CHAPTER XVI

LINEAR ALGEBRA

§1. The Scope of Linear Algebra and Its Apparatus

Linear functions and matrices. Among the functions of a single variable, by far the simplest is the so-called *linear function* $l(x) = ax + b$. Its graph is, of course, the simplest of curves, namely the straight line.

All the same, the linear function is one of the most important. This is due to the fact that every "smooth" curve on a small segment is like a straight line, and the less curved the segment is, the nearer it comes to a straight line. In the language of the theory of the functions, this means that every "smooth" (continuously differentiable) function is, for a small change of the independent variable, close to a linear function. The linear function can be characterized by the fact that its increment is proportional to the increment of the independent variable. Indeed: $\Delta l(x) = l(x_0 + \Delta x) - l(x_0) = a(x_0 + \Delta x) + b - (ax_0 + b) = a\,\Delta x$. Conversely, if $\Delta l(x) = a\,\Delta x$, then $l(x) - l(x_0) = a(x - x_0)$ and $l(x) = ax + l(x_0) - ax_0 = ax + b$, where $b = l(x_0) - ax_0$. But from the differential calculus, we know that in the increment of an arbitrary differentiable function we can single out in a natural way the principal part, the so-called differential of the function, which is proportional to the increment of the independent variable, and that the increment of the function differs from its differential by an infinitesimal of higher order than the increment of the independent variable. Thus, a differentiable function is, for an infinitely small change of the independent variable, really close to a linear function to within an infinitesimal of higher order.

The situation is similar with functions of several variables.

A *linear function of several variables* is a function of the form $a_1x_1 + a_2x_2 + \cdots + a_nx_n + b$. If $b = 0$, the linear function is said to be

homogeneous. A linear function of several variables is characterized by the following two properties:

1. The increment of a linear function, computed under the assumption that only one of the independent variables receives some increment while the values of the remaining variables are unchanged, is proportional to the increment of this independent variable.

2. The increment of a linear function, computed under the assumption that all the independent variables obtain increments, is equal to the algebraic sum of the increments obtained by changing each variable separately.

The linear function of several variables plays the same role among all the functions of these variables as the linear function of one variable among all the functions of one variable. For every "smooth" function (i.e., a function having continuous partial derivatives with respect to all variables) is close to some linear function for small changes of the independent variables. In fact, the increment of such a function $w = f(x_1, x_2, \cdots, x_n)$ is equal, to within infinitesimals of higher order, to the total differential $(\partial f/\partial x_1)\,dx_1 + \cdots + (\partial f/\partial x_n)\,dx_n$, which is a linear homogeneous function of the increments $dx_1, \cdots, dx_n$ of the independent variables. Hence it follows that the function w itself, which is equal to the sum of its initial value and its increment, can be expressed in terms of its independent variables for small changes of these in the form of a linear inhomogeneous function to within infinitesimals of higher orders.

Problems whose solution requires the investigation of functions of several variables arise in connection with the study of the dependence of one quantity on several factors. A problem is called *linear* if the dependence under consideration turns out to be linear. By the properties of linear functions that we have indicated earlier, a linear problem can be characterized by the following properties.

1. *The property of proportionality*. The result of the action of each separate factor is proportional to its value.

2. *The property of independence*. The total result of an action is equal to the sum of the results of the actions of the separate factors.

The fact that every "smooth" function can be replaced in a first approximation by a linear one, for small changes of the variables, is a reflection of a general principle, namely that every problem on the change of some quantity under the action of several factors can be regarded in a first approximation, for small actions, as a linear problem, i.e., as having the properties of independence and proportionality. It often turns out that this attitude gives an adequate result for practical purposes (the classical theory of elasticity, the theory of small oscillations, etc.)

The physical quantities to be studied are often characterized by certain

numbers (a force by the three projections on the coordinate axes, the tension at a given point of an elastic body by the six components of the so-called stress tensor, etc.). Hence there arises the necessity of considering simultaneously several functions of several variables, and, in a first approximation, of several linear functions.

A linear function of one variable is so simple in its properties that it does not require any special study. Things are different with linear functions of several variables, where the presence of many variables introduces some special features. The situation is still more complicated when we go from a single function of several variables $x_1, x_2, \cdots, x_n$ to a set of several functions $y_1, y_2, \cdots, y_m$ of the same variables. As a "first approximation" there appears here a set of linear functions:

$$\begin{aligned} y_1 &= a_{11}x_1 + \cdots + a_{1n}x_n + b_1, \\ y_2 &= a_{21}x_1 + \cdots + a_{2n}x_n + b_2, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_m &= a_{m1}x_1 + \cdots + a_{mn}x_n + b_m. \end{aligned}$$

A set of linear functions is already a rather complicated mathematical object, and the study of its full of interesting and nontrivial results.

The study of linear functions and their systems also constitutes the initial object of that branch of algebra that is called linear algebra.

Historically, the first task of linear algebra is that of solving a system of linear equations:

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2, \\ &\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

The simplest case of this problem is treated in a school course on elementary algebra. The problem of finding methods for the simplest possible and least laborious numerical solution of systems for large n still attracts the close attention of many researchers, because the numerical solution of such systems enters as an important constituent part into many calculations and investigations.

Linear homogeneous functions are also known as *linear forms*. A given system of linear forms

$$\begin{aligned} y_1 &= a_{11}x_1 + \cdots + a_{1n}x_n, \\ &\cdots\cdots\cdots\cdots\cdots \\ y_m &= a_{m1}x_1 + \cdots + a_{mn}x_n \end{aligned}$$

is completely described by its system of coefficients, since the properties

of such a system of forms depend only on the numerical values of the coefficients and the names of the variables are inessential.

For example, the system of forms

$$\begin{matrix} 3x_1 + x_2 - x_3, \\ 2x_1 + x_2 + 3x_3, \\ x_1 - x_2 - x_3 \end{matrix} \quad \text{and} \quad \begin{matrix} 3t_1 + t_2 - t_3, \\ 2t_1 + t_2 + 3t_3, \\ t_1 - t_2 - t_3 \end{matrix}$$

obviously have identical properties and need not be regarded as essentially distinct.

The set of coefficients of a system of linear forms can be given in a natural way in the form of a rectangular array

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}.$$

Such arrays bear the name of *matrices*. The numbers a_{ij} are called the elements of the matrix. The need of considering matrices arises necessarily from the very scope of linear algebra.

Important special cases of matrices are the matrices that consist of a single column, which are simply called columns, those that consist of a single row, called rows, and finally the square matrices, i.e., those in which the number of rows is equal to the number of columns. The number of rows (or columns) of a square matrix is called its *order*. The "matrix" (a) consisting of a single number is identified with that number.

In connection with the simplest operations on a set of linear forms, it is natural to define operations on matrices.

Suppose that two systems of linear forms are given,

$$\begin{matrix} y_1 = a_{11}x_1 + \cdots + a_{1n}x_n, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_m = a_{m1}x_1 + \cdots + a_{mn}x_n, \end{matrix}$$

and

$$\begin{matrix} z_1 = b_{11}x_1 + \cdots + b_{1n}x_n, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ z_m = b_{m1}x_1 + \cdots + b_{mn}x_n. \end{matrix}$$

Let us add these forms,

$$\begin{matrix} y_1 + z_1 = (a_{11} + b_{11})\,x_1 + \cdots + (a_{1n} + b_{1n})\,x_n, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_m + z_m = (a_{m1} + b_{m1})\,x_1 + \cdots + (a_{mn} + b_{mn})\,x_n. \end{matrix}$$

It is natural to say that the matrix of the system of forms so obtained

$$\begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{bmatrix}$$

is the sum of the matrices

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \cdots & \cdots & \cdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix}.$$

Similarly, the product of the matrix

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

by the number c is defined as the matrix of the coefficients in the system of forms $cy_1, cy_2, \cdots, cy_m$, where $y_1, y_2, \cdots, y_m$ are the forms whose coefficients constitute the matrix

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}.$$

From this definition it is clear that

$$c \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} ca_{11} & \cdots & ca_{1n} \\ \cdots & \cdots & \cdots \\ ca_{m1} & \cdots & ca_{mn} \end{bmatrix}.$$

Finally, the operation of multiplication of a matrix by a matrix is defined as follows. Suppose that

$$\begin{matrix} z_1 = a_{11}y_1 + \cdots + a_{1m}y_m, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ z_k = a_{k1}y_1 + \cdots + a_{km}y_m \end{matrix} \tag{1}$$

and

$$\begin{matrix} y_1 = b_{11}x_1 + \cdots + b_{1n}x_n, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_m = b_{m1}x_1 + \cdots + b_{mn}x_n. \end{matrix}$$

When we substitute in (1) the expressions of $y_1, y_2, \cdots, y_m$ in terms of

$x_1, x_2, \cdots, x_n$, we find that $z_1, z_2, \cdots, z_k$ can also be expressed in terms of $x_1, x_2, \cdots, x_n$ by linear forms

$$z_1 = c_{11}x_1 + \cdots + c_{1n}x_n,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$z_k = c_{k1}x_1 + \cdots + c_{kn}x_n.$$

The matrix of coefficients

$$\begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \cdots & \cdots & \cdots \\ c_{k1} & \cdots & c_{kn} \end{bmatrix}$$

is called the product of the matrices

$$\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \cdots & \cdots & \cdots \\ a_{k1} & \cdots & a_{km} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \cdots & \cdots & \cdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix}$$

and is denoted by

$$\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \cdots & \cdots & \cdots \\ a_{k1} & \cdots & a_{km} \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \cdots & \cdots & \cdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix}.$$

It is easy to calculate how the elements of the product of two matrices are expressed in terms of the elements of its factors. The element c_{ij} is the coefficient of x_j in the expression for z_i in terms of $x_1, x_2, \cdots, x_n$.

But $z_i = a_{i1}y_1 + \cdots + a_{im}y_m$ and

$$y_1 = \cdots + b_{1j}x_j + \cdots,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$y_m = \cdots + b_{mj}x_j + \cdots.$$

Therefore,

$$z_i = \cdots + (a_{i1}b_{1j} + \cdots + a_{im}b_{mj})\,x_j + \cdots,$$

hence

$$c_{ij} = a_{i1}b_{1j} + \cdots + a_{im}b_{mj}.$$

Thus, the element in the ith row and the jth column of the product of two matrices is equal to the sum of the products of the elements of the ith row of the first factor by the corresponding elements of the jth column of the second factor. For example:

$$\begin{pmatrix} 2 & 1 & 3 \\ 3 & 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 1 & 2 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 2\cdot 3 + 1\cdot 1 + 3\cdot 2 & 2\cdot 1 + 1\cdot 2 + 3\cdot 4 \\ 3\cdot 3 + 1\cdot 1 + 1\cdot 2 & 3\cdot 1 + 1\cdot 2 + 1\cdot 4 \end{pmatrix} = \begin{pmatrix} 13 & 16 \\ 12 & 9 \end{pmatrix}$$

Although a matrix is, so to speak, a "composite" object and many elements enter into its formation, it is useful and convenient to denote it by a single letter and to preserve the usual notation for operations of addition and multiplication. We shall use the capital letters of the Latin alphabet to denote matrices. The application of such a concise notation brings simplicity and lucidity into the theory of matrices by embracing in short formulas, that remind us of the formulas of ordinary algebra, complicated relations connecting a set of numbers, namely the elements of the matrix that occur in these formulas. Thus, for example, the set of linear forms

$$a_{11}x_1 + \cdots + a_{1n}x_n,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$a_{m1}x_1 + \cdots + a_{mn}x_n$$

appears in matrix notation as AX, where A is the coefficient matrix and X the "column" formed by the variables $x_1, x_2, \cdots, x_n$. The system of linear equations

$$a_{11}x_1 + \cdots + a_{1n}x_n = b_1,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$a_{m1}x_1 + \cdots + a_{mn}x_n = b_m$$

is written as

$$AX = B,$$

where A is the coefficient matrix, X the column of the unknowns, and B the column of the absolute terms.

The fundamental operations on matrices, namely addition and multiplication, are, of course, not always defined. The operation of addition makes sense for matrices of *equal structure*, i.e., having the same number of rows and of columns. As the result of addition, we obtain a matrix of the same structure. The operation of multiplication makes sense if the number of columns of the first matrix is equal to the number of rows of the second. As the result we obtain a matrix in which the number of rows is equal to the number of rows of the first factor and the number of columns is equal to the number of columns of the second factor.

The operations on square matrices are subject to most of the laws for operations on number, but some of the laws turn out to be violated.

Let us enumerate the fundamental properties for operations on matrices:

1. $A + B = B + A$ (commutative law for addition).
2. $(A + B) + C = A + (B + C)$ (associative law for addition).
3. $c(A + B) = cA + cB$
3′. $(c_1 + c_2)A = c_1A + c_2A$ (distributive laws for multiplication by a number. Here c, c_1, c_2 are numbers and not matrices).

4. $(c_1c_2)\,A = c_1(c_2A)$ (associative law for multiplication by a number).

5. There exists a "null" matrix

$$O = \begin{pmatrix} 0 \cdots 0 \\ \cdots\cdots \\ 0 \cdots 0 \end{pmatrix}$$

such that $A + O = A$ for every matrix A.

6. $c \cdot O = 0 \cdot A = O$; conversely, if $cA = O$, then $c = 0$ or $A = O$ (here c is a number).

7. For every matrix A there exists an opposite matrix $-A$, i.e., such that $A + (-A) = O$.

8. $(A + B) \cdot C = AC + BC$ }
8'. $C(A + B) = CA + CB$ } (distributive laws for addition and multiplication of matrices).

9. $(AB)\,C = A(BC)$ (associative law for multiplication).

10. $(cA)\,B = A(cB) = c(AB)$.

These properties hold not only for square matrices but also for arbitrary rectangular matrices with the sole proviso that the operations that occur in each of the numbered formulas must be defined. For square matrices of equal order this proviso is automatically fulfilled.

All these properties of the operations are similar to the properties of operations on numbers.

We shall now point out two peculiarities of the operations on matrices. First, the commutative law for the multiplication of matrices, even square ones, need not hold; i.e., AB is not always equal to BA. For example:

$$\begin{pmatrix} 3 & -2 \\ -1 & 4 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} -3 & 2 \\ 11 & 6 \end{pmatrix};$$

$$\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}\begin{pmatrix} 3 & -2 \\ -1 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 6 \\ 7 & 2 \end{pmatrix}.$$

Second, the product of two numbers is, of course, equal to zero if and only if one of the factors is equal to zero. This theorem is well known to be fundamental in the theory of algebraic equations. But under multiplication of matrices it turns out to be false. For the product of two matrices may be equal to the null matrix, although neither factor is equal to the null matrix. For example:

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

Let us mention yet another property of the multilplication of matrices. The matrix $\bar{A}$ is called the *transpose* of A if in every row of $\bar{A}$ there stand the elements of the corresponding column of A in the same order. For example, for the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

the transpose is the matrix

$$\bar{A} = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}.$$

The operation of multiplication is connected with that of transposition by the formula

$$\overline{AB} = \bar{B}\bar{A},$$

which is easily verified on the basis of the multiplication rule for matrices.

The theory of matrices forms an indispensable part of linear algebra in that it plays the role of an apparatus for stating and solving its problems.

Geometric analogies in linear algebra. Apart from the earlier described source for the emergence of the ideas and problems of linear algebra, there are also the needs of mathematical analysis and geometry, in particular analytic geometry, that lead to the development of linear algebra and, in turn, enrich it by important ideas and analogies. It is well known that the analytic geometry of the plane, and, in an even greater measure, of the space, as far as the theory of straight lines and planes is concerned, makes use of the apparatus of linear algebra in its simplest form. For a straight line in the plane is given by a linear equation in two variables that links the two coordinates of an arbitrary point of the line. A plane in space is given by a linear equation in three variables (the coordinates of an arbitrary point of this plane), a line in space by two linear equations.

A special simplicity and clarity is, of course, brought into analytical geometry and consequently into the theory of the simplest systems of linear equations by the use of a concept of a vector. Now a similar simplicity and clarity is brought into linear algebra, in particular into the general theory of systems of linear equations, by the use of the concept of a vector in a generalized sense. The way to this generalization is the following. A vector (in space) is given by three numbers, namely its three projections on the coordinate axes. Every triplet of real numbers in turn can be represented geometrically in the form of a vector (in space).

