < +\infty$ then $R_A = S^{-1}$, where these operators are defined in the first and second part of the proof of Theorem 8.17, respectively. ■

An immediate consequence of the last corollary is

Corollary 8.7. *Suppose that $A \in L_0(X)$. If*

$$A \in \begin{cases} D^+(X \to X) \\ D^-(X \to X) \end{cases} \quad \textit{then } A \textit{ has a } \begin{cases} \textit{left} \\ \textit{right} \end{cases} \textit{ regularizer to the ideal } K(X \to X)$$

of all finite dimensional operators belonging to $L_0(X)$, respectively.

Similarly, as Theorem 8.17 implies Corollaries 8.6 and 8.7, Theorem 8.18 implies the following

Corollary 8.8. *If $A \in D(X \to X) \cap L_0(X)$ then A has a simple regularizer to the ideal $K(X \to X)$ of all finite dimensional operators belonging to $L_0(X)$.*

Corollary 8.9. *The algebra $L_0(X)$ is regularizable.*

Let be given an algebra $\mathcal{X}(X) \subset L_0(X)$ of linear operators and a left (right) ideal $\mathcal{J} \subset \mathcal{X}(X)$ $*$). Then $\mathcal{J}$ is said to be *positive* (*negative*) *semi-Fredholm ideal* if $\alpha_{I+T} < +\infty$ ($\beta_{I+T} < +\infty$) for every $T \in \mathcal{J}$. The algebra $\mathcal{X}(X)$ is left (right) regularizable to a left (right) ideal $\mathcal{J} \subset \mathcal{X}(X)$ if every operator $A \in \mathcal{X}(X)$ with a finite nullity (deficiency) has a left (right) regularizer to the ideal $\mathcal{J}$.

Theorem 8.19. *Suppose that an algebra $\mathcal{X}(X) \subset L_0(X)$ of linear operators is left (right) regularizable to a positive (negative) semi-Fredholm ideal $\mathcal{J} \subset \mathcal{X}(X)$. Then*

(i) the operator $A + T_0$ has a finite nullity (deficiency) for every operator $A \in \mathcal{X}(X)$ with a finite nullity (deficiency) and for every $T_0 \in \mathcal{J}$;

(ii) every operator belonging to a positive (negative) semi-Fredholm ideal $\mathcal{J}_1 \subset \mathcal{X}(X)$ is a perturbation of the class of all linear operators with a finite nullity (deficiency) belonging to $\mathcal{X}(X)$.

Proof. (i) By our assumption, there is an operator $R_A \in \mathcal{X}(X)$ such that $R_A A = I + T$ ($AR_A = I + T$, respectively), where $T \in \mathcal{J}$. Then for every $T_0 \in \mathcal{J}$

$$R_A(A + T_0) = R_A A + R_A T_0 = I + T + R_A T_0 = I + T_1, \quad \textit{where } T_1 = T + R_A T_0 \in \mathcal{J},$$

$*$) cf. for instance, Jacobson J[1], also PRR[1].

$$(A+T_0)R_A = AR_A + T_0R_A = I + T + T_0R_a = I + T_2, \quad \textit{where } T_2 = T + T_0R_A \in \mathcal{J},$$

respectively. By the assumption about the ideal $\mathcal{J}$, it follows that the operator $I + T_1$ has a finite nullity (the operator $I + T_2$ has a finite deficiency). This, and Theorem 8.2 together imply that the operator $A + T_0$ has a finite nullity (deficiency).

(ii) Let $\tilde{\mathcal{J}} = \mathcal{J} + \mathcal{J}_1$. In a similar way, as in the proof of Theorem 8.14, we show that $\tilde{\mathcal{J}}$ is a positive (negative) semi-Fredholm ideal in $\mathcal{X}(X)$ and that the algebra $\mathcal{X}(X)$ is left (right) regularizable to the ideal $\tilde{\mathcal{J}}$. This, and Point (i) together imply that operators belonging to $\tilde{\mathcal{J}}$, in particular, operators belonging to $\mathcal{J}_1$, are perturbations of the class of all operators with a finite nullity (deficiency) belonging to $\mathcal{X}(X)$. ∎

Theorem 8.20. *If an algebra $\mathcal{X}(X) \subset L_0(X)$ is regularizable and $T = A + B$ for every $T \in \mathcal{X}(X)$, where the operators A, B have finite nullities (deficiencies) then the set Π of all perturbations of the class of all operators with a finite nullity (deficiency) is a left (right) ideal in the algebra $\mathcal{X}(X)$.*

Proof. Theorem 8.5 implies that $A + B \in \Pi$ whenever $A, B \in \Pi$. Let $V \in \mathcal{X}(X)$ be a perturbation of the class of all operators with a finite d-characteristic and let $A \in \mathcal{X}(X)$ have a finite nullity (deficiency). Let $B \in \mathcal{X}(X)$ be an operator such that the superposition BV (VB, respectively) is well defined. Since the algebra $\mathcal{X}(X)$ is regularizable, we conclude that there an operator $R_B \in \mathcal{X}(X)$ such that

$$BR_B = I + K \quad (R_BB = I + K < \textit{ respectively}) \quad \textit{where } K \in K_{\mathcal{X}}(X).$$

Theorems 8.1 and 8.6 together imply that the operator

$$A + BV = (BR_B - K)A + BV = B(R_BA + V) - KA$$

$$(A + VB = A(R_BB_K) + VB = (AR_B + V)B_AK, \quad \textit{respectively})$$

has a finite nullity (deficiency). Then the operator BV (VB, respectively) is a perturbation of the class of all operators with a finite nullity (deficiency). Our assumption that every operator belonging to $\mathcal{X}(X)$ is a sum of two operators with a finite nullity (deficiency) and the additivity of the set of perturbations together imply our conclusion. ∎

Corollary 8.10 *If every operator belonging to a regularizable algebra $\mathcal{X}(X) \subset L_0(X)$ of linear operators is a sum of two operators with a finite d-characteristic then the set of all perturbations of the class of all operators with a finite d-characteristic belonging to the algebra $\mathcal{X}(X)$ is a quasi-Fredholm ideal in $\mathcal{X}(X)$. By Point (iii) of Theorem 8.15, this ideal is the maximal quasi-Fredholm ideal in $\mathcal{X}(X)$.*

The following theorem can be proved (cf. PRR[1], pp. 61-63):

Theorem 8.21. *Every operator belonging to the algebra $L_0(X)$ is a sum of two isomorphisms. Even more, if the bases in spaces X and Y are equipotent then every operator $A \in L_0(X \to Y)$ is a sum of two isomorphisms of the space X onto the space Y.*

If $\mathcal{X}(X)$ is an arbitrary algebra of linear operators, then the theorem which says that every operator $A \in \mathcal{X}(X)$ is sum of two operators with a finite d-characteristic, is not true. This is shown by the following example.

