THE THEORY OF MATRICES

BY

F. R. GANTMACHER

VOLUME ONE

CHELSEA PUBLISHING COMPANY
NEW YORK, N. Y.

COPYRIGHT © 1959, BY CHELSEA PUBLISHING COMPANY

© 1959, CHELSEA PUBLISHING COMPANY

COPYRIGHT 1959, BY CHELSEA PUBLISHING COMPANY

COPYRIGHT © 1960, BY CHELSEA PUBLISHING COMPANY

LIBRARY OF CONGRESS CATALOG CARD NUMBER 59-11779

THE PRESENT WORK, PUBLISHED IN TWO VOLUMES, IS AN ENGLISH TRANSLATION BY K. A. HIRSCH, OF THE RUSSIAN-LANGUAGE BOOK TEORIYA MATRITS BY F. R. GANTMACHER (Гантмахер)

DEDALUS - Acervo - IAG

512.831
Gan T
v.1

The theory of matrices,

30200006220

SBN 8284-0131-4

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

THE MATRIX CALCULUS is widely applied nowadays in various branches of mathematics, mechanics, theoretical physics, theoretical electrical engineering, etc. However, neither in the Soviet nor the foreign literature is there a book that gives a sufficiently complete account of the problems of matrix theory and of its diverse applications. The present book is an attempt to fill this gap in the mathematical literature.

The book is based on lecture courses on the theory of matrices and its applications that the author has given several times in the course of the last seventeen years at the Universities of Moscow and Tiflis and at the Moscow Institute of Physical Technology.

The book is meant not only for mathematicians (undergraduates and research students) but also for specialists in allied fields (physics, engineering) who are interested in mathematics and its applications. Therefore the author has endeavoured to make his account of the material as accessible as possible, assuming only that the reader is acquainted with the theory of determinants and with the usual course of higher mathematics within the programme of higher technical education. Only a few isolated sections in the last chapters of the book require additional mathematical knowledge on the part of the reader. Moreover, the author has tried to keep the individual chapters as far as possible independent of each other. For example, Chapter V, *Functions of Matrices*, does not depend on the material contained in Chapters II and III. At those places of Chapter V where fundamental concepts introduced in Chapter IV are being used for the first time, the corresponding references are given. Thus, a reader who is acquainted with the rudiments of the theory of matrices can immediately begin with reading the chapters that interest him.

The book consists of two parts, containing fifteen chapters.

In Chapters I and III, information about matrices and linear operators is developed *ab initio* and the connection between operators and matrices is introduced.

Chapter II expounds the theoretical basis of Gauss's elimination method and certain associated effective methods of solving a system of n linear equations, for large n. In this chapter the reader also becomes acquainted with the technique of operating with matrices that are divided into rectangular 'blocks.'

In Chapter IV we introduce the extremely important 'characteristic' and 'minimal' polynomials of a square matrix, and the 'adjoint' and 'reduced adjoint' matrices.

In Chapter V, which is devoted to functions of matrices, we give the general definition of $f(A)$ as well as concrete methods of computing it—where $f(\lambda)$ is a function of a scalar argument λ and A is a square matrix. The concept of a function of a matrix is used in §§ 5 and 6 of this chapter for a complete investigation of the solutions of a system of linear differential equations of the first order with constant coefficients. Both the concept of a function of a matrix and this latter investigation of differential equations are based entirely on the concept of the minimal polynomial of a matrix and—in contrast to the usual exposition—do not use the so-called theory of elementary divisors, which is treated in Chapters VI and VII.

These five chapters constitute a first course on matrices and their applications. Very important problems in the theory of matrices arise in connection with the reduction of matrices to a normal form. This reduction is carried out on the basis of Weierstrass' theory of elementary divisors. In view of the importance of this theory we give two expositions in this book: an analytic one in Chapter VI and a geometric one in Chapter VII. We draw the reader's attention to §§ 7 and 8 of Chapter VI, where we study effective methods of finding a matrix that transforms a given matrix to normal form. In § 8 of Chapter VII we investigate in detail the method of A. N. Krylov for the practical computation of the coefficients of the characteristic polynomial.

In Chapter VIII certain types of matrix equations are solved. We also consider here the problem of determining all the matrices that are permutable with a given matrix and we study in detail the many-valued functions of matrices $\sqrt[m]{A}$ and $\ln A$.

Chapters IX and X deal with the theory of linear operators in a unitary space and the theory of quadratic and hermitian forms. These chapters do not depend on Weierstrass' theory of elementary divisors and use, of the preceding material, only the basic information on matrices and linear operators contained in the first three chapters of the book. In § 9 of Chapter X we apply the theory of forms to the study of the principal oscillations of a system with n degrees of freedom. In § 11 of this chapter we give an account of Frobenius' deep results on the theory of Hankel forms. These results are used later, in Chapter XV, to study special cases of the Routh-Hurwitz problem.

The last five chapters form the second part of the book [the second volume, in the present English translation]. In Chapter XI we determine normal forms for complex symmetric, skew-symmetric, and orthogonal matrices and establish interesting connections of these matrices with real matrices of the same classes and with unitary matrices.

In Chapter XII we expound the general theory of pencils of matrices of the form $A + \lambda B$, where A and B are arbitrary rectangular matrices of the same dimensions. Just as the study of regular pencils of matrices $A + \lambda B$ is based on Weierstrass' theory of elementary divisors, so the study of singular pencils is built upon Kronecker's theory of minimal indices, which is, as it were, a further development of Weierstrass's theory. By means of Kronecker's theory—the author believes that he has succeeded in simplifying the exposition of this theory—we establish in Chapter XII canonical forms of the pencil of matrices $A + \lambda B$ in the most general case. The results obtained there are applied to the study of systems of linear differential equations with constant coefficients.

In Chapter XIII we explain the remarkable spectral properties of matrices with non-negative elements and consider two important applications of matrices of this class: 1) homogeneous Markov chains in the theory of probability and 2) oscillatory properties of elastic vibrations in mechanics. The matrix method of studying homogeneous Markov chains was developed in the book [25] by V. I. Romanovskii and is based on the fact that the matrix of transition probabilities in a homogeneous Markov chain with a finite number of states is a matrix with non-negative elements of a special type (a 'stochastic' matrix).

The oscillatory properties of elastic vibrations are connected with another important class of non-negative matrices—the 'oscillation matrices.' These matrices and their applications were studied by M. G. Kreĭn jointly with the author of this book. In Chapter XIII, only certain basic results in this domain are presented. The reader can find a detailed account of the whole material in the monograph [7].

In Chapter XIV we compile the applications of the theory of matrices to systems of differential equations with variable coefficients. The central place (§§ 5-9) in this chapter belongs to the theory of the multiplicative integral (Produktintegral) and its connection with Volterra's infinitesimal calculus. These problems are almost entirely unknown in Soviet mathematical literature. In the first sections and in § 11, we study reducible systems (in the sense of Lyapunov) in connection with the problem of stability of motion; we also give certain results of N. P. Erugin. Sections 9-11 refer to the analytic theory of systems of differential equations. Here we clarify an inaccuracy in Birkhoff's fundamental theorem, which is usually applied to the investigation of the solution of a system of differential equations in the neighborhood of a singular point, and we establish a canonical form of the solution in the case of a regular singular point.

In § 12 of Chapter XIV we give a brief survey of some results of the fundamental investigations of I. A. Lappo-Danilevskiĭ on analytic functions of several matrices and their applications to differential systems.

The last chapter, Chapter XV, deals with the applications of the theory of quadratic forms (in particular, of Hankel forms) to the Routh-Hurwitz problem of determining the number of roots of a polynomial in the right half-plane ($\operatorname{Re} z > 0$). The first sections of the chapter contain the classical treatment of the problem. In § 5 we give the theorem of A. M. Lyapunov in which a stability criterion is set up which is equivalent to the Routh-Hurwitz criterion. Together with the stability criterion of Routh-Hurwitz we give, in § 11 of this chapter, the comparatively little known criterion of Liénard and Chipart in which the number of determinant inequalities is only about half of that in the Routh-Hurwitz criterion.

At the end of Chapter XV we exhibit the close connection between stability problems and two remarkable theorems of A. A. Markov and P. L. Chebyshev, which were obtained by these celebrated authors on the basis of the expansion of certain continued fractions of special types in series of decreasing powers of the argument. Here we give a matrix proof of these theorems.

This, then, is a brief summary of the contents of this book.

F. R. Gantmacher

PUBLISHERS' PREFACE

THE PUBLISHERS WISH TO thank Professor Gantmacher for his kindness in communicating to the translator new versions of several paragraphs of the original Russian-language book.

The Publishers also take pleasure in thanking the VEB Deutscher Verlag der Wissenschaften, whose many published translations of Russian scientific books into the German language include a counterpart of the present work, for their kind spirit of cooperation in agreeing to the use of their formulas in the preparation of the present work.

No material changes have been made in the text in translating the present work from the Russian except for the replacement of several paragraphs by the new versions supplied by Professor Gantmacher. Some changes in the references and in the Bibliography have been made for the benefit of the English-language reader.

CONTENTS

PREFACE iii
PUBLISHERS' PREFACE vi

I. MATRICES AND OPERATIONS ON MATRICES 1
§ 1. Matrices. Basic notation 1
§ 2. Addition and multiplication of rectangular matrices 3
§ 3. Square matrices 12
§ 4. Compound matrices. Minors of the inverse matrix 19

II. THE ALGORITHM OF GAUSS AND SOME OF ITS APPLICATIONS 23
§ 1. Gauss's elimination method 23
§ 2. Mechanical interpretation of Gauss's algorithm 28
§ 3. Sylvester's determinant identity 31
§ 4. The decomposition of a square matrix into triangular factors 33
§ 5. The partition of a matrix into blocks. The technique of operating with partitioned matrices. The generalized algorithm of Gauss 41

III. LINEAR OPERATORS IN AN n-DIMENSIONAL VECTOR SPACE 50
§ 1. Vector spaces 50
§ 2. A linear operator mapping an n-dimensional space into an m-dimensional space 55
§ 3. Addition and multiplication of linear operators 57
§ 4. Transformation of coordinates 59
§ 5. Equivalent matrices. The rank of an operator. Sylvester's inequality 61
§ 6. Linear operators mapping an n-dimensional space into itself 66

§ 7. Characteristic values and characteristic vectors of a linear operator ... 69
§ 8. Linear operators of simple structure ... 72

IV. THE CHARACTERISTIC POLYNOMIAL AND THE MINIMAL POLYNOMIAL OF A MATRIX ... 76

§ 1. Addition and multiplication of matrix polynomials ... 76
§ 2. Right and left division of matrix polynomials ... 77
§ 3. The generalized Bézout theorem ... 80
§ 4. The characteristic polynomial of a matrix. The adjoint matrix ... 82
§ 5. The method of Faddeev for the simultaneous computation of the coefficients of the characteristic polynomial and of the adjoint matrix ... 87
§ 6. The minimal polynomial of a matrix ... 89

V. FUNCTIONS OF MATRICES ... 95

§ 1. Definition of a function of a matrix ... 95
§ 2. The Lagrange-Sylvester interpolation polynomial ... 101
§ 3. Other forms of the definition of $f(A)$. The components of the matrix A ... 104
§ 4. Representation of functions of matrices by means of series 110
§ 5. Application of a function of a matrix to the integration of a system of linear differential equations with constant coefficients ... 116
§ 6. Stability of motion in the case of a linear system ... 125

VI. EQUIVALENT TRANSFORMATIONS OF POLYNOMIAL MATRICES. ANALYTIC THEORY OF ELEMENTARY DIVISORS ... 130

§ 1. Elementary transformations of a polynomial matrix ... 130
§ 2. Canonical form of a λ-matrix ... 134
§ 3. Invariant polynomials and elementary divisors of a polynomial matrix ... 139
§ 4. Equivalence of linear binomials ... 145
§ 5. A criterion for similarity of matrices ... 147
§ 6. The normal forms of a matrix ... 149
§ 7. The elementary divisors of the matrix $f(A)$ ... 153

§ 8. A general method of constructing the transforming matrix 159
§ 9. Another method of constructing a transforming matrix ... 164

VII. THE STRUCTURE OF A LINEAR OPERATOR IN AN n-DIMENSIONAL SPACE ... 175

§ 1. The minimal polynomial of a vector and a space (with respect to a given linear operator) ... 175
§ 2. Decomposition into invariant subspaces with co-prime minimal polynomials ... 177
§ 3. Congruence. Factor space ... 181
§ 4. Decomposition of a space into cyclic invariant subspaces ... 184
§ 5. The normal form of a matrix ... 190
§ 6. Invariant polynomials. Elementary divisors ... 193
§ 7. The Jordan normal form of a matrix ... 200
§ 8. Krylov's method of transforming the secular equation ... 202

VIII. MATRIX EQUATIONS ... 215

§ 1. The equation $AX = XB$ ... 215
§ 2. The special case $A = B$. Commuting matrices ... 220
§ 3. The equation $AX - XB = C$ ... 225
§ 4. The scalar equation $f(X) = O$ ... 225
§ 5. Matrix polynomial equations ... 227
§ 6. The extraction of m-th roots of a non-singular matrix ... 231
§ 7. The extraction of m-th roots of a singular matrix ... 234
§ 8. The logarithm of a matrix ... 239

IX. LINEAR OPERATORS IN A UNITARY SPACE ... 242

§ 1. General considerations ... 242
§ 2. Metrization of a space ... 243
§ 3. Gram's criterion for linear dependence of vectors ... 246
§ 4. Orthogonal projection ... 248
§ 5. The geometrical meaning of the Gramian and some inequalities ... 250
§ 6. Orthogonalization of a sequence of vectors ... 256
§ 7. Orthonormal bases ... 262
§ 8. The adjoint operator ... 265

§ 9. Normal operators in a unitary space ... 268
§ 10. The spectra of normal, hermitian, and unitary operators ... 270
§ 11. Positive-semidefinite and positive-definite hermitian operators ... 274
§ 12. Polar decomposition of a linear operator in a unitary space. Cayley's formulas ... 276
§ 13. Linear operators in a euclidean space ... 280
§ 14. Polar decomposition of an operator and the Cayley formulas in a euclidean space ... 286
§ 15. Commuting normal operators ... 290

X. QUADRATIC AND HERMITIAN FORMS ... 294

§ 1. Transformation of the variables in a quadratic form ... 294
§ 2. Reduction of a quadratic form to a sum of squares. The law of inertia ... 296
§ 3. The methods of Lagrange and Jacobi of reducing a quadratic form to a sum of squares ... 299
§ 4. Positive quadratic forms ... 304
§ 5. Reduction of a quadratic form to principal axes ... 308
§ 6. Pencils of quadratic forms ... 310
§ 7. Extremal properties of the characteristic values of a regular pencil of forms ... 317
§ 8. Small oscillations of a system with n degrees of freedom ... 326
§ 9. Hermitian forms ... 331
§ 10. Hankel forms ... 338

BIBLIOGRAPHY ... 351

INDEX ... 369

CHAPTER I

MATRICES AND OPERATIONS ON MATRICES

§ 1. Matrices. Basic Notation

1. Let F be a given number field.[1]

DEFINITION 1: *A rectangular array of numbers of the field* F

$$\begin{Vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{Vmatrix} \tag{1}$$

is called a matrix. *When $m = n$, the matrix is called square and the number* ***m, equal to n, is called its order.*** *In the general case the matrix is called* ***rectangular*** *(of dimension $m \times n$). The numbers that constitute the matrix* ***are called its elements.***

NOTATION: In the double-subscript notation for the elements, the first subscript always denotes the row and the second subscript the column containing the given element.

As an alternative to the notation (1) for a matrix we shall also use the abbreviation

$$\|a_{ik}\| \quad (i = 1, 2, \dots, m;\ k = 1, 2, \dots, n). \tag{2}$$

Often the matrix (1) will also be denoted by a single letter, for example A. If A is a square matrix of order n, then we shall write $A = |\,|a_{ik}|\,|_1^n$. The determinant of a square matrix $A = \|a_{ik}\|_1^n$ will be denoted by $|a_{ik}|_1^n$ or by $|A|$.

[1] A *number field* is defined as an arbitrary collection of numbers within which the four operations of addition, subtraction, multiplication, and division by a non-zero number can always be carried out.

Examples of number fields are: the set of all rational numbers, the set of all real numbers, and the set of all complex numbers.

All the numbers that will occur in the sequel are assumed to belong to the number field given initially.

We introduce a concise notation for determinants formed from elements of the given matrix:

$$A\begin{pmatrix} i_1 & i_2 \dots i_p \\ k_1 & k_2 \dots k_p \end{pmatrix} = \begin{vmatrix} a_{i_1 k_1} & a_{i_1 k_2} \dots a_{i_1 k_p} \\ a_{i_2 k_1} & a_{i_2 k_2} \dots a_{i_2 k_p} \\ \dots\dots\dots\dots\dots \\ a_{i_p k_1} & a_{i_p k_2} \dots a_{i_p k_p} \end{vmatrix}. \tag{3}$$

The determinant (3) is called a *minor* of A of order p, provided $1 \leqq i_1 < i_2 < \dots < i_p \leqq m$ and $1 \leqq k_1 < k_2 < \dots < k_p \leqq n$. A rectangular matrix $A = \| a_{ik} \|$ $(i = 1, 2, \dots, m;\ k = 1, 2, \dots, n)$ has $\binom{m}{p} \cdot \binom{n}{p}$ minors of order p

$$A\begin{pmatrix} i_1 & i_2 \dots i_p \\ k_1 & k_2 \dots k_p \end{pmatrix} \quad \left(\begin{matrix} 1 \leqq i_1 < i_2 < \dots < i_p \leqq m \\ 1 \leqq k_1 < k_2 < \dots < k_p \leqq n \end{matrix};\ p \leqq m, n \right). \tag{3'}$$

The minors (3′) in which $i_1 = k_1$, $i_2 = k_2$, $\dots$, $i_p = k_p$, are called *principal* minors.

In the notation (3) the determinant of a square matrix $A = \| a_{ik} \|_1^n$ can be written as follows:

$$|A| = A\begin{pmatrix} 1 & 2 \dots n \\ 1 & 2 \dots n \end{pmatrix}.$$

The largest among the orders of the non-zero minors generated by a matrix is called the *rank* of the matrix. If r is the rank of a rectangular matrix A of dimension $m \times n$, then obviously $r \leqq \min(m, n)$.

A rectangular matrix consisting of a single column

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

is called a *column matrix* and will be denoted by $(x_1, x_2, \dots, x_n)$.

A rectangular matrix consisting of a single row

$$\| z_1, z_2, \dots, z_n \|$$

is called a *row matrix* and will be denoted by $[z_1, z_2, \dots, z_n]$.

A square matrix in which all the elements outside the main diagonal are zero

$$\begin{Vmatrix} d_1 & 0 \dots 0 \\ 0 & d_2 \dots 0 \\ \dots\dots\dots \\ 0 & 0 \dots d_n \end{Vmatrix}$$

is called a *diagonal matrix* and is denoted by[2] $\| d_i \delta_{ik} \|_1^n$ or by

$$\{ d_1, d_2, \dots, d_n \}.$$

Suppose that m quantities $y_1, y_2, \dots, y_m$ have linear and homogeneous expressions in terms of n other quantities $x_1, x_2, \dots, x_n$:

$$\left.\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ &\dots\dots\dots\dots\dots\dots\dots\dots \\ y_m &= a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n, \end{aligned}\right\} \tag{4}$$

or more concisely,

$$y_i = \sum_{k=1}^{n} a_{ik}x_k \qquad (i = 1, 2, \dots, m). \tag{4'}$$

The transformation of the quantities $x_1, x_2, \dots, x_n$ into the quantities $y_1, y_2, \dots, y_m$ by means of the formulas (4) is called a *linear transformation*.

The coefficients of this transformation form a rectangular matrix (1) of dimension $m \times n$.

The linear transformation (4) determines the matrix (1) uniquely, and vice versa.

In the next section we shall define the basic operations on rectangular matrices using the properties of the linear transformations (4) as our starting point.

§ 2. Addition and Multiplication of Rectangular Matrices

We shall define the basic operations on matrices: addition of matrices, multiplication of a matrix by a number, and multiplication of matrices.

1. Suppose that the quantities $y_1, y_2, \dots, y_m$ are expressed in terms of the quantities $x_1, x_2, \dots, x_n$ by means of the linear transformation

$$y_i = \sum_{k=1}^{n} a_{ik}x_k \qquad (i = 1, 2, \dots, m) \tag{5}$$

[2] Here δ_{ik} is the Kronecker symbol: $\delta_{ik} = \begin{cases} 1 & (i = k), \\ 0 & (i \neq k). \end{cases}$

and the quantities $z_1, z_2, \ldots, z_m$ in terms of the same quantities $x_1, x_2, \ldots, x_n$ by means of the transformation

$$z_i = \sum_{k=1}^{n} b_{ik} x_k \qquad (i = 1, 2, \ldots, m). \tag{6}$$

Then

$$y_i + z_i = \sum_{k=1}^{n} (a_{ik} + b_{ik})\, x_k \qquad (i = 1, 2, \ldots, m). \tag{7}$$

In accordance with this, we formulate the following definition.

Definition 2: *The sum of two rectangular matrices* $A = \| a_{ik} \|$ *and* $B = \| b_{ik} \|$, *both of dimension* $m \times n$, *is the matrix* $C = \| c_{ik} \|$, *of the same dimension, whose elements are the sums of the corresponding elements of the given matrices:*

$$C = A + B,$$

where

$$c_{ik} = a_{ik} + b_{ik} \quad (i = 1, 2, \ldots, m;\ k = 1, 2, \ldots, n).$$

The operation of forming the sum of given matrices is called addition.

Example.

$$\begin{Vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{Vmatrix} + \begin{Vmatrix} c_1 & c_2 & c_3 \\ d_1 & d_2 & d_3 \end{Vmatrix} = \begin{Vmatrix} a_1 + c_1 & a_2 + c_2 & a_3 + c_3 \\ b_1 + d_1 & b_2 + d_2 & b_3 + d_3 \end{Vmatrix}.$$

According to Definition 2, only rectangular matrices of equal dimension can be added.

By virtue of the same definition, the coefficient matrix of the transformation (7) is the sum of the coefficient matrices of the transformations (5) and (6).

From the definition of matrix addition it follows immediately that this operation has the properties of commutativity and associativity:

1. $A + B = B + A$;
2. $(A + B) + C = A + (B + C)$.

Here A, B, and C are arbitrary rectangular matrices all of equal dimension.

The operation of addition of matrices extends in a natural way to the case of an arbitrary finite number of summands.

2. Let us multiply the quantities $y_1, y_2, \ldots, y_m$ in the transformation (5) by some number α of F. Then

$$\alpha y_i = \sum_{k=1}^{n} (\alpha a_{ik})\, x_k \qquad (i = 1, 2, \ldots, m).$$

In accordance with this, we formulate the following definition.

Definition 3. *The product of a matrix* $A = \| a_{ik} \|$ $(i = 1, 2, \ldots, m;\ k = 1, 2, \ldots, n)$ *by a number* α *of* F *is the matrix* $C = \| c_{ik} \|$ $(i = 1, 2, \ldots, m;\ k = 1, 2, \ldots, n)$ *whose elements are obtained from the corresponding elements of* A *by multiplication by* α:

$$C = \alpha A,$$

where

$$c_{ik} = \alpha a_{ik} \qquad (i = 1, 2, \ldots, m;\ k = 1, 2, \ldots, n).$$

The operation of forming the product of a matrix by a number is called multiplication of the matrix by the number.

Example.

$$\alpha \begin{Vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{Vmatrix} = \begin{Vmatrix} \alpha a_1 & \alpha a_2 & \alpha a_3 \\ \alpha b_1 & \alpha b_2 & \alpha b_3 \end{Vmatrix}.$$

It is easy to see that

1. $\alpha(A + B) = \alpha A + \alpha B$,
2. $(\alpha + \beta) A = \alpha A + \beta A$,
3. $(\alpha\beta) A = \alpha(\beta A)$.

Here A and B are rectangular matrices of equal dimension and α and β are numbers of F.

The *difference* $A - B$ of two rectangular matrices of equal dimension is defined by

$$A - B = A + (-1) B.$$

If A is a square matrix of order n and α a number of F, then[3]

$$| \alpha A | = \alpha^n | A |.$$

3. Suppose that the quantities $z_1, z_2, \ldots, z_m$ are expressed in terms of the quantities $y_1, y_2, \ldots, y_n$ by the transformation

$$z_i = \sum_{k=1}^{n} a_{ik} y_k \qquad (i = 1, 2, \ldots, m) \tag{8}$$

and that the quantities $y_1, y_2, \ldots, y_n$ are expressed in terms of the quantities $x_1, x_2, \ldots, x_q$ by the formulas

$$y_k = \sum_{j=1}^{q} b_{kj} x_j \qquad (k = 1, 2, \ldots, n) \tag{9}$$

Then on substituting these expressions for the y_k $(k = 1, 2, \ldots, n)$ in (8) we can express $z_1, z_2, \ldots, z_m$ in terms of $x_1, x_2, \ldots, x_q$ by means of the composite transformation:

[3] Here the symbols $| A |$ and $| \alpha A |$ denote the determinants of the matrices A and αA (see p. 1).

$$z_i=\sum_{k=1}^{n}a_{ik}\sum_{j=1}^{q}b_{kj}x_j=\sum_{j=1}^{q}\Big(\sum_{k=1}^{n}a_{ik}b_{kj}\Big)x_j \qquad (i=1, 2, \ldots, m). \tag{10}$$

In accordance with this we formulate the following definition.

DEFINITION 4. *The product of two rectangular matrices*

$$A=\begin{Vmatrix} a_{11} & a_{12}\ldots a_{1n}\\ a_{21} & a_{22}\ldots a_{2n}\\ \ldots\ldots\ldots\ldots\\ a_{m1} & a_{m2}\ldots a_{mn}\end{Vmatrix}, \qquad B=\begin{Vmatrix} b_{11} & b_{12}\ldots b_{1q}\\ b_{21} & b_{22}\ldots b_{2q}\\ \ldots\ldots\ldots\ldots\\ b_{n1} & b_{n2}\ldots b_{nq}\end{Vmatrix}$$

is the matrix

$$C=\begin{Vmatrix} c_{11} & c_{12}\ldots c_{1q}\\ c_{21} & c_{22}\ldots c_{2q}\\ \ldots\ldots\ldots\ldots\\ c_{m1} & c_{m2}\ldots c_{mq}\end{Vmatrix}$$

in which the element c_{ij} at the intersection of the i-th row and the j-th column is the 'product'[4] of the i-th row of the first matrix A into the j-th column of the second matrix B:

$$c_{ij}=\sum_{k=1}^{n}a_{ik}b_{kj} \qquad (i=1, 2, \ldots, m;\ j=1, 2, \ldots, q). \tag{11}$$

The operation of forming the product of given matrices is called matrix multiplication.

Example.

$$\begin{Vmatrix} a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\end{Vmatrix}\begin{Vmatrix} c_1 & d_1 & e_1 & f_1\\ c_2 & d_2 & c_2 & f_2\\ c_3 & d_3 & e_3 & f_3\end{Vmatrix}=$$

$$=\begin{Vmatrix} a_1c_1+a_2c_2+a_3c_3 & a_1d_1+a_2d_2+a_3d_3 & a_1e_1+a_2e_2+a_3e_3 & a_1f_1+a_2f_2+a_3f_3\\ b_1c_1+b_2c_2+b_3c_3 & b_1d_1+b_2d_2+b_3d_3 & b_1e_1+b_2e_2+b_3e_3 & b_1f_1+b_2f_2+b_3f_3\end{Vmatrix}.$$

By Definition 4 the coefficient matrix of the transformation (10) is the product of the coefficient matrices of (8) and (9).

Note that *the operation of multiplication of two rectangular matrices can only be carried out when the number of columns of the first factor is equal to the number of rows of the second.* In particular, multiplication is always possible when both factors are square matrices of one and the same order.

[4] The product of two sequences of numbers $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ is defined as the sum of the products of the corresponding numbers: $\sum_{i=1}^{n}a_ib_i$.

The reader should observe that even in this special case the multiplication of matrices does not have the property of commutativity. For example,

$$\begin{Vmatrix} 1 & 2\\ 3 & 4\end{Vmatrix}\begin{Vmatrix} 2 & 0\\ 3 & -1\end{Vmatrix}=\begin{Vmatrix} 8 & -2\\ 18 & -4\end{Vmatrix}, \quad \text{but} \quad \begin{Vmatrix} 2 & 0\\ 3 & -1\end{Vmatrix}\begin{Vmatrix} 1 & 2\\ 3 & 4\end{Vmatrix}=\begin{Vmatrix} 2 & 4\\ 0 & 2\end{Vmatrix}.$$

If $AB=BA$, then the matrices A and B are called *permutable* or *commuting*.

Example. The matrices

$$A=\begin{Vmatrix} 1 & 2\\ -2 & 0\end{Vmatrix} \quad \text{and} \quad B=\begin{Vmatrix} -3 & 2\\ -2 & -4\end{Vmatrix}$$

are permutable, because

$$AB=\begin{Vmatrix} -7 & -6\\ 6 & -4\end{Vmatrix} \quad \text{and} \quad BA=\begin{Vmatrix} -7 & -6\\ 6 & -4\end{Vmatrix}.$$

It is very easy to verify the *associative* property of matrix multiplication and also the *distributive* property of multiplication with respect to addition:

1. $(AB)C=A(BC)$,
2. $(A+B)C=AC+BC$,
3. $A(B+C)=AB+AC$.

The definition of matrix multiplication extends in a natural way to the case of several factors.

When we make use of the multiplication of rectangular matrices, we can write the linear transformation

$$\begin{aligned} y_1&=a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n\\ y_2&=a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ y_m&=a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n\end{aligned}$$

as a single matrix equation

$$\begin{Vmatrix} y_1\\ y_2\\ \vdots\\ y_m\end{Vmatrix}=\begin{Vmatrix} a_{11} & a_{12}\ldots a_{1n}\\ a_{21} & a_{22}\ldots a_{2n}\\ \ldots\ldots\ldots\ldots\\ a_{m1} & a_{m2}\ldots a_{mn}\end{Vmatrix}\begin{Vmatrix} x_1\\ x_2\\ \vdots\\ x_n\end{Vmatrix},$$

or in abbreviated form,

$$y=Ax.$$

Here $x=(x_1, x_2, \ldots, x_n)$ and $y=(y_1, y_2, \ldots, y_m)$ are column matrices and $A=\| a_{ik} \|$ is a rectangular matrix of dimension $m \times n$.

Let us treat the special case when in the product $C=AB$ the second factor is a square diagonal matrix $B=\{d_1, d_2, \ldots, d_n\}$. Then it follows from (11) that

$$c_{ij}=a_{ij}d_j \qquad (i=1, 2, \ldots, m;\ j=1, 2, \ldots, n),$$

i.e.,

$$\begin{Vmatrix} a_{11} & a_{12} \ldots a_{1n} \\ a_{21} & a_{22} \ldots a_{2n} \\ \ldots\ldots\ldots\ldots \\ a_{m1} & a_{m2} \ldots a_{mn} \end{Vmatrix} \begin{Vmatrix} d_1 & 0 \ldots 0 \\ 0 & d_2 \ldots 0 \\ \ldots\ldots\ldots \\ 0 & 0 \ldots d_n \end{Vmatrix} = \begin{Vmatrix} a_{11}d_1 & a_{12}d_2 \ldots a_{1n}d_n \\ a_{21}d_1 & a_{22}d_2 \ldots a_{2n}d_n \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}d_1 & a_{m2}d_2 \ldots a_{mn}d_n \end{Vmatrix}.$$

Similarly,

$$\begin{Vmatrix} d_1 & 0 \ldots 0 \\ 0 & d_2 \ldots 0 \\ \ldots\ldots\ldots \\ 0 & 0 \ldots d_m \end{Vmatrix} \begin{Vmatrix} a_{11} & a_{12} \ldots a_{1n} \\ a_{21} & a_{22} \ldots a_{2n} \\ \ldots\ldots\ldots\ldots \\ a_{m1} & a_{m2} \ldots a_{mn} \end{Vmatrix} = \begin{Vmatrix} d_1a_{11} & d_1a_{12} \ldots d_1a_{1n} \\ d_2a_{21} & d_2a_{22} \ldots d_2a_{2n} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ d_ma_{m1} & d_ma_{m2} \ldots d_ma_{mn} \end{Vmatrix}.$$

Hence: *When a rectangular matrix A is multiplied on the right (left) by a diagonal matrix $\{d_1, d_2, \ldots\}$, then the columns (rows) of A are multiplied by $d_1, d_2, \ldots$, respectively.*

4. Suppose that a square matrix $C=\| c_{ij} \|_1^m$ is the product of two rectangular matrices $A=\| a_{ik} \|$ and $B=\| b_{kj} \|$ of dimension $m \times n$ and $n \times m$, respectively:

$$\begin{Vmatrix} c_{11} \ldots c_{1m} \\ \ldots\ldots\ldots \\ c_{m1} \ldots c_{mm} \end{Vmatrix} = \begin{vmatrix} a_{11}\ a_{12} \ldots a_{1n} \\ \ldots\ldots\ldots\ldots \\ a_{m1}\, a_{m2} \ldots a_{mn} \end{vmatrix} \begin{Vmatrix} b_{11} \ldots b_{1m} \\ b_{21} \ldots b_{2m} \\ \ldots\ldots\ldots \\ b_{n1} \ldots b_{nm} \end{Vmatrix}, \tag{12}$$

i.e.,

$$c_{ij}=\sum_{\alpha=1}^{n} a_{i\alpha}b_{\alpha j} \qquad (i, j=1, 2, \ldots, m). \tag{13}$$

We shall establish the important *Binet-Cauchy formula*, which expresses the determinant $|C|$ in terms of the minors of A and B:

$$\begin{vmatrix} c_{11} \ldots c_{1m} \\ \ldots\ldots\ldots \\ c_{m1} \ldots c_{mm} \end{vmatrix} = \sum_{1 \le k_1 < k_2 < \cdots < k_m \le n} \begin{vmatrix} a_{1k_1} \ldots a_{1k_m} \\ \ldots\ldots\ldots \\ a_{mk_1} \ldots a_{mk_m} \end{vmatrix} \begin{vmatrix} b_{k_1 1} \ldots b_{k_1 m} \\ \ldots\ldots\ldots \\ b_{k_m 1} \ldots b_{k_m m} \end{vmatrix} \tag{14}$$

or, in the notation[5] of page 2,

$$C\begin{pmatrix} 1 & 2 \ldots m \\ 1 & 2 \ldots m \end{pmatrix} = \sum_{1 \le k_1 < k_2 < \cdots < k_m \le n} A\begin{pmatrix} 1 & 2 \ldots m \\ k_1 & k_2 \ldots k_m \end{pmatrix} B\begin{pmatrix} k_1 & k_2 \ldots k_m \\ 1 & 2 \ldots m \end{pmatrix}. \tag{14'}$$

According to this formula *the determinant of C is the sum of the products of all possible minors of the maximal (m-th) order[5] of A into the corresponding minors of the same order of B.*

Derivation of the Binet-Cauchy formula. By (13) the determinant of C can be represented in the form

$$\begin{aligned} \begin{vmatrix} c_{11} \ldots c_{1m} \\ \ldots\ldots \\ c_{m1} \ldots c_{mm} \end{vmatrix} &= \begin{vmatrix} \sum_{\alpha_1=1}^{n} a_{1\alpha_1}b_{\alpha_1 1} \ \ldots\ \sum_{\alpha_m=1}^{n} a_{1\alpha_m}b_{\alpha_m m} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \sum_{\alpha_1=1}^{n} a_{m\alpha_1}b_{\alpha_1 1} \ \ldots\ \sum_{\alpha_m=1}^{n} a_{m\alpha_m}b_{\alpha_m m} \end{vmatrix} \\ &= \sum_{\alpha_1, \ldots, \alpha_m=1}^{n} \begin{vmatrix} a_{1\alpha_1}b_{\alpha_1 1} \ \ldots\ a_{1\alpha_m}b_{\alpha_m m} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m\alpha_1}b_{\alpha_1 1} \ \ldots\ a_{m\alpha_m}b_{\alpha_m m} \end{vmatrix} \\ &= \sum_{\alpha_1, \ldots, \alpha_m=1}^{n} A\begin{pmatrix} 1 & 2 & \ldots & m \\ \alpha_1 & \alpha_2 & \ldots & \alpha_m \end{pmatrix} b_{\alpha_1 1}\, b_{\alpha_2 2} \ \ldots\ b_{\alpha_m m}. \end{aligned} \tag{15}$$

If $m > n$, then among the numbers $\alpha_1, \alpha_2, \ldots, \alpha_m$ there are always at least two that are equal, so that every summand on the right-hand side of (15) is zero. Hence in this case $|C|=0$.

Now let $m \le n$. Then in the sum on the right-hand side of (15) all those summands will be zero in which at least two of the subscripts $\alpha_1, \alpha_2, \ldots, \alpha_m$ are equal. All the remaining summands of (15) can be split into groups of $m!$ terms each by combining into one group those summands that differ from each other only in the order of the subscripts $\alpha_1, \alpha_2, \ldots, \alpha_m$ (so that

[5] When $m > n$, the matrices A and B do not have minors of order m. In that case the right-hand sides of (14) and (14') are to be replaced by zero.

within each such group the subscripts $\alpha_1, \alpha_2, \ldots, \alpha_m$ have one and the same set of values). Now within one such group the sum of the corresponding terms is[6]

$$\sum \varepsilon(\alpha_1, \alpha_2, \ldots, \alpha_m) A \begin{pmatrix} 1 & 2 & \ldots & m \\ k_1 & k_2 & \ldots & k_m \end{pmatrix} b_{\alpha_1 1} b_{\alpha_2 2} \cdots b_{\alpha_m m} =$$
$$= A \begin{pmatrix} 1 & 2 & \ldots & m \\ k_1 & k_2 & \ldots & k_m \end{pmatrix} \sum \varepsilon(\alpha_1, \alpha_2, \ldots, \alpha_m) b_{\alpha_1 1} b_{\alpha_2 2} \cdots b_{\alpha_m m}$$
$$= A \begin{pmatrix} 1 & 2 & \ldots & m \\ k_1 & k_2 & \ldots & k_m \end{pmatrix} B \begin{pmatrix} k_1 & k_2 & \ldots & k_m \\ 1 & 2 & \ldots & m \end{pmatrix}.$$

Hence from (15) we obtain (14′).

Example 1.

$$\begin{Vmatrix} a_1c_1 + a_2c_2 + \cdots + a_nc_n & a_1d_1 + a_2d_2 + \cdots + a_nd_n \\ b_1c_1 + b_2c_2 + \cdots + b_nc_n & b_1d_1 + b_2d_2 + \cdots + b_nd_n \end{Vmatrix} = \begin{vmatrix} a_1 & a_2 \ldots a_n \\ b_1 & b_2 \ldots b_n \end{vmatrix} \begin{Vmatrix} c_1 & d_1 \\ c_2 & d_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ c_n & d_n \end{Vmatrix}.$$

Therefore formula (14) yields the so-called *Cauchy identity*

$$\begin{vmatrix} a_1c_1 + a_2c_2 + \cdots + a_nc_n & a_1d_1 + a_2d_2 + \cdots + a_nd_n \\ b_1c_1 + b_2c_2 + \cdots + b_nc_n & b_1d_1 + b_2d_2 + \cdots + b_nd_n \end{vmatrix} = \sum_{1 \leqq i < k \leqq n} \begin{vmatrix} a_i & a_k \\ b_i & b_k \end{vmatrix} \begin{vmatrix} c_i & d_i \\ c_k & d_k \end{vmatrix}. \quad (16)$$

Setting $a_i = c_i$, $b_i = d_i$ $(i = 1, 2, \ldots, n)$ in this identity, we obtain:

$$\begin{vmatrix} a_1^2 + a_2^2 + \cdots + a_n^2 & a_1b_1 + a_2b_2 + \cdots + a_nb_n \\ a_1b_1 + a_2b_2 + \cdots + a_nb_n & b_1^2 + b_2^2 + \cdots + b_n^2 \end{vmatrix} = \sum_{1 \leqq i < k \leqq n} \begin{vmatrix} a_i & a_k \\ b_i & b_k \end{vmatrix}^2.$$

If a_i and b_i $(i = 1, 2, \ldots, n)$ are real numbers, we deduce the well-known inequality

$$(a_1b_1 + a_2b_2 + \cdots + a_nb_n)^2 \leq (a_1^2 + a_2^2 + \cdots + a_n^2)(b_1^2 + b_2^2 + \cdots + b_n^2). \quad (17)$$

Here the equality sign holds if and only if all the numbers a_i are proportional to the corresponding numbers b_i $(i = 1, 2, \ldots, n)$.

Example 2.

$$\begin{Vmatrix} a_1c_1 + b_1d_1 & \ldots & a_1c_n + b_1d_n \\ \cdots\cdots\cdots & & \cdots\cdots\cdots \\ a_nc_1 + b_nd_1 & \ldots & a_nc_n + b_nd_n \end{Vmatrix} = \begin{Vmatrix} a_1 & b_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ a_n & b_n \end{Vmatrix} \begin{Vmatrix} c_1 \ldots c_n \\ d_1 \ldots d_n \end{Vmatrix}.$$

[6] Here $k_1 < k_2 < \ldots < k_m$ is the normal order of the subscripts $\alpha_1, \alpha_2, \ldots, \alpha_m$ and $\varepsilon(\alpha_1, \alpha_2, \ldots, \alpha_m) = (-1)^N$, where N is the number of transpositions of the indices needed to put the permutation $\alpha_1, \alpha_2, \ldots, \alpha_m$ into normal order.

Therefore for $n > 2$

$$\begin{vmatrix} a_1c_1 + b_1d_1 & \ldots & a_1c_n + b_1d_n \\ \cdots\cdots\cdots & & \cdots\cdots\cdots \\ a_nc_1 + b_nd_1 & \ldots & a_nc_n + b_nd_n \end{vmatrix} = 0.$$

Let us consider the special case where A and B are square matrices of one and the same order n. When we set $m = n$ in (14′), we arrive at the well-known multiplication theorem for determinants:

$$C \begin{pmatrix} 1 & 2 \ldots n \\ 1 & 2 \ldots n \end{pmatrix} = A \begin{pmatrix} 1 & 2 \ldots n \\ 1 & 2 \ldots n \end{pmatrix} B \begin{pmatrix} 1 & 2 \ldots n \\ 1 & 2 \ldots n \end{pmatrix}$$

or, in another notation,

$$|C| = |AB| = |A| \cdot |B|. \quad (18)$$

Thus, ***the determinant of the product of two square matrices is equal to the product of the determinants of the factors.***

5. The Binet-Cauchy formula enables us, in the general case also, to express the minors of the product of two rectangular matrices in terms of the minors of the factors. Let

$$A = \|a_{ik}\|, \quad B = \|b_{kj}\|, \quad C = \|c_{ij}\|$$
$$(i = 1, 2, \ldots, m;\ k = 1, 2, \ldots, n;\ j = 1, 2, \ldots, q)$$

and

$$C = AB.$$

We consider an arbitrary minor of C:

$$C \begin{pmatrix} i_1 & i_2 \ldots i_p \\ j_1 & j_2 \ldots j_p \end{pmatrix} \quad \left(1 \leqq \begin{matrix} i_1 < i_2 < \cdots < i_p \leqq m \\ j_1 < j_2 < \cdots < j_p \leqq q \end{matrix};\ p \leqq m \text{ and } p \leqq q \right).$$

The matrix formed from the elements of this minor is the product of two rectangular matrices

$$\begin{Vmatrix} a_{i_1 1} & a_{i_1 2} & \ldots & a_{i_1 n} \\ \cdots\cdots\cdots & & & \\ a_{i_p 1} & a_{i_p 2} & \ldots & a_{i_p n} \end{Vmatrix}, \qquad \begin{Vmatrix} b_{1 j_1} & \ldots & b_{1 j_p} \\ b_{2 j_1} & \ldots & b_{2 j_p} \\ \cdots\cdots & & \cdots\cdots \\ b_{n j_1} & \ldots & b_{n j_p} \end{Vmatrix}.$$

Therefore, by applying the Binet-Cauchy formula, we obtain:[7]

$$C\begin{pmatrix} i_1 & i_2 \dots i_p \\ j_1 & j_2 \dots j_p \end{pmatrix} = \sum_{1 \leqq k_1 < k_2 < \dots < k_p \leqq n} A\begin{pmatrix} i_1 & i_2 \dots i_p \\ k_1 & k_2 \dots k_p \end{pmatrix} B\begin{pmatrix} k_1 & k_2 \dots k_p \\ j_1 & j_2 \dots j_p \end{pmatrix}. \quad (19)$$

For $p=1$ formula (19) goes over into (11). For $p>1$ formula (19) is a natural generalization of (11).

We mention another consequence of (19).

The rank of the product of two rectangular matrices does not exceed the rank of either factor.

If $C=AB$ and r_A, r_B, r_C are the ranks of A, B, C, then

$$r_C \leqq \min(r_A, r_B).$$

§ 3. Square Matrices

1. The square matrix of order n in which the main diagonal consists entirely of units and all the other elements are zero is called the *unit matrix* and is denoted by $E^{(n)}$ or simply by E. The name 'unit matrix' is connected with the following property of E: For every rectangular matrix

$$A = \|a_{ik}\| \qquad (i=1, 2, \dots, m;\ k=1, 2, \dots, n)$$

we have

$$E^{(m)}A = AE^{(n)} = A.$$

Clearly

$$E^{(n)} = \|\delta_{ik}\|_1^n.$$

Let $A=\|a_{ik}\|_1^n$ be a square matrix. Then the *power of the matrix* is defined in the usual way:

$$A^p = \underbrace{AA \cdots A}_{p \text{ times}} \qquad (p=1, 2, \dots); \qquad A^0 = E.$$

From the associative property of matrix multiplication it follows that

$$A^pA^q = A^{p+q}.$$

Here p and q are arbitrary non-negative integers.

[7] It follows from the Binet-Cauchy formula that the minors of order p in C for $p>n$ (if minors of such orders exist) are all zero. In that case the right-hand side of (19) is to be replaced by zero. See footnote 5, p. 9.

We consider a polynomial (integral rational function) with coefficients in the field F:

$$f(t) = \alpha_0 t^m + \alpha_1 t^{m-1} + \dots + \alpha_m.$$

Then by $f(A)$ we shall mean the matrix

$$f(A) = \alpha_0 A^m + \alpha_1 A^{m-1} + \dots + \alpha_m E.$$

We define in this way a *polynomial in a matrix.*

Suppose that $f(t)$ is the product of two polynomials $g(t)$ and $h(t)$:

$$f(t) = g(t)h(t). \quad (21)$$

The polynomial $f(t)$ is obtained from $g(t)$ and $h(t)$ by multiplication term by term and collection of similar terms. In this we make use of the multiplication rule for powers: $t^p \cdot t^q = t^{p+q}$. Since all these operations remain valid when the scalar t is replaced by the matrix A, it follows from (21) that

$$f(A) = g(A)h(A).$$

Hence, in particular,[8]

$$g(A)h(A) = h(A)g(A); \quad (22)$$

i.e., *two polynomials in one and the same matrix are always permutable.*

Examples.

Let the sequence of elements a_{ik} for which $k-i=p$ $(i-k=p)$ in a rectangular matrix $A=\|a_{ik}\|$ be called the p-th *superdiagonal (subdiagonal)* of the matrix. We denote by $H^{(n)}$ the square matrix order n in which all the elements of the first superdiagonal are units and all the other elements are zero. The matrix $H^{(n)}$ will also be denoted simply by H. Then

$$H = H^{(n)} = \begin{Vmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 \\ 0 & 0 & 0 & \dots & 0 \end{Vmatrix}, \qquad H^2 = \begin{Vmatrix} 0 & 0 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 0 & \dots & 0 \end{Vmatrix}, \dots,$$

$$H^p = 0 \qquad (p \geqq n).$$

[8] Since each of these products is equal to one and the same $f(A)$, by virtue of the fact that $h(t)\,g(t) = f(t)$. It is worth mentioning that the substitution of matrices in an algebraic identity in *several* variables is not valid. The substitution of matrices that commute with one another, however, is allowable in this case.

By these equations, if

$$f(t)=a_0+a_1t+a_2t^2+\cdots+a_{n-1}t^{n-1}+\cdots$$

is a polynomial in t, then

$$f(H)=a_0E+a_1H+a_2H^2+\cdots=\begin{Vmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-1} \\ 0 & a_0 & a_1 & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & a_2 \\ \cdot & & & \cdot & a_1 \\ 0 & 0 & 0 & \cdots & a_0 \end{Vmatrix}.$$

Similarly, if F is the square matrix of order n in which all the elements of the first subdiagonal are units and all others are zero, then

$$f(F)=a_0E+a_1F+a_2F^2+\cdots=\begin{Vmatrix} a_0 & 0 & \cdots & 0 \\ a_1 & a_0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ a_{n-1} & \cdots & a_1 & a_0 \end{Vmatrix}.$$

We leave it to the reader to verify the following properties of the matrices H and F:

1. *When an arbitrary rectangular matrix A of dimension $m \times n$ is multiplied on the left by the matrix H (or F) of order m, then all the rows of A are shifted upward (or downward) by one place, the first (last) row of A disappears, and the last (first) row of the product is filled by zeros.* For example,

$$\begin{Vmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{Vmatrix}\begin{Vmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \end{Vmatrix}=\begin{Vmatrix} b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \\ 0 & 0 & 0 & 0 \end{Vmatrix},$$

$$\begin{Vmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{Vmatrix}\begin{Vmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \end{Vmatrix}=\begin{Vmatrix} 0 & 0 & 0 & 0 \\ a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \end{Vmatrix}.$$

2. *When an arbitrary rectangular matrix A of dimension $m \times n$ is multiplied on the right by the matrix H (or F) of order n, then all the columns of A are shifted to the right (left) by one place, the last (first) column of A disappears, and the first (last) column of the product is filled by zeros.* For example,

$$\begin{Vmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \end{Vmatrix}\begin{Vmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{Vmatrix}=\begin{Vmatrix} 0 & a_1 & a_2 & a_3 \\ 0 & b_1 & b_2 & b_3 \\ 0 & c_1 & c_2 & c_3 \end{Vmatrix},$$

$$\begin{Vmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \end{Vmatrix}\begin{Vmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{Vmatrix}=\begin{Vmatrix} a_2 & a_3 & a_4 & 0 \\ b_2 & b_3 & b_4 & 0 \\ c_2 & c_3 & c_4 & 0 \end{Vmatrix}.$$

2. A square matrix A is called *singular* if $|A|=0$. Otherwise A is called ***non-singular***.

Let $A=\|a_{ik}\|_1^n$ be a non-singular matrix ($|A| \neq 0$). Let us consider the linear transformation with coefficient matrix A

$$y_i=\sum_{k=1}^{n} a_{ik}x_k \quad (i=1, 2, \ldots, n). \tag{23}$$

When we regard (23) as equations for $x_1, x_2, \ldots, x_n$ and observe that the determinant of the system of equations (23) is, by assumption, different from zero, then we can express $x_1, x_2, \ldots, x_n$ in terms of $y_1, y_2, \ldots, y_n$ by means of the well-known formulas:

$$x_i=\frac{1}{|A|}\begin{vmatrix} a_{11} & \cdots & a_{1,i-1} & y_1 & a_{1,i+1} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2,i-1} & y_2 & a_{2,i+1} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{n,i-1} & y_n & a_{n,i+1} & \cdots & a_{nn} \end{vmatrix}\equiv\sum_{k=1}^{n} a_{ik}^{(-1)}y_k \quad (i=1, 2, \ldots, n). \tag{24}$$

We have thus obtained the 'inverse' transformation of the transformation (23). The coefficient matrix of this transformation

$$A^{-1}=\|a_{ik}^{(-1)}\|_1^n$$

will be called the *inverse matrix* of A. From (24) it is easy to see that

$$a_{ik}^{(-1)}=\frac{A_{ki}}{|A|} \quad (i, k=1, 2, \ldots, n), \tag{25}$$

where A_{ki} is the algebraic complement (the cofactor) of the element a_{ki} in the determinant $|A|$ $(i, k=1, 2, \ldots, n)$.

For example, if

$$A=\begin{Vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{Vmatrix} \text{ and } |A|\neq 0,$$

then

$$A^{-1}=\frac{1}{|A|}\begin{Vmatrix} b_2c_3-b_3c_2 & a_3c_2-a_2c_3 & a_2b_3-a_3b_2 \\ b_3c_1-b_1c_3 & a_1c_3-a_3c_1 & a_3b_1-a_1b_3 \\ b_1c_2-b_2c_1 & a_2c_1-a_1c_2 & a_1b_2-a_2b_1 \end{Vmatrix}.$$

By forming the composite transformation of the given transformation (23) and the inverse (24), in either order, we obtain in both cases the identity transformation (with the unit matrix as coefficient matrix); therefore

$$AA^{-1}=A^{-1}A=E. \tag{26}$$

The validity of equation (26) can also be established by direct multiplication of the matrices A and A^{-1}. In fact, by (25) we have[9]

$$[AA^{-1}]_{ij}=\sum_{k=1}^{n} a_{ik}a_{kj}^{(-1)}=\frac{1}{|A|}\sum_{k=1}^{n} a_{ik}A_{jk}=\delta_{ij} \qquad (i,j=1,2,\ldots,n).$$

Similarly,

$$[A^{-1}A]_{ij}=\sum_{k=1}^{n} a_{ij}^{(-1)}a_{kj}=\frac{1}{|A|}\sum_{k=1}^{n} A_{ki}a_{kj}=\delta_{ij} \qquad (i,j=1,2,\ldots,n).$$

It is easy to see that the matrix equations

$$AX=E \text{ and } XA=E \qquad (|A|\neq 0) \tag{27}$$

have no solutions other than $X=A^{-1}$. For by multiplying both sides of the first (second) equation on the left (right) by A^{-1} and using the associative property of matrix multiplication we obtain from (26) in both cases:[10]

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

[9] Here we make use of the well-known property of determinants that the sum of the products of the elements of an arbitrary column into the cofactors of the elements of that column is equal to the value of the determinant and the sum of the products of the elements of a column into the cofactors of the corresponding element of another column is zero.

[10] If A is a singular matrix, then the equations (27) have no solution. For if one of these equations had a solution $X=\|x_{ik}\|_1^n$, then we would have by the multiplication theorem of determinants (see formula (18)) that $|A|\cdot|X|=|E|=1$, and this is impossible when $|A|=0$.

In the same way it can be shown that each of the matrix equations

$$AX=B,\ XA=B \qquad (|A|\neq 0), \tag{28}$$

where X and B are rectangular matrices of equal dimensions and A is a square matrix of appropriate order, have one and only one solution,

$$X=A^{-1}B \quad \text{and} \quad X=BA^{-1}, \tag{29}$$

respectively. The matrices (29) are the 'left' and the 'right' quotients on 'dividing' B by A. From (28) and (29) we deduce (see p. 12) that $r_B\leqq r_X$ and $r_X\leqq r_B$, so that $r_X=r_B$. On comparing this with (28), we have:

When a rectangular matrix is multiplied on the left or on the right by a non-singular matrix, the rank of the original matrix remains unchanged.