For vectors the operations of addition ("by the parallelogram rule") and multiplication by a number are defined. These operations are defined in accordance with similar operations on forces, velocities, accelerations, and other physical quantities that can be represented by means of vectors.

If vectors are given by their coordinates (i.e., their projections on the coordinate axes), then the operations of addition and multiplication by a number performed on vectors correspond to the analogous operations on the rows (or columns) of their coordinates.

Thus, it is convenient to interpret a row or column of three elements geometrically as a vector in three-dimensional space, and then the basic operations on "rows" (or "columns") are interpreted by the corresponding operations on vectors in space, so that the algebra of rows (or columns) of three elements formally does not differ at all from the algebra of the vectors of three-dimensional space. This circumstance makes it natural to introduce a geometric terminology into linear algebra.

A column (or row) of n numbers

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

is regarded as a "vector", i.e., as an element of some "n-dimensional vector space." The sum of the vectors

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

is taken to be the vector

$$\begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \cdots\cdots \\ x_n + y_n \end{pmatrix};$$

the product of the vector

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

by the number c is taken to be the vector

$$\begin{pmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{pmatrix}.$$

The set of all vectors (columns) forms, by definition, the n-dimensional arithmetical vector space.

Together with the n-dimensional arithmetical vector space we can introduce the concept of an n-dimensional point space, by associating with each column of n real numbers a geometrical image, namely a point. Then the n-dimensional vector space is defined in the following way. With every pair of points A and B we associate the vector $\overrightarrow{AB}$ leading from A to B by taking as its coordinates (its projections on the coordinate axes), by definition, the difference of the corresponding coordinates of the points B and A. Two vectors are taken to be equal if their corresponding coordinates are equal, just as in three-dimensional space we regard vectors as equal if one of them is obtained from the other by a parallel shift.

Between the vectors of an n-dimensional vector space and the points of an n-dimensional point space, there exists a natural one-to-one correspondence.

The point

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

is taken as the "origin of coordinates," and to every point there corresponds the vector that joins the origin to that point. Then we associate with every vector the point that is the end point of this vector, assuming that its beginning coincides with the origin. The introduction of the point space creates new analogies that enable us to "see" better in n-dimensional space.

However, in the further generalizations (§2) a rigorous definition of a point space becomes rather more complicated, and we shall, therefore, not make use of this concept. The reader who wishes to use the analogies arising from the investigation of a point space should visualize the elements of a vector space as vectors emanating from the origin of coordinates.

The introduction of a geometric terminology enables us to use in linear algebra analogies based on the geometric intuition which originates in the study of the geometry of three-dimensional space.

Of course, these analogies must be used with a certain care, bearing in mind that every intuitive-geometric argument can be checked in a strictly logical way applying only precise definitions of "geometric" concepts and rigorous proofs of theorems.

A characteristic feature of the elements of an n-dimensional vector space is the existence of the operations of addition and multiplication by a number, with properties reminiscent of the operations on numbers. Namely, as we have already mentioned in the account of the properties of operations on matrices, for the operation of addition the commutative and associative laws are satisfied, the distributive laws (for multiplication by a number) hold, the operation of addition has a unique inverse, and the product of a vector by a number gives the null vector if and only if either the vector is the null vector or the number is zero.

However, not only these columns (and rows) possess the features referred to. Such features also belong to the set of matrices of equal structure and to physical vector quantities: forces, velocities, accelerations, etc. They also belong to some mathematical objects of an altogether different nature, for example: the set of all polynomials in one variable, the set of all continuous functions on a given interval $[a, b]$, the set of all solutions of a linear homogeneous differential equation, etc.

This circumstance motivates a further generalization of a vector space, namely the introduction of general linear spaces. The elements of such generalized spaces may be arbitrary mathematical or physical objects for which the operations of addition and multiplication by a number are defined in a natural fashion. Such a very general and abstract approach to the concept of a linear space does not bring any complications into the theory, as we have seen earlier: Every linear space (of course, n-dimensional; the meaning of this will be clarified in the next section) does not differ in its structure and its properties from the arithmetical linear space, but the field of applicability is considerably extended by this generalization and it becomes possible to apply the methods of linear algebra to a very wide range of problems of theoretical science.

§2. Linear Spaces

Definition of a linear space. We now proceed to a rigorous definition of a linear space.

A *linear space* is a collection of objects of an arbitrary nature for which the concepts of a sum and of a product by a number make sense and which satisfy the following postulates:

1. $(X + Y) + Z = X + (Y + Z)$.
2. There exists a "null" element $\mathbf{0}$ such that $X + \mathbf{0} = X$ for every X.
3. For every element X there exists an opposite $-X$ such that $X + (-X) = \mathbf{0}$.
4. $X + Y = Y + X$.
5. $1 \cdot X = X$.
6. $c_1(c_2X) = c_1c_2X$.
7. $(c_1 + c_2)X = c_1X + c_2X$.
8. $c(X + Y) = cX + cY$.

Here X, Y, Z are elements of the linear space; 1, c_1, c_2, c are numbers.

These postulates (which are also called the *axioms of a linear space*) are very natural and constitute a formal account of those properties of the operations of addition and multiplication by a number that are necessarily linked with the concept of these operations in whatever generalized sense they are to be understood. Operations having one physical meaning or another are, in fact, treated as addition and multiplication by a number in all cases when these operations satisfy the postulates 1–8.

We mention some consequences of these axioms:

a. The null element $\mathbf{0}$ of the space is unique, i.e., there exists only one element satisfying axiom 2.

b. The opposite element of a given element X is unique.

c. "Subtraction" has a meaning; i.e., when a sum and one of the summands is given, the other summand is always defined, in fact, uniquely: If $X + Z = Y$, then $Z = Y + (-X)$.

d. $0 \cdot X = c \cdot \mathbf{0} = \mathbf{0}$.

e. If $cX = \mathbf{0}$, then either $c = 0$, or $X = \mathbf{0}$.

f. $-X = (-1)X$.

The proof of these consequences are very simple and will be omitted. In what follows the elements of a linear space will be called vectors.

Linear dependence and independence of vectors. We now proceed to the important concept of linear dependence and independence of vectors.

The vector $c_1X_1 + c_2X_2 + \cdots + c_mX_m$ with arbitrary numerical values of the coefficients $c_1, c_2, \cdots, c_m$ is called a linear combination of the vectors $X_1, X_2, \cdots, X_m$. If among the vectors $X_1, X_2, \cdots, X_m$ there is at least one that is a linear combination of the remaining ones, then the vectors $X_1, X_2, \cdots, X_m$ are called linearly dependent. But if none of the

vectors $X_1, X_2, \cdots, X_m$ is a linear combination of the remaining ones, then the vectors $X_1, X_2, \cdots, X_m$ are called linearly independent.

It is easy to see that for linear independence of the vectors $X_1, X_2, \cdots, X_m$ it is necessary and sufficient that the relation $c_1X_1 + c_2X_2 + \cdots + c_mX_m = 0$ should hold for $c_1 = c_2 = \cdots = c_m = 0$ only.

For vectors of the ordinary three-dimensional space the concepts of linear dependence and independence have a simple geometrical meaning.

Suppose two vectors X_1 and X_2 are given. Their linear dependence means that one of the vectors is a "linear combination" of the other, i.e., that they differ simply by a numerical factor. This means that both vectors belong to a common straight line, i.e., that they have equal or opposite direction.

Conversely, if two vectors are contained in one straight line, then they are linearly dependent. Consequently linear independence of two vectors X_1 and X_2 means that these vectors cannot be placed on one straight line; their directions are essentially distinct.

Let us now investigate what linear dependence and independence of three vectors means. Suppose that the vectors X_1, X_2 and X_3 are linearly dependent and, for the sake of definiteness, that the vector X_3 is a linear combination of the vectors X_1 and X_2. Then X_3 obviously lies in a plane containing the vectors X_1 and X_2; i.e., all three vectors X_1, X_2, X_3 belong to one plane. It is easy to see that if the vectors X_1, X_2, X_3 lie in one plane, then they are linearly dependent. For if the vectors X_1 and X_2 do not lie on one line, then X_3 can be decomposed with respect to X_1 and X_2, i.e., represented as a linear combination of X_1 and X_2. But if X_1 and X_2 lie on one line, then already X_1 and X_2 are linearly dependent.

Thus, linear dependence of three vectors X_1, X_2, X_3 is equivalent to the fact that they lie in one plane. Therefore X_1, X_2, X_3 are linearly independent if and only if they do not belong to one plane.

Four vectors in three-dimensional space are always linearly dependent. For if the vectors X_1, X_2, X_3 are linearly dependent, then the vectors X_1, X_2, X_3, X_4 are also linearly dependent for any X_4. But if X_1, X_2, X_3 are linearly independent, then they do not lie in one plane and every vector X_4 can be decomposed with respect to X_1, X_2, X_3, i.e., represented as a linear combination of them.

The preceding arguments can be generalized in the following way.

In three-dimensional space the vectors $X_1, X_2, \cdots, X_k$ $(k \geqslant 3)$ are linearly dependent if and only if they belong to a space (straight line or plane) of a dimension less than k.

In what follows we shall see after a rigorous definition of subspace and dimension that also in the general case linear dependence of the vectors $X_1, X_2, \cdots, X_k$ is equivalent to the fact that they belong to a space whose dimension is less than k; i.e., the "geometrical" meaning of linear dependence remains the same as for vectors in three-dimensional space.

The following theorem plays a fundamental role in the theory of linear spaces. If the vectors $X_1, X_2, \cdots, X_m$ are linear combinations of the vectors $Y_1, Y_2, \cdots, Y_k$ and $m > k$, then $X_1, X_2, \cdots, X_m$ are linearly dependent (theorem on the linear dependence of linear combinations).

For $k = 1$ the theorem is obvious. For $k > 1$ it is easily proved by the method of mathematical induction with respect to k.

Basis and dimension of a space. In three-dimensional space any three vectors X_1, X_2, X_3 that do not lie in one plane (i.e., that are linearly independent) form a *basis* of the space, which means that every vector of the space can be decomposed with respect to X_1, X_2, X_3, i.e., represented as a linear combination of them.

General linear vector spaces can be divided into two types.

It can happen that a space contains an arbitrarily large number of linearly independent vectors. Such spaces are called infinite-dimensional and their study leads to a branch of linear algebra that is the topic of a special mathematical discipline, functional analysis (see Chapter XIX).

A linear space is called finite-dimensional if there exists a finite bound for the number of linearly independent vectors, i.e., a number n such that there exist in the space n linearly independent vectors, but that any vectors more than n in number are linearly dependent. The number n is called the *dimension* of the space.

Thus, the space of vectors of the ordinary geometrical three-dimensional space is three-dimensional also in the sense of the general definition we have given. For in the three-dimensional geometric space, there exist many triplets of linearly independent vectors, but any four vectors are linearly dependent.

The space of n-term columns is n-dimensional in the sense of our definition. For there are n linearly independent vectors in the space, for example

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}, \cdots, \quad e_n = \begin{pmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ 1 \end{pmatrix} \tag{2}$$

but every vector

$$\begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{pmatrix}$$

of the space is a linear combination of them, namely: $x_1e_1 + x_2e_2 + \cdots + x_ne_n$. Therefore by the theorem of linear dependence of linear combinations any vectors more than n in number are linearly dependent.

Polynomials in one variable form a linear space. For there is a natural definition of the operations of addition and of multiplication by a number for polynomials, and they satisfy the axioms 1–8. However, this space is infinite-dimensional, since the vectors $1, x, \cdots, x^N$ are linearly independent for any N. But the set of polynomials whose degree does not exceed a given number N form a finite-dimensional space whose dimension is $N+1$. For the vectors $1, x, \cdots, x^N$ are linearly independent and their number is $N+1$. Now every polynomial whose degree does not exceed N is a linear combination of $1, x, \cdots, x^N$ so that by the theorem on linear independence any polynomials of degree $\leqslant N$, if they are more than $N+1$ in number, are linearly dependent.

We now introduce the important concept of a *basis* for an n-dimensional space. A basis is defined as a set of linearly independent vectors of the space such that every vector of the space is a linear combination of vectors of this set. Thus, in the space of columns a basis is, for example, the set of vectors (2). In the space of polynomials of degree $\leqslant N$ the "vectors" $1, x, \cdots, x^N$ can be taken as a basis. In the three-dimensional geometrical space any triplet of linearly independent vectors plays the role of a basis.

In an n-dimensional linear space, every set of n linearly independent vectors (and the existence of at least one such set is part of the definition of an n-dimensional space) form a basis of the space. For let $e_1, e_2, \cdots, e_n$ be linearly independent vectors of an n-dimensional linear space and X an arbitrary vector of the space. Then the vectors $X, e_1, \cdots, e_n$ are linearly dependent (since their number is more than n), i.e., there are numbers $c, c_1, c_2, \cdots, c_n$, not all equal to zero, such that $cX + c_1e_1 + \cdots + c_ne_n = 0$. Here $c \neq 0$, because if we had $c = 0$, then the vectors $e_1, e_2, \cdots, e_n$ would be linearly dependent. Therefore $X = -(c_1/c)\,e_1 - \cdots - (c_n/c)\,e_n$; i.e., every vector of the space is a linear combination of the vectors $e_1, e_2, \cdots, e_n$.

Every basis of an n-dimensional linear space consists of exactly n vectors. For the vectors of a basis are linearly independent and therefore their number cannot be larger than n. On the other hand, let $e_1, e_2, \cdots, e_k$ be an arbitrary basis of an n-dimensional space. We have already established that $k \leqslant n$. But every vector of the space, by definition of a basis, is a linear combination of the vectors $e_1, e_2, \cdots, e_k$ and by the theorem on linear dependence of linear combinations any vectors, more than k in number, are linearly dependent, from which it follows that the dimension n of the space is not larger than the number k of vectors of a basis. Thus, $k = n$, and this is what we had to prove.

We now introduce *coordinates* of a vector with respect to a given basis $e_1, e_2, \cdots, e_n$. As we have shown earlier, every vector X is a linear combination of the vectors of the basis. This representation is unique. For suppose that the vector X is expressed in terms of the basis $e_1, e_2, \cdots, e_n$ in two ways:

$$X = x_1e_1 + x_2e_2 + \cdots + x_ne_n,$$

$$X = x_1'e_1 + x_2'e_2 + \cdots + x_n'e_n.$$

Then $(x_1 - x_1')\,e_1 + (x_2 - x_2')\,e_2 + \cdots + (x_n - x_n')\,e_n = 0$, and from this it follows by the linear independence of $e_1, e_2, \cdots, e_n$ that $x = x_1', \cdots, x_n = x_n'$.

The coefficients $x_1, x_2, \cdots, x_n$ in the decomposition of an arbitrary vector X in terms of the vectors of a basis are called the *coordinates* of X in this basis. In this way every vector, *once a basis of the space is chosen*, can in a natural manner be associated with the row (or column) of its coordinates and vice versa: every row (or column) of n numbers can be regarded as the set of coordinates of a certain vector.

The operations of addition of vectors and multiplication of a vector by a number correspond to the similar operations on the rows (or columns) of their coordinates.

Therefore every n-dimensional linear space, irrespective of the nature of its elements (they may be functions, matrices, any physical quantities whatsoever, etc.), does not differ at all from the space of rows (or columns) with respect to these operations. Thus, as we have already mentioned, the generalized axiomatic approach to the concept of a linear space does not lead to any complications in comparison with the treatment of the space as a space of rows, but it extends the domain of applicability of this concept considerably.

An identity of properties of two sets of objects in relation to a given system of operations (or arbitrary other relations between their elements) is called in mathematics an isomorphism. An exact definition of isomorphism of algebraic systems will be given in Chapter XX. Using this term we can say that all n-dimensional linear spaces, irrespective of the nature of their elements, are isomorphic to one another and isomorphic to a single model, namely the space of rows.

Subspaces. A set of vectors of an n-dimensional linear space R_n, satisfying the condition that every linear combination of arbitrary vectors of the set under consideration also belongs to it, is called a *subspace* of the space. Obviously a subspace of the space R_n is itself a linear space and has, therefore, bases and a dimension. It is also obvious that the dimension of the subspace does not exceed the dimension of the whole space and can be equal to it if and only if the subspace coincides with the whole space.

Examples of subspaces of the three-dimensional vector space are the planes and lines we have studied, to within a translation, more accurately, the sets of all vectors that lie in a plane or on a line.