Example 8.5. Let X be the linear space of all continuous complex-valued functions defined on the whole complex plane. Let $\mathcal{X}(X)$ be the algebra of operators of multiplication by a complex polynomial $p(z)$. If a polynomial $p(z)$ is not a constant then the corresponding operator P has the infinite deficiency β_P. Indeed, the fundamental theorem of algebra implies that there is a number $z_0 \in \mathbb{C}$ such that $p(z_0) = 0$. Observe that for every $x \in X$ there is a constant $c > 0$ such that

$$|x(z) - x(z_0)| \leq c|z - z_0| \quad for \ |z - z_0| < 1 \quad whenever \ x \in \mathcal{E}_P = PX.$$

Write $x_\alpha(z) = |z - z_0|^\alpha$, where $0 < \alpha < 1$. Then $x_\alpha \notin \operatorname{lin}\{\mathcal{E}_P, x_\beta, \beta > \alpha\}$ and $\beta_P = +\infty$. Hence $D(X \to X) = \{aI : a \in \mathbb{C}\}$. We therefore conclude that an operator $P \in \mathcal{X}(X)$, which is not of the form aI, cannot be written as the sum of two operators with a finite d-characteristic. □

Theorems 8.10, 8.11, 8.12, 8.13, 8.14, 8.15, 8.19 and 8,20 and Corollaries 8.5, 8.7, 8.8, 8.9, 8.10, where there are considered algebras $\mathcal{X}(X)$ of linear operators mapping a linear space X into itself, can be generalized for the case of linear operators mapping a linear space into another. Even more, they can be generalized for the class of all linear operators mapping a linear space X_α into a linear space X_β for arbitrary $\alpha, \beta \in \mathcal{A}$, where $\{X_\alpha\}_{\alpha \in \mathcal{A}}$ is a class of linear spaces. In order to do it, there is necessary to introduce and to examine an algebraic structure much more general than an algebra (cf. PR[5]).

Exercises.

Exercise 8.1. Let X be the space (s) of all real sequences. Define the operators S_l and S_r of the *shifts to left* and *to right* by means of formulae

$$S_l x = \{x_2, x_3, ...\}, \quad S_r x = \{0, x_1, x_2, ...\} \quad for \ x = \{x_1, x_2, ...\}.$$

Prove that the operators S_l and S_r have the index different than zero. □

Exercise 8.2. In the space $C[0, 1]$ consider an integral equation with a degenerate kernel:

$$x(t) + \lambda \int_0^t tsx(s)ds = \mu t, \tag{8.17}$$

where λ, μ are real parameters. Prove that

(i) Equation (8.17) has a unique solution for every $\mu \in \mathbb{R}$ whenever $\lambda \neq -3$;

(ii) if $\lambda = -3$ then solution to Equation (8.17) exist if and only if $\mu = \frac{3}{2}$;

(iii) determine these solutions. □

Exercise 8.3. In the space $C[0, \frac{\pi}{2}]$ consider an integral equation with a degenerate kernel:

$$x(t) - \lambda \int_0^{\pi/2} \sin\ (t + s)x(s)ds = \sin\ 2t - \frac{3}{2}, \tag{8.18}$$

where $\lambda \in \mathbb{R}$ is a parameter. Prove that

(i) if $\lambda \neq \frac{2}{1 \pm \pi}$ then Equation (8.18) has a unique solution;

(ii) if $\lambda = \frac{2}{1 \pm \pi}$ then there exist at least two solutions to Equation (8.18);

(iii) determine the nullity of the integral operator appearing in Equation (8.18). □

Exercise 8.4. Let $T \in L_0(X)$ and let λ_0 be an eigenvalue of T. A principal space X_{λ_0} corresponding to the eigenvalue λ_0 is said to be *splittable* if

$$X = X_{\lambda_0} \oplus N_{\lambda_)}, \tag{8.19}$$

where N_{λ_0} is an invariant subspace for the operator T, i.e. $TN_{\lambda_0} \subset N_{\lambda_0}$ and $N_{\lambda_0} = (T - \lambda_0 I)N_{\lambda_0}$. Prove that

(i) if the subspace X_{λ_0} is finite dimensional then the decomposition (8.19) is uniquely determined;

(ii) if X_{λ_0} is a finite dimensional splittable principal space then the operator $A - \lambda_0 I$ has a finite dimensional characteristic. □

Exercise 8.5. Suppose that $A_1, ..., A_n \in L_0(X)$, $A = A_1 ... A_n$ and $\alpha_A = \beta_A = 0$. Prove that $\alpha_{A_j} = 0$ and $\beta_{A_j} = 0$ for $j = 1, ..., n$. □

Exercise 8.6. Prove that λ is a regular value of the operator T^n, where $T \in L_0(X)$, if and only if the n-th roots $\lambda_1, ..., \lambda_n$ of the number λ are regular values of the operator T (cf. Exercise 8.5). □

Exercise 8.7. The spectrum *) of an operator is said to be *discrete* if is either finite or a denumerable sequence $\{\lambda\}$ tending to zero . Let $T \in L_0(X)$. Prove that the following conditions are equivalent:

(i) The operator T has a discrete spectrum.

(ii) There is a positive integer n such that the operator T^n has a discrete spectrum.

(iii) For every positive integer n the operator T^n has a discrete spectrum. □

Exercise 8.8. Let $T \in L_0(X)$. Prove the following theorem: if there is a positive integer N such that for all $n > N$ the operator T^n has a discrete spectrum and the operator $I - T^n$ has a finite dimensional characteristic and the index zero, then the operator $I - T$ has a finite dimensional characteristic and the index zero (cf. Exercise 8.7). □

Exercise 8.9. Define a shift operator by the formula: $(Sx)(t) = x(t-h)$ for $x \in X$, where X is the linear space of all functions continuous and bounded for $t \in \mathbb{R}$ and the point $h \in \mathbb{R}$ is fixed. Prove that

(i) the operator S is an isomorphism;

(ii) the operator $I - S$ has not a finite dimensional characteristic. □

---

*) cf. Chapter 5.

Chapter 9.

Index of conjugate operators. Generalized Fredholm alternative.

In Chapter 2 we have denoted by X' the space conjugate to a space X, i.e. the space of all linear functionals defined on the space X. A subspace $\Xi \subset X$ is said to be *total* if $\xi(x) = 0$ for every $\xi \in \Xi$ implies $x = 0$ (where $x \in X$).

Theorem 9.1. *The space X' is total.*

Proof. Let $x \neq 0$ be an arbitrary element of the space X. Denote by X_0 a one-dimensional subspace spanned by the element x, i.e. $X_0 = \{tx : t \text{ is a scalar}\}$. Consider a functional f_0 defined on the subspace X_0 by means of the formula: $f_0(tx) = t$. Clearly, f_0 is a linear functional. Corollary 2.2 implies that the functional f_0 can be extended to a functional f defined on the whole space X. Then $f \in X'$ and $f(x) = 1 \neq 0$. The arbitrariness of the element $x \in X$ implies that the space X' is total. ∎

Observe that elements $x \in X$ can be treated as functionals defined on a total subspace $\Xi \subset X'$, since a mapping

$$F_x(\xi) = \xi(x) \quad \textit{for every } x \in X \textit{ and } \xi \in \Xi.$$

Therefore, if we denote by Ξ' the space of all linear functionals defined on Ξ then the whole space X is mapped in a one-to-one way into the subspace Ξ'. This mapping is said to be the *canonical embedding* and is denoted by κ. The image κX of the space X by this mapping is a total space of functionals defined on the space Ξ, for the condition $\xi(x)$ for all $x \in X$ implies $\xi = 0$.

In the sequel every total subspace of the space X' will be also called a *conjugate space* with X.

Let be given two linear spaces X and Y, both over the same field $\mathbb{F}$ of scalars. Suppose that $\mathrm{H} \subset Y'$ is a conjugate space *). Then to every operator $A \in L(X \to Y)$ there corresponds the operator ηA, whose domain is the space H and the set of values is X', defined by means of the equality

$$(\eta A)x = \eta(Ax) \quad \textit{for all } x \in \mathcal{D}_A \textit{ and } \eta \in \mathrm{H}$$

(cf. Formula (2.18)).

The operator ηA is said to be a *conjugate operator* with A and will be denoted by A'. Then, by definition,

$$A'\eta = \eta A \quad \textit{for every } \eta \in \mathrm{H}.$$

Similarly, as in Chapter 2, we prove that $I' = I$ and if the sum $A + B$ is well defined then $(A + B)' = A' + B'$ (cf. Formulae (2.20), (2.21)).