Note that (26) implies $|A|\cdot|A^{-1}|=1$, i.e.

$$|A^{-1}|=\frac{1}{|A|}.$$

For any two non-singular matrices we have

$$(AB)^{-1}=B^{-1}A^{-1}. \tag{30}$$

3. All the matrices of order n form a ring[11] with unit element $E^{(n)}$.

Since in this ring the operation of multiplication by a number of $\mathfrak{P}$ is defined, and since there exists a basis of n^2 linearly independent matrices in terms of which all the matrices of order n can be expressed linearly,[12] the ring of matrices of order n is an algebra.[13]

[11] A *ring* is a collection of elements in which two operations are defined and can always be carried out uniquely: the 'addition' of two elements (with the commutative and associative properties) and the 'multiplication' of two elements (with the associative and distributive properties with respect to addition); moreover, the addition is reversible. See, for example, van der Waerden, *Modern Algebra*, § 14.

[12] For, an arbitrary matrix $A=\|a_{ik}\|_1^n$ with elements in $\mathfrak{P}$ can be represented in the form $A=\sum_{i,k=1}^{n} a_{ik}E_{ik}$, where E_{ik} is the matrix of order n in which there is a 1 at the intersection of the i-th row and the k-th column and all the other elements are zeros.

[13] See, for example, van der Waerden, *Modern Algebra*, § 17.

All the square matrices of order n form a commutative group with respect to the operation of addition.[14] All the non-singular matrices of order n form a (non-commutative) group with respect to the operation of multiplication.

A square matrix $A = \| a_{ik} \|_1^n$ is called *upper triangular* (*lower triangular*) if all the elements below (above) the main diagonal are zero:

$$A = \begin{Vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & & \cdot \\ 0 & 0 & \dots & a_{nn} \end{Vmatrix}, \quad A = \begin{Vmatrix} a_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & 0 \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \cdot & & \cdot & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{Vmatrix}.$$

A diagonal matrix is a special case both of an upper triangular matrix and a lower triangular matrix.

Since the determinant of a triangular matrix is equal to the product of its diagonal elements, a triangular (and, in particular, a diagonal) matrix is non-singular if and only if all its diagonal elements are different from zero.

It is easy to verify that the sum and the product of two diagonal (upper triangular, lower triangular) matrices is a diagonal (upper triangular, lower triangular) matrix and that the inverse of a non-singular diagonal (upper triangular, lower triangular) matrix is a matrix of the same type. Therefore:

1. All the diagonal matrices of order n form a commutative group under the operation of addition, as do all the upper triangular matrices or all the lower triangular matrices.

2. All the non-singular diagonal matrices form a commutative group under multiplication.

3. All the non-singular upper (lower) triangular matrices form a (non-commutative) group under multiplication.

4. We conclude this section with a further important operation on matrices —*transposition*.

[14] A *group* is a set of objects in which an operation is defined which associates with any two elements a and b of the set a well-defined third element $a * b$ of the same set provided that

1) the operation has the associative property $((a * b) * c = a * (b * c))$,

2) there exists a unit element e in the set $(a * e = e * a = a)$, and

3) for every element a of the set there exists an inverse element a^{-1} $(a * a^{-1} = a^{-1} * a = e)$.

A group is called *commutative*, or abelian, if the group operation has the commutative property. Concerning the group concept see, for example, [53], pp. 245ff.

If $A = \| a_{ik} \|$ $(i = 1, 2, \dots, m;\ k = 1, 2, \dots, n)$, then the *transpose* A^T is defined as $A^\mathsf{T} = \| a_{ik}^\mathsf{T} \|$, where $a_{ki}^\mathsf{T} = a_{ik}$ $(i = 1, 2, \dots, m;\ k = 1, 2, \dots, n)$. If A is of dimension $m \times n$, then A^T is of dimension $n \times m$.

It is easy to verify the following properties:[15]

1. $(A + B)^\mathsf{T} = A^\mathsf{T} + B^\mathsf{T}$,
2. $(\alpha A)^\mathsf{T} = \alpha A^\mathsf{T}$,
3. $(AB)^\mathsf{T} = B^\mathsf{T} A^\mathsf{T}$,
4. $(A^{-1})^\mathsf{T} = (A^\mathsf{T})^{-1}$.

If a square matrix $S = \| s_{ij} \|_1^n$ coincides with its transpose $(S^\mathsf{T} = S)$, then it is called *symmetric*. In a symmetric matrix elements that are symmetrically placed with respect to the main diagonal are equal. Note that the product of two symmetric matrices is not, in general, symmetric. By 3., this holds if and only if the two given symmetric matrices are permutable.

If a square matrix $K = \| k_{ij} \|_1^n$ differs from its transpose by a factor -1 $(K^\mathsf{T} = -K)$, then it is called *skew-symmetric*. In a skew-symmetric matrix any two elements that are symmetrical to the main diagonal differ from each other by a factor -1 and the diagonal elements are zero. From 3. it follows that the product of two permutable skew-symmetric matrices is a symmetric matrix.[16]

§ 4. Compound Matrices. Minors of the Inverse Matrix

1. Let $A = \| a_{ik} \|_1^n$ be a given matrix. We consider all possible minors of A of order p $(1 \leqq p \leqq n)$:

$$A \begin{pmatrix} i_1 & i_2 & \dots & i_p \\ k_1 & k_2 & \dots & k_p \end{pmatrix} \qquad \left(1 \leqq \begin{matrix} i_1 < i_2 < \dots < i_p \\ k_1 < k_2 < \dots < k_p \end{matrix} \leqq n \right). \tag{31}$$

The number of these minors is N^2, where $N = \binom{n}{p}$ is the number of combinations of n objects taken p at a time. In order to arrange the minors (31) in a square array, we enumerate in some definite order—lexicographic order, for example—all the N combinations of p indices selected from among the indices $1, 2, \dots, n$.

[15] In formulas 1., 2., 3., A and B are arbitrary rectangular matrices for which the corresponding operations are feasible. In 4., A is an arbitrary square non-singular matrix.

[16] As regards the representation of a square matrix A in the form of a product of two symmetric matrices $(A = S_1 S_2)$ or two skew-symmetric matrices $(A = K_1 K_2)$, see [357].

If the combinations of indices $i_1 < i_2 < \ldots < i_p$ and $k_1 < k_2 < \ldots < k_p$ have the numbers α and β, then the minors (31) will also be denoted as follows:

$$a_{\alpha\beta} = A\begin{pmatrix} i_1 & i_2 & \ldots & i_p \\ k_1 & k_2 & \ldots & k_p \end{pmatrix}.$$

By giving to α and β independently all the values from 1 to N, we obtain all the minors of $A = \| a_{ik} \|_1^n$ of order p.

The square matrix of order N

$$\mathfrak{A}_p = \| a_{\alpha\beta} \|_1^N$$

is called the p-th *compound matrix* of $A = \| a_{ik} \|_1^n$; p can take the values $1, 2, \ldots, n$. Here $\mathfrak{A}_1 = A$, and $\mathfrak{A}_n$ consists of the single element $| A |$.

Note. The order of enumeration of the combination of indices is fixed once and for all and does not depend on the choice of A.

Example. Let

$$A = \begin{Vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{Vmatrix}.$$

We enumerate all combinations of the indices 1, 2, 3, 4 taken two at a time by arranging them in the following order:

$$(12) \quad (13) \quad (14) \quad (23) \quad (24) \quad (34).$$

Then

$$\mathfrak{A}_2 = \begin{Vmatrix} A\binom{1\ 2}{1\ 2} & A\binom{1\ 2}{1\ 3} & A\binom{1\ 2}{1\ 4} & A\binom{1\ 2}{2\ 3} & A\binom{1\ 2}{2\ 4} & A\binom{1\ 2}{3\ 4} \\ A\binom{1\ 3}{1\ 2} & A\binom{1\ 3}{1\ 3} & A\binom{1\ 3}{1\ 4} & A\binom{1\ 3}{2\ 3} & A\binom{1\ 3}{2\ 4} & A\binom{1\ 3}{3\ 4} \\ A\binom{1\ 4}{1\ 2} & A\binom{1\ 4}{1\ 3} & A\binom{1\ 4}{1\ 4} & A\binom{1\ 4}{2\ 3} & A\binom{1\ 4}{2\ 4} & A\binom{1\ 4}{3\ 4} \\ A\binom{2\ 3}{1\ 2} & A\binom{2\ 3}{1\ 3} & A\binom{2\ 3}{1\ 4} & A\binom{2\ 3}{2\ 3} & A\binom{2\ 3}{2\ 4} & A\binom{2\ 3}{3\ 4} \\ A\binom{2\ 4}{1\ 2} & A\binom{2\ 4}{1\ 3} & A\binom{2\ 4}{1\ 4} & A\binom{2\ 4}{2\ 3} & A\binom{2\ 4}{2\ 4} & A\binom{2\ 4}{3\ 4} \\ A\binom{3\ 4}{1\ 2} & A\binom{3\ 4}{1\ 3} & A\binom{3\ 4}{1\ 4} & A\binom{3\ 4}{2\ 3} & A\binom{3\ 4}{2\ 4} & A\binom{3\ 4}{3\ 4} \end{Vmatrix}$$

We mention some properties of compound matrices:

1. *From $C = AB$ it follows that $\mathfrak{C}_p = \mathfrak{A}_p \cdot \mathfrak{B}_p (p = 1, 2, \ldots, n)$.*

For when we express the minors of order p $(1 \leqq p \leqq n)$ of the matrix product C, by formula (19), in terms of the minors of the same order of the factors, then we have:

$$C\begin{pmatrix} i_1 & i_2 & \ldots & i_p \\ k_1 & k_2 & \ldots & k_p \end{pmatrix} = \sum_{1 \leqq l_1 < l_2 < \cdots < l_p \leqq n} A\begin{pmatrix} i_1 & i_2 & \ldots & i_p \\ l_1 & l_2 & \ldots & l_p \end{pmatrix} B\begin{pmatrix} l_1 & l_2 & \ldots & l_p \\ k_1 & k_2 & \ldots & k_p \end{pmatrix}$$
$$\left(1 \leqq \begin{matrix} i_1 < i_2 < \cdots < i_p \\ k_1 < k_2 < \cdots < k_p \end{matrix} \leqq n\right). \tag{32}$$

Obviously, in the notation of this section, equation (32) can be written as follows:

$$c_{\alpha\beta} = \sum_{\lambda=1}^{N} a_{\alpha\lambda} b_{\lambda\beta} \qquad (\alpha, \beta = 1, 2, \ldots, N)$$

(here α, β, and λ are the numbers of the combinations of indices $i_1 < i_2 < \ldots < i_p$; $k_1 < k_2 < \ldots < k_p$; $l_1 < l_2 < \ldots < l_p$). Hence

$$\mathfrak{C}_p = \mathfrak{A}_p \mathfrak{B}_p \qquad (p = 1, 2, \ldots, n).$$

2. *From $B = A^{-1}$ it follows that $\mathfrak{B}_p = \mathfrak{A}_p^{-1}$ $(p = 1, 2, \ldots, n)$.*

This result follows immediately from the preceding one when we set $C = E$ and bear in mind that $\mathfrak{E}_p$ is the unit matrix of order $N = \binom{n}{p}$

From 2. there follows an important formula that expresses the minors of the inverse matrix in terms of the minors of the given matrix:

If $B = A^{-1}$, then for arbitrary $\left(1 \leqq \begin{matrix} i_1 < i_2 < \cdots < i_p \\ k_1 < k_2 < \cdots < k_p \end{matrix} \leqq n\right)$

$$B\begin{pmatrix} i_1 & i_2 & \ldots & i_p \\ k_1 & k_2 & \ldots & k_p \end{pmatrix} = \frac{(-1)^{\sum_{\nu=1}^{p} i_\nu + \sum_{\nu=1}^{p} k_\nu}}{A\begin{pmatrix} 1 & 2 & \ldots & n \\ 1 & 2 & \ldots & n \end{pmatrix}} A\begin{pmatrix} k_1' & k_2' & \ldots & k_{n-p}' \\ i_1' & i_2' & \ldots & i_{n-p}' \end{pmatrix}, \tag{33}$$

where $i_1 < i_2 < \cdots < i_p$ and $i_1' < i_2' < \cdots < i_{n-p}'$ form a complete system of indices $1, 2, \ldots, n$, as do $k_1 < k_2 < \cdots < k_p$ and $k_1' < k_2' < \cdots < k_{n-p}'$.

For it follows from $AB = E$ that

$$\mathfrak{A}_p \mathfrak{B}_p = \mathfrak{E}_p$$

or in more explicit form:

$$\sum_{\alpha=1}^{N} a_{\gamma\alpha} b_{\alpha\beta} = \delta_{\gamma\beta} = \begin{cases} 1 & (\gamma = \beta), \\ 0 & (\gamma \neq \beta). \end{cases} \tag{34}$$

Equations (34) can also be written as follows:

$$\sum_{1 \leq i_1 < i_2 < \cdots < i_p \leq n} A\begin{pmatrix} j_1 j_2 \cdots j_p \\ i_1 i_2 \cdots i_p \end{pmatrix} B\begin{pmatrix} i_1 i_2 \cdots i_p \\ k_1 k_2 \cdots k_p \end{pmatrix} = \begin{cases} 1, & \text{if} \quad \sum_{\nu=1}^{p} (j_\nu - k_\nu)^2 = 0, \\ 0, & \text{if} \quad \sum_{\nu=1}^{p} (j_\nu - k_\nu)^2 > 0 \end{cases} \tag{34'}$$

$$\left(1 \leq \begin{matrix} j_1 < j_2 < \ldots < j_p \\ k_1 < k_2 < \ldots < k_p \end{matrix} \leq n\right).$$

On the other hand, when we apply the well-known Laplace expansion to the determinant $|A|$, we obtain

$$\sum_{1 \leq i_1 < i_2 < \cdots < i_p \leq n} A\begin{pmatrix} j_1 j_2 \cdots j_p \\ i_1 i_2 \cdots i_p \end{pmatrix} \cdot (-1)^{\sum_{\nu=1}^{p} i_\nu + \sum_{\nu=1}^{p} k_\nu} A\begin{pmatrix} k'_1 k'_2 \cdots k'_{n-p} \\ i'_1 i'_2 \cdots i'_{n-p} \end{pmatrix} =$$

$$= \begin{cases} |A|, & \text{if} \quad \sum_{\nu=1}^{p} (j_\nu - k_\nu)^2 = 0, \\ 0, & \text{if} \quad \sum_{\nu=1}^{p} (j_\nu - k_\nu)^2 > 0, \end{cases} \tag{35}$$

where $i_1 < i_2 < \cdots < i_p$ and $i'_1 < i'_2 < \cdots < i'_{n-p}$ form a complete system of indices $1, 2, \ldots, n$, as do $k_1 < k_2 < \cdots < k_p$ and $k'_1 < k'_2 < \cdots < k'_{n-p}$. Comparison of (35) with (34') and (34) shows that the equations (34) are satisfied if we take together with $b_{\alpha\beta}$ not $B\begin{pmatrix} i_1 i_2 \cdots i_p \\ k_1 k_2 \cdots k_p \end{pmatrix}$ but rather

$$\frac{(-1)^{\sum_{\nu=1}^{p} i_\nu + \sum_{\nu=1}^{p} k_\nu} A\begin{pmatrix} k'_1 k'_2 \cdots k'_{n-p} \\ i'_1 i'_2 \cdots i'_{n-p} \end{pmatrix}}{A\begin{pmatrix} 1 & 2 \ldots n \\ 1 & 2 \ldots n \end{pmatrix}}.$$

Since the elements $b_{\alpha\beta}$ of the inverse matrix of $\mathfrak{A}_p$ are uniquely determined by (34), equation (33) must hold.

CHAPTER II

THE ALGORITHM OF GAUSS AND SOME OF ITS APPLICATIONS

§ 1. Gauss's Elimination Method

1. Let

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

be a system of n linear equations in n unknowns $x_1, x_2, \ldots, x_n$ with right-hand sides $y_1, y_2, \ldots, y_n$.

In matrix form this system may be written as

$$Ax = y. \tag{1'}$$

Here $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ are columns and $A = \|a_{ik}\|_1^n$ is the square coefficient matrix.

If A is non-singular, then we can rewrite this as

$$x = A^{-1}y, \tag{2}$$

or in explicit form:

$$x_i = \sum_{k=1}^{n} a_{ik}^{(-1)} y_k \qquad (i = 1, 2, \ldots, n). \tag{2'}$$

Thus, the task of computing the elements of the inverse matrix $A^{-1} = \|a_{ik}^{(-1)}\|_1^n$ is equivalent to the task of solving the system of equations (1) for *arbitrary* right-hand sides $y_1, y_2, \ldots, y_n$. The elements of the inverse matrix are determined by the formulas (25) of Chapter I. However, the actual computation of the elements of A^{-1} by these formulas is very tedious for large n. Therefore, effective methods of computing the elements of an inverse matrix—and hence of solving a system of linear equations—are of great practical value.[1]

[1] For a detailed account of these methods, we refer the reader to the book by Faddeev [15] and the group of papers that appeared in *Uspehi Mat. Nauk*, Vol. 5, 3 (1950).

In the present chapter we expound the theoretical basis of some of these methods; they are variants of Gauss's elimination method, whose acquaintance the reader first made in his algebra course at school.

2. Suppose that in the system of equations (1) we have $a_{11} \neq 0$. We eliminate x_1 from all the equations beginning with the second by adding to the second equation the first multipled by $-\frac{a_{21}}{a_{11}}$, to the third the first multiplied by $-\frac{a_{31}}{a_{11}}$, and so on. The system (1) has now been replaced by the equivalent system

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= y_1 \\ a_{22}^{(1)}x_2 + \cdots + a_{2n}^{(1)}x_n &= y_2^{(1)} \\ \cdots\cdots\cdots\cdots\cdots & \\ a_{n2}^{(1)}x_2 + \cdots + a_{nn}^{(1)}x_n &= y_n^{(1)} \end{aligned}\right\} \tag{3}$$

The coefficients of the unknowns and the constant terms of the last $n-1$ equations are given by the formulas

$$a_{ij}^{(1)} = a_{ij} - \frac{a_{i1}}{a_{11}}a_{1j}, \qquad y_i^{(1)} = y_i - \frac{a_{i1}}{a_{11}}y_1 \qquad (i, j = 2, \ldots, n). \tag{3'}$$

Suppose that $a_{22}^{(1)} \neq 0$. Then we eliminate x_2 in the same way from the last $n-2$ equations of the system (3) and obtain the system

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= y_1 \\ a_{22}^{(1)}x_2 + a_{23}^{(1)}x_3 + \cdots + a_{2n}^{(1)}x_n &= y_2^{(1)} \\ a_{33}^{(2)}x_3 + \cdots + a_{3n}^{(2)}x_n &= y_3^{(2)} \\ \cdots\cdots\cdots\cdots\cdots & \\ a_{n3}^{(2)}x_3 + \cdots + a_{nn}^{(2)}x_n &= y_n^{(2)}. \end{aligned}\right\} \tag{4}$$

The new coefficients and the new right-hand sides are connected with the preceding ones by the formulas:

$$a_{ij}^{(2)} = a_{ij}^{(1)} - \frac{a_{i2}^{(1)}}{a_{22}^{(1)}}a_{2j}^{(1)}, \quad y_i^{(2)} = y_i^{(1)} - \frac{a_{i2}^{(1)}}{a_{22}^{(1)}}y_2^{(1)} \qquad (i, j = 3, \ldots, n). \tag{5}$$

Continuing the algorithm, we go in $n-1$ steps from the original system (1) to the triangular recurrent system

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= y_1 \\ a_{22}^{(1)}x_2 + a_{23}^{(1)}x_3 + \cdots + a_{2n}^{(1)}x_n &= y_2^{(1)} \\ a_{33}^{(2)}a_3 + \cdots + a_{3n}^{(2)}x_n &= y_3^{(2)} \\ \cdots\cdots\cdots\cdots & \\ a_{nn}^{(n-1)}x_n &= y_n^{(n-1)}. \end{aligned}\right\} \tag{6}$$

This reduction can be carried out if and only if in the process all the numbers $a_{11}, a_{22}^{(1)}, a_{33}^{(2)}, \ldots, a_{n-1,n-1}^{(n-2)}$ turn out to be different from zero.

This *algorithm of Gauss* consists of operations of a simple type such as can easily be carried out by present-day computing machines.

3. Let us express the coefficients and the right-hand sides of the reduced system in terms of the coefficients and the right-hand sides of the original system (1). We shall not assume here that in the reduction process all the numbers $a_{11}, a_{22}^{(1)}, a_{33}^{(2)}, \ldots, a_{n-1,n-1}^{(n-2)}$ turn out to be different from zero; we consider the general case, in which the first p of these numbers are different from zero:

$$a_{11} \neq 0, \ a_{22}^{(1)} \neq 0, \ \ldots, \ a_{pp}^{(p-1)} \neq 0 \qquad (p \leqq n-1). \tag{7}$$

This enables us (at the p-th step of the reduction) to put the original system of equations into the form

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots\cdots\cdots + a_{1n}x_n &= y_1 \\ a_{22}^{(1)}x_2 + \cdots\cdots\cdots + a_{2n}^{(1)}x_n &= y_2^{(1)} \\ \cdots\cdots\cdots\cdots\cdots & \\ a_{pp}^{(p-1)}x_p + \cdots\cdots + a_{pn}^{(p-1)}x_n &= y_p^{(p-1)} \\ a_{p+1,p+1}^{(p)}x_{p+1} + \cdots + a_{p+1,n}^{(p)}x_n &= y_{p+1}^{(p)} \\ \cdots\cdots\cdots\cdots\cdots & \\ a_{n,p+1}^{(p)}x_{p+1} + \cdots + a_{nn}^{(p)}x_n &= y_n^{(p)}. \end{aligned}\right\} \tag{8}$$

We denote the coefficient matrix of this system of equations by G_p:

$$G_p = \begin{Vmatrix} a_{11} & a_{12} & \ldots & a_{1p} & a_{1,p+1} & \ldots & a_{1n} \\ 0 & a_{22}^{(1)} & \ldots & a_{2p}^{(1)} & a_{2,p+1}^{(1)} & \ldots & a_{2n}^{(1)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \ldots & a_{pp}^{(p-1)} & a_{p,p+1}^{(p-1)} & \ldots & a_{pn}^{(p-1)} \\ 0 & 0 & \ldots & 0 & a_{p+1,p+1}^{(p)} & \ldots & a_{p+1,n}^{(p)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \ldots & 0 & a_{n,p+1}^{(p)} & \ldots & a_{nn}^{(p)} \end{Vmatrix}. \tag{9}$$

The transition from A to G_p is effected as follows: To every row of A in succession from the second to the n-th there are added some preceding rows (from the first p) multiplied by certain factors. Therefore all the minors of order h contained in the first h rows of A and G_p are equal:

$$A\begin{pmatrix} 1 & 2 & \ldots & h \\ k_1 & k_2 & \ldots & k_h \end{pmatrix} = G_p\begin{pmatrix} 1 & 2 & \ldots & h \\ k_1 & k_2 & \ldots & k_h \end{pmatrix} \quad \begin{pmatrix} 1 \leqq k_1 < k_2 < \cdots < k_h \leqq n \\ h = 1, 2, \ldots, n \end{pmatrix}. \tag{10}$$

From these formulas we find, by taking into account the structure (9) of G_p,

$$A\begin{pmatrix}1 & 2 & \dots & p\\ 1 & 2 & \dots & p\end{pmatrix}=a_{11}a_{22}^{(1)}\cdot\dots\cdot a_{pp}^{(p-1)}, \tag{11}$$

$$A\begin{pmatrix}1 & 2 & \dots & p & i\\ 1 & 2 & \dots & p & k\end{pmatrix}=a_{11}a_{22}^{(1)}\cdots a_{pp}^{(p-1)}a_{ik}^{(p)} \quad (i,k=p+1,\dots,n), \tag{12}$$

When we divide the second of these equations by the first, we obtain the fundamental formulas[2]

$$a_{ik}^{(p)}=\frac{A\begin{pmatrix}1 & 2 & \dots & p & i\\ 1 & 2 & \dots & p & k\end{pmatrix}}{A\begin{pmatrix}1 & 2 & \dots & p\\ 1 & 2 & \dots & p\end{pmatrix}} \quad (i,k=p+1,\dots,n). \tag{13}$$

If the conditions (7) hold for a given value of p, then they also hold for every smaller value of p. Therefore the formulas (13) are valid not only for the given value of p but also for all smaller values of p. The same holds true of (11). Hence instead of this formula we can write the equations

$$A\begin{pmatrix}1\\ 1\end{pmatrix}=a_{11}, \quad A\begin{pmatrix}1 & 2\\ 1 & 2\end{pmatrix}=a_{11}a_{22}^{(1)}, \quad A\begin{pmatrix}1 & 2 & 3\\ 1 & 2 & 3\end{pmatrix}=a_{11}a_{22}^{(1)}a_{33}^{(2)}, \quad \dots. \tag{14}$$

Thus, the conditions (7), i.e., the necessary and sufficient conditions for the feasibility of the first p steps in Gauss's algorithm, can be written in the form of the following inequalities:

$$A\begin{pmatrix}1\\ 1\end{pmatrix}\neq 0, \quad A\begin{pmatrix}1 & 2\\ 1 & 2\end{pmatrix}\neq 0, \dots, \quad A\begin{pmatrix}1 & 2 & \dots & p\\ 1 & 2 & \dots & p\end{pmatrix}\neq 0 \tag{15}$$

From (14) we then find:

$$a_{11}=A\begin{pmatrix}1\\ 1\end{pmatrix}, \; a_{22}^{(1)}=\frac{A\begin{pmatrix}1 & 2\\ 1 & 2\end{pmatrix}}{A\begin{pmatrix}1\\ 1\end{pmatrix}}, \; a_{33}^{(2)}=\frac{A\begin{pmatrix}1 & 2 & 3\\ 1 & 2 & 3\end{pmatrix}}{A\begin{pmatrix}1 & 2\\ 1 & 2\end{pmatrix}}, \dots, \; a_{pp}^{(p-1)}=\frac{A\begin{pmatrix}1 & 2 & \dots & p\\ 1 & 2 & \dots & p\end{pmatrix}}{A\begin{pmatrix}1 & 2 & \dots & p-1\\ 1 & 2 & \dots & p-1\end{pmatrix}}. \tag{16}$$

In order to eliminate $x_1, x_2, \dots, x_p$ *consecutively* by Gauss's algorithm it is necessary that all the values (16) should be different from zero, i.e., that the inequalities (15) should hold. However, the formulas for $a_{ik}^{(p)}$ make sense if only the last of the conditions (15) holds.

[2] See [181], p. 89.

4. Suppose the coefficient matrix of the system of equations (1) to be of rank r. Then, by a suitable permutation of the equations and a renumbering of the unknowns, we can arrange that the following inequalities hold:

$$A\begin{pmatrix}1 & 2 & \dots & j\\ 1 & 2 & \dots & j\end{pmatrix}\neq 0 \qquad (j=1,2,\dots,r). \tag{17}$$

This enables us to eliminate $x_1, x_2, \dots, x_r$ consecutively and to obtain the system of equations

$$\left.\begin{aligned}
a_{11}x_1+a_{12}x_2+\dots+a_{1n}x_n &= y_1\\
a_{22}^{(1)}x_2+\dots+a_{2n}^{(1)}x_n &= y_2^{(1)}\\
\dots\dots\dots

If $n = r$, i.e. if the matrix $A = \| a_{ik} \|_1^n$ is non-singular, and

$$A\begin{pmatrix} 1 & 2 & \dots & j \\ 1 & 2 & \dots & j \end{pmatrix} \neq 0 \qquad (j = 1, 2, \dots, n),$$

then we can eliminate $x_1, x_2, \dots, x_{n-1}$ in succession by means of Gauss's algorithm and reduce the system of equations to the form (6).

§ 2. Mechanical Interpretation of Gauss's Algorithm

1. We consider an arbitrary elastic statical system S supported on edges (for example, a string, a rod, a multispan rod, a membrane, a lamina, or a discrete system) and choose n points (1), (2), . . . , (n) on it. We shall consider the displacements (sags) $y_1, y_2, \dots, y_n$ of the points (1), (2), . . . , (n) of S under the action of forces $F_1, F_2, \dots, F_n$ applied at these points.

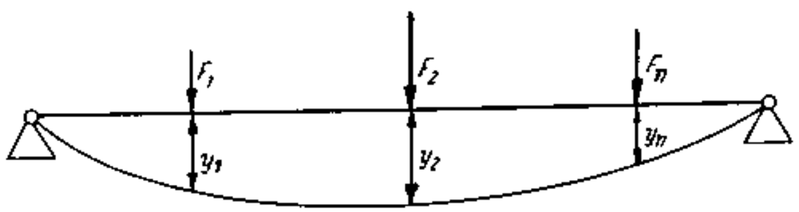

Fig. 1

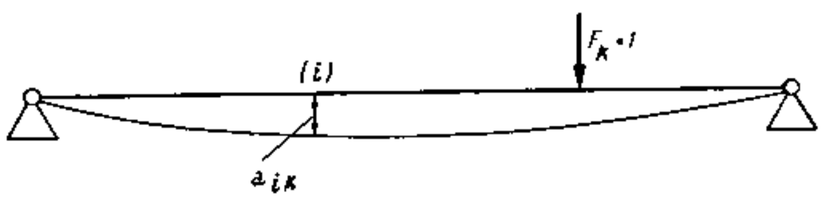

Fig. 2

We assume that the forces and the displacements are parallel to one and the same direction and are determined, therefore, by their algebraic magnitudes (Fig. 1). Moreover, we assume the *principle of linear superposition of forces*:

1. *Under the combined action of two systems of forces the corresponding displacements are added together.*

2. *When the magnitudes of all the forces are multiplied by one and the same real number, then all the displacements are multiplied by the same number.*

We denote by a_{ik} the coefficient of influence of the point (k) on the point (i), i.e., the displacement of (i) under the action of a unit force applied at (k) ($i, k = 1, 2, \dots, n$) (Fig. 2). Then under the combined action of the forces $F_1, F_2, \dots, F_n$ the displacements $y_1, y_2, \dots, y_n$ are determined by the formulas

$$\sum_{k=1}^{n} a_{ik} F_k = y_i \qquad (i = 1, 2, \dots, n). \tag{23}$$

Comparing (23) with the original system (1), we can interpret the task of solving the system of equations (1) as follows:

The displacements $y_1, y_2, \dots, y_n$ being given, we are required to find the corresponding forces $F_1, F_2, \dots, F_n$.

We denote by S_p the statical system that is obtained from S by introducing p fixed hinged supports at the points (1), (2), . . . , (p) ($p \leqq n$). We denote the coefficients of influence for the remaining movable points ($p+1$), . . . , (n) of the system S_p by

$$a_{ik}^{(p)} \qquad (i, k = p+1, \dots, n)$$

(see Fig. 3 for $p = 1$).

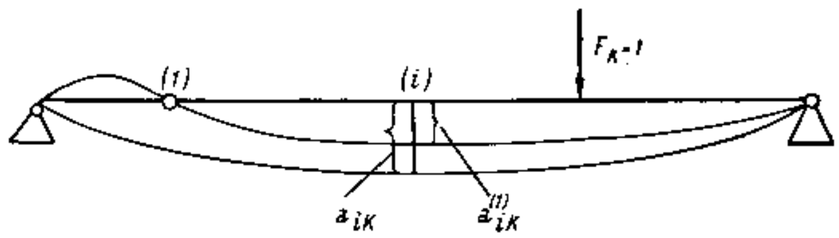

Fig. 3

The coefficient $a_{ik}^{(p)}$ can be regarded as the displacement at the point (i) of S under the action of a unit force at (k) and of the reactions $R_1, R_2, \dots, R_p$ at the fixed points (1), (2), . . . , (p). Therefore

$$a_{ik}^{(p)} = R_1 a_{i1} + \dots + R_p a_{ip} + a_{ik}. \tag{24}$$

On the other hand, under the same forces the displacements of the system S at the points (1), (2), . . . , (p) are zero:

$$\left.\begin{aligned} R_1 a_{11} + \dots + R_p a_{1p} + a_{1k} &= 0 \\ \dots\dots\dots\dots\dots\dots\dots & \\ R_1 a_{p1} + \dots + R_p a_{pp} + a_{pk} &= 0. \end{aligned}\right\} \tag{25}$$

If

$$A\begin{pmatrix}1 & 2 & \dots & p\\ 1 & 2 & \dots & p\end{pmatrix}\neq 0,$$

then we can determine $R_1, R_2, \dots R_p$ from (25) and substitute the expressions so obtained in (24). This elimination of $R_1, R_2, \dots, R_p$ can be carried out as follows. To the system of equations (25) we adjoin (24) written in the form

$$R_1 a_{i1} + \dots + R_p a_{ip} + a_{ik} - a_{ik}^{(p)} = 0. \qquad (24')$$

Regarding (25) and (24′) as a system of $p+1$ homogeneous equations with non-zero solutions $R_1, R_2, \dots, R_p, R_{p+1} = 1$, we see that the determinant of the system must be zero:

$$\begin{vmatrix} a_{11} & \dots & a_{1p} & a_{1k}\\ \dots & \dots & \dots & \dots\\ a_{p1} & \dots & a_{pp} & a_{pk}\\ a_{i1} & \dots & a_{ip} & a_{ik} - a_{ik}^{(p)}\end{vmatrix} = 0.$$

Hence

$$a_{ik}^{(p)} = \frac{A\begin{pmatrix}1 & 2 \dots p & i\\ 1 & 2 \dots p & k\end{pmatrix}}{A\begin{pmatrix}1 & 2 \dots p\\ 1 & 2 \dots p\end{pmatrix}} \qquad (i, k = p+1, \dots, n). \qquad (26)$$

These formulas express the coefficients of influence of the 'support' system S_p in terms of those of the original system S.

But formulas (26) coincide with formulas (13) of the preceding section. Therefore *for every* p $(\leqq n-1)$ *the coefficients* $a_{ik}^{(p)}$ $(i, k = p+1, \dots, n)$ *in the algorithm of Gauss are the coefficients of influence of the support system* S_p.

The truth of this fundamental proposition can also be ascertained by purely mechanical considerations without recourse to the algebraic derivation of formulas (13). For this purpose we consider, to begin with, the special case of a single support: $p = 1$ (Fig. 3). In this case, the coefficients of influence of the system S_1 are given by the formulas (we put $p=1$ in (26)):

$$a_{ik}^{(1)} = \frac{A\begin{pmatrix}1 & i\\ 1 & k\end{pmatrix}}{A\begin{pmatrix}1\\ 1\end{pmatrix}} = a_{ik} - \frac{a_{i1}}{a_{11}} a_{1k} \qquad (i, k = 1, 2, \dots, n)$$

These formulas coincide with the formulas (3′).

Thus, if the coefficients a_{ik} $(i, k = 1, 2, \dots, n)$ in the system of equations (1) are the coefficients of influence of the statical system S, then the coefficients $a_{ik}^{(1)}$ $(i, k = 2, \dots, n)$ in Gauss's algorithm are the coefficients of influence of the system S_1. Applying the same reasoning to the system S_1 and introducing a second support at the point (2) in this system, we see that the coefficients $a_{ik}^{(2)}$ $(i, k = 3, \dots, n)$ in the system of equations (4) are the coefficients of influence of the support system S_2 and, in general, for every p $(\leqq n-1)$ the coefficients $a_{ik}^{(p)}$ $(i, k = p+1, \dots, n)$ in Gauss's algorithm are the coefficients of influence of the support system S_p.

From mechanical considerations it is clear that *the successive introduction of* p *supports is equivalent to the simultaneous introduction of these supports.*

Note. We wish to point out that in the mechanical interpretation of the elimination algorithm it was not necessary to assume that the points at which the displacements are investigated coincide with the points at which the forces $F_1, F_2, \dots, F_n$ are applied. We can assume that $y_1, y_2, \dots, y_n$ are the displacements of the points (1), (2), . . . , (n) and that the forces $F_1, F_2, \dots, F_n$ are applied at the points (1′), (2′), . . . , (n'). Then a_{ik} is the coefficient of influence of the point (k') on the point (k). In that case we must consider instead of the support at the point (j) a generalized support at the points (j), (j') under which the displacement at the point (j) is maintained all the time equal to zero at the expense of a suitably chosen auxiliary force R_j at the point (j'). The conditions that allow us to introduce p generalized supports at the points (1), (1′); (2), (2′), . . . ; (p), (p'), i.e., that allow us to satisfy the conditions $y_1 = 0, y_2 = 0, \dots, y_p = 0$ for arbitrary $F_{p+1}, \dots, F_n$ at the expense of suitable $R_1 = F_1, \dots, R_p = F_p$, can be expressed by the inequality

$$A\begin{pmatrix}1 & 2 \dots p\\ 1 & 2 \dots p\end{pmatrix}\neq 0.$$

§ 3. Sylvester's Determinant Identity

1. In § 1, a comparison of the matrices A and G_p led to equations (10) and (11).

These equations enable us to give an easy proof of the important *determinant identity of Sylvester.* For from (10) and (11) we find:

$$|A| = A\begin{pmatrix}1 & 2 & \dots & n\\ 1 & 2 & \dots & n\end{pmatrix} = A\begin{pmatrix}1 & 2 & \dots & p\\ 1 & 2 & \dots & p\end{pmatrix}\begin{vmatrix} a_{p+1,p+1}^{(p)} & \dots & a_{p+1,n}^{(p)}\\ \dots & \dots & \dots\\ a_{n,p+1}^{(p)} & \dots & a_{nn}^{(p)}\end{vmatrix}. \qquad (27)$$

We introduce borderings of the minor $A\begin{pmatrix}1 & 2 \dots p\\ 1 & 2 \dots p\end{pmatrix}$ by the determinants

$$b_{ik}=A\begin{pmatrix}1 & 2 \dots p & i\\ 1 & 2 \dots p & k\end{pmatrix} \qquad (i,\ k=p+1,\ \dots,\ n).$$

The matrix formed from these determinants will be denoted by

$$B=\|b_{ik}\|_{p+1}^{n}.$$

Then by formulas (13)

$$\begin{vmatrix} a_{p+1,\,p+1}^{(p)} & \cdots & a_{p+1,\,n}^{(p)} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ a_{n,\,p+1}^{(p)} & \cdots & a_{nn}^{(p)} \end{vmatrix} = \frac{\begin{vmatrix} b_{p+1,\,p+1} & \cdots & b_{p+1,\,n} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ b_{n,\,p+1} & \cdots & b_{nn} \end{vmatrix}}{\left[A\begin{pmatrix}1 & 2 & \dots & p\\ 1 & 2 & \dots & p\end{pmatrix}\right]^{n-p}} = \frac{|B|}{\left[A\begin{pmatrix}1 & 2 & \dots & p\\ 1 & 2 & \dots & p\end{pmatrix}\right]^{n-p}}.$$

Therefore equation (27) can be rewritten as follows:

$$|B|=\left[A\begin{pmatrix}1 & 2 & \dots & p\\ 1 & 2 & \dots & p\end{pmatrix}\right]^{n-p-1}|A|. \tag{28}$$

This is Sylvester's determinant identity. It expresses the determinant $|B|$ formed from the bordered determinants in terms of the original determinant and the bordered minor.

We have established equation (28) for a matrix $A\begin{pmatrix}1 & 2 \dots p\\ 1 & 2 \dots p\end{pmatrix}$ whose elements satisfy the inequalities

$$A\begin{pmatrix}1 & 2 & \dots & j\\ 1 & 2 & \dots & j\end{pmatrix}\neq 0 \tag{29}$$

$$(j=1,\ 2,\ \dots,\ p).$$

However, we can show by a 'continuity argument' that this restriction may be removed and that Sylvester's identity holds for an arbitrary matrix $A=\|a_{ik}\|_1^n$. For suppose that the inequalities (29) do not hold. We introduce the matrix

$$A_\varepsilon=A+\varepsilon E.$$

Obviously $\lim\limits_{\varepsilon\to 0} A_\varepsilon=A$. On the other hand, the minors

$$A_\varepsilon\begin{pmatrix}1 & 2 \dots j\\ 1 & 2 \dots j\end{pmatrix}=\varepsilon^j+\cdots$$

$$(j=1,\ 2,\ \dots,\ p)$$

are p polynomials in ε that do not vanish identically. Therefore we can choose a sequence $\varepsilon_m\to 0$ such that

$$A_{\varepsilon_m}\begin{pmatrix}1 & 2 \dots j\\ 1 & 2 \dots j\end{pmatrix}\neq 0 \qquad (j=1,\ 2,\ \dots,\ p;\ m=1,\ 2,\ \dots).$$

We can write down the identity (28) for the matrices A_{ε_m}. Taking the limit $m\to\infty$ on both sides of this identity, we obtain Sylvester's identity for the limit matrix[3] $A=\lim\limits_{m\to\infty} A_{\varepsilon_m}$.

If we apply the identity (28) to the determinant

$$A\begin{pmatrix}1 & 2 \dots p & i_1 & i_2 & \dots i_q\\ 1 & 2 \dots p & k_1 & k_2 & \dots k_q\end{pmatrix} \qquad \left(p<\begin{matrix}i_1<i_2<\cdots<i_q\\ k_1<k_2<\cdots<k_q\end{matrix}\leqq n\right)$$

then we obtain a form of Sylvester's identity particularly convenient for applications

$$B\begin{pmatrix}i_1 & i_2 & \dots i_q\\ k_1 & k_2 & \dots k_q\end{pmatrix}=\left[A\begin{pmatrix}1 & 2 \dots p\\ 1 & 2 \dots p\end{pmatrix}\right]^{q-1}A\begin{pmatrix}1 & 2 \dots p & i_1 & i_2 & \dots i_q\\ 1 & 2 \dots p & k_1 & k_2 & \dots k_q\end{pmatrix}. \tag{30}$$

§ 4. The Decomposition of a Square Matrix into Triangular Factors

1. Let $A=\|a_{ik}\|_1^n$ be a given matrix of rank r. We introduce the following notation for the successive principal minors of the matrix

$$D_k=A\begin{pmatrix}1 & 2 \dots k\\ 1 & 2 \dots k\end{pmatrix} \qquad (k=1,\ 2,\ \dots,\ n).$$

Let us assume that the conditions for the feasibility of Gauss's algorithm are satisfied:

$$D_k\neq 0 \qquad (k=1,\ 2,\ \dots,\ r).$$

We denote by G the coefficient matrix of the system of equations (18) to which the system

$$\sum_{k=1}^{n} a_{ik}x_k=y_i \qquad (i=1,\ 2,\ \dots,\ n)$$

[3] By the *limit* (for $p\to\infty$) *of a sequence of matrices* $X_p=\|x_{ik}^{(p)}\|_1^n$ we mean the matrix $X=\|x_{ik}\|_1^n$, where $x_{ik}=\lim\limits_{p\to\infty} x_{ik}^{(p)}$ $(i,\ k=1,\ 2,\ \dots,\ n)$.

has been reduced by the elimination method of Gauss. The matrix G is of upper triangular form and the elements of its first r rows are determined by the formulas (13), while the elements of the last $n - r$ rows are all equal to zero:[4]

$$G=\begin{Vmatrix} a_{11} & a_{12} & \dots & a_{1r} & a_{1,r+1} & \dots & a_{1n} \\ 0 & a_{22}^{(1)} & \dots & a_{2r}^{(1)} & a_{2,r+1}^{(1)} & \dots & a_{2n}^{(1)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{rr}^{(r-1)} & a_{r,r+1}^{(r-1)} & \dots & a_{rn}^{(r-1)} \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{Vmatrix}.$$

The transition from A to G is effected by a certain number N of operations of the following type: to the i-th row of the matrix we add the j-th row $(j < i)$, after a preliminary multiplication by some number α. Such an operation is equivalent to the multiplication on the left of the matrix to be transformed by the matrix

$$\begin{matrix} & & & (j) & & (i) & \\ \begin{Vmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \dots & & \dots & & \dots & & \dots \\ 0 & \dots & \alpha & \dots & 1 & \dots & 0 \\ \dots & & \dots & & \dots & & \dots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{Vmatrix} & & & & & & \end{matrix}. \tag{31}$$

In this matrix the main diagonal consists entirely of units, and all the remaining elements, except α, are zero.

Thus,

$$G = W_N \cdots W_2 W_1 A,$$

where each matrix $W_1, W_2, \dots, W_N$ is of the form (31) and is therefore a lower triangular matrix with diagonal elements equal to 1.

[4] See formulas (19). G coincides with the matrix G_p (p. 25) for $p = r$.

Let

$$W = W_N \cdots W_2 W_1. \tag{32}$$

Then

$$G = WA. \tag{33}$$

We shall call W the *transforming* matrix for A in Gauss's elimination method. Both matrices G and W are uniquely determined by A. From (32) it follows that W is lower triangular with diagonal elements equal to 1.

Since W is non-singular, we obtain from (33):

$$A = W^{-1}G. \tag{33'}$$

We have thus represented A in the form of a product of a lower triangular matrix W^{-1} and an upper triangular matrix G. The problem of decomposing a matrix A into factors of this type is completely answered by the following theorem:

Theorem 1: *Every matrix $A = \|a_{ik}\|_1^n$ of rank r in which the first r successive principal minors are different from zero*

$$D_k = A\begin{pmatrix} 1 & 2 & \dots & k \\ 1 & 2 & \dots & k \end{pmatrix} \neq 0 \quad \text{for} \quad k = 1, 2, \dots, r \tag{34}$$

can be represented in the form of a product of a lower triangular matrix B and an upper triangular matrix C

$$A = BC = \begin{Vmatrix} b_{11} & 0 & \dots & 0 \\ b_{21} & b_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{Vmatrix} \begin{Vmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ 0 & c_{22} & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & c_{nn} \end{Vmatrix}. \tag{35}$$

Here

$$b_{11}c_{11} = D_1, \quad b_{22}c_{22} = \frac{D_2}{D_1}, \quad \dots, \quad b_{rr}c_{rr} = \frac{D_r}{D_{r-1}}. \tag{36}$$

The values of the first r diagonal elements of B and C can be chosen arbitrarily subject to the conditions (36).

When the first r diagonal elements of B and C are given, then the elements of the first r rows of B and of the first r columns of C are uniquely determined, and are given by the following formulas:

$$b_{gk} = b_{kk}\frac{A\begin{pmatrix} 1 & 2 & \dots & k-1 & g \\ 1 & 2 & \dots & k-1 & k \end{pmatrix}}{A\begin{pmatrix} 1 & 2 & \dots & k \\ 1 & 2 & \dots & k \end{pmatrix}}, \quad c_{kg} = c_{kk}\frac{A\begin{pmatrix} 1 & 2 & \dots & k-1 & k \\ 1 & 2 & \dots & k-1 & g \end{pmatrix}}{A\begin{pmatrix} 1 & 2 & \dots & k \\ 1 & 2 & \dots & k \end{pmatrix}} \tag{37}$$

$$(g = k, k+1, \dots, n; \quad k = 1, 2, \dots, r).$$

If $r < n$ ($|A| = 0$), then all the elements in the last $n - r$ rows of B can be put equal to zero and all the elements of the last $n - r$ columns of C can be chosen arbitrarily; or, conversely, the last $n - r$ rows of C can be filled with zeros and the last $n - r$ rows of B can be chosen arbitrarily.

Proof. That a representation of a matrix satisfying conditions (34) can be given in the form of a product (35) has been proved above (see (33′)).

Now let B and C be arbitrary lower and upper triangular matrices whose product is A. Making use of the formulas for the minors of the product of two matrices we find:

$$A\begin{pmatrix}1 & 2 & \dots & k-1 & g\\ 1 & 2 & \dots & k-1 & k\end{pmatrix} = \sum_{\alpha_1<\alpha_2<\dots<\alpha_k} B\begin{pmatrix}1 & 2 & \dots & k-1 & g\\ \alpha_1 & \alpha_2 & \dots & \alpha_{k-1} & \alpha_k\end{pmatrix} C\begin{pmatrix}\alpha_1 & \alpha_2 & \dots & \alpha_k\\ 1 & 2 & \dots & k\end{pmatrix} \tag{38}$$
$$(g = k, k+1, \dots, n;\ k = 1, 2, \dots, r).$$

Since C is an upper triangular matrix, the first k columns of C contain only one non-vanishing minor of order k, namely $C\begin{pmatrix}1 & 2 & \dots & k\\ 1 & 2 & \dots & k\end{pmatrix}$. Therefore, equation (38) can be written as follows:

$$A\begin{pmatrix}1 & 2 & \dots & k-1 & g\\ 1 & 2 & \dots & k-1 & k\end{pmatrix} = B\begin{pmatrix}1 & 2 & \dots & k-1 & g\\ 1 & 2 & \dots & k-1 & k\end{pmatrix} C\begin{pmatrix}1 & 2 & \dots & k\\ 1 & 2 & \dots & k\end{pmatrix}$$
$$= b_{11}b_{22}\cdots b_{k-1,k-1}b_{gk}c_{11}c_{22}\cdots c_{kk} \tag{39}$$
$$(g = k, k+1, \dots, n;\ k = 1, 2, \dots, r).$$

We put $g = k$ in this equation, obtaining

$$b_{11}b_{22}\cdots b_{kk}c_{11}c_{22}\cdots c_{kk} = D_k \qquad (k = 1, 2, \dots, r), \tag{40}$$

and relations (36) follow.

Without violating equation (35) we may multiply the matrix B in that equation on the right by an arbitrary non-singular diagonal matrix $M = \|\mu_i \delta_{ik}\|_1^n$, while multiplying C at the same time on the left by $M^{-1} = \|\mu_i^{-1}\delta_{ik}\|_1^n$. But this is equivalent to multiplying the columns of B by $\mu_1, \mu_2, \dots, \mu_n$, respectively, and the rows of C by $\mu_1^{-1}, \mu_2^{-1}, \dots, \mu_n^{-1}$. We may therefore give arbitrary values to the diagonal elements $b_{11}, b_{22}, \dots, b_{rr}$ and $c_{11}, c_{22}, \dots, c_{rr}$, provided they satisfy (36).

Further, from (39) and (40) we find:

$$b_{gk} = b_{kk}\frac{A\begin{pmatrix}1 & 2 & \dots & k-1 & g\\ 1 & 2 & \dots & k-1 & k\end{pmatrix}}{A\begin{pmatrix}1 & 2 & \dots & k\\ 1 & 2 & \dots & k\end{pmatrix}} \qquad (g = k, k+1, \dots, n;\ k = 1, 2, \dots, r),$$

i.e., the first formulas in (37). The second formulas in (37), for the elements of C, are established similarly.

We observe that in the multiplication of B and C the elements b_{kg} of the last $n - r$ columns of B and the elements c_{gk} of the last $n - r$ rows of C are multiplied only among each other. We have seen that all the elements of the last $n - r$ rows of C may be chosen to be zero.[5] But as a consequence, the elements of the last $n - r$ columns of B may be chosen arbitrarily. Clearly the product of B and C does not change if we choose the last $n - r$ columns of B to be zeros and choose the elements of the last $n - r$ rows of C arbitrarily.

This completes the proof of the theorem.

From this theorem there follow a number of interesting corollaries.

Corollary 1: *The elements of the first r columns of B and the first r rows of C are connected with the elements of A by the recurrence relations*

$$\left.\begin{aligned} b_{ik} &= \frac{a_{ik} - \sum_{j=1}^{k-1} b_{ij}c_{jk}}{c_{kk}} \qquad (i \geqq k;\ i = 1, 2, \dots, n;\ k = 1, 2, \dots, r),\\ c_{ik} &= \frac{a_{ik} - \sum_{j=1}^{i-1} b_{ij}c_{jk}}{b_{ii}} \qquad (i \leqq k;\ i = 1, 2, \dots, r;\ k = 1, 2, \dots, n). \end{aligned}\right\} \tag{41}$$

The relations (41) follow immediately from the matrix equation (35); they can be used to advantage in the actual computation of the elements of B and C.

Corollary 2: *If $A = \|a_{ik}\|_1^n$ is a non-singular matrix ($r = n$) satisfying (34), then the matrices B and C in the representation (35) are uniquely determined as soon as the diagonal elements of these matrices are chosen in accordance with (36).*

Corollary 3: *If $S = \|s_{ik}\|_1^n$ is a symmetric matrix of rank r and*

$$D_k = S\begin{pmatrix}1 & 2 & \dots & k\\ 1 & 2 & \dots & k\end{pmatrix} \neq 0 \qquad (k = 1, 2, \dots, r),$$

then

$$S = B\,B',$$

where $B = \|b_{ik}\|_1^n$ is a lower triangular matrix in which

[5] This follows from the representation (33′). Here, as we have shown already, arbitrary values may be given to the diagonal elements $b_{11}, \dots, b_{rr}, c_{11}, \dots, c_{rr}$ provided (36) is satisfied by the introduction of suitable factors $\mu_1, \mu_2, \dots, \mu_r$.

$$b_{gk}=\begin{cases}\dfrac{1}{\sqrt{D_kD_{k-1}}}=A\begin{pmatrix}1&2&\dots&k-1&g\\1&2&\dots&k-1&k\end{pmatrix} & (g=k,k+1,\dots,n;\ k=1,2,\dots,r),\\ 0 & (g=k,k+1,\dots,n;\ k=r+1,\dots,n).\end{cases} \tag{42}$$

2. In the representation (35) let the elements of the last $n-r$ columns of C be zero. Then we may set

$$B=F\begin{Vmatrix} b_{11} & & & & 0\\ & \ddots & & & \\ & & b_{rr} & & \\ & & & 0 & \\ & & & & \ddots\\ 0 & & & & 0\end{Vmatrix},\qquad C=\begin{Vmatrix} c_{11} & & & & 0\\ & \ddots & & & \\ & & c_{rr} & & \\ & & & 0 & \\ & & & & \ddots\\ 0 & & & & 0\end{Vmatrix}L\ , \tag{43}$$

where F and L are upper and lower triangular matrices respectively; the first r diagonal elements of F and L are 1 and the elements of the last $n-r$ columns of F and the last $n-r$ rows of L can be chosen completely arbitrarily. Substituting (43) for B and C in (35) and using (36), we obtain the following theorem:

Theorem 2: *Every matrix* $A=\|a_{ik}\|_1^n$ *of rank* r *in which*

$$D_k=A\begin{pmatrix}1&2&\dots&k\\1&2&\dots&k\end{pmatrix}\neq 0 \qquad \text{for } k=1,2,\dots,r$$

can be represented in the form of a product of a lower triangular matrix F, *a diagonal matrix* D, *and an upper triangular matrix* L:

$$A=FDL=\begin{Vmatrix}1&0&\dots&0\\ f_{21}&1&\dots&0\\ \dots&\dots&\dots&\dots\\ f_{n1}&f_{n2}&\dots&1\end{Vmatrix}\begin{Vmatrix}D_1&&&&\\ &\frac{D_2}{D_1}&&&\\ &&\ddots&&\\ &&&\frac{D_r}{D_{r-1}}&\\ &&&&0\\ &&&&&\ddots\\ &&&&&&0\end{Vmatrix}\begin{Vmatrix}1&l_{12}&\dots&l_{1n}\\ 0&1&\dots&l_{2n}\\ \dots&\dots&\dots&\dots\\ 0&0&\dots&1\end{Vmatrix}, \tag{44}$$

where

$$f_{gk}=\frac{A\begin{pmatrix}1&2&\dots&k-1&g\\1&2&\dots&k-1&k\end{pmatrix}}{A\begin{pmatrix}1&2&\dots&k\\1&2&\dots&k\end{pmatrix}},\qquad l_{kg}=\frac{A\begin{pmatrix}1&2&\dots&k-1&k\\1&2&\dots&k-1&g\end{pmatrix}}{A\begin{pmatrix}1&2&\dots&k\\1&2&\dots&k\end{pmatrix}}, \tag{45}$$

$$(g=k+1,\dots,n;\ k=1,2,\dots,r),$$

and f_{gk} and l_{kg} *are arbitrary for* $g=k+1,\dots,n;\ k=r+1,\dots,n$.