Very frequently we have to investigate the subspaces "spanned" by a system of vectors. These subspaces are defined as follows. Suppose that a system of linearly independent or dependent vectors $X_1, X_2, \cdots, X_m$ of the space R_n is given. Then the set of all linear combinations of these vectors $\{c_1X_1 + c_2X_2 + \cdots + c_mX_m\}$ forms a subspace of R_n which is called the subspace spanned by the vectors $X_1, X_2, \cdots, X_m$.

The dimension of this subspace is called the *rank* of the system of vectors $X_1, X_2, \cdots, X_m$. It is easy to see that the rank of a system of vectors is equal to the maximal number of linearly independent vectors contained in the system.

The "set" consisting only of the null vector formally satisfies the conditions imposed on a subspace. The dimension of this subspace is taken to be zero.

If two subspaces of a space R_n are given, then we can form from them in a natural manner two other subspaces, their vector sum (or union) and their intersection.

The *vector sum* of two subspaces P and Q is defined as the set of all sums of vectors belonging to the subspaces P and Q. The vector sum can also be regarded as the subspace spanned by the union of the bases of the subspaces P and Q.

The *intersection* of two subspaces is defined as the set of all vectors that belong to both subspaces. For example, the vector sum of two planes (i.e., two-dimensional vector subspaces) in the ordinary three-dimensional space is the whole space (provided only that the planes do not coincide) and their intersection is a straight line (under the same proviso).

The dimensions p and q of the two given subspaces, the dimension t of their vector sum, and the dimension s of their intersection satisfy the following interesting relation:

$$p + q = t + s.$$

We omit the proof of this statement.

From this relation we can make certain deductions concerning the intersection of subspaces in special cases. For example, two noncoincident planes (i.e., two-dimensional subspaces) in a space of four dimensions intersect in general only in a point (the dimension of their intersection is zero) and two planes intersect in a straight line only if their vector sum is three-dimensional, i.e., if both planes belong to some three-dimensional subspace. For in this case $t + s = 2 + 2 = 4$, from which it follows that $s = 1$ only when $t = 3$.

Complex linear spaces. In the description of the space of rows and of the general linear space, we have not specified what sort of numbers we are dealing with in the definition of the operation of multiplication of a vector by a number. Since we have started out from a generalization of the ordinary vectors, i.e., the directed segments in the geometrical three-dimensional space, we have had in mind arbitrary real numbers. The so-constructed linear spaces, which are called real linear spaces, generalize in the most natural way the three-dimensional space of ordinary vectors. However, in many problems of contemporary mathematics it turns out to be useful to consider a *complex linear space*. By this we mean a collection of objects for which the operations of addition and of multiplication by an arbitrary complex number are defined so that these operations satisfy all the axioms 1–8. As an example of a complex space, we can take the space of rows whose elements are arbitrary complex numbers.

Formally the theory of complex spaces does not differ essentially from the theory of real spaces.

However, even a two-dimensional complex space does not have an intuitive geometric interpretation. The fact is that an n-dimensional complex space can also be regarded as a real one, in view of the fact that since the operation of multiplication by an arbitrary complex number is defined for it, the operation of multiplication by a real number is defined just as well. But the dimension of a complex n-dimensional space regarded as a real one is equal to $2n$, i.e., twice as much. For if $e_1, e_2, \cdots, e_n$ is a basis of the complex space, then we can take as a basis of the same space, regarded as a real one, for example the vectors

$$e_1, ie_1, e_2, ie_2, \cdots, e_n, ie_n, \quad \text{where} \quad i = \sqrt{-1}.$$

Therefore a two-dimensional complex space can be interpreted as a real one, but four-dimensional.

Furthermore, the theory of linear spaces does not undergo any changes formally if as the collection of numbers by which the "vectors" of the space may be multiplied we take an arbitrary set of numbers, other than that of all real or all complex numbers, provided only that the results of the basic arithmetical operations (addition, subtraction, multiplication, and division) performed on numbers of the set again belong to the set. A set of numbers satisfying these postulates is called a number field. (This concept will be studied in more detail in Chapter XX.) As an example of a number field, we can take the field of rational numbers.

In some parts of algebra that are close to the theory of numbers, the theory of linear spaces over an arbitrary field is successfully applied.

The n-dimensional Euclidean space. Some important concepts of the ordinary vector space have not yet been generalized in the preceding account, in particular, the concept of the length of a vector and the angle between vectors. It is well known that in analytic geometry problems relating to the intersection of lines and planes, parallelism, and many others make no use of these concepts. The properties of a space whose description does not require the concepts of the length of a vector and of angle can be characterized as the properties that remain unchanged under arbitrary affine transformations [see Chapter III, §11]. For this reason linear spaces in which the concept of the length of a vector is not defined are called *affine spaces*.

However, many problems of mathematics require generalizations of the concepts of the length of a vector and of an angle to n-dimensional spaces. These generalizations proceed by means of an analogy with the theory of vectors in a plane or in space.

Let us consider, first of all, the real space of rows. The length of the vector $X = (x_1, x_2, \cdots, x_n)$ is defined to be the number

$$|X| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

This is quite natural, since for $n = 2$ and $n = 3$ the length of a vector is computed precisely by this formula in terms of its coordinates with respect to Cartesian coordinate axes.

The concept of the angle between vectors is introduced in a natural way by the following considerations. In a plane and in space the angle between the vectors X and Y is the angle at the vertex A in the triangle with the sides $AB = |X|$, $AC = |Y|$ and $BC = |X - Y|$.

In an n-dimensional space it is natural to take this as the definition of the angle between vectors, i.e., to proceed as if we could "draw" a pair of vectors in an n-dimensional space and "place" them in a plane preserving their lengths and the angle between them. However, such a definition would lack rigor; the existence of a triangle ABC with the lengths of the vectors $|X|$, $|Y|$, and $|X - Y|$ is needed in the proof.

Disregarding this inaccuracy, we introduce a formula for the computation of an angle. By a well-known formula of trigonometry we have

$$BC^2 = AB^2 + AC^2 - 2AB \cdot AC \cos\phi;$$

hence,

$$\cos\phi = \frac{|X|^2 + |Y|^2 - |X - Y|^2}{2|X| \cdot |Y|}$$

$$= \frac{x_1^2 + \cdots + x_n^2 + y_1^2 + \cdots + y_n^2 - (x_1 - y_1)^2 - \cdots - (x_n - y_n)^2}{2|X| \cdot |Y|}$$

$$= \frac{x_1y_1 + \cdots + x_ny_n}{|X| \cdot |Y|}.$$

If we retain the term "scalar product," as in three-dimensional space, for the product of the lengths of vectors by the cosine of the angle between them, we find that the scalar product of the vectors is computed by the formula

$$X \cdot Y = x_1y_1 + \cdots + x_ny_n,$$

which for $n = 2$ and $n = 3$ coincides with the well-known formulas for the scalar product of ordinary vectors.

Strictly speaking, the expression $x_1y_1 + \cdots + x_ny_n$ should be taken as the definition of the *scalar product* (because there is a lack of rigor in the definition of the scalar product by means of the angle) and then the angle between vectors can be defined by the formula

$$\cos\phi = \frac{X \cdot Y}{|X| \cdot |Y|}. \tag{3}$$

This is what we shall do.

To justify this definition of an angle we have to show that the absolute value of the right-hand side of formula (3) does not exceed 1, i.e., that $(X \cdot Y)^2 \leqslant |X|^2 \cdot |Y|^2$.

In expanded form this inequality becomes

$$(x_1y_1 + \cdots + x_ny_n)^2 < (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2).$$

It is known as the Cauchy-Bunjakovskiĭ inequality and can be proved directly, by a fairly tedious computation. We shall prove it by the following indirect argument.

First of all we mention that the scalar multiplication of vectors has the following properties:

1′. $X \cdot X = |X|^2 > 0$ for $X \neq 0$.

2′. $X \cdot Y = Y \cdot X$.

3′. $(cX) \cdot Y = c(X \cdot Y)$.

4′. $(X_1 + X_2) \cdot Y = X_1 \cdot Y + X_2 \cdot Y$.

That these properties hold follows immediately from the expression of the scalar product in terms of the coordinates.

We now introduce the vector $Y + tX$, where t is an arbitrary real number. We have $|Y + tX|^2 \geqslant 0$, because the square of the length of a vector cannot be negative. But by the properties of the scalar product $|Y + tX|^2 = |Y|^2 + 2tX \cdot Y + t^2|X|^2$. Moreover, it is known that a quadratic trinomial is nonnegative for all values of the real variable t if and only if its roots are imaginary or equal, i.e., if its discriminant is

negative or zero. But the discriminant of the trinomial $|Y|^2 + 2tX \cdot Y + t^2|X|^2$ is equal to $4(X \cdot Y)^2 - 4|X|^2|Y|^2$ so that $(X \cdot Y)^2 - |X|^2|Y|^2 \leqslant 0$, and this is equivalent to the Cauchy-Bunjakovskiĭ inequality.

From this inequality it follows that

$$\frac{|X \cdot Y|}{|X||Y|} \leqslant 1$$

and therefore the definition of an angle by the formula (3) is justified.

Furthermore, it is easy to deduce the inequalities

$$|X| - |Y| \leqslant |X \pm Y| \leqslant |X| + |Y|,$$

which imply, in particular, the existence of a triangle with the sides $|X|$, $|Y|$ and $|X - Y|$, so that the nonrigorous definition of an angle given previously, which was based on geometric intuition, now also becomes valid.

Axiomatic definition of an *n*-dimensional Euclidean space. In the preceding section we have introduced the concepts of the length of a vector, of angle, and of the scalar product in the space of rows. In the general axiomatic definition of an n-dimensional real linear space, these concepts are also defined axiomatically, and the concept of a scalar product comes first.

Scalar multiplication of vectors of a linear real space is the name for an operation which associates with every pair of vectors X and Y a real number, their so-called scalar product $X \cdot Y$, where this operation must satisfy the following postulates (axioms):

1′. $X \cdot X > 0$ for $X \neq 0$, $0 \cdot 0 = 0$.

2′. $X \cdot Y = Y \cdot X$.

3′. $(cX) \cdot Y = c(X \cdot Y)$.

4′. $(X_1 + X_2) \cdot Y = X_1 \cdot Y + X_2 \cdot Y$.

Furthermore, by the length of a vector we mean the number $\sqrt{X \cdot X}$, by the cosine of the angle between the vectors X and Y the number $X \cdot Y/|X| \cdot |Y|$. To justify this latter inequality, it is necessary to establish the Cauchy-Bunjakovskiĭ inequality $(X \cdot Y)^2 \leqslant |X|^2|Y|^2$. But this can be done exactly as we have done it in the preceding section. In our proof we have made use only of the properties 1′, 2′, 3′ and 4′ of the scalar product, the specific nature of the space of rows playing no role in the proof. A real linear space in which a scalar multiplication satisfying the axioms 1′–4′ has been introduced is called a *Euclidean space*.

In various concrete linear spaces that are studied in mathematics, scalar products are introduced by various methods whose choice is dictated by the nature of the problem. For example, in the spaces whose elements are the functions of one variable $X(t)$ defined on a given interval $a \leqslant t \leqslant b$, the scalar product of two elements $X(t)$ and $Y(t)$ is often taken to be the number $\int_a^b X(t)\,Y(t)\,dt$ or $\int_a^b X(t)\,Y(t)\,p(t)\,dt$, where $p(t)$ is some positive function. It is easy to see that all the axioms 1′–4′ are satisfied for either of these definitions.

Orthogonality; orthonormal bases. Two vectors of a Euclidean space are called *orthogonal* (or perpendicular) if their scalar product is zero. It is easy to see that pairwise orthogonal nonzero vectors are always linearly independent. For suppose that $X_1, X_2, \cdots, X_m$ are pairwise orthogonal nonzero vectors and that $c_1X_1 + c_2X_2 + \cdots + c_mX_m = 0$.

By the property of the scalar product $X_1(c_1X_1 + c_2X_2 + \cdots + c_mX_m) = c_1|X_1|^2 = 0$; hence $c_1 = 0$. In the same way we can show that $c_2 = \cdots = c_m = 0$. Therefore $X_1, \cdots, X_m$ are linearly independent.

From what we have proved, it follows that in an n-dimensional space there cannot be more than n pairwise orthogonal nonzero vectors and that every set of n pairwise orthogonal vectors forms a basis of the space. If, moreover, the lengths of all the n pairwise orthogonal vectors are 1, then the basis they form is called *orthonormal*.

It is not difficult to show, but we shall omit the proof, that a Euclidean space has orthonormal bases, in fact infinitely many. Moreover, if in a space R some subspace P is chosen, then an orthonormal basis of the subspace can be extended to an orthonormal basis of the whole space by adjoining certain vectors.

It is often convenient to define vectors in a Euclidean space by their coordinates in an arbitrary orthonormal basis, because in this case we obtain a particularly simple expression for the scalar product. For if a vector X has the coordinates $(x_1, x_2, \cdots, x_n)$ in the orthonormal basis $e_1, e_2, \cdots, e_n$ and the vector Y the coordinates $(y_1, y_2, \cdots, y_n)$, i.e.,

$$X = x_1e_1 + x_2e_2 + \cdots + x_ne_n \quad \text{and} \quad Y = y_1e_1 + y_2e_2 + \cdots + y_ne_n,$$

then by the property of the scalar product

$$\begin{aligned} X \cdot Y = {} & x_1y_1e_1e_1 + x_1y_2e_1e_2 + \cdots + x_1y_ne_1e_n \\ & + x_2y_1e_1e_1 + x_2y_2e_2e_2 + \cdots + x_2y_ne_2e_n \\ & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ & + x_ny_1e_ne_1 + x_ny_2e_ne_2 + \cdots + x_ny_ne_ne_n \\ = {} & x_1y_1 + x_2y_2 + \cdots + x_ny_n, \end{aligned}$$

since $e_i e_k = 0$ for $i \neq k$ and $e_i e_i = |e_i|^2 = 1$ for every $i = 1, 2, \cdots, n$. In particular, $X \cdot X = x_1^2 + x_2^2 + \cdots + x_n^2$.

Thus, the length of a vector and the scalar product are expressed in terms of the coordinates of an orthonormal basis by the same formulas as in the space of rows.

The transition from one of the models of a Euclidean space, namely the space of rows, to the general axiomatically defined Euclidean space does not introduce any complications, but extends the domain of applicability of the theory.

Now let us deal with the problem of orthogonal projection of vectors on a subspace. Let R_n be an n-dimensional Euclidean space and P_m an m-dimensional subspace of it. Further, let $e_1, e_2, \cdots, e_m, f_1, \cdots, f_{n-m}$ be an orthonormal basis of R_n including an orthonormal basis of the subspace P_m. The subspace Q_{n-m} spanned by the vectors $f_1, f_2, \cdots, f_{n-m}$ is called the *orthogonal complement* of the subspace P_m. Its dimension is $n - m$. The orthogonal complement Q_{n-m} can be characterized as the subspace consisting of all vectors that are orthogonal to every vector of the subspace P_m.

Every vector Z belonging to R_n can be expressed uniquely as a sum of vectors X and Y of which one belongs to P_m, the other to Q_{n-m}. This is clear, because the vector Z can be expressed uniquely in the form

$$Z = x_1 e_1 + \cdots + x_m e_m + y_1 f_1 + \cdots + y_{n-m} f_{n-m},$$

so that $X = x_1 e_1 + \cdots + x_m e_m$, $Y = y_1 f_1 + \cdots + y_{n-m} f_{n-m}$.

The vector X is called the *orthogonal projection* of Z onto P_m.

Unitary spaces. The concepts of the length of a vector and the scalar product of vectors can also be defined in a complex space. As before, the concept of the scalar multiplication is put first, and this is defined as follows. With every pair X and Y of vectors of a complex space we associate a complex (not necessarily real) number, their so-called scalar product $X \cdot Y$. The operation of scalar multiplication must satisfy the following axioms:

1″. $X \cdot X$ is real and positive for $X \neq 0$, $0 \cdot 0 = 0$.

2″. $Y \cdot X = (X \cdot Y)'$. Here the prime denotes transition to the conjugate complex number.

3″. $(cX) \cdot Y = c(X \cdot Y)$ for an arbitrary complex c.

4″. $(X_1 + X_2) \cdot Y = X_1 \cdot Y + X_2 \cdot Y$ (distributive law).