---

*) Here and in the sequel H denotes the Greek capital letter "eta".

Let $\Xi \subset X'$ be an arbitrary conjugate space. Consider operators A' as defined for functionals $\eta \in \mathrm{H}$ such that $A'\eta = \eta A \in \Xi$. In this manner to every operator $A \in L(X \to Y)$ there corresponds an operator $A' \in L(\mathrm{H} \to \Xi)$. With this general formulation it may happens that the operator A' is defined on the set $\{0\}$ only. We therefore shall consider in the sequel only operators $A \in L_0(X \to Y)$ such that $A' \in L_0(\mathrm{H} \to \Xi)$, i.e. operators $A \in L_0(X \to Y)$ such that $A'\eta \in \Xi$ for every $\eta \in \mathrm{H}$. The set of all these operators will be denoted by $L_0(X \to Y, \mathrm{H} \to \Xi)$. Then, by definition,

$$L_0(X \to Y, \mathrm{H} \to \Xi) = \tag{9.1}$$

$$= \{A \in L_0(X \to Y) : A'\eta = \eta A \in \Xi \ for\ every\ \eta \in \mathrm{H}\}.$$

Clearly, this set is a linear space. The space $L_0(X \to Y, \mathrm{H} \to \Xi)$ will be denoted shortly by $L_0(X, \Xi)$. Clearly, the space $L_0(X, \Xi)$ is an algebra for $(AB)' = B'A'$ for all $A, B \in L_0(X, \Xi)$ (cf. Formula (2.22)).

Theorem 9.2. *A finite dimensional operator K defined by means of the equality*

$$Kx = \sum_{j=1}^{n} f_j(x)y_j \quad for\ x \in X,$$

where $f_1, ..., f_n \in X'$ and $y_1, ..., y_n \in Y$ are linearly independent, belongs to the space $L_0(X \to X, \Xi \to \Xi)$ if and only if $f_j \in \Xi$ $(j = 1, ..., n)$.

Proof. Indeed, suppose that $\xi \in \mathrm{H}$. If $f_1, ..., f_n \in \Xi$ then $K'\xi = \xi K = \sum_{j=1}^{n} \xi(y_j)f_j \in \Xi$. ■

For an arbitrary subset $E \subset Y$ and an arbitrary conjugate space $\mathrm{H} \subset Y'$ write

$$E^{\perp} = \{\eta \in \mathrm{H} : \eta(y) = 0 \ for\ all\ y \in E\}. \tag{9.2}$$

The set $E^{\perp}$ is said to be an H-*orthogonal complement* of the set E.

Theorem 9.3. *If $A \in L_0(X \to Y, \mathrm{H} \to \Xi)$ then $\alpha_A \le \beta_{A'}$.*

Proof. Clearly, $\alpha_{A'} = \dim\, \mathcal{E}_A^{\perp} = \dim\, (AX)^{\perp}$. On the other hand, every functional $\eta \in \mathcal{E}_A^{\perp}$ induces a functional in the quotient space $Y/\mathcal{E}_A$. If $\beta_A = \dim\, Y/\mathcal{A} < +\infty$ then the dimension of the conjugate space $(Y/\mathcal{E}_A)'$ is equal to β_A (cf. Corollary 2.3). Then $\alpha_{A'} \le \beta_A$. If $\beta_A = +\infty$ then this inequality is always satisfied. ■

A subspace $E \subset Y$ is said to be H-*describable* if $(E^{\perp})^{\perp} = E$, where

$$(E^{\perp})^{\perp} = \{y \in Y : \eta(y) = 0 \ for\ \eta \in E^{\perp}\} \tag{9.3}$$

and $\mathrm{H} \subset Y'$ is a conjugate space.

An operator $A \in L_0(X \to Y, \mathrm{H} \to \Xi)$ is said to be H-*solvable* if its range $\mathcal{E}_A$ is H-describable.

Clearly, if an operator A is H-solvable then dim $\mathcal{E}_A^{\perp} = \beta_A$. This, and the proof of Theorem 9.3 together imply

Theorem 9.4. *If an operator* $A \in L_0(X \to Y, \mathrm{H} \to \Xi)$ *is* H*-solvable then* $\alpha_{A'} = \beta_A$.

Corollary 9.1. *If* $A \in L_0(X \to Y, \mathrm{H} \to \Xi)$ *and* $\mathcal{E}_A = Y$ *then the operator* $A' \in L_0(\mathrm{H} \to \Xi, X \to Y)$ *maps* H *into* Ξ *in a one-to-one way.*

Proof. Indeed, by our assumption, $\alpha_{A'} = \beta_A = 0$. Then ker $A' = \{0\}$ and the operator A' is invertible on the space H. ■

Corollary 9.2. *If an operator* $A \in L_0(X \to Y, \mathrm{H} \to \Xi)$ *is an isomorphism the conjugate operator* A' *is also an isomorphism.*

Let E be a subspace of a linear space X. Let φ_E be a mapping of the space X into the quotient space X/E such that to every element $x \in X$ there corresponds a coset $[x] = x + E$ induced by the element x, i.e. $\varphi_E x = x + E$ $(x \in X)$. Clearly, the mapping φ_E is a linear operator.

Corollary 9.3. *Suppose that* E *is a subspace of a linear space* X *and that* $\Xi \subset X'$ *is a conjugate space. If* $H \subset (X/E)'$ *is a conjugate space satisfying the condition* $\varphi'_E H \subset \Xi$ *then the operator* φ'_E *defined on the whole space* H *maps* H *in a one-to-one way into the space* $E^{\perp}$.

Proof. By definition, $\mathcal{D}_{\varphi_E} = X$, $\mathcal{E}_{\varphi_E} = \varphi_E X = X/E$. This, and Corollary 9.1 together imply that $\mathcal{D}_{\varphi_E} = \mathrm{H}$ and that φ'_E is a one-to-one mapping of H into Ξ. Suppose that $\xi \in \varphi'_E \mathrm{H}$, however, $\xi \notin E^{\perp}$. Then there is an element $x \in E$ such that $\xi(x) \neq 0$. Since $\xi = \varphi'_E \eta$ for an $\eta \in \mathrm{H}$ and $\varphi_E x = 0$ for every $x \in X$, we conclude that

$$\xi(x) = (\varphi'_E \eta)x = \eta(\varphi_E x) = \eta(0) = 0,$$

which contradicts to our assumption that $\xi(x) = 0$. ■

Theorem 9.5. *Every operator* $A \in L_0(X \to Y)$ *is* Y'*-solvable.*

Proof. Let $y_0 \in Y$ be an arbitrary element which does not belong to the range $\mathcal{E}_A$ of the operator A. Write

$$Y_0 = \mathrm{lin}\ \{y_0 + \mathcal{E}_A\} = \mathrm{lin}\ \{ay + z,\ where\ z = Ax,\ x \in X,\ a\ is\ a\ scalar\}.$$

Define a functional η_0 by means of the formula: $\eta_0(u) = a$ for $u \in Y_0$. By Corollary 2.2, the functional η_0 can be extended to a functional $\tilde{\eta}_0$ defined on the whole space Y. Observe that, by definition, $\tilde{\eta}_0(\mathcal{E}_A) = \eta_0(\mathcal{E}_A) = 0$. Hence $\tilde{\eta}_0 \in \mathcal{E}_A^{\perp}$. Then for every $y_0 \neq\in \mathcal{E}_A$ there is a functional $\tilde{\eta} \in \mathcal{E}_A^{\perp}$ such that $\eta_0(y_0) = 1$. This implies that

$$\mathcal{E}_A = \{y \in Y : \eta(y) = 0\ for\ all\ \eta \in \mathcal{E}_A^{\perp}\} = (\mathcal{E}_A^{\perp})^{\perp},$$

which proves that the operator A is Y'-solvable. ■

Theorems 9.5 and 9.4 together imply

Corollary 9.4. *If $A \in L_0(X \to Y, Y' \to X')$ then $\alpha_{A'} = \beta_A$.*