3. The elimination method of Gauss, when applied to a matrix $A=\|a_{ik}\|_1^n$ of rank r for which $D_k\neq 0$ $(k=1,2,\dots,r)$, yields two matrices: a lower triangular matrix W with diagonal elements 1 and an upper triangular matrix G in which the first r diagonal elements are $D_1,\frac{D_2}{D_1},\dots,\frac{D_r}{D_{r-1}}$ and the last $n-r$ rows consist entirely of zeros. G is the *Gaussian form* of the matrix A; W is the transforming matrix.

For actual computation of the elements of W we recommend the following device.

We obtain the matrix W when we apply to the unit matrix E all the transformations (given by $W_1,\dots,W_N$) that we have performed on A in the algorithm of Gauss (in this case we shall have instead of the product WA, equal to G, the product WE, equal to W). Let us, therefore, write the unit matrix E on the right of A:

$$\begin{Vmatrix}a_{11}&\dots&a_{1n}&1&\dots&0\\ \dots&\dots&\dots&\dots&\dots&\dots\\ \dots&\dots&\dots&\dots&\dots&\dots\\ a_{n1}&\dots&a_{nn}&0&\dots&1\end{Vmatrix}. \tag{46}$$

By applying all the transformations of the algorithm of Gauss to this rectangular matrix we obtain a rectangular matrix consisting of the two square matrices G and W:

$$(G,W).$$

Thus, *the application of Gauss's algorithm to the matrix* (46) *gives the matrices* G *and* W *simultaneously.*

If A is non-singular, so that $|A|\neq 0$, then $|G|\neq 0$ as well. In this case, (33) implies that $A^{-1}=G^{-1}W$. Since G and W are determined by means of the algorithm of Gauss, the task of finding the inverse matrix A^{-1} reduces to determining G^{-1} and multiplying G^{-1} by W.

Although there is no difficulty in finding the inverse matrix G^{-1} once the matrix G has been determined, because G is triangular, the operations involved can nevertheless be avoided. For this purpose we introduce, together with the matrices G and W, similar matrices G_1 and W_1 for the transposed matrix $A^{\mathsf T}$. Then $A^{\mathsf T}=W_1^{-1}G_1$, i.e.,

$$A=G_1^{\mathsf T}W_1^{\mathsf T-1}. \tag{47}$$

Let us compare (33′) with (44):

$$A=W^{-1}G,\ A=FDL.$$

These equations may be regarded as two distinct decompositions of the form (35); here we take the product DL as the second factor C. Since the first r diagonal elements of the first factors are the same (they are equal to 1), their first r columns coincide. But then, since the last $n - r$ columns of F may be chosen arbitrarily, we chose them such that

$$F = W^{-1}. \tag{48}$$

On the other hand, a comparison of (47) with (44),

$$A = G_1^{\mathsf{T}} W_1^{\mathsf{T}-1}, \quad A = FDL,$$

shows that we may also select the arbitrary elements of L in such a way that

$$L = W_1^{\mathsf{T}-1}. \tag{49}$$

Replacing F and L in (44) by their expressions (48) and (49), we obtain

$$A = W^{-1} D W_1^{\mathsf{T}-1}. \tag{50}$$

Comparing this equation with (33′) and (47) we find:

$$G = DW_1^{\mathsf{T}-1}, \quad G_1^{\mathsf{T}} = W^{-1}D. \tag{51}$$

We now introduce the diagonal matrix

$$\widehat{D} = \left\{\frac{1}{D_1}, \frac{D_1}{D_2}, \ldots, \frac{D_{r-1}}{D_r}, 0, \ldots, 0\right\}. \tag{52}$$

Since

$$D = D\widehat{D}D,$$

it follows from (50) and (51) that

$$A = G_1^{\mathsf{T}} \widehat{D} G. \tag{53}$$

Formula (53) shows that the decomposition of A into triangular factors can be obtained by applying the algorithm of Gauss to the matrices A and A^{T}.

Now let A be non-singular ($r = n$). Then $|D| \neq 0$, $\widehat{D} = D^{-1}$. Therefore it follows from (50) that

$$A^{-1} = W_1^{\mathsf{T}} \widehat{D} W. \tag{54}$$

This formula yields an effective computation of the inverse matrix A^{-1} by the application of Gauss's algorithm to the rectangular matrices

$$(A, E) \qquad (A^{\mathsf{T}}, E).$$

If, in particular, we take as our A a symmetrical matrix S, then G_1 coincides with G and W_1 with W, and therefore formulas (53) and (54) assume the form

$$S = G^{\mathsf{T}} \widehat{D} G, \tag{55}$$

$$S^{-1} = W^{\mathsf{T}} \widehat{D} W. \tag{56}$$

§ 5. The Partition of a Matrix into Blocks. The Technique of Operating with Partitioned Matrices. The Generalized Algorithm of Gauss

It often becomes necessary to use matrices that are partitioned into rectangular parts—'cells' or 'blocks.' In the present section we deal with such partitioned matrices.

1. Let a rectangular matrix

$$A = \|a_{ik}\| \qquad (i = 1, 2, \ldots, m;\ k = 1, 2, \ldots, n) \tag{57}$$

be given.

By means of horizontal and vertical lines we dissect A into rectangular blocks:

$$A = \begin{pmatrix} \overbrace{A_{11}}^{n_1} & \overbrace{A_{12}}^{n_2} \ldots \overbrace{A_{1t}}^{n_t} \\ A_{21} & A_{22} \ldots A_{2t} \\ \ldots\ldots\ldots\ldots \\ A_{s1} & A_{s2} \ldots A_{st} \end{pmatrix} \begin{matrix} \} m_1 \\ \} m_2 \\ \vdots \\ \} m_s \end{matrix}. \tag{58}$$

We shall say of matrix (58) that it is *partitioned* into *st blocks*, or *cells* $A_{\alpha\beta}$ of dimensions $m_\alpha \times n_\beta$ ($\alpha = 1, 2, \ldots, s$; $\beta = 1, 2, \ldots, t$), or that it is represented in the form of a *partitioned*, or blocked, matrix. Instead of (58) we shall simply write

$$A = (A_{\alpha\beta}) \qquad (\alpha = 1, 2, \ldots, s;\ \beta = 1, 2, \ldots, t). \tag{59}$$

In the case $s = t$ we shall use the following notation:

$$A = (A_{\alpha\beta})_1^s. \tag{60}$$

Operations on partitioned matrices are performed according to the same formal rules as in the case in which we have numerical elements instead of blocks. For example, let A and B be two rectangular matrices of equal dimensions partitioned into blocks in exactly the same way:

$$A=(A_{\alpha\beta}),\ B=(B_{\alpha\beta}) \qquad (\alpha=1, 2, \ldots, s;\ \beta=1, 2, \ldots, t). \tag{61}$$

It is easy to verify that

$$A+B=(A_{\alpha\beta}+B_{\alpha\beta}) \qquad (\alpha=1, 2, \ldots, s;\ \beta=1, 2, \ldots, t). \tag{62}$$

We have to consider multiplication of partitioned matrices in more detail. We know (see Chapter I, p. 6) that for the multiplication of two rectangular matrices A and B the length of the rows of the first factor A must be the same as the height of the columns of the second factor B. For 'block' multiplication of these matrices we require, in addition, that the partitioning into blocks be such that the horizontal dimensions in the first factor are the same as the corresponding vertical dimensions in the second:

$$A=\begin{pmatrix} \overset{n_1}{\widetilde{A}_{11}} & \overset{n_2}{\widetilde{A}_{12}} & \cdots & \overset{n_t}{\widetilde{A}_{1t}} \\ A_{21} & A_{22} & \cdots & A_{2t} \\ \cdots & \cdots & \cdots & \cdots \\ A_{s1} & A_{s2} & \cdots & A_{st} \end{pmatrix} \begin{matrix} \}\, m_1 \\ \}\, m_2 \\ \vdots \\ \}\, m_s \end{matrix}, \qquad B=\begin{pmatrix} \overset{p_1}{\widetilde{B}_{11}} & \overset{p_2}{\widetilde{B}_{12}} & \cdots & \overset{p_u}{\widetilde{B}_{1u}} \\ B_{21} & B_{22} & \cdots & B_{2u} \\ \cdots & \cdots & \cdots & \cdots \\ B_{t1} & B_{t2} & \cdots & B_{tu} \end{pmatrix} \begin{matrix} \}\, n_1 \\ \}\, n_2 \\ \vdots \\ \}\, n_t \end{matrix}. \tag{63}$$

Then it is easy to verify that

$$AB=C=(C_{\alpha\beta}),\ \text{where}\quad C_{\alpha\beta}=\sum_{\delta=1}^{t} A_{\alpha\delta}B_{\delta\beta} \qquad \begin{pmatrix}\alpha=1, 2, \ldots, s \\ \beta=1, 2, \ldots, u\end{pmatrix}. \tag{64}$$

We mention separately the special case in which one of the factors is a *quasi-diagonal* matrix. Let A be quasi-diagonal, i.e., let $s=t$ and $A_{\alpha\beta}=O$ for $\alpha\neq\beta$. In this case formula (64) gives

$$C_{\alpha\beta}=A_{\alpha\alpha}B_{\alpha\beta} \qquad (\alpha=1, 2, \ldots, s;\ \beta=1, 2, \ldots, u). \tag{65}$$

When a partitioned matrix is multiplied on the left by a quasi-diagonal matrix, then the rows of the matrix are multiplied on the left by the corresponding diagonal blocks of the quasi-diagonal matrix.

Now let B be a quasi-diagonal matrix, i.e., let $t=u$ and $B_{\alpha\beta}=O$ for $\alpha\neq\beta$. Then we obtain from (64):

$$C_{\alpha\beta}=A_{\alpha\beta}B_{\beta\beta} \qquad (\alpha=1, 2, \ldots, s;\ \beta=1, 2, \ldots, u). \tag{66}$$

When a partitioned matrix is multiplied on the right by a quasi-diagonal matrix, then all the columns of the partitioned matrix are multiplied on the right by the corresponding diagonal cells of the quasi-diagonal matrix.

Note that the multiplication of *square* partitioned matrices of one and the same order is always feasible if the factors are split into equal quadratic schemes of blocks and there are square matrices on the diagonal places in each factor.

The partitioned matrix (58) is called *upper (lower) quasi-triangular* if $s=t$ and all $A_{\alpha\beta}=O$ for $\alpha>\beta$ ($\alpha<\beta$). A quasi-diagonal matrix is a special case of a quasi-triangular matrix.

From the formulas (64) it is easy to see that:

The product of two upper (lower) quasi-triangular matrices is itself an upper (lower) quasi-triangular matrix;[6] *the diagonal cells of the product are obtained by multiplying the corresponding diagonal cells of the factors.*

For when we set $s=t$ in (64) and

$$A_{\alpha\beta}=O,\quad B_{\alpha\beta}=O \quad \text{for} \quad \alpha<\beta,$$

we find

$$\left.\begin{aligned} &C_{\alpha\beta}=O \qquad \text{for} \quad \alpha<\beta \\ \text{and} \qquad &C_{\alpha\alpha}=A_{\alpha\alpha}B_{\alpha\alpha} \end{aligned}\right\} \quad (\alpha, \beta=1, 2, \ldots, s).$$

The case of lower quasi-triangular matrices is treated similarly.

We mention a rule for the calculation of the determinant of a quasi-triangular matrix. This rule can be obtained from the Laplace expansion.

If A is a quasi-triangular matrix (in particular, a quasi-diagonal matrix), then the determinant of the matrix is equal to the product of the determinant of the diagonal cells:

$$|A|=|A_{11}|\,|A_{22}|\cdots|A_{ss}|. \tag{67}$$

2. Let a partitioned matrix

$$A=\begin{pmatrix} \overset{n_1}{\widetilde{A}_{11}} & \overset{n_2}{\widetilde{A}_{12}} & \cdots & \overset{n_t}{\widetilde{A}_{1t}} \\ A_{21} & A_{22} & \cdots & A_{2t} \\ \cdots & \cdots & \cdots & \cdots \\ A_{s1} & A_{s2} & \cdots & A_{st} \end{pmatrix} \begin{matrix} \}\, m_1 \\ \}\, m_2 \\ \vdots \\ \}\, m_s \end{matrix}. \tag{68}$$

[6] It is assumed here that the block multiplication is feasible.

be given. To the α-th row of submatrices we add the β-th row, multiplied on the left by a rectangular matrix X of dimension $m_\alpha \times n_\beta$. We obtain a partitioned matrix

$$B = \begin{pmatrix} A_{11} & \dots & A_{1t} \\ \dots & \dots & \dots \\ A_{\alpha 1} + XA_{\beta 1} & \dots & A_{\alpha t} + XA_{\beta t} \\ \dots & \dots & \dots \\ A_{\beta 1} & \dots & A_{\beta t} \\ \dots & \dots & \dots \\ A_{s1} & \dots & A_{st} \end{pmatrix}. \tag{69}$$

We introduce an auxiliary square matrix V, which we give in the form of a square scheme of blocks:

$$V = \begin{array}{c} \begin{matrix} \widehat{m_1} & \dots & \widehat{m_\alpha} & \dots & \widehat{m_\beta} & \dots & \widehat{m_s} \end{matrix} \\ \begin{pmatrix} \widehat{E} & \dots & \widehat{O} & \dots & \widehat{O} & \dots & \widehat{O} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ O & \dots & E & \dots & X & \dots & O \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ O & \dots & O & \dots & E & \dots & O \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ O & \dots & O & \dots & O & \dots & E \end{pmatrix} \begin{matrix} \} m_1 \\ \vdots \\ \} m_\alpha \\ \vdots \\ \} m_\beta \\ \vdots \\ \} m_s \end{matrix} \end{array} \tag{70}$$

In the diagonal blocks of V there are unit matrices of order $m_1, m_2, \dots, m_s$, respectively; all the non-diagonal blocks of V are equal to zero except the block X that lies at the intersection of the α-th row and β-th column.

It is easy to see that

$$VA = B. \tag{71}$$

As V is non-singular, we have[7] for the ranks of A and B:

$$r_A = r_B. \tag{72}$$

In the special case where A is a square matrix, we have from (70):

$$|V||A| = |B|. \tag{73}$$

But the determinant of the quasi-triangular matrix V is 1:

$$|V| = 1. \tag{74}$$

Hence

$$|A| = |B|. \tag{75}$$

[7] See p. 12.

The same conclusion holds when we add to an arbitrary columu of (68) another column multiplied on the right by a rectangular matrix x of suitable dimensions.

The results obtained can be formulated as the following theorem.

Theorem 3: *If to the α-th row (column) of the blocks of the partitioned matrix A we add the β-th row (column) multiplied on the left (right) by a rectangular matrix X of the corresponding dimensions, then the rank of A remains unchanged under this transformation and, if A is a square matrix, the determinant of A is also unchanged.*

3. We now consider the special case in which the diagonal block A_{11} in A is square and non-singular ($|A_{11}| \neq 0$).

To the α-th row of A we add the first row multiplied on the left by $-A_{\alpha 1}A_{11}^{-1}$ ($\alpha = 2, \dots, s$). We thus obtain the matrix

$$B_1 = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1t} \\ O & A_{22}^{(1)} & \dots & A_{2t}^{(1)} \\ \dots & \dots & \dots & \dots \\ O & A_{s2}^{(1)} & \dots & A_{st}^{(1)} \end{pmatrix}, \tag{76}$$

where

$$A_{\alpha\beta}^{(1)} = -A_{\alpha 1}A_{11}^{-1}A_{1\beta} + A_{\alpha\beta} \qquad (\alpha = 2, \dots, s;\ \beta = 2, \dots, t). \tag{77}$$

If the matrix $A_{22}^{(1)}$ is square and non-singular, then the process can be continued. In this way we arrive at the *generalized algorithm of Gauss*.

Let A be a square matrix. Then

$$|A| = |B_1| = |A_{11}| \begin{vmatrix} A_{22}^{(1)} & \dots & A_{2t}^{(1)} \\ \dots & \dots & \dots \\ A_{s2}^{(1)} & \dots & A_{st}^{(1)} \end{vmatrix}. \tag{78}$$

Formula (78) reduces the computation of the determinant $|A|$, consisting of st blocks to the computation of a determinant of lower order consisting of $(s-1)\cdot(t-1)$ blocks.[8]

Let us consider a determinant Δ partitioned into four blocks:

$$\Delta = \begin{vmatrix} A & B \\ C & D \end{vmatrix}. \tag{79}$$

where A and D are square matrices.

Suppose $|A| \neq 0$. Then from the second row we subtract the first multiplied on the left by CA^{-1}. We obtain

[8] If $A_{22}^{(1)}$ is a square matrix and $|A_{22}^{(1)}| \neq 0$, then this determinant of $(s-1)$ $(t-1)$ blocks can again be subjected to such a transformation, etc.

$$\Delta=\begin{vmatrix} A & B \\ O & D-CA^{-1}B \end{vmatrix}=|A|\,|D-CA^{-1}B|. \tag{I}$$

Similarly, if $|D|\neq 0$, we subtract from the first row in Δ the second multiplied on the left by BD^{-1}, obtaining

$$\Delta=\begin{vmatrix} A-BD^{-1}C & O \\ C & D \end{vmatrix}=|A-BD^{-1}C|\,|D|. \tag{II}$$

In the special case in which all four matrices A, B, C, D are square (of one and the same order n), we deduce from (I) and (II) the *formulas of Schur*, which reduce the computation of a determinant of order $2n$ to the computation of a determinant of order n:

$$\Delta=|AD-ACA^{-1}B| \qquad (A\neq 0), \tag{Ia}$$

$$\Delta=|AD-BD^{-1}CD| \qquad (D\neq 0). \tag{IIa}$$

If the matrices A and C are permutable, then it follows from (Ia) that

$$\Delta=|AD-CB| \qquad (\text{provided } AC=CA). \tag{Ib}$$

Similarly, if C and D are permutable, then

$$\Delta=|AD-BC| \qquad (\text{provided } CD=DC). \tag{IIb}$$

Formula (Ib) was obtained under the assumption $|A|\neq 0$, and (IIb) under the assumption $|D|\neq 0$. However, these restrictions can be removed by continuity arguments.

From formulas (I)-(IIb) we can obtain another six formulas by replacing A and D on the right-hand sides simultaneously by B and C.

Example.

$$\Delta=\begin{vmatrix} 1 & 0 & b_1 & b_2 \\ 0 & 1 & b_3 & b_4 \\ c_1 & c_2 & d_1 & d_2 \\ c_3 & c_4 & d_3 & d_4 \end{vmatrix}.$$

By formula (Ib),

$$\Delta=\begin{vmatrix} d_1-c_1b_1-c_2b_3 & d_2-c_1b_2-c_2b_4 \\ d_3-c_3b_1-c_4b_3 & d_4-c_3b_2-c_4b_4 \end{vmatrix}.$$

4. From Theorem 3 there follows also

Theorem 4: *If a rectangular matrix R is represented in partitioned form*

$$R=\begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{80}$$

where A is a square non-singular matrix of order n ($|A|\neq 0$), then the rank of R is equal to n if and only if

$$D=CA^{-1}B. \tag{81}$$

Proof. We subtract from the second row of blocks of R the first, multiplied on the left by CA^{-1}. Then we obtain the matrix

$$T=\begin{pmatrix} A & B \\ O & D-CA^{-1}B \end{pmatrix}. \tag{82}$$

By Theorem 3, the matrices R and T have the same rank. But the rank of T coincides with the rank of A (namely, n) if and only if $D-CA^{-1}B=O$, i.e., when (80) holds. This proves the theorem.

From Theorem 4 there follows an algorithm[9] for the construction of the inverse matrix A^{-1} and, more generally, the product $CA^{-1}B$, where B and C are rectangular matrices of dimensions $n\times p$ and $q\times n$.

By means of Gauss's algorithm,[10] we reduce the matrix

$$\begin{pmatrix} A & B \\ -C & O \end{pmatrix} \qquad (|A|\neq 0) \tag{83}$$

to the form

$$\begin{pmatrix} G & B_1 \\ O & X \end{pmatrix}. \tag{84}$$

We will show that

$$X=CA^{-1}B. \tag{85}$$

For, the same transformation that was applied to the matrix (83) reduces the matrix

[9] See [181].

[10] We do not apply here the entire algorithm of Gauss to the matrix (83) but only the first n steps of the algorithm, where n is the order of the matrix. This can be done if the conditions (15) hold for $p=n$. But if these conditions do not hold, then, since $|A|\neq 0$, we may renumber the first n rows (or the first n columns) of the matrix (83) so that the n steps of Gauss's algorithm turn out to be feasible. Such a modified Gaussian algorithm is sometimes applied even when the conditions (15), with $p=n$, are satisfied.

$$\begin{pmatrix} A & B \\ -C & -CA^{-1}B \end{pmatrix} \tag{86}$$

to the form

$$\begin{pmatrix} G & B_1 \\ O & X-CA^{-1}B \end{pmatrix}. \tag{87}$$

By Theorem 4, the matrix (86) is of rank n (n is the order of A). But then (87) must also be of rank n. Hence $X - CA^{-1}B = O$, i.e., (85) holds.

In particular, if $B = y$, where y is a column matrix, and $C = E$, then

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

Therefore, when we apply Gauss's algorithm to the matrix

$$\begin{pmatrix} A & y \\ -E & O \end{pmatrix},$$

we obtain the solution of the system of equations

$$Ax = y.$$

Further, if in (83) we set $B = C = E$, then by applying the algorithm of Gauss to the matrix

$$\begin{pmatrix} A & E \\ -E & O \end{pmatrix},$$

we obtain

$$\begin{pmatrix} G & W \\ O & X \end{pmatrix},$$

where

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

Let us illustrate this method by finding A^{-1} in the following example.

Example. Let

$$A = \begin{Vmatrix} 2 & 1 & 1 \\ 1 & 0 & 2 \\ 3 & 1 & 2 \end{Vmatrix}.$$

It is required to compute A^{-1}.

We apply a somewhat modified elimination method[11] to the matrix

[11] See the preceding footnote.

$$\left\|\begin{array}{rrrrrr} 2 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 & 1 & 0 \\ 3 & 1 & 2 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \end{array}\right\|.$$

To all the rows we add certain multiples of the second row and we arrange that all the elements of the first column, except the second, become zero. Then we add to all the rows, except the second, the third row multiplied by certain factors and see to it that in the second column all the elements, except the second and third, become zero. Then we add to the last three rows the first row with suitable factors and obtain a matrix of the form

$$\left\|\begin{array}{cccccc} * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ 0 & 0 & 0 & -2 & -1 & 2 \\ 0 & 0 & 0 & 4 & 1 & -3 \\ 0 & 0 & 0 & 1 & 1 & -1 \end{array}\right\|.$$

Therefore

$$A^{-1} = \left\|\begin{array}{rrr} -2 & -1 & 2 \\ 4 & 1 & -3 \\ 1 & 1 & -1 \end{array}\right\|.$$

CHAPTER III

LINEAR OPERATORS IN AN n-DIMENSIONAL VECTOR SPACE

Matrices constitute the fundamental analytic apparatus for the study of linear operators in an n-dimensional space. The study of these operators, in turn, enables us to divide all matrices into classes and to exhibit the significant properties that all matrices of one and the same class have in common.

In the present chapter we shall expound the simpler properties of linear operators in an n-dimensional space. The investigation will be continued in Chapters VII and IX.

§ 1. Vector Spaces

1. Let $\boldsymbol{R}$ be a set of arbitrary elements $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \ldots$ in which two operations are defined:[1] the operation of 'addition' and the operation of 'multiplication by a number of the field F.' We postulate that these operations can always be performed uniquely in $\boldsymbol{R}$ and that the following rules hold for arbitrary elements $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$ of $\boldsymbol{R}$ and numbers α, β of F:

1. $\boldsymbol{x}+\boldsymbol{y}=\boldsymbol{y}+\boldsymbol{x}$.
2. $(\boldsymbol{x}+\boldsymbol{y})+\boldsymbol{z}=\boldsymbol{x}+(\boldsymbol{y}+\boldsymbol{z})$.
3. There exists an element $\boldsymbol{o}$ in $\boldsymbol{R}$ such that the product of the number 0 with any element $\boldsymbol{x}$ of $\boldsymbol{R}$ is equal to $\boldsymbol{o}$:

   $0\cdot\boldsymbol{x}=\boldsymbol{o}$.
4. $1\cdot\boldsymbol{x}=\boldsymbol{x}$.
5. $\alpha(\beta\boldsymbol{x})=(\alpha\beta)\boldsymbol{x}$.
6. $(\alpha+\beta)\boldsymbol{x}=\alpha\boldsymbol{x}+\beta\boldsymbol{x}$.
7. $\alpha(\boldsymbol{x}+\boldsymbol{y})=\alpha\boldsymbol{x}+\alpha\boldsymbol{y}$.

[1] These operations will be denoted by the usual signs '+' and '·'; the latter sign will sometimes be omitted.

DEFINITION 1: *A set $\boldsymbol{R}$ of elements in which two operations—'addition' of elements and 'multiplication of elements of $\boldsymbol{R}$ by a number of* F*'—can always be performed uniquely and for which postulates* 1.-7. *hold is called a vector space (over the field* F*) and the elements are called vectors.*[2]

DEFINITION 2. *The vectors $\boldsymbol{x}, \boldsymbol{y}, \ldots, \boldsymbol{u}$ of $\boldsymbol{R}$, are called linearly dependent if there exist numbers $\alpha, \beta, \ldots, \delta$ in* F*, not all zero, such that*

$$\alpha\boldsymbol{x}+\beta\boldsymbol{y}+\cdots+\delta\boldsymbol{u}=\boldsymbol{o}. \tag{1}$$

If such a linear dependence does not hold, then the vectors $\boldsymbol{x}, \boldsymbol{y}, \ldots, \boldsymbol{u}$ are called linearly independent.

If the vectors $\boldsymbol{x}, \boldsymbol{y}, \ldots, \boldsymbol{u}$ are linearly dependent, then one of the vectors can be repesented as a linear combination, with coefficients in R, of the remaining ones. For example, if $\alpha\neq 0$ in (1), then

$$\boldsymbol{x}=-\frac{\beta}{\alpha}\boldsymbol{y}-\cdots-\frac{\delta}{\alpha}\boldsymbol{u}.$$

DEFINITION 3. *The space $\boldsymbol{R}$ is called finite-dimensional and the number n is called the dimension of the space if there exist n linearly independent vectors in $\boldsymbol{R}$, while any $n+1$ vectors in $\boldsymbol{R}$ are linearly dependent. If the space contains linearly independent systems of an arbitrary number of vectors, then it is called infinite-dimensional.*

In this book we shall study mainly finite-dimensional spaces.

DEFINITION 4. *A system of n linearly independent vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ of an n-dimensional space, given in a definite order, is called a basis of the space.*

2. *Example* 1. The set of all ordinary vectors (directed geometrical segments) is a three-dimensional vector space. The part of this space that consists of the vectors parallel to some plane is a two-dimensional space, and all the vectors parallel to a given line form a one-dimensional vector space.

Example 2. Let us call a column $x=(x_1, x_2, \ldots, x_n)$ of n numbers of F a vector (where n is a fixed number). We define the basic operations as operations on column matrices:

[2] It is easy to see that all the usual properties of the operations of addition and of multiplication by a number follow from properties 1.-7. For example, for arbitrary $\boldsymbol{x}$ of $\boldsymbol{R}$ we have:

$$\boldsymbol{x}+\boldsymbol{o}=\boldsymbol{x}\ [\boldsymbol{x}+\boldsymbol{o}=1\cdot\boldsymbol{x}+0\cdot\boldsymbol{x}=(1+0)\cdot\boldsymbol{x}=1\cdot\boldsymbol{x}=\boldsymbol{x}];$$
$$\boldsymbol{x}+(-\boldsymbol{x})=\boldsymbol{o}, \text{ where } -\boldsymbol{x}=(-1)\cdot\boldsymbol{x};$$

etc.

$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n),$$
$$\alpha(x_1, x_2, \ldots, x_n) = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n).$$

The null vector is the column $(0, 0, \ldots, 0)$. It is easy to verify that all the postulates 1.-7. are satisfied. The vectors form an n-dimensional space. As a basis of the space we can take, for example, the column of unit matrices of order n:

$$(1, 0, \ldots, 0), (0, 1, \ldots, 0), \ldots, (0, 0, \ldots, 1).$$

The space thus defined is often called the *n-dimensional number space.*

Example 3. The set of all infinite sequences $(x_1, x_2, \ldots, x_n, \ldots)$ in which the operations are defined in a natural way, i.e.,

$$(x_1, x_2, \ldots, x_n, \ldots) + (y_1, y_2, \ldots, y_n, \ldots) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n, \ldots),$$
$$\alpha(x_1, x_2, \ldots, x_n, \ldots) = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n, \ldots),$$

is an infinite-dimensional space.

Example 4. The set of polynomials $a_0 + a_1 t + \ldots + a_{n-1} t^{n-1}$ of degree $< n$ with coefficients in F is an n-dimensional vector space.[3] As a basis of this space we can take, say, the system of powers $t^0, t^1, \ldots, t^{n-1}$.

The set of all such polynomials (without a bound on the degree) form an infinite-dimensional space.

Example 5. The set of all functions defined on a closed interval $[a, b]$ form an infinite-dimensional space.

3. Let the vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ forms a basis of an n-dimensional vector space $\boldsymbol{R}$ and let $\boldsymbol{x}$ be an arbitrary vector of the space. Then the vectors $\boldsymbol{x}, \boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ are linearly dependent (because there are $n + 1$ of them):

$$\alpha_0 \boldsymbol{x} + \alpha_1 \boldsymbol{e}_1 + \alpha_2 \boldsymbol{e}_2 + \cdots + \alpha_n \boldsymbol{e}_n = \boldsymbol{o},$$

where at least one of the numbers $\alpha_0, \alpha_1, \ldots, \alpha_n$ is different from zero. But in this case we must have $\alpha_0 \neq 0$, since the vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ cannot be linearly dependent. Therefore

$$\boldsymbol{x} = x_1 \boldsymbol{e}_1 + x_2 \boldsymbol{e}_2 + \cdots + x_n \boldsymbol{e}_n \tag{2}$$

where $x_i = -\alpha_i/\alpha_0$ $(i = 1, 2, \ldots, n)$.

Note that the numbers $x_1, x_2, \ldots, x_n$ are uniquely determined when the vector $\boldsymbol{x}$ and the basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ are given. For if there is another decomposition of $\boldsymbol{x}$ besides (2),

$$\boldsymbol{x} = x_1' \boldsymbol{e}_1 + x_2' \boldsymbol{e}_2 + \cdots + x_n' \boldsymbol{e}_n, \tag{3}$$

[3] The basic operations are taken to be ordinary addition of polynomials and multiplication of a polynomial by a number.

then, by subtracting (2) from (3), we obtain

$$(x_1' - x_1) \boldsymbol{e}_1 + (x_2' - x_2) \boldsymbol{e}_2 + \cdots + (x_n' - x_n) \boldsymbol{e}_n = \boldsymbol{o},$$

and since the vectors of a basis are linearly dependent, it follows that

$$x_1' - x_1 = x_2' - x_2 = \cdots = x_n' - x_n = 0,$$

i.e.,

$$x_1' = x_1, \; x_2' = x_2, \; \ldots, \; x_n' = x_n. \tag{4}$$

The numbers $x_1, x_2, \ldots, x_n$ are called the *coordinates* of $\boldsymbol{x}$ in the basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$.

If

$$\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{e}_i \text{ and } \boldsymbol{y} = \sum_{i=1}^{n} y_i \boldsymbol{e}_i,$$

then

$$\boldsymbol{x} + \boldsymbol{y} = \sum_{i=1}^{n} (x_i + y_i) \boldsymbol{e}_i \text{ and } \alpha \boldsymbol{x} = \sum_{i=1}^{n} \alpha x_i \boldsymbol{e}_i.$$

i.e., *the coordinates of a sum of vectors are obtained by addition of the corresponding coordinates of the summands and the product of a vector by a number α is obtained by multiplying all the coordinates of the vector by α.*

4. Let the vectors

$$\boldsymbol{x}_k = \sum_{i=1}^{n} x_{ik} \boldsymbol{e}_i$$

be linearly dependent, i.e.,

$$\sum_{k=1}^{m} c_k \boldsymbol{x}_k = \boldsymbol{o}, \tag{5}$$

where at least one of the numbers $c_1, c_2, \ldots, c_m$ is not equal to zero.

If a vector is the null vector, then all its components are zero. Hence the vector equation (5) is equivalent to the following system of scalar equations:

$$\left.\begin{aligned} c_1 x_{11} + c_2 x_{12} + \cdots + c_m x_{1m} &= 0 \\ c_1 x_{21} + c_2 x_{22} + \cdots + c_m x_{2m} &= 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ c_1 x_{n1} + c_2 x_{n2} + \cdots + c_m x_{nm} &= 0. \end{aligned}\right\} \tag{6}$$

As is well known, this system of homogeneous linear equations for $c_1, c_2, \ldots, c_m$ has a non-zero solution if and only if the rank of the coefficient matrix is less than the number of unknowns, i.e., less than m. A necessary and sufficient condition for the independence of the vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ is, therefore, that this rank should be m.

Thus, the following theorem holds:

Theorem 1: *In order that the vectors $x_1, x_2, \ldots, x_m$ be linearly independent it is necessary and sufficient that the rank r of the matrix formed from the coordinates of these vectors in an arbitrary basis*

$$\begin{Vmatrix} x_{11} & x_{12} & \ldots & x_{1m} \\ x_{21} & x_{22} & \ldots & x_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_{n1} & x_{n2} & \ldots & x_{nm} \end{Vmatrix} \tag{7}$$

be equal to m, i.e., to the number of vectors.

Note. The linear independence of the vectors $x_1, x_2, \ldots, x_m$ means that the columns of the matrix (7) are linearly independent, since the k-th column consists of the coordinates of x_k $(k = 1, 2, \ldots, m)$. By the theorem, therefore, if the columns of a matrix are linearly independent, then the rank of the matrix is equal to the number of columns. Hence it follows that in an arbitrary rectangular matrix the maximal number of linearly independent columns is equal to the rank of the matrix. Moreover, if we transpose the matrix, i.e., change the rows into columns and the columns into rows, then the rank obviously remains unchanged. Hence *in a rectangular matrix the number of linearly independent columns is always equal to the number of linearly independent rows and equal to the rank of the matrix.*[4]

5. If in an n-dimensional space a basis $e_1, e_2, \ldots, e_n$ has been chosen, then to every vector x there corresponds uniquely the column $x = (x_1, x_2, \ldots, x_n)$, where $x_1, x_2, \ldots, x_n$ are the coordinates of x in the given basis. Thus, the choosing of a basis establishes a one-to-one correspondence between the vectors of an arbitrary n-dimensional vector space $\boldsymbol{R}$ and the vectors of the n-dimensional number space $\boldsymbol{R}'$ considered in Example 2. Here the sum of vectors in $\boldsymbol{R}$ corresponds to the sum of the corresponding vectors of $\boldsymbol{R}'$. The analogous correspondence holds for the product of a vector by a number α of F. In other words, an arbitrary n-dimensional vector space is *isomorphic* to the n-dimensional number space, and therefore *all vector spaces of the same number n of dimensions over the same number field* F *are isomorphic.* This means that to within isomorphism there exists only one n-dimensional vector space for a given number field.

[4] This proposition follows from Theorem 1, in the proof of which we have started from the well-known property of a system of linear homogeneous equations: a non-zero solution exists only when the rank of the coefficient matrix is less than the number of unknowns. For a proof of Theorem 1 independent of this property, see § 5.

The reader may ask why we have introduced an 'abstract' n-dimensional space if it coincides to within isomorphism with the n-dimensional number space. Indeed, we could have defined a vector as a system of n numbers given in a definite order and could have introduced the operations on these vectors in the very way it was done in Example 2. But we would then have mixed up properties of vectors that do not depend on the choice of a basis with properties of a particular basis. For example, the fact that all the coordinates of a vector are zero is a property of the vector itself; it does not depend on the choice of basis. But the equality of all its coordinates is not a property of the vector itself, because it disappears under a change of basis. *The axiomatic definition of a vector space immediately singles out the properties of vectors that do not depend on the choice of a basis.*

§ 2. A Linear Operator Mapping an n-Dimensional Space into an m-Dimensional Space

1. We consider a linear transformation

$$\left.\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ y_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n, \end{aligned}\right\} \tag{8}$$

whose coefficients belong to the number field F as well as two vector spaces over F: an n-dimensional space $\boldsymbol{R}$ and an m-dimensional space $\boldsymbol{S}$. We choose a basis $e_1, e_2, \ldots, e_n$ in $\boldsymbol{R}$ and a basis $g_1, g_2, \ldots, g_m$ in $\boldsymbol{S}$. Then the transformation (8) associates with every vector $x = \sum_{i=1}^{n} x_i e_i$ of $\boldsymbol{R}$ a certain vector $y = \sum_{k=1}^{m} y_k g_k$ of $\boldsymbol{S}$, i.e., the transformation (8) determines a certain operator $\boldsymbol{A}$ that sets up a correspondence between the vector x and the vector y : $y = Ax$. It is easy to see that this operator $\boldsymbol{A}$ has the property of linearity, which we formulate as follows:

Definition 5: *An operator $\boldsymbol{A}$ mapping $\boldsymbol{R}$ into $\boldsymbol{S}$, i.e., associating with every vector x of $\boldsymbol{R}$ a certain vector $y = Ax$ of $\boldsymbol{S}$ is called linear if for arbitrary x_1, x_2 of $\boldsymbol{R}$ and α of* F

$$\boldsymbol{A}(x_1 + x_2) = \boldsymbol{A}x_1 + \boldsymbol{A}x_2, \quad \boldsymbol{A}(\alpha x_1) = \alpha \boldsymbol{A}x_1. \tag{9}$$

Thus, the transformation (8), for a given basis in $\boldsymbol{R}$ and a given basis in $\boldsymbol{S}$, determines a linear operator mapping $\boldsymbol{R}$ into $\boldsymbol{S}$.

We shall now show the converse, i.e., that for an arbitrary linear operator $\boldsymbol{A}$ mapping $\boldsymbol{R}$ into $\boldsymbol{S}$ and arbitrary bases $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ in $\boldsymbol{R}$ and $\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_m$ in $\boldsymbol{S}$, there exists a rectangular matrix with elements in F

$$\begin{Vmatrix} a_{11} & a_{12} \ldots a_{1n} \\ a_{21} & a_{22} \ldots a_{2n} \\ \ldots\ldots\ldots\ldots \\ a_{m1} & a_{m2} \ldots a_{mn} \end{Vmatrix} \tag{10}$$

such that the linear transformation (8) formed by means of this matrix expresses the coordinates of the transformed vector $\boldsymbol{y} = \boldsymbol{Ax}$ in terms of the coordinates of the original vector $\boldsymbol{x}$.

Let us, in fact, apply the operator $\boldsymbol{A}$ to the basis vector $\boldsymbol{e}_k$ and let the coordinates in the basis $\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_m$ of the vector $\boldsymbol{Ae}_k$ thus obtained be denoted by $a_{1k}, a_{2k}, \ldots, a_{mk}$ $(k = 1, 2, \ldots, n)$:

$$\boldsymbol{Ae}_k = \sum_{i=1}^{m} a_{ik}\boldsymbol{g}_i \qquad (k = 1, 2, \ldots, n). \tag{11}$$

Multiplying both sides of (11) by x_k and summing from 1 to n, we obtain

$$\sum_{k=1}^{n} x_k \boldsymbol{Ae}_k = \sum_{i=1}^{m} \Big(\sum_{k=1}^{n} a_{ik}x_k\Big)\boldsymbol{g}_i;$$

hence

$$\boldsymbol{y} = \boldsymbol{Ax} = \boldsymbol{A}\Big(\sum_{k=1}^{n} x_k\boldsymbol{e}_k\Big) = \sum_{k=1}^{n} x_k\boldsymbol{Ae}_k = \sum_{i=1}^{m} y_i\boldsymbol{g}_i,$$

where

$$y_i = \sum_{k=1}^{n} a_{ik}x_k \qquad (i = 1, 2, \ldots, m),$$

and this is what we had to show.

Thus, for given bases of $\boldsymbol{R}$ and $\boldsymbol{S}$: *to every linear operator $\boldsymbol{A}$ mapping $\boldsymbol{R}$ into $\boldsymbol{S}$ there corresponds a rectangular matrix of dimension $m \times n$ and, conversely, to every such matrix there corresponds a linear operator mapping $\boldsymbol{R}$ into $\boldsymbol{S}$.*

Here, in the matrix A corresponding to the operator $\boldsymbol{A}$, *the k-th column consists of the coordinates of the vector $\boldsymbol{Ae}_k$* $(k = 1, 2, \ldots, n)$.

We denote by $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_m)$ the coordinate columns of the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. Then the vector equation

$$\boldsymbol{y} = \boldsymbol{Ax}$$

corresponds to the matrix equation

$$y = Ax,$$

which is the matrix form of the transformation (8).

Example. We consider the set of all polynomials in t of degree $\leqq n - 1$ with coefficients in F. This set forms an n-dimensional vector space $\boldsymbol{R}_n$ (see Example 4., p. 52). Similarly, the polynomials in t of degree $\leqq n - 2$ with coefficients in F form a space $\boldsymbol{R}_{n-1}$. The differentiation operator $\frac{d}{dt}$ associates with every polynomial of $\boldsymbol{R}_n$ a certain polynomial in $\boldsymbol{R}_{n-1}$. Thus, this operator maps $\boldsymbol{R}_n$ into $\boldsymbol{R}_{n-1}$. The differentiation operator is linear, since

$$\frac{d}{dt}[\varphi(t) + \psi(t)] = \frac{d\varphi(t)}{dt} + \frac{d\psi(t)}{dt}, \quad \frac{d}{dt}[\alpha\varphi(t)] = \alpha\frac{d\varphi(t)}{dt}.$$

In $\boldsymbol{R}_n$ and $\boldsymbol{R}_{n-1}$ we choose bases consisting of powers of t:

$$t^0 = 1, t, \ldots, t^{n-1} \qquad \text{and} \qquad t^0 = 1, t, \ldots, t^{n-2}.$$

Using formulas (11), we construct the rectangular matrix of dimension $(n - 1 \times n)$ corresponding to the differentiation operator $\frac{d}{dt}$ in these bases:

$$\begin{Vmatrix} 0 & 1 & 0 \ldots & 0 \\ 0 & 0 & 2 \ldots & 0 \\ \ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots \\ 0 & 0 & 0 \ldots & n-1 \end{Vmatrix}.$$

§ 3. Addition and Multiplication of Linear Operators

1. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two linear operators mapping $\boldsymbol{R}$ into $\boldsymbol{S}$ and let the corresponding matrices be

$$A = \|a_{ik}\|, \quad B = \|b_{ik}\| \qquad (i = 1, 2, \ldots, m;\ k = 1, 2, \ldots, n).$$

Definition 6: *The sum of the operators $\boldsymbol{A}$ and $\boldsymbol{B}$ is the operator $\boldsymbol{C}$ defined by the equation*[5]

$$\boldsymbol{Cx} = \boldsymbol{Ax} + \boldsymbol{Bx} \quad (\boldsymbol{x} \in \boldsymbol{R}). \tag{12}$$

On the basis of this definition it is easy to verify that the sum $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$ of the linear operators $\boldsymbol{A}$ and $\boldsymbol{B}$ is itself a linear operator. Furthermore,

$$\boldsymbol{Ce}_k = \boldsymbol{Ae}_k + \boldsymbol{Be}_k = \sum_{i=1}^{n} (a_{ik} + b_{ik})\,\boldsymbol{e}_k.$$

[5] $\boldsymbol{x} \in \boldsymbol{R}$ means that the element $\boldsymbol{x}$ belongs to the set $\boldsymbol{R}$. It is assumed that (12) holds for arbitrary $\boldsymbol{x}$ in $\boldsymbol{R}$.

Hence it follows that the operator $\boldsymbol{C}$ corresponds to the matrix $C = \| c_{ik} \|$, where $c_{ik} = a_{ik} + b_{ik}$ $(i = 1, 2, \ldots, m;\ k = 1, 2, \ldots, n)$, i.e., the operator $\boldsymbol{C}$ corresponds to the matrix

$$C = A + B. \tag{13}$$

We would come to the same conclusion starting from the matrix equation

$$Cx = Ax + Bx \tag{14}$$

(x is the coordinate column of the vector $\boldsymbol{x}$) corresponding to the vector equation (12). Since x is an arbitrary column, (13) follows from (14).

2. Let $\boldsymbol{R}$, $\boldsymbol{S}$, and $\boldsymbol{T}$ be three vector spaces of dimension q, n, and m, and let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two linear operators, of which $\boldsymbol{B}$ maps $\boldsymbol{R}$ into $\boldsymbol{S}$ and $\boldsymbol{A}$ maps $\boldsymbol{S}$ into $\boldsymbol{T}$; in symbols:

$$\boldsymbol{R} \overset{\boldsymbol{B}}{\to} \boldsymbol{S} \overset{\boldsymbol{A}}{\to} \boldsymbol{T}.$$

Definition 7. *The product of the operators $\boldsymbol{A}$ and $\boldsymbol{B}$ is the operator $\boldsymbol{C}$ for which*

$$\boldsymbol{Cx} = \boldsymbol{A}(\boldsymbol{Bx}) \quad (\boldsymbol{x} \in \boldsymbol{R}). \tag{15}$$

holds for every $\boldsymbol{x}$ of $\boldsymbol{R}$.

The operator $\boldsymbol{C}$ maps $\boldsymbol{R}$ into $\boldsymbol{T}$:

$$\boldsymbol{R} \overset{\boldsymbol{C} = \boldsymbol{AB}}{\longrightarrow} \boldsymbol{T}.$$

From the linearity of the operators $\boldsymbol{A}$ and $\boldsymbol{B}$ follows the linearity of $\boldsymbol{C}$. We choose arbitrary bases in $\boldsymbol{R}$, $\boldsymbol{S}$, and $\boldsymbol{T}$ and denote by A, B, and C the matrices corresponding, in this choice of basis, to the operators $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$. Then the vector equations

$$\boldsymbol{z} = \boldsymbol{Ay},\ \boldsymbol{y} = \boldsymbol{Bx},\ \boldsymbol{z} = \boldsymbol{Cx} \tag{16}$$

correspond to the matrix equations:

$$z = Ay,\ y = Bx,\ z = Cx,$$

where x, y, z are the coordinate columns of the vectors $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$. Hence

$$Cx = A(Bx) = (AB)x$$

and as the column x is arbitrary

$$C = AB. \tag{17}$$

Thus, the product $\boldsymbol{C} = \boldsymbol{AB}$ of the operators $\boldsymbol{A}$ and $\boldsymbol{B}$ corresponds to the matrix $C = \| c_{ij} \|$ $(i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, q)$, which is the product of the matrices A and B.

We leave it to the reader to show that the operator[6]

$$\boldsymbol{C} = \alpha\boldsymbol{A} \qquad (\alpha \in \mathrm{F})$$

corresponds to the matrix

$$C = \alpha A.$$

Thus we see that in Chapter I the operations on matrices were so defined that the sum $\boldsymbol{A} + \boldsymbol{B}$, the product $\boldsymbol{AB}$, and the product $\alpha\boldsymbol{A}$ correspond to the matrices $A + B$, AB, and αA, respectively, where A and B are the matrices corresponding to the operators $\boldsymbol{A}$ and $\boldsymbol{B}$, and α is a number of F.

§ 4. Transformation of Coordinates

1. In an n-dimensional vector space we consider two bases: $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ (the 'old' basis) and $\boldsymbol{e}_1^*, \boldsymbol{e}_2^*, \ldots, \boldsymbol{e}_n^*$ the 'new' basis).

The mutual disposition of the basis vectors is determined if the coordinates of the vectors of the basis are given relative to the other basis.

We set

$$\left.\begin{aligned} \boldsymbol{e}_1^* &= t_{11}\boldsymbol{e}_1 + t_{21}\boldsymbol{e}_2 + \cdots + t_{n1}\boldsymbol{e}_n \\ \boldsymbol{e}_2^* &= t_{12}\boldsymbol{e}_1 + t_{22}\boldsymbol{e}_2 + \cdots + t_{n2}\boldsymbol{e}_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \boldsymbol{e}_n^* &= t_{1n}\boldsymbol{e}_1 + t_{2n}\boldsymbol{e}_2 + \cdots + t_{nn}\boldsymbol{e}_n \end{aligned}\right\} \tag{18}$$

or in abbreviated form,

$$\boldsymbol{e}_k^* = \sum_{i=1}^{n} t_{ik}\boldsymbol{e}_i \quad (k = 1, 2, \ldots, n). \tag{18'}$$

We shall now establish the connection between the coordinates of one and the same vector in the two different bases.

Let $x_1, x_2, \ldots, x_n$ and $x_1^*, x_2^*, \ldots, x_n^*$ be the coordinates of the vector $\boldsymbol{x}$ relative to the 'old' and the 'new' bases, respectively:

$$\boldsymbol{x} = \sum_{i=1}^{n} x_i\boldsymbol{e}_i = \sum_{k=1}^{n} x_k^*\boldsymbol{e}_k^*. \tag{19}$$

In (19) we substitute for the vectors $\boldsymbol{e}_k^*$ the expressions given for them in (18). We obtain:

[6] I.e., the operator for which $\boldsymbol{Cx} = \alpha\boldsymbol{Ax}$ $(\boldsymbol{x} \in \boldsymbol{R})$.

$$x=\sum_{k=1}^{n} x_k^* \sum_{i=1}^{n} t_{ik}\, e_i=\sum_{i=1}^{n}\Big(\sum_{k=1}^{n} t_{ik}\, x_k^*\Big)\, e_i.$$

Comparing this with (19) and bearing in mind that the coordinates of a vector are uniquely determined when the vector and the basis are given, we find:

$$x_i=\sum_{k=1}^{n} t_{ik}\, x_k^* \quad (i=1, 2, \ldots, n). \tag{20}$$

or in explicit form:

$$\left.\begin{aligned} x_1 &= t_{11}\, x_1^* + t_{12} x_2^* + \cdots + t_{1n} x_n^* \\ x_2 &= t_{21}\, x_1^* + t_{22} x_2^* + \cdots + t_{2n} x_n^* \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n &= t_{n1}\, x_1^* + t_{n2} x_2^* + \cdots + t_{nn} x_n^*. \end{aligned}\right\} \tag{21}$$

Formulas (21) determine the transformation of the coordinates of a vector on transition from one basis to another. They express the 'old' coordinates in terms of the 'new' ones. The matrix

$$T=\| t_{ik} \|_1^n \tag{22}$$

is called the *matrix of the coordinate transformation* or the *transforming matrix*. Its k-th column consists of the 'old' coordinates of the k-th 'new' basis vector. This follows from formulas (18) or immediately from (21) if we set in the latter $x_k^*=1$, $x_i^*=0$ for $i\neq k$.

Note that the matrix T is non-singular, i.e.,

$$|\, T\,| \neq 0. \tag{23}$$

For when we set in (21) $x_1=x_2=\ldots=x_n=0$, we obtain a system of n linear homogeneous equations in the n unknowns $x_1^*, x_2^*, \ldots, x_n^*$ with determinant $|\, T\,|$. This system can only have the zero solution $x_1^*=0$, $x_2^*=0$, $\ldots$, $x_n^*=0$, since otherwise (19) would imply a linear dependence among the vectors $e_1^*, e_2^*, \ldots, e_n^*$. Therefore $|\, T\,| \neq 0$.[7]

We now introduce the column matrices $x=(x_1, x_2, \ldots, x_n)$ and $x^*=(x_1^*, x_2^*, \ldots, x_n^*)$. Then the formulas (21) for the coordinate transformation can be written in the form of the following matrix equation:

$$x=Tx^*. \tag{24}$$

Multiplying both sides of this equation by T^{-1}, we obtain the expression for the inverse transformation

$$x^*=T^{-1}x. \tag{25}$$

[7] The inequality (23) also follows from Theorem 1 (p. 54), because the elements of T are the 'old' coordinates of the linearly independent vectors $e_1^*, e_2^*, \ldots, e_n^*$.

§ 5. Equivalent Matrices. The Rank of an Operator. Sylvester's Inequality

1. Let $\boldsymbol{R}$ and $\boldsymbol{S}$ be two vector spaces of dimension n and m, respectively, over the number field F and let $\boldsymbol{A}$ be a linear operator mapping $\boldsymbol{R}$ into $\boldsymbol{S}$. In the present section we shall make clear how the matrix A corresponding to the given linear operator $\boldsymbol{A}$ changes when the bases in $\boldsymbol{R}$ and $\boldsymbol{S}$ are changed.

We choose arbitrary bases $e_1, e_2, \ldots, e_n$ in $\boldsymbol{R}$ and $g_1, g_2, \ldots, g_m$ in $\boldsymbol{S}$. In these bases the operator $\boldsymbol{A}$ corresponds to a matrix $A=\| a_{ik} \|$ $(i=1, 2, 3, \ldots, m;\ k=1, 2, \ldots, n)$. To the vector equation

$$y=Ax \tag{26}$$

there corresponds the matrix equation

$$y=Ax, \tag{27}$$

where x and y are the coordinate columns for the vectors x and y in the bases $e_1, e_2, \ldots, e_n$ and $g_1, g_2, \ldots, g_m$.

We now choose other bases $e_1^*, e_2^*, \ldots, e_n^*$ and $g_1^*, g_2^*, \ldots, g_m^*$ in $\boldsymbol{R}$ and $\boldsymbol{S}$. In the new bases we shall have x^*, y^*, A^* instead of x, y, A. Here

$$y^*=A^*x^*. \tag{28}$$

Let us denote by Q and N the non-singular square matrices of order n and m, respectively, that realize the coordinate tranformations in the spaces $\boldsymbol{R}$ and $\boldsymbol{S}$ on transition from the old bases to the new ones (see § 4):

$$x=Qx^*, \quad y=Ny^*. \tag{29}$$

Then we obtain from (27) and (29):

$$y^*=N^{-1}y=N^{-1}Ax=N^{-1}AQx^*. \tag{30}$$

Setting $P=N^{-1}$, we find from (28) and (30):

$$A^*=PAQ. \tag{31}$$

Definition 8: *Two rectangular matrices A and B of the same dimension are called equivalent if there exist two non-singular matrices P and Q such that*[8]

$$B=PAQ. \tag{32}$$

[8] If the matrices A and B are of dimension $m\times n$, then in (32) the square matrix P is of order m, and Q of order n. If the elements of the equivalent matrices A and B belong to some number field, then P and Q may be chosen such that their elements belong to the same number field.

From (31) it follows that two matrices corresponding to one and the same linear operator $\boldsymbol{A}$ for different choices of bases in $\boldsymbol{R}$ and $\boldsymbol{S}$ are always equivalent. It is easy to see that, conversely, if a matrix A corresponds to the operator $\boldsymbol{A}$ for certain bases in $\boldsymbol{R}$ and $\boldsymbol{S}$, and if a matrix B is equivalent to A, then it corresponds to the same linear operator for certain other bases in $\boldsymbol{R}$ and $\boldsymbol{S}$.

Thus, to every linear operator mapping $\boldsymbol{R}$ into $\boldsymbol{S}$ there corresponds a class of equivalent matrices with elements in F.