In the space of rows with complex elements the scalar product of the vectors $(x_1, \cdots, x_n)$ and $(y_1, \cdots, y_n)$ can be taken to be the number $x_1 y_1' + \cdots + x_n y_n'$. It is easy to verify that all the axioms 1″-4″ are satisfied for this definition.

The length of a vector is defined as the number $\sqrt{X \cdot X}$. The concept of angle between vectors is not defined.

A complex linear space with a scalar product satisfying the axioms 1″-4″ is called a *unitary space*.

§3. Systems of Linear Equations

Systems of two equations with two unknowns and of three equations with three unknowns. A system of two linear equations with two unknowns appears in the following general form

$$a_1 x + b_1 y = c_1,$$
$$a_2 x + b_2 y = c_2.$$

To solve this system we multiply the first equation by b_2, the second by $-b_1$ and add. We obtain

$$(a_1 b_2 - a_2 b_1)\, x = c_1 b_2 - c_2 b_1.$$

Similarly, by multiplying the first equation by $-a_2$, the second by a_1, and adding, we obtain

$$(a_1 b_2 - a_2 b_1)\, y = a_1 c_2 - a_2 c_1.$$

From these equations it is easy to determine x and y, if only the expression $a_1 b_2 - a_2 b_1$ formed from the coefficients of the unknown x and y is different from zero. This expression is called the *determinant* of the matrix

$$\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}$$

formed from the coefficients of the system. The determinant is denoted:

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}.$$

From the definition it follows that the determinant is computed by the scheme

which requires no further explanation.

Let us return to the solution of the system. The expressions $c_1b_2 - c_2b_1$ and $a_1c_2 - a_2c_1$ also appear as determinants in accordance with our definition, namely,

$$\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}.$$

Thus, if the determinant

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$$

is different from zero, then we obtain the following formulas for the solution of the system:

$$x = \frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}; \quad y = \frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}. \tag{4}$$

Strictly speaking, these arguments are not complete. The operations on the equations that we have carried out to deduce the formulas for the solution of the system make sense only under the assumption that x and y are in fact numbers that form a solution of the system. The logical substance of our argument is the following: If the determinant of the coefficients of the system is not zero and the solution of the system exists, then it can be computed by the formulas (4). Therefore it is still necessary to verify that the values of the unknown that we have found do in fact satisfy both equations of the system. This can be done without any difficulty.

Thus, if the determinant of the matrix of the coefficients of the system is different from zero, then the system has a unique solution given by the formulas (4).

For a system of three equations with three unknowns

$$\begin{aligned} a_1x + b_1y + c_1z &= d_1, \\ a_2x + b_2y + c_2z &= d_2, \\ a_3x + b_3y + c_3z &= d_3, \end{aligned}$$

it is easy to carry out similar arguments and computations; for this purpose it is sufficient to add up the equations after multiplying them by factors such that after addition two of the unknowns disappear. To make the unknown y and z disappear, we have to take for these factors $b_2c_3 - b_3c_2$, $b_3c_1 - b_1c_3$ and $b_1c_2 - b_2c_1$, as is easy to verify by computation.

We obtain the result that if the expression

$$\Delta = a_1b_2c_3 - a_1b_3c_2 + a_2b_3c_1 - a_2b_1c_3 + a_3b_1c_2 - a_3b_2c_1$$

is different from zero, then the system has a unique solution obtained by the formulas

$$x = \frac{\Delta_1}{\Delta}, \quad y = \frac{\Delta_2}{\Delta}, \quad z = \frac{\Delta_3}{\Delta},$$

where $\Delta_1, \Delta_2, \Delta_3$ are the expressions obtained from Δ by replacing the coefficients of the corresponding unknown by the absolute terms.

The expression Δ is called the determinant of the matrix

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix}$$

and is denoted by

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

For the computation of a determinant the following scheme is useful:

In the first of these schemes, the lines (a diagonal and two triangles) connect the positions of the elements whose product occurs in the composition of the determinant with a plus sign; and in the second scheme, they connect the terms occuring in the determinant with a minus sign.

For systems of two equations with two unknowns and of three equations with three unknowns, we have obtained entirely similar results. In both cases the system has a unique solution, provided the determinant of the matrix of the coefficients is different from zero. The formulas for the solution are also similar: In the denominator of each of the unknowns stands the determinant of the matrix of coefficients, and in the numerators the determinants of the matrices that arise from the matrix of coefficients by replacing the coefficients of the unknown to be computed by the absolute terms.

An immediate generalization of these results to systems of n equations in n unknowns for arbitrary n is somewhat difficult. It becomes comparatively easy by an indirect method: First we generalize the concept of a

determinant to square matrices of arbitrary order, and having studied the properties of determinants we apply their theory to the investigation of systems of equations.

Determinants of the *n*th order. When we consider the explicit expression for the determinants

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1b_2 - a_2b_1$$

and

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1b_2c_3 - a_1b_3c_2 + a_2b_3c_1 - a_2b_1c_3 + a_3b_1c_2 - a_3b_2c_1 ,$$

we notice that in every term there occurs as a factor exactly one element from each row and one from each column of the determinant, and that all possible products of this form occur in the determinant with a plus or a minus sign. This property is at the bottom of the generalization of the concept of the determinant to square matrices of arbitrary order. In fact, the determinant of a square matrix of order n or, briefly, a *determinant of the nth order* is defined as the algebraic sum of all possible products of the elements of the matrix, precisely one from each row and one from each column; these products are given plus or minus signs by a certain well-defined rule. This rule is somewhat complicated to explain, and we shall not dwell on its formulation. It is sufficient to mention that it is arranged in such a way that the following important basic properties of a determinant are secured:

1. When two rows are interchanged, the determinant changes its sign.

For determinants of order 2 and 3, this property is easy to verify by an immediate computation. In the general case it is proved on the basis of the rule for the signs that we have not formulated here.

Determinants have quite a number of other remarkable properties that enable us to apply determinants successfully in diverse theoretical and numerical calculations, notwithstanding the fact that determinants are extraordinarily cumbersome: A determinant of order n contains, as is easy to see, $n!$ terms, each term consists of n factors, and the factors are provided with their signs according to a complicated rule.

We now proceed to enumerate the basic properties of determinants but omit their detailed proofs. The first of these properties has been formulated.

2. A determinant does not change when its matrix is transposed, i.e.,

when the rows are replaced by the columns, preserving their order. The proof is based on a detailed study of the rule of the distribution of signs in the terms of the determinant. This property enables us to transfer every statement concerning the rows of the determinant to a statement on columns.

3. A determinant is a linear function of the elements of each row (or column). In detail,

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{i1} & \cdots & a_{in} \\ \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} , \tag{5}$$

where $A_{i1}, A_{i2}, \cdots, A_{in}$ are expressions that do not depend on the elements of the ith row.

This property follows evidently from the fact that every term contains one and only one factor from each row, in particular the ith row.

The equation (5) is called the expansion of the determinant with respect to the elements of the ith row, and the coefficients $A_{i1}, A_{i2}, \cdots, A_{in}$ are called the algebraic complements of the elements $a_{i1}, a_{i2}, \cdots, a_{in}$ of the determinant.

4. The algebraic complement A_{ij} of the element a_{ij} is equal, apart from the sign, to the so-called minor Δ_{ij} of the determinant, i.e., the determinant of order $(n - 1)$ that arises from the given one by crossing out the ith row and jth column. To obtain the algebraic complement the minor must be taken with the sign $(-1)^{i+j}$. The properties 3 and 4 reduce the computation of a determinant of order n to the computation of n determinants of order $n - 1$.

The fundamental properties that we have enumerated have a number of interesting consequences. We now mention some of these.

5. A determinant with two equal rows is zero.

For if a determinant has two equal rows, then the determinant does not change when they are interchanged, because the rows are identical; but on the other hand, by our first property it should change its sign. Therefore it is equal to zero.

6. The sum of the products of the elements of any row by the algebraic complements of another row is zero.

For such a sum is the result of expanding a determinant with two equal rows with respect to one of them.

7. A common factor of the elements of any row can be taken before the determinant sign.

This follows from property 3.

8. A determinant with two proportional rows is zero.

It is sufficient to take out the factor of proportionality and then we have a determinant with two equal rows.

9. A determinant does not change if we add to the elements of any one row numbers that are proportional to the elements of another row.

For by property 3 the modified determinant is equal to the sum of the original determinant and a determinant with two proportional rows, which is zero.

The last property gives us a good method for computing determinants. Without changing the value of a determinant, we can transform its matrix by applying this rule so that in one row (or column) all the elements except one become zero. Then by expanding the determinant with respect to the elements of this row (column) we reduce the computation of a determinant of order n to that of a *single* determinant of order $n-1$, namely, the algebraic complement of the only nonzero element of the row in question.

For example, suppose we have to compute the determinant

$$\Delta = \begin{vmatrix} 1 & 1 & -1 & 2 \\ 2 & -1 & 1 & 1 \\ -1 & 2 & 0 & 1 \\ 1 & 1 & -2 & 1 \end{vmatrix}.$$

We add to the second column the first, multiplied by -1, to the third, the first, and to the fourth, the first, multiplied by -2 and obtain

$$\Delta = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & -3 & 3 & -3 \\ -1 & 3 & -1 & 3 \\ 1 & 0 & -1 & -1 \end{vmatrix}.$$

Expanding Δ by the elements of the first row, we obtain

$$\Delta = 1 \cdot (-1)^{1+1} \begin{vmatrix} -3 & 3 & -3 \\ 3 & -1 & 3 \\ 0 & -1 & -1 \end{vmatrix}.$$

Finally, we add to the first row the second and expand with respect to the elements of the first column; we obtain

$$\Delta = \begin{vmatrix} 0 & 2 & 0 \\ 3 & -1 & 3 \\ 0 & -1 & -1 \end{vmatrix} = 3 \cdot (-1)^{1+2} \begin{vmatrix} 2 & 0 \\ -1 & -1 \end{vmatrix} = -3 \cdot (-2) = 6.$$

The determinant of a matrix A is denoted by $|A|$.

In conclusion, we mention a further very important property of determinants.

The determinant of the product of two square matrices is equal to the product of the determinants of the factors, i.e., in short notation, $|AB| = |A||B|$.

This property enables us in particular to multiply determinants of equal order by the rule for multiplication of matrices.

Systems of n linear equations in n unknowns. With the apparatus of determinants, it is now easy to generalize the results obtained earlier for systems of two equations with two unknowns and of three equations with three unknowns to systems of n equations with n unknowns, under the assumption that the determinant of the coefficient matrix is different from zero.

Let

$$\begin{gathered} a_{11}x_1 + \cdots + a_{1j}x_j + \cdots + a_{1n}x_n = b_1, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + \cdots + a_{nj}x_j + \cdots + a_{nn}x_n = b_n \end{gathered}$$

be such a system. We denote by Δ the determinant of the coefficient matrix of the system. By assumption it is different from zero. Furthermore, we denote by A_{ij} the algebraic complement of the element a_{ij}. We multiply the first equation by A_{1j}, the second by A_{2j}, $\cdots$, the nth by A_{nj} and add. We obtain

$$\Delta x_j = b_1 A_{1j} + \cdots + b_n A_{nj}.$$

For the coefficients of all the unknowns except x_j vanish, because they appear as sums of the products of the algebraic complements of the elements of the jth column with the elements of other columns (property 6, applied to columns); but the coefficients of the unknown x_j is equal to the sum of the products of the elements of the jth column with their algebraic complements, i.e., is equal to Δ.

Thus,

$$x_j = \frac{b_1 A_{1j} + \cdots + b_n A_{nj}}{\Delta} \quad \text{for all} \quad j = 1, 2, \cdots, n. \tag{6}$$

As we have said, these arguments are valid only if we understand by $x_1, x_2, \cdots, x_n$ a solution of the system, the existence of which must be assumed in the first instance.

Hence the result of the argument is the following.

If a solution of the system exists, then it is given by the formulas (6) and is therefore unique.

To complete the exposition, it is necessary to prove the existence of a solution, and this can be done by substituting the values we have found for the unknowns into all the equations of the original system. It is easy to verify by using the same property of the determinant (but this time applied to the rows) that these values in fact satisfy all the equations.

Thus the following theorem is true: If the determinant of the coefficient matrix of a system of n equations with n unknowns is different from zero, then the system has a unique solution given by the formulas (6).

These formulas can be transformed by remarking that the sum $b_1A_{1j} + \cdots + b_nA_{nj}$ can be written in the form of a determinant, namely:

$$\Delta_j = b_1A_{1j} + \cdots + b_nA_{nj} = \begin{vmatrix} a_{11} & \cdots & b_1 & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & \cdots & b_n & \cdots & a_{nn} \end{vmatrix}$$

(the absolute terms occur in the jth column).

Hence the results we have stated for systems of equations with two and three unknowns have been completely generalized to a system of n equations, and even the formulas for the solution are formally exactly the same.

We mention one corollary of this theorem: If it is known of a system of equations that it has no solution at all or that the solution is not unique, then the determinant of the coefficient matrix is equal to zero.

This corollary is particularly often applied to homogeneous systems, i.e., those in which the absolute terms $b_1, b_2, \cdots, b_n$ are all zero. Homogeneous systems always have the obvious "trivial" solution $x_1 = x_2 = \cdots = x_n = 0$.

If a homogeneous system has, apart from the trivial one, also a nontrivial solution, then its determinant is zero.

This statement opens up the possibility of using the theory of determinants in other branches of mathematics and its applications.

Let us consider, for example, a problem in analytical geometry: to find the equation of the plane passing through three given points (x_1, y_1, z_1), (x_2, y_2, z_2) and (x_3, y_3, z_3) that do not lie on one line.

From elementary geometry it is known that such a plane exists. Suppose that its equation is of the form $Ax + By + Cz + D = 0$. Then

$$\begin{aligned} Ax_1 + By_1 + Cz_1 + D &= 0, \\ Ax_2 + By_2 + Cz_2 + D &= 0, \\ Ax_3 + By_3 + Cz_3 + D &= 0. \end{aligned}$$

Let x, y, z be the coordinates of an arbitrary point in that plane. Then we also have

$$Ax + By + Cz + D = 0.$$

We regard these four equations as a system of linear homogeneous equations for the coefficients A, B, C, D of the required plane. This system has a nontrivial solution, because the required plane exists. Therefore the determinant of the system is zero; i.e.,

$$\begin{vmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x & y & z & 1 \end{vmatrix} = 0. \tag{7}$$

Now this is the equation of the required plane. For it is an equation of the first degree in x, y, z, a fact which follows from the linearity of the determinant with respect to the elements of the last row.

By making use of the fact that the given points do not lie on one line, it is easy to verify that not all the coefficients of this equation are zero. Consequently the equation (7) is indeed the equation of a plane. This plane passes through the given points, because their coordinates obviously satisfy the equation.

Matrix notation for a system of n equations in n unknowns. A system of n linear equations in n unknowns

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1, \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{n1}x_1 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

can be written in matrix notation in the form of a single equation

$$AX = B.$$

Here A denotes the coefficient matrix, X the column formed by the unknowns, and B the column of the absolute terms.

The solution of the system (if the determinant of the matrix A is different from zero) can be written explicitly as follows [see formula (6)]:

$$\begin{aligned} x_1 &= \frac{A_{11}}{\Delta}b_1 + \frac{A_{21}}{\Delta}b_2 + \cdots + \frac{A_{n1}}{\Delta}b_n, \\ x_2 &= \frac{A_{12}}{\Delta}b_1 + \frac{A_{22}}{\Delta}b_2 + \cdots + \frac{A_{n2}}{\Delta}b_n, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n &= \frac{A_{1n}}{\Delta}b_1 + \frac{A_{2n}}{\Delta}b_2 + \cdots + \frac{A_{nn}}{\Delta}b_n \end{aligned}$$

or in matrix form

$$X = \begin{bmatrix} \frac{A_{11}}{\Delta} & \frac{A_{21}}{\Delta} & \cdots & \frac{A_{n1}}{\Delta} \\ \frac{A_{12}}{\Delta} & \frac{A_{22}}{\Delta} & \cdots & \frac{A_{n2}}{\Delta} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{A_{1n}}{\Delta} & \frac{A_{2n}}{\Delta} & \cdots & \frac{A_{nn}}{\Delta} \end{bmatrix} B.$$

The matrix standing as first factor on the right-hand side of the equation is called the *inverse* to the matrix A and is denoted by A^{-1}. Using this notation we obtain the solution of the system $AX = B$ in the following simple and natural form, which recalls the formula for the solution of a single equation in one unknown:

$$X = A^{-1}B.$$

We can easily give another proof of the result obtained, in terms of the algebra of matrices.