Theorem 9.4 does not hold without the assumption that the operator under question is H-solvable, even if $X = Y$, $\Xi = \mathrm{H}$. This is shown by the following

Example 9.1. Suppose that $X = Y = C^\infty[0,1]$ and that

$$\Xi = \mathrm{H} = \{\xi : \xi = \xi(x) = \int_0^t x(t)\tilde{\xi}(t)dt, \ where\ \tilde{\xi} \in C^\infty[0,1],\ \tilde{\xi}^{(n)}(0) = 0\ (n \in \mathbb{N})\},$$

where by $\tilde{\xi}^{(n)}$ is denoted the nth derivative of the function $\tilde{\xi}$. Let

$$(Ax)(t) = y(t) = \int_t^1 x(s)ds \quad for\ x \in C^\infty[0,1].$$

Clearly, $\alpha_A = 0$, $\beta_A > 0$ (cf. Example 8.2). Let a functional $\eta \in \mathrm{H}$ be given by a function $\tilde{\eta}$. Then for every $x \in C^\infty[0,1]$ we have

$$(A'\eta)x = \eta(Ax) = \int_0^1 \left[\int_t^1 x(s)ds\right]\tilde{\eta}(t)dt =$$

$$= \left[\int_0^t x(s)ds\Big(\int_0^t \tilde{\eta}(s)ds\Big)\right]_0^1 + \int_0^1 x(t)\left[\int_0^t \tilde{\eta}(s)ds\right]dt =$$

$$= \int_0^1 x(t)\left[\int_0^t \tilde{\eta}(s)ds\right]dt.$$

Then the conjugate operator A' maps the functional η defined by the formula

$$\eta(x) = \int_0^t \tilde{\eta}(t)x(t)dt$$

into a functional ξ given by the formula

$$\xi(x) = \int_0^t \tilde{\xi}(t)x(t)dt, \quad where \quad \tilde{\xi}(t) = \int_0^t \tilde{\eta}(s)ds \quad (x \in X).$$

The operator A' is a one-to-one mapping of the space H into itself. Hence $\beta_{A'} = \alpha_{A'} = 0$, i.e. $\alpha_{A'} < \beta_A$. □

Let $A \in L_0(X \to Y, \mathrm{H} \to \Xi)$. According to the admitted convention, the operator A' naps the space H into the space Ξ. The spaces Y and X may be treated as spaces of functionals over H and Ξ, respectively. The definition of a conjugate operator immediately implies that the operator $A'' = (A')'$ conjugate with $A' \in L_0(\mathrm{H} \to \Xi, X \to Y)$ is equal to the operator A. This, and the change of roles of the operators A and A' in Theorem 9.3 together imply

Theorem 9.6. *If $A \in L_0(X \to Y, \mathrm{H} \to \Xi)$ then $\alpha_A \le \beta_{A'}$.*

In order to obtain a theorem dual to Corollary 9.4, we should say what it means that the operator A' is X-describable. Write

$$\mathcal{A} = \{x \in X : \xi(x) = 0 \ for\ all\ \xi \in \mathcal{E}_{A'}\}.$$

Then

$$\mathcal{E}_{A'} = \{\xi \in \Xi : \xi(x) = 0 \ for\ all\ x \in \mathcal{A}\}.$$

But this fact that $\xi \in \mathcal{E}_{A'}$ implies that $\xi = A'\eta$ for an $\eta \in \mathrm{H}$. Then $\xi(x) = \eta(x)$. Since the space H is total, we conclude that $\mathcal{A} = \{x \in X : Ax = 0\} = \ker A$. This implies that $\mathcal{E}_{A'} = \ker\ A^{\perp}$. WE therefore obtain the following

Theorem 9.7. *If $A \in L_0(X \to Y, \mathrm{H} \to \Xi)$ and $\mathcal{E}_{A'} = \ker\ A^{\perp}$ then $\alpha_A = \beta_{A'}$.*

Theorem 9.8. *If $A \in L_0(X \to Y, Y' \to X')$ then $\alpha_A = \beta_{A'}$.*

Proof. Decompose spaces X and Y onto direct sums: $X = \ker\ A \oplus V_1$, $Y = \mathcal{E}_A \oplus V_2$. The operator A is a one-to-one mapping of the subspace V_1 onto the subspace $\mathcal{E}_A$. This, and Corollary 9.2 together imply that the operator A' maps the space of all linear functionals defined on the subspace $\mathcal{E}_A$ onto the space V_1' of all linear functionals defined on the subspace V_1. Every linear functional defined on the subspace V_1 can be extended to the whole space X in such a manner that $f(x) = 0$ for $x \in \ker\ A$ (cf. Corollary 2.2). Then $\beta_{A'} = \dim X/V_1 = \dim \ker A = \alpha_A$. ■

The above considerations show us that the equalities $\beta_{A'} = \alpha_A$, $\alpha_{A'} = \beta_A$ do not hold always. Hence also the equality $\kappa_{A'} = -\kappa_A$ does not hold always. We therefore shall introduce a new notion.

Let $A \in L_0(X \to Y, \mathrm{H} \to \Xi)$. Write $\beta_A^{\mathrm{H}} = \alpha_{A'}$. An ordered pair $(\alpha_A, \beta_A^{\mathrm{H}})$ is said to be a *d_{H}-characteristic* of the operator A. A d_{H}-characteristic is *finite* (*semi-finite*) if $\alpha_A < +\infty$ and $\beta_A^{\mathrm{H}} < +\infty$ (either $\alpha_A < +\infty$ or $\beta_A^{\mathrm{H}} < +\infty$, respectively). For operators having either finite or semi-finite d_{H}-characteristic we may define H-*index* κ_A^{H} in the following way:

$$\kappa_A^{\mathrm{H}} = \begin{cases} \beta_A^{\mathrm{H}} - \alpha_A & \text{if } \alpha_A < +\infty,\ \beta_A^{\mathrm{h}} = \alpha_{A'} < +\infty, \\ +\infty & \text{if } \alpha_A < +\infty,\ \beta_A^{\mathrm{H}} = +\infty, \\ -\infty & \text{if } \alpha_A = +\infty,\ \beta_A^{\mathrm{H}} < +\infty. \end{cases}$$

Theorem 9.3 implies that $\beta_A^{\mathrm{H}} = \alpha_{A'} \le \beta_A$. Then

$$\kappa_A^{\mathrm{H}} \le \kappa_A \quad for\ A \in L_0(X \to Y, \mathrm{H} \to \Xi). \tag{9.4}$$

It is easy to see that the pair $(\alpha_{A'}, \alpha_A)$ is a d_X-characteristic of the operator A'. Then

$$\kappa_A^{\mathrm{H}} = -\kappa_{A'}^{X} \quad for\ A \in L_0(X \to Y, \mathrm{H} \to \Xi). \tag{9.5}$$

Corollary 9.4 and Theorem 9.8 immediately imply

Corollary 9.5. *If $A \in L_0(X \to Y, Y' \to X')$ then the d_X-characteristic of the operator A is equal to its d-characteristic, hence $\kappa_A^{Y'} = \kappa_A$.*

Example 9.1 showed that a d_{H}-characteristic of an operator A is not always equal to its d-characteristic, even if $X = Y$, $\Xi = \mathrm{H}$. This example also implies that for the H-index of a superposition of operators a formula similar to Formula (2.2) does not hold. Indeed, if the space X and the operator A are defined as in Example 9.1 and $B = \frac{\mathrm{d}}{\mathrm{d}t}$ then $AB = -I$. Hence $\kappa_{BA}^{\mathrm{H}} = \kappa_{BA} = 0$. On the other hand, $\kappa_B^{\mathrm{H}} = \kappa_B =$ - 1 and $\kappa_A^{\mathrm{H}} = 0$. Then

$$\kappa_{BA}^{\mathrm{H}} \neq \kappa_A^{\mathrm{H}} + \kappa_B^{\mathrm{H}} = -1.$$

Suppose that the operator $A \in L_0(X \to Y, Y' \to X')$ has a finite d-characteristic (nullity, deficiency) and the d_{H}-characteristic of the operator A is equal to its d-characteristic. Then A is said to be a Φ_{H}-*operator* (Φ_{H}^+-*operator*, Φ_{H}^--*operator*, respectively).