2. The following theorem establishes a criterion for the equivalence of two matrices:

Theorem 2: *Two rectangular matrices of the same dimension are equivalent if and only if they have the same rank.*

Proof. The condition is *necessary*. When a rectangular matrix is multiplied by an arbitrary non-singular square matrix (on the right or left), then its rank does not change (see Chapter I, p. 17). Therefore it follows from (32) that

$$r_A = r_B .$$

The condition is *sufficient*. Let A be a rectangular matrix of dimension $m \times n$. It determines a linear operator $\boldsymbol{A}$ mapping the space $\boldsymbol{R}$ with the basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ into the space $\boldsymbol{S}$ with the basis $\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_m$. Let r denote the number of linearly independent vectors among the vectors $\boldsymbol{Ae}_1, \boldsymbol{Ae}_2, \ldots, \boldsymbol{Ae}_n$. Without loss of generality we may assume that the vectors $\boldsymbol{Ae}_1, \boldsymbol{Ae}_2, \ldots, \boldsymbol{Ae}_r$ are linearly independent[9] and that the remaining $\boldsymbol{Ae}_{r+1}, \boldsymbol{Ae}_{r+2}, \ldots, \boldsymbol{Ae}_n$ are expressed linearly in terms of them:

$$\boldsymbol{Ae}_k = \sum_{j=1}^{r} c_{kj} \boldsymbol{Ae}_j \qquad (k = r+1, \ldots, n). \tag{33}$$

We define a new basis in $\boldsymbol{R}$ as follows:

$$\boldsymbol{e}_i^* = \begin{cases} \boldsymbol{e}_i & (i = 1, 2, \ldots, r), \\ \boldsymbol{e}_i - \sum_{j=1}^{r} c_{ij} \boldsymbol{e}_j & (i = r+1, \ldots, n). \end{cases} \tag{34}$$

Then by (33),

$$\boldsymbol{Ae}_k^* = \boldsymbol{o} \qquad (k = r+1, \ldots, n). \tag{35}$$

Next, we set

$$\boldsymbol{Ae}_j^* = \boldsymbol{g}_j^* \qquad (j = 1, 2, \ldots, r). \tag{36}$$

[9] This can be achieved by a suitable numbering of the basis vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$.

The vectors $\boldsymbol{g}_1^*, \boldsymbol{g}_2^*, \ldots, \boldsymbol{g}_r^*$ are linearly independent. We supplement them with suitable vectors $\boldsymbol{g}_{r+1}^*, \boldsymbol{g}_{r+2}^*, \ldots, \boldsymbol{g}_m^*$ to obtain a basis $\boldsymbol{g}_1^*, \boldsymbol{g}_2^*, \ldots, \boldsymbol{g}_m^*$ of $\boldsymbol{S}$.

The matrix corresponding to the same operator $\boldsymbol{A}$ in the new bases $\boldsymbol{e}_1^*, \boldsymbol{e}_2^*, \ldots, \boldsymbol{e}_n^*$, $\boldsymbol{g}_1^*, \boldsymbol{g}_2^*, \ldots, \boldsymbol{g}_m^*$ has now, by (35) and (36), the form

$$I_r = \begin{Vmatrix} \overbrace{1}^{r} & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \end{Vmatrix}. \tag{37}$$

Along the main diagonal of I_r, starting at the top, there are r units; all the remaining elements of I_r are zeros. Since the matrices A and I_r correspond to one and the same operator $\boldsymbol{A}$, they are equivalent. As we have proved, equivalent matrices have the same rank. Hence the rank of the original matrix A is r.

We have shown that an arbitrary rectangular matrix of rank r is equivalent to the 'canonical' matrix I_r. But I_r is completely determined by specifying its dimensions $m \times n$ and the number r. Therefore all rectangular matrices of given dimension $m \times n$ and of given rank r are equivalent to one and the same matrix I_r and consequently to each other. This completes the proof of the theorem.

3. Let $\boldsymbol{A}$ be a linear operator mapping an n-dimensional space $\boldsymbol{R}$ into an n-dimensional space $\boldsymbol{S}$. The set of all vectors of the form $\boldsymbol{Ax}$, where $\boldsymbol{x} \in \boldsymbol{R}$, forms a vector space.[10] This space will be denoted by $\boldsymbol{AR}$; it is part of the space $\boldsymbol{S}$ or, as we shall say, is a *subspace* of $\boldsymbol{S}$.

Together with the subspace $\boldsymbol{AR}$ of $\boldsymbol{S}$ we consider the set of all vectors $\boldsymbol{x} \in \boldsymbol{R}$ that satisfy the equation

$$\boldsymbol{Ax} = \boldsymbol{o} \tag{38}$$

These vectors also form a subspace of $\boldsymbol{R}$, which we shall denote by $\boldsymbol{N_A}$.

[10] The set of vectors of the form $\boldsymbol{Ax}$ ($\boldsymbol{x} \in \boldsymbol{R}$) satisfies the postulates 1.-7. of § 1, because the sum of two such vectors and the product of such a vector by a number are also vectors of this form.

Definition 9: *If a linear operator $\boldsymbol{A}$ maps $\boldsymbol{R}$ into $\boldsymbol{S}$, then the dimension r of the space $\boldsymbol{AR}$ is called the rank of $\boldsymbol{A}$,*[11] *and the dimension d of the space $\boldsymbol{N_A}$ consisting of all vectors $\boldsymbol{x} \in \boldsymbol{R}$ that satisfy the condition* (38) *is called the defect, or nullity, of $\boldsymbol{A}$.*

Among all the equivalent rectangular matrices that describe a given operator $\boldsymbol{A}$ in distinct bases there occurs the canonical matrix I_r (see (37)). We denote the corresponding bases of $\boldsymbol{R}$ and $\boldsymbol{S}$ by $\boldsymbol{e_1^*}, \boldsymbol{e_2^*}, \ldots, \boldsymbol{e_n^*}$ and $\boldsymbol{g_1^*}, \boldsymbol{g_2^*}, \ldots, \boldsymbol{g_m^*}$. Then

$$\boldsymbol{Ae_1^* = g_1^*}, \ \ldots, \ \boldsymbol{Ae_r^* = g_r^*}, \quad \boldsymbol{Ae_{r+1}^* = \ldots = Ae_n^* = o}.$$

From the definition of $\boldsymbol{AR}$ and $\boldsymbol{N_A}$ it follows that the vectors $\boldsymbol{g_1^*}, \boldsymbol{g_2^*}, \ldots, \boldsymbol{g_r^*}$ form a basis of $\boldsymbol{AR}$ and that the vectors $\boldsymbol{e_{r+1}^*}, \boldsymbol{e_{r+2}^*}, \ldots, \boldsymbol{e_n^*}$ form a basis of $\boldsymbol{N_A}$. Hence it follows that r is the rank of the operator $\boldsymbol{A}$ and that

$$d = n - r. \tag{39}$$

If A is an arbitrary matrix corresponding to $\boldsymbol{A}$, then it is equivalent to I_r and therefore has the same rank r. Thus, *the rank of an operator $\boldsymbol{A}$ coincides with the rank of the rectangular matrix A*

$$A = \begin{Vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{Vmatrix}$$

determined by $\boldsymbol{A}$ in arbitrary bases $\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_n} \in \boldsymbol{R}$ and $\boldsymbol{g_1}, \boldsymbol{g_2}, \ldots, \boldsymbol{g_m} \in \boldsymbol{S}$.

The columns of A are formed by the coordinate vectors $\boldsymbol{A_1 e_1}, \ldots, \boldsymbol{A_n e_n}$. Since it follows from $\boldsymbol{x} = \sum_{i=1}^{n} \boldsymbol{x_i e_i}$ that $\boldsymbol{Ax} = \sum_{i=1}^{n} \boldsymbol{x_i Ae_i}$, the rank of $\boldsymbol{A}$, i.e., the dimension of $\boldsymbol{RA}$, is equal to the maximal number of linearly independent vectors among $\boldsymbol{Ae_1}, \boldsymbol{Ae_2}, \ldots, \boldsymbol{Ae_n}$. Thus:

The rank of a matrix coincides with the number of linearly independent columns of the matrix.

Since under transposition the rows of a matrix become its columns and the rank remains unchanged:

[11] The dimension of the space $\boldsymbol{AR}$ never exceeds the dimension of $\boldsymbol{R}$, so that $r \leqq n$. This follows from the fact that the equation $\boldsymbol{x} = \sum_{i=1}^{n} \boldsymbol{x_i e_i}$ (where $\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_n}$ is a basis of $\boldsymbol{R}$) implies the equation $\boldsymbol{Ax} = \sum_{i=1}^{n} \boldsymbol{x_i Ae_i}$.

The number of linearly independent rows of a matrix is also equal to the rank of the matrix.[12]

4. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two linear operators and let $\boldsymbol{C} = \boldsymbol{AB}$ be their product. Suppose that the operator $\boldsymbol{B}$ maps $\boldsymbol{R}$ into $\boldsymbol{S}$ and that the operator $\boldsymbol{A}$ maps $\boldsymbol{S}$ into $\boldsymbol{T}$. Then the operator $\boldsymbol{C}$ maps $\boldsymbol{R}$ into $\boldsymbol{T}$:

$$\boldsymbol{R} \xrightarrow{\boldsymbol{B}} \boldsymbol{S} \xrightarrow{\boldsymbol{A}} \boldsymbol{T}, \quad \boldsymbol{R} \xrightarrow{\boldsymbol{C}} \boldsymbol{T}.$$

We introduce the matrices A, B, C corresponding to $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ in some choice of bases in $\boldsymbol{R}$, $\boldsymbol{S}$, and $\boldsymbol{T}$. Then the matrix equation $C = AB$ will correspond to the operator equation $\boldsymbol{C} = \boldsymbol{AB}$.

We denote by r_A, r_B, r_C the ranks of the operators $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ or, what is the same, of the matrices A, B, C. These numbers determine the dimensions of the subspaces $\boldsymbol{AS}$, $\boldsymbol{BR}$, $\boldsymbol{A(BR)}$. Since $\boldsymbol{BR} \subset \boldsymbol{S}$, we have $\boldsymbol{A(BR)} \subset \boldsymbol{AS}$.[13] Moreover, the dimension of $\boldsymbol{A(BR)}$ cannot exceed the dimension of $\boldsymbol{BR}$.[14] Therefore

$$r_C \leqq r_A, \quad r_C \leqq r_B.$$

These inequalities were obtained in Chapter I, § 2 from the formula for the minors of a product of two matrices.

Let us regard $\boldsymbol{A}$ as an operator mapping $\boldsymbol{BR}$ into $\boldsymbol{T}$. Then the rank of this operator is equal to the dimension of the space $\boldsymbol{A(BR)}$, i.e., to r_C. Therefore, by applying (39) we obtain

$$r_C = r_B - d_1, \tag{40}$$

where d_1 is the maximal number of linearly independent vectors of $\boldsymbol{BR}$ that satisfy the equation

$$\boldsymbol{Ax} = \boldsymbol{o}. \tag{41}$$

But all the solutions of this equation that belong to $\boldsymbol{S}$ form a subspace of dimension d, where

$$d = n - r_A \tag{42}$$

is the defect of the operator $\boldsymbol{A}$ mapping $\boldsymbol{S}$ into $\boldsymbol{T}$. Since $\boldsymbol{BR} \subset \boldsymbol{S}$,

$$d_1 \leqq d. \tag{43}$$

From (40), (42), and (43) we find:

$$r_A + r_B - n \leqq r_C.$$

[12] In § 1 we reached these conclusions on the basis of other arguments (see p. 54).

[13] $\boldsymbol{R} \subset \boldsymbol{S}$ means that the set $\boldsymbol{R}$ forms part of the set $\boldsymbol{S}$.

[14] See Footnote 11.

Thus we have obtained *Sylvester's inequality* for the rank of the product of two rectangular matrices A and B of dimensions $m \times n$ and $n \times q$:

$$r_A + r_B - n \leqq r_{AB} \leqq \min(r_A, r_B). \tag{44}$$

§ 6. Linear Operators Mapping an n-Dimensional Space into Itself

1. A linear operator mapping the n-dimensional vector space $\boldsymbol{R}$ into itself (here $\boldsymbol{R} \equiv \boldsymbol{S}$ and $n = m$) will be referred to simply as a *linear operator in* $\boldsymbol{R}$.

The sum of two linear operators in $\boldsymbol{R}$ and the product of such an operator by a number are also linear operators in $\boldsymbol{R}$. Multiplication of two such operators is always feasible, and this product is also a linear operator in $\boldsymbol{R}$. Hence the linear operators in $\boldsymbol{R}$ form a ring.[15] This ring has an identity operator, namely the operator $\boldsymbol{E}$ for which

$$\boldsymbol{E}x = x \qquad (x \in \boldsymbol{R}). \tag{45}$$

For every operator $\boldsymbol{A}$ in $\boldsymbol{R}$ we have

$$\boldsymbol{E}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{E} = \boldsymbol{A}.$$

If $\boldsymbol{A}$ is a linear operator in $\boldsymbol{R}$, then the powers $\boldsymbol{A}^2 = \boldsymbol{A}\boldsymbol{A}$, $\boldsymbol{A}^3 = \boldsymbol{A}\boldsymbol{A}\boldsymbol{A}$, and in general $\boldsymbol{A}^m = \underbrace{\boldsymbol{A}\boldsymbol{A}\cdots\boldsymbol{A}}_{m \text{ times}}$ have a meaning. We set $\boldsymbol{A}^0 = \boldsymbol{E}$. Then it is easy to see that for all non-negative integers p and q we have

$$\boldsymbol{A}^p \boldsymbol{A}^q = \boldsymbol{A}^{p+q}.$$

Let $f(t) = \alpha_0 t^m + \alpha_1 t^{m-1} + \cdots + \alpha_{m-1} t + \alpha_m$ be a polynomial in a scalar argument t with coefficients in the field F. Then we set:

$$f(\boldsymbol{A}) = \alpha_0 \boldsymbol{A}^m + \alpha_1 \boldsymbol{A}^{m-1} + \cdots + \alpha_{m-1}\boldsymbol{A} + \alpha_m \boldsymbol{E}. \tag{46}$$

Here $f(\boldsymbol{A})g(\boldsymbol{A}) = g(\boldsymbol{A})f(\boldsymbol{A})$ for any two polynomials $f(t)$ and $g(t)$.

Let

$$y = \boldsymbol{A}x \qquad (x, y \in \boldsymbol{R}).$$

We denote by $x_1, x_2, \ldots, x_n$ the coordinates of the vector x in an arbitrary basis $e_1, e_2, \ldots, e_n$ and by $y_1, y_2, \ldots, y_n$ the coordinates of y in the same basis. Then

$$y_i = \sum_{k=1}^{n} a_{ik} x_k \qquad (i = 1, 2, \ldots, n). \tag{47}$$

[15] This ring is in fact an algebra. See Chapter I, p. 17.

In the basis $e_1, e_2, \ldots, e_n$ the linear operator $\boldsymbol{A}$ corresponds to a square matrix $A = \| a_{ik} \|_1^n$.[16] We remind the reader (see § 2.) that in the k-th column of this matrix are to be found the coordinates of the vector $\boldsymbol{A}e_k$ $(k = 1, 2, \ldots, n)$. Introducing the coordinate columns $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$, we can write the transformation (47) in matrix form

$$y = Ax. \tag{48}$$

The sum and product of two operators $\boldsymbol{A}$ and $\boldsymbol{B}$ correspond to the sum and product of the corresponding square matrices $A = \| a_{ik} \|_1^n$ and $B = \| b_{ik} \|_1^n$. The product $a\boldsymbol{A}$ corresponds to the matrix aA. The identity operator $\boldsymbol{E}$ corresponds to the square unit matrix $E = \| \delta_{ik} \|_1^n$. Thus, the choice of a basis establishes *an isomorphism between the ring of linear operators in* $\boldsymbol{R}$ *and the ring of square matrices of order* n *with elements in* F. In this isomorphism the polynomial $f(\boldsymbol{A})$ corresponds to the matrix $f(A)$.

Let us consider, apart from the basis $e_1, e_2, \ldots, e_n$, another basis $e_1^*, e_2^*, \ldots, e_n^*$ of $\boldsymbol{R}$. Then, in analogy with (48), we have

$$y^* = A^* x^*, \tag{49}$$

where x^*, y^* are the column matrices formed from the coordinates of the vectors x, y in the basis $e_1^*, e_2^*, \ldots, e_n^*$ and $A^* = \| a_{ik}^* \|_1^n$ is the square matrix corresponding to the operator $\boldsymbol{A}$ in this basis. We rewrite in matrix form the formulas for the transformation of coordinates

$$x = Tx^*, \quad y = Ty^*. \tag{50}$$

Then from (48) and (50) we find:

$$y^* = T^{-1}ATx^*;$$

and a comparison with (49) gives:

$$A^* = T^{-1}AT. \tag{51}$$

Formula (51) is a special case of (31) on p. 61 (namely, $P = T^{-1}$ and $Q = T$).

Definition 10: *Two matrices* A *and* B *connected by the relation*

$$B = T^{-1}AT. \tag{51'}$$

where T *is a non-singular matrix, are called similar.*[17]

[16] See § 2 of this chapter. In this case the spaces $\boldsymbol{R}$ and $\boldsymbol{S}$ coincide; in the same way, the bases $e_1, e_2, \ldots, e_n$ and $g_1, g_2, \ldots, g_m$ of these spaces are identified.

Thus, we have shown that *two matrices corresponding to one and the same linear operator in* $\boldsymbol{R}$ *for distinct bases are similar* and the matrix T linking these matrices coincides with the matrix of the coordinate transformation in the transition from the first basis to the second (see (50)).

In other words, to a linear operator in $\boldsymbol{R}$ there corresponds a whole class of similar matrices; they represent the given operator in various bases.

In studying properties of a linear operator in $\boldsymbol{R}$, we are at the same time studying the matrix properties that are common to the whole class of similar matrices, that is, that remain unchanged, or invariant, under transition from a given matrix to a similar one.

We note at once that two similar matrices always have the same determinant. For it follows from (51′) that

$$|B| = |T|^{-1}|A||T| = |A|. \tag{52}$$

The equation $|B| = |A|$ is a necessary, but not a sufficient condition for the similarity of the matrices A and B.

In Chapter VI we shall establish a criterion for the similarity of two matrices, i.e., we shall give necessary and sufficient conditions for two square matrices of order n to be similar.

In accordance with (52) we may define the determinant $|A|$ of a linear operator $\boldsymbol{A}$ in $\boldsymbol{R}$ as the determinant of an arbitrary matrix corresponding to the given operator.

If $|\boldsymbol{A}| = 0$ $(\neq 0)$, then the operator $\boldsymbol{A}$ is called *singular* (*non-singular*). In accordance with this definition a singular (non-singular) operator corresponds to a singular (non-singular) matrix in any basis. For a singular operator:

1) There always exists a vector $\boldsymbol{x} \neq \boldsymbol{o}$ such that $\boldsymbol{Ax} = \boldsymbol{o}$;
2) $\boldsymbol{AR}$ is a proper part of $\boldsymbol{R}$.

For a non-singular operator:

1) $\boldsymbol{Ax} = \boldsymbol{o}$ implies that $\boldsymbol{x} = \boldsymbol{o}$;
2) $\boldsymbol{AR} \equiv \boldsymbol{R}$, i.e., the vectors of the form $\boldsymbol{Ax}$ $(\boldsymbol{x} \in \boldsymbol{R})$ fill out the whole space $\boldsymbol{R}$.

In other words, a linear operator in $\boldsymbol{R}$ is singular or non-singular depending on whether its defect is positive or zero.

[17] The matrix T can always be chosen such that its elements belong to the same basic number field F as those of A and B. It is easy to verify the three properties of similar matrices:

Reflexivity (a matrix A is always similar to itself);
Symmetry (if A is similar to B, then B is similar to A); and
Transitivity (if A is similar to B, and B to C, then A is similar to C).

§ 7. Characteristic Values and Characteristic Vectors of a Linear Operator

1. An important role in the study of the structure of a linear operator $\boldsymbol{A}$ in $\boldsymbol{R}$ is played by the vectors $\boldsymbol{x}$ for which

$$\boldsymbol{Ax} = \lambda \boldsymbol{x} \qquad (\lambda \in \mathrm{F}, \quad \boldsymbol{x} \neq \boldsymbol{o}) \tag{53}$$

Such vectors are called *characteristic vectors* and the numbers λ corresponding to them are called *characteristic values* or *characteristic roots* of the operator $\boldsymbol{A}$ (or of the matrix A).†

In order to find the characteristic values and characteristic vectors of an operator $\boldsymbol{A}$ we choose an arbitrary basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ in $\boldsymbol{R}$. Let $\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{e}_i$ and let $A = \| a_{ik} \|_1^n$ be the matrix corresponding to $\boldsymbol{A}$ in the basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$. Then if we equate the corresponding coordinates of the vectors on the left-hand and right-hand sides of (53), we obtain a system of scalar equations

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= \lambda x_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= \lambda x_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= \lambda x_n, \end{aligned}\right\} \tag{54}$$

which can also be written as

$$\left.\begin{aligned} (a_{11} - \lambda)x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + (a_{22} - \lambda)x_2 + \cdots + a_{2n}x_n &= 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + (a_{nn} - \lambda)x_n &= 0 \end{aligned}\right\} \tag{55}$$

Since the required vector must not be the null vector, at least one of its coordinates $x_1, x_2, \ldots, x_n$ must be different from zero.

In order that the system of linear homogeneous equations (55) should have a non-zero solution it is necessary and sufficient that the determinant of the system be zero:

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} - \lambda & \ldots & a_{2n} \\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} - \lambda \end{vmatrix} = 0. \tag{56}$$

† Other terms in use for the former are: *proper vector, latent vector, eigenvector.*
Other terms for the latter are: *proper value, latent value, latent root, latent number, characteristic number, eigenvalue,* etc.

The equation (56) is an algebraic equation of degree n in λ. Its coefficients belong to the same number field F as the elements of the matrix $A = \| a_{ik} \|_1^n$.

Equation (56) occurs in various problems of geometry, mechanics, astronomy, and physics and is known as the *characteristic equation* or the *secular equation*[18] of the matrix $A = \| a_{ik} \|_1^n$ (the left-hand side is called the *characteristic polynomial*).

Thus, every characteristic value λ of a linear operator $\boldsymbol{A}$ is a root of the characteristic equation (56). And conversely, if a number λ is a root of (56), then for this value λ the system (55) and hence (54) has a non-zero solution $x_1, x_2, \ldots, x_n$, i.e., to this number λ there corresponds a characteristic vector $\boldsymbol{x} = \sum x_i \boldsymbol{e}_i$ of the operator $\boldsymbol{A}$.

From what we have shown, it follows that every linear operator $\boldsymbol{A}$ in $\boldsymbol{R}$ has not more than n distinct characteristic values.

If F is the field of complex numbers, then every linear operator in $\boldsymbol{R}$ always has at least one characteristic vector in $\boldsymbol{R}$ corresponding to a characteristic value λ.[19] This follows from the fundamental theorem of algebra, according to which an algebraic equation (56) in the field of complex numbers always has at least one root.

Let us write (56) in explicit form

$$|A - \lambda E| \equiv (-\lambda)^n + S_1(-\lambda)^{n-1} + S_2(-\lambda)^{n-2} + \cdots + S_{n-1}(-\lambda) + S_n = 0. \quad (57)$$

It is easy to see that here

$$S_1 = \sum_{i=1}^{n} a_{ii}, \quad S_2 = \sum_{1 \leq i < k \leq n} A \begin{pmatrix} i & k \\ i & k \end{pmatrix}, \quad \cdots \quad (58)$$

and, in general, S_p is the sum of the principal minors of order p of the matrix $A = \| a_{ik} \|_1^n$ $(p = 1, 2, \ldots, n)$.[20] In particular, $S_n = |A|$.

We denote by $\tilde{A}$ the matrix corresponding to the same operator $\boldsymbol{A}$ in another basis. $\tilde{A}$ is similar to A:

[18] The name is due to the fact that this equation occurs in the study of secular perturbations of the planets.

[19] This proposition is valid even in the more general case in which F is an arbitrary algebraically closed field, i.e., a field that contains the roots of all algebraic equations with coefficients in the field.

[20] The power $(-\lambda)^{n-p}$ occurs only in those terms of the characteristic determinant (56) that contain precisely $n - p$ of the diagonal elements, say,

$$a_{j_1 j_1} - \lambda, \; a_{j_2 j_2} - \lambda, \; \ldots, \; a_{j_{n-p} j_{n-p}} - \lambda.$$

The product of these diagonal elements occurs in the expansion of the determinant (56) with a factor in which the term free of λ is the principal minor

$$A \begin{pmatrix} i_1 & i_2 & \ldots & i_p \\ i_1 & i_2 & \ldots & i_p \end{pmatrix},$$

where $i_1, i_2, \ldots, i_p$ together with $j_1, j_2, \ldots, j_{n-p}$ forms a complete set of indices $1, 2, \ldots, n$; hence in the development of (56) we have

$$|A - \lambda E| = (a_{j_1 j_1} - \lambda)(a_{j_2 j_2} - \lambda) \cdots (a_{j_{n-p} j_{n-p}} - \lambda) A \begin{pmatrix} i_1 & i_2 & \ldots & i_p \\ i_1 & i_2 & \ldots & i_p \end{pmatrix} + (*)$$

When we take all possible combinations $j_1, j_2, \ldots, j_{n-p}$ of $n - p$ of the indices $1, 2, \ldots, n$, we obtain for the coefficient S_p of $(-\lambda)^{n-p}$ the sum of all principal minors of order p in A.

$$\tilde{A} = T^{-1} A T.$$

Hence

$$\tilde{A} - \lambda E = T^{-1}(A - \lambda E) T$$

and therefore

$$|\tilde{A} - \lambda E| = |A - \lambda E|. \quad (59)$$

Thus, similar matrices A and $\tilde{A}$ have the same characteristic polynomial. This polynomial is sometimes called the characteristic polynomial of the operator $\boldsymbol{A}$ and is denoted by $|\boldsymbol{A} - \lambda \boldsymbol{E}|$

If $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \ldots$ are linearly independent characteristic vectors of an operator $\boldsymbol{A}$ corresponding to one and the same characteristic λ, and $\alpha, \beta, \gamma, \ldots$ are arbitrary numbers of F, *then the vector $\alpha\boldsymbol{x} + \beta\boldsymbol{y} + \gamma\boldsymbol{z} + \cdots$ is either equal to zero or is also a characteristic vector of $\boldsymbol{A}$ corresponding to the same λ.*

For from

$$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}, \; \boldsymbol{A}\boldsymbol{y} = \lambda\boldsymbol{y}, \; \boldsymbol{A}\boldsymbol{z} = \lambda\boldsymbol{z}, \; \ldots$$

it follows that

$$\boldsymbol{A}(\alpha\boldsymbol{x} + \beta\boldsymbol{y} + \gamma\boldsymbol{z} + \cdots) = \lambda(\alpha\boldsymbol{x} + \beta\boldsymbol{y} + \gamma\boldsymbol{z} + \cdots).$$

In other words, linearly independent characteristic vectors corresponding to one and the same characteristic value λ form a basis of a 'characteristic' subspace each vector of which is a characteristic vector for the same λ. In particular, each characteristic vector generates a one-dimensional subspace, a 'characteristic' direction.

However, if characteristic vectors of a linear operator $\boldsymbol{A}$ correspond to distinct characteristic values, then a linear combination of these characteristic vectors is not, in general, a characteristic vector of $\boldsymbol{A}$.

The significance of the characteristic vectors and characteristic numbers for the study of linear operators will be illustrated in the next section by the example of operators of simple structure.

§ 8. Linear Operators of Simple Structure

1. We begin with the following lemma.

Lemma: *Characteristic vectors belonging to pairwise distinct characteristic values are always linearly independent.*

Proof. Let

$$\boldsymbol{A}\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i \quad (\boldsymbol{x}_i \neq \boldsymbol{o};\ \lambda_i \neq \lambda_k \text{ for } i \neq k;\ i, k = 1, 2, \ldots, m) \tag{60}$$

and

$$\sum_{i=1}^{m} c_i \boldsymbol{x}_i = \boldsymbol{o}. \tag{61}$$

Applying the operator $\boldsymbol{A}$ to both sides we obtain:

$$\sum_{i=1}^{m} c_i \lambda_i \boldsymbol{x}_i = \boldsymbol{o}. \tag{62}$$

We multiply both sides of (61) by λ_1 and subtract (61) from (62) term by term. Then we obtain

$$\sum_{i=2}^{m} c_i (\lambda_i - \lambda_1) \boldsymbol{x}_i = \boldsymbol{o}. \tag{63}$$

We can say that (63) is obtained from (61) by termwise application of the operator $\boldsymbol{A} - \lambda_1 \boldsymbol{E}$. If we apply the operators $\boldsymbol{A} - \lambda_2 \boldsymbol{E}, \ldots, \boldsymbol{A} - \lambda_{m-1} \boldsymbol{E}$ to (63) term by term, we are led to the following equation:

$$c_m (\lambda_m - \lambda_{m-1})(\lambda_m - \lambda_{m-2}) \cdots (\lambda_m - \lambda_1) \boldsymbol{x}_m = \boldsymbol{o},$$

so that $c_m = 0$. Since any of the summands in (61) can be put last, we have in (61)

$$c_1 = c_2 = \ldots = c_m = 0,$$

i.e., there is no linear dependence among the vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$. This proves the lemma.

If the characteristic equation of an operator has n distinct roots and these roots belong to F, then by the lemma the characteristic vectors belonging to these roots are linearly independent.

Definition 11: *A linear operator $\boldsymbol{A}$ in $\boldsymbol{R}$ is said to be an operator of simple structure if $\boldsymbol{A}$ has n linearly independent characteristic vectors in $\boldsymbol{R}$, where n is the dimension of $\boldsymbol{R}$.*

Thus, a linear operator in $\boldsymbol{R}$ has simple structure if all the roots of the characteristic equation are distinct and belong to F. However, these conditions are not necessary. There exist linear operators of simple structure whose characteristic polynomial has multiple roots.

Let us consider an arbitrary linear operator $\boldsymbol{A}$ of simple structure. We denote by $\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_n$ a basis of $\boldsymbol{R}$ consisting of characteristic vectors of the operator, i.e.,

$$\boldsymbol{A}\boldsymbol{g}_k = \lambda_k \boldsymbol{g}_k \quad (k = 1, 2, \ldots, n).$$

If

$$\boldsymbol{x} = \sum_{k=1}^{n} x_k \boldsymbol{g}_k,$$

then

$$\boldsymbol{A}\boldsymbol{x} = \sum_{k=1}^{n} x_k \boldsymbol{A}\boldsymbol{g}_k = \sum_{k=1}^{n} \lambda_k x_k \boldsymbol{g}_k.$$

The effect of the operator $\boldsymbol{A}$ of simple structure on the vector $\boldsymbol{x} = \sum_{k=1}^{n} x_k \boldsymbol{g}_k$ may be put into words as follows:

In the n-dimensional space $\boldsymbol{R}$ there exist n linearly independent 'directions' along which the operator $\boldsymbol{A}$ of simple structure realizes a 'dilatation' with coefficients $\lambda_1, \lambda_2, \ldots, \lambda_n$. An arbitrary vector $\boldsymbol{x}$ may be decomposed into components along these characteristic directions. These components are subject to the corresponding 'dilatations' and their sum then gives the vector $\boldsymbol{A}\boldsymbol{x}$.

It is easy to see that to the operator $\boldsymbol{A}$ in a 'characteristic' basis $\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_n$ there corresponds the diagonal matrix

$$\tilde{A} = \| \lambda_i \delta_{ik} \|_1^n.$$

If we denote by A the matrix corresponding to $\boldsymbol{A}$ in an arbitrary basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$, then

$$A = T \| \lambda_i \delta_{ik} \|_1^n T^{-1}. \tag{64}$$

A matrix that is similar (p. 68) to a diagonal matrix is called a *matrix of simple structure*. Thus, to an operator of simple structure there corresponds in any basis a matrix of simple structure, and vice versa.

2. The matrix T in (64) realizes the transition from the basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ to the basis $\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_n$. The k-th column of T contains the coordinates of a characteristic vector $\boldsymbol{g}_k$ (with respect to $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$) that corresponds to the characteristic value λ_k of A $(k = 1, 2, \ldots, n)$. The matrix T is called the *fundamental* matrix for A.

We rewrite (64) as follows:

$$A = TLT^{-1} \quad (L = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}). \tag{64'}$$

On going over to the p-th compound matrices ($1 \leqq p \leqq n$), we obtain (see Chapter I, § 4):

$$\mathfrak{A}_p = \mathfrak{T}_p \mathfrak{L}_p \mathfrak{T}_p^{-1}. \tag{65}$$

$\mathfrak{L}_p$ is a diagonal matrix of order N ($N = \binom{n}{p}$) along whose main diagonal are all the possible products of $\lambda_1, \lambda_2, \ldots, \lambda_n$ taken p at a time. A comparison of (65) with (64') yields the following theorem:

Theorem 3: *If a matrix $A = \| a_{ik} \|_1^n$ has simple structure, then for every $p \leqq n$ the compound matrix $\mathfrak{A}_p$ also has simple structure; moreover, the characteristic values of $\mathfrak{A}_p$ are all the possible products $\lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_p}$ ($1 \leqq i_1 < i_2 < \ldots < i_p \leqq n$) of p of the characteristic values $\lambda_1, \lambda_2, \ldots, \lambda_n$ of A, and the fundamental matrix of $\mathfrak{A}_p$ is the compound $\mathfrak{T}_p$ of the fundamental matrix T of A.*

Corollary: *If a characteristic value λ_k of a matrix of simple structure $A = \| a_{ik} \|_1^n$ corresponds to a characteristic vector with the coordinates $t_{1k}, t_{2k}, \ldots, t_{nk}$ ($k = 1, 2, \ldots, n$) and if $T = \| t_{ik} \|_1^n$, then the characteristic value $\lambda_{k_1}\lambda_{k_2}\cdots\lambda_{k_p}$ ($1 \leqq k_1 < k_2 < \ldots < k_p \leqq n$) of $\mathfrak{A}_p$ corresponds to the characteristic vector with coordinates*

$$T\begin{pmatrix} i_1 & i_2 & \ldots & i_p \\ k_1 & k_2 & \ldots & k_p \end{pmatrix} \quad (1 \leqq i_1 < i_2 < \cdots < i_p \leqq n). \tag{66}$$

An arbitrary matrix $A = \| a_{ik} \|_1^n$ may be represented in the form of a sequence of matrices A_m ($m \to \infty$) each of which does not have multiple characteristic values and, therefore, has simple structure. The characteristic values $\lambda_1^{(m)}, \lambda_2^{(m)}, \ldots, \lambda_n^{(m)}$ of the matrix A_m converge for $m \to \infty$ to the characteristic values $\lambda_1, \lambda_2, \ldots, \lambda_n$ of A,

$$\lim_{m \to \infty} \lambda_k^{(m)} = \lambda_k \qquad (k = 1, 2, \ldots, n).$$

Hence

$$\lim_{m \to \infty} \lambda_{k_1}^{(m)} \lambda_{k_2}^{(m)} \cdots \lambda_{k_p}^{(m)} = \lambda_{k_1}\lambda_{k_2} \cdots \lambda_{k_p} \quad (1 \leqq k_1 < k_2 < \cdots < k_p \leqq n).$$

Moreover, since $\lim_{m \to \infty} \mathfrak{A}_{(m)p} = \mathfrak{A}_p$, we deduce from Theorem 3:

Theorem 4 (Kronecker): *If $\lambda_1, \lambda_2, \ldots, \lambda_n$ is a complete system of characteristic values of an arbitrary matrix A, then a complete system of characteristic values of the compound matrix $\mathfrak{A}_p$ consists of all possible products of the numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$ taken p at a time ($p = 1, 2, \ldots, n$).*

In the present section we have investigated operators and matrices of simple structure. The study of the structure of operators and matrices of general type will be resumed in Chapters VI and VII.

CHAPTER IV

THE CHARACTERISTIC POLYNOMIAL AND THE MINIMAL POLYNOMIAL OF A MATRIX

Two polynomials are associated with every square matrix: the characteristic polynomial and the minimal polynomial. These polynomials play an important role in various problems of the theory of matrices. For example, the concept of a function of a matrix, which we shall introduce in the next chapter, will be based entirely on the concept of the minimal polynomial. In the present chapter, the properties of the characteristic polynomial and the minimal polynomial are studied. A prerequisite to this investigation is some basic information about polynomials with matrix coefficients and operations on them.

§ 1. Addition and Multiplication of Matrix Polynomials

1. We consider a square *polynomial* matrix $A(\lambda)$, i.e., a square matrix whose elements are polynomials in λ (with coefficients in the given number field F):

$$A(\lambda) = \| a_{ik}(\lambda) \|_1^n = \| a_{ik}^{(0)}\lambda^m + a_{ik}^{(1)}\lambda^{m-1} + \cdots + a_{ik}^{(m)} \|_1^n . \tag{1}$$

The matrix $A(\lambda)$ can be represented in the form of a polynomial with matrix coefficients arranged with respect to the powers of λ:

$$A(\lambda) = A_0\lambda^m + A_1\lambda^{m-1} + \cdots + A_m , \tag{2}$$

where

$$A_j = \| a_{ik}^{(j)} \|_1^n \qquad (j = 0, 1, \ldots, m) . \tag{3}$$

The number m is called the *degree* of the polynomial, provided $A_0 \neq O$. The number n is called the *order* of the polynomial. The polynomial (1) is called *regular* if $|A_0| \neq 0$.

A polynomial with matrix coefficients will sometimes be called a *matrix polynomial*. In contrast to a matrix polynomial an ordinary polynomial with scalar coefficients will be called a *scalar polynomial*.

We shall now consider the fundamental operations on matrix polynomials. Let two matrix polynomials $A(\lambda)$ and $B(\lambda)$ of the same order be given. We denote by m the larger of their degrees. These polynomials can be written in the form

$$A(\lambda) = A_0\lambda^m + A_1\lambda^{m-1} + \cdots + A_m ,$$
$$B(\lambda) = B_0\lambda^m + B_1\lambda^{m-1} + \cdots + B_m .$$

Then

$$A(\lambda) \pm B(\lambda) = (A_0 \pm B_0)\lambda^m + (A_1 \pm B_1)\lambda^{m-1} + \cdots + (A_m \pm B_m),$$

i.e.: *The sum (difference) of two matrix polynomials of the same order can be represented in the form of a polynomial whose degree does not exceed the larger of the degrees of the given polynomials.*

Let $A(\lambda)$ and $B(\lambda)$ be two matrix polynomials of the same order n and of respective degrees m and p:

$$A(\lambda) = A_0\lambda^m + A_1\lambda^{m-1} + \cdots + A_m \qquad (A_0 \neq O),$$
$$B(\lambda) = B_0\lambda^p + B_1\lambda^{p-1} + \cdots + B_p \qquad (B_0 \neq O).$$

Then

$$A(\lambda)\,B(\lambda) = A_0B_0\lambda^{m+p} + (A_0B_1 + A_1B_0)\lambda^{m+p-1} + \cdots + A_mB_p . \tag{4}$$

If we multiply $B(\lambda)$ by $A(\lambda)$ (i.e., interchange the order of the factors), then we obtain, in general, a different polynomial.

2. The multiplication of matrix polynomials has a specific property. In contrast to the product of scalar polynomials, the product (4) of matrix polynomials may have a degree less than $m + p$, i.e., less than the sum of the degrees of the factors. For, in (4) the product A_0B_0 may be the null matrix even though $A_0 \neq O$, $B_0 \neq O$. However, if at least one of the matrices A_0 and B_0 is non-singular, then it follows from $A_0 \neq O$ and $B_0 \neq O$ that $A_0B_0 \neq O$. Thus: *The product of two matrix polynomials is a polynomial whose degree is less than or equal to the sum of the degrees of the factors. If at least one of the two factors is regular, then the degree of the product is always equal to the sum of the degrees of the factors.*

§ 2. Right and Left Division of Matrix Polynomials

1. Let $A(\lambda)$ and $B(\lambda)$ be two matrix polynomials of the same order n, and let $B(\lambda)$ be regular:

$$A(\lambda) = A_0\lambda^m + A_1\lambda^{m-1} + \cdots + A_m \qquad (A_0 \neq O),$$
$$B(\lambda) = B_0\lambda^p + B_1\lambda^{p-1} + \cdots + B_p \qquad (|B_0| \neq 0).$$

We shall say that the matrix polynomials $Q(\lambda)$ and $R(\lambda)$ are the *right quotient* and the *right remainder*, respectively, of $A(\lambda)$ on division by $B(\lambda)$ if

$$A(\lambda) = Q(\lambda)\, B(\lambda) + R(\lambda) \tag{5}$$

and if the degree of $R(\lambda)$ is less than that of $B(\lambda)$.

Similarly, we shall call the polynomials $\widehat{Q}(\lambda)$ and $\widehat{R}(\lambda)$ the *left quotient* and the *left remainder* of $A(\lambda)$ on division by $B(\lambda)$ if

$$A(\lambda) = B(\lambda)\, \widehat{Q}(\lambda) + \widehat{R}(\lambda) \tag{6}$$

and if the degree of $\widehat{R}(\lambda)$ is less than that of $B(\lambda)$.

The reader should note that in the '*right*' division (i.e., when the right quotient and the right remainder are to be found) in (5) the quotient $Q(\lambda)$ is multiplied by the 'divisor' $B(\lambda)$ on the *right*, and in the '*left*' division in (6) the quotient $\widehat{Q}(\lambda)$ is multiplied by the divisor $B(\lambda)$ on the *left*. The polynomials $\widehat{Q}(\lambda)$ and $\widehat{R}(\lambda)$ do not, in general, coincide with $Q(\lambda)$ and $R(\lambda)$.

2. We shall now show that *both right and left division of matrix polynomials of the same order are always possible and unique, provided the divisor is a regular polynomial.*

Let us consider the right division of $A(\lambda)$ by $B(\lambda)$. If $m < p$, we can set $Q(\lambda) = O$, $R(\lambda) = A(\lambda)$. If $m \geqq p$, we apply the usual scheme for the division of a polynomial by a polynomial in order to find the quotient $Q(\lambda)$ and the remainder $R(\lambda)$. We 'divide' the highest term of the dividend $A_0\lambda^m$ by the highest term of the divisor $B_0\lambda^p$. We obtain the highest term $A_0B_0^{-1}\lambda^{m-p}$ of the required quotient. We multiply this term on the right by the divisor $B(\lambda)$ and subtract the product so obtained from $A(\lambda)$. Thus we find the 'first remainder' $A^{(1)}(\lambda)$:

$$A(\lambda) = A_0B_0^{-1}\lambda^{m-p}\, B(\lambda) + A^{(1)}(\lambda)\,. \tag{7}$$

The degree $m^{(1)}$ of $A^{(1)}(\lambda)$ is less than m:

$$A^{(1)}(\lambda) = A_0^{(1)}\lambda^{m^{(1)}} + \cdots \qquad (A_0^{(1)} \neq O,\ m^{(1)} < m)\,. \tag{8}$$

If $m^{(1)} \geqq p$, then we repeat the process and obtain:

$$\left.\begin{aligned} A^{(1)}(\lambda) &= A_0^{(1)}\, B_0^{-1}\lambda^{m^{(1)}-p}\, B(\lambda) + A^{(2)}(\lambda)\,, \\ A^{(2)}(\lambda) &= A_0^{(2)}\lambda^{m^{(2)}} + \cdots \qquad (m^{(2)} < m^{(1)}),\end{aligned}\right\} \tag{9}$$

etc.

Since the degrees of $A(\lambda)$, $A^{(1)}(\lambda)$, $A^{(2)}(\lambda)$, ... decrease, at some stage we arrive at a remainder $R(\lambda)$ whose degree is less than p. Then it follows from (7) and (9) that

$$A(\lambda) = Q(\lambda)B(\lambda) + R(\lambda),$$

where

$$Q(\lambda) = A_0B_0^{-1}\lambda^{m-p} + A_0^{(1)}B_0^{-1}\lambda^{m^{(1)}-p} + \cdots. \tag{10}$$

We shall now prove the *uniqueness* of the right division. Suppose we have simultaneously

$$A(\lambda) = Q(\lambda)\, B(\lambda) + R(\lambda) \tag{11}$$

and

$$A(\lambda) = Q^*(\lambda)\, B(\lambda) + R^*(\lambda), \tag{12}$$

where the degrees of $R(\lambda)$ and $R^*(\lambda)$ are less than that of $B(\lambda)$, i.e., less than p. Subtracting (11) from (12) term by term we obtain

$$[Q(\lambda) - Q^*(\lambda)]\, B(\lambda) = R^*(\lambda) - R(\lambda)\,. \tag{13}$$

If we had $Q(\lambda) - Q^*(\lambda) \not\equiv O$, then the degree on the left-hand side of (13) would be the sum of the degrees of $B(\lambda)$ and $Q(\lambda) - Q^*(\lambda)$, because $|B_0| \neq 0$, and would therefore be at least equal to p. This is impossible, since the degree of the polynomial on the right-hand side of (13) is less than p. Thus, $Q(\lambda) - Q^*(\lambda) \equiv O$, and then it follows from (13) that $R(\lambda) - R^*(\lambda) \equiv O$, i.e.,

$$Q(\lambda) = Q^*(\lambda)\,, \qquad R(\lambda) = R^*(\lambda)\,.$$

The existence and uniqueness of the left quotient and left remainder is established similarly.[1]

---

[1] Note that the possibility and uniqueness of the left division of $A(\lambda)$ by $B(\lambda)$ follows from that of the right division of the transposed matrices $A^{\mathsf{T}}(\lambda)$ and $B^{\mathsf{T}}(\lambda)$. (The regularity of $B(\lambda)$ implies that of $B^{\mathsf{T}}(\lambda)$.) For from

$$A^{\mathsf{T}}(\lambda) = Q_1(\lambda)\, B^{\mathsf{T}}(\lambda) + R_1(\lambda)$$

it follows (see Chapter I, p. 19) that

$$A(\lambda) = B(\lambda)\, Q_1^{\mathsf{T}}(\lambda) + R_1^{\mathsf{T}}(\lambda)\,. \tag{6'}$$

By the same reasoning, the left division of $A(\lambda)$ by $B(\lambda)$ is unique; for if it were not, then the right division of $A^{\mathsf{T}}(\lambda)$ by $B^{\mathsf{T}}(\lambda)$ would not be unique.

Comparison of (6) and (6') gives

$$\widehat{Q}(\lambda) = Q_1^{\mathsf{T}}(\lambda), \quad \widehat{R}(\lambda) = R_1^{\mathsf{T}}(\lambda)\,.$$

Example.

$$A(\lambda)=\begin{Vmatrix}\lambda^3+\lambda & 2\lambda^3+\lambda^2\\ -\lambda^3-2\lambda^2+1 & 3\lambda^2+\lambda\end{Vmatrix}$$

$$=\overbrace{\begin{Vmatrix}1 & 2\\ -1 & 3\end{Vmatrix}}^{A_0}\lambda^3+\begin{Vmatrix}0 & 1\\ -2 & 0\end{Vmatrix}\lambda^3+\begin{Vmatrix}1 & 0\\ 0 & 1\end{Vmatrix}\lambda+\begin{Vmatrix}0 & 0\\ 1 & 0\end{Vmatrix},$$

$$B(\lambda)=\begin{Vmatrix}2\lambda^3+3 & -\lambda^2+1\\ -\lambda^3-1 & \lambda^3+2\end{Vmatrix}=\overbrace{\begin{Vmatrix}2 & -1\\ -1 & 1\end{Vmatrix}}^{B_0}\lambda^3+\begin{Vmatrix}3 & 1\\ -1 & 2\end{Vmatrix},$$

$$|B_0|=1,\ B_0^{-1}=\begin{Vmatrix}1 & 1\\ 1 & 2\end{Vmatrix},\ A_0B_0^{-1}=\begin{Vmatrix}3 & 5\\ 2 & 5\end{Vmatrix},\ A_0B_0^{-1}B(\lambda)=\begin{Vmatrix}\lambda^3+4 & 2\lambda^3+13\\ -\lambda^3+1 & 3\lambda^3+12\end{Vmatrix},$$

$$A^{(1)}(\lambda)=\begin{Vmatrix}\lambda^3+\lambda & 2\lambda^3+\lambda^2\\ -\lambda^3-2\lambda^3+1 & 3\lambda^3+\lambda\end{Vmatrix}-\begin{Vmatrix}\lambda^3+4\lambda & 2\lambda^3+13\lambda\\ -\lambda^3+\lambda & 3\lambda^3+12\lambda\end{Vmatrix}$$

$$=\begin{Vmatrix}-3\lambda & \lambda^3-13\lambda\\ -2\lambda^3-\lambda+1 & -11\lambda\end{Vmatrix},$$

$$A^{(1)}(\lambda)=\begin{Vmatrix}0 & 1\\ -2 & 0\end{Vmatrix}\lambda^3+\begin{Vmatrix}-3 & -13\\ -1 & -11\end{Vmatrix}\lambda+\begin{Vmatrix}0 & 0\\ 1 & 0\end{Vmatrix},$$

$$A_0^{(1)}B_0^{-1}=\begin{Vmatrix}0 & 1\\ -2 & 0\end{Vmatrix}\cdot\begin{Vmatrix}1 & 1\\ 1 & 2\end{Vmatrix}=\begin{Vmatrix}1 & 2\\ -2 & -2\end{Vmatrix},$$

$$A_0^{(1)}B_0^{-1}B(\lambda)=\begin{Vmatrix}1 & 2\\ -2 & -2\end{Vmatrix}\cdot\begin{Vmatrix}2\lambda^3+3 & -\lambda^2+1\\ -\lambda^2-1 & \lambda^3+2\end{Vmatrix}=\begin{Vmatrix}1 & \lambda^3+5\\ -2\lambda^3-4 & -6\end{Vmatrix},$$

$$R(\lambda)=A^{(1)}(\lambda)-A_0^{(1)}B_0^{-1}B(\lambda)$$

$$=\begin{Vmatrix}-3\lambda & \lambda^3-13\lambda\\ -2\lambda^3-\lambda+1 & -11\lambda\end{Vmatrix}-\begin{Vmatrix}1 & \lambda^3+5\\ -2\lambda^3-4 & -6\end{Vmatrix}=\begin{Vmatrix}-3\lambda-1 & -13\lambda-5\\ -\lambda+5 & -11\lambda+6\end{Vmatrix},$$

$$Q(\lambda)=A_0B_0^{-1}\lambda+A_0^{(1)}B_0^{-1}=\begin{Vmatrix}3 & 5\\ 2 & 5\end{Vmatrix}\lambda+\begin{Vmatrix}1 & 2\\ -2 & -2\end{Vmatrix}=\begin{Vmatrix}3\lambda+1 & 5\lambda+2\\ 2\lambda-2 & 5\lambda-2\end{Vmatrix}.$$

As an exercise, the reader should verify that

$$A(\lambda)=Q(\lambda)B(\lambda)+R(\lambda).$$

§ 3. The Generalized Bézout Theorem

1. We consider an arbitrary matrix polynomial of order n

$$F(\lambda)=F_0\lambda^m+F_1\lambda^{m-1}+\cdots+F_m\qquad(F_0\neq O).\tag{14}$$

This polynomial can also be written as follows:

$$F(\lambda)=\lambda^mF_0+\lambda^{m-1}F_1+\cdots+F_m.\tag{15}$$

For a scalar λ, both ways of writing give the same result. However, if we substitute for the scalar argument λ a square matrix A of order n, then the results of the substitution in (14) and (15) will, in general, be distinct, since the powers of A need not be permutable with the matrix coefficients $F_0, F_1, \ldots, F_m$.

We set

$$F(A)=F_0A^m+F_1A^{m-1}+\cdots+F_m\tag{16}$$

and

$$\widehat{F}(A)=A^mF_0+A^{m-1}F_1+\cdots+F_m,\tag{17}$$

and call $F(A)$ the *right value* and $\widehat{F}(A)$ the *left value* of $F(\lambda)$ on substitution of A for λ.[2]

We divide $F(\lambda)$ by the binomial $\lambda E-A$. In this case the right remainder $R(\lambda)$ and left remainder $\widehat{R}(\lambda)$ will not depend on λ. To determine the right remainder we use the usual division scheme:

$$\begin{aligned}F(\lambda)&=F_0\lambda^m+F_1\lambda^{m-1}+\cdots+F_m\\ &=F_0\lambda^{m-1}(\lambda E-A)+(F_0A+F_1)\lambda^{m-1}+F_2\lambda^{m-2}+\cdots\\ &=[F_0\lambda^{m-1}+(F_0A+F_1)\lambda^{m-2}](\lambda E-A)+(F_0A^2+F_1A+F_2)\lambda^{m-2}+F_3\lambda^{m-3}+\cdots\\ &=[F_0\lambda^{m-1}+(F_0A+F_1)\lambda^{m-2}+\cdots\\ &\qquad+F_0A^{m-1}+F_1A^{m-2}+\cdots+F_{m-1}](\lambda E-A)\\ &\qquad+F_0A^m+F_1A^{m-1}+\cdots+F_m.\end{aligned}$$

Thus we have found that

$$R=F_0A^m+F_1A^{m-1}+\cdots+F_m=F(A).\tag{18}$$

Similarly

$$\widehat{R}=\widehat{F}(A).\tag{19}$$

This proves

Theorem 1 (The Generalized Bézout Theorem): *When the matrix polynomial $F(\lambda)$ is divided on the right by the binomial $\lambda E-A$, the remainder is $F(A)$; when it is divided on the left, the remainder is $\widehat{F}(A)$.*

[2] In the 'right' value $F(A)$ the powers of A are at the right of the coefficients; in the 'left' value $\widehat{F}(A)$, at the left.

2. From this theorem it follows that:

A polynomial $F(\lambda)$ is divisible by the binomial $\lambda E - A$ on the right (left) without remainder if and only if $F(A) = O$ ($\widehat{F}(A) = O$).

Example. Let $A = \| a_{ik} \|_1^n$ and let $f(\lambda)$ be a polynomial in λ. Then

$$F(\lambda) = f(\lambda)E - f(A)$$

is divisible by $\lambda E - A$ (both on the right and on the left) without remainder. This follows immediately from the generalized Bézout Theorem, because in this case $F(A) = \widehat{F}(A) = O$.

§ 4. The Characteristic Polynomial of a Matrix. The Adjoint Matrix

1. We consider a matrix $A = \| a_{ik} \|_1^n$. The *characteristic matrix* of A is $\lambda E - A$. The determinant of the characteristic matrix

$$\Delta(\lambda) = |\lambda E - A| = |\lambda\delta_{ik} - a_{ik}|_1^n,$$

is a scalar polynomial in λ and is called the *characteristic polynomial* of A (see Chapter III, § 7).[3]

The matrix $B(\lambda) = \| b_{ik}(\lambda) \|_1^n$, where $b_{ik}(\lambda)$ is the algebraic complement of the element $\lambda\delta_{ik} - a_{ik}$ in the determinant $\Delta(\lambda)$ is called the *adjoint matrix* of A.