For this purpose we must first of all mention the special role of the matrix

$$E = \begin{pmatrix} 1 & 0 & 0 \cdots 0 \\ 0 & 1 & 0 \cdots 0 \\ \cdots & \cdots & \cdots \\ 0 & 0 & 0 \cdots 1 \end{pmatrix},$$

the so-called *unit matrix*.

The unit matrix plays among square matrices the same role as the number 1 among numbers. In fact, for every matrix A the following equations hold: $AE = A$ and $EA = A$. This is easy to verify by the rule for the multiplication of matrices.

The matrix A^{-1} defined previously, the inverse to A, plays in relation to it a similar role to that played by the inverse of a given number:

$$AA^{-1} = A^{-1}A = E.$$

The validity of these equations can be verified by the rule for the multiplication of matrices and by the properties 3 and 6 of a determinant.

Knowing these properties of the unit and the inverse matrix we can obtain the solution of the system $AX = B$ in the following way.

Suppose that $AX = B$. Then $A^{-1}(AX) = A^{-1}B$. But $A^{-1}(AX) = (A^{-1}A)X = EX = X$ and therefore $X = A^{-1}B$.

Suppose now that $X = A^{-1}B$, then $AX = AA^{-1}B = EB = B$.

Thus the "equation" $AX = B$ has the unique solution $X = A^{-1}B$, provided only that A^{-1} exists.

We have established the existence of the inverse matrix A^{-1} for A under the assumption that the determinant of A is different from zero. This condition is not only sufficient but also necessary for the existence of the inverse matrix. For suppose that the matrix A has an inverse A^{-1} such that $AA^{-1} = E$. Then by the property of the determinant of the product of two matrices

$$|A||A^{-1}| = |E| = 1,$$

and from this it follows that the determinant of A is not zero.

A matrix whose determinant is different from zero is called *nondegenerate* or *nonsingular*. We have therefore established that an inverse matrix always exists for nondegenerate matrices and only for them.

The introduction of the concept of an inverse matrix turns out to be useful not only in the theory of systems of linear equations but also in many other problems of linear algebra.

In conclusion, we mention that the formulas we have derived for the solution of linear systems are an irreplaceable tool in theoretical considerations but are not convenient for the numerical solution of systems. As we have already mentioned, various methods and computational schemes have been worked out for the numerical solution of systems, and in view of the great importance of this problem for practical investigations, the work of simplifying the numerical solution of systems (especially with large numbers of unknowns) is intensively pursued even at present.

The general case of systems of linear equations. We now turn to the investigation of systems of linear equations in the most general case when it is not assumed that the number of equations is equal to the number of unknowns. In such a general setting it cannot be expected, naturally, that a solution of the system always exists or that, in case it exists, it turns out to be unique. It is natural to assume that if the number of equations is less than the number of unknowns the system has infinitely many solutions. For example, two equations of the first degree in three unknowns are satisfied by the coordinates of every point on the straight line that is the intersection of the planes defined by the equations. However it can happen in this case that the system has no solution at all, namely when the planes are parallel. And if the number of equations is greater than the number of unknowns, then as a rule the system has no solution. However, in this case it is possible that the system has solutions, even infinitely many.

In order to investigate the existence and the character of the set of solutions of a system in this general setting, we turn to a "geometrical" interpretation of the system.

We interpret the system of equations

$$\begin{aligned} a_{11}x_1 &+ a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ a_{21}x_1 &+ a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 &+ a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{aligned} \tag{8}$$

in the m-dimensional space of columns in the form

$$x_1A_1 + x_2A_2 + \cdots + x_nA_n = B.$$

Here $A_1, A_2, \cdots, A_n$ denote the columns of the coefficients of the corresponding unknowns and B the column of the absolute terms.

In this interpretation the problem of the existence of a solution of the system turns into the problem of whether the given vector B is a linear combination of the vectors $A_1, A_2, \cdots, A_n$.

The answer to this problem is almost obvious. For the vector B to be a linear combination of the vectors $A_1, A_2, \cdots, A_n$ it is necessary and sufficient that B should be contained in the subspace spanned by $A_1, A_2, \cdots, A_n$ or, in other words, that the subspaces spanned by the systems of vectors $A_1, A_2, \cdots, A_n$ and $A_1, A_2, \cdots, A_n, B$ should coincide.

Since the first of these subspaces is contained in the second, they coincide if and only if their dimensions are equal. We recall that the dimension of the subspace spanned by a given system of vectors is called the rank of this system. Thus, a necessary and sufficient condition for the existence of a solution of the system $x_1A_1 + x_2A_2 + \cdots + x_nA_n = B$ is the equality of the ranks of the vector systems $A_1, A_2, \cdots, A_n$ and $A_1, A_2, \cdots, A_n, B$.

It can be proved, but we shall not do it here, that the rank of a system of vectors is equal to the rank of the matrix formed from the coordinates of these vectors. Here we understand by the *rank of a matrix* the largest order of a nonzero determinant that can be formed from the given matrix by omitting part of its rows and columns.

Since the coordinates of the vectors $A_1, A_2, \cdots, A_n$ (in the natural basis for the space of columns) are the coefficients of the system and the coordinates of the vector B its absolute terms, we obtain the following final formulation of the condition for the existence of a solution of a system.

For the existence of at least one solution of the system of linear equations

$$\begin{aligned} a_{11}x_1 &+ \cdots + a_{1n}x_n = b_1, \\ &\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 &+ \cdots + a_{mn}x_n = b_m \end{aligned}$$

it is necessary and sufficient that the rank of the matrix formed from the coefficients of the system should be equal to the rank of the matrix formed from the coefficients and the absolute terms.

Now let us investigate the character of the set of solutions if they exist. Let $x_1^0, x_2^0, \cdots, x_n^0$ be any solution of the system (8). We set $x_1 = x_1^0 + y_1$, $x_2 = x_2^0 + y_2, \cdots, x_n = x_n^0 + y_n$. Then in view of the fact that $x_1^0, x_2^0, \cdots, x_n^0$ form a solution of the system (8), the new unknowns $y_1, y_2, \cdots, y_n$ must satisfy the homogeneous system

$$\begin{aligned} a_{11}y_1 &+ \cdots + a_{1n}y_n = 0, \\ &\cdots\cdots\cdots\cdots\cdots \\ a_{m1}y_1 &+ \cdots + a_{mn}y_n = 0 \end{aligned} \tag{9}$$

with the same coefficient matrix. Conversely, if we add to the original solution $x_1^0, x_2^0, \cdots, x_n^0$ of the system (8) an arbitrary solution of the homogeneous system (9), then we obtain another solution of the system (8).

Thus, in order to obtain the general solution of the system (8), it is only necessary to take an arbitrary particular solution of it and to add it to the general solution of the homogeneous system (9).

In this way the problem of the character of the set of solutions of the system (8) is reduced to the same problem for the homogeneous system (9). We shall consider this problem in the next section.

Homogeneous systems. We shall interpret the homogeneous system of linear equations

$$\begin{aligned} a_{11}y_1 &+ a_{12}y_2 + \cdots + a_{1n}y_n = 0, \\ a_{21}y_1 &+ a_{22}y_2 + \cdots + a_{2n}y_n = 0, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}y_1 &+ a_{m2}y_2 + \cdots + a_{mn}y_n = 0 \end{aligned}$$

in the n-dimensional Euclidean space. (Here we assume that the coefficients of the system are real. For systems with complex coefficients, we can give a similar interpretation in unitary space and obtain similar results.)

Let $A_1', A_2', \cdots, A_m', Y$ be the vectors of a Euclidean space whose coordinates in an orthonormal basis are, respectively,

$$(a_{11}, a_{12}, \cdots, a_{1n}), (a_{21}, a_{22}, \cdots, a_{2n}), \cdots, (a_{m1}, a_{m2}, \cdots, a_{mn}),$$
$$(y_1, y_2, \cdots, y_n).$$

Then the system assumes the form

$$A_1'Y = 0, A_2'Y = 0, \cdots, A_m'Y = 0;$$

i.e., every solution of the system determines a vector orthogonal to all the vectors formed by the coefficients of the various equations.

Therefore, the set of solutions forms the subspace that is the orthogonal complement to the subspace spanned by the vectors $A_1', A_2', \cdots, A_m'$. The dimension of the latter subspace is equal to the rank r of the matrix formed from the coefficients of the system. The dimension of the orthogonal complement, i.e., of the "solution space," is then equal to $n - r$.

Now every subspace has a basis, i.e., a system of linearly independent vectors equal in number to the dimension of the subspace and such that their linear combinations fill the whole subspace. Therefore, among the solutions of a homogeneous system there exist $n - r$ linearly independent solutions such that all the solutions of the system are linear combinations of them. Here n denotes the number of unknowns and r the rank of the coefficient matrix.

Thus, the structure of the solutions of a homogeneous system and, consequently, also of an inhomogeneous system is completely clarified. In particular, a homogeneous system has the unique trivial solution $x_1 = x_2 = \cdots = x_n = 0$ if and only if the rank of the coefficient matrix is equal to the number of unknowns. By what we have said at the end of the preceding paragraph, the same condition is also the condition for uniqueness of the solution for systems of inhomogeneous equations (providing the consistency condition is satisfied).

Our investigation of these systems shows clearly how the introduction of generalized geometrical concepts leads to simplicity and lucidity in a complicated algebraic problem.

§4. Linear Transformations

Definition and examples. In many mathematical investigations it becomes necessary to change the variables, i.e., to go over from one system of variables $x_1, x_2, \cdots, x_n$ to another $y_1, y_2, \cdots, y_n$, connected with the first by means of a functional dependence:

$$\begin{aligned} y_1 &= \phi_1(x_1, x_2, \cdots, x_n), \\ y_2 &= \phi_2(x_1, x_2, \cdots, x_n), \\ &\cdots\cdots\cdots\cdots \\ y_n &= \phi_n(x_1, x_2, \cdots, x_n). \end{aligned}$$

For example, if the variables are the coordinates of a point in a plane or in space, then the transition from one system of coordinates to another system gives rise to a transformation of coordinates that is defined by the

expressions for the original coordinates in terms of the new ones or vice versa.

Moreover, a transformation of variables arises in studying the changes due to a transition from one position or configuration to another for objects whose position or configuration is described by the values of the variables. As a typical example of this kind of tranformation, we can take the change of coordinates of the points of some body under deformations.

An abstractly given transformation of a system of n variables is usually interpreted as a transformation (deformation) of an n-dimensional space, i.e., as an association between each vector of the space (or part of it) with coordinates $x_1, x_2, \cdots, x_n$ and a corresponding vector with coordinates $y_1, y_2, \cdots, y_n$.

As we have said previously, every "smooth" function (having continuous partial derivatives) of several variables is close to a linear function for small changes of these variables. Therefore every "smooth" transformation (i.e., one for which the functions $\phi_1, \phi_2, \cdots, \phi_n$ in its analytical expression have continuous partial derivatives) is close to a linear transformation in a small part of the space:

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + b_1, \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + b_2, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_n &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n + b_n. \end{aligned} \tag{10}$$

This circumstance alone makes the study of the properties of linear transformations one of the most important problems of mathematics. For example, from the theory of n linear equations in n unknowns, we know that a necessary and sufficient condition for the system of equations (10) with respect to $x_1, x_2, \cdots, x_n$ to have a unique solution, i.e., for the corresponding linear transformation to be invertible, is the nonvanishing of the determinant of the coefficients. This circumstance is the foundation of an important theorem of analysis: For a transformation

$$\begin{aligned} y_1 &= \phi_1(x_1, x_2, \cdots, x_n), \\ y_2 &= \phi_2(x_1, x_2, \cdots, x_n), \\ &\cdots\cdots\cdots\cdots \\ y_n &= \phi_n(x_1, x_2, \cdots, x_n), \end{aligned}$$

which is smooth in the neighborhood of a given point, to have a smooth

inverse transformation it is necessary and sufficient that at the given point the determinant

$$\begin{vmatrix} \frac{\partial \phi_1}{\partial x_1} & \cdots & \frac{\partial \phi_1}{\partial x_n} \\ \frac{\partial \phi_2}{\partial x_1} & \cdots & \frac{\partial \phi_2}{\partial x_n} \\ \cdots\cdots & & \\ \frac{\partial \phi_n}{\partial x_1} & \cdots & \frac{\partial \phi_n}{\partial x_n} \end{vmatrix}$$

should be different from zero.

The study of the general linear transformation (10) essentially reduces to the study of the homogeneous transformation with the same coefficients

$$\begin{aligned} y_1 &= a_{11}x_1 + \cdots + a_{1n}x_n , \\ y_2 &= a_{21}x_1 + \cdots + a_{2n}x_n , \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ y_n &= a_{n1}x_1 + \cdots + a_{nn}x_n , \end{aligned} \tag{11}$$

and in what follows, when speaking of linear tranformations, we shall always have homogeneous transformations in mind.

Linear transformations of an n-dimensional space can also be defined by their intrinsic properties, apart from the formulas (11) that connect the coordinates of corresponding points. Such a coordinate-free definition of the concept of a linear transformation is useful in that it does not depend on the choice of a basis. This definition is as follows.

A linear transformation of an n-dimensional linear space is a function $Y = A(X)$ where X and Y are vectors. This function satisfies the postulate of linearity

$$A(c_1X_1 + c_2X_2) = c_1A(X_1) + c_2A(X_2). \tag{12}$$

In what follows, when speaking of a linear transformation of a space, we shall understand it in the sense of this definition.

This definition is equivalent to the preceding one in terms of coordinates. For the function $Y = A(X)$, which to the vector X with the coordinates $x_1, x_2, \cdots, x_n$ associates the vector Y with the coordinates $y_1, y_2, \cdots, y_n$ in such a way that the coordinates $y_1, y_2, \cdots, y_n$ are expressed in terms of the coordinates $x_1, x_2, \cdots, x_n$ in the form of linear homogeneous functions, obviously satisfies the postulate (12). Conversely, if the function $Y = A(X)$ satisfies the postulate (12) and if $e_1, e_2, \cdots, e_n$ is an arbitrary basis of the space, then

$$A(x_1e_1 + x_2e_2 + \cdots + x_ne_n) = x_1A(e_1) + x_2A(e_2) + \cdots + x_nA(e_n).$$

We denote the coordinates (in the same basis) of the vector $A(e_j)$ by $a_{1j}, \cdots, a_{nj}; j = 1, \cdots, n$. Then the coordinates of the vector $Y = A(X)$ are

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n , \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n , \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ y_n &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n . \end{aligned}$$

Thus, to every linear transformation of a linear space there corresponds a certain square matrix with respect to a given basis. This transformation can be written in matrix language in the form $Y = AX$. Here X is the column of the coordinates of the original vector, Y the column of the coordinates of the transformed vector, and A the coefficient matrix of the transformation. The columns of the matrix A are formed by the coordinates of those vectors into which the vectors of the basis are transformed. In accordance with the matrix notation, we shall subsequently often write a linear transformation itself in the form $Y = AX$, omitting the parentheses.

From the formula

$$A(x_1e_1 + x_2e_2 + \cdots + x_ne_n) = x_1A(e_1) + x_2A(e_2) + \cdots + x_nA(e_n)$$

it follows that the whole space is mapped under a linear transformation into the subspace spanned by the vectors $A(e_1), \cdots, A(e_n)$. The dimension of this subspace is equal to the rank of the system of vectors $A(e_1)$, $A(e_2), \cdots, A(e_n)$, or, what is the same, to the rank of the matrix formed by their coordinates, i.e., the rank of the matrix A associated with the transformation. This subspace coincides with the whole space if and only if the rank of the matrix A is equal to n, i.e., if the determinant of A is different from zero. In this case the linear transformation is called *nonsingular* or *nondegenerate*.

From the theory of systems of linear equations, we know that nondegenerate transformations are uniquely invertible and that the coordinates of the original vector are expressed in terms of the coordinates of the transformed vector by the formula $\mathbf{X} = \mathbf{A}^{-1}\mathbf{Y}$.