Theorem 9.9. *If an operator $K \in L_0(X, \Xi)$ is finite dimensional then the operator $I + K$ is a Φ_Ξ-operator.*

Proof. Let $Kx = \sum\limits_{j=1}^{N} f_j(x)x_j$ for $x \in X$, where $f_1, ..., f_n \in X'$, $x_1, ..., x_n \in X$ are linearly independent. We have shown (in the proof of Theorem 8.4) that $\alpha_{I+K} = \beta_{I+K} = n - k$, where k is the rank of the matrix $[\xi_i(x_j) + \delta_{ij}]_{i,j=1,...,n}$. Consider the conjugate operator $I + K'$, where $\xi K' = \sum\limits_{j=1}^{n} \xi(x_j)f_j$. Clearly, $\alpha_{I+K'} = \beta_{I+K'} = n - k'$, where k' is the rank of the matrix

$$[\xi_j(x_i) + \delta_{ij}]_{i,j=1,...,n} = [\xi_i(x_j) + \delta_{ij}]_{i,j=1,...,n}^T.$$

Then $k' = k$. Hence $\beta_{I+K}^{\mathrm{H}} = \alpha_{I+K'} = n - k' = n - k = \alpha_{I+K} = \beta_{I+K}$. ■

Let X be a linear space and let $X_0 \subset X$ be its subspace. Let Ξ_0 be a family (non-necessarily linear) of linear functionals defined on the space X. The subspace X_0 is said to be *described* by the family Ξ_0 when $\xi_0(x) = 0$ for all $\xi_0 \in \Xi_0$ if and only if $x \in X_0$, i.e. when

$$X_0 = \{x \in X : \xi_0(x) = 0 \quad \mathit{for\ all}\ \xi_0 \in \Xi_0\}.$$

In other words: a subspace $X_0 \subset X$ is described by a family Ξ_0 if and only if X_0 is Ξ_0-describable and $X_0^\perp = \mathrm{lin}\ \Xi_0$.

An operator $A \in L_0(X \to Y, \mathrm{H} \to \Xi)$ is H-solvable if and only if its range $\mathcal{E}_A$ can be described by a family $\mathrm{H}_0 \subset \mathrm{H}$. It is easy to verify that an operator $A \in L_0(X \to Y, \mathrm{H} \to \Xi)$ with a finite d-characteristic is a Φ_{H}-operator if and only if its range $\mathcal{E}_A$ can be described by a finite system of linear functionals. If a subspace $X_0 \subset X$ can be described by a finite system of linear functionals $\Xi_0 \subset \Xi$ then every subspace $X_1 \subset X$ containing X_0, i.e. such that $X_0 \subset X_1 \subset X$, can be described by a finite system $\Xi_1 \subset \Xi_0 \subset \Xi$ of linear functionals.

An operator $T \in L_0(X \to Y, \mathrm{H} \to \Xi)$ satisfies the *Fredholm alternative* if three following conditions hold:

(a) The homogeneous equation $(I+T)x = 0$ has the finite number k of linearly independent solutions.

(b) The conjugate homogeneous equation $(I' + T')\xi = 0$ has the finite number k' of linearly independent solutions and $k' = k$.

(c) The equation

$$(I+T)x = y, \qquad y \in Y,$$

has a solution if and only if $\xi(y) = 0$ for every solution ξ of the conjugate homogeneous equation $(I' + T')\xi = 0$.

This alternative, i.e. Conditions (a), (b), (c), have been proved firstly for some integral equations by Swedish mathematician Ivar Fredholm in the years 1901-1904.

The Fredholm alternative here was presented in a traditional way. Now, using notions introduced in Chapter 8 and in this chapter, we can formulate it in another way. Namely,

An operator $T \in L_0(X \to Y, \mathrm{H} \to \Xi)$ satisfies the *Fredholm alternative* if three following conditions hold:

(a') $\alpha_{I+T} < +\infty$;

(b') $\alpha_{I'+T'} = \alpha_{I+T}$;

(c') the space $\mathcal{E}_{I+T} = (I+T)X$ is described by the family $\mathcal{Z} = \{\xi_1, ..., \xi_n\}$ of linear functionals such that $\ker\,(I' + T') = \mathrm{lin}\,\mathcal{Z}$ and $n = \alpha_{I+T} = \alpha_{I'+T'}$.

The above conditions immediately imply

Theorem 9.10. *An operator $T \in L_0(X \to Y, \mathrm{H} \to \Xi)$ satisfies the Fredholm alternative if and only if the operator $I+T$ is a Φ_{H}-operator and $\kappa^{\mathrm{H}}_{I+T} = 0$.*

An operator $T \in L_0(X \to Y, \mathrm{H} \to \Xi)$ satisfies the *generalized Fredholm alternative* (otherwise we say: T is a *Noether operator*) if

(i) $\alpha_{I+T} < +\infty$;

(ii) $\alpha_{I'+T'} < +\infty$;

(iii) the space $\mathcal{E}_{I+T} = (I+T)X$ is described by the family $\mathcal{Z} = \{\xi_1, ..., \xi_n\}$ of linear functionals such that $\ker\,(I' + T') = \mathrm{lin}\,\mathcal{Z}$ and $n' = \alpha_{I'+T'}$.

In other words, *in the generalized Fredholm alternative we do not assume that the nullities of the operator $I+T$ and its conjugate $I' + T'$ are equal.*

Conditions (i), (ii), (iii) of the generalized Fredholm alternative together imply

Theorem 9.11. *An operator $T \in L_0(X \to Y, \mathrm{H} \to \Xi)$ satisfies the generalized Fredholm alternative if and only if the operator $I+T$ is a Φ_{H}-operator.*

Note that, by definitions, an operator $T \in L_0(X \to Y, \mathrm{H} \to \Xi)$ is a Noether operator if and only if $I + T$ is a Φ_{H}-operator.

Theorems 9.9 and 9.10 together imply that finite dimensional operators satisfy the Fredholm alternative. The operator $D - I$, where the operator D is defined in Example 8.2, satisfies the generalized Fredholm alternative.