By way of example, for the matrix

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

we have:

$$\lambda E - A = \begin{Vmatrix} \lambda - a_{11} & -a_{12} & -a_{13} \\ -a_{21} & \lambda - a_{22} & -a_{23} \\ -a_{31} & -a_{32} & \lambda - a_{33} \end{Vmatrix},$$

$$\Delta(\lambda) = |\lambda E - A| = \lambda^3 - (a_{11} + a_{22} + a_{33})\lambda^2 + \ldots,$$

$$B(\lambda) = \begin{Vmatrix} \lambda^2 - (a_{22} + a_{33})\lambda + a_{22}a_{33} - a_{23}a_{32} & * & * \\ a_{21}\lambda + a_{23}a_{31} - a_{21}a_{33} & * & * \\ a_{31}\lambda + a_{21}a_{32} - a_{22}a_{31} & * & * \end{Vmatrix}.$$

[3] This polynomial differs by the factor $(-1)^n$ from the polynomial $\Delta(\lambda)$ introduced in Chapter III, § 7.

These definitions imply the following identities in λ:

$$(\lambda E - A)B(\lambda) = \Delta(\lambda)E, \tag{20}$$

$$B(\lambda)(\lambda E - A) = \Delta(\lambda)E. \tag{20'}$$

The right-hand sides of these equations can be regarded as polynomials with matrix coefficients (each of these coefficients is the product of a scalar and the unit matrix E). The polynomial matrix $B(\lambda)$ can also be represented in the form of a polynomial arranged with respect to the powers of λ. Equations (20) and (20') show that $\Delta(\lambda)E$ is divisible on the right and on the left by $\lambda E - A$ without remainder. By the Generalized Bézout Theorem, this is only possible when the remainder $\Delta(A)E = \Delta(A)$ is the null matrix. Thus we have proved:

Theorem 2 (Hamilton-Cayley): *Every square matrix A satisfies its characteristic equation*, i.e.

$$\Delta(A) = O. \tag{21}$$

Example.

$$A = \begin{Vmatrix} 2 & 1 \\ -1 & 3 \end{Vmatrix},$$

$$\Delta(\lambda) = \begin{vmatrix} \lambda - 2 & -1 \\ 1 & \lambda - 3 \end{vmatrix} = \lambda^2 - 5\lambda + 7,$$

$$\Delta(A) = A^2 - 5A + 7E = \begin{Vmatrix} 3 & 5 \\ -5 & 8 \end{Vmatrix} - 5\begin{Vmatrix} 2 & 1 \\ -1 & 3 \end{Vmatrix} + 7\begin{Vmatrix} 1 & 0 \\ 0 & 1 \end{Vmatrix} = \begin{Vmatrix} 0 & 0 \\ 0 & 0 \end{Vmatrix} = O.$$

2. We denote by $\lambda_1, \lambda_2, \ldots, \lambda_n$ all the characteristic values of A, i.e., all the roots of the characteristic polynomial $\Delta(\lambda)$ (each λ_i is repeated as often as its multiplicity as a root of $\Delta(\lambda)$ requires). Then

$$\Delta(\lambda) = |\lambda E - A| = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n). \tag{22}$$

Let $g(\mu)$ be an arbitrary scalar polynomial. We wish to find the characteristic values of $g(A)$. For this purpose we split $g(\mu)$ into linear factors

$$g(\mu) = a_0(\mu - \mu_1)(\mu - \mu_2)\cdots(\mu - \mu_l). \tag{23}$$

On both sides of this identity we substitute the matrix A for μ:

$$g(A) = a_0(A - \mu_1 E)(A - \mu_2 E)\cdots(A - \mu_l E). \tag{24}$$

Passing to determinants on both sides of (24) and using (22) and (23) we find

$$|g(A)| = a_0^n |A-\mu_1 E|\,|A-\mu_2 E| \cdots |A-\mu_l E|$$
$$= (-1)^{nl} a_0^n \Delta(\mu_1)\Delta(\mu_2)\cdots\Delta(\mu_l)$$
$$= (-1)^{nl} a_0^n \prod_{i=1}^{l}\prod_{k=1}^{n}(\mu_i-\lambda_k) = g(\lambda_1)\,g(\lambda_2)\cdots g(\lambda_n).$$

If in the equation

$$|g(A)| = g(\lambda_1)\,g(\lambda_2)\cdots g(\lambda_n) \tag{25}$$

we replace the polynomial $g(\mu)$ by $\lambda - g(\mu)$, where λ is some parameter, we find:

$$|\lambda E - g(A)| = [\lambda - g(\lambda_1)]\,[\lambda - g(\lambda_2)]\cdots[\lambda - g(\lambda_n)]. \tag{26}$$

This leads to the following theorem.

Theorem 3: *If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are all the characteristic values (with the proper multiplicities) of a matrix A and if $g(\mu)$ is a scalar polynomial, then $g(\lambda_1), g(\lambda_2), \ldots, g(\lambda_n)$ are the characteristic values of $g(A)$.*

In particular, if A has the characteristic values $\lambda_1, \lambda_2, \ldots, \lambda_n$, then A^k has the characteristic values $\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k$ $(k = 0, 1, 2, \ldots)$.

3. We shall now derive an effective formula expressing the adjoint matrix $B(\lambda)$ in terms of the characteristic polynomial $\Delta(\lambda)$.

Let

$$\Delta(\lambda) = \lambda^n - p_1\lambda^{n-1} - p_2\lambda^{n-2} - \cdots - p_n. \tag{27}$$

The difference $\Delta(\lambda) - \Delta(\mu)$ is divisible by $\lambda - \mu$ without remainder Therefore

$$\delta(\lambda,\mu) = \frac{\Delta(\lambda)-\Delta(\mu)}{\lambda-\mu} = \lambda^{n-1} + (\mu - p_1)\lambda^{n-2} + (\mu^2 - p_1\mu - p_2)\lambda^{n-3} + \cdots \tag{28}$$

is a polynomial in λ and μ.

The identity

$$\Delta(\lambda) - \Delta(\mu) = \delta(\lambda,\mu)(\lambda-\mu) \tag{29}$$

will still hold if we replace λ and μ by the permutable matrices λE and A. Since by the Hamilton-Cayley Theorem $\Delta(A) = O$,

$$\Delta(\lambda)E = \delta(\lambda E, A)(\lambda E - A). \tag{30}$$

Comparing (20′) with (30), we obtain by virtue of the uniqueness of the quotient the required formula

$$B(\lambda) = \delta(\lambda E, A). \tag{31}$$

Hence by (28)

$$B(\lambda) = E\lambda^{n-1} + B_1\lambda^{n-2} + B_2\lambda^{n-3} + \cdots + B_{n-1}, \tag{32}$$

where

$$B_1 = A - p_1 E, \quad B_2 = A^2 - p_1 A - p_2 E, \quad \ldots$$

and, in general,

$$B_k = A^k - p_1 A^{k-1} - p_2 A^{k-2} - \cdots - p_k E \qquad (k = 1, 2, \ldots, n-1). \tag{33}$$

The matrices $B_1, B_2, \ldots, B_{n-1}$ can be computed in succession, starting from the recurrence relation

$$B_k = AB_{k-1} - p_k E \qquad (k = 1, 2, \ldots, n-1;\ B_0 = E). \tag{34}$$

Moreover,[4]

$$AB_{n-1} - p_n E = O. \tag{35}$$

The relations (34) and (35) follow immediately from (20) if we equate the coefficients of equal powers of λ on both sides.

If A is non-singular, then

$$p_n = (-1)^{n-1}|A| \neq 0,$$

and it follows from (35) that

$$A^{-1} = \frac{1}{p_n} B_{n-1}. \tag{36}$$

Let λ_0 be a characteristic value of A, so that $\Delta(\lambda_0) = 0$. Substituting the value λ_0 in (20), we find:

$$(\lambda_0 E - A)\,B(\lambda_0) = O. \tag{37}$$

Let us assume that $B(\lambda_0) \neq O$ and denote by b an arbitrary non-zero column of this matrix. Then from (37) we have $(\lambda_0 E - A)b = O$ or

$$Ab = \lambda_0 b. \tag{38}$$

Therefore every non-zero column of $B(\lambda_0)$ determines a characteristic vector corresponding to the characteristic value λ_0.[5]

Thus:

---

[4] From (34) follows (33). If we substitute in (35) the expression for B_{n-1} given in (33), we obtain $\Delta(A) = 0$. This approach to the Hamilton-Cayley Theorem does not require the Generalized Bézout Theorem explicitly, but contains this theorem implicitly.

[5] See Chapter III, § 7. If to the characteristic value λ_0 there correspond d_0 linearly independent characteristic vectors ($n - d_0$ is the rank of $\lambda_0 E - A$), then the rank of $B(\lambda_0)$ does not exceed d_0. In particular, if only one characteristic direction corresponds to λ_0, then in $B(\lambda_0)$ the elements of any two columns are proportional.

If the coefficients of the characteristic polynomial are known, then the adjoint matrix can be found by formula (31). *If the given matrix A is non-singular, then the inverse matrix A^{-1} can be found by formula* (36). *If λ_0 is a characteristic value of A, then the non-zero columns of $B(\lambda_0)$ are characteristic vectors of A for $\lambda=\lambda_0$.*

Example.

$$A=\begin{Vmatrix} 2 & -1 & 1 \\ 0 & 1 & 1 \\ -1 & 1 & 1 \end{Vmatrix},$$

$$\Delta(\lambda)=|\lambda E-A|=\begin{vmatrix} \lambda-2 & 1 & -1 \\ 0 & \lambda-1 & -1 \\ 1 & -1 & \lambda-1 \end{vmatrix}=\lambda^3-4\lambda^2+5\lambda-2,$$

$$\delta(\lambda,\mu)=\frac{\Delta(\lambda)-\Delta(\mu)}{\lambda-\mu}=\lambda^2+\lambda(\mu-4)+\mu^2-4\mu+5,$$

$$B(\lambda)=\delta(\lambda E, A)=\lambda^2 E+\lambda\underbrace{(A-4E)}_{B_1}+\underbrace{A^2-4A+5E}_{B_2}.$$

But

$$B_1=A-4E=\begin{Vmatrix} -2 & -1 & 1 \\ 0 & -3 & 1 \\ -1 & 1 & -3 \end{Vmatrix},\quad B_2=AB_1+5E=\begin{Vmatrix} 0 & 2 & -2 \\ -1 & 3 & -2 \\ 1 & -1 & 2 \end{Vmatrix},$$

$$B(\lambda)=\begin{Vmatrix} \lambda^2-2\lambda & -\lambda+2 & \lambda-2 \\ -1 & \lambda^2-3\lambda+3 & \lambda-2 \\ -\lambda+1 & \lambda-1 & \lambda^2-3\lambda+2 \end{Vmatrix},$$

$$|A|=2,\quad A^{-1}=\frac{1}{2}B_2=\begin{Vmatrix} 0 & 1 & -1 \\ -\frac{1}{2} & \frac{3}{2} & -1 \\ \frac{1}{2} & -\frac{1}{2} & 1 \end{Vmatrix}.$$

Furthermore,

$$\Delta(\lambda)=(\lambda-1)^2(\lambda-2).$$

The first column of the matrix $B(+1)$ gives the characteristic vector $(+1, +1, 0)$ for the characteristic value $\lambda=1$.

The first column of the matrix $B(+2)$ gives the characteristic vector $(0, +1, +1)$ corresponding to the characteristic value $\lambda=2$.

§ 5. The Method of Faddeev for the Simultaneous Computation of the Coefficients of the Characteristic Polynomial and of the Adjoint Matrix

1. D. K. Faddeev[6] has suggested a method for the simultaneous determination of the scalar coefficients $p_1, p_2, \ldots, p_n$ of the characteristic polynomial

$$\Delta(\lambda)=\lambda^n-p_1\lambda^{n-1}-p_2\lambda^{n-2}-\cdots-p_n \tag{39}$$

and of the matrix coefficients $B_1, B_2, \ldots, B_{n-1}$ of the adjoint matrix $B(\lambda)$.

In order to explain the method of Faddeev[7] we introduce the concept of the trace (or spur) of a matrix.

By the *trace* $\operatorname{tr} A$ of a matrix $A=\| a_{ik} \|_1^n$ we mean the sum of the diagonal elements of the matrix:

$$\operatorname{tr} A=\sum_{i=1}^{n} a_{ii}. \tag{40}$$

It is easy to see that

$$\operatorname{tr} A=p_1=\sum_{i=1}^{n}\lambda_i, \tag{41}$$

if $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the characteristic values of A, i.e., if

$$\Delta(\lambda)=(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n). \tag{42}$$

Since by Theorem 3 A^k has the characteristic values $\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k$ $(k=0, 1, 2, \ldots,)$, we have

$$\operatorname{tr} A^k=s_k=\sum_{i=1}^{n}\lambda_i^k \qquad (k=0, 1, 2, \ldots). \tag{43}$$

The sums s_k $(k=1, 2, \ldots, n)$ of powers of the roots of the polynomial (39) are connected with the coefficients by Newton's formulas[8]

$$kp_k=s_k-p_1s_{k-1}-\cdots-p_{k-1}s_1 \qquad (k=1, 2, \ldots, n). \tag{44}$$

If the traces $s_1, s_2, \ldots, s_n$ of the matrices $A, A^2, \ldots, A^n$ are computed, then the coefficients $p_1, p_2, \ldots, p_n$ can be determined from (44). This is the method of Leverrier for the determination of the coefficients of the characteristic polynomial from the traces of the powers of the matrix.

2. Faddeev has proposed to compute successively, instead of the traces of the powers $A, A^2, \ldots, A^n$, the traces of certain other matrices $A_1, A_2, \ldots, A_n$

---

[6] See [14], p. 160.

[7] In Chapter VII, § 8, we shall discuss another effective method, due to A. N. Krylov, of computing the coefficients of the characteristic polynomial.

[8] See, for example, G. Chrystal, *Textbook of Algebra*, Vol. I, pp. 436ff.

and so to determine $p_1, p_2, \ldots, p_n$ and $B_1, B_2, \ldots, B_n$ by the following formulas:

$$\left.\begin{array}{lll} A_1 = A, & p_1 = \operatorname{tr} A_1, & B_1 = A_1 - p_1 E, \\ A_2 = AB_1, & p_2 = \frac{1}{2}\operatorname{tr} A_2, & B_2 = A_2 - p_2 E, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & & \\ A_{n-1} = AB_{n-2}, & p_{n-1} = \frac{1}{n-1}\operatorname{tr} A_{n-1}, & B_{n-1} = A_{n-1} - p_{n-1}E, \\ A_n = AB_{n-1}, & p_n = \frac{1}{n}\operatorname{tr} A_n, & B_n = A_n - p_n E = O. \end{array}\right\} \quad (45)$$

The last equation $B_n = A_n - p_n E = O$ may be used to check the computation.

In order to convince ourselves that the numbers $p_1, p_2, \ldots, p_n$ and the matrices $B_1, B_2, \ldots, B_{n-1}$ that are determined successively by (45) are, in fact, the coefficients of $\Delta(\lambda)$ and $B(\lambda)$, we note that the following formulas for A_k and B_k $(k = 1, 2, \ldots, n)$ follow from (45):

$$A_k = A^k - p_1 A^{k-1} - \cdots - p_{k-1}A, \quad B_k = A^k - p_1 A^{k-1} - \cdots - p_{k-1}A - p_k E. \quad (46)$$

Equating the traces on the left-hand and right-hand sides of the first of these formulas, we obtain

$$k p_k = s_k - p_1 s_{k-1} - \cdots - p_{k-1} s_1.$$

But these formulas coincide with Newton's formulas (44) by which the coefficients of the characteristic polynomial $\Delta(\lambda)$ are determined successively. Therefore the numbers $p_1, p_2, \ldots, p_n$ determined by (45) are also the coefficients of $\Delta(\lambda)$. But then the second of formulas (46) coincide with formulas (33) by which the matrix coefficients $B_1, B_2, \ldots, B_{n-1}$ of the adjoint matrix $B(\lambda)$ are determined. Therefore, formulas (45) also determine the coefficients $B_1, B_2, \ldots, B_{n-1}$ of the matrix polynomial $B(\lambda)$.

Example.[9]

$$A = \begin{Vmatrix} 2 & -1 & 1 & 2 \\ 0 & 1 & 1 & 0 \\ -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ \hline 2 & 2 & 4 & 3 \end{Vmatrix}, \quad p_1 = \operatorname{tr} A = 4, \quad B_1 = A - 4E = \begin{Vmatrix} -2 & -1 & 1 & 2 \\ 0 & -3 & 1 & 0 \\ -1 & 1 & -3 & 1 \\ 1 & 1 & 1 & -4 \end{Vmatrix};$$

[9] As a check on the computation, we write under each matrix A_1, A_2, A_3 a row whose elements are the sums of the elements above it. The product of this row of 'column-sums' of the first factor into the columns of the second factor must give the elements of the column-sum of the product.

$$A_2 = AB_1 = \begin{Vmatrix} -3 & 4 & 0 & -3 \\ -1 & -2 & -2 & 1 \\ 2 & 0 & -2 & -5 \\ -3 & -3 & -1 & 3 \\ \hline -5 & -1 & -5 & -4 \end{Vmatrix}, \quad p_2 = \frac{1}{2}\operatorname{tr} A_2 = -2, \quad B_2 = A_2 + 2E = \begin{Vmatrix} -1 & 4 & 0 & -3 \\ -1 & 0 & -2 & 1 \\ 2 & 0 & 0 & -5 \\ -3 & -3 & -1 & 3 \end{Vmatrix};$$

$$A_3 = AB_2 = \begin{Vmatrix} -5 & 2 & 0 & -2 \\ 1 & 0 & -2 & -4 \\ -1 & -7 & -3 & 4 \\ 0 & 4 & -2 & -7 \\ \hline -5 & -1 & -7 & -9 \end{Vmatrix}, \quad p_3 = \frac{1}{3}\operatorname{tr} A_3 = -5, \quad B_3 = A_3 + 5E = \begin{Vmatrix} 0 & 2 & 0 & -2 \\ 1 & 5 & -2 & 4 \\ -1 & -7 & 2 & 4 \\ 0 & 4 & -2 & -2 \end{Vmatrix};$$

$$A_4 = AB_3 = \begin{Vmatrix} -2 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & -2 \end{Vmatrix}, \quad p_4 = -2.$$

$$\Delta(\lambda) = \lambda^4 - 4\lambda^3 + 2\lambda^2 + 5\lambda + 2,$$

$$|A| = 2, \quad A^{-1} = \frac{1}{p_4} B_3 = \begin{Vmatrix} 0 & -1 & 0 & 1 \\ -\frac{1}{2} & -\frac{5}{2} & 1 & -2 \\ \frac{1}{2} & \frac{7}{2} & -1 & -2 \\ 0 & -2 & 1 & 1 \end{Vmatrix}.$$

Note. If we wish to determine p_1, p_2, p_3, p_4 and only the first columns of B_1, B_2, B_3, it is sufficient to compute in A_2 the elements of the first column and only the diagonal elements of the remaining columns, in A_3 only the elements of the first column, and in A_4 only the first two elements of the first column.

§ 6. The Minimal Polynomial of a Matrix

1. Definition 1: *A scalar polynomial $f(\lambda)$ is called an annihilating polynomial of the square matrix A if*

$$f(A) = O.$$

An annihilating polynomial $\psi(\lambda)$ of least degree with highest coefficient 1 is called a *minimal polynomial* of A.

By the Hamilton-Cayley Theorem the characteristic polynomial $\Delta(\lambda)$ is an annihilating polynomial of A. However, as we shall show below, it is not, in general, a minimal polynomial.

Let us divide an arbitrary annihilating polynomial $f(\lambda)$ by a minimal polynomial

$$f(\lambda) = \psi(\lambda)\, q(\lambda) + r(\lambda),$$

where the degree of $r(\lambda)$ is less than that of $\psi(\lambda)$. Hence we have:

$$f(A)=\psi(A)\,q(A)+r(A).$$

Since $f(A)=O$ and $\psi(A)=O$, it follows that $r(A)=O$. But the degree of $r(\lambda)$ is less than that of the minimal polynomial $\psi(\lambda)$. Therefore $r(\lambda)\equiv 0$.[10] Hence: *Every annihilating polynomial of a matrix is divisible without remainder by the minimal polynomial.*

Let $\psi_1(\lambda)$ and $\psi_2(\lambda)$ be two minimal polynomials of one and the same matrix. Then each is divisible without remainder by the other, i.e., the polynomials differ by a constant factor. This constant factor must be 1, because the highest coefficients in $\psi_1(\lambda)$ and $\psi_2(\lambda)$ are 1. Thus we have proved the *uniqueness of the minimal polynomial* of a given matrix A.

2. We shall now derive a formula connecting the minimal polynomial with the characteristic polynomial.

We denote by $D_{n-1}(\lambda)$ the greatest common divisor of all the minors of order $n-1$ of the characteristic matrix $\lambda E-A$, i.e., of all the elements of the matrix $B(\lambda)=\| b_{ik}(\lambda) \|_1^n$ (see the preceding section). Then

$$B(\lambda)=D_{n-1}(\lambda)\,C(\lambda), \tag{47}$$

where $C(\lambda)$ is a certain polynomial matrix, the *'reduced' adjoint matrix* of $\lambda E-A$. From (20) and (47) we have:

$$\Delta(\lambda)\,E=(\lambda E-A)\,C(\lambda)\,D_{n-1}(\lambda). \tag{48}$$

Hence it follows that $\Delta(\lambda)$ is divisible without remainder by $D_{n-1}(\lambda)$:[11]

$$\frac{\Delta(\lambda)}{D_{n-1}(\lambda)}=\psi(\lambda), \tag{49}$$

where $\psi(\lambda)$ is some polynomial. The factor $D_{n-1}(\lambda)$ in (48) may be cancelled on both sides:[12]

$$\psi(\lambda)\,E=(\lambda E-A)\,C(\lambda). \tag{50}$$

[10] Otherwise there would exist an annihilating polynomial of degree less than that of the minimal polynomial.

[11] We could also verify this immediately by expanding the characteristic determinant $\Delta(\lambda)$ with respect to the elements of an arbitrary row.

[12] In this case we have, apart from (50), also the identity (see (20′))

$$\psi(\lambda)\,E=C(\lambda)\,(\lambda E-A),$$

i.e., $C(\lambda)$ is at one and the same time the left quotient and right quotient of $\psi(\lambda)E$ on division by $\lambda E-A$.

Since $\psi(\lambda)E$ is divisible on the left without remainder by $\lambda E-A$, it follows by the Generalized Bézout Theorem that

$$\psi(A)=O.$$

Thus, the polynomial $\psi(\lambda)$ defined by (49) is an annihilating polynomial of A. Let us show that it is the minimal polynomial.

We denote the minimal polynomial by $\psi^*(\lambda)$. Then $\psi(\lambda)$ is divisible by $\psi^*(\lambda)$ without remainder:

$$\psi(\lambda)=\psi^*(\lambda)\,\chi(\lambda). \tag{51}$$

Since $\psi^*(A)=O$, by the Generalized Bézout Theorem the matrix polynomial $\psi^*(\lambda)E$ is divisible on the left by $\lambda E-A$ without remainder:

$$\psi^*(\lambda)\,E=(\lambda E-A)\,C^*(\lambda). \tag{52}$$

From (51) and (52) it follows that

$$\psi(\lambda)\,E=(\lambda E-A)\,C^*(\lambda)\,\chi(\lambda). \tag{53}$$

The identities (50) and (53) show that $C(\lambda)$ as well as $C^*(\lambda)\chi(\lambda)$ are left quotients of $\psi(\lambda)E$ on division by $\lambda E-A$. By the uniqueness of division

$$C(\lambda)=C^*(\lambda)\,\chi(\lambda).$$

Hence it follows that $\chi(\lambda)$ is a common divisor of all the elements of the polynomial matrix $C(\lambda)$. But, on the other hand, the greatest common divisor of all the elements of the reduced adjoint matrix $C(\lambda)$ is equal to 1, because the matrix was obtained from $B(\lambda)$ by division by $D_{n-1}(\lambda)$. Therefore $\chi(\lambda)=\text{const}$. Since the highest coefficients of $\psi(\lambda)$ and $\psi^*(\lambda)$ are equal, we have in (51) $\chi(\lambda)=1$, i.e., $\psi(\lambda)=\psi^*(\lambda)$, and this is what we had to prove.

We have established the following formula for the minimal polynomial:

$$\psi(\lambda)=\frac{\Delta(\lambda)}{D_{n-1}(\lambda)}. \tag{54}$$

3. For the reduced adjoint matrix $C(\lambda)$ we have a formula analogous to (31) (p. 84):

$$C(\lambda)=\Psi(\lambda E, A); \tag{55}$$

where the polynomial $\Psi(\lambda,\mu)$ is defined by the equation[13]

[13] Formula (55) can be deduced in the same way as (31). On both sides of the identity $\psi(\lambda)-\psi(\mu)=(\lambda-\mu)\Psi(\lambda,\mu)$ we substitute for λ and μ the matrices λE and A and compare the matrix equation so obtained with (50).

$$\Psi(\lambda, \mu) = \frac{\psi(\lambda) - \psi(\mu)}{\lambda - \mu}. \tag{56}$$

Moreover,

$$(\lambda E - A)\, C(\lambda) = \psi(\lambda)\, E. \tag{57}$$

Going over to determinants on both sides of (57), we obtain

$$\Delta(\lambda)\, |C(\lambda)| = [\psi(\lambda)]^n. \tag{58}$$

Thus, $\Delta(\lambda)$ is divisible without remainder by $\psi(\lambda)$ and some power of $\psi(\lambda)$ is divisible without remainder by $\Delta(\lambda)$, i.e., the sets of all the distinct roots of the polynomials $\Delta(\lambda)$ and $\psi(\lambda)$ are eqnal. In other words: *All the distinct characteristic values of A are roots of $\psi(\lambda)$.*

If

$$\begin{gathered}\Delta(\lambda) = (\lambda - \lambda_1)^{n_1} (\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_s)^{n_s} \\ (\lambda_i \neq \lambda_j \text{ for } i \neq j;\ n_i > 0,\ i, j = 1, 2, \ldots, s),\end{gathered} \tag{59}$$

then

$$\psi(\lambda) = (\lambda - \lambda_1)^{m_1} (\lambda - \lambda_2)^{m_2} \cdots (\lambda - \lambda_s)^{m_s}, \tag{60}$$

where

$$0 < m_k \leqq n_k \qquad (k = 1, 2, \ldots, s). \tag{61}$$

4. We mention one further property of the matrix $C(\lambda)$. Let λ_0 be an arbitrary characteristic value of $A = \| a_{ik} \|_1^n$. Then $\psi(\lambda_0) = 0$ and therefore, by (57),

$$(\lambda_0 E - A)\, C(\lambda_0) = O. \tag{62}$$

Note that $C(\lambda_0) \neq O$ always holds, for otherwise all the elements of the reduced adjoint matrix $C(\lambda)$ would be divisible without remainder by $\lambda - \lambda_0$, and this is impossible.

We denote by c an arbitrary non-zero column of $C(\lambda_0)$. Then from (62)

$$(\lambda_0 E - A)\, c = o,$$

i.e.,

$$Ac = \lambda_0 c. \tag{63}$$

In other words, every non-zero column of $C(\lambda_0)$ (and such a column always exists) determines a characteristic vector for $\lambda = \lambda_0$.

Example.

$$A = \begin{Vmatrix} 3 & -3 & 2 \\ -1 & 5 & -2 \\ -1 & 3 & 0 \end{Vmatrix},$$

$$\Delta(\lambda) = \begin{Vmatrix} \lambda - 3 & 3 & -2 \\ 1 & \lambda - 5 & 2 \\ 1 & -3 & \lambda \end{Vmatrix} = \lambda^3 - 8\lambda^2 + 20\lambda - 16 = (\lambda - 2)^2 (\lambda - 4),$$

$$\delta(\lambda, \mu) = \frac{\Delta(\mu) - \Delta(\lambda)}{\mu - \lambda} = \mu^2 + \mu(\lambda - 8) + \lambda^2 - 8\lambda + 20,$$

$$B(\lambda) = A^2 + (\lambda - 8) A + (\lambda^2 - 8\lambda + 20) E$$

$$= \begin{Vmatrix} 10 & -18 & 12 \\ -6 & 22 & -12 \\ -6 & 18 & -8 \end{Vmatrix} + (\lambda - 8) \begin{Vmatrix} 3 & -3 & 2 \\ -1 & 5 & -2 \\ -1 & 3 & 0 \end{Vmatrix} + (\lambda^2 - 8\lambda + 20) \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{Vmatrix}$$

$$= \begin{Vmatrix} \lambda^2 - 5\lambda + 6 & -3\lambda + 6 & 2\lambda - 4 \\ -\lambda + 2 & \lambda^2 - 3\lambda + 2 & -2\lambda + 4 \\ -\lambda + 2 & 3\lambda - 6 & \lambda^2 - 8\lambda + 2 \end{Vmatrix}.$$

All the elements of the matrix $B(\lambda)$ are divisible by $D_2(\lambda) = \lambda - 2$. Cancelling this factor, we have:

$$C(\lambda) = \begin{Vmatrix} \lambda - 3 & -3 & 2 \\ -1 & \lambda - 1 & -2 \\ -1 & 3 & \lambda - 6 \end{Vmatrix}$$

and

$$\psi(\lambda) = \frac{\Delta(\lambda)}{\lambda - 2} = (\lambda - 2)(\lambda - 4).$$

In $C(\lambda)$ we substitute for λ the value $\lambda_0 = 2$:

$$C(2) = \begin{Vmatrix} -1 & -3 & 2 \\ -1 & 1 & -2 \\ -1 & 3 & -4 \end{Vmatrix}.$$

The first column gives us the characteristic vector (1, 1, 1,) for $\lambda_0 = 2$. The second column gives us the characteristic vector (— 3, 1, 3) for the same characteristic value $\lambda_0 = 2$. The third column is a linear combination of the first two.

Similarly, setting $\lambda_0 = 4$, we find from the first column of the matrix $C(4)$ the characteristic vector (1, — 1, — 1) corresponding to the characteristic value $\lambda_0 = 4$.

The reader should note that $\psi(\lambda)$ and $C(\lambda)$ could have been determined by a different method.

To begin with, let us find $D_2(\lambda)$. $D_2(\lambda)$ can only have 2 and 4 as its roots. For $\lambda = 4$ the second order minor

$$\begin{vmatrix} 1 & \lambda - 5 \\ 1 & -3 \end{vmatrix} = -\lambda + 2$$

of $\Delta(\lambda)$ does not vanish. Therefore $D_2(4) \neq 0$. For $\lambda = 2$ the columns of $\Delta(\lambda)$ become proportional. Therefore all the minors of order two in $\Delta(\lambda)$

vanish for $\lambda=2$: $D_2(2)=0$. Since the minor to be computed is of the first degree, $D_2(\lambda)$ cannot be divisible by $(\lambda-2)^2$. Therefore

$$D_2(\lambda)=\lambda-2.$$

Hence

$$\psi(\lambda)=\frac{\Delta(\lambda)}{\lambda-2}=(\lambda-2)(\lambda-4)=\lambda^2-6\lambda+8,$$

$$\psi(\lambda,\mu)=\frac{\psi(\mu)-\psi(\lambda)}{\mu-\lambda}=\mu+\lambda-6,$$

$$C(\lambda)=\psi(\lambda E, A)=A+(\lambda-6)E=\begin{Vmatrix}\lambda-3 & -3 & 2\\ -1 & \lambda-1 & -2\\ -1 & 3 & \lambda-6\end{Vmatrix}.$$

CHAPTER V

FUNCTIONS OF MATRICES

§ 1. Definition of a Function of a Matrix

1. Let $A=\|a_{ik}\|_1^n$ be a square matrix and $f(\lambda)$ a function of a scalar argument λ. We wish to define what is to be meant by $f(A)$, i.e., we wish to extend the function $f(\lambda)$ to a matrix value of the argument.

We already know the solution of this problem in the simplest special case where $f(\lambda)=\gamma_0\lambda^l+\gamma_1\lambda^{l-1}+\cdots+\gamma_l$ is a polynomial in λ. In this case, $f(A)=\gamma_0A^l+\gamma_1A^{l-1}+\cdots+\gamma_lE$. Starting from this special case, we shall obtain a definition of $f(A)$ in the general case.

We denote by

$$\psi(\lambda)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_s)^{m_s} \tag{1}$$

the minimal polynomial[1] of A (where $\lambda_1,\lambda_2,\ldots,\lambda_s$ are all the distinct characteristic values of A). The degree of this polynomial is $m=\sum_{k=1}^{s}m_k$.

Let $g(\lambda)$ and $h(\lambda)$ be two polynomials such that

$$g(A)=h(A). \tag{2}$$

Then the difference $d(\lambda)=g(\lambda)-h(\lambda)$, as an annihilating polynomial for A, is divisible by $\psi(\lambda)$ without remainder; we shall write this as follows:

$$g(\lambda)\equiv h(\lambda)\ (\mathrm{mod}\,\psi(\lambda)). \tag{3}$$

Hence by (1)

$$d(\lambda_k)=0,\quad d'(\lambda_k)=0,\quad\ldots,\quad d^{(m_k-1)}(\lambda_k)=0\qquad(k=1,2,\ldots,s),$$

i.e.,

$$g(\lambda_k)=h(\lambda_k),\quad g'(\lambda_k)=h'(\lambda_k),\ldots,g^{(m_k-1)}(\lambda_k)=h^{(m_k-1)}(\lambda_k)\qquad(k=1,2,\ldots,s). \tag{4}$$

[1] See Chapter IV, § 6.

The m numbers

$$f(\lambda_k),\quad f'(\lambda_k),\quad \ldots,\quad f^{(m_k-1)}(\lambda_k)\qquad (k=1,2,\ldots,s) \tag{5}$$

will be called *the values of the function $f(\lambda)$ on the spectrum of the matrix A* and the set of all these values will be denoted symbolically by $f(\Lambda_A)$. If for a function $f(\lambda)$ the values (5) exist (i.e., have meaning), then we shall say that *the function $f(\lambda)$ is defined on the spectrum of the matrix A.*

Equation (4) shows that the polynomials $g(\lambda)$ and $h(\lambda)$ have the same values on the spectrum of A. In symbols:

$$g(\Lambda_A)=h(\Lambda_A).$$

Our argument is reversible: from (4) follows (3) and therefore (2).

Thus, given a matrix A, the values of the polynomial $g(\lambda)$ on the spectrum of A determine the matrix $g(A)$ completely, i.e., all polynomials $g(\lambda)$ that assume the same values on the spectrum of A have one and the same matrix value $g(A)$.

We postulate that the definition of $f(A)$ in the general case be subject to the same principle: *The values of the function $f(\lambda)$ on the spectrum of the matrix A must determine $f(A)$ completely, i.e., all functions $f(\lambda)$ having the same values on the spectrum of A must have the same matrix value $f(A)$.*

But then it is obvious that for the general definition of $f(A)$ it is sufficient to look for a polynomial[2] $g(\lambda)$ that assumes the same values on the spectrum of A as $f(\lambda)$ does and to set:

$$f(A)=g(A).$$

We are thus led to the following definition:

DEFINITION 1: *If the function $f(\lambda)$ is defined on the spectrum of the matrix A, then*

$$f(\Lambda_A)=g(\Lambda_A),$$

where $g(\lambda)$ is an arbitrary polynomial that assumes on the spectrum of A the same values as does $f(\lambda)$:

$$f(A)=g(A).$$

Among all the polynomials with complex coefficients that assume on the spectrum of A the same values as $f(\lambda)$ there is one and only one polynomial

[2] It will be proved in § 2 that such an interpolation polynomial always exists and an algorithm for the computation of the coefficients of the interpolation polynomial of least degree will be given.

$r(\lambda)$ that is of degree less than m.[3] This polynomial $r(\lambda)$ is uniquely determined by the interpolation conditions:

$$r(\lambda_k)=f(\lambda_k),\quad r'(\lambda_k)=f'(\lambda_k),\quad \ldots,\quad r^{(m_k-1)}(\lambda_k)=f^{(m_k-1)}(\lambda_k)\qquad (k=1,2,\ldots,s). \tag{6}$$

The polynomial $r(\lambda)$ is called the *Lagrange-Sylvester interpolation polynomial* for $f(\lambda)$ on the spectrum of A. Definition 1 can also be formulated as follows:

DEFINITION 1′: *Let $f(\lambda)$ be a function defined on the spectrum of a matrix A and $r(\lambda)$ the corresponding Lagrange-Sylvester interpolation polynomial. Then*

$$f(A)=r(A).$$

Note. If the minimal polynomial $\psi(\lambda)$ of a matrix A has no multiple roots[4] (in (1) $m_1=m_2=\ldots=m_s=1$; $s=m$), then for $f(A)$ to have a meaning it is sufficient that $f(\lambda)$ be defined at the characteristic values $\lambda_1,\lambda_2,\ldots,\lambda_m$. But if $\psi(\lambda)$ has multiple roots, then for some characteristic values the derivatives of $f(\lambda)$ up to a certain order (see (6)) must be defined as well.

Example 1: Let us consider the matrix[5]

$$H=\begin{Vmatrix} \overbrace{0 & 1 & 0 & \ldots & 0}^{n} \\ 0 & 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 \\ 0 & 0 & 0 & \ldots & 0 \end{Vmatrix}.$$

Its minimal polynomial is λ^n. Therefore the values of $f(\lambda)$ on the spectrum of H are the numbers $f(0), f'(0), \ldots, f^{(n-1)}(0)$, and the polynomial $r(\lambda)$ is of the form

$$r(\lambda)=f(0)+\frac{f'(0)}{1!}\lambda+\cdots+\frac{f^{(n-1)}(0)}{(n-1)!}\lambda^{n-1}.$$

Therefore

[3] This polynomial is obtained from any other polynomial having the same spectral values by taking the remainder on division by $\psi(\lambda)$ of that polynomial.

[4] In Chapter VI it will be shown that A is a matrix of simple structure (see Chapter III, § 8) in this case, and this case only.

[5] The properties of the matrix H were worked out in the example on pp. 13-14.

$$f(H)=f(0)E+\frac{f'(0)}{1!}H+\cdots+\frac{f^{(n-1)}(0)}{(n-1)!}H^{n-1}=\begin{Vmatrix} f(0) & \frac{f'(0)}{1!} & \cdot & \cdot & \cdot & \frac{f^{(n-1)}(0)}{(n-1)!} \\ 0 & f(0) & \cdot & & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot & f'(0) \\ \cdot & & & & \cdot & 1! \\ 0 & 0 & \cdot & \cdot & \cdot & f(0) \end{Vmatrix}$$

Example 2: Let us consider the matrix

$$J=\begin{Vmatrix} \overbrace{\lambda_0 & 1 & 0 & \cdot & \cdot & 0}^{n} \\ 0 & \lambda_0 & 1 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & 1 \\ 0 & 0 & 0 & \cdot & \cdot & \lambda_0 \end{Vmatrix}.$$

Note that $J=\lambda_0 E+H$, so that $J-\lambda_0 E=H$. The minimal polynomial of J is clearly $(\lambda-\lambda_0)^n$. The interpolation polynomial $r(\lambda)$ of $f(\lambda)$ is given by the equation

$$r(\lambda)=f(\lambda_0)+\frac{f'(\lambda_0)}{1!}(\lambda-\lambda_0)+\cdots+\frac{f^{(n-1)}(\lambda_0)}{(n-1)!}(\lambda-\lambda_0)^{n-1}$$

Therefore

$$f(J)=r(J)=f(\lambda_0)E+\frac{f'(\lambda_0)}{1!}H+\cdots+\frac{f^{(n-1)}(\lambda_0)}{(n-1)!}H^{n-1}$$

$$=\begin{Vmatrix} f(\lambda_0) & \frac{f'(\lambda_0)}{1!} & \cdot & \cdot & \cdot & \frac{f^{(n-1)}(\lambda_0)}{(n-1)!} \\ 0 & f(\lambda_0) & \cdot & & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot & \frac{f'(\lambda_0)}{1!} \\ \cdot & & & & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & f(\lambda_0) \end{Vmatrix}.$$

2. We mention two properties of functions of matrices.

1. *If two matrices A and B are similar and T transforms A into B,*

$$B=T^{-1}AT,$$

then the matrices $f(A)$ and $f(B)$ are also similar and T transforms $f(A)$ into $f(B)$,

$$f(B)=T^{-1}f(A)T.$$

For two similar matrices have equal minimal polynomials,[6] so that $f(\lambda)$ assumes the same values on the spectrum of A and of B. Therefore there exists an interpolation polynomial $r(\lambda)$ such that $f(A)=r(A)$ and $f(B)=r(B)$. But then it follows[6] from the equation $r(B)=T^{-1}r(A)T$ that

$$f(B)=T^{-1}f(A)T.$$

2. *If A is a quasi-diagonal matrix*

$$A=\{A_1, A_2, \ldots, A_u\},$$

then

$$f(A)=\{f(A_1), f(A_2), \ldots, f(A_u)\}.$$

Let us denote by $r(\lambda)$ the Lagrange-Sylvester interpolation polynomial of $f(\lambda)$ on the spectrum of A. Then it is easy to see that

$$f(A)=r(A)=\{r(A_1), r(A_2), \ldots, r(A_u)\}. \tag{7}$$

On the other hand, the minimal polynomial $\psi(\lambda)$ of A is an annihilating polynomial for each of the matrices $A_1, A_2, \ldots, A_u$. Therefore it follows from the equation

$$f(A_A)=r(A_A)$$

that

$$f(A_{A_1})=r(A_{A_1}), \ldots, f(A_{A_u})=r(A_{A_u}).$$

Therefore

$$f(A_1)=r(A_1), \ldots, f(A_u)=r(A_u),$$

and equation (7) can be written as follows:

$$f(A)=\{f(A_1), f(A_2), \ldots, f(A_u)\}. \tag{8}$$

Example 1: If the matrix A is of simple structure

$$A=T\{\lambda_1, \lambda_2, \ldots, \lambda_n\}T^{-1},$$

then

$$f(A)=T\{f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_n)\}T^{-1}.$$

$f(A)$ has meaning if the function $f(\lambda)$ is defined at $\lambda_1, \lambda_2, \ldots, \lambda_n$.

[6] From $B=T^{-1}AT$ it follows that $B^k=T^{-1}A^kT$ $(k=0, 1, 2, \ldots)$. Hence for every polynomial $g(\lambda)$ we have $g(B)=T^{-1}g(A)T$. Therefore it follows from $g(A)=O$ that $g(B)=O$, and vice versa.

Example 2: Let J be a matrix of the following quasi-diagonal form

$$J=\begin{Vmatrix} \overbrace{\begin{matrix}\lambda_1 & 1 & 0 & \dots & 0\\ 0 & \lambda_1 & 1 & \dots & 0\\ \dots & \dots & \dots & \dots & \dots\\ 0 & 0 & 0 & \dots & 1\\ 0 & 0 & 0 & \dots & \lambda_1\end{matrix}}^{\nu_1} & & \\ & \ddots & \\ & & \overbrace{\begin{matrix}\lambda_u & 1 & 0 & \dots & 0\\ 0 & \lambda_u & 1 & \dots & 0\\ \dots & \dots & \dots & \dots & \dots\\ 0 & 0 & 0 & \dots & 1\\ 0 & 0 & 0 & \dots & \lambda_u\end{matrix}}^{\nu_u} \end{Vmatrix}$$

All the elements in the non-diagonal blocks are zero. By (8) (see also the example on pp. 12-13),

$$f(J)=\begin{Vmatrix} f(\lambda_1) & \frac{f'(\lambda_1)}{1!} & \dots & \frac{f^{(\nu_1-1)}(\lambda_1)}{(\nu_1-1)!} & \\ 0 & f(\lambda_1) & \dots & \vdots & \\ \vdots & \ddots & \ddots & \frac{f'(\lambda_1)}{1!} & \\ 0 & 0 & \dots & f(\lambda_1) & \\ & & & & \ddots \\ & & & & \begin{matrix} f(\lambda_u) & \frac{f'(\lambda_u)}{1!} & \dots & \frac{f^{(\nu_u-1)}(\lambda_u)}{(\nu_u-1)!}\\ 0 & f(\lambda_u) & \dots & \vdots\\ \vdots & \ddots & \ddots & \frac{f'(\lambda_u)}{1!}\\ 0 & 0 & \dots & f(\lambda_u)\end{matrix} \end{Vmatrix}$$

OBSERVATORIO DE S. PAULO-BIBLIOTECA-N.º

Here, as in the matrix J, all the elements in the non-diagonal blocks are also zero.[7]

[7] It will be established later (Chapter VI, § 6 or Chapter VII, § 7) that an arbitrary matrix $A=\| a_{ik} \|_1^n$ is always similar to some matrix of the form J : $A=TJT^{-1}$. Therefore (see 1, on p. 98) we always have $f(A)=Tf(J)T^{-1}$.

§ 2. The Lagrange-Sylvester Interpolation Polynomial

1. To begin with, we consider the case in which the characteristic equation $|\lambda E-A|=0$ has no multiple roots. The roots of this equation—the characteristic values of the matrix A—will be denoted by $\lambda_1, \lambda_2, \dots, \lambda_n$. Then

$$\psi(\lambda)=|\lambda E-A|=(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n),$$

and condition (6) can be written as follows:

$$r(\lambda_k)=f(\lambda_k) \qquad (k=1, 2, \dots, n).$$

In this case, $r(\lambda)$ is the ordinary Lagrange interpolation polynomial for the function $f(\lambda)$ at the points $\lambda_1, \lambda_2, \dots, \lambda_n$:

$$r(\lambda)=\sum_{k=1}^{n}\frac{(\lambda-\lambda_1)\cdots(\lambda-\lambda_{k-1})(\lambda-\lambda_{k+1})\cdots(\lambda-\lambda_n)}{(\lambda_k-\lambda_1)\cdots(\lambda_k-\lambda_{k-1})(\lambda_k-\lambda_{k+1})\cdots(\lambda_k-\lambda_n)}f(\lambda_k).$$

By Definition 1′

$$f(A)=r(A)=\sum_{k=1}^{n}\frac{(A-\lambda_1E)\cdots(A-\lambda_{k-1}E)(A-\lambda_{k+1}E)\cdots(A-\lambda_nE)}{(\lambda_k-\lambda_1)\cdots(\lambda_k-\lambda_{k-1})(\lambda_k-\lambda_{k+1})\cdots(\lambda_k-\lambda_n)}f(\lambda_k).$$

2. Let us assume now that the characteristic polynomial has multiple roots, but that the minimal polynomial, which is a divisor of the characteristic polynomial, has only simple roots:[8]

$$\psi(\lambda)=(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_m).$$

In this case (as in the preceding one) all the exponents m_k in (1) are equal to 1, and the equation (6) takes the form

$$r(\lambda_k)=f(\lambda_k) \qquad (k=1, 2, \dots, m).$$

$r(\lambda)$ is again the ordinary Lagrange interpolation polynomial and

$$f(A)=\sum_{k=1}^{m}\frac{(A-\lambda_1E)\cdots(A-\lambda_{k-1}E)(A-\lambda_{k+1}E)\cdots(A-\lambda_mE)}{(\lambda_k-\lambda_1)\cdots(\lambda_k-\lambda_{k-1})(\lambda_k-\lambda_{k+1})\cdots(\lambda_k-\lambda_m)}f(\lambda_k).$$

3. We now consider the general case:

$$\psi(\lambda)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_s)^{m_s} \qquad (m_1+m_2+\dots+m_s=m).$$

We represent the rational function $\frac{r(\lambda)}{\psi(\lambda)}$, where the degree of $r(\lambda)$ is less than the degree of $\psi(\lambda)$, as a sum of partial fractions:

[8] See footnote 4.

$$\frac{r(\lambda)}{\psi(\lambda)}=\sum_{k=1}^{s}\left[\frac{\alpha_{k1}}{(\lambda-\lambda_k)^{m_k}}+\frac{\alpha_{k2}}{(\lambda-\lambda_k)^{m_k-1}}+\cdots+\frac{\alpha_{km_k}}{\lambda-\lambda_k}\right], \tag{9}$$

where α_{kj} $(j=1, 2, \ldots, m_k;\ k=1, 2, \ldots, s)$ are certain constants.

In order to determine the numerators α_{kj} of the partial fractions we multiply both sides of (9) by $(\lambda-\lambda_k)^{m_k}$ and denote by $\psi_k(\lambda)$ the polynomial $\frac{\psi(\lambda)}{(\lambda-\lambda_k)^{m_k}}$. Then we obtain:

$$\frac{r(\lambda)}{\psi_k(\lambda)}=\alpha_{k1}+\alpha_{k2}(\lambda-\lambda_k)+\cdots+\alpha_{km_k}(\lambda-\lambda_k)^{m_k-1}+ \\ +(\lambda-\lambda_k)^{m_k}\varrho_k(\lambda) \qquad (k=1, 2, \ldots, s), \tag{10}$$

where $\varrho_k(\lambda)$ is a rational function, regular for $\lambda=\lambda_k$.[9]

Hence

$$\left.\begin{aligned}\alpha_{k1}&=\left[\frac{r(\lambda)}{\psi_k(\lambda)}\right]_{\lambda=\lambda_k},\\ \alpha_{k2}=\left[\frac{r(\lambda)}{\psi_k(\lambda)}\right]'_{\lambda=\lambda_k}&=r(\lambda_k)\left[\frac{1}{\psi_k(\lambda)}\right]'_{\lambda=\lambda_k}+r'(\lambda_k)\frac{1}{\psi_k(\lambda_k)},\ \ldots (k=1, 2, \ldots s).\end{aligned}\right\} \tag{11}$$

Formulas (11) show that the numerators α_{kj} on the right-hand side of (9) are expressible in terms of the values of the polynomial $r(\lambda)$ on the spectrum of A, and these values are known: they are equal to the corresponding values of the function $f(\lambda)$ and its derivatives. Therefore

$$\alpha_{k1}=\frac{f(\lambda_k)}{\psi_k(\lambda_k)},\ \alpha_{k2}=f(\lambda_k)\left[\frac{1}{\psi_k(\lambda)}\right]'_{\lambda=\lambda_k}+f'(\lambda_k)\frac{1}{\psi_k(\lambda_k)},\ \ldots \\ (k=1, 2, \ldots, s). \tag{12}$$

Formulas (12) may be abbreviated as follows:

$$\alpha_{kj}=\frac{1}{(j-1)!}\left[\frac{f(\lambda)}{\psi_k(\lambda)}\right]^{(j-1)}_{\lambda=\lambda_k} \qquad (j=1, 2, \ldots, m_k;\quad k=1, 2, \ldots, s). \tag{13}$$

When all the α_{kj} have been found, we can determine $r(\lambda)$ from the following formula, which is obtained from (9) by multiplying both sides by $\psi(\lambda)$:

$$r(\lambda)=\sum_{k=1}^{s}[\alpha_{k1}+\alpha_{k2}(\lambda-\lambda_k)+\cdots+\alpha_{km_k}(\lambda-\lambda_k)^{m_k-1}]\psi_k(\lambda). \tag{14}$$

In this formula the expression in brackets that multiplies $\psi_k(\lambda)$ is, by (13), equal to the sum of the first m_k terms of the Taylor expansion of $f(\lambda)$ in powers of $(\lambda-\lambda_k)$.

[9] I.e., that does not become infinite for $\lambda=\lambda_k$.

Note. The Lagrange-Sylvester interpolation polynomial can be obtained by a limiting process from the Lagrange interpolation polynomial.

Let

$$\psi(\lambda)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_s)^{m_s} \quad (m=\sum_{k=1}^{s}m_k).$$

We denote the Lagrange interpolation polynomial constructed for the m points

$$\lambda_1^{(1)}, \lambda_1^{(2)}, \ldots, \lambda_1^{(m_1)};\ \lambda_2^{(1)}, \lambda_2^{(2)}, \ldots, \lambda_2^{(m_2)};\ \ldots;\ \lambda_s^{(1)}, \lambda_s^{(2)}, \ldots, \lambda_s^{(m_s)}$$

by

$$L(\lambda)=L\begin{pmatrix}\lambda_1^{(1)}, \ldots, \lambda_1^{(m_1)};\ \ldots;\ \lambda_s^{(1)}, \ldots, \lambda_s^{(m_s)};\\ f(\lambda_1^{(1)}), \ldots, f(\lambda_1^{(m_1)});\ \ldots;\ f(\lambda_s^{(1)}), \ldots, f(\lambda_s^{(m_s)});\end{matrix}\lambda\Big).$$

Then it is not difficult to show that the required Lagrange-Sylvester polynomial is determined by the formula

$$r(\lambda)=\lim_{\substack{\lambda_1^{(1)}, \ldots, \lambda_1^{(m_1)}\to\lambda_1\\ \cdots\cdots\cdots\\ \lambda_s^{(1)}, \ldots, \lambda_s^{(m_s)}\to\lambda_s}} L(\lambda).$$

Example:

$$\psi(\lambda)=(\lambda-\lambda_1)^2(\lambda-\lambda_2)^3 \qquad (m=5).$$

Then

$$\frac{r(\lambda)}{\psi(\lambda)}=\frac{\alpha}{(\lambda-\lambda_1)^2}+\frac{\beta}{\lambda-\lambda_1}+\frac{\gamma}{(\lambda-\lambda_2)^3}+\frac{\delta}{(\lambda-\lambda_2)^2}+\frac{\varepsilon}{\lambda-\lambda_2}.$$

Hence

$$r(\lambda)=[\alpha+\beta(\lambda-\lambda_1)](\lambda-\lambda_2)^3+[\gamma+\delta(\lambda-\lambda_2)+\varepsilon(\lambda-\lambda_2)^2](\lambda-\lambda_1)^2$$

and therefore

$$r(A)=[\alpha E+\beta(A-\lambda_1E)](A-\lambda_2E)^3+[\gamma E+\delta(A-\lambda_2E)+\varepsilon(A-\lambda_2E)^2](A-\lambda_1E)^2.$$

α, β, γ, δ, and ε can be found from the following formulas:

$$\alpha=\frac{f(\lambda_1)}{(\lambda_1-\lambda_2)^3},\quad \beta=-\frac{3}{(\lambda_1-\lambda_2)^4}f(\lambda_1)+\frac{1}{(\lambda_1-\lambda_2)^3}f'(\lambda_1),$$
$$\gamma=\frac{f(\lambda_2)}{(\lambda_2-\lambda_1)^2},\quad \delta=-\frac{2}{(\lambda_2-\lambda_1)^3}f(\lambda_2)+\frac{1}{(\lambda_2-\lambda_1)^2}f'(\lambda_2),$$
$$\varepsilon=-\frac{3}{(\lambda_2-\lambda_1)^4}f(\lambda_2)-\frac{2}{(\lambda_2-\lambda_1)^3}f'(\lambda_2)+\frac{1}{2}\frac{1}{(\lambda_2-\lambda_1)^2}f''(\lambda_2).$$

§ 3. Other Forms of the Definition of $f(A)$. The Components of the Matrix A

1. Let us return to the formula (14) for $r(\lambda)$. When we substitute in (14) the expressions (12) for the coefficients α and combine the terms that contain one and the same value of the function $f(\lambda)$ or of one of its derivatives, we represent $r(\lambda)$ in the form

$$r(\lambda)=\sum_{k=1}^{s}\left[f(\lambda_k)\,\varphi_{k1}(\lambda)+f'(\lambda_k)\,\varphi_{k2}(\lambda)+\cdots+f^{(m_k-1)}(\lambda_k)\,\varphi_{k\,m_k}(\lambda)\right]. \quad (15)$$

Here $\varphi_{kj}(\lambda)$ $(j=1, 2, \ldots, m_k;\ k=1, 2, \ldots, s)$ are easily computable polynomials in λ of degree less than m. These polynomials are completely determined when $\psi(\lambda)$ is given and do not depend on the choice of the function $f(\lambda)$. The number of these polynomials is equal to the number of values of the function $f(\lambda)$ on the spectrum of A, i.e., equal to m (m is the degree of the minimal polynomial $\psi(\lambda)$). The functions $\varphi_{kj}(\lambda)$ represent the Lagrange-Sylvester interpolation polynomial for the function whose values on the spectrum of A are all equal to zero with the exception of $f^{(j-1)}(\lambda_k)$, which is equal to 1.