A transformation whose matrix has the determinant zero is called *singular* or *degenerate*. A degenerate transformation is not invertible. This follows from the theory of linear transformations or more intuitively from the fact that it transforms the whole space into part of it.

As a first example of a nondegenerate transformation, we take the identity transformation that maps every vector into itself. The matrix of the identity transformation in any basis is the unit matrix E. A nonsingular transformation is also given by a similarity that consists in multiplying

all the vectors of the space by one and the same number. The matrix of a similarity transformation does not depend on the choice of a basis and has the form aE, where a is the similarity factor.

An important special case of nondegenerate transformation are the orthogonal transformations. The concept of an *orthogonal transformation* has a meaning when applied to a Euclidean space and is defined as a linear transformation preserving the lengths of vectors. An orthogonal transformation is a generalization to n-dimensional space of a rotation of the space around the fixed origin of coordinates or a rotation combined with a reflection in an arbitrary plane passing through the origin.

It is easy to see that under an orthogonal transformation not only the lengths of vectors are preserved but also scalar products and that, consequently, orthogonal transformations carry an orthogonal basis of the space into a system of pairwise orthogonal unit vectors which in turn is then necessarily also a basis.

The matrix connected with an orthogonal transformation with respect to an orthonormal basis has the following specific properties.

First, the sum of the squares of the elements of each column is 1, since these sums are the squares of the lengths of the vectors into which the vectors of the given basis are transformed. Second, the sums of the products of corresponding elements taken from distinct columns are zero, since these sums are the scalar products of the vectors into which the vectors of the basis are transformed.

In matrix notation both these properties can be written by the single formula

$$\bar{P}P = E.$$

Here P is the matrix of the orthogonal transformation (with respect to an orthonormal basis), and $\bar{P}$ is its transposed matrix, i.e., the matrix whose rows are the columns of P in the same order.

For the diagonal elements of the matrix $\bar{P}P$ are by the rule for the multiplication of matrices equal to the sum of the squares of the elements of the corresponding column of P, and the nondiagonal elements are equal to the sum of the products of the corresponding elements taken from distinct columns of P.

As an example for a degenerate transformation, we can take the orthogonal projection of all vectors of a Euclidean space onto some subspace (see §2). For in this transformation the whole space is mapped onto part of it.

Transformation of coordinates. We now consider the problem of the transformation of coordinates in n-dimensional space, i.e., the problem

how the coordinates of vectors are changed on transition from one basis to another.

Let the original basis be $e_1, e_2, \cdots, e_n$ and let $f_1, f_2, \cdots, f_n$ be an arbitrary other basis of the space. Suppose further that

$$C = \begin{bmatrix} c_{11}, \cdots, c_{1n} \\ \cdots\cdots\cdots \\ c_{n1}, \cdots, c_{nn} \end{bmatrix}$$

is the matrix whose columns are the coordinates of the vectors of the new basis $f_1, f_2, \cdots, f_n$ with respect to the original one. The matrix C is obviously nondegenerate, because the vectors $f_1, f_2, \cdots, f_n$ are linearly independent. It is called the *matrix of the coordinate transformation*.

We denote by $x_1, x_2, \cdots, x_n$ the coordinates of a certain vector X with respect to the basis $e_1, e_2, \cdots, e_n$ and by $x'_1, x'_2, \ldots, x'_n$ the coordinates of the same vector with respect to the basis $f_1, f_2, \cdots, f_n$. Then $X = x'_1 f_1 + x'_2 f_2 + \cdots + x'_n f_n$ and therefore the coordinates of the vector X with respect to the original basis form the column

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} c_{11}x'_1 + c_{12}x'_2 + \cdots + c_{1n}x'_n \\ c_{21}x'_1 + c_{22}x'_2 + \cdots + c_{2n}x'_n \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ c_{n1}x'_1 + c_{n2}x'_2 + \cdots + c_{nn}x'_n \end{bmatrix} = \begin{bmatrix} c_{11} \cdots c_{1n} \\ c_{21} \cdots c_{2n} \\ \cdots\cdots\cdots \\ c_{n1} \cdots c_{nn} \end{bmatrix} \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{bmatrix}.$$

Thus, the original coordinates are expressed linearly and homogeneously in terms of the transformation with the matrix C.

The formulas that express the connection between the coordinates with respect to the original and the transformed basis coincide formally with the formulas that link the coordinates of corresponding vectors in a nondegenerate linear transformation of the space. This circumstance enables us to interpret an abstractly given linear homogeneous transformation of variables with a nondegenerate matrix either as a transformation of coordinates or as a transformation of the space. In each concrete case the choice of one of these two interpretations is determined by the context of the problem under consideration.

Let us now deal with the question how the matrix of a linear transformation of the space is changed under a coordinate transformation.

Suppose that the given linear transformation has the matrix A in the basis $e_1, e_2, \cdots, e_n$ so that the column Y of the coordinates of the transformed vector is linked with the column X of the original one by the formula

$$Y = AX.$$

Suppose now that a transformation of coordinates with the matrix C is made; X', Y' denote, respectively, the columns of the coordinates of the original and the transformed vectors with respect to the new basis. Then $X = CX'$, $Y = CY'$ and hence

$$Y' = C^{-1}Y = C^{-1}AX = C^{-1}ACX'.$$

Thus, the matrix of our transformation with respect to the new basis is $C^{-1}AC$.

Two matrices A and B connected by the relation $B = C^{-1}AC$, where C is a nonsingular matrix, are called *similar*. One and the same linear transformation corresponds with respect to various bases to a class of pairwise similar matrices.

Eigenvectors and eigenvalues of a linear transformation. A very important class of linear transformations consists of the transformations that come about in the following way.

Let $e_1, e_2, \cdots, e_n$ be arbitrary linearly independent vectors of the space. Suppose that under the transformation they are multiplied by certain numbers $\lambda_1, \lambda_2, \cdots, \lambda_n$. If the vectors $e_1, e_2, \cdots, e_n$ are taken as a basis of the space, then the transformation can be described by

$$\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

The transformations of this class have a simple and intuitive geometrical meaning (of course, only for real spaces and for $n = 2$ or $n = 3$). Namely, if all the numbers λ_i are positive, then the transformation that we describe consists in a stretching (or compressing) of the space in the directions of the vectors $e_1, e_2, \cdots, e_n$ with coefficients $\lambda_1, \lambda_2, \cdots, \lambda_n$. If some of the λ_i are negative, then the deformation of the space is accompanied by a change of direction of some of the vectors $e_1, e_2, \cdots, e_n$ into the opposite. Finally, if for example $\lambda_1 = 0$, then a projection of the space parallel to e_1 takes place onto the subspace spanned by $e_2, \cdots, e_n$ with a subsequent deformation in these directions.

The class of transformations we have considered is important, because in spite of its simplicity it is very general. In fact, it can be established that every linear transformation satisfying certain not very severe restrictions belongs to our class; i.e., we can find for it a basis in which it is described by a diagonal matrix.

The restrictions to be imposed on the transformation become particularly

clear if we consider linear transformation of a complex space. In what follows this will be assumed.

We introduce the following definition.

A nonzero vector X which under a linear transformation A of the space goes into a collinear vector λX is called an ***eigenvector of the transformation***. In other words, a nonzero vector X is an eigenvector of the transformation A if and only if $AX = \lambda X$. The number λ is called an ***eigenvalue*** of the transformation A.

It is obvious that if in some basis a transformation has a diagonal matrix, then this basis consists of eigenvectors and the diagonal elements are eigenvalues. Conversely, if there exists in the space a basis consisting of eigenvectors of the transformation A, then in this basis the matrix of the transformation A is diagonal and consists of the eigenvalues corresponding to the vectors of the basis.

We now proceed to study the properties of eigenvectors and eigenvalues. With this aim we write the definition of an eigenvector in coordinate notation. Let A be the matrix corresponding to the transformation A with respect to a certain basis and X the column of coordinates of the vector X in the same basis. The equation $AX = \lambda X$ in coordinate notation is written as $AX = \lambda X$ or

$$(A - \lambda E)X = 0.$$

In expanded form this equation turns into the system

$$\begin{aligned} (a_{11} - \lambda)x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0, \\ a_{21}x_1 + (a_{22} - \lambda)x_2 + \cdots + a_{2n}x_n &= 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + (a_{nn} - \lambda)x_n &= 0. \end{aligned}$$

We can regard this system of equations as a system of linear and homogeneous equations for $x_1, x_2, \cdots, x_n$. We are interested in the case when this system has a nontrivial solution, because the coordinates of an eigenvector must not all be equal to zero. Now we know that a necessary and sufficient condition for the existence of a nontrivial solution of a system of linear homogeneous equations is that the rank of the coefficient matrix should be less than the number of unknowns, and this is equivalent to the vanishing of the determinant of the system

$$\begin{vmatrix} a_{11} - \lambda & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} = 0.$$

Thus, all eigenvalues of the transformation A are roots of the polynomial $|A - \lambda E|$ and, conversely, every root of this polynomial is an eigenvalue of the transformation since to every root there corresponds at least one eigenvector. The polynomial $|A - \lambda E|$ is called the *characteristic polynomial* of the matrix A. The equation $|A - \lambda E| = 0$ is called the *characteristic* or *secular equation* and its roots *characteristic numbers* of the matrix.*

By the fundamental theorem of higher algebra (Chapter IV), every polynomial has at least one root; therefore every linear transformation has at least one eigenvalue and hence at least one eigenvector. But, of course, it is quite possible that even in the case when the transformation can be expressed by a real matrix it turns out that all or some of its eigenvalues are complex. Consequently, the theorem on the existence of (real) eigenvalues and eigenvectors for an arbitrary linear transformation is not true in a real space. For example, the transformation of the plane that consists in a rotation around the origin of coordinates by any angle other than 180° changes the directions of all the vectors of the plane so that there are no eigenvectors for this transformation.

The roots of the characteristic polynomial of a matrix A are the eigenvalues of the transformation A; therefore matrices that correspond to one and the same transformation in distinct bases have identical sets of roots of the characteristic polynomial. This leads to the plausible asssertion that the characteristic polynomial of a linear transformation also depends on the transformation only and not on the choice of a basis. This can be verified by the following elegant calculation, which is based on the properties of operations on matrices and determinants.

We know that if a matrix A corresponds to a transformation A in some basis, then in any other basis the transformation A has a similar matrix $C^{-1}AC$, where C is some nonsingular matrix. But

$$|C^{-1}AC - \lambda E| = |C^{-1}AC - C^{-1}\lambda EC| = |C^{-1}(A - \lambda E)C|$$
$$= |C^{-1}||C||A - \lambda E| = |C^{-1}C||A - \lambda E| = |A - \lambda E|.$$

Thus, matrices corresponding to one and the same transformation A in distinct bases have in fact one and the same characteristic polynomial, which can therefore be called the polynomial of the transformation.

We shall now make the assumption that all the eigenvalues of the transformation A are distinct. Let us prove that the eigenvectors, one for each eigenvalue, are linearly independent. For if we suppose that some

* The name "secular equation" has arisen in celestial mechanics in connection with the problem of the so-called secular disturbances in the motions of the planets.

of them, say $e_1, \cdots, e_k$, are linearly independent and the remaining ones, among them e_{k+1}, are linear combinations of these, then

$$e_{k+1} = c_1e_1 + c_2e_2 + \cdots + c_ke_k. \tag{13}$$

When we apply the linear transformation to both sides of this equation, we obtain

$$Ae_{k+1} = c_1Ae_1 + c_2Ae_2 + \cdots + c_kAe_k,$$

from which it follows by the definition of an eigenvector that

$$\lambda_{k+1}e_{k+1} = c_1\lambda_1e_1 + c_2\lambda_2e_2 + \cdots + c_k\lambda_ke_k.$$

When we multiply equation (13) by λ_{k+1} and subtract from it the equation just obtained, we have

$$c_1(\lambda_{k+1} - \lambda_1)e_1 + c_2(\lambda_{k+1} - \lambda_2)e_2 + \cdots + c_k(\lambda_{k+1} - \lambda_k)e_k = 0,$$

hence it follows by the linear independence of $e_1, e_2, \cdots, e_k$ that

$$c_1(\lambda_{k+1} - \lambda_1) = c_2(\lambda_{k+1} - \lambda_2) = \cdots = c_k(\lambda_{k+1} - \lambda_k) = 0.$$

But we had assumed that all the eigenvalues are distinct and the vectors are chosen one for each eigenvalue. Therefore $\lambda_{k+1} - \lambda_1 \neq 0$, $\lambda_{k+1} - \lambda_2 \neq 0$, $\cdots$, $\lambda_{k+1} - \lambda_k \neq 0$ and the equation (13) is impossible, since the coefficients $c_1, c_2, \cdots, c_k$ cannot all be zero.

Now it is clear that if all the eigenvalues of a linear transformation are distinct, then there exists a basis in which the matrix of the transformation has diagonal form. For we can choose as such a basis a system of eigenvectors, one for each eigenvalue. As we have shown, they are linearly independent and their number is equal to the dimension of the space; i.e., they do in fact form a basis.

The theorem we have proved can be stated in terms of the theory of matrices as follows. If all the eigenvalues of a matrix are distinct, then the matrix is similar to the diagonal matrix whose diagonal elements are these eigenvalues.

The problem of transforming the matrix of a linear transformation to its simplest form is considerably more complicated if there are equal ones among the roots of the characteristic polynomial. We shall confine ourselves to a short account of the final result.

A "canonical box" of order m is defined as a matrix of the form

$$I_{m,\lambda_i} = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{bmatrix}.$$

All the unnamed elements are equal to zero.

A canonical Jordan matrix is defined as a matrix in which there are "canonical boxes" along the main diagonal and all the remaining elements are zero:

$$\begin{bmatrix} \boxed{I_{m_1,\lambda_1}} & & & \\ & \boxed{I_{m_2,\lambda_2}} & & \\ & & \ddots & \\ & & & \boxed{I_{m_k,\lambda_k}} \end{bmatrix}.$$

The numbers λ_i in the distinct "boxes" are not necessarily pairwise distinct. Every matrix can be reduced to a canonical Jordan matrix similar to it. The proof of this theorem is rather complicated. We ought to mention that this theorem plays an important role in many applications of algebra to other problems of mathematics, in particular in the theory of systems of linear differential equations.

A matrix can be reduced to diagonal form if and only if the orders m_i of all boxes are equal to 1.

§5. Quadratic Forms

Definition and simplest properties. A *quadratic form* is a homogeneous polynomial of degree 2 in several variables.

A quadratic form in n variables $x_1, x_2, \cdots, x_n$ consists of terms of two types: squares of the variables and products of two variables, both with certain coefficients. A quadratic form can be written in the following quadratic scheme:

$$\begin{aligned} f(x_1, x_2, \cdots, x_n) = \quad & a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1n}x_1x_n \\ & + a_{21}x_2x_1 + a_{22}x_2^2 + \cdots + a_{2n}x_2x_n \\ & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ & + a_{n1}x_nx_1 + a_{n2}x_nx_2 + \cdots + a_{nn}x_n^2. \end{aligned}$$

Pairs of similar terms $a_{12}x_1x_2$ and $a_{21}x_2x_1$, etc., are written with equal coefficients so that each of them gets half the coefficient of the corresponding product of the variables. Thus, every quadratic form is uniquely connected with its coefficient matrix, which is symmetric.

A quadratic form can conveniently be represented in the following

matrix notation. We denote by X the column of the variables $x_1, x_2, \cdots, x_n$ by $\tilde{X}$ the row $(x_1, x_2, \cdots, x_n)$, i.e., the transposed matrix of X. Then

$$\begin{aligned} f(x_1, x_2, \cdots, x_n) = \quad & x_1(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n) \\ & + x_2(a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n) + \cdots \\ & + x_n(a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n) \\ = (x_1, x_2, \cdots, x_n) & \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{bmatrix} \\ = (x_1, x_2, \cdots, x_n) & \begin{bmatrix} a_{11}a_{12} \cdots a_{1n} \\ a_{21}a_{22} \cdots a_{2n} \\ \cdots\cdots\cdots\cdots \\ a_{n1}a_{n2} \cdots a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \tilde{X}AX. \end{aligned}$$

Quadratic forms occur in many branches of mathematics and its applications.