Corollary 9.5 implies

Corollary 9.6. *If an operator $A \in L_0(X \to Y, Y' \to X')$ has a finite d-characteristic then the operator $T = A - I$ satisfies the generalized Fredholm alternative,*

Theorem 9.12. *If $A \in D(X \to Y) \cap L_0(X \to Y)$ then for every conjugate space $\Xi \subset X'$ there is a Φ_Ξ-operator $B \in L_0(Y \to X)$ such that the operators $AB - I$ and $BA - I$ are finite dimensional in the spaces Y and X, respectively.*

Proof. By our assumption, $\alpha_A = \dim \ker A < +\infty$. Let $\alpha_A = n$. Consider a system $\{f_1, .., f_n\} \subset \Xi$ of linear functionals whose restrictions to the subspace $\ker A$ are linearly independent. Let $V = \{x \in X : f_j(x) = 0 \ for \ j = 1, ..., n\}$. Decompose the space X onto the direct sum: $X = \ker A \oplus V$. The restriction of the operator A to the subspace V is invertible and maps V into the set $\mathcal{E}_A$. Let A_1^{-1} be its inverse defined on the subspace $\mathcal{E}_A$. Decompose the space Y onto the direct sum: $Y = \mathcal{E}_A \oplus V_1$ and define an operator B in the following way:

$$By = \begin{cases} A_1^{-1} & \text{for } y \in \mathcal{E}_A, \\ 0 & \text{for } y \in V_1. \end{cases}$$

Since the set $\mathcal{E}_B = V$ can be described by a finite system of linear functionals, we conclude that B is a Φ_Ξ-operator.

The operator $P_{\mathcal{E}_A} = AB$ is a projector onto the set $\mathcal{E}_{\mathcal{B}}$ while the operator $P_V = BA$ is a projector onto the subspace V. Since α_A and β_A are finite, these operators differ from the identity operator only by finite dimensional operators. The subspace V can be described by a finite system of linear functionals belonging to the space Ξ. Hence $P_V = I - K$, where the operator $K \in L_0(X, \Xi)$ is finite dimensional. Similarly, $P_{\mathcal{E}_A} = I - K_1$ where the operator K_1 is finite dimensional. In particular, if A is a Φ_{H}-operator then $K_1 \in L_0(Y, \mathrm{H})$. ■

Theorem 9.13. *If a Φ_{H}-operator $B \in L_0(X \to Y)$, a Φ_Σ-operator $A \in L_0(Y \to Z)$ and $A'\mathrm{H} \subset \Sigma$ then the superposition is a Φ_Σ-operator and*

$$\kappa_{AB}^{\Sigma} = \kappa_A^{\Sigma} + \kappa_B^{\mathrm{H}}. \tag{9.6}$$

Proof. Similarly, as in the proof of Theorem 8.1, decompose the space Y onto the direct sum (8.4): $Y = \mathcal{E}_B \oplus V_2 \oplus V_3$. Since B is a Φ_{H}-operator, the space $\mathcal{E}_B$ can be described by a finite system of linear functionals of dimension equal to the dimension of the space $V_2 \oplus V_3$. Then every space $Y_0 \supset \mathcal{E}_B$ can be described by a finite system of linear functionals. In particular, the space $\mathcal{E}_B \oplus V_2$ can be described by a finite system $f_1, ..., f_{n_3}$ of linear functionals, where $n_3 = \dim V_3$. The set $\mathcal{E}_A$ can be described by a finite system $g_1, ..., g_{\beta_A}$ of linear functionals, because A is a Φ_Σ-operator. Since $A'\mathrm{H} \subset \Sigma$, Formula (8.4′) implies that $\mathcal{E}_A = \mathcal{E}_{AB} \oplus AV_3$.

Then the set $\mathcal{E}_{AB}$ can be described by a finite system $g_1, ..., g_{\beta_A}, f_1A, ..., f_{n_3}A$ of linear functionals. This, and Theorem 8.1 together imply that

$$\kappa^{\Sigma}_{AB} = \kappa_{AB} = \kappa_A + \kappa_B = \kappa^{\Sigma}_A + \kappa^{\mathrm{H}}_B.$$

■

Theorem 9.14. *Suppose that $B \in L(X \to Y)$ and $A \in L_0(Y \to Z, \mathrm{H} \to \Sigma)$. If $\beta_{AB} < +\infty$ and $\beta^{\Sigma}_{AB} = \beta_{AB}$ then $\beta_A < +\infty$ and $\beta^{\mathrm{H}}_A = \beta_A$.*

Proof. Theorem 8.2 implies that $\beta_A < +\infty$ for $\beta_{AB} < +\infty$. By the assumption that $\beta^{\Sigma}_{AB} = \beta_{AB}$, it follows that the set $\mathcal{E}_{AB}$ can be described by a finite system of linear functionals. Then every space containing $\mathcal{E}_{AB}$, in particular the space $\mathcal{E}_A = \mathcal{E}_{AB} \oplus AV_3$ (cf. Formula (8.4′)), can be described by a finite system of linear functionals. Then $\beta^{\mathrm{H}}_A = \beta_A$. ■

Theorem 9.14 and Corollary 8.1 immediately imply

Corollary 9.7. *If $A \in L_0(X \to Y)$, $B \in L_0(Y \to X)$, AB is a Φ_{Ξ}-operator and B is a Φ_{H}-operator then A is a Φ_{Ξ}-operator.*

Let $P(X, \Xi) \subset L_0(X)$ be an arbitrary algebra of linear operators and let $K_P(X, \Xi)$ be the set of all finite dimensional operators contained in $P(X, \Xi)$. A proper ideal $\mathcal{J} \subset P(X, \Xi)$ is said to be a *Ξ-quasi-Fredholm ideal* if the $I + T$ is a Φ_{Ξ}-operator whenever $T \in \mathcal{J}$.

Theorem 9.9 implies

Corollary 9.8. *If $P(X, \Xi) \subset L_0(X)$ is an arbitrary algebra of linear operators then the set $K_P(X, \Xi)$ of all finite dimensional operators contained in $P(X, \Xi)$ is a Ξ-quasi-Fredholm ideal in $P(X, \Xi)$.*

Theorem 9.15. *Suppose that $A \in L_0(X \to Y, \mathrm{H} \to \Xi)$ and that there is an operator $R_A \in L_0(X \to Y, \mathrm{H} \to \Xi)$ such that*

$$AR_A - I \in \mathcal{J}_{\mathrm{H}}, \qquad R_AA - I \in \mathcal{J}_{\Xi},$$

where $\mathcal{J}_{\mathrm{H}}$ is a H-quasi-Fredholm ideal and $\mathcal{J}_{\Xi}$ is a Ξ-quasi-Fredholm ideal. Then A is a Φ_{Ξ}-operator.

Proof. By the assumptions, AR_A is a Φ_{H}-operator and R_AA is a Φ_{Ξ}-operator. This, and Corollary 9.7 together imply that A is a $\Phi - \Xi$-operator and R_A is a Φ_{H}-operator. ■

Theorem 9.15 implies

Corollary 9.9. *If an operator A belonging to an algebra $P(X, \Xi) \subset L_0(X, \Xi)$ of linear operators has a simple regularizer to a Ξ-quasi-Fredholm ideal $\mathcal{J} \subset P(X, \Xi)$ then A is a Φ_{Ξ}-operator.*

Theorem 9.16. *Every quasi-Fredholm ideal $\mathcal{J}$ contained in a regularizable algebra $P(X, \Xi) \subset L_0(X, \Xi)$ of linear operators is a Ξ-quasi-Fredholm ideal.*

Proof. Suppose that $T \in \mathcal{J}$. Then the operator $I + T$ has a finite d-characteristic. Since the algebra $P(X, \Xi)$ is regularizable, the operator $I+T$ has a simple regularizer to the ideal $K_P(X, \Xi)$ of all finite dimensional operators contained in $P(X, \Xi)$ which, by Corollary 9.8, is a Ξ-quasi-Fredholm ideal. Corollary 9.9 implies that $I + T$ is a Φ_Ξ-operator. This, and the arbitrariness of the operator $T \in \mathcal{J}$ implies that $\mathcal{J}$ is a Ξ-quasi-Fredholm ideal. ■