All the polynomials $\varphi_{kj}(\lambda)$ $(j=1, 2, \ldots, m_k;\ k=1, 2, \ldots, s)$ are linearly independent. For suppose that

$$\sum_{k=1}^{s}\sum_{j=1}^{m_k}c_{kj}\varphi_{kj}(\lambda)=0.$$

Let us determine the interpolation polynomial $r(\lambda)$ from the m conditions:

$$r^{(j-1)}(\lambda_k)=c_{kj} \qquad (j=1, 2, \ldots, m_k;\ k=1, 2, \ldots, s). \quad (16)$$

Then by (15) and (16)

$$r(\lambda)=\sum_{k=1}^{s}\sum_{j=1}^{m_k}c_{kj}\varphi_{kj}(\lambda)=0$$

and, therefore, by (16)

$$c_{kj}=0 \qquad (j=1, 2, \ldots, m_k;\ k=1, 2, \ldots, s).$$

From (15) we deduce the *fundamental formula for* $f(A)$:

$$f(A)=\sum_{k=1}^{s}\left[f(\lambda_k)\,Z_{k1}+f'(\lambda_k)\,Z_{k2}+\cdots+f^{(m_k-1)}(\lambda_k)\,Z_{km_k}\right], \quad (17)$$

where

$$Z_{kj}=\varphi_{kj}(A) \quad (j=1, 2, \ldots, m_k;\ k=1, 2, \ldots, s). \quad (18)$$

The matrices Z_{kj} are completely determined when A is given and do not depend on the choice of the function $f(\lambda)$. On the right-hand side of (17) the function $f(\lambda)$ is represented only by its values on the spectrum of A.

The matrices Z_{kj} $(j=1, 2, \ldots, m_k;\ k=1, 2, \ldots, s)$ will be called the *constituent matrices* or *components* of the given matrix A.

The components Z_{kj} are linearly independent.

For suppose that

$$\sum_{k=1}^{s}\sum_{j=1}^{m_k}c_{kj}Z_{kj}=O.$$

Then by (18)

$$\chi(A)=O, \quad (19)$$

where

$$\chi(\lambda)=\sum_{k=1}^{s}\sum_{j=1}^{m_k}c_{kj}\varphi_{kj}(\lambda). \quad (20)$$

Since by (20) the degree of $\chi(\lambda)$ is less than m, the degree of the minimal polynomial $\psi(\lambda)$, it follows from (19) that

$$\chi(\lambda)\equiv 0.$$

But then, since the m functions $\varphi_{kj}(\lambda)$ are linearly independent, (20) implies that

$$c_{kj}=0 \qquad (j=1, 2, \ldots, m_k;\ k=1, 2, \ldots, s),$$

and this is what we had to prove.

2. From the linear independence of the constituent matrices Z_{kj} it follows, among other things, that none of these matrices can be zero. Let us also note that any two components Z_{kj} are permutable among each other and with A, because they are all scalar polynomials in A.

The formula (17) for $f(A)$ is particularly convenient to use when it is necessary to deal with several functions of one and the same matrix A, or when the function $f(\lambda)$ depends not only on λ, but also on some parameter t. In the latter case, the components Z_{kj} on the right-hand side of (17) do not depend on t, and the parameter t enters only into the scalar coefficients of the matrices.

In the example at the end of § 2, where $\psi(\lambda)=(\lambda-\lambda_1)^2(\lambda-\lambda_2)^3$, we may represent $r(\lambda)$ in the form

$$r(\lambda)=f(\lambda_1)\,\varphi_{11}(\lambda)+f'(\lambda_1)\,\varphi_{12}(\lambda)+f(\lambda_2)\,\varphi_{21}(\lambda)+f'(\lambda_2)\,\varphi_{22}(\lambda)+f''(\lambda_2)\,\varphi_{23}(\lambda),$$

where

$$\varphi_{11}(\lambda)=\left(\frac{\lambda-\lambda_2}{\lambda_1-\lambda_2}\right)^3\left[1-\frac{3(\lambda-\lambda_1)}{\lambda_1-\lambda_2}\right],\qquad \varphi_{12}(\lambda)=\frac{(\lambda-\lambda_1)(\lambda-\lambda_2)^3}{(\lambda_1-\lambda_2)^3},$$

$$\varphi_{21}(\lambda)=\left(\frac{\lambda-\lambda_1}{\lambda_2-\lambda_1}\right)^2\left[1-\frac{2(\lambda-\lambda_2)}{\lambda_2-\lambda_1}+\frac{3(\lambda-\lambda_2)^2}{(\lambda_2-\lambda_1)^2}\right],$$

$$\varphi_{22}(\lambda)=\frac{(\lambda-\lambda_1)^2(\lambda-\lambda_2)}{(\lambda_2-\lambda_1)^2}\left[1-\frac{2(\lambda-\lambda_2)}{\lambda_2-\lambda_1}\right],$$

$$\varphi_{23}(\lambda)=\frac{(\lambda-\lambda_1)^2(\lambda-\lambda_2)^2}{2(\lambda_2-\lambda_1)^2}.$$

Therefore

$$f(A)=f(\lambda_1)Z_{11}+f'(\lambda_1)Z_{12}+f(\lambda_2)Z_{21}+f'(\lambda_2)Z_{22}+f''(\lambda_2)Z_{23}$$

where

$$Z_{11}=\varphi_{11}(A)=\frac{1}{(\lambda_1-\lambda_2)^3}(A-\lambda_2E)^2\left[E-\frac{3}{\lambda_1-\lambda_2}(A-\lambda_1E)\right],$$

$$Z_{12}=\varphi_{12}(A)=\frac{1}{(\lambda_1-\lambda_2)^3}(A-\lambda_1E)(A-\lambda_2E)^2,\quad \ldots.$$

3. When the matrix A is given and its components have actually to be found, we can set in the fundamental formula (17) $f(\mu)=\frac{1}{\lambda-\mu}$, where λ is a parameter. Then we obtain

$$(\lambda E-A)^{-1}=\frac{C(\lambda)}{\psi(\lambda)}=\sum_{k=1}^{s}\left[\frac{Z_{k1}}{\lambda-\lambda_k}+\frac{1!\,Z_{k2}}{(\lambda-\lambda_k)^2}+\cdots+\frac{(m_k-1)!\,Z_{km_k}}{(\lambda-\lambda_k)^{m_k}}\right],\qquad (21)$$

where $C(\lambda)$ is the reduced adjoint matrix of $\lambda E-A$ (Chapter IV, § 6).[10]

The matrices $(j-1)!\,Z_{kj}$ are the numerators of the partial fractions in the decomposition (21), and by analogy with (9) they may be expressed by the values of $C(\lambda)$ on the spectrum of A by formulas similar to (11):

$$(m_k-1)!\,Z_{km_k}=\frac{C(\lambda_k)}{\psi_k(\lambda)},\qquad (m_k-2)!\,Z_{k\,m_k-1}=\left[\frac{C(\lambda)}{\psi_k(\lambda)}\right]'_{\lambda=\lambda_k},\quad \ldots.$$

Hence

$$Z_{kj}=\frac{1}{(j-1)!\,(m_k-j)!}\left[\frac{C(\lambda)}{\psi_k(\lambda)}\right]^{(m_k-j)}_{\lambda=\lambda_k}\quad (j=1,2,\ldots,m_k;\ k=1,2,\ldots,s).\qquad (22)$$

When we replace the constituent matrices in (17) by their expressions (22), we can represent the fundamental formula (17) in the form

[10] For $f(\mu)=\frac{1}{\lambda-\mu}$ we have $f(A)=(\lambda E-A)^{-1}$. For $f(A)=r(A)$, where $r(\mu)$ is the Lagrange-Sylvester interpolation polynomial. From the fact that $f(\mu)$ and $r(\mu)$ coincide on the spectrum of A it follows that $(\lambda-\mu)r(\mu)$ and $(\lambda-\mu)f(\mu)=1$ coincide on this spectrum. Hence $(\lambda E-A)r(A)=(\lambda E-A)f(A)=E$.

$$f(A)=\sum_{k=1}^{s}\frac{1}{(m_k-1)!}\left[\frac{C(\lambda)}{\psi_k(\lambda)}f(\lambda)\right]^{(m_k-1)}_{\lambda=\lambda_k}.\qquad (23)$$

Example 1:[11]

$$A=\left\|\begin{array}{ccc|c} 2 & -1 & 1 & 2\\ 0 & 1 & 1 & 2\\ -1 & 1 & 1 & 1\end{array}\right\|,\quad \lambda E-A=\left\|\begin{array}{ccc} \lambda-2 & 1 & -1\\ 0 & \lambda-1 & -1\\ 1 & -1 & \lambda-1\end{array}\right\|.$$

In this case $\Delta(\lambda)=|\lambda E-A|=(\lambda-1)^2(\lambda-2)$. Since the minor of the element in the first row and second column of $\lambda E-A$ is equal to 1, we have $D_2(\lambda)=1$ and, therefore,

$$\psi(\lambda)=\Delta(\lambda)=(\lambda-1)^2(\lambda-2)=\lambda^3-4\lambda^2+5\lambda-2,$$

$$\Psi(\lambda,\mu)=\frac{\psi(\mu)-\psi(\lambda)}{\mu-\lambda}=\mu^2+(\lambda-4)\mu+\lambda^2-4\lambda+5$$

and

$$C(\lambda)=\Psi(\lambda E,A)=A^2+(\lambda-4)A+(\lambda^2-4\lambda+5)E$$

$$=\left\|\begin{array}{ccc|c} 3 & -2 & 2 & 3\\ -1 & 2 & 2 & 3\\ -3 & 3 & 1 & 1\end{array}\right\|+(\lambda-4)\left\|\begin{array}{ccc|c} 2 & -1 & 1 & 2\\ 0 & 1 & 1 & 2\\ -1 & 1 & 1 & 1\end{array}\right\|+(\lambda^2-4\lambda+5)\left\|\begin{array}{ccc} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{array}\right\|.$$

The fundamental formula has in this case the form

$$f(A)=f(1)Z_{11}+f'(1)Z_{12}+f(2)Z_{21}.\qquad (24)$$

Setting $f(\mu)=\frac{1}{\lambda-\mu}$, we find:

$$(\lambda E-A)^{-1}=\frac{C(\lambda)}{\psi(\lambda)}=\frac{Z_{11}}{\lambda-1}+\frac{Z_{12}}{(\lambda-1)^2}+\frac{Z_{21}}{\lambda-2};$$

hence

$$Z_{11}=-C(1)-C'(1),\qquad Z_{12}=-C(1),\qquad Z_{21}=C(2).$$

We now use the above expression for $C(\lambda)$, compute Z_{11}, Z_{12}, Z_{21}, and substitute the results obtained in (24):

[11] The elements of the sum column are printed in italics and are used for checking the computation. When we multiply the rows of A into the sum column of B we obtain the sum column of AB.

$$f(A)=f(1)\begin{vmatrix}1&0&0\\1&0&0\\1&-1&1\end{vmatrix}+f'(1)\begin{vmatrix}1&-1&1\\1&-1&1\\0&0&0\end{vmatrix}+f(2)\begin{vmatrix}0&0&0\\-1&1&0\\-1&1&0\end{vmatrix}$$

$$=\begin{vmatrix}f(1)+f'(1)&-f'(1)&f'(1)\\f(1)+f'(1)-f(2)&-f'(1)+f(2)&f'(1)\\f(1)-f(2)&-f(1)+f(2)&f(1)\end{vmatrix}. \tag{25}$$

Example 2: Let us show that we can determine $f(A)$ starting only from the fundamental formula. Again let

$$A=\begin{vmatrix}2&-1&1\\0&1&1\\-1&1&1\end{vmatrix},\quad \psi(\lambda)=(\lambda-1)^2(\lambda-2).$$

Then

$$f(A)=f(1)Z_1+f'(1)Z_2+f(2)Z_3. \tag{24'}$$

In (24') we substitute for $f(\lambda)$ in succession 1, $\lambda-1$, $(\lambda-1)^2$:

$$Z_1+Z_2+Z_3=E=\begin{vmatrix}1&0&0\\0&1&0\\0&0&1\end{vmatrix},$$

$$Z_2+Z_3=A-E=\begin{vmatrix}1&-1&1\\0&0&1\\-1&1&0\end{vmatrix}\begin{matrix}1\\1\\0\end{matrix},$$

$$Z_3=(A-E)^2=\begin{vmatrix}0&0&0\\-1&1&0\\-1&1&0\end{vmatrix}\begin{matrix}0\\0\\0\end{matrix}.$$

Computing the third equation from the first two term by term, we can determine all the Z. Substituting in (24'), we obtain the expression for $f(A)$.

4. The examples we have analyzed illustrate three methods of practical computation of $f(A)$. In the first method, we found the interpolation polynomial $r(\lambda)$ and put $f(A)=r(A)$. In the second method, we made use of the decomposition (21) and expressed the components Z_{kj} in (17) by the values of the reduced adjoint matrix $C(\lambda)$ on the spectrum of A. In the third method, we started from the fundamental formula (17) and substituted in succession certain simple polynomials for $f(\lambda)$; from the linear equations so obtained we determined the constituent matrices Z_{kj}.

The third method is perhaps the most convenient for practical purposes. In the general case it can be stated as follows:

In (17) we substitute for $f(\lambda)$ successively certain polynomials $g_1(\lambda)$, $g_2(\lambda)$, ..., $g_m(\lambda)$:

$$g_i(A)=\sum_{k=1}^{s}\left[g_i(\lambda_k)Z_{k1}+g_i'(\lambda_k)Z_{k2}+\cdots+g_i^{(m_k-1)}(\lambda_k)Z_{k\,m_k}\right]$$
$$(i=1,2,\ldots,m). \tag{26}$$

From the m equations (26) we determine the matrices Z_{kj} and substitute the expressions so obtained in (17).

The result of eliminating Z_{kj} from the $(m+1)$ equations (26) and (17) can be written in the form

$$\begin{vmatrix}f(A)&f(\lambda_1)\ldots f^{(m_1-1)}(\lambda_1)\ldots f(\lambda_s)\ldots f^{(m_s-1)}(\lambda_s)\\g_1(A)&g_1(\lambda_1)\ldots g_1^{(m_1-1)}(\lambda_1)\ldots g_1(\lambda_s)\ldots g_1^{(m_s-1)}(\lambda_s)\\\cdot&\cdot\\\cdot&\cdot\\\cdot&\cdot\\g_m(A)&g_m(\lambda_1)\ldots g_m^{(m_1-1)}(\lambda_1)\ldots g_m(\lambda_s)\ldots g_m^{(m_s-1)}(\lambda_s)\end{vmatrix}=0.$$

Expanding this determinant with respect to the elements of the first column, we obtain the required expression for $f(A)$. As the factor of $f(A)$ we have here the determinant $\Delta=\left|g_i^{(\rho)}(\lambda_k)\right|$ (in the i-th row of Δ there are found the values of the polynomial $g_i(\lambda)$ on the spectrum of A; $i=1,2,\ldots,m$). In order to determine $f(A)$ we must have $\Delta\neq 0$. This will be so if no linear combination[12] of the polynomials vanishes completely on the spectrum of A, i.e., is divisible by $\psi(\lambda)$.

The condition $\Delta\neq 0$ is always satisfied when the degrees of the polynomial $g_1(\lambda)$, $g_2(\lambda)$, ..., $g_m(\lambda)$ are $0, 1, \ldots, m-1$, respectively.[13]

5. In conclusion, we mention that high powers of a matrix A^n can be conveniently computed by formula (17) by setting $f(\lambda)$ equal to λ^n.[14]

Example: Given the matrix $A=\begin{Vmatrix}5&-4\\4&-3\end{Vmatrix}$ it is required to compute the elements of A^{100}. The minimal polynomial of the matrix is $\psi(\lambda)=(\lambda-1)^2$.

[12] With coefficients not all equal to zero.

[13] In the last example, $m=3$, $g_1(\lambda)=1$, $g_2(\lambda)=\lambda-1$, $g_3(\lambda)=(\lambda-1)^2$.

[14] Formula (17) may also be used to compute the inverse matrix A^{-1}, by setting $f(\lambda)=\frac{1}{\lambda}$ or, what is the same, by setting $\lambda=0$ in (21).

The fundamental formula is

$$f(A)=f(1)Z_1+f'(1)Z_2.$$

Replacing $f(\lambda)$ successively by 1 and $\lambda-1$, we obtain:

$$Z_1=E,\quad Z_2=A-E.$$

Therefore

$$f(A)=f(1)E+f'(1)(A-E).$$

Setting $f(\lambda)=\lambda^{100}$, we find

$$A^{100}=E+100(A-E)=\begin{Vmatrix}1&0\\0&1\end{Vmatrix}+100\begin{Vmatrix}4&-4\\4&-4\end{Vmatrix}=\begin{Vmatrix}401&-400\\400&-399\end{Vmatrix}.$$

§ 4. Representation of Functions of Matrices by means of Series

1. Let $A=\|a_{ik}\|_1^n$ be a matrix with the minimal polynomial (1):

$$\psi(\lambda)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_s)^{m_s}\quad\left(m=\sum_{k=1}^{s}m_k\right).$$

Furthermore, let $f(\lambda)$ be a function and let $f_1(\lambda), f_2(\lambda), \ldots, f_p(\lambda), \ldots$ be a sequence of functions defined on the spectrum of A.

We shall say that *the sequence of functions $f_p(\lambda)$ converges for $p\to\infty$ to some limit on the spectrum of A* if the limits

$$\lim_{p\to\infty}f_p(\lambda_k),\quad \lim_{p\to\infty}f_p'(\lambda_k),\quad \ldots,\quad \lim_{p\to\infty}f_p^{(m_k-1)}(\lambda_k)\qquad(k=1,2,\ldots,s)$$

exist.

We shall say that *the sequence of functions $f_p(\lambda)$ converges for $p\to\infty$ to the function $f(\lambda)$ on the spectrum of A*, and we shall write

$$\lim_{p\to\infty}f_p(A_A)=f(A_A)$$

if

$$\lim_{p\to\infty}f_p(\lambda_k)=f(\lambda_k),\quad \lim_{p\to\infty}f_p'(\lambda_k)=f'(\lambda_k),\quad \ldots,\quad \lim_{p\to\infty}f_p^{(m_k-1)}(\lambda_k)=f^{(m_k-1)}(\lambda_k)$$
$$(k=1,2,\ldots,s).$$

The fundamental formula

$$f(A)=\sum_{k=1}^{s}\left[f(\lambda_k)Z_{k1}+f'(\lambda_k)Z_{k2}+\cdots+f^{(m_k-1)}(\lambda_k)Z_{km_k}\right]$$

expresses $f(A)$ in terms of the values of $f(\lambda)$ on the spectrum of A. If we regard the matrix as a vector in a space $\boldsymbol{R}_{n^2}$ of dimension n^2, then it follows from the fundamental formula, by the linear independence of the matrices Z_{kj}, that all the $f(A)$ (for given A) form an m-dimensional subspace of $\boldsymbol{R}_{n^2}$

with basis Z_{kj} $(j=1,2,\ldots,m_k;\ k=1,2,\ldots,s)$. In this basis the 'vector' $f(A)$ has as its coordinates the m values of the function $f(\lambda)$ on the spectrum of A.

These considerations make the following theorem perfectly obvious:

THEOREM 1: *A sequence of matrices $f_p(A)$ converges for $p\to\infty$ to some limit if and only if the sequence $f_p(\lambda)$ converges for $p\to\infty$ on the spectrum of A to a limit, i.e., the limits*

$$\lim_{p\to\infty}f_p(A)\quad\textit{and}\quad\lim_{p\to\infty}f_p(A_A)$$

always exist simultaneously. Moreover, the equation

$$\lim_{p\to\infty}f_p(A_A)=f(A_A)\tag{27}$$

implies that

$$\lim_{p\to\infty}f_p(A)=f(A)\tag{28}$$

and conversely.

Proof. 1) If the values of $f_p(\lambda)$ converge on the spectrum of A for $p\to\infty$ to limit values, then from the formulas

$$f_p(A)=\sum_{k=1}^{s}\left[f_p(\lambda_k)Z_{k1}+f_p'(\lambda_k)Z_{k2}+\cdots+f_p^{(m_k-1)}(\lambda_k)Z_{km_k}\right]\tag{29}$$

there follows the existence of the limit $\lim_{p\to\infty}f_p(A)$. On the basis of this formula and of (17) we deduce (28) from (27).

2) Suppose, conversely, that $\lim_{p\to\infty}f_p(A)$ exists. Since the m constituent matrices Z are linearly independent, we can express, by (29), the m values of $f_p(\lambda)$ on the spectrum of A (as a linear form) by the m elements of the matrix $f_p(A)$. Hence the existence of the limit $\lim_{p\to\infty}f_p(A_A)$ follows, and (27) holds in the presence of (28).

According to this theorem, if a sequence of polynomials $g_p(\lambda)$ $(p=1,2,3,\ldots)$ converges to the function $f(\lambda)$ on the spectrum of A, then

$$\lim_{p\to\infty}g_p(A)=f(A).$$

2. This formula underlines the naturalness and generality of our definition of $f(A)$. $f(A)$ is always obtained from the $g_p(A)$ by passing to the limit $p\to\infty$, provided only that the sequence of polynomials $g_p(\lambda)$ converges to $f(\lambda)$ on the spectrum of A. The latter condition is necessary for the existence of the limit $\lim_{p\to\infty}g_p(A)$.

We shall say that *the series* $\sum_{p=0}^{\infty} u_p(\lambda)$ *converges on the spectrum of* A *to the function* $f(\lambda)$ and we shall write

$$f(A_A)=\sum_{p=0}^{\infty} u_p(A_A), \tag{30}$$

if all the functions occurring here are defined on the spectrum of A and the following equations hold:

$$f(\lambda_k)=\sum_{p=0}^{\infty} u_p(\lambda_k),\quad f'(\lambda_k)=\sum_{p=0}^{\infty} u_p'(\lambda_k),\quad \ldots,\ f^{(m_k-1)}(\lambda_k)=\sum_{p=0}^{\infty} u_p^{(m_k-1)}(\lambda_k)$$
$$(k=1,2,\ldots,s),$$

where the series on the right-hand sides of these equations converge. In other words, if we set

$$s_p(\lambda)=\sum_{q=0}^{p} u_q(\lambda)\qquad (p=0,1,2,\ldots),$$

then (30) is equivalent to

$$f(A_A)=\lim_{p\to\infty} s_p(A_A). \tag{31}$$

It is obvious that the theorem just proved can be stated in the following equivalent form:

THEOREM 1′: *The series* $\sum_{p=0}^{\infty} u_p(A)$ *converges to a matrix if and only if the series* $\sum_{p=0}^{\infty} u_p(\lambda)$ *converges on the spectrum of* A. *Moreover, the equation*

$$f(A_A)=\sum_{p=0}^{\infty} u_p(A_A)$$

implies that

$$f(A)=\sum_{p=0}^{\infty} u_p(A),$$

and conversely.

3. Suppose a power series is given with the circle of convergence $|\lambda-\lambda_0|<R$ and the sum $f(\lambda)$:

$$f(\lambda)=\sum_{p=0}^{\infty}\alpha_p(\lambda-\lambda_0)^p\qquad (|\lambda-\lambda_0|<R). \tag{32}$$

Since a power series may be differentiated term by term any number of times within the circle of convergence, (32) converges on the spectrum of any matrix whose characteristic values lie within the circle of convergence.

Thus we have:

THEOREM 2: *If the function* $f(\lambda)$ *can be expanded in a power series in the circle* $|\lambda-\lambda_0|<r$,

$$f(\lambda)=\sum_{p=0}^{\infty}\alpha_p(\lambda-\lambda_0)^p, \tag{33}$$

then this expansion remains valid when the scalar argument λ *is replaced by a matrix* A *whose characteristic values lie within the circle of convergence.*

Note. In this theorem we may allow a characteristic value λ_k of A to fall on the circumference of the circle of convergence; but we must then postulate in addition that the series (33), differentiated m_k-1 times term by term, should converge at the point $\lambda=\lambda_k$. It is well known that this already implies the convergence of the j times differentiated series (33) at the point λ_k to $f^{(j)}(\lambda_k)$ for $j=0,1,\ldots,m_k-1$.

The theorem just proved leads, for example, to the following expansions:[15]

$$e^A=\sum_{p=0}^{\infty}\frac{A^p}{p!},\quad \cos A=\sum_{p=0}^{\infty}\frac{(-1)^p}{(2p)!}A^{2p},\quad \sin A=\sum_{p=0}^{\infty}(-1)^p\frac{A^{2p+1}}{(2p+1)!},$$
$$\cosh A=\sum_{p=0}^{\infty}\frac{A^{2p}}{(2p)!},\quad \sinh A=\sum_{p=0}^{\infty}\frac{A^{2p+1}}{(2p+1)!},$$
$$(E-A)^{-1}=\sum_{p=0}^{\infty}A^p\qquad (|\lambda_k|<1;\ k=1,2,\ldots,s),$$
$$\ln A=\sum_{p=1}^{\infty}\frac{(-1)^{p-1}}{p}(A-E)^p\qquad (|\lambda_k-1|<1;\ k=1,2,\ldots,s)$$

(by $\ln\lambda$ we mean here the so-called principal value of the many-valued function $\operatorname{Ln}\lambda$, i.e., that branch for which $\operatorname{Ln}1=0$).

Let $G(u_1,u_2,\ldots,u_l)$ be a polynomial in $u_1,u_2,\ldots,u_l$; let $f_1(\lambda),f_2(\lambda),\ldots,f_l(\lambda)$ be functions of λ defined on the spectrum of the matrix A, and let

$$g(\lambda)\equiv G[f_1(\lambda),f_2(\lambda),\ldots,f_l(\lambda)].$$

Then from

$$g(A_A)=0 \tag{34}$$

there follows:

$$G[f_1(A),f_2(A),\ldots,f_l(A)]=O. \tag{35}$$

[15] The expansions in the first two rows hold for an arbitrary matrix A.

For let us denote by $f_1(\lambda), f_2(\lambda), \ldots, f_l(\lambda)$ the Lagrange-Sylvester interpolation polynomials for $r_1(\lambda), r_2(\lambda), \ldots, r_l(\lambda)$, and let us set:

$$G[f_1(A), f_2(A), \ldots, f_l(A)] = G[r_1(A), r_2(A), \ldots, r_l(A)] = h(A) = O,$$

Then (34) implies

$$h(\lambda) = G[r_1(\lambda), r_2(\lambda), \ldots, r_l(\lambda)],$$

Hence it follows that

$$h(A_A) = 0. \tag{36}$$

and this is what we had to show.

This result allows us to extend identities between functions of a scalar variable to matrix values of the argument.

For example, from

$$\cos^2\lambda + \sin^2\lambda = 1$$

we obtain for an arbitrary matrix A

$$\cos^2 A + \sin^2 A = E$$

(in this case $G(u_1, u_2) = u_1^2 + u_2^2 - 1$, $f_1(\lambda) = \cos\lambda$, and $f_2(\lambda) = \sin\lambda$).

Similarly, for every matrix A

$$e^A e^{-A} = E,$$

i.e.,

$$e^{-A} = (e^A)^{-1}$$

Further, for every matrix A

$$e^{iA} = \cos A + i \sin A$$

Let A be a non-singular matrix ($|A| \neq 0$). We denote by $\sqrt[n]{\lambda}$ the single-valued branch of the many-valued function $\sqrt[n]{\lambda}$ that is defined in a domain not containing the origin and containing all the characteristic values of A. Then $\sqrt[n]{A}$ has a meaning. From $(\sqrt[n]{\lambda})^n - \lambda = 0$ it now follows that

$$(\sqrt[n]{A})^n = A.$$

Let $f(\lambda) = \frac{1}{\lambda}$ and let $A = \| a_{ik} \|_1^n$ be a non-singular matrix. Then $f(\lambda)$ is defined as the spectrum of A, and in the equation

$$\lambda f(\lambda) = 1$$

we can therefore replace λ by A:

i.e.,[16]

$$A \cdot f(A) = E,$$

$$f(A) = A^{-1}.$$

Denoting by $r(\lambda)$ the interpolation polynomial for the function $1/\lambda$ we may represent the inverse matrix A^{-1} in the form of a polynomial in A:

[16] We have already made use of this on p. 109. See footnote 10.

$$A^{-1} = r(A).$$

Let us consider a rational function $\varrho(\lambda) = \frac{g(\lambda)}{h(\lambda)}$, where $g(\lambda)$ and $h(\lambda)$ are co-prime polynomials in λ. This function is defined on the spectrum of A if and only if the characteristic values of A are not roots of $h(\lambda)$, i.e.,[17] if $|h(A)| \neq 0$. Under this assumption we may replace λ by A in the identity

$$\varrho(\lambda) h(\lambda) = g(\lambda),$$

obtaining:

$$\varrho(A) h(A) = g(A).$$

Hence

$$\varrho(A) = g(A)[h(A)]^{-1} = [h(A)]^{-1} g(A). \tag{37}$$

Notes. 1) If $\boldsymbol{A}$ is a linear operator in an n-dimensional space $\boldsymbol{R}$, then $f(\boldsymbol{A})$ is defined exactly like $f(A)$:

$$f(\boldsymbol{A}) = r(\boldsymbol{A}),$$

where $r(\lambda)$ is the Lagrange-Sylvester interpolation polynomial for $f(\lambda)$ on the spectrum of the operator $\boldsymbol{A}$ (the spectrum of $\boldsymbol{A}$ is determined by the minimal annihilating polynomial $\psi(\lambda)$ of $\boldsymbol{A}$).

According to this definition, if the matrix $A = \| a_{ik} \|_1^n$ corresponds to the operator $\boldsymbol{A}$ in some basis of the space, then in the same basis the matrix $f(A)$ corresponds to the operator $f(\boldsymbol{A})$. All the statements of this chapter in which there occurs a matrix A remain valid after replacement of the matrix A by the operator $\boldsymbol{A}$.

2) We can also define[18] a function of a matrix $f(A)$ starting from the characteristic polynomial

$$\Delta(\lambda) = \prod_{k=1}^{s} (\lambda - \lambda_k)^{n_k}$$

instead of the minimal polynomial

$$\psi(\lambda) = \prod_{k=1}^{s} (\lambda - \lambda_k)^{m_k}.$$

[17] See (25) on p. 84.

[18] See, for example, MacMillan, W. D., *Dynamics of Rigid Bodies* (New York, 1936).

We have then to set $f(A) = g(A)$, where $g(\lambda)$ is an interpolation polynomial of degree less than n modulo $\Delta(\lambda)$ of the function $f(\lambda)$.[19] The formulas (17), (21), and (23) are to be replaced by the following[20]

$$f(A) = \sum_{k=1}^{s} \left[f(\lambda_k) \widehat{Z}_{k1} + f'(\lambda_k) \widehat{Z}_{k2} + \cdots + f^{(n_k-1)}(\lambda_k) \widehat{Z}_{kn_k} \right], \tag{17'}$$

$$(\lambda E - A)^{-1} = \frac{B(\lambda)}{\Delta(\lambda)} = \sum_{k=1}^{s} \left[\frac{\widehat{Z}_{k1}}{\lambda - \lambda_k} + \frac{1!\widehat{Z}_{k2}}{(\lambda-\lambda_k)^2} + \cdots + \frac{(n_k-1)!\widehat{Z}_{kn_k}}{(\lambda-\lambda_k)^{n_k}} \right], \tag{21'}$$

$$f(A) = \sum_{k=1}^{s} \frac{1}{(n_k-1)!} \left[\frac{B(\lambda)}{\Delta_k(\lambda)} f(\lambda) \right]^{(n_k-1)}_{\lambda=\lambda_k}, \tag{23'}$$

where

$$\Delta_k(\lambda) = \frac{\Delta(\lambda)}{(\lambda-\lambda_k)^{n_k}} \quad (k = 1, 2, \ldots, s).$$

However, in (17') the values $f^{(m_k)}(\lambda_k), f^{(m_k+1)}(\lambda_k), \ldots, f^{(n_k-1)}(\lambda_k)$ occur only fictitiously, because a comparison of (21) with (21') yields:

$$\widehat{Z}_{k1} = Z_{k1}, \ldots, \widehat{Z}_{km_k} = Z_{km_k}, \widehat{Z}_{km_k+1} = \cdots = \widehat{Z}_{kn_k} = O.$$

§ 5. Application of a Function of a Matrix to the Integration of a System of Linear Differential Equations with Constant Coefficients

1. We begin by considering a system of homogeneous linear differential equations of the first order with constant coefficients:

$$\left.\begin{aligned} \frac{dx_1}{dt} &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ \frac{dx_2}{dt} &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ &\cdots\cdots\cdots\cdots\cdots \\ \frac{dx_n}{dt} &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n, \end{aligned}\right\} \tag{38}$$

where t is the independent variable, $x_1, x_2, \ldots, x_n$ are unknown functions of t, and a_{ik} $(i, k = 1, 2, \ldots, n)$ are complex numbers.

We introduce the square matrix $A = \| a_{ik} \|_1^n$ of the coefficients and the column matrix $x = (x_1, x_2, \ldots, x_n)$. Then the system (38) can be written in the form of a single matrix differential equation

[19] The polynomial $g(\lambda)$ is not uniquely determined by the equation $f(A) = g(A)$ and the condition 'degree less than n.'

[20] The special case of (23') in which $f(\lambda) \equiv \lambda^n$ is sometimes called Perron's formula (see [40], pp. 25-27).

$$\frac{dx}{dt} = Ax. \tag{39}$$

Here, and in what follows, we mean by the *derivative of a matrix* that matrix which is obtained from the given one by replacing all its elements by their derivatives. Therefore $\frac{dx}{dt}$ is the column matrix with the elements $\frac{dx_1}{dt}, \frac{dx_2}{dt}, \ldots, \frac{dx_n}{dt}$.

We shall seek a solution of the system of differential equations satisfying the following initial conditions:

$$x_1|_{t=0} = x_{10}, \; x_2|_{t=0} = x_{20}, \; \ldots, \; x_n|_{t=0} = x_{n0}$$

or, briefly,

$$x|_{t=0} = x_0. \tag{40}$$

Let us expand the unknown column x into a MacLaurin series in powers of t:

$$x = x_0 + \dot{x}_0 t + \ddot{x}_0 \frac{t^2}{2!} + \cdots \quad \left(\dot{x}_0 = \frac{dx}{dt}\Big|_{t=0}, \; \ddot{x}_0 = \frac{d^2x}{dt^2}\Big|_{t=0}, \ldots \right). \tag{41}$$

Then by successive differentiations we find from (39):

$$\frac{d^2x}{dt^2} = A\frac{dx}{dt} = A^2x, \quad \frac{d^3x}{dt^3} = A\frac{d^2x}{dt^2} = A^3x, \quad \ldots. \tag{42}$$

Substituting the value $t = 0$ in (39) and (42), we obtain:

$$\dot{x}_0 = Ax_0, \quad \ddot{x}_0 = A^2x_0, \ldots.$$

Now the series (41) can be written as follows:

$$x = x_0 + tAx_0 + \frac{t^2}{2!}A^2x_0 + \cdots = e^{At}x_0. \tag{43}$$

By direct substitution in (39) we see[21] that (43) is a solution of the differential equation (39). Setting $t = 0$ in (43), we find:

$$x|_{t=0} = x_0.$$

Thus, the formula (43) gives us the solution of the given system of differential equations satisfying the initial conditions (40).

Let us set $f(\lambda) = e^{\lambda t}$ in (17). Then

$$e^{At} = \| q_{ik}(t) \|_1^n = \sum_{k=1}^{s} (Z_{k1} + Z_{k2}t + \cdots + Z_{km_k}t^{m_k-1}) e^{\lambda_k t}. \tag{44}$$

[21] $\frac{d}{dt}(e^{At}) = \frac{d}{dt}\left(E + At + \frac{A^2t^2}{2!} + \cdots\right) = A + A^2t + \frac{A^3t^2}{2!} + \cdots = Ae^{At}$.

The solution (43) may then be written in the following form:

$$\left.\begin{aligned} x_1 &= q_{11}(t)\,x_{10} + q_{12}(t)\,x_{20} + \cdots + q_{1n}(t)\,x_{n0} \\ x_2 &= q_{21}(t)\,x_{10} + q_{22}(t)\,x_{20} + \cdots + q_{2n}(t)\,x_{n0} \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n &= q_{n1}(t)\,x_{10} + q_{n2}(t)\,x_{20} + \cdots + q_{nn}(t)\,x_{n0}, \end{aligned}\right\} \tag{45}$$

where $x_{10}, x_{20}, \ldots, x_{n0}$ are constants equal to the initial values of the unknown functions $x_1, x_2, \ldots, x_n$.

Thus, the integration of the given system of differential equations reduces to the computation of the elements of the matrix e^{At}.

If $t = t_0$ is taken as the initial value of the argument, then (43) is to be replaced by the formula

$$x = e^{A(t-t_0)}\,x_0. \tag{46}$$

Example.

$$\begin{aligned} \frac{dx_1}{dt} &= 3x_1 - x_2 + x_3, \\ \frac{dx_2}{dt} &= 2x_1 \qquad - x_3, \\ \frac{dx_3}{dt} &= x_1 - x_2 + 2x_3. \end{aligned}$$

The coefficient matrix is

$$A = \begin{Vmatrix} 3 & -1 & 1 \\ 2 & 0 & 1 \\ 1 & -1 & 2 \end{Vmatrix}.$$

We form the characteristic determinant

$$\Delta(\lambda) = -\begin{vmatrix} 3-\lambda & 1 & 1 \\ 2 & -\lambda & 1 \\ 1 & -1 & 2-\lambda \end{vmatrix} = (\lambda-1)(\lambda-2)^2.$$

The greatest common divisor of the minors of order 2 is $D_2(\lambda) = 1$. Therefore

$$\varphi(\lambda) = \Delta(\lambda) = (\lambda-1)(\lambda-2)^2.$$

The fundamental formula is

$$f(A) = f(1)\,Z_1 + f(2)\,Z_2 + f'(2)\,Z_3.$$

For $f(\lambda)$ we choose in succession 1, $\lambda - 2$, $(\lambda-2)^2$. We obtain:

$$Z_1 + Z_2 = E = \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix},$$

$$-Z_1 + Z_3 = A - 2E = \left\| \begin{matrix} 1 & -1 & 1 \\ 2 & -2 & 1 \\ 1 & -1 & 0 \end{matrix} \right\| \begin{matrix} 1 \\ 1 \\ 0 \end{matrix},$$

$$Z_1 = (A-2E)^2 = \left\| \begin{matrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{matrix} \right\| \begin{matrix} 0 \\ 0 \\ 0 \end{matrix}.$$

Hence we determine Z_1, Z_2, Z_3 and substitute in the fundamental formula

$$f(A) = f(1)\begin{Vmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{Vmatrix} + f(2)\begin{Vmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \end{Vmatrix} + f'(2)\begin{Vmatrix} 1 & -1 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 0 \end{Vmatrix}.$$

If we now replace $f(\lambda)$ by $e^{\lambda t}$, we obtain:

$$e^{At} = e^t\begin{Vmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{Vmatrix} + e^{2t}\begin{Vmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \end{Vmatrix} + te^{2t}\begin{Vmatrix} 1 & -1 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 0 \end{Vmatrix} = \begin{Vmatrix} (1+t)e^{2t} & -te^{2t} & te^{2t} \\ -e^t + (1+t)e^{2t} & e^t - te^{2t} & te^{2t} \\ -e^t + e^{2t} & e^t - e^{2t} & e^{2t} \end{Vmatrix}.$$

Thus

$$\begin{aligned} x_1 &= C_1(1+t)e^{2t} - C_2te^{2t} + C_3te^{2t}, \\ x_2 &= C_1[-e^t + (1+t)e^{2t}] + C_2(e^t - te^{2t}) + C_3te^{2t}, \\ x_3 &= C_1(-e^t + e^{2t}) + C_2(e^t - e^{2t}) + C_3e^{2t}, \end{aligned}$$

where

$$C_1 = x_{10}, \quad C_2 = x_{20}, \quad C_3 = x_{30}.$$

2. We now consider a system of inhomogeneous linear differential equations with constant coefficients:

$$\left.\begin{aligned} \frac{dx_1}{dt} &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + f_1(t) \\ \frac{dx_2}{dt} &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + f_2(t) \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \frac{dx_n}{dt} &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n + f_n(t), \end{aligned}\right\} \tag{47}$$

where $f_i(t)$ $(i=1, 2, \ldots, n)$ are continuous functions in the interval $t_0 \leqq t \leqq t_1$. Denoting by $f(t)$ the column matrix with the elements $f_1(t), f_2(t), \ldots, f_n(t)$ and again setting $A = \| a_{ik} \|_1^n$, we write the system (47) as follows:

$$\frac{dx}{dt} = Ax + f(t). \tag{48}$$

We replace x by a new column z of unknown functions, connected with x by the relation

$$x = e^{At}z \tag{49}$$

Differentiating (49) term by term and substituting the expression for $\frac{dx}{dt}$ in (48) we find:[22]

$$e^{At}\frac{dz}{dt} = f(t). \tag{50}$$

Hence[23]

$$z(t) = c + \int_{t_0}^{t} e^{-A\tau} f(\tau)\, d\tau \tag{51}$$

and so by (49)

$$x = e^{At}\left[c + \int_{t_0}^{t} e^{-A\tau} f(\tau)\, d\tau\right] = e^{At}c + \int_{t_0}^{t} e^{A(t-\tau)} f(\tau)\, d\tau; \tag{52}$$

where c is a column with arbitrary constant elements.

When we give to the argument t in (52) the value t_0 we find $c = e^{-At_0}x_0$; so that (52) can be written in the following form:

$$x = e^{A(t-t_0)}x_0 + \int_{t_0}^{t} e^{A(t-\tau)} f(\tau)\, d\tau. \tag{53}$$

[22] See footnote 21.

[23] If a matrix function of a scalar argument is given, $B(\tau) = \| b_{ik}(\tau) \|$ $(i=1, 2, \ldots, m;\ k=1, 2, \ldots, n;\ t_1 \leqq \tau \leqq t_2)$, then the *integral* $\int_{t_1}^{t_2} B(\tau)\, d\tau$ is defined in the natural way:

$$\int_{t_1}^{t_2} B(\tau)\, d\tau = \left\| \int_{t_1}^{t_2} b_{ik}(\tau)\, d\tau \right\| \qquad (i=1, 2, \ldots, m;\ k=1, 2, \ldots, n).$$

Setting $e^{At} = \| q_{ik}(t) \|_1^n$, we can write the solution (53) in expanded form:

$$\left.\begin{aligned}
x_1 &= q_{11}(t-t_0)\, x_{10} + \cdots + q_{1n}(t-t_0)\, x_{n0} + \\
&\quad + \int_{t_0}^{t} [q_{11}(t-\tau) f_1(\tau) + \cdots + q_{1n}(t-\tau) f_n(\tau)]\, d\tau \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
x_n &= q_{n1}(t-t_0)\, x_{10} + \cdots + q_{nn}(t-t_0)\, x_{n0} + \\
&\quad + \int_{t_0}^{t} [q_{n1}(t-\tau) f_1(\tau) + \cdots + q_{nn}(t-\tau) f_n(\tau)]\, d\tau.
\end{aligned}\right\} \tag{54}$$

3. As an example we consider *the motion of a heavy material point in a vacuum near the surface of the earth, taking the motion of the earth into account.* It is known[24] that in this case the acceleration of the point relative to the earth is determined by the constant force of gravity $m\boldsymbol{g}$ and the inertial Coriolis force $-2m\omega \times \boldsymbol{v}$ ($\boldsymbol{v}$ is the velocity of the point relative to the earth, ω the constant angular velocity of the earth). Therefore the differential equation of motion of the point has the form[25]

$$\frac{d\boldsymbol{v}}{dt} = \boldsymbol{g} - 2\omega \times \boldsymbol{v}. \tag{55}$$

We define a linear operator $\boldsymbol{A}$ in three-dimensional euclidean space by the equation

$$\boldsymbol{A}\boldsymbol{x} = -2\omega \times \boldsymbol{x} \tag{56}$$

and write instead of (55)

$$\frac{d\boldsymbol{v}}{dt} = \boldsymbol{A}\boldsymbol{v} + \boldsymbol{g}. \tag{57}$$

Comparing (57) with (48), we easily find from (53):

$$\boldsymbol{v} = e^{\boldsymbol{A}t}\boldsymbol{v}_0 + \int_0^t e^{\boldsymbol{A}\tau}\, d\tau \cdot \boldsymbol{g} \qquad (\boldsymbol{v}_0 = \boldsymbol{v}|_{t=0}).$$

Integrating term by term, we determine the radius vector of the motion of the point:

$$\boldsymbol{r} = \boldsymbol{r}_0 + \int_0^t e^{\boldsymbol{A}\tau}\, d\tau \boldsymbol{v}_0 + \int_0^t \int_0^\tau e^{\boldsymbol{A}\sigma}\, d\sigma\, d\tau\, \boldsymbol{g}, \tag{58}$$

where

$$\boldsymbol{r}_0 = \boldsymbol{r}|_{t=0} \quad \text{and} \quad \boldsymbol{v}_0 = \boldsymbol{v}|_{t=0}.$$

[24] See A. Sommerfeld, *Lectures on Theoretical Physics*, Vol. I (Mechanics), § 30.

[25] Here the symbol $\times$ denotes the vector product.

Substituting for e^{At} the series

$$E + A\frac{t}{1!} + A^2\frac{t^2}{2!} + \cdots$$

and replacing A by its expression from (56), we have:

$$r = r_0 + v_0 t + \frac{1}{2} g t^2 - \omega \times \left(v_0 t^2 + \frac{1}{3} g t^3 \right) + \omega \times \left[\omega \times \left(\frac{2}{3} v_0 t^3 + \frac{1}{6} g t^4 \right) \right] + \cdots .$$

Considering that the angular velocity ω is small (for the earth, $\omega \approx 7.3 \times 10^{-5}$ sec^{-1}), we neglect the terms containing the second and higher powers of ω; for the additional displacement of the point due to the motion of the earth we then obtain the approximate formula

$$d = -\omega \times \left(v_0 t^2 + \frac{1}{3} g t^3 \right).$$

Returning to the exact solution (58), let us compute e^{At}. As a preliminary we establish that the minimal polynomial of the operator A has the form

$$\psi(\lambda) = \lambda(\lambda^2 + 4\omega^2).$$

For we find from (56)

$$A^2 x = 4\omega \times (\omega \times x) = 4(\omega x)\omega - 4\omega^2 x,$$
$$A^3 x = -2\omega \times A^2 x = 8\omega^2 (\omega \times x).$$

Hence and from (56) it follows that the operators E, A, A^2 are linearly independent and that

$$A^3 + 4\omega^2 A = O.$$

The minimal polynomial $\psi(\lambda)$ has the simple roots $0, 2\omega i, -2\omega i$. The Lagrange interpolation formula for e^{At} has the form

$$1 + \frac{\sin 2\omega t}{2\omega}\lambda + \frac{1 - \cos 2\omega t}{4\omega^2}\lambda^2.$$

Then

$$e^{At} = E + \frac{\sin 2\omega t}{2\omega} A + \frac{1 - \cos 2\omega t}{4\omega^2} A^2.$$

Substituting this expression for e^{At} in (58) and replacing the operator A by its expression from (56), we find

$$r = r_0 + v_0 t + g\frac{t^2}{2} - \omega \times \left(\frac{1 - \cos 2\omega t}{2\omega^2} v_0 + \frac{2\omega t - \sin 2\omega t}{4\omega^3} g \right) + \omega \times \left[\omega \times \left(\frac{2\omega t - \sin 2\omega t}{2\omega^3} v_0 + \frac{-1 + 2\omega^2 t^2 + \cos 2\omega t}{4\omega^4} g \right) \right]. \quad (59)$$

Let us consider the special case $v_0 = o$. When we expand the triple vector product we obtain:

$$r = r_0 + g\frac{t^2}{2} + \frac{2\omega t - \sin 2\omega t}{4\omega^3}(g \times \omega) + \frac{\cos 2\omega t - 1 + 2\omega^2 t^2}{4\omega^3}(g \sin\varphi\, \omega - \omega g),$$

where φ is the geographical latitude of the point whose motion we are considering. The term

$$\frac{2\omega t - \sin 2\omega t}{4\omega^3}(g \times \omega)$$

represents the eastward displacement perpendicular to the plane of the meridian, and the last term on the right-hand side of the last formula gives the displacement in the meridian plane perpendicular to, and away from, the earth's axis.

4. Suppose now that the following system of linear differential equations of the second order is given:

$$\left.\begin{aligned} &\frac{d^2x_1}{dt^2} + a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ &\frac{d^2x_2}{dt^2} + a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0 \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\frac{d^2x_n}{dt^2} + a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = 0, \end{aligned}\right\} \quad (60)$$

where the a_{ik} $(i, k = 1, 2, \ldots, n)$ are constant coefficients. Introducing again the column $x = (x_1, x_2, \ldots, x_n)$ and the square matrix $A = \| a_{ik} \|_1^n$, we rewrite (60) in matrix form

$$\frac{d^2x}{dt^2} + Ax = O. \quad (60')$$

We consider, to begin with, the case in which $|A| \neq 0$. If $n = 1$, i.e., if x and A are scalars and $A \neq 0$, the general solution of the equation (60) can be written in the form

$$x = \cos(\sqrt{A}\, t)\, x_0 + (\sqrt{A})^{-1} \sin(\sqrt{A}\, t)\, \dot{x}_0 \quad (61)$$

where $x_0 = x|_{t=0}$ and $\dot{x}_0 = \frac{dx}{dt}\Big|_{t=0}$.

By direct verification we see that (61) is a solution of (60) for arbitrary n, where x is a column and A a non-singular square matrix.[26] Here we use the formulas

[26] By $\sqrt{A}$ we mean a matrix whose square is equal to A. $\sqrt{A}$, we know, exists when $|A| \neq 0$ (see p. 114).

$$\left.\begin{aligned}\cos\left(\sqrt{A}\,t\right) &= E - \frac{1}{2!}At^2 + \frac{1}{4!}A^2t^4 - \cdots,\\ \left(\sqrt{A}\right)^{-1}\sin\left(\sqrt{A}\,t\right) &= Et - \frac{1}{3!}At^3 + \frac{1}{5!}A^2t^5 - \cdots.\end{aligned}\right\} \tag{62}$$

Formula (61) comprises all solutions of the system (60) or (60′), as the initial values x_0 and $\dot{x}_0$ may be chosen arbitrarily.

The right-hand sides of the formulas (62) have a meaning even when $|A| = 0$. Therefore (61) is the general solution of the given system of differential equations also when $|A| = 0$, provided only that the functions $\cos(\sqrt{A}t)$ and $(\sqrt{A})^{-1}\sin(\sqrt{A}t)$, which are part of this expression, are interpreted as the right-hand sides of the formulas (62).

We leave it to the reader to verify that the general solution of the inhomogeneous system

$$\frac{d^2x}{dt^2} + Ax = f(t) \tag{63}$$

satisfying the initial conditions $x|_{t=0} = x_0$ and $\left.\frac{dx}{dt}\right|_{t=0} = \dot{x}_0$ can be written in the form

$$x = \cos\left(\sqrt{A}\,t\right)x_0 + \left(\sqrt{A}\right)^{-1}\sin\left(\sqrt{A}\,t\right)\dot{x}_0 + \\ + \left(\sqrt{A}\right)^{-1}\int_0^t \sin\left[\sqrt{A}\,(t-\tau)\right]f(\tau)\,d\tau. \tag{64}$$

If $t = t_0$ is taken as the initial time, then in (61) and (64) $\cos(\sqrt{A}t)$ and $\sin(\sqrt{A}t)$ must be replaced by $\cos\left(\sqrt{A}\,(t-t_0)\right)$ and $\sin\left(\sqrt{A}\,(t-t_0)\right)$, and $\int_0^t$ by $\int_{t_0}^t$.

In the special case

$$f(t) = h\sin(pt+\alpha)$$

(h is a constant column, and p and α are numbers), (64) can be replaced by:

$$x = \cos\left(\sqrt{A}\,t\right)c + \left(\sqrt{A}\right)^{-1}\sin\left(\sqrt{A}\,t\right)d + (A - p^2E)^{-1}h\sin(pt+\alpha),$$

where c and d are columns with arbitrary constant elements. This formula has meaning when p^2 is not a characteristic value of the matrix A ($|A - p^2E| \neq 0$).

§ 6. Stability of Motion in the Case of a Linear System

1. Let $x_1, x_2, \ldots, x_n$ be parameters that characterize the displacement of 'perturbed' motion of a given mechanical system from an original motion,[27] and suppose that these parameters satisfy a system of differential equations of the first order:

$$\frac{dx_i}{dt} = f_i(x_1, x_2, \ldots, x_n, t) \qquad (i = 1, 2, \ldots, n); \tag{65}$$

the independent variable t in these equations is the time, and the right-hand sides $f_i(x_1, x_2, \ldots, x_n, t)$ are continuous functions of the variables $x_1, \ldots, x_n$ in some domain containing the point $x_1 = 0, x_2 = 0, \ldots, x_n = 0$) for all $t \geqq t_0$ (t_0 is the initial time).

We now introduce the definition of stability of motion according to Lyapunov.[28]

The motion to be investigated is called *stable* if for every $\varepsilon > 0$ we can find a $\delta > 0$ such that for arbitrary initial values of the parameters $x_{10}, x_{20}, \ldots, x_{n0}$ (for $t = t_0$) with moduli less than δ the parameters $x_1, x_2, \ldots, x_n$ remain of moduli less than ε for the whole time of the motion ($t \geqq t_0$), i.e., if for every $\varepsilon > 0$ we can find a $\delta > 0$ such that from

$$|x_{i0}| < \delta \qquad (i = 1, 2, \ldots, n) \tag{66}$$

it follows that

$$|x_i(t)| < \varepsilon \qquad (t \geqq t_0). \tag{67}$$

If, in addition, for some $\delta > 0$ we always have $\lim_{t\to+\infty} x_i(t) = 0$ $(i = 1, 2, \ldots, n)$ as long as $|x_{i0}| < \delta$ $(i = 1, 2, \ldots, n)$, then the motion is called *asymptotically stable.*

We now consider a linear system, i.e., that special case when (65) is a system of linear homogeneous differential equations

$$\frac{dx_i}{dt} = \sum_{k=1}^{n} p_{ik}(t)\,x_k, \tag{68}$$

where the $p_{ik}(t)$ are continuous functions for $t \geqq t_0$ $(i, k = 1, 2, \ldots, n)$.

In matrix form the system (68) can be written as follows:

---

[27] In these parameters, the motion to be studied is characterized by constant zero values $x_1 = 0, x_2 = 0, \ldots, x_n = 0$. Therefore in the mathematical treatment of the problem we speak of the 'stability' of the zero solution of the system (65) of differential equations.

[28] See [14], p. 13; [9], pp. 10-11; or [36], pp. 11-12. See also [3].

$$\frac{dx}{dt} = P(t)\,x, \tag{68'}$$

where x is the column matrix with the elements $x_1, x_2, \ldots, x_n$ and $P(t) = \| p_{ik}(t) \|_1^n$ is the coefficient matrix.