In the theory of numbers and in crystallography, we consider quadratic forms under the proviso that the variables $x_1, x_2, \cdots, x_n$ assume only integral values. In analytical geometry a quadratic form arises in setting up the equation of a curve (or surface) of the second order. In mechanics and physics a quadratic form appears in the expression for the kinetic energy of a system in terms of the components of the generalized velocities, etc. Furthermore, a study of quadratic forms is necessary even in analysis for the investigation of functions of several variables in connection with problems where it is important to clarify how a given function in the neighborhood of a given point deviates from its approximating linear function. An example of problems of this type is the investigation of maxima and minima of a function.

Let us consider, for example, the problem of finding the maxima and minima of a function of two variables $w = f(x, y)$ having continuous partial derivatives of the third order. A necessary condition for the point (x_0, y_0) to give a maximum or minimum of the function w is that the partial derivatives of the first order should vanish at the point (x_0, y_0). Let us assume that this condition is satisfied. We give to the variables x and y small increments h and k and consider the corresponding increment of the function $\Delta w = f(x_0 + h, y_0 + k) - f(x_0, y_0)$. By Taylor's formula this increment is equal, to within terms of higher order of smallness, to the quadratic form $\frac{1}{2}(rh^2 + 2shk + tk^2)$, where r, s and t are the values of the second derivatives $\partial^2 w/\partial x^2$, $\partial^2 w/\partial x \partial y$, $\partial^2 w/\partial y^2$, computed at the point (x_0, y_0). If this quadratic form is positive for all values of h and k (except

$h = k = 0$), then the function w has a minimum at the point (x_0, y_0); if it is negative, then a maximum. Finally, if the form assumes positive as well as negative values, then there is neither a maximum nor a minimum. Functions with a larger number of variables can be investigated in a similar way.

The study of quadratic forms essentially consists in the investigation of the problem of the equivalence of forms under one set or another of linear transformations of the variables. Two quadratic forms are called *equivalent* if one of them can be carried into the other by one of the transformations of the given set. Closely connected with the problem of equivalence is the problem of *reduction* of a form, i.e., its transformation into a certain form, as simple as possible.

In various problems connected with quadratic forms, we consider various sets of admissible transformations of the variables.

In problems of analysis arbitrary nonsingular transformations of the variables are admitted; for the purposes of analytical geometry, the greatest interest lies in orthogonal transformations, i.e., those that correspond to the transition from one system of variable Cartesian coordinates to another. Finally, in the theory of numbers and in crystallography, we consider linear transformations with integer coefficients and with a determinant equal to 1.

Here we shall consider two of these problems: the problem of the reduction of a quadratic form to the simplest possible form by means of arbitrary nonsingular transformations and the same problem for orthogonal transformations. Above all, we shall find out how the matrix of the quadratic form is changed under a linear transformation of the variables.

Suppose that $f(x_1, x_2, \cdots, x_n) = \tilde{X}AX$, where A is the symmetric matrix of the coefficients of the form and X is the column of the variables.

We take a linear transformation of the variables, writing it briefly $X = CX'$. Here C denotes the coefficient matrix of this transformation and X' the column of the new variables. Then $\tilde{X} = \tilde{X}'\tilde{C}$ and consequently $\tilde{X}AX = \tilde{X}'(\tilde{C}AC)X'$ so that the matrix of the transformed quadratic form is $\tilde{C}AC$.

The matrix $\tilde{C}AC$ is automatically symmetric, as is easy to verify. Thus, the problem of reducing a quadratic form to the simplest possible form is equivalent to the task of reducing a symmetric matrix to the simplest form by multiplying it on the left and on the right by transposed matrices.

Transformation of a quadratic form to the canonical form by successive completion of squares. We shall show that every (real) quadratic form can be reduced to a sum of squares of new variables with certain coefficients by a real nonsingular linear transformation.

To prove this we shall show first of all that if the form is not identically zero, then we can make the coefficient of the square of the first variable different from zero by the application of a nonsingular transformation of the variables.

For suppose that

$$\begin{aligned} f(x_1, x_2, \cdots, x_n) = {} & a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1n}x_1x_n \\ & + a_{21}x_2x_1 + a_{22}x_2^2 + \cdots + a_{2n}x_2x_n \\ & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ & + a_{n1}x_nx_1 + a_{n2}x_nx_2 + \cdots + a_{nn}x_n^2. \end{aligned}$$

If $a_{11} \neq 0$, then no transformation is required. If $a_{11} = 0$, but some one of the diagonal coefficients $a_{kk} \neq 0$, then we set $x_1 = x_k'$, $x_k = x_1'$ equating the remaining original variables to the corresponding new ones. This nonsingular transformation achieves our object. Finally, if all the diagonal coefficients are equal to zero, then at least one of the nondiagonal coefficients is different from zero, for example a_{12}.

By taking the nonsingular transformation

$$\begin{aligned} x_1 &= x_1', \\ x_2 &= x_1' + x_2' \end{aligned}$$

and equating the remaining original variables to the new ones, we achieve our aim.

Thus, without loss of generality we can assume that $a_{11} \neq 0$.

Let us now separate out the square of a linear function such that all the terms containing x_1 occur in this square.

This can easily be done. For

$$\begin{aligned} f(x_1, x_2, \cdots, x_n) = {} & a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1n}x_1x_n \\ & + a_{21}x_2x_1 + a_{22}x_2^2 + \cdots + a_{2n}x_2x_n \\ & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ & + a_{n1}x_nx_1 + a_{n2}x_nx_2 + \cdots + a_{nn}x_n^2 \\ = {} & a_{11}\left[x_1 + \frac{a_{12}}{a_{11}}x_2 + \cdots + \frac{a_{1n}}{a_{11}}x_n\right]^2 - a_{11}\left[\frac{a_{12}}{a_{11}}x_2 + \cdots + \frac{a_{1n}}{a_{11}}x_n\right]^2 \\ & + a_{22}x_2^2 + \cdots + a_{2n}x_2x_n \\ & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ & + a_{n2}x_nx_2 + \cdots + a_{nn}x_n^2. \end{aligned}$$

By removing the parentheses in the second summand and collecting similar terms, we obtain

$$f(x_1, x_2, \cdots, x_n) = a_{11}\left[x_1 + \frac{a_{12}}{a_{11}}x_2 + \cdots + \frac{a_{1n}}{a_{11}}x_n\right]^2 + f_1(x_2, \cdots, x_n),$$

where f_1 is a form in $n-1$ variables.

The transformation

$$\begin{aligned} x_1 + \frac{a_{12}}{a_{11}}x_2 + \cdots + \frac{a_{1n}}{a_{11}}x_n &= x_1', \\ x_2 &= x_2', \\ \cdots\cdots\cdots\cdots\cdots & \\ x_n &= x_n', \end{aligned}$$

is obviously nonsingular. By making this transformation we reduce our form to

$$a_{11}x_1'^2 + f_1(x_2', \cdots, x_n').$$

Continuing the process in a similar manner we reduce the form to the required "canonical" form

$$\alpha_1 z_1^2 + \alpha_2 z_2^2 + \cdots + \alpha_n z_n^2.$$

Here $z_1, z_2, \cdots, z_n$ are the new variables introduced at the last step.

The law of inertia of quadratic forms. In the reduction of a quadratic form to canonical form there always is a considerable arbitrariness in the choice of the transformation of variables that brings about this reduction. This arbitrariness comes in, among other things, from the fact that we can precede the previous method of successive separation of squares by an arbitrary nonsingular transformation of the variables.

However, notwithstanding this arbitrariness, as a result of the reduction of the given form we obtain an almost unique canonical quadratic form independent of the choice of the reducing transformation. Namely the number of squares of the new variables that occur with positive, negative, and zero coefficients is always one and the same, irrespective of the method of reduction. This theorem is known as the *law of inertia* of quadratic forms. We shall not prove it here.

The law of interia of a quadratic form solves the problem of equivalence of a real quadratic form under all nonsingular transformations. For two forms are equivalent if and only if their reduction to canonical form leads to forms with the same number of squares with positive, negative, and zero coefficients.

Of special interest for the applications are the quadratic forms that under reduction to canonical form turn into a sum of squares of the new variables with all the coefficients *positive*. Such forms are called *positive definite*.

Positive-definite quadratic forms are characterized by the property that their values for real values of the variables not all equal to zero are always positive.

Orthogonal transformation of quadratic forms to canonical form. Of special interest among all possible methods of reducing a quadratic form to canonical form are the orthogonal transformations, i.e., those that can be obtained by a linear transformation of the variables with an orthogonal matrix. Such transformations are of interest, for example, in analytical geometry, in the problem of reducing the general equation of a curve or surface of the second order to canonical form.

In order to convince ourselves of the possibility of such a transformation, it is convenient to regard the quadratic form as a function of vectors in a Euclidean space by considering the variables $x_1, x_2, \cdots, x_n$ as the coordinates of a variable vector in an orthonormal basis. Then an orthogonal transformation of the variables can be interpreted as a transition from one orthonormal basis to another.

With the quadratic form

$$\begin{aligned} f(x_1, x_2, \cdots, x_n) = {} & a_{11}x_1^2 + \cdots + a_{1n}x_1x_n \\ & \cdots\cdots\cdots\cdots\cdots\cdots \\ & + a_{n1}x_nx_1 + \cdots + a_{nn}x_n^2 \end{aligned}$$

we connect the linear transformation A that has with respect to the chosen basis the matrix

$$A = \begin{bmatrix} a_{11} \cdots a_{1n} \\ \cdots\cdots \\ a_{n1} \cdots a_{nn} \end{bmatrix}.$$

Then our quadratic form can be regarded as the scalar product $AX \cdot X$ (where X is the vector with the coordinates $x_1, x_2, \cdots, x_n$), and its coefficients a_{ij} are the scalar product $Ae_i \cdot e_j$, where $e_1, e_2, \cdots, e_n$ is the chosen orthonormal basis.

It is easy to see that in consequence of the symmetry of the matrix A the following equation holds for arbitrary vectors X and Y

$$AX \cdot Y = X \cdot AY.$$

We shall show first of all that the transformation A has at least one real eigenvalue and eigenvector corresponding to it.

For this purpose we consider the values of the form $AX \cdot X$ under the assumption that the vector X ranges over the unit sphere, i.e., the set of all unit vectors. Under these conditions the form $AX \cdot X$ will have a maximum. Let us show that this maximum $AX \cdot X$ is an eigenvalue of the tranformation A and the vector X_0 for which this maximum is assumed is a corresponding eigenvector; i.e., $AX_0 = \lambda_1 X_0$.

The proof of this statement is by an indirect method, namely, by establishing that the vector AX_0 is orthogonal to all vectors orthogonal to X_0.

We note that for an arbitrary vector Z we have the inequality $AZ \cdot Z \leqslant \lambda_1 | Z |^2$. This is obvious from the fact that $X = Z/| Z |$ is a unit vector and λ_1 the maximum of the values of the form $AX \cdot X$ on the unit sphere. We consider $Z = X_0 + \epsilon Y$, where ϵ is a real number and Y an arbitrary vector orthogonal to the vector X_0. Then

$$AZ \cdot Z = (AX_0 + \epsilon AY) \cdot (X_0 + \epsilon Y) = AX_0 \cdot X_0 + 2\epsilon AX_0 \cdot Y + \epsilon^2 AY \cdot Y$$
$$= \lambda_1 + 2\epsilon AX_0 \cdot Y + \epsilon^2 AY \cdot Y.$$

Moreover,

$$| Z |^2 = (X_0 + \epsilon Y) \cdot (X_0 + \epsilon Y) = | X_0 |^2 + \epsilon^2 | Y |^2 = 1 + \epsilon^2 | Y |^2,$$

because

$$X_0 \cdot Y = 0, \quad | X_0 |^2 = 1.$$

Therefore,

$$\lambda_1 + 2\epsilon AX_0 \cdot Y + \epsilon^2 AY \cdot Y \leqslant \lambda_1 + \epsilon^2 \lambda_1 | Y |^2,$$

and from this we obtain on dividing by ϵ^2

$$\frac{2}{\epsilon} AX_0 \cdot Y < \lambda_1 | Y |^2 - AY \cdot Y. \tag{14}$$

This last inequality must be satisfied for arbitrary real ϵ of sufficiently small absolute value.

But it can only be satisfied under the condition $AX_0 \cdot Y = 0$, because if $AX_0 \cdot Y > 0$, then the inequality (14) is impossible for sufficiently small positive ϵ, and if $AX_0 \cdot Y < 0$, then it is impossible for negative ϵ of sufficiently small absolute value. Thus, $AX_0 \cdot Y = 0$, i.e., AX_0 is in fact orthogonal to every vector orthogonal to X_0. Therefore AX_0 and X_0 are collinear; i.e., $AX_0 = \lambda' X_0$, where λ' is a real number. The fact that $\lambda' = \lambda_1$ is easy to verify, for

$$\lambda_1 = AX_0 \cdot X_0 = \lambda' X_0 \cdot X_0 = \lambda'.$$

Now it is easy to show that every quadratic form can in fact be reduced to canonical form by an orthogonal transformation.

Let $e_1, e_2, \cdots, e_n$ be the original orthonormal basis of the space and $f_1, f_2, \cdots, f_n$ a new orthonormal basis in which the first vector f_1 is an eigenvector X_0 of the transformation A. Let $x_1, x_2, \cdots, x_n$ be the coordinates of the vector X in the original basis and $x_1', x_2', \cdots, x_n'$ its coordinates in the new basis. Then

$$\begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = P \begin{bmatrix} x_1' \\ x_2' \\ \cdot \\ \cdot \\ \cdot \\ x_n' \end{bmatrix}$$

where P is an orthogonal matrix.

Let us carry out the transition to the new variables in the quadratic form $AX \cdot X$. In the new variables the quadratic form has the coefficients $a_{ij}' = Af_i \cdot f_j$. Therefore

$$a_{11}' = Af_1 \cdot f_1 = \lambda_1 f_1 \cdot f_1 = \lambda_1,$$
$$a_{1j}' = a_{j1}' = Af_1 \cdot f_j = \lambda_1 f_1 \cdot f_j = 0 \quad \text{for} \quad j \neq 1;$$

i.e., the form is now

$$\lambda_1 x_1'^2 + \phi(x_2', \cdots, x_n').$$

Thus, by means of an orthogonal transformation we have succeeded in separating out one square of a new variable.

By repeating the same arguments with the new form $\phi(x_2', \cdots, x_n')$, etc., we eventually arrive at the conclusion that the form turns out to be reduced to canonical form by a chain of orthogonal transformations. But it is obvious that a chain of orthogonal transformations is equivalent to a single orthogonal transformation. This concludes the proof of the theorem.

§6. Functions of Matrices and Some of Their Applications

Functions of matrices. The applications of linear algebra to other branches of mathematics are very numerous and diverse. It is not an exaggeration to say that the ideas and results of linear algebra are used in a large part of contemporary mathematics and theoretical physics in one form or another, particularly in the form of matrix calculus.

We shall deal briefly with one of the methods of applying the matrix calculus to the theory of ordinary differential equations. Here functions of matrices play an important role.

For this purpose we consider the values of the form $AX \cdot X$ under the assumption that the vector X ranges over the unit sphere, i.e., the set of all unit vectors. Under these conditions the form $AX \cdot X$ will have a maximum. Let us show that this maximum $AX \cdot X$ is an eigenvalue of the tranformation A and the vector X_0 for which this maximum is assumed is a corresponding eigenvector; i.e., $AX_0 = \lambda_1 X_0$.

The proof of this statement is by an indirect method, namely, by establishing that the vector AX_0 is orthogonal to all vectors orthogonal to X_0.

We note that for an arbitrary vector Z we have the inequality $AZ \cdot Z \leqslant \lambda_1 |Z|^2$. This is obvious from the fact that $X = Z/|Z|$ is a unit vector and λ_1 the maximum of the values of the form $AX \cdot X$ on the unit sphere. We consider $Z = X_0 + \epsilon Y$, where ϵ is a real number and Y an arbitrary vector orthogonal to the vector X_0. Then

$$AZ \cdot Z = (AX_0 + \epsilon AY) \cdot (X_0 + \epsilon Y) = AX_0 \cdot X_0 + 2\epsilon AX_0 \cdot Y + \epsilon^2 AY \cdot Y$$
$$= \lambda_1 + 2\epsilon AX_0 \cdot Y + \epsilon^2 AY \cdot Y.$$

Moreover,

$$|Z|^2 = (X_0 + \epsilon Y) \cdot (X_0 + \epsilon Y) = |X_0|^2 + \epsilon^2 |Y|^2 = 1 + \epsilon^2 |Y|^2,$$

because

$$X_0 \cdot Y = 0, \quad |X_0|^2 = 1.$$

Therefore,

$$\lambda_1 + 2\epsilon AX_0 \cdot Y + \epsilon^2 AY \cdot Y \leqslant \lambda_1 + \epsilon^2 \lambda_1 |Y|^2,$$

and from this we obtain on dividing by ϵ^2

$$\frac{2}{\epsilon} AX_0 \cdot Y < \lambda_1 |Y|^2 - AY \cdot Y. \tag{14}$$

This last inequality must be satisfied for arbitrary real ϵ of sufficiently small absolute value.