Theorem 9.17. *Suppose that an algebra $\mathcal{X}(X)$ of linear operators contains the ideal $K(X, \Xi)$ of all finite dimensional operators belonging to $L_0(X, \Xi)$ and that $I + K$ is a Φ_Ξ-operator for every $K \in K(X, \Xi)$. Then $\mathcal{X}(X) \subset L_0(X, \Xi)$.*

Proof. Suppose that an operator $T \in \mathcal{X}(X)$ does not preserve the conjugate space Ξ, i.e. there is a linear functional $\xi \in \Xi$ such that $\eta = T'\xi \notin \Xi$. Let an operator $P \in K(X, \Xi)$ be of the form $Px = \xi(x)x_0$ for $x \in X$. Then $PTx = \xi(Tx)x_0 = \eta(x)x_0$ for $x \in X$. Choose an x_0 such that $\eta(x_0) \neq 0$. Let

$$Kx = \frac{\eta(x)}{\eta(x_0)} x_0 \qquad for \quad x \in X.$$

By this definition, K is a finite dimensional operator belonging to $\mathcal{X}(X)$. But $\mathcal{E}_{I-K} = \{x \in X : \eta(x) = 0\}$. Indeed,

$$\eta(x - Kx) = \eta(x) - \frac{\eta(x)}{\eta(x_0)} \eta(x_0) \qquad for \quad x \in X.$$

However, by definition, $\eta \notin \Xi$. This implies that the operator $I - K$ is not a Φ_Ξ-operator, a contradiction with our assumption. Hence every operator $T \in \mathcal{X}(X)$ preserves the conjugate space. ■

Suppose that we are given an algebra $\mathcal{X}(X) \subset L_0(X, \Xi)$ of linear operators. Denote by $P'(X, \Xi)$ the set of all operators conjugate to operators belonging to $P(X, \Xi)$. It is easy to verify that this set is also an algebra. Let $\mathcal{J}$ be an ideal in the algebra $P(X, \Xi)$. Then the set $\mathcal{J}'$ of all operators conjugate to operators belonging to $\mathcal{J}$ is an ideal in the algebra $P'(X.\Xi)$. Indeed, if $A \in P(X, \Xi)$, $T_1, T_2 \in \mathcal{J}$ and a_1, a_2 are scalars then $A' \in P(X, \Xi)$, $T_1', T_2' \in \mathcal{J}'$ and

$$a_1 T_1' + a_2 T_2' = (a_1 T_1 + a_2 T_2)' \in \mathcal{J}',$$

$$A' T_1' = (T_1 A)' \in \mathcal{J}', \qquad T_2' A' = (A T_2) \in \mathcal{J}'.$$

Hence, if the operator A has a left regularizer to the ideal $\mathcal{J}$, i.e. $R_A A = I + T$, where $T \in \mathcal{J}$, then

$$A' R_A' = (R_A A)' = (I + T)' = I' + T', \quad where \quad T' \in \mathcal{J}',$$

i.e. R_A' is a right regularizer of the operator A' to the ideal $\mathcal{J}'$. Similarly, If R_A is a right regularizer of the operator A to the ideal $\mathcal{J}$ then R_A' is a left regularizer of the operator A' to the ideal $\mathcal{J}'$. In particular, if the operator A has a simple regularizer to the ideal $\mathcal{J}$ then the operator A' has a simple regularizer to the ideal $\mathcal{J}'$ and $R_{A'} = R_A'$.

Theorem 9.18. *If an operator $A \in D(X \to Y)$ is a Φ_{H}-operator then every operator $K \in K(X \to Y, \mathrm{H} \to \Xi)$ is a Φ_{H}-perturbation of the operator A.*

Proof. Since $K \in K(X \to Y, \mathrm{H} \to \Xi)$, we have

$$Kx = \sum_{j=1}^{n} g_j(x)y_j, \quad for\ x \in X, \quad where\ g_j \in \Xi,\ y_j \in Y \quad (j = 1, ..., n).$$

Theorem 8.6 implies that $A + K \in D(X \to Y)$. Let $Z = \{x : g_j(Ax) = 0\ (j = 1, ..., n)\}$. Clearly, $(A + K)x = Ax$ for $x \in Z$. But

$$AZ = \mathcal{E}_A \cap \{y : g_j(y) = 0\ (j = 1, ..., n)\}.$$

By our assumption, A is a Φ_{H}-operator. Then the set $\mathcal{E}_A$ can be described by a finite system of linear functionals. Hence the set $AZ \supset \mathcal{E}_A$ can be described by a finite set of linear functionals. Furthermore, since $\mathcal{E}_{A+K} \supset AZ$, we conclude that the set $\mathcal{E}_{A+K}$ can be also described by a finite set of linear functionals. This proves that $A+K$ is a Φ_{H}-operator. Then K is a Φ_{H}-perturbation of the operator A. ■

Theorem 9.19. *If an algebra $P(X, \Xi) \subset L_0(X, \Xi)$ of linear operators is regularizable to a Ξ-quasi-Fredholm ideal $\mathcal{J} \subset P(X, \Xi)$ then all operators belonging to the ideal $\mathcal{J}$ are perturbations of the class of all Φ_{Ξ}-operators belonging to the algebra $P(X, \Xi)$.*

Proof. If $A \in P(X, \Xi)$ is an arbitrary Φ_{Ξ}-operator then it has a simple regularizer R_A to the ideal $\mathcal{J}$, i.e. we have

$$R_A A = I + T_1, \quad AR_A = I + T_2, \quad where\ T_1, T_2 \in \mathcal{J}.$$

Suppose that $T \in \mathcal{J}$. Then

$$R_A(A + T) = I + R_A T + T_1 = I + T_3, \quad where\ T_3 \in \mathcal{J},$$

$$(A + T)R_A = I + TR_A + T_2 = I + T_4, \quad where\ T_4 \in \mathcal{J}.$$

Since $\mathcal{J}$ is a Ξ-quasi-Fredholm ideal, the operators $R_A(A + T)$ and $(A + T)R_A$ are Φ_{Ξ}-operators. By Theorem 9.18, $A + T$ is a Φ_{Ξ}-operator. The arbitrariness of the operator $A \in P(X, \Xi)$ implies that every operator $T \in \mathcal{J}$ is a perturbation of the class of all Φ_{Ξ}-operators belonging to the algebra $P(X, \Xi)$. ■

Theorem 9.20. *Suppose that all assumptions of Theorem 9.19 are satisfied. Then every operator belonging to a Ξ-quasi=Fredholm ideal $\mathcal{J}_1 \subset P(X, \Xi)$ is a perturbation of the class of all Φ_{Ξ}-operators belonging to the algebra $P(X, \Xi)$.*

Proof. Let $\tilde{\mathcal{J}} = \mathcal{J} + \mathcal{J}_1$. Similarly, as in the proof of Theorem 8.14, we show that $\tilde{\mathcal{J}}$ is a Ξ-quasi-Fredholm ideal. Theorem 9.19 implies that all operators belonging to $\tilde{\mathcal{J}}$, in particular, operators belonging to $\mathcal{J}_1$, are perturbations of the class of all Φ_{Ξ}-operators belonging to the algebra $P(X, \Xi)$. ■

A Ξ-quasi-Fredholm ideal $\mathcal{J}$ is said to be a Ξ-*Fredholm ideal* if $\kappa^{\Xi}_{I+T} = 0$ whenever $T \in \mathcal{J}$. Since for Φ_{Ξ}-operators their Ξ-index is equal to their index, Theorems 8.14 and 9.20 together immediately imply

Corollary 9.10. *If $\mathcal{J}$ and $\mathcal{J}_1$ are Ξ-Fredholm ideals in an algebra $P(X,\Xi) \subset L_0(X,\Xi)$ of linear operators regularizable to the ideal $\mathcal{J}$ then all operators belonging to the ideal $\tilde{\mathcal{J}} = \mathcal{J} + \mathcal{J}_1$ are perturbations of the class of all Φ_{Ξ}-operators belonging to the algebra $P(X,\Xi)$ and these perturbations preserve the Ξ-index, i.e.*

$$\kappa^{\Xi}_{A+T} = \kappa^{\Xi}_{A}$$

for every Φ_{Ξ}-operator $A \in P(X,\Xi)$ and for every $T \in \mathcal{J}_1$.