We denote by

$$q_{1j}(t),\ q_{2j}(t),\ \ldots,\ q_{nj}(t) \qquad (j = 1, 2, \ldots, n) \tag{69}$$

n linearly independent solutions of (68).[29] The matrix $Q(t) = \| q_{ij} \|_1^n$ whose columns are these solutions is called an *integral matrix* of the system (68).

Every solution of the system of linear homogeneous differential equations is obtained as a linear combination of n linearly independent solutions with constant coefficients:

$$x_i = \sum_{j=1}^{n} c_j q_{ij}(t) \qquad (i = 1, 2, \ldots, n),$$

or in matrix form,

$$x = Q(t)\,c, \tag{70}$$

where c is the column matrix whose elements are arbitrary constants $c_1, c_2, \ldots, c_n$.

We now choose the special integral matrix for which

$$Q(t_0) = E; \tag{71}$$

in other words, in the choice of n linearly independent solutions of (69) we shall start from the following special initial conditions:[30]

$$q_{ij}(t_0) = \delta_{ij} = \begin{cases} 0 & (i \neq j), \\ 1 & (i = j) \end{cases} \qquad (i, j = 1, 2, \ldots, n).$$

Then setting $t = t_0$ in (70), we find from (71):

$$x_0 = c,$$

and therefore formula (70) assumes the form

$$x = Q(t)\,x_0 \tag{72}$$

or, in expanded form,

$$x_i = \sum_{j=1}^{n} q_{ij}(t)\,x_{j0} \qquad (i = 1, 2, \ldots, n). \tag{72'}$$

[29] Here the second subscript j denotes the number of the solution.

[30] Arbitrary initial conditions determine uniquely a certain solution of a given system.

We consider three cases:

1. $Q(t)$ is a *bounded matrix* in the interval $(t_0, +\infty)$, i.e., there exists a number M such that

$$|q_{ij}(t)| \leqq M \qquad (t \geqq t_0;\ i, j = 1, 2, \ldots, n).$$

In this case it follows from (72') that

$$|x_i(t)| \leqq nM \max |x_{j0}|.$$

The condition of stability is satisfied. (It is sufficient to take $\delta < \frac{\varepsilon}{nM}$ in (66) and (67).) *The motion characterized by the zero solution* $x_1 = 0, x_2 = 0, \ldots, x_n = 0$ *is stable.*

2. $\lim_{t \to +\infty} Q(t) = O$. In this case the matrix $Q(t)$ is bounded in the interval $(t_0, +\infty)$ and therefore, as we have already explained, the motion is stable. Moreover, it follows from (72) that

$$\lim_{t \to +\infty} x(t) = o.$$

for every x_0. *The motion is asymptotically stable.*

3. $Q(t)$ is an *unbounded matrix* in the interval $(t_0, +\infty)$. This means that at least one of the functions $q_{ij}(t)$, say $q_{hk}(t)$, is not bounded in the interval. We take the initial conditions $x_{10} = 0,\ x_{20} = 0, \ldots,\ x_{k-1,0} = 0,$ $x_{k0} \neq 0,\ x_{k+1,0} = 0, \ldots, x_{n0} = 0$. Then

$$x_h(t) = q_{hk}(t)\,x_{k0}.$$

However small in modulus x_{k0} may be, the function $x_h(t)$ is unbounded. The condition (67) is not satisfied for any δ. *The motion is unstable.*

2. We now consider the special case where the coefficients in the system (68) are constants:

$$P(t) = P = \text{const}. \tag{73}$$

We have then (see § 5)

$$x = e^{P(t - t_0)}\,x_0. \tag{74}$$

Comparing (74) with (72), we find that in this case

$$Q(t) = e^{P(t - t_0)}. \tag{75}$$

We denote by

$$\psi(\lambda) = (\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \cdots (\lambda - \lambda_s)^{m_s}$$

the minimal polynomial of the coefficient matrix P.

For the investigation of the integral matrix (75) we apply formula (17) on p. 104. In this case $f(\lambda)=e^{\lambda(t-t_0)}$ (t is regarded as a parameter), $f^{(j)}(\lambda_k)=(t-t_0)^j e^{\lambda_k(t-t_0)}$. Formula (17) yields

$$e^{P(t-t_0)}=\sum_{k=1}^{s}[Z_{k1}+Z_{k2}(t-t_0)+\cdots+Z_{km_k}(t-t_0)^{m_k-1}]e^{\lambda_k(t-t_0)}. \quad (76)$$

We consider three cases:

1. $\operatorname{Re}\lambda_k \leqq 0$ $(k=1,2,\ldots,s)$; and moreover, for all λ_k with $\operatorname{Re}\lambda_k=0$ the corresponding $m_k=1$ (i.e., pure imaginary characteristic values are simple roots of the minimal polynomial).

2. $\operatorname{Re}\lambda_k<0$ $(k=1,2,\ldots,s)$.

3. For some k we have $\operatorname{Re}\lambda_k>0$; or $\operatorname{Re}\lambda_k=0$, but $m_k>1$.

From the formula (76) it follows that in the first case the matrix $Q(t)=e^{P(t-t_0)}$ is bounded in the interval $(t_0,+\infty)$, in the second case $\lim_{t\to+\infty} e^{P(t-t_0)}=0$, and in the third case the matrix $e^{P(t-t_0)}$ is not bounded in the interval $(t_0,+\infty)$.[31]

Therefore in the first case the motion $(x_1=0, x_2=0, \ldots, x_n=0)$ is stable, in the second case it is asymptotically stable, and in the third case it is unstable.

[31] Special consideration is only required in the case when in (76) for $e^{P(t-t_0)}$ there occur several terms of maximal growth (for $t\to+\infty$), i.e., with maximal $\operatorname{Re}\lambda_k=\alpha_0$ and (for the given $\operatorname{Re}\lambda_k=\alpha_0$) maximal value $m_k=m_0$. The expression (76) can be represented in the form

$$e^{P(t-t_0)}=e^{\alpha_0(t-t_0)}(t-t_0)^{m_0-1}\Big[\sum_{j=1}^{r}Z_{kjm_0}e^{i\beta_j(t-t_0)}+(*)\Big],$$

where $\beta_1,\beta_2,\ldots,\beta_r$ are distinct real numbers and $(*)$ denotes a matrix that tends to zero as $t\to+\infty$. From this representation it follows that the matrix $e^{P(t-t_0)}$ is not bounded for $\alpha_0+m_0-1>0$, because the matrix $\sum_{j=1}^{r}Z_{kjm_0}e^{i\beta_j(t-t_0)}$ cannot converge for $t\to+\infty$. We can see this by showing that

$$f(t)=\sum_{j=1}^{r}c_j e^{i\beta_j t},$$

where c_j are complex numbers and β_j real and distinct numbers, can converge to zero for $t\to+\infty$ only when $f(t)\equiv 0$. But, in fact, it follows from $\lim_{t\to+\infty} f(t)=0$ that

$$\sum_{j=1}^{r}|c_j|^2=\lim_{T\to+\infty}\frac{1}{T}\int_0^T|f(t)|^2\,dt=0$$

and therefore

$$c_1=c_2=\cdots=c_n=0.$$

The results of the investigation may be formulated in the form of the following theorem:[32]

THEOREM 3: *The zero solution of the linear system* (68) *for* $P=\text{const.}$ *is stable in the sense of Lyapunov if*

1) *the real parts of all the characteristic values of* P *are negative or zero,*

2) *those characteristic values whose real part is zero, i.e., the pure imaginary characteristic values (if any such exist), are simple roots of the minimal polynomial of* P;

and it is unstable if at least one of the conditions 1), 2) *is violated.*

The zero solution of the linear system (68) *is asymptotically stable if and only if all the characteristic values of* P *have negative real parts.*

The considerations above enable us to make a statement about the nature of the integral matrix $e^{P(t-t_0)}$ in the general case of arbitrary characteristic values of the constant matrix P.

THEOREM 4: *The integral matrix* $e^{P(t-t_0)}$ *of the linear system* (68) *for* $P=\text{const.}$ *is always representable in the form*

$$e^{P(t-t_0)}=Z_-(t)+Z_0+Z_+(t),$$

where

1) $\lim_{t\to+\infty} Z_-(t)=O$,

2) Z_0 *is either constant or is a bounded matrix in the interval* $(t_0,+\infty)$ *that does not have a limit for* $t\to+\infty$,

3) $Z_+(t)=O$ *or* $Z_+(t)$ *is an unbounded matrix in the interval* $(t_0,+\infty)$.

Proof. On the right-hand side of (76) we divide all the summands into three groups. We denote by $Z_-(t)$ the sum of all the terms containing the factors $e^{\lambda_k(t-t_0)}$, with $\operatorname{Re}\lambda_k<0$. We denote by Z_0 the sum of all those matrices Z_{k1} for which $\operatorname{Re}\lambda_k=0$. We denote by $Z_+(t)$ the sum of all the remaining terms. It is easy to see that $Z_-(t)$, $Z_0(t)$, and $Z_+(t)$ have the properties 1), 2), 3) of the theorem.

[32] On the question of sharpening the criteria of stability and instability for quasi-linear systems (i.e., of non-linear systems that become linear after neglecting the non-linear terms), see further Chapter XIV, § 3.

CHAPTER VI

EQUIVALENT TRANSFORMATIONS OF POLYNOMIAL MATRICES. ANALYTIC THEORY OF ELEMENTARY DIVISORS

The first three sections of this chapter deal with the theory of equivalent polynomial matrices. On the basis of this, we shall develop, in the next three sections, the analytical theory of elementary divisors, i.e., the theory of the reduction of a constant (non-polynomial) square matrix A to a normal form $\tilde{A}$ ($A = T\tilde{A}T^{-1}$). In the last two sections of the chapter two methods for the construction of the transforming matrix T will be given.

§ 1. Elementary Transformations of a Polynomial Matrix

1. DEFINITION 1: *A polynomial matrix, or λ-matrix, is a rectangular matrix $A(\lambda)$ whose elements are polynomials in λ:*

$$A(\lambda) = \|a_{ik}(\lambda)\| = \|a_{ik}^{(0)}\lambda^l + a_{ik}^{(1)}\lambda^{l-1} + \cdots + a_{ik}^{(l)}\| \qquad \begin{matrix}(i = 1, 2, \ldots, m;\\ k = 1, 2, \ldots, n);\end{matrix}$$

here l is the largest of the degrees of the polynomials $a_{ik}(\lambda)$.

Setting

$$A_j = \|a_{ik}^{(j)}\| \qquad (i = 1, 2, \ldots, m;\ k = 1, 2, \ldots, n;\ j = 0, 1, \ldots, l),$$

we may represent the polynomial matrix $A(\lambda)$ in the form of a matrix polynomial in λ, i.e., in the form of a polynomial in λ with matrix coefficients:

$$A(\lambda) = A_0\lambda^l + A_1\lambda^{l-1} + \cdots + A_{l-1}\lambda + A_l.$$

We introduce the following *elementary operations* on a polynomial matrix $A(\lambda)$:

1. Multiplication of any row, for example the i-th, by a number $c \neq 0$.

2. Addition to any row, for example the i-th, of any other row, for example the j-th, multiplied by any arbitrary polynomial $b(\lambda)$.

3. Interchange of any two rows, for example the i-th and the j-th.

We leave it to the reader to verify that the operations 1., 2., 3. are equivalent to a multiplication of the polynomial matrix $A(\lambda)$ on the *left* by the following square matrices of order m, respectively:[1]

$$S' = \begin{Vmatrix} 1 & & & & & & & & & 0 \\ & \ddots & & & & & & & & \\ & & 1 & & & & & & & \\ & & & c & & & & & & \\ & & & & 1 & & & & & \\ & & & & & \ddots & & & & \\ & & & & & & 1 & & & \\ & & & & & & & 1 & & \\ & & & & & & & & \ddots & \\ 0 & & & & & & & & & 1 \end{Vmatrix}, \quad S'' = \begin{Vmatrix} 1 & & & & & & & & & 0 \\ & \ddots & & & & & & & & \\ & & 1 & & & & & & & \\ & & & 1 & & b(\lambda) & & & & \\ & & & & 1 & & & & & \\ & & & & & \ddots & & & & \\ & & & & & & 1 & & & \\ & & & & & & & 1 & & \\ & & & & & & & & \ddots & \\ 0 & & & & & & & & & 1 \end{Vmatrix},$$

$$S''' = \begin{Vmatrix} 1 & & & & & & & & & 0 \\ & \ddots & & & & & & & & \\ & & 1 & & & & & & & \\ & & & 0 & & 1 & & & & \\ & & & & 1 & & & & & \\ & & & & & \ddots & & & & \\ & & & & & & 1 & & & \\ & & & 1 & & 0 & & & & \\ & & & & & & & 1 & & \\ & & & & & & & & \ddots & \\ 0 & & & & & & & & & 1 \end{Vmatrix}; \qquad (1)$$

in other words, as the result of applying the operations 1., 2., 3. the matrix $A(\lambda)$ is transformed into $S' \cdot A(\lambda)$, $S'' \cdot A(\lambda)$, and $S''' \cdot A(\lambda)$, respectively. The operations of type 1., 2., 3. are therefore called *left elementary* operations.

In the same way we define the *right elementary* operations on a polynomial matrix (these are performed not on the rows, but on the columns);[2] the matrices (of order n) corresponding to them are:

[1] In the matrices (1) all the elements that are not shown are 1 on the main diagonal and 0 elsewhere.

[2] See footnote 1.

$$T' = \begin{Vmatrix} 1 & & & & & & & & & 0 \\ & \ddots & & & & & & & & \\ & & \ddots & & & & & & & \\ & & & \ddots & & & & & & \\ & & & & c & & & & & \\ & & & & & \ddots & & & & \\ & & & & & & \ddots & & & \\ & & & & & & & \ddots & & \\ 0 & & & & & & & & & 1 \end{Vmatrix} \begin{matrix} \\ \\ \\ \\ (i) \\ \\ \\ \\ \\ \end{matrix},$$

$$T'' = \begin{Vmatrix} 1 & & & & & & & & & 0 \\ & \ddots & & & & & & & & \\ & & \ddots & & & & & & & \\ & & & 1 & & & & & & \\ & & & & \ddots & & & & & \\ & & & b(\lambda) & & \ddots & & & & \\ & & & & & & \ddots & & & \\ & & & & & & & \ddots & & \\ 0 & & & & & & & & & 1 \end{Vmatrix} \begin{matrix} \\ \\ \\ (i) \\ \\ (j) \\ \\ \\ \\ \end{matrix},$$

$$T''' = \begin{Vmatrix} 1 & & & & & & & & & 0 \\ & \ddots & & & & & & & & \\ & & \ddots & & & & & & & \\ & & & 0 & & 1 & & & & \\ & & & & \ddots & & & & & \\ & & & 1 & & 0 & & & & \\ & & & & & & \ddots & & & \\ & & & & & & & \ddots & & \\ 0 & & & & & & & & & 1 \end{Vmatrix} \begin{matrix} \\ \\ \\ (i) \\ \\ (j) \\ \\ \\ \\ \end{matrix}.$$

The result of applying a right elementary operation is equivalent to multiplying the matrix $A(\lambda)$ on the *right* by the corresponding matrix T.

Note that T' and T''' coincide with S' and S''' and that T'' coincides with S'' when the indices i and j are interchanged in these matrices. The matrices of type S', S'', S''' (or, what is the same, T', T'', T''') will be called *elementary matrices.*

The determinant of every elementary matrix does not depend on λ and is different from zero. Therefore each left (right) elementary operation has an inverse operation which is also a left (right) elementary operation.[3]

Definition 2: *Two polynomial matrices $A(\lambda)$ and $B(\lambda)$ are called* 1) *left-equivalent*, 2) *right-equivalent*, 3) *equivalent if one of them can be obtained from the other by means of* 1) *left-elementary*, 2) *right elementary*, 3) *left and right elementary operations, respectively.*[4]

[3] It follows from this that if a matrix $B(\lambda)$ is obtained from $A(\lambda)$ by means of left (right; left and right) elementary operations, then $A(\lambda)$ can, conversely, be obtained from $B(\lambda)$ by means of elementary operations of the same type. The left elementary operations form a group, as do the right elementary operations.

[4] From the definition it follows that only matrices of the same dimensions can be left-equivalent, right-equivalent, or simply equivalent.

Let $B(\lambda)$ be obtained from $A(\lambda)$ by means of the left elementary operations corresponding to $S_1, S_2, \ldots, S_p$. Then

$$B(\lambda) = S_p S_{p-1} \cdots S_1 A(\lambda). \tag{2}$$

Denoting the product $S_p S_{p-1} \cdots S_1$ by $P(\lambda)$, we write (2) in the form

$$B(\lambda) = P(\lambda) A(\lambda), \tag{3}$$

where $P(\lambda)$, like each of the matrices $S_1, S_2, \ldots, S_p$ has a *constant*[5] non-zero determinant.

In the next section we shall prove that every square λ-matrix $P(\lambda)$ with a constant non-zero determinant can be represented in the form of a product of elementary matrices. Therefore (3) is equivalent to (2) and signifies left equivalence of the matrices $A(\lambda)$ and $B(\lambda)$.

In the case of right equivalence of the polynomial matrices $A(\lambda)$ and $B(\lambda)$ we shall have instead of (3) the equation

$$B(\lambda) = A(\lambda) Q(\lambda) \tag{3'}$$

and in the case of (two-sided) equivalence the equation

$$B(\lambda) = P(\lambda) A(\lambda) Q(\lambda). \tag{3''}$$

Here again, $P(\lambda)$ and $Q(\lambda)$ are matrices with non-zero determinants, independent of λ.

Thus, Definition 2 can be replaced by an equivalent definition.

Definition 2': *Two rectangular λ-matrices $A(\lambda)$ and $B(\lambda)$ are called* 1) *left-equivalent*, 2) *right-equivalent*, 3) *equivalent if*

1) $B(\lambda) = P(\lambda) A(\lambda)$,
2) $B(\lambda) = A(\lambda) Q(\lambda)$,
3) $B(\lambda) = P(\lambda) A(\lambda) Q(\lambda)$,

respectively, where $P(\lambda)$ and $Q(\lambda)$ are polynomial square matrices with constant non-zero determinants.

2. All the concepts introduced above are illustrated in the following important example.

We consider a system of m linear homogeneous differential equations of order l with constant coefficients, where $x_1, x_2, \ldots, x_n$ are n unknown functions of the independent variable t:

$$\left.\begin{aligned} a_{11}(D)\, x_1 + a_{12}(D)\, x_2 + \cdots + a_{1n}(D)\, x_n &= 0 \\ a_{21}(D)\, x_1 + a_{22}(D)\, x_2 + \cdots + a_{2n}(D)\, x_n &= 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{m1}(D)\, x_1 + a_{m2}(D)\, x_2 + \cdots + a_{mn}(D)\, x_n &= 0; \end{aligned}\right\} \tag{4}$$

[5] I.e., independent of λ.

here

$$a_{ik}(D)=a_{ik}^{(0)}D^l+a_{ik}^{(1)}D^{l-1}+\cdots+a_{ik}^{(l)} \quad (i=1, 2, \ldots, m;\ k=1, 2, \ldots, n)$$

is a polynomial in D with constant coefficients; $D=\frac{d}{dt}$ is the differential operator.

The matrix of operator coefficients

$$A(D)=\|a_{ik}(D)\| \qquad (i=1, 2, \ldots, m;\ k=1, 2, \ldots, n)$$

is a polynomial matrix, or D-matrix.

Clearly, the left elementary operation 1. on the matrix $A(D)$ signifies term-by-term multiplication of the i-th differential equation of the system by the number $c\neq 0$. The left elementary operation 2. signifies the term-by-term addition to the i-th equation of the j-th equation which has previously been subjected to the differential operator $b(D)$. The left elementary operation 3. signifies an interchange of the i-th and j-th equation.

Thus, if we replace in (4) the matrix $A(D)$ of operator coefficients by a left-equivalent matrix $B(D)$, we obtain a deduced system of equations. Since, conversely, by the same reasoning, the original system is a consequence of the new system, the two systems of equations are equivalent.[6]

It is not difficult in this example to interpret the right elementary operations as well. The first of them signifies the introduction of a new unknown function $\bar{x}_i=\frac{1}{c}x_i$ for the unknown function x_i; the second signifies the introduction of a new unknown function $\bar{x}_j=x_j+b(D)x_i$ (instead of x_j); the third signifies the interchange of the terms in the equations that contain x_i and x_j (i.e., $\bar{x}_i=x_j$, $\bar{x}_j=x_i$).

§ 2. Canonical Form of a λ-Matrix

1. To begin with, we shall examine what comparatively simple form we can obtain for a rectangular polynomial matrix $A(\lambda)$ by means of left elementary operations only.

Let us assume that the first column of $A(\lambda)$ contains elements not identically equal to zero. Among them we choose a polynomial of least degree and by a permutation of the rows we make it into the element $a_{11}(\lambda)$. Then we divide $a_{i1}(\lambda)$ by $a_{11}(\lambda)$; we denote quotient and remainder by $q_{i1}(\lambda)$ and $r_{i1}(\lambda)$ $(i=2, \ldots, m)$:

[6] Here it is assumed that the unknown functions $x_1, x_2, \ldots, x_n$ are such that their derivatives of all orders, as far as they occur in the transformations, exist. With this restriction, two systems of equations with left-equivalent matrices $A(D)$ and $B(D)$ have the same solutions.

$$a_{i1}(\lambda)=a_{11}(\lambda)q_{i1}(\lambda)+r_{i1}(\lambda) \qquad (i=2, \ldots, m).$$

Now we subtract from the i-th row the first row multiplied by $q_{i1}(\lambda)$ $(i=2, \ldots, m)$. If not all the remainders $r_{i1}(\lambda)$ are identically equal to zero, then we choose one of them that is not equal to zero and is of least degree and put it into the place of $a_{11}(\lambda)$ by a permutation of the rows. As the result of all these operations, the degree of the polynomial $a_{11}(\lambda)$ is reduced.

Now we repeat this process. Since the degree of the polynomial $a_{11}(\lambda)$ is finite, this must come to an end at some stage—i.e., at this stage all the elements $a_{21}(\lambda), a_{31}(\lambda), \ldots, a_{m1}(\lambda)$ turn out to be identically equal to zero.

Next we take the element $a_{22}(\lambda)$ and apply the same procedure to the rows numbered $2, 3, \ldots, m$, achieving $a_{32}(\lambda)=\ldots=a_{m2}(\lambda)=0$. Continuing still further, we finally reduce the matrix $A(\lambda)$ to the following form:

$$\begin{Vmatrix} b_{11}(\lambda) & b_{12}(\lambda) & \ldots & b_{1m}(\lambda) & \ldots & b_{1n}(\lambda) \\ 0 & b_{22}(\lambda) & \ldots & b_{2m}(\lambda) & \ldots & b_{2n}(\lambda) \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & b_{mm}(\lambda) & \ldots & b_{mn}(\lambda) \end{Vmatrix}, \quad (m\leq n)$$

$$\begin{Vmatrix} b_{11}(\lambda) & b_{12}(\lambda) & \ldots & b_{1n}(\lambda) \\ 0 & b_{22}(\lambda) & \ldots & b_{2n}(\lambda) \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & b_{nn}(\lambda) \\ 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 0 \end{Vmatrix}. \quad (m\geq n) \tag{5}$$

If the polynomial $b_{22}(\lambda)$ is not identically equal to zero, then by applying a left elementary operation of the second type we can make the degree of the element $b_{12}(\lambda)$ less than the degree of $b_{22}(\lambda)$ (if $b_{22}(\lambda)$ is of degree zero, then $b_{12}(\lambda)$ becomes identically equal to zero). In the same way, if $b_{33}(\lambda)\equiv 0$, then by left elementary operations of the second type we make the degrees of the elements $b_{13}(\lambda)$, $b_{23}(\lambda)$ less than the degree of $b_{33}(\lambda)$ without changing the elements $b_{12}(\lambda)$, etc.

We have established the following theorem:

THEOREM 1: *An arbitrary rectangular polynomial matrix of dimension $m\times n$ can always be brought into the form (5) by means of left elementary operations, where the polynomials $b_{1k}(\lambda), b_{2k}(\lambda), \ldots, b_{k-1,k}(\lambda)$ are of degree less than that of $b_{kk}(\lambda)$, provided $b_{kk}(\lambda)\not\equiv 0$, and are all identically equal to zero if $b_{kk}(\lambda)=\text{const.}\neq 0$ $(k=2, 3, \ldots, \min(m, n))$.*

Similarly, we prove

THEOREM 2: *An arbitrary rectangular polynomial matrix of dimension $m\times n$ can always be brought into the form*

$$\left\|\begin{matrix} c_{11}(\lambda) & 0 & \dots & 0 & 0 \dots 0 \\ c_{21}(\lambda) & c_{22}(\lambda) & \dots & 0 & 0 \dots 0 \\ \dots & \dots & \dots & \dots & \dots \\ c_{m1}(\lambda) & c_{m2}(\lambda) & \dots & c_{mm}(\lambda) & 0 \dots 0 \end{matrix}\right\|, \qquad \left\|\begin{matrix} c_{11}(\lambda) & 0 & \dots & 0 \\ c_{21}(\lambda) & c_{22}(\lambda) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ c_{n1}(\lambda) & c_{n2}(\lambda) & \dots & c_{nn}(\lambda) \\ \dots & \dots & \dots & \dots \\ c_{m1}(\lambda) & c_{m2}(\lambda) & \dots & c_{mn}(\lambda) \end{matrix}\right\| \tag{6}$$

$(m \leqq n)$ $\qquad$ $(m \geqq n)$

by means of right elementary operations, where the polynomials $c_{k1}(\lambda)$, $c_{k2}(\lambda)$, $\dots$, $c_{k,k-1}(\lambda)$ *are of degree less than that of* $c_{kk}(\lambda)$, *provided* $c_{kk}(\lambda) \not\equiv 0$, *and all are identically equal to zero if* $c_{kk}(\lambda) = \text{const.} \neq 0$ $(k = 2, 3, \dots, \min(m, n))$.

2. From Theorems 1 and 2 we deduce the corollary:

Corollary: *If the determinant of a square polynomial matrix* $P(\lambda)$ *does not depend on* λ *and is different from zero, then the matrix can be represented in the form of a product of a finite number of elementary matrices.*

For by Theorem 1 the matrix $P(\lambda)$ can be brought into the form

$$\left\|\begin{matrix} b_{11}(\lambda) & b_{12}(\lambda) & \dots & b_{1n}(\lambda) \\ 0 & b_{22}(\lambda) & \dots & b_{2n}(\lambda) \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & b_{nn}(\lambda) \end{matrix}\right\|, \tag{7}$$

by left elementary operations, where n is the order of $P(\lambda)$. Since in the application of elementary operations to a square polynomial matrix the determinant of the matrix is only multiplied by constant non-zero factors, the determinant of the matrix (7), like that of $P(\lambda)$, does not depend on λ and is different from 0, i.e.,

$$b_{11}(\lambda)\ b_{22}(\lambda) \cdots b_{nn}(\lambda) = \text{const.} \neq 0.$$

Hence

$$b_{kk}(\lambda) = \text{const.} \neq 0 \quad (k = 1, 2, \dots, n).$$

But then, also by Theorem 1, the matrix (7) has the diagonal form $\| b_k \delta_{ik} \|_1^n$ and can therefore be reduced to the unit matrix E by means of left elementary operations of type 1. But then, conversely, the unit matrix E can be transformed into $P(\lambda)$ by means of the left elementary operations whose matrices are $S_1, S_2, \dots, S_p$. Therefore

$$P(\lambda) = S_p S_{p-1} \cdots S_1 E = S_p S_{p-1} \cdots S_1.$$

As we pointed out on p. 133, from this corollary there follows the equivalence of the two Definitions 2 and 2′ of equivalence of polynomial matrices.

3. Let us return to our example of the system of differential equations (4). We apply Theorem 1 to the matrix $\| a_{ik}(D) \|$ of operator coefficients. As we have shown on p. 135, the system (4) is then replaced by an equivalent system

$$\left.\begin{aligned} b_{11}(D)\, x_1 + b_{12}(D)\, x_2 + \dots + b_{1s}(D)\, x_s &= -b_{1,s+1}(D)\, x_{s+1} - \dots - b_{1n}(D)\, x_n, \\ b_{22}(D)\, x_2 + \dots + b_{2s}(D)\, x_s &= -b_{2,s+1}(D)\, x_{s+1} - \dots - b_{2n}(D)\, x_n, \\ \dots\dots\dots\

If at least one of the elements $a_{ik}(\lambda)$ $(i=2,\ldots,m;\ k=2,\ldots,n)$ is not divisible without remainder by $a_{11}(\lambda)$, then by adding to the first column that column which contains such an element we arrive at the preceding case and can therefore again replace the element $a_{11}(\lambda)$ by a polynomial of smaller degree.

Since the original element $a_{11}(\lambda)$ had a definite degree and since the process of reducing this degree cannot be continued indefinitely, we must, after a finite number of elementary operations, obtain a matrix of the form

$$\begin{Vmatrix} a_1(\lambda) & 0 & \ldots & 0 \\ 0 & b_{22}(\lambda) & \ldots & b_{2n}(\lambda) \\ \ldots & \ldots & \ldots & \ldots \\ 0 & b_{m2}(\lambda) & \ldots & b_{mn}(\lambda) \end{Vmatrix}, \tag{8}$$

in which all the elements $b_{ik}(\lambda)$ are divisible without remainder by $a_1(\lambda)$. If among these elements $b_{ik}(\lambda)$ there is one not identically equal to zero, then continuing the same reduction process on the rows numbered $2,\ldots,m$ and the columns $2,\ldots,n$, we reduce the matrix (8) to the form

$$\begin{Vmatrix} a_1(\lambda) & 0 & 0 & \ldots & 0 \\ 0 & a_2(\lambda) & 0 & \ldots & 0 \\ 0 & 0 & c_{33}(\lambda) & \ldots & c_{3n}(\lambda) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & c_{m3}(\lambda) & \ldots & c_{mn}(\lambda) \end{Vmatrix},$$

where $a_2(\lambda)$ is divisible without remainder by $a_1(\lambda)$ and all the polynomials $c_{ik}(\lambda)$ are divisible without remainder by $a_2(\lambda)$. Continuing the process further, we finally arrive at a matrix of the form

$$\begin{Vmatrix} a_1(\lambda) & 0 & \ldots & 0 & 0 \ldots 0 \\ 0 & a_2(\lambda) & \ldots & 0 & 0 \ldots 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & a_s(\lambda) & 0 \ldots 0 \\ 0 & 0 & \ldots & 0 & 0 \ldots 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 0 & 0 \ldots 0 \end{Vmatrix}, \tag{9}$$

where the polynomials $a_1(\lambda),\ a_2(\lambda),\ \ldots,\ a_s(\lambda)$ $(s \leqq \min(m,n))$ are not identically equal to zero and each is divisible by the preceding one.

By multiplying the first s rows by suitable non-zero numerical factors, we can arrange that the highest coefficients of the polynomials $a_1(\lambda)$, $a_2(\lambda),\ldots,a_s(\lambda)$ are equal to 1.

Definition 3: *A rectangular polynomial matrix is called a canonical diagonal matrix if it is of the form* (9), *where* 1) *the polynomials* $a_1(\lambda)$, $a_2(\lambda),\ldots,a_s(\lambda)$ *are not identically equal to zero and* 2) *each of the polynomials* $a_2(\lambda),\ldots,a_s(\lambda)$ *is divisible by the preceding. Moreover, it is assumed that the highest coefficients of all the polynomials* $a_1(\lambda),\ a_2(\lambda),\ldots,$ $a_s(\lambda)$ *are equal to* 1.

Thus, we have proved that: *An arbitrary rectangular polynomial matrix* $A(\lambda)$ *is equivalent to a canonical diagonal matrix.* In the next section we shall prove that: *The polynomials* $a_1(\lambda),\ a_2(\lambda),\ \ldots,\ a_s(\lambda)$ *are uniquely determined by the given matrix* $A(\lambda)$; and we shall set up formulas that connect these polynomials with the elements of $A(\lambda)$.

§ 3. Invariant Polynomials and Elementary Divisors of a Polynomial Matrix

1. We introduce the concept of invariant polynomials of a λ-matrix $A(\lambda)$.

Let $A(\lambda)$ be a polynomial matrix of rank r, i.e., the matrix has minors of order r not identically equal to zero, but all the minors of order greater than r are identically equal to zero in λ. We denote by $D_j(\lambda)$ the greatest common divisor of all the minors of order j in $A(\lambda)$ $(j=1,2,\ldots,r)$.[7] Then it is easy to see that in the series

$$D_r(\lambda),\quad D_{r-1}(\lambda),\ldots,D_1(\lambda),\quad D_0(\lambda)\equiv 1$$

each polynomial is divisible by the preceding one.[8] The corresponding quotients will be denoted by $i_1(\lambda), i_2(\lambda),\ldots,i_r(\lambda)$:

$$i_1(\lambda)=\frac{D_r(\lambda)}{D_{r-1}(\lambda)},\quad i_2(\lambda)=\frac{D_{r-1}(\lambda)}{D_{r-2}(\lambda)},\ \ldots,\ i_r(\lambda)=\frac{D_1(\lambda)}{D_0(\lambda)}=D_1(\lambda). \tag{10}$$

Definition 4: *The polynomials* $i_1(\lambda), i_2(\lambda),\ldots,i_r(\lambda)$ *defined by* (10) *are called the invariant polynomials of the rectangular matrix* $A(\lambda)$.

The term 'invariant polynomial' is explained by the following arguments. Let $A(\lambda)$ and $B(\lambda)$ be two equivalent polynomial matrices. Then they are obtained from one another by means of elementary operations. But an easy verification shows immediately that the elementary operations

[7] We take the highest coefficient in $D_j(\lambda)$ to be 1 $(j=1,2,\ldots,r)$.

[8] If we apply the Bézout decomposition with respect to the elements of any row to an arbitrary minor of order j, then every term in the decomposition is divisible by $D_{j-1}(\lambda)$; therefore every minor of order j, and hence $D_j(\lambda)$, is divisible by $D_{j-1}(\lambda)$ $(j=2,3,\ldots,r)$.

change neither the rank of $A(\lambda)$ nor the polynomials $D_1(\lambda)$, $D_2(\lambda)$, ..., $D_r(\lambda)$. For when we apply to the identity (3'') the formula that expresses a minor of a product of matrices by the minors of the factors (see p. 12), we obtain for an arbitrary minor of $B(\lambda)$ the expression

$$B\begin{pmatrix} j_1 & j_2 \dots j_p \\ k_1 & k_2 \dots k_p \end{pmatrix}; \lambda\Big) = \sum_{\substack{1 \leqq \alpha_1 < \alpha_2 < \dots < \alpha_p \leqq m \\ 1 \leqq \beta_1 < \beta_2 < \dots < \beta_p \leqq m}} P\begin{pmatrix} j_1 & j_2 \dots j_p \\ \alpha_1 & \alpha_2 \dots \alpha_p \end{pmatrix} A\begin{pmatrix} \alpha_1 & \alpha_2 \dots \alpha_p \\ \beta_1 & \beta_2 \dots \beta_p \end{pmatrix}; \lambda\Big) Q\begin{pmatrix} \beta_1 & \beta_2 \dots \beta_p \\ k_1 & k_2 \dots k_p \end{pmatrix}$$
$$(p = 1, 2, \dots, \min(m, n)).$$

Hence it follows that all the minors of order r or greater of the matrix $B(\lambda)$ are zero, so that we have for the rank r^* of $B(\lambda)$:

$$r^* \leqq r.$$

Moreover, it follows from the same formula that $D_p^*(\lambda)$, the greatest common divisor of all the minors of order p of $B(\lambda)$, is divisible by $D_p(\lambda)$ $(p = 1, 2, \dots, \min(m, n))$. But the matrices $A(\lambda)$ and $B(\lambda)$ can exchange roles. Therefore $r \leqq r^*$ and $D_p(\lambda)$ is divisible by $D_p^*(\lambda)$ $(p = 1, 2, \dots, \min(m, n))$. Hence[9]

$$r = r^*, \quad D_1^*(\lambda) = D_1(\lambda), \quad D_2^*(\lambda) = D_2(\lambda), \dots, D_r^*(\lambda) = D_r(\lambda).$$

Since elementary operations do not change the polynomials $D_1(\lambda)$, $D_2(\lambda)$, ..., $D_r(\lambda)$, they also leave the polynomials $i_1(\lambda)$, $i_2(\lambda)$, ..., $i_r(\lambda)$ defined by (10) unchanged.

Thus, the polynomials $i_1(\lambda)$, $i_2(\lambda)$, ..., $i_r(\lambda)$ remain *invariant* on transition from one matrix to another equivalent one.

If the polynomial matrix has the canonical diagonal form (9), then it is easy to see that for this matrix

$$D_1(\lambda) = a_1(\lambda), \quad D_2(\lambda) = a_1(\lambda)\, a_2(\lambda), \quad \dots, \quad D_r(\lambda) = a_1(\lambda)\, a_2(\lambda) \cdots a_r(\lambda).$$

But then, by (10), the diagonal polynomials in (9) $a_1(\lambda)$, $a_2(\lambda)$, ..., $a_r(\lambda)$ coincide with the invariant polynomials

$$i_1(\lambda) = a_r(\lambda), \; i_2(\lambda) = a_{r-1}(\lambda), \; \dots, \; i_r(\lambda) = a_1(\lambda). \tag{11}$$

Here $i_1(\lambda)$, $i_2(\lambda)$, ..., $i_r(\lambda)$ are at the same time the invariant polynomials of the original matrix $A(\lambda)$, because it is equivalent to (9).

The results obtained can be stated in the form of the following theorem.

[9] The highest coefficients in $D_p(\lambda)$ and $D_p^*(\lambda)$ $(p = 1, 2, \dots, r)$ are 1.

Theorem 3: *The rectangular polynomial matrix $A(\lambda)$ is always equivalent to a canonical diagonal matrix*

$$\begin{Vmatrix} i_r(\lambda) & 0 & \dots & 0 & 0 \dots 0 \\ 0 & i_{r-1}(\lambda) & \dots & 0 & 0 \dots 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & i_1(\lambda) & 0 \dots 0 \\ 0 & 0 & \dots & 0 & 0 \dots 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 \dots 0 \end{Vmatrix}. \tag{12}$$

Moreover, r must here be the rank of $A(\lambda)$ and $i_1(\lambda)$, $i_2(\lambda)$, ..., $i_r(\lambda)$ the invariant polynomials of $A(\lambda)$ defined by (10).

Corollary 1: *Two rectangular matrices of the same dimension $A(\lambda)$ and $B(\lambda)$ are equivalent if and only if they have the same invariant polynomials.*

The sufficiency of the condition was explained above. The necessity follows from the fact that two polynomial matrices having the same invariant polynomials are equivalent to one and the same canonical diagonal matrix and, therefore, to each other. Thus: *The invariant polynomials form a complete system of invariants of a λ-matrix.*

Corollary 2: *In the sequence of invariant polynomials*

$$i_1(\lambda) = \frac{D_r(\lambda)}{D_{r-1}(\lambda)}, \quad i_2(\lambda) = \frac{D_{r-1}(\lambda)}{D_{r-2}(\lambda)}, \quad \dots, \quad i_r(\lambda) = \frac{D_1(\lambda)}{D_0(\lambda)} \qquad (D_0(\lambda) \equiv 1) \tag{13}$$

every polynomial from the second onwards divides the preceding one.

This statement does not follow immediately from (13). It does follow from the fact that the polynomials $i_1(\lambda)$, $i_2(\lambda)$, ..., $i_r(\lambda)$ coincide with the polynomials $a_r(\lambda)$, $a_{r-1}(\lambda)$, ..., $a_1(\lambda)$ of the canonical diagonal matrix (9).

2. We now indicate a method of computing the invariant polynomials of a quasi-diagonal λ-matrix if the invariant polynomials of the matrices in the diagonal blocks are known.

Theorem 4: *If in a quasi-diagonal rectangular matrix*

$$C(\lambda) = \begin{Vmatrix} A(\lambda) & O \\ O & B(\lambda) \end{Vmatrix}$$

every invariant polynomial of $A(\lambda)$ divides every invariant polynomial of $B(\lambda)$, then the set of invariant polynomials of $C(\lambda)$ is the union of the invariant polynomials of $A(\lambda)$ and $B(\lambda)$.

Proof. We denote by $i'_1(\lambda), i'_2(\lambda), \ldots, i'_r(\lambda)$ and $i''_1(\lambda), i''_2(\lambda), \ldots, i''_q(\lambda)$, respectively, the invariant polynomials of the λ-matrices $A(\lambda)$ and $B(\lambda)$. Then[10]

$$A(\lambda) \sim \{i'_r(\lambda), \ldots, i'_1(\lambda), 0, \ldots, 0\}, \quad B(\lambda) \sim \{i''_q(\lambda), \ldots, i''_1(\lambda), 0, \ldots, 0\}$$

and therefore

$$C(\lambda) \sim \{i'_r(\lambda), \ldots, i'_1(\lambda), \quad i''_q(\lambda), \ldots, i''_1(\lambda), 0, \ldots, 0\}. \tag{14}$$

The λ-matrix on the right-hand side of this relation is of canonical diagonal form. By Theorem 3 the diagonal elements of this matrix that are not identically equal to zero then form a complete system of invariants of the polynomial matrix $C(\lambda)$. This proves the theorem.

In order to determine the invariant polynomials of $C(\lambda)$ in the general case of arbitrary invariant polynomials of $A(\lambda)$ and $B(\lambda)$ we make use of the important concept of elementary divisors.

We decompose the invariant polynomials $i_1(\lambda), i_2(\lambda), \ldots, i_r(\lambda)$ into irreducible factors over the given number field F:[11]

$$\begin{aligned} i_1(\lambda) &= [\varphi_1(\lambda)]^{c_1} [\varphi_2(\lambda)]^{c_2} \cdots [\varphi_s(\lambda)]^{c_s}, \\ i_2(\lambda) &= [\varphi_1(\lambda)]^{d_1} [\varphi_2(\lambda)]^{d_2} \cdots [\varphi_s(\lambda)]^{d_s}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ i_r(\lambda) &= [\varphi_1(\lambda)]^{l_1} [\varphi_2(\lambda)]^{l_2} \cdots [\varphi_s(\lambda)]^{l_s}. \end{aligned} \qquad \begin{pmatrix} c_k \geqq d_k \geqq \cdots \geqq l_k \geqq 0; \\ k = 1, 2, \ldots, s \end{pmatrix} \tag{15}$$

Here $\varphi_1(\lambda), \varphi_2(\lambda), \ldots, \varphi_s(\lambda)$ are all the distinct factors irreducible over F (and with highest coefficient 1) that occur in $i_1(\lambda), i_2(\lambda), \ldots, i_r(\lambda)$.

DEFINITION 5: *All the powers among* $[\varphi_1(\lambda)]^{c_1}, \ldots, [\varphi_s(\lambda)]^{l_s}$ *in* (15), *as far as they are distinct from* 1, *are called the elementary divisors of the matrix* $A(\lambda)$ *in the field* F.[12]

THEOREM 5: *The set of elementary divisors of the rectangular quasi-diagonal matrix*

$$C(\lambda) = \left\| \begin{matrix} A(\lambda) & O \\ O & B(\lambda) \end{matrix} \right\|$$

is always obtained by combining the elementary divisors of $A(\lambda)$ *with those of* $B(\lambda)$.

[10] The symbol $\sim$ denotes here the equivalence of matrices; and braces { }, a diagonal rectangular matrix of the form (12).

[11] Some of the exponents $c_k, d_k, \ldots, l_k$ $(k = 1, 2, \ldots, s)$ may be equal to zero.

[12] The formulas (15) enable us to define not only the elementary divisors of $A(\lambda)$ in the field F in terms of the invariant polynomials but also, conversely, the invariant polynomials in terms of the elementary divisors.

Proof. We decompose the invariant polynomials of $A(\lambda)$ and $B(\lambda)$ into irreducible factors over F:[13]

$$\begin{aligned} i'_1(\lambda) &= [\varphi_1(\lambda)]^{c'_1} [\varphi_2(\lambda)]^{c'_2} \cdots [\varphi_s(\lambda)]^{c'_s}, \quad i''_1(\lambda) = [\varphi_1(\lambda)]^{c''_1} [\varphi_2(\lambda)]^{c''_2} \cdots [\varphi_s(\lambda)]^{c''_s}, \\ i'_2(\lambda) &= [\varphi_1(\lambda)]^{d'_1} [\varphi_2(\lambda)]^{d'_2} \cdots [\varphi_s(\lambda)]^{d'_s}, \quad i''_2(\lambda) = [\varphi_1(\lambda)]^{d''_1} [\varphi_2(\lambda)]^{d''_2} \cdots [\varphi_s(\lambda)]^{d''_s}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ i'_r(\lambda) &= [\varphi_1(\lambda)]^{h'_1} [\varphi_2(\lambda)]^{h'_2} \cdots [\varphi_s(\lambda)]^{h'_s}, \quad i''_q(\lambda) = [\varphi_1(\lambda)]^{g''_1} [\varphi_2(\lambda)]^{g''_2} \cdots [\varphi_s(\lambda)]^{g''_s}. \end{aligned}$$

We denote by

$$c_1 \geqq d_1 \geqq \cdots \geqq l_1 > 0, \tag{16}$$

all the non-zero numbers among $c'_1, d'_1, \ldots, h'_1, c''_1, d''_1, \ldots, g''_1$.

Then the matrix $C(\lambda)$ is equivalent to the matrix (14), and by a permutation of rows and of columns the latter can be brought into 'diagonal' form

$$\{[\varphi_1(\lambda)]^{c_1} \cdot (*), [\varphi_1(\lambda)]^{d_1} \cdot (*), \ldots, [\varphi_1(\lambda)]^{l_1} \cdot (*), (**), \ldots, (**)\} \tag{17}$$

where we have denoted by $(*)$ polynomials that are prime to $\varphi_1(\lambda)$ and by $(**)$ polynomials that are either prime to $\varphi_1(\lambda)$ or identically equal to zero. From the form of the matrix (17) we deduce immediately the following decomposition of the polynomials $D_r(\lambda)$, $D_{r-1}(\lambda), \ldots$ and $i_1(\lambda), i_2(\lambda), \ldots$ of the matrix $C(\lambda)$:

$$D_r(\lambda) = [\varphi_1(\lambda)]^{c_1+d_1+\cdots+l_1} \cdot (*), \quad D_{r-1}(\lambda) = [\varphi_1(\lambda)]^{d_1+\cdots+l_1} \cdot (*), \ldots,$$
$$i_1(\lambda) = [\varphi_1(\lambda)]^{c_1} (*), \quad i_2(\lambda) = [\varphi_1(\lambda)]^{d_1} (*), \ldots.$$

Hence it follows that $[\varphi_1(\lambda)]^{c_1}$, $[\varphi_1(\lambda)]^{d_1}, \ldots, [\varphi_1(\lambda)]^{l_1}$, i.e., all the powers

$$[\varphi_1(\lambda)]^{c'_1}, \ldots, [\varphi_1(\lambda)]^{h'_1}, [\varphi_1(\lambda)]^{c''_1}, \ldots, [\varphi_1(\lambda)]^{g''_1},$$

as far as they are distinct from 1, are elementary divisors of $C(\lambda)$.

The elementary divisors of $C(\lambda)$ that are powers of $\varphi_2(\lambda)$ are determined similarly, etc. This completes the proof of the theorem.

Note. The theory of equivalence for integral matrices (i.e., matrices whose elements are integers) can be constructed along similar lines. Here in 1., 2. (see pp. 130-31) $c = \pm 1$, $b(\lambda)$ is to be replaced by an integer, and in (3), (3'), (3''), in place of $P(\lambda)$ and $Q(\lambda)$ there are integral matrices with determinants equal to ± 1.

[13] If any irreducible polynomial $\varphi_k(\lambda)$ occurs as a factor in some invariant polynomials, but not in others, then in the latter we write $\varphi_k(\lambda)$ with a zero exponent.

3. Suppose now that $A = \| a_{ik} \|_1^n$ is a matrix with elements in the field F. We form its characteristic matrix

$$\lambda E - A = \begin{Vmatrix} \lambda - a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \dots & -a_{2n} \\ \dots & \dots & \dots & \dots \\ -a_{n1} & -a_{n2} & \dots & \lambda - a_{nn} \end{Vmatrix}. \tag{18}$$

The characteristic matrix is a λ-matrix of rank n. Its invariant polynomials

$$i_1(\lambda) = \frac{D_n(\lambda)}{D_{n-1}(\lambda)}, \quad i_2(\lambda) = \frac{D_{n-1}(\lambda)}{D_{n-2}(\lambda)}, \ \dots, \ i_n(\lambda) = \frac{D_1(\lambda)}{D_0(\lambda)} \ (D_0(\lambda) \equiv 1), \tag{19}$$

are called the *invariant polynomials of the matrix* A and the corresponding elementary divisors in F are called the *elementary divisors of the matrix* A *in the field* F. A knowledge of the invariant polynomials (and, hence, of the elementary divisors) of A enables us to investigate its structure. Therefore practical methods of computing the invariant polynomials of a matrix are of interest. The formulas (19) give an algorithm for computing these polynomials, but for large n this algorithm is very cumbrous.

Theorem 3 gives another method of computing invariant polynomials, based on the reduction of the characteristic matrix (18) to canonical diagonal form by means of elementary operations.

Example:

$$A = \begin{Vmatrix} 3 & 1 & 0 & 0 \\ -4 & -1 & 0 & 0 \\ 6 & 1 & 2 & 1 \\ -14 & -5 & -1 & 0 \end{Vmatrix}, \qquad \lambda E - A = \begin{Vmatrix} \lambda - 3 & -1 & 0 & 0 \\ 4 & \lambda + 1 & 0 & 0 \\ -6 & -1 & \lambda - 2 & -1 \\ 14 & 5 & 1 & \lambda \end{Vmatrix}.$$

In the characteristic matrix $\lambda E - A$ we add to the fourth row the third multiplied by λ:

$$\begin{Vmatrix} \lambda - 3 & -1 & 0 & 0 \\ 4 & \lambda + 1 & 0 & 0 \\ -6 & -1 & \lambda - 2 & -1 \\ 14 - 6\lambda & 5 - \lambda & \lambda^2 - 2\lambda + 1 & 0 \end{Vmatrix}.$$

Now adding to the first three columns the fourth, multiplied by -6, -1, and $\lambda - 2$, respectively, we obtain

$$\begin{Vmatrix} \lambda - 3 & -1 & 0 & 0 \\ 4 & \lambda + 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 14 - 6\lambda & 5 - \lambda & \lambda^2 - 2\lambda + 1 & 0 \end{Vmatrix}.$$

We add to the first column the second multiplied by $\lambda - 3$:

$$\begin{Vmatrix} 0 & -1 & 0 & 0 \\ \lambda^2 - 2\lambda + 1 & \lambda + 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ -\lambda^2 + 2\lambda - 1 & 5 - \lambda & \lambda^2 - 2\lambda + 1 & 0 \end{Vmatrix}.$$

To the second and fourth rows we add the first multiplied by $\lambda + 1$ and $5 - \lambda$, respectively; we obtain

$$\begin{Vmatrix} 0 & -1 & 0 & 0 \\ \lambda^2 - 2\lambda + 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ -\lambda^2 + 2\lambda - 1 & 0 & \lambda^2 - 2\lambda + 1 & 0 \end{Vmatrix}.$$

To the second row we add the fourth; then we multiply the first and third rows by -1. After permuting some rows and columns we obtain:

$$\begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & (\lambda - 1)^2 & 0 \\ 0 & 0 & 0 & (\lambda - 1)^2 \end{Vmatrix}.$$

The matrix has two elementary divisors $(\lambda - 1)^2$ and $(\lambda - 1)^2$.

§ 4. Equivalence of Linear Binomials

1. In the preceding sections we have considered rectangular λ-matrices. In the present section we consider two square λ-matrices $A(\lambda)$ and $B(\lambda)$ of order n in which all the elements are of degree not higher than 1 in λ. These polynomial matrices may be represented in the form of matrix binomials:

$$A(\lambda) = A_0\lambda + A_1, \ B(\lambda) = B_0\lambda + B_1.$$

We shall assume that these binomials are of degree 1 and regular, i.e., that $|A_0| \neq 0$, $|B_0| \neq 0$ (see p. 76).

The following theorem gives a criterion for the equivalence of such binomials:

THEOREM 6: *If two regular binomials of the first degree* $A_0\lambda + A_1$ and $B_0\lambda + B_1$ *are equivalent, then they are strictly equivalent, i.e., in the identity*

$$B_0\lambda + B_1 = P(\lambda)(A_0\lambda + A_1)Q(\lambda) \tag{20}$$

the matrices $P(\lambda)$ *and* $Q(\lambda)$*—with constant non-zero determinants—can be replaced by constant non-singular matrices* P *and* Q*:*[14]

$$B_0\lambda + B_1 = P(A_0\lambda + A_1)Q. \tag{21}$$

[14] The identity (21) is equivalent to the two matrix equations: $B_0 = PA_0Q$ and $B_1 = PA_1Q$.

Proof. Since the determinant of $P(\lambda)$ does not depend on λ and is different from zero,[15] the inverse matrix $M(\lambda) = P^{-1}(\lambda)$ is also a polynomial matrix. With the help of this matrix we write (20) in the form

$$M(\lambda)(B_0\lambda + B_1) = (A_0\lambda + A_1)Q(\lambda). \tag{22}$$

Regarding $M(\lambda)$ and $Q(\lambda)$ as matrix polynomials, we divide $M(\lambda)$ on the left by $A_0\lambda + A_1$ and $Q(\lambda)$ on the right by $B_0\lambda + B_1$:

$$M(\lambda) = (A_0\lambda + A_1)S(\lambda) + M, \tag{23}$$

$$Q(\lambda) = T(\lambda)(B_0\lambda + B_1) + Q; \tag{24}$$

here M and Q are constant square matrices (independent of λ) of order n. We substitute these expressions for $M(\lambda)$ and $Q(\lambda)$ in (22). After a few small transformations, we obtain

$$(A_0\lambda + A_1)[T(\lambda) - S(\lambda)](B_0\lambda + B_1) = M(B_0\lambda + B_1) - (A_0\lambda + A_1)Q. \tag{25}$$

The difference in the brackets must be identically equal to zero; for otherwise the product on the left-hand side of (25) would be of degree $\geqq 2$, while the polynomial on the right-hand side of the equation is of degree not higher than 1. Therefore

$$S(\lambda) = T(\lambda); \tag{26}$$

But then we obtain from (25):

$$M(B_0\lambda + B_1) = (A_0\lambda + A_1)Q. \tag{27}$$

We shall now show that M is a non-singular matrix. For this purpose we divide $P(\lambda)$ on the left by $B_0\lambda + B_1$:

$$P(\lambda) = (B_0\lambda + B_1)U(\lambda) + P. \tag{28}$$

From (22), (23), and (28) we deduce:

$$\begin{aligned} E = M(\lambda)P(\lambda) &= M(\lambda)(B_0\lambda + B_1)U(\lambda) + M(\lambda)P \\ &= (A_0\lambda + A_1)Q(\lambda)U(\lambda) + (A_0\lambda + A_1)S(\lambda)P + MP \\ &= (A_0\lambda + A_1)[Q(\lambda)U(\lambda) + S(\lambda)P] + MP. \end{aligned} \tag{29}$$

[15] The equivalence of the binomials $A_0\lambda + A_1$ and $B_0\lambda + B_1$ means that an identity (20) exists in which $|P(\lambda)| = \text{const.} \neq 0$ and $|Q(\lambda)| = \text{const.} \neq 0$. However, in this case the last relations follow from (20) itself. For the determinants of regular binomials of the first degree are of degree n:

$$|A_0\lambda + A_1| = |A_0|\lambda^n + \ldots, |B_0\lambda + B_1| = |B_0|\lambda^n + \ldots; |A_0| \neq 0, |B_0| \neq 0.$$

Therefore it follows from

$$|B_0\lambda + B_1| = |P(\lambda)|\,|A_0\lambda + A_1|\,|Q(\lambda)|$$

that

$$|P(\lambda)| = \text{const.} \neq 0, \quad |Q(\lambda)| = \text{const.} \neq 0.$$

Since the last term of this chain of equations must be of degree zero in λ (because it is equal to E), the expression in brackets must be identically equal to zero. But then from (29)

$$MP = E, \tag{30}$$

so that $|M| \neq 0$ and $M^{-1} = P$.