But it can only be satisfied under the condition $AX_0 \cdot Y = 0$, because if $AX_0 \cdot Y > 0$, then the inequality (14) is impossible for sufficiently small positive ϵ, and if $AX_0 \cdot Y < 0$, then it is impossible for negative ϵ of sufficiently small absolute value. Thus, $AX_0 \cdot Y = 0$, i.e., AX_0 is in fact orthogonal to every vector orthogonal to X_0. Therefore AX_0 and X_0 are collinear; i.e., $AX_0 = \lambda' X_0$, where λ' is a real number. The fact that $\lambda' = \lambda_1$ is easy to verify, for

$$\lambda_1 = AX_0 \cdot X_0 = \lambda' X_0 \cdot X_0 = \lambda'.$$

Now it is easy to show that every quadratic form can in fact be reduced to canonical form by an orthogonal transformation.

Let $e_1, e_2, \cdots, e_n$ be the original orthonormal basis of the space and $f_1, f_2, \cdots, f_n$ a new orthonormal basis in which the first vector f_1 is an eigenvector X_0 of the transformation A. Let $x_1, x_2, \cdots, x_n$ be the coordinates of the vector X in the original basis and $x_1', x_2', \cdots, x_n'$ its coordinates in the new basis. Then

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = P \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{bmatrix}$$

where P is an orthogonal matrix.

Let us carry out the transition to the new variables in the quadratic form $AX \cdot X$. In the new variables the quadratic form has the coefficients $a_{ij}' = Af_i \cdot f_j$. Therefore

$$a_{11}' = Af_1 \cdot f_1 = \lambda_1 f_1 \cdot f_1 = \lambda_1,$$
$$a_{1j}' = a_{j1}' = Af_1 \cdot f_j = \lambda_1 f_1 \cdot f_j = 0 \quad \text{for} \quad j \neq 1;$$

i.e., the form is now

$$\lambda_1 x_1'^2 + \phi(x_2', \cdots, x_n').$$

Thus, by means of an orthogonal transformation we have succeeded in separating out one square of a new variable.

By repeating the same arguments with the new form $\phi(x_2', \cdots, x_n')$, etc., we eventually arrive at the conclusion that the form turns out to be reduced to canonical form by a chain of orthogonal transformations. But it is obvious that a chain of orthogonal transformations is equivalent to a single orthogonal transformation. This concludes the proof of the theorem.

§6. Functions of Matrices and Some of Their Applications

Functions of matrices. The applications of linear algebra to other branches of mathematics are very numerous and diverse. It is not an exaggeration to say that the ideas and results of linear algebra are used in a large part of contemporary mathematics and theoretical physics in one form or another, particularly in the form of matrix calculus.

We shall deal briefly with one of the methods of applying the matrix calculus to the theory of ordinary differential equations. Here functions of matrices play an important role.

First of all we define the powers of a square matrix A. We set $A^0 = E$, $A^1 = A$, $A^2 = AA$, $A^3 = A^2A$, $A^4 = A^3A$, etc. By the associative law it is easy to show that $A^mA^n = A^{m+n}$ for arbitrary natural numbers m and n. The operations of addition and of multiplication by a number have been defined for matrices. This enables us to define in a natural way the meaning of a polynomial (in one variable) of a matrix. For if $\phi(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n$, then we set (by definition) $\phi(A) = a_0A^n + a_1A^{n+1} + \cdots + a_nE$. Thus, the concept of the simplest function of a matrix argument, namely the polynomial, is defined.

By means of a limit process it is easy to generalize the concept of a function of a matrix argument to a considerably wider class of functions than polynomials in one variable. Without treating this problem in all its generality we shall restrict ourselves here to the examination of analytic functions.

First of all we introduce the concept of the limit of a sequence of matrices. A sequence of matrices

$$A_1 = \begin{bmatrix} a_{11}^{(1)} \cdots a_{1n}^{(1)} \\ \cdots\cdots\cdots \\ a_{n1}^{(1)} \cdots a_{nn}^{(1)} \end{bmatrix}, \quad A_2 = \begin{bmatrix} a_{11}^{(2)} \cdots a_{1n}^{(2)} \\ \cdots\cdots\cdots \\ a_{n1}^{(2)} \cdots a_{nn}^{(2)} \end{bmatrix}, \cdots$$

is said to *converge* to the matrix

$$A = \begin{bmatrix} a_{11} \cdots a_{1n} \\ \cdots\cdots \\ a_{n1} \cdots a_{nn} \end{bmatrix}$$

(or to have the matrix A as its limit) if $\lim_{k\to\infty} a_{ij}^{(k)} = a_{ij}$ for all i, j. Further, the sum of a series $A_1 + A_2 + \cdots + A_k + \cdots$ is defined as the limit of the sums of its segments $\lim_{k\to\infty}(A_1 + A_2 + \cdots + A_k)$ if this limit exists.

Let $f(z)$ be an analytic function regular in a neighborhood of $z = 0$. Then $f(z)$ can be expanded in a power series

$$f(z) = a_0 + a_1z + a_2z^2 + \cdots + a_kz^k + \cdots.$$

For every square matrix A it is natural to set

$$f(A) = a_0E + a_1A + a_2A^2 + \cdots + a_kA^k + \cdots.$$

Now it turns out that this series converges for all matrices A whose eigenvalues lie within the circle of convergence of the power series $a_0 + a_1z + \cdots + a_kz^k + \cdots$.

Of interest for the applications are the elementary functions of matrices.

For example, the geometric series $E + A + A^2 + \cdots + A^k + \cdots$ is convergent for matrices whose eigenvalues are of absolute value less than 1, and the sum of the series is the matrix $(E - A)^{-1}$, in complete accordance with the formula

$$1 + x + \cdots + x^k + \cdots = \frac{1}{1 - x}.$$

The representation of $(E - A)^{-1}$ in the form of an infinite series gives an effective method for the approximate solution of systems of linear equations whose coefficient matrix is close to the unit matrix.

For when we write such a system in the form

$$(E - A)X = B,$$

we obtain

$$X = (E - A)^{-1}B = B + AB + A^2B + \cdots \tag{15}$$

and this gives a convenient formula for the solution of the system, if only the series (15) converges sufficiently fast.

It is useful to consider the binomial series

$$(E + A)^m = E + \frac{m}{1}A + \frac{m(m-1)}{2!}A^2 + \cdots$$

which can be applied (provided the eigenvalues of A are less than 1 in modulus) not only for natural exponents m but also for fractional and negative exponents.

Particularly important for the applications is the exponential matrix function

$$e^A = E + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots.$$

The series defining the exponential function converges for every matrix A. The exponential matrix function has properties reminding us of properties of the ordinary exponential function. For example, if A and B commute under multiplication, i.e., $AB = BA$, then $e^{A+B} = e^A \cdot e^B$. However, for noncommuting A and B the formula ceases to be true.

Application to the theory of systems of ordinary linear differential equations. In the theory of systems of ordinary linear differential equations it is appropriate to consider matrices whose elements are functions of an independent variable:

$$U(t) = \begin{bmatrix} a_{11}(t) \cdots a_{1n}(t) \\ \cdots\cdots\cdots\cdots \\ a_{m1}(t) \cdots a_{mn}(t) \end{bmatrix}.$$

For such matrices the concept of the derivative with respect to the argument t is defined in a natural way, namely:

$$\frac{dU(t)}{dt} = \begin{bmatrix} a'_{11}(t) \cdots a'_{1n}(t) \\ \cdots\cdots\cdots \\ a'_{m1}(t) \cdots a'_{mn}(t) \end{bmatrix}.$$

It is not difficult to verify that some elementary formulas of differentiation are valid for matrices. For example,

$$\frac{d(U+V)}{dt} = \frac{dU}{dt} + \frac{dV}{dt},$$

$$\frac{d(cU)}{dt} = c\frac{dU}{dt},$$

$$\frac{d(UV)}{dt} = \frac{dU}{dt}V + U\frac{dV}{dt}.$$

(The multiplication must be carried out strictly in the order as given in the formula!)

A system of ordinary linear homogeneous differential equations

$$\begin{aligned} \frac{dy_1}{dt} &= a_{11}(t)\,y_1 + a_{12}(t)\,y_2 + \cdots + a_{1n}(t)\,y_n, \\ \frac{dy_2}{dt} &= a_{21}(t)\,y_1 + a_{22}(t)\,y_2 + \cdots + a_{2n}(t)\,y_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \frac{dy_n}{dt} &= a_{n1}(t)\,y_1 + a_{n2}(t)\,y_2 + \cdots + a_{nn}(t)\,y_n \end{aligned}$$

can be written in this notation in the form

$$\frac{dY}{dt} = A(t)\,Y,$$

where

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \qquad A(t) = \begin{bmatrix} a_{11}(t) \cdots a_{1n}(t) \\ \cdots\cdots\cdots \\ a_{n1}(t) \cdots a_{nn}(t) \end{bmatrix},$$

i.e, in a form similar to a single linear homogeneous differential equation.

If the coefficients of the system are constants, i.e., if the matrix A is constant, then the solution of the system also looks outwardly like the solution of the equation $y' = ay$. For in this case $Y = e^{At}C$, where C is a column of arbitrary constants.

The solution in this form is very convenient for computations. The fact is that for an arbitrary analytic function $f(z)$ we have the equation

$$f(B^{-1}LB) = B^{-1}f(L)\,B.$$

Since every matrix can be reduced to the canonical Jordan form (see §4), the computation of a function of an arbitrary matrix reduces to the computation of a function of a canonical matrix, which is easy to carry out. Therefore, if $A = B^{-1}LB$, where L is a canonical matrix, then

$$Y = e^{At}C = B^{-1}e^{Lt}BC = B^{-1}e^{Lt}C',$$

where $C' = BC$ is a column of arbitrary constants.

From this formula it is easy to derive an explicit expression for all the components of the required column Y.

The Soviet mathematician I. A. Lappo-Danilevskiĭ has successfully developed the apparatus of the theory of matrix functions and was the first to apply it to the investigation of systems of differential equations, including those with variable coefficients. His results count among the most brilliant achievements of mathematics in the last fifty years.

Suggested Reading

G. Birkhoff and S. MacLane, *A survey of modern algebra*, Macmillan, New York, 1941.

V. N. Faddeeva, *Computational methods of linear algebra*. Translated by C. D. Benster, Dover, New York, 1959.

F. R. Gantmacher, *Applications of the theory of matrices*. Translated by J. L. Brenner, Interscience, New York, 1959.

——, *The theory of matrices*, Vols. 1, 2. Translated by K. A. Hirsch, Chelsea New York, 1959.

[This book and the one above are translations from the Russian original, published in 1953, but with various additions, some of which were supplied by the author. The two volumes contain a great deal of material not available in other texts.]

I. M. Gel'fand, *Lectures on linear algebra*, Interscience, New York, 1961.

F. Klein, *Elementary mathematics from an advanced standpoint*. Vol. I. *Arithmetic, algebra, analysis*. Translated by E. R. Hedrick and C. A. Noble, Dover, New York, 1953.

O. Schreier and E. Sperner, *Introduction to modern algebra and matrix theory*, Chelsea, New York, 1951.

V. I. Smirnov, *Linear algebra and group theory*. Translated and revised by R. A. Silverman, McGraw-Hill, New York, 1961.

For such matrices the concept of the derivative with respect to the argument t is defined in a natural way, namely:

$$\frac{dU(t)}{dt} = \begin{bmatrix} a'_{11}(t) \cdots a'_{1n}(t) \\ \cdots\cdots\cdots \\ a'_{m1}(t) \cdots a'_{mn}(t) \end{bmatrix}.$$

It is not difficult to verify that some elementary formulas of differentiation are valid for matrices. For example,

$$\frac{d(U+V)}{dt} = \frac{dU}{dt} + \frac{dV}{dt},$$

$$\frac{d(cU)}{dt} = c\frac{dU}{dt},$$

$$\frac{d(UV)}{dt} = \frac{dU}{dt}V + U\frac{dV}{dt}.$$

(The multiplication must be carried out strictly in the order as given in the formula!)

A system of ordinary linear homogeneous differential equations

$$\begin{aligned} \frac{dy_1}{dt} &= a_{11}(t)\,y_1 + a_{12}(t)\,y_2 + \cdots + a_{1n}(t)\,y_n, \\ \frac{dy_2}{dt} &= a_{21}(t)\,y_1 + a_{22}(t)\,y_2 + \cdots + a_{2n}(t)\,y_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \frac{dy_n}{dt} &= a_{n1}(t)\,y_1 + a_{n2}(t)\,y_2 + \cdots + a_{nn}(t)\,y_n \end{aligned}$$

can be written in this notation in the form

$$\frac{dY}{dt} = A(t)\,Y,$$

where

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \qquad A(t) = \begin{bmatrix} a_{11}(t) \cdots a_{1n}(t) \\ \cdots\cdots\cdots \\ a_{n1}(t) \cdots a_{nn}(t) \end{bmatrix},$$

i.e, in a form similar to a single linear homogeneous differential equation.

If the coefficients of the system are constants, i.e., if the matrix A is constant, then the solution of the system also looks outwardly like the solution of the equation $y' = ay$. For in this case $Y = e^{At}C$, where C is a column of arbitrary constants.

The solution in this form is very convenient for computations. The fact is that for an arbitrary analytic function $f(z)$ we have the equation

$$f(B^{-1}LB) = B^{-1}f(L)\,B.$$

Since every matrix can be reduced to the canonical Jordan form (see §4), the computation of a function of an arbitrary matrix reduces to the computation of a function of a canonical matrix, which is easy to carry out. Therefore, if $A = B^{-1}LB$, where L is a canonical matrix, then

$$Y = e^{At}C = B^{-1}e^{Lt}BC = B^{-1}e^{Lt}C',$$

where $C' = BC$ is a column of arbitrary constants.

From this formula it is easy to derive an explicit expression for all the components of the required column Y.

The Soviet mathematician I. A. Lappo-Danilevskiĭ has successfully developed the apparatus of the theory of matrix functions and was the first to apply it to the investigation of systems of differential equations, including those with variable coefficients. His results count among the most brilliant achievements of mathematics in the last fifty years.

Suggested Reading

G. Birkhoff and S. MacLane, *A survey of modern algebra*, Macmillan, New York, 1941.

V. N. Faddeeva, *Computational methods of linear algebra*. Translated by C. D. Benster, Dover, New York, 1959.

F. R. Gantmacher, *Applications of the theory of matrices*. Translated by J. L. Brenner, Interscience, New York, 1959.

——, *The theory of matrices*, Vols. 1, 2. Translated by K. A. Hirsch, Chelsea New York, 1959.

[This book and the one above are translations from the Russian original, published in 1953, but with various additions, some of which were supplied by the author. The two volumes contain a great deal of material not available in other texts.]

I. M. Gel'fand, *Lectures on linear algebra*, Interscience, New York, 1961.

F. Klein, *Elementary mathematics from an advanced standpoint*. Vol. I. *Arithmetic, algebra, analysis*. Translated by E. R. Hedrick and C. A. Noble, Dover, New York, 1953.

O. Schreier and E. Sperner, *Introduction to modern algebra and matrix theory*, Chelsea, New York, 1951.

V. I. Smirnov, *Linear algebra and group theory*. Translated and revised by R. A. Silverman, McGraw-Hill, New York, 1961.

H. W. Turnbull and A. C. Aitken, *An introduction to the theory of canonical matrices*, Dover, New York, 1961.

B. L. van der Waerden, *Modern algebra*, Vol. I. Translated from the second revised German edition by Fred Blum. With revisions and additions by the author. Frederick Ungar, New York, 1949.