It has been shown in Example 9.1 that d_{Ξ}-characteristic may change together with a change of a conjugate space Ξ. However, two following theorems show that in some cases a change of spaces X and Ξ does not imply a change of the corresponding d_{Ξ}-characteristic. This fact plays an important role, in particular, when solving integral equations.

Theorem 9.21 (First reduction theorem). *Suppose that $\Xi \subset X'$ is a conjugate space and that an operator $A = I + T$, where $T \in L_0(X)$, has a finite d_{Ξ}-characteristic. Let X_0 be an arbitrary subspace of the space X containing the set TX and let Ξ_0 be an arbitrary subspace of the space Ξ containing the set $\Xi T = T'\Xi$:*

$$TX \subset X_0 \subset X, \qquad \Xi T \subset \Xi_0 \subset \Xi.$$

Then the operator A restricted to the subspace X_0 has a finite d_{Ξ_0}-characteristic which is equal to a d_{Ξ}-characteristic of the operator A on the whole space X.

Proof. It is an immediate consequence of the fact that all solutions of the equation $(I+T)x = 0$ in the space X belong to the subspace X_0, i.e. $\ker\,(I+T) \subset X_0$. Similarly, all solutions of the conjugate equation $\xi(I+T) = (I'+T')\xi = 0$ in the space Ξ belong to the space Ξ_0, i.e. $\ker\,(I'+T') \subset \Xi_0$. ∎

Theorem 9.22 (Second reduction theorem). *Let X_0 be a subspace of a linear space X and let Ξ_0 be a subspace of a conjugate space $\Xi \subset X'$. Suppose that an operator $A \in L_0(X_0,\Xi_0)$ has a simple regularizer R_A such that*

$$AR_A = I + T_1, \qquad R_A A = I + T_2,$$

where the operators T_1, T_2 can be extended to operators $\tilde{T}_1$, $\tilde{T}_2 \in L_0(X,\Xi)$ such that both, $I+\tilde{T}_1$, $I+\tilde{T}_2$ are Φ_{Ξ}-operators. Then the operator A is a Φ_{Ξ_0}-operator.

Proof. The operators $I+\tilde{T}_1$, $I+\tilde{T}_2$ are defined on the whole space X and, by our assumption, they are Φ_{Ξ}-operators. Theorem 9,21 implies that the operators $I+T_1$, $I+T_2$ are Φ_{Ξ_0}-operators. Then AR_A and R_AA are also Φ_{Ξ_0}-operators. This, and Corollary 9.7 together imply that A is a Φ_{Ξ_0}-operator. ∎

Exercises.

Exercise 9.1. An operator B is said to be *generalized inverse* for an operator A (otherwise: *almost inverse of* A) if

$$ABA = A \quad \textit{and} \quad BAB = B. \tag{9.7}$$

Prove that

(i) If the operator A has a generalized inverse B then B has a generalized inverse A.

(ii) The operators A and B determined in Theorem 9.12 are generalized inverses each to another.

(iii) If $A \in D(X \to Y) \cap L_0(X \to Y)$ then to every conjugate space $\Xi \subset X'$ there is a Φ_Ξ-operator $B \in D(X \to Y) \cap L_0(X \to Y)$ which is a generalized inverse of A. □

Exercise 9.2. Suppose that $T \in L_0(X)$ and there is a positive integer m such that $I - T^m$ is a Φ_Ξ-operator for a conjugate space $\Xi \subset X'$. Prove that $I - T$ is a Φ_Ξ-operator. □

Exercise 9.3. Are theorems similar to Theorems 9.13, 9.19, 9.20 and Corollary 9.10 true for Φ_Ξ^+-operators and Φ_Ξ^--operators? □

REFERENCES

R. Bellman

B[1] *Introduction to Matrix Analysis.* McGraw-Hill, New York 1960.

K. Borsuk

B[2] *Geometria Analityczna w n Wymiarach.* (In Polish; Analytic Geometry in n-dimensions), Monografie Matematyczne 12, PWN-Polish Scientific Publishers, Warszawa, 1950; English ed.: *Multidemensional Analytic Geometry.* Monografie Matematyczne 50, PWN-Polish Scientific Publishers, Warszawa, 1969.*

K. Borsuk and W. Szmielew

B[3] *Foundations of Geometry.* North Holland, Amsterdam, 1960.

M. Gel'fand

G[1] *Lectures on Linear Algebra.* Dover Publications, New York, 1989.

Ch. Hermite

H[1] Sur la formule d'interpolation de Lagrange. Journ. Reine Angew. Math., 84(1878), 70-78.

N. Jacobson

J[1] *Structure of rings.* Amer. Math. Soc., Providence, R.I., 1956.

I. Kaplansky

K[1] Rings with a polynomial identity. Bull. Amer. Math. Soc., 54(1948), 575-580.

S. Lang

L[1] *Linear Algebra.* Addison-Wesley, Reading, Ma, 1966; 3rd ed. Springer-Verlag, New York, 1987.

P. D. Lax

L[2] *Linear Algebra.* John Wiley & Sons, New York, 1997.

Nguyen Van Mau

N[1] *Algebraic Elements and Relative Boundary Value Problems in Linear Spaces.* Vietnam National University Publishers, Hanoi, 2005.

V. V. Prostakov

P[1] *Problems and Theorems in Linear Algebra.* Translations of Math. Monographs 134. Amer. Math. Soc., Providence, R.I., 1994.

D. Przeworska-Rolewicz

PR[1] Équations avec opérations algébriques. Studia Math., 22(1963), 337-367.*

PR[2] *Przestrzenie Liniowe i Operatory Liniowe.* (In Polish; Linear Spaces and Linear Operators), Wydawnictwa Naukowo-Techniczne, Warszawa, 1977.

* This publication is available in The Virtual Mathematical Library (Wirtualna Biblioteka Naukowa), cf. http: // matwbn.icm.edu.pl

PR[3] *Algebraic Analysis.* PWN-Polish Scientific Publishers and D. Reidel, Warszawa-Dordrecht, 1988.

PR[4] *Logarithms and Antilogarithms. An Algebraic Analysis Approach.* With Appendix by Z. Binderman. Kluwer Academic Publishers, Dordrecht, 1998.

PR[5] Algebraic analysis in structures with the Kaplansky-Jacobson property. Studia Math., 168(2)(2005), 165-186.

D. Przeworska-Rolewicz and S. Rolewicz.
PRR[1] *Equations in linear spaces.* Monografie Matematyczne 47, PWN-Polish Scientific Publishers, Warszawa, 1968. *

M. Tasche
T[1] Abstrakte Differentialgleichungen mit algebraische Operatoren. Wiss. Zeitschrift der Universität Rostock, 24(1975), 1231-1236.

* This publication is available in The Virtual Mathematical Library (Wirtualna Biblioteka Naukowa), cf. http: // matwbn.icm.edu.pl

INDEX

J

K

L