Multiplying both sides of (27) on the left by P, we obtain:

$$B_0\lambda + B_1 = P(A_0\lambda + A_1)Q.$$

The fact that P is non-singular follows from (30). That P and Q are non-singular also follows directly from (21), since this identity implies

$$B_0 = PA_0Q$$

and therefore

$$|P|\,|A_0|\,|Q| = |B_0| \neq 0.$$

This completes the proof of the theorem.

Note. From the proof it follows (see (24) and (28)) that the constant matrices P and Q by which we have replaced the λ-matrices $P(\lambda)$ and $Q(\lambda)$ in (20) can be taken as the left and right remainders, respectively, of $P(\lambda)$ and $Q(\lambda)$ on division by $B_0\lambda + B_1$.

§ 5. A Criterion for Similarity of Matrices

1. Let $A = \| a_{ik} \|_1^n$ be a matrix with numerical elements from the field F. Its characteristic matrix $\lambda E - A$ is a λ-matrix of rank n and therefore has n invariant polynomials (see § 3)

$$i_1(\lambda), i_2(\lambda), \ldots, i_n(\lambda).$$

The following theorem shows that these invariant polynomials determine the original matrix A to within similarity transformations.

Theorem 7: *Two matrices $A = \| a_{ik} \|_1^n$ and $B = \| b_{ik} \|_1^n$ are similar ($B = T^{-1}AT$) if and only if they have the same invariant polynomials or, what is the same, the same elementary divisors in the field* F.

Proof. The condition is *necessary.* For if the matrices A and B are similar, then there exists a non-singular matrix T such that

$$B = T^{-1}AT.$$

Hence

$$\lambda E - B = T^{-1}(\lambda E - A)T.$$

This equation shows that the characteristic matrices $\lambda E - A$ and $\lambda E - B$ are equivalent and therefore have the same invariant polynomials.

The condition is *sufficient*. Suppose that the characteristic matrices $\lambda E - A$ and $\lambda E - B$ have the same invariant polynomials. Then these λ-matrices are equivalent (see Corollary 1 to Theorem 3) and there exist, in consequence, two polynomial matrices $P(\lambda)$ and $Q(\lambda)$ such that

$$\lambda E - B = P(\lambda)(\lambda E - A)Q(\lambda). \tag{31}$$

Applying Theorem 6 to the matrix binomials $\lambda E - A$ and $\lambda E - B$, we may replace in (31) the λ-matrices $P(\lambda)$ and $Q(\lambda)$ by constant matrices P and Q:

$$\lambda E - B = P(\lambda E - A)Q; \tag{32}$$

moreover, P and Q may be taken (see the Note on p. 147) as the left remainder and the right remainder, respectively, of $P(\lambda)$ and $Q(\lambda)$ on division by $\lambda E - B$, i.e., by the Generalized Bézout Theorem, we may set:[16]

$$P = \widehat{P}(B),\ Q = Q(B) \tag{33}$$

Equating coefficients of the powers of λ on both sides of (32), we obtain:

$$B = PAQ, \quad E = PQ,$$

i.e.,

$$B = T^{-1}AT,$$

where

$$T = Q = P^{-1}$$

This proves the theorem.

2. *Note.* We have incidentally established the following result, which we state separately:

Supplement to Theorem 7. *If* $A = \| a_{ik} \|_1^n$ *and* $B = \| b_{ik} \|_1^n$ *are two similar matrices,*

$$B = T^{-1}AT, \tag{34}$$

then we can choose as the transforming matrix T *the matrix*

$$T = Q(B) = [\widehat{P}(B)]^{-1}, \tag{35}$$

where $P(\lambda)$ *and* $Q(\lambda)$ *are polynomial matrices in the identity*

$$\lambda E - B = P(\lambda)(\lambda E - A)Q(\lambda)$$

which connects the equivalent characteristic matrices $\lambda E - A$ *and* $\lambda E - B$*; in* (35) $Q(B)$ *denotes the right value of the matrix polynomial* $Q(\lambda)$*, and* $\widehat{P}(B)$ *the left value of* $P(\lambda)$*, when the argument is replaced by* B*.*

[16] We recall that $\widehat{P}(B)$ is the left value of the polynomial $P(\lambda)$ and $Q(B)$ the right value of $Q(\lambda)$, when λ is replaced by B (see p. 81).

§ 6. The Normal Forms of a Matrix

1. Let

$$g(\lambda) = \lambda^m + \alpha_1\lambda^{m-1} + \cdots + \alpha_{m-1}\lambda + \alpha_m$$

be a polynomial with coefficients in F.

We consider the square matrix of order m

$$L = \begin{Vmatrix} 0 & 0 & \ldots & 0 & -\alpha_m \\ 1 & 0 & \ldots & 0 & -\alpha_{m-1} \\ 0 & 1 & \ldots & 0 & -\alpha_{m-2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 & -\alpha_1 \end{Vmatrix}. \tag{36}$$

It is not difficult to verify that $g(\lambda)$ is the characteristic polynomial of L:

$$|\lambda E - L| = \begin{vmatrix} \lambda & 0 & 0 & \ldots & 0 & \alpha_m \\ -1 & \lambda & 0 & \ldots & 0 & \alpha_{m-1} \\ 0 & -1 & \lambda & \ldots & 0 & \alpha_{m-2} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & -1 & \alpha_1 + \lambda \end{vmatrix} = g(\lambda).$$

On the other hand, the minor of the element α_m in the characteristic determinant is equal to ± 1. Therefore $D_{m-1}(\lambda) = 1$ and $i_1(\lambda) = \frac{D_m(\lambda)}{D_{m-1}(\lambda)} = D_m(\lambda) = g(\lambda)$, $i_2(\lambda) = \ldots = i_m(\lambda) = 1$.

Thus, L has a single invariant polynomial different from 1, namely $g(\lambda)$.

We shall call L the *companion matrix* of the polynomial $g(\lambda)$.

Let $A = \| a_{ik} \|_1^n$ be a matrix with the invariant polynomials

$$i_1(\lambda), i_2(\lambda), \ldots, i_t(\lambda), i_{t+1}(\lambda) = 1, \ldots, i_n(\lambda) = 1. \tag{37}$$

Here the polynomials $i_1(\lambda), i_2(\lambda), \ldots, i_t(\lambda)$ have positive degrees and, from the second onwards, each divides the preceding one. We denote the companion matrices of these polynomials by $L_1, L_2, \ldots, L_t$.

Then the quasi-diagonal matrix of order n

$$L_I = \{L_1, L_2, \ldots, L_t\} \tag{38}$$

has the polynomials (37) as its invariant polynomials (see Theorem 4 on p. 141). Since the matrices A and L_I have the same invariant polynomials, they are similar, i.e., there always exists a non-singular matrix U ($|U| \neq 0$) such that

$$A = UL_{\mathrm{I}}U^{-1}. \tag{I}$$

The matrix L_{I} is called the *first natural normal form* of the matrix A. This normal form is characterized by: 1) the quasi-diagonal form (38), 2) the special structure of the diagonal blocks (36), and 3) the additional condition: in the sequence of characteristic polynomials of the diagonal blocks every polynomial from the second onwards divides the preceding one.[17]

2. We now denote by

$$\chi_1(\lambda), \chi_2(\lambda), \ldots, \chi_u(\lambda) \tag{39}$$

the elementary divisors of $A = \| a_{ik} \|_1^n$ in the number field F. The corresponding companion matrices will be denoted by

$$L^{(1)}, L^{(2)}, \ldots, L^{(u)}.$$

Since $\chi_j(\lambda)$ is the only elementary divisor of $L^{(j)}$ $(j = 1, 2, \ldots, u)$,[18] the quasi-diagonal matrix

$$L_{\mathrm{II}} = \{L^{(1)}, L^{(2)}, \ldots, L^{(u)}\} \tag{40}$$

has, by Theorem 5, the polynomials (39) as its elementary divisors.

The matrices A and L_{II} have the same elementary divisors in F. Therefore the matrices are similar, i.e., there always exists a non-singular matrix V $(|V| \neq 0)$ such that

$$A = VL_{\mathrm{II}}V^{-1}. \tag{II}$$

The matrix L_{II} is called the *second natural normal form* of the matrix A. This normal form is characterized by: 1) the quasi-diagonal form (40), 2) the special structure of the diagonal blocks (36), and 3) the additional condition: the characteristic polynomial of each diagonal block is a power of an irreducible polynomial over F.

Note. The elementary divisors of a matrix A, in contrast to the invariant polynomials, are essentially connected with the given number field F. If we choose instead of the original field F another number field (which also contains the elements of the given matrix A), then the elementary divisors may change. Together with the elementary divisors, the second natural normal form of a matrix also changes.

[17] From the conditions 1), 2), 3) it follows automatically that the characteristic polynomials of the diagonal blocks in L_{I} are the invariant polynomials of the matrix L_{I} and, hence, of A.

[18] $\chi_j(\lambda)$ is the only invariant polynomial of $L^{(j)}$ and is at the same time a power of a polynomial irreducible over F.

Suppose, for example, that $A = \| a_{ik} \|_1^n$ is a matrix with real elements. The characteristic polynomial of the matrix then has real coefficients. But this polynomial may have complex roots. If F is the field of real numbers, then among the elementary divisors there may also be powers of irreducible quadratic trinomials with real coefficients. If F is the field of complex numbers, then every elementary divisor has the form $(\lambda - \lambda_0)^p$.

3. Let us assume now that the number field F contains not only the elements of A, but also the characteristic values of the matrix.[19] Then the elementary divisors of A have the form[20]

$$(\lambda - \lambda_1)^{p_1}, (\lambda - \lambda_2)^{p_2}, \ldots, (\lambda - \lambda_u)^{p_u} \qquad (p_1 + p_2 + \cdots + p_u = n). \tag{41}$$

We consider one of these elementary divisors:

$$(\lambda - \lambda_0)^p$$

and associate with it the following matrix of order p:

$$\begin{Vmatrix} \lambda_0 & 1 & 0 & \ldots & 0 \\ 0 & \lambda_0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 \\ 0 & 0 & 0 & \ldots & \lambda_0 \end{Vmatrix} = \lambda_0 E^{(p)} + H^{(p)}. \tag{42}$$

It is easy to verify that this matrix has only the one elementary divisor $(\lambda - \lambda_0)^p$. The matrix (42) will be called the *Jordan block* corresponding to the elementary divisor $(\lambda - \lambda_0)^p$.

The Jordan blocks corresponding to the elementary divisors (41) will be denoted by

$$J_1, J_2, \ldots, J_u.$$

Then the quasi-diagonal matrix

$$J = \{J_1, J_2, \ldots, J_u\}$$

has the powers (41) as its elementary divisors.

The matrix J can also be written in the form

$$J = \{\lambda_1 E_1 + H_1, \lambda_2 E_2 + H_2, \ldots, \lambda_u E_u + H_u\};$$

where

$$E_k = E^{(p_k)}, H_k = H^{(p_k)} \qquad (k = 1, 2, \ldots, u).$$

[19] This always holds for an arbitrary matrix A if F is the field of complex numbers.

[20] Among the numbers $\lambda_1, \lambda_2, \ldots, \lambda_u$ there may be some that are equal.

Since the matrices A and J have the same elementary divisors, they are similar, i.e., there exists a non-singular matrix $T(|T| \neq 0)$ such that

$$A = TJT^{-1} = T\{\lambda_1 E_1 + H_1,\ \lambda_2 E_2 + H_2,\ \ldots,\ \lambda_u E_u + H_u\}T^{-1}. \quad \text{(III)}$$

The matrix J is called the *Jordan normal form* or simply *Jordan form* of A. The Jordan normal form is characterized by its quasi-diagonal form and by the special structure (42) of the diagonal blocks.

The following scheme describes the Jordan matrix J for the elementary divisors $(\lambda-\lambda_1)^2$, $(\lambda-\lambda_2)^3$, $\lambda-\lambda_3$, $(\lambda-\lambda_4)^2$:

$$J = \left\|\begin{array}{cccccccc} \lambda_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_4 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_4 \end{array}\right\|. \quad (43)$$

If (and only if) all the elementary divisors of a matrix A are of the first degree, the Jordan form is a diagonal matrix, and in this case we have:

$$A = T\{\lambda_1, \lambda_2, \ldots, \lambda_n\}T^{-1}. \quad (44)$$

Thus: *A matrix A has simple structure* (see Chapter III, § 8) *if and only if all its elementary divisors are of the first degree.*[21]

Instead of the Jordan block (42) sometimes the '*lower*' *Jordan block* of order p is used:

$$\left\|\begin{array}{ccccccc} \lambda_0 & 0 & . & . & . & 0 & 0 \\ 1 & \lambda_0 & . & . & . & 0 & 0 \\ 0 & . & . & & & . & . \\ . & . & . & . & & . & . \\ . & & & . & . & . & . \\ . & & & & . & \lambda_0 & 0 \\ 0 & . & . & . & 0 & 1 & \lambda_0 \end{array}\right\| = \lambda_0 E^{(p)} + F^{(p)}.$$

This matrix also has the single elementary divisor $(\lambda-\lambda_0)^p$ only. To the elementary divisors (41) there corresponds the *lower Jordan matrix.*[22]

[21] The elementary divisors of degree 1 are often called 'linear' or 'simple' elementary divisors.

[22] The matrix J is often called the *upper Jordan matrix*, in contrast to the lower Jordan matrix $J_{(1)}$.

$$J_{(1)} = \{\lambda_1 E_1 + F_1,\ \lambda_2 E_2 + F_2,\ \ldots,\ \lambda_u E_u + F_u\}$$
$$(E_k = E^{(p_k)},\ F_k = F^{(p_k)};\ k = 1, 2, \ldots, u).$$

An arbitrary matrix A having the elementary divisors (41) is always similar to $J_{(1)}$, i.e., there exists a non-singular matrix T_1 $(|T_1| \neq 0)$ such that

$$A = T_1 J_{(1)} T_1^{-1} = T_1\{\lambda_1 E_1 + F_1,\ \lambda_2 E_2 + F_2,\ \ldots,\ \lambda_u E_u + F_u\}T_1^{-1}. \quad \text{(IV)}$$

We also note that if $\lambda_0 \neq 0$, each of the two matrices

$$\lambda_0(E^{(p)} + H^{(p)}), \qquad \lambda_0(E^{(p)} + F^{(p)})$$

has only the single elementary divisor $(\lambda-\lambda_0)^p$. Therefore for a *non-singular* matrix A having the elementary divisors (41) we have, apart from (III) and (IV), the representations

$$A = T_2\{\lambda_1(E_1 + H_1), \lambda_2(E_2 + H_2), \ldots, \lambda_u(E_u + H_u)\}T_2^{-1}, \quad \text{(V)}$$
$$A = T_3\{\lambda_1(E_1 + F_1), \lambda_2(E_2 + F_2), \ldots, \lambda_u(E_u + F_u)\}T_3^{-1}. \quad \text{(VI)}$$

§ 7. The Elementary Divisors of the Matrix $f(A)$

1. In this section we consider the following problem:

Given the elementary divisors (in the field of complex numbers) of a matrix $A = \|a_{ik}\|_1^n$ *and given a function* $f(\lambda)$ *defined on the spectrum of* A, *to determine the elementary divisors (in the field of complex numbers) of the matrix* $f(A)$.

The matrix $f(A)$ does not alter if we replace the function $f(\lambda)$ by a polynomial that assumes on the spectrum of A the same values as $f(\lambda)$ (see Chapter V, § 1). Without loss of generality we may therefore assume in what follows that $f(\lambda)$ is a polynomial.

We denote by

$$(\lambda-\lambda_1)^{p_1},\ (\lambda-\lambda_2)^{p_2}, \ldots, (\lambda-\lambda_u)^{p_u}$$

the elementary divisors of A.[23] Thus A is similar to the Jordan matrix

$$A = TJT^{-1},$$

and so

$$f(A) = Tf(J)T^{-1}.$$

[23] Among the numbers $\lambda_1, \lambda_2, \ldots, \lambda_u$ there may be some that are equal.

Moreover,

$$J=\{J_1, J_2, \ldots, J_u\}, \quad J_i=\lambda_i E^{(p_i)}+H^{(p_i)} \quad (i=1, 2, \ldots, u)$$

and

$$f(J)=\{f(J_1), f(J_2), \ldots, f(J_u)\}, \tag{45}$$

where (see Example 2 on p. 100)

$$f(J_i)=\begin{Vmatrix} f(\lambda_i) & \frac{f'(\lambda_i)}{1!} & \cdots & \frac{f^{(p_i-1)}(\lambda_i)}{(p_i-1)!} \\ 0 & f(\lambda_i) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{f'(\lambda_i)}{1!} \\ 0 & 0 & \cdots & f(\lambda_i) \end{Vmatrix}. \tag{46}$$

Since the similar matrices $f(A)$ and $f(J)$ have the same elementary divisors, we shall from now on consider $f(J)$ instead of $f(A)$.

2. Let us determine the defect[24] d of $f(A)$ or, what is the same, of $f(J)$. The defect of a quasi-diagonal matrix is equal to the sum of the defects of the various diagonal blocks and the defect of $f(J_i)$ (see 46)) is equal to the smaller of the numbers k_i and p_i, where k_i is the multiplicity of λ_i as a root of $f(\lambda)$,[25] so that

$$f(\lambda_i)=f'(\lambda_i)=\cdots=f^{(k_i-1)}(\lambda_i)=0, \ f^{(k_i)}(\lambda_i)\neq 0 \qquad (i=1, 2, \ldots, u).$$

We have thus arrived at the following theorem:

THEOREM 8: *The defect of the matrix $f(A)$, where A has the elementary divisors*

$$(\lambda-\lambda_1)^{p_1}, (\lambda-\lambda_2)^{p_2}, \ldots, (\lambda-\lambda_u)^{p_u} \tag{47}$$

is given by the formula

$$d=\sum_{i=1}^{u}\min(k_i, p_i); \tag{48}$$

[24] $d=n-r$, where r is the rank of $f(A)$. If the elementary divisors of a matrix are known, then the defect of the matrix is determined as the number of elementary divisors corresponding to the characteristic value 0, i.e., as the number of elementary divisors of the form λ^m.

[25] k_i may be equal to zero; in that case $f(\lambda_i)\neq 0$.

here k_i is the multiplicity of λ_i as root of $f(\lambda)$ $(i=1, 2, \ldots, u)$.[26]

As an application of this theorem we shall determine all the elementary divisors of an arbitrary matrix $A=\|a_{ik}\|_1^n$ that corresponds to a characteristic value λ_0:

$$\underbrace{\lambda-\lambda_0, \ldots, \lambda-\lambda_0}_{g_1}; \quad \underbrace{(\lambda-\lambda_0)^2, \ldots, (\lambda-\lambda_0)^2}_{g_2}; \quad \ldots; \quad \underbrace{(\lambda-\lambda_0)^m, \ldots, (\lambda-\lambda_0)^m}_{g_m},$$

where $g_i\geqq 0$ $(i=1, 2, \ldots, m-1)$, $g_m>0$, provided the defects

$$d_1, d_2, \ldots, d_m$$

of the matrices

$$A-\lambda_0 E, (A-\lambda_0 E)^2, \ldots, (A-\lambda_0 E)^m$$

are given.

For this purpose we note that $(A-\lambda_0 E)^j=f_j(A)$, where $f_j(\lambda)=(\lambda-\lambda_0)^j$ $(j=1, 2, \ldots, m)$. In order to determine the defect of $(A-\lambda_0 E)^j$ we have, therefore, to set $k_i=j$ in (48) for the elementary divisors corresponding to the characteristic value λ_0 and $k_i=0$ for all the other terms $(j=1, 2, \ldots, m)$. Thus we obtain the formulas

$$\left.\begin{aligned} g_1+g_2+g_3+\cdots+g_m&=d_1, \\ g_1+2g_2+2g_3+\cdots+2g_m&=d_2, \\ g_1+2g_2+3g_3+\cdots+3g_m&=d_3, \\ &\cdots\cdots\cdots\cdots\cdots \\ g_1+2g_2+3g_3+\cdots+mg_m&=d_m. \end{aligned}\right\} \tag{49}$$

Hence[27]

$$g_j=2d_j-d_{j-1}-d_{j+1} \qquad (j=1, 2, \ldots, m; \ d_0=0, \ d_{m+1}=d_m). \tag{50}$$

3. Let us return to the basic problem of determining the elementary divisors of the matrix $f(A)$. As we have mentioned above, the elementary divisors of $f(A)$ coincide with those of $f(J)$ and the elementary divisors of a quasi-diagonal matrix coincide with those of the diagonal blocks (see Theorem 5). Therefore the problem reduces to finding the elementary divisors of a matrix C of regular triangular form:

[26] In the general case, where $f(\lambda)$ is not a polynomial, then $\min(k_i, p_i)$ in (48) has to be interpreted as the number p_i if

$$f(\lambda_i)=f'(\lambda_i)=\ldots=f^{(p_i)}(\lambda_i)=0$$

and as the number $k_i\leqq p_i$ if

$$f(\lambda_i)=f'(\lambda_i)=\ldots=f^{(k_i-1)}(\lambda_i)=0, \quad f^{(k_i)}(\lambda_i)\neq 0$$

$(i=1, 2, \ldots, u)$.

[27] The number m is characterized by the fact that $d_{m-1}<d_m=d_{m+j}$ $(j=1, 2, \ldots)$.

$$C=\sum_{k=0}^{p-1}a_kH^k=\begin{Vmatrix} a_0 & a_1 & \cdots & a_{p-1} \\ 0 & a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & a_1 \\ 0 & 0 & \cdots & a_0 \end{Vmatrix}. \tag{51}$$

We consider separately two cases:

1. $a_1 \neq 0$. The characteristic polynomial of C is obviously equal to

$$D_p(\lambda)=(\lambda-a_0)^p.$$

Since $D_{p-1}(\lambda)$ divides $D_p(\lambda)$ without remainder, we have

$$D_{p-1}(\lambda)=(\lambda-a_0)^g \quad (g\leqq p).$$

Here $D_{p-1}(\lambda)$ denotes the greatest common divisor of the minors of order $p-1$ in the characteristic matrix

$$\lambda E-C=\begin{Vmatrix} \lambda-a_0 & -a_1 & \cdots & -a_{p-1} \\ 0 & \lambda-a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & -a_1 \\ 0 & 0 & \cdots & \lambda-a_0 \end{Vmatrix}.$$

It is easy to see that when the minor of the zero element marked by '+' is expanded, every term contains at least one factor $\lambda-a_0$, except the product of the elements on the main diagonal, which is $(-a_1)^{p-1}$ and is therefore in our case different from zero. But since $D_{p-1}(\lambda)$ must be a power of $\lambda-a_0$, we see that $g=0$. But then it follows from

$$D_p(\lambda)=(\lambda-a_0)^p, \quad D_{p-1}(\lambda)=1$$

that C has only the one elementary divisor $(\lambda-a_0)^p$.

2. $a_1=\ldots=a_{k-1}=0$, $a_k\neq 0$. In this case,

$$C=a_0E+a_kH^k+\cdots+a_{p-1}H^{p-1}.$$

Therefore for the positive integer j the defect of the matrix

$$(C-a_0E)^j=a_k^jH^{kj}+\cdots$$

is given by

$$d_j=\begin{cases} kj, & \text{when} \quad kj\leqq p, \\ p, & \text{when} \quad kj>p. \end{cases}$$

We set

$$p=qk+h \quad (0\leqq h<k). \tag{52}$$

Then[28]

$$d_1=k,\ d_2=2k,\ \ldots,\ d_q=qk,\ d_{q+1}=p. \tag{53}$$

Therefore we have by (50)

$$g_1=\cdots=g_{q-1}=0, \quad g_q=k-h,\ g_{q+1}=h.$$

Thus, the matrix C has the elementary divisors

$$\underbrace{(\lambda-a_0)^{q+1},\ \ldots,\ (\lambda-a_0)^{q+1}}_{h}, \quad \underbrace{(\lambda-a_0)^q,\ \ldots,\ (\lambda-a_0)^q}_{k-h}, \tag{54}$$

where the integers $q>0$ and $h\geqq 0$ are determined by (52).

4. Now we are in a position to ascertain what elementary divisors the matrix $f(J)$ has (see (45) and (46)). To each elementary divisor of A

$$(\lambda-\lambda_0)^p$$

there corresponds in $f(J)$ the diagonal cell

$$f(\lambda_0E+H)=\sum_{i=0}^{p-1}\frac{f^{(i)}(\lambda_0)}{i!}H^i=\begin{Vmatrix} f(\lambda_0) & \frac{f'(\lambda_0)}{1!} & \cdots & \frac{f^{(p-1)}(\lambda_0)}{(p-1)!} \\ 0 & f(\lambda_0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{f'(\lambda_0)}{1!} \\ 0 & 0 & \cdots & f(\lambda_0) \end{Vmatrix}. \tag{55}$$

Clearly the problem reduces to finding the elementary divisors of a cell of the form (55). But the matrix (55) is of the regular triangular form (51), where

$$a_0=f(\lambda_0),\ a_1=f'(\lambda_0),\ a_2=\frac{f''(\lambda_0)}{2!},\ \ldots.$$

Thus we arrive at the theorem:

[28] In this case the number $q+1$ plays the role of m in (49) and (50).

Theorem 9: *The elementary divisors of the matrix $f(A)$ are obtained from those of A in the following way: To an elementary divisor*

$$(\lambda-\lambda_0)^p \tag{56}$$

of A for $p=1$ or for $p>1$ and $f'(\lambda_0)\neq 0$ there corresponds a single elementary divisor

$$\big(\lambda-f(\lambda_0)\big)^p \tag{57}$$

of $f(A)$; for $p>1$ and $f(\lambda_0)=\ldots=f^{(k-1)}(\lambda_0)=0$, $f^{(k)}(\lambda_0)\neq 0$ $(k<p)$ to the elementary divisor (56) of A there correspond the following elementary divisors of $f(A)$:

$$\underbrace{\big(\lambda-f(\lambda_0)\big)^{q+1},\ldots,\big(\lambda-f(\lambda_0)\big)^{q+1}}_{h},\quad\underbrace{\big(\lambda-f(\lambda_0)\big)^{q},\ldots,\big(\lambda-f(\lambda_0)\big)^{q}}_{k-h}, \tag{58}$$

where

$$p=qk+h,\quad 0\leqq q,\quad 0\leqq h<k;$$

finally, for $p>1$, $f'(\lambda_0)=\ldots=f^{(p-1)}(\lambda_0)=0$, to the elementary divisor (56) there correspond p elementary divisors of the first degree of $f(A)$:[29]

$$\lambda-f(\lambda_0),\ldots,\lambda-f(\lambda_0). \tag{59}$$

We note the following special cases of this theorem.

1. *If $\lambda_1,\lambda_2,\ldots,\lambda_n$ are the characteristic values of A, then $f(\lambda_1)$, $f(\lambda_2)$, $\ldots$, $f(\lambda_n)$ are the characteristic values of $f(A)$.* (In both sequences each characteristic value is repeated as often as its multiplicity as a root of the characteristic equation indicates.)[30]

2. *If the derivative $f'(\lambda)$ is not zero on the spectrum of A,*[31] *then in going from A to $f(A)$ the elementary divisors are not 'split up' i.e., if A has the elementary divisors*

$$(\lambda-\lambda_1)^{p_1},\ (\lambda-\lambda_2)^{p_2},\ \ldots,\ (\lambda-\lambda_n)^{p_n},$$

then $f(A)$ has the elementary divisors

$$\big(\lambda-f(\lambda_1)\big)^{p_1},\ \big(\lambda-f(\lambda_2)\big)^{p_2},\ \ldots,\ \big(\lambda-f(\lambda_n)\big)^{p_n}.$$

[29] (57) is obtained from (58) by setting $k=1$; (59) is obtained from (58) by setting $k=p$ or $k>p$.

[30] Statement 1. was established separately in Chapter IV, p. 84.

[31] I.e., $f'(\lambda_i)\neq 0$ for those λ_i that are multiple roots of the minimal polynomial.

§ 8. A General Method of Constructing the Transforming Matrix

In many problems in the theory of matrices and its applications it is sufficient to know the normal form into which a given matrix $A=\|a_{ik}\|_1^n$ can be carried by similarity transformations. The normal form is completely determined by the invariant polynomials of the characteristic matrix $\lambda E-A$. To find the latter, we can use the defining formulas (see (10) on p. 139) or the reduction of the characteristic matrix $\lambda E-A$ to canonical diagonal form by elementary transformations.

In some problems, however, it is necessary to know not only the normal form $\tilde{A}$ of the given matrix A, but also a non-singular transforming matrix T.

1. An immediate method of determining T consists in the following. The equation

$$A=T\tilde{A}T^{-1}$$

can be written as:

$$AT-T\tilde{A}=0.$$

This matrix equation in T is equivalent to a system of n^2 linear homogeneous equations in the n^2 unknown coefficients of T. The determination of a transforming matrix reduces to the solution of this system of n^2 equations. Moreover, we have to choose from the set of all solutions one for which $|T|\neq 0$. The existence of such a solution is certain, since A and $\tilde{A}$ have the same invariant polynomials.[32]

Note that whereas the normal form is uniquely determined by the matrix A,[33] for the transforming matrix T we always have an innumerable set of values that are given by

$$T=UT_1, \tag{60}$$

where T_1 is one of the transforming matrices and U is an arbitrary matrix that is permutable with A.[34]

[32] From this fact follows the similarity of $\tilde{A}$ and A.

[33] This statement is unconditionally true as regards the first natural normal form. As far as the second normal form or the Jordan normal form is concerned, they are uniquely determined to within the order of the diagonal blocks.

[34] The formula (60) may be replaced by

$$T=T_1V,$$

where V is an arbitrary matrix permutable with $\tilde{A}$.

The method proposed above for determining a transforming matrix T is simple enough in concept but of little use in practice, since it requires a great many computations (even for $n=4$ we have to solve 16 linear equations).

2. We proceed to explain a more efficient method of constructing the transforming matrix T. This method is based on the Supplement to Theorem 7 (p. 148). According to this, we can choose as the transforming matrix

$$T=Q(\tilde{A}), \tag{61}$$

provided

$$\lambda E-\tilde{A}=P(\lambda)(\lambda E-A)Q(\lambda).$$

The latter equation expresses the equivalence of the characteristic matrices $\lambda E-A$ and $\lambda E-\tilde{A}$. Here $P(\lambda)$ and $Q(\lambda)$ are polynomial matrices with constant non-zero determinants.

For the actual process of finding $Q(\lambda)$ we reduce the two λ-matrices $\lambda E-A$ and $\lambda E-\tilde{A}$ to canonical form by means of the corresponding elementary transformations

$$\{i_n(\lambda), i_{n-1}(\lambda), \ldots, i_1(\lambda)\}=P_1(\lambda)(\lambda E-A)Q_1(\lambda) \tag{62}$$

$$\{i_n(\lambda), i_{n-1}(\lambda), \ldots, i_1(\lambda)\}=P_2(\lambda)(\lambda E-\tilde{A})Q_2(\lambda) \tag{63}$$

where

$$Q_1(\lambda)=T_1T_2,\ldots,T_{p_1},\quad Q_2(\lambda)=T_1^*T_2^*\ldots T_{p_2}^*, \tag{64}$$

and where $T_1,\ldots,T_{p_1}, T_1^*,\ldots,T_{p_2}^*$ are the elementary matrices corresponding to the elementary operations on the columns of the λ-matrices $\lambda E-A$ and $\lambda E-\tilde{A}$. From (62), (63), and (64) it follows that

$$\lambda E-\tilde{A}=P(\lambda)(\lambda E-A)Q(\lambda),$$

where

$$Q(\lambda)=Q_1(\lambda)Q_2^{-1}(\lambda)=T_1T_2\cdots T_{p_1}T_{p_2}^{*-1}T_{p_2-1}^{*-1}\cdots T_1^{*-1}. \tag{65}$$

We can compute the matrix $Q(\lambda)$ by applying successively to the columns of the unit matrix E the elementary operations with the matrices $T_1,\ldots,T_{p_1}, T_{p_2}^{*-1},\ldots,T_1^{*-1}$. After this (in accordance with (61)) we replace the argument λ in $Q(\lambda)$ by the matrix $\tilde{A}$.

Example.

$$A=\begin{Vmatrix} 1 & -1 & 1 & -1\\ -3 & 3 & -5 & 4\\ 8 & -4 & 3 & -4\\ 15 & -10 & 11 & -11 \end{Vmatrix}.$$

Let us introduce a symbolic notation for the right elementary operations and the corresponding matrices (see pp. 130-131):

$$T'=[(c)\,i],\quad T''=[i+(b(\lambda))\,j],\quad T'''=[ij].$$

In transforming the characteristic matrix $\lambda E-A$ into normal diagonal form we shall at the same time keep a record of the elementary right operations to be performed, i.e., the operations on the columns:

$$\lambda E-A=\begin{Vmatrix} \lambda-1 & 1 & -1 & 1\\ 3 & \lambda-3 & 5 & -4\\ -8 & 4 & \lambda-3 & 4\\ -15 & 10 & -11 & \lambda+11 \end{Vmatrix},\ \begin{Vmatrix} 0 & 0 & 0 & 1\\ 4\lambda-1 & \lambda+1 & 1 & -4\\ -4\lambda-4 & 0 & \lambda+1 & 4\\ -\lambda^2-10\lambda-4 & -\lambda-1 & \lambda & \lambda+11 \end{Vmatrix},$$

$$\begin{Vmatrix} 1 & 0 & 0 & 0\\ -4 & \lambda+1 & 1 & 4\lambda-1\\ 4 & 0 & \lambda+1 & -4\lambda-4\\ \lambda+11 & -\lambda-1 & \lambda & -\lambda^2-10\lambda-4 \end{Vmatrix};$$

$$\begin{Vmatrix} 1 & 0 & 0 & 0\\ 0 & \lambda+1 & 1 & 4\lambda-1\\ 0 & 0 & \lambda+1 & -4\lambda-4\\ 0 & -\lambda-1 & \lambda & -\lambda^2-10\lambda-4 \end{Vmatrix},\ \begin{Vmatrix} 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & -\lambda^2-2\lambda-1 & \lambda+1 & -4\lambda^2-7\lambda-3\\ 0 & -\lambda^2-2\lambda-1 & \lambda & -5\lambda^2-9\lambda-4 \end{Vmatrix},$$

$$\begin{Vmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & \lambda+1 & -\lambda^2-2\lambda-1 & -4\lambda^2-7\lambda-3\\ 0 & \lambda & -\lambda^2-2\lambda-1 & -5\lambda^2-9\lambda-4 \end{Vmatrix};$$

$$\begin{Vmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & \lambda^2+2\lambda+1 & 4\lambda^2+7\lambda+3\\ 0 & 0 & \lambda^2+2\lambda+1 & 5\lambda^2+9\lambda+4 \end{Vmatrix},\ \begin{Vmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & \lambda^2+2\lambda+1 & -\lambda^2-3\lambda-2\\ 0 & 0 & \lambda^2+2\lambda+1 & -\lambda-1 \end{Vmatrix},$$

$$\begin{Vmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & -\lambda^2-2\lambda-1 & \lambda+1\\ 0 & 0 & \lambda^2+2\lambda+1 & -\lambda^2-3\lambda-2 \end{Vmatrix};$$

$$\begin{Vmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & \lambda+1 & -\lambda^2-2\lambda-1\\ 0 & 0 & -\lambda^2-3\lambda-2 & \lambda^2+2\lambda+1 \end{Vmatrix},\ \begin{Vmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & \lambda+1 & 0\\ 0 & 0 & -\lambda^2-3\lambda-2 & -(\lambda+1)^3 \end{Vmatrix},$$

$$\begin{Vmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & \lambda+1 & 0\\ 0 & 0 & 0 & (\lambda+1)^3 \end{Vmatrix}.$$

Here

$$Q_1(\lambda)=[1+(1-\lambda)\,4]\,[2-4]\,[3+4]\,[14]\,[4-(\lambda+1)\,3]\,[4+(1-4\lambda)\,3]\,[23]\times\\ \times[4-(5)\,3]\,[43]\,[4+(\lambda+1)\,3].$$

We have found the invariant polynomials $(\lambda+1)^3$, $(\lambda+1)$, 1, and 1 of A. The matrix has two elementary divisors, $(\lambda+1)^3$ and $(\lambda+1)$. Therefore the Jordan normal form is

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

By elementary operations we bring the matrix $\lambda E - J$ into normal diagonal form

$$\lambda E-J=\begin{Vmatrix} \lambda+1 & -1 & 0 & 0 \\ 0 & \lambda+1 & -1 & 0 \\ 0 & 0 & \lambda+1 & 0 \\ 0 & 0 & 0 & \lambda+1 \end{Vmatrix},\quad \begin{Vmatrix} \lambda+1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & (\lambda+1)^2 & \lambda+1 & 0 \\ 0 & 0 & 0 & \lambda+1 \end{Vmatrix},$$

$$\begin{Vmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ (\lambda+1)^3 & (\lambda+1)^2 & \lambda+1 & 0 \\ 0 & 0 & 0 & \lambda+1 \end{Vmatrix};$$

$$\begin{Vmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ (\lambda+1)^3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda+1 \end{Vmatrix},\quad \begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & (\lambda+1)^3 \\ 0 & 0 & \lambda+1 & 0 \end{Vmatrix},$$

$$\begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \lambda+1 & 0 \\ 0 & 0 & 0 & (\lambda+1)^3 \end{Vmatrix}.$$

Here

$$Q_2(\lambda)=[2+(\lambda+1)3]\ [1+(\lambda+1)2]\ [12]\ [23]\ [34].$$

Therefore

$$\begin{aligned} Q(\lambda)&=Q_1(\lambda)\,Q_2^{-1}(\lambda)\\ &=[1+(1-\lambda)4]\ [2-4]\ [3+4]\ [14]\ [2-(\lambda+1)3]\ [4+(1-4\lambda)3]\ [23]\ [4-(5)3]\times\\ &\quad\times[43]\ [4+(\lambda+1)3]\ [34]\ [23]\ [12]\ [1-(\lambda+1)2]\ [2-(\lambda+1)3]. \end{aligned}$$

We apply these elementary operations successively to the unit matrix E:

$$E=\begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix},\quad \begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1-\lambda & -1 & 1 & 1 \end{Vmatrix},$$

$$\begin{Vmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & -1 & 1 & 1-\lambda \end{Vmatrix},\quad \begin{Vmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & -\lambda-1 & 1 & 0 \\ 1 & -\lambda-2 & 1 & 1-\lambda \end{Vmatrix};$$

$$\begin{Vmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & -\lambda-1 & 1 & 1-4\lambda \\ 1 & -\lambda-2 & 1 & 2-5\lambda \end{Vmatrix},\quad \begin{Vmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & -\lambda-1 & 1-4\lambda \\ 1 & 1 & -\lambda-2 & 2-5\lambda \end{Vmatrix},\quad \begin{Vmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -5 \\ 0 & 1 & -\lambda-1 & \lambda+6 \\ 1 & 1 & -\lambda-2 & 12 \end{Vmatrix};$$

$$\begin{Vmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & -5 & 1 \\ 0 & 1 & \lambda+6 & -\lambda-1 \\ 1 & 1 & 12 & -\lambda-2 \end{Vmatrix},\quad \begin{Vmatrix} 0 & 0 & 1 & \lambda+1 \\ 0 & 0 & -5 & -5\lambda-4 \\ 0 & 1 & \lambda+6 & \lambda^2+6\lambda+5 \\ 1 & 1 & 12 & 11\lambda+10 \end{Vmatrix},\quad \begin{Vmatrix} \lambda+1 & 0 & 0 & 1 \\ -5\lambda-4 & 0 & 0 & -5 \\ \lambda^2+6\lambda+5 & 0 & 1 & \lambda+6 \\ 11\lambda+10 & 1 & 1 & 12 \end{Vmatrix},$$

$$\begin{Vmatrix} \lambda+1 & 0 & 0 & 1 \\ -5\lambda-4 & 0 & 0 & -5 \\ \lambda^2+6\lambda+5 & -\lambda-1 & 1 & \lambda+6 \\ 10\lambda+9 & -\lambda & 1 & 12 \end{Vmatrix}.$$

Thus

$$Q(\lambda)=\begin{Vmatrix} \lambda+1 & 0 & 0 & 1 \\ -5\lambda-4 & 0 & 0 & -5 \\ \lambda^2+6\lambda+5 & -\lambda-1 & 1 & \lambda+6 \\ 10\lambda+9 & -\lambda & 1 & 12 \end{Vmatrix}$$

$$=\begin{Vmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{Vmatrix}\lambda^2+\begin{Vmatrix} 1 & 0 & 0 & 0 \\ -5 & 0 & 0 & 0 \\ 6 & -1 & 0 & 1 \\ 10 & -1 & 0 & 0 \end{Vmatrix}\lambda+\begin{Vmatrix} 1 & 0 & 0 & 1 \\ -4 & 0 & 0 & -5 \\ 5 & -1 & 1 & 6 \\ 9 & 0 & 1 & 12 \end{Vmatrix}.$$

Observing that

$$J^2=\begin{Vmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix},$$

we have

$$T=Q(J)=\begin{Vmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{Vmatrix}\cdot\begin{Vmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix}+$$

$$+\begin{Vmatrix} 1 & 0 & 0 & 0 \\ -5 & 0 & 0 & 0 \\ 6 & -1 & 0 & 1 \\ 12 & -1 & 0 & 0 \end{Vmatrix}\cdot\begin{Vmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{Vmatrix}+\begin{Vmatrix} 1 & 0 & 0 & 1 \\ -4 & 0 & 0 & -5 \\ 5 & -1 & 1 & 6 \\ 9 & 0 & 1 & 12 \end{Vmatrix}$$

$$=\begin{Vmatrix} 0 & 1 & 0 & 1 \\ 1 & -5 & 0 & -5 \\ 0 & 4 & 1 & 5 \\ -1 & 11 & 0 & 12 \end{Vmatrix}.$$

Check:

$$AT=\begin{Vmatrix} 0 & -1 & 1 & -1 \\ -1 & 6 & -5 & 5 \\ 0 & -4 & 3 & -5 \\ 1 & -12 & 11 & -12 \end{Vmatrix},\quad TJ=\begin{Vmatrix} 0 & -1 & 1 & -1 \\ -1 & 6 & -5 & 5 \\ 0 & -4 & 3 & -5 \\ 1 & -12 & 11 & -12 \end{Vmatrix},$$

i.e., $AT=TJ$.

$$|T|=\begin{vmatrix} 0 & 1 & 0 & 1 \\ 1 & -5 & 0 & -5 \\ 0 & 4 & 1 & 5 \\ -1 & 11 & 0 & 12 \end{vmatrix}=-1\neq 0.$$

Therefore

$$A=TJT^{-1}.$$

§ 9. Another Method of Constructing a Transforming Matrix

1. We shall now explain another method of constructing a transforming matrix which often leads to fewer computations than the method of the preceding section. However, we shall apply this second method only when the Jordan normal form and the elementary divisors

$$(\lambda-\lambda_1)^{p_1},\ (\lambda-\lambda_2)^{p_2},\ \ldots \tag{66}$$

of the given matrix A are known.

Let $A=TJT^{-1}$, where

$$J=\{\lambda_1E^{(p_1)}+H^{(p_1)},\ \lambda_2E^{(p_2)}+H^{(p_2)},\ldots\}=\begin{Vmatrix} \overbrace{\begin{matrix} \lambda_1 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_1 \end{matrix}}^{p_1} & & \\ & \overbrace{\begin{matrix} \lambda_2 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_2 \end{matrix}}^{p_2} & \\ & & \ddots \end{Vmatrix}.$$

Then denoting the k-th column of T by t_k $(k=1,2,\ldots,n)$, we replace the matrix equation

$$AT=TJ$$

by the equivalent system of equations

$$At_1=\lambda_1t_1,\ At_2=\lambda_1t_2+t_1,\ldots,At_{p_1}=\lambda_1t_{p_1}+t_{p_1-1} \tag{67}$$

$$At_{p_1+1}=\lambda_2t_{p_1+1},\ At_{p_1+2}=\lambda_2t_{p_1+2}+t_{p_1+1},\ldots,\ At_{p_1+p_2}=\lambda_2t_{p_1+p_2}+t_{p_1+p_2-1} \tag{68}$$

. . . . . .

which we rewrite as follows:

$$(A-\lambda_1E)t_1=O,\ (A-\lambda_1E)t_2=t_1,\ldots,(A-\lambda_1E)t_{p_1}=t_{p_1-1} \tag{67'}$$

$$(A-\lambda_2E)t_{p_1+1}=O,\ (A-\lambda_2E)t_{p_1+2}=t_{p_1+1},\ldots,(A-\lambda_2E)t_{p_1+p_2}=t_{p_1+p_2-1} \tag{68'}$$

. . . . . .

Thus, all the columns of T are split into 'Jordan chains' of columns: $[t_1,t_2,\ldots,t_{p_1}]$, $[t_{p_1+1},t_{p_1+2},\ldots,t_{p_1+p_2}]$, ...

To every Jordan block of J (or, what is the same, to every elementary divisor (66)) there corresponds its Jordan chain of columns. Each Jordan chain of columns is characterized by a system of equations of type (67), (68), etc.

The task of finding a transforming matrix T reduces to that of finding the Jordan chains that would give in all n linearly independent columns.

We shall show that these Jordan chains of columns can be determined by means of the reduced adjoint matrix $C(\lambda)$ (see Chapter IV, § 6).

For the matrix $C(\lambda)$ we have the identity

$$(\lambda E-A)C(\lambda)=\psi(\lambda)E. \tag{69}$$

where $\psi(\lambda)$ is the minimal polynomial of A.

Let

$$\psi(\lambda)=(\lambda-\lambda_0)^m\chi(\lambda)\qquad(\chi(\lambda_0)\neq 0).$$

We differentiate the identity (69) term by term $m-1$ times:

$$\left.\begin{aligned} (\lambda E-A)C'(\lambda)+C(\lambda)&=\psi'(\lambda)E \\ (\lambda E-A)C''(\lambda)+2C'(\lambda)&=\psi''(\lambda)E \\ \ldots\ldots\ldots\

In (71) we replace the matrices (72) by their k-th columns ($k = 1, 2, \ldots, n$). We obtain:

$$(A - \lambda_0 E) C_k = o,\ (A - \lambda_0 E) D_k = C_k, \ldots, (A - \lambda_0 E) K_k = G_k \quad (k = 1, 2, \ldots, n). \tag{73}$$

Since $C = C(\lambda_0) \neq O$,[35] we can find a k ($\leqq n$) such that

$$C_k \neq o. \tag{74}$$

Then the m columns

$$C_k, D_k, F_k, \ldots, G_k, K_k \tag{75}$$

are linearly independent. For let

$$\gamma C_k + \delta D_k + \ldots + \varkappa K_k = o. \tag{76}$$

Multiplying both sides of (76) successively by $A - \lambda_0 E, \ldots, (A - \lambda_0 E)^{m-1}$, we obtain

$$\delta C_k + \ldots + \varkappa G_k = o, \ldots, \varkappa C_k = o. \tag{77}$$

From (76) and (77) we find by (74):

$$\gamma = \delta = \ldots = \varkappa = 0.$$

Since the linearly independent columns (75) satisfy the system of equations (73), they form a Jordan chain of vectors corresponding to the elementary divisor $(\lambda - \lambda_0)^m$ (compare (73) with (67′)).

If $C_k = o$ for some k, but $D_k \neq o$, then the columns $D_k, \ldots, G_k, K_k$ form a Jordan chain of $m - 1$ vectors, etc.

2. We shall now show first of all how to construct a transforming matrix T in the case where the elementary divisors of A are pairwise co-prime:

$$(\lambda - \lambda_1)^{m_1}, \ldots, (\lambda - \lambda_s)^{m_s},$$
$$(\lambda_i \neq \lambda_j \text{ for } i \neq j;\ i, j = 1, 2, \ldots, s).$$

With the elementary divisor $(\lambda - \lambda_j)^{m_j}$ we associate the Jordan chain of columns

$$C^{(j)}, D^{(j)}, \ldots, G^{(j)}, K^{(j)},$$

constructed as indicated above. Then

$$(A - \lambda_j E) C^{(j)} = o,\ (A - \lambda_j E) D^{(j)} = C^{(j)}, \ldots, (A - \lambda_j E) K^{(j)} = G^{(j)}. \tag{78}$$

When we give to j the values $1, 2, \ldots, s$, we obtain s Jordan chains containing n columns in all. These columns are lincarly independent.

[35] From $C(\lambda_0) = O$ it would follow that all the elements of $C(\lambda)$ have a common divisor of positive degree, in contradiction to the definition of $C(\lambda)$.

For, suppose that

$$\sum_{j=1}^{s} [\gamma_j C^{(j)} + \delta_j D^{(j)} + \ldots + \varkappa_j K^{(j)}] = o. \tag{79}$$

We multiply both sides of (79) on the left by

$$(A - \lambda_1 E)^{m_1} \cdots (A - \lambda_{j-1} E)^{m_{j-1}} (A - \lambda_j E)^{m_j - 1} (A - \lambda_{j+1} E)^{m_{j+1}} \cdots (A - \lambda_s E)^{m_s} \tag{80}$$

and obtain

$$\varkappa_j = 0.$$

Replacing $m_j - 1$ successively by $m_j - 2$, $m_j - 3$, ... in (80), we find:

$$\gamma_j = \delta_j = \ldots = \varkappa_j = 0 \quad (j = 1, 2, \ldots, s),$$

and this is what we had to prove.

We define the matrix T by the formula

$$T = (C^{(1)}, D^{(1)}, \ldots, K^{(1)}; C^{(2)}, D^{(2)}, \ldots, K^{(2)}; \ldots; C^{(s)}, D^{(s)}, \ldots, K^{(s)}). \tag{81}$$

Example.

$$A = \begin{Vmatrix} 8 & 3 & -10 & -3 \\ 3 & -1 & -4 & 2 \\ 2 & 3 & -2 & -4 \\ 2 & -1 & -3 & 2 \\ 1 & 2 & -1 & -3 \end{Vmatrix};$$

$$\varphi(\lambda) = \Delta(\lambda) = (\lambda - 1)^2 (\lambda + 1)^2 = \lambda^4 - 2\lambda^2 + 1.$$

elementary divisors: $(\lambda - 1^2), (\lambda + 1)^2$,

$$\Psi(\lambda, \mu) = \frac{\varphi(\mu) - \varphi(\lambda)}{\mu - \lambda} = \mu^3 + \lambda\mu^2 + (\lambda^2 - 2)\mu + \lambda^3 - 2\lambda,$$

$$\begin{array}{cccc} 3 & 2 & 2 & 1 \\ \cdot & \cdot & \cdot & \cdot \\ 1 & 4 & 0 & 2 \end{array}$$

$$C(\lambda) = \Psi(\lambda E, A) = A^3 + \lambda A^2 + (\lambda^2 - 2) A + (\lambda^3 - 2\lambda) E.$$

We make up the first column $C_1(\lambda)$:

$$C_1(\lambda) = [A^3]_1 + \lambda [A^2]_1 + (\lambda^2 - 2) A_1 + (\lambda^3 - 2\lambda) E_1.$$

For the computation of the first column of A^2 we multiply all the rows of A into the first column of A. We obtain:[36] $[A^2]_1 = (1, 4, 0, 2)$. Multiplying all the rows of A into this column, we find: $[A^3]_1 = (3, 6, 2, 3)$.

Therefore

$$C_1(\lambda) = \begin{Vmatrix} 3 \\ 6 \\ 2 \\ 3 \end{Vmatrix} + \lambda \begin{Vmatrix} 1 \\ 4 \\ 0 \\ 2 \end{Vmatrix} + (\lambda^2 - 2) \begin{Vmatrix} 3 \\ 2 \\ 2 \\ 1 \end{Vmatrix} + (\lambda^3 - 2\lambda) \begin{Vmatrix} 1 \\ 0 \\ 0 \\ 0 \end{Vmatrix} = \begin{Vmatrix} \lambda^3 + 3\lambda^2 - \lambda - 3 \\ 2\lambda^2 + 4\lambda + 2 \\ 2\lambda^2 - 2 \\ \lambda^3 + 2\lambda + 1 \end{Vmatrix}$$

[36] The columns into which we multiply the rows are written underneath the rows of A. The elements of the row of column-sums are set up in italics, for checking.

Hence $C_1(1) = (0, 8, 0, 4)$ and $C_1'(1) = (8, 8, 4, 4)$. As $C_1(-1) = (0, 0, 0, 0)$, we pass on to the second column and, proceeding as before, we find: $C_2(-1) = (-4, 0, -4, 0)$ and $C_2'(-1) = (4, -4, 4, -4)$. We set up the matrix:

$$(C_1(1), C_1'(1); C_2(-1), C_2'(-1)) = \begin{Vmatrix} 0 & 8 & -4 & 4 \\ 8 & 8 & 0 & -4 \\ 0 & 4 & -4 & 4 \\ 4 & 4 & 0 & -4 \end{Vmatrix}.$$

We cancel[37] 4 in the first two columns and -4 in the last two columns.

$$T = \begin{Vmatrix} 0 & 2 & 1 & -1 \\ 2 & 2 & 0 & 1 \\ 0 & 1 & 1 & -1 \\ 1 & 1 & 0 & 1 \end{Vmatrix}.$$

We leave it to the reader to verify that

$$AT = T \cdot \begin{Vmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{Vmatrix}$$

3. Coming now to the general case, we shall investigate the Jordan chains of vectors corresponding to a characteristic value λ_0 for which there are p elementary divisors $(\lambda - \lambda_0)^m$, q elementary divisors $(\lambda - \lambda_0)^{m-1}$, r elementary divisors $(\lambda - \lambda_0)^{m-2}$, etc.

As a preliminary to this, we establish some properties of the matrices

$$C = C(\lambda_0),\ D = C'(\lambda_0),\ F = \frac{1}{2!}C''(\lambda_0),\ \ldots,\ K = \frac{1}{(m-1)!}C^{(m-1)}(\lambda_0). \tag{82}$$

1. *The matrices* (82) *can be represented in the form of polynomials in* A:

$$C = h_1(A),\ D = h_2(A),\ \ldots,\ K = h_m(A), \tag{83}$$

where

$$h_i(\lambda) = \frac{\psi(\lambda)}{(\lambda - \lambda_0)^i} \quad (i = 1, 2, \ldots, m). \tag{84}$$

For

$$C(\lambda) = \Psi(\lambda E, A),$$

where

$$\Psi(\lambda, \mu) = \frac{\psi(\mu) - \psi(\lambda)}{\mu - \lambda}.$$

[37] A Jordan chain remains a Jordan chain when all its columns are multiplied by a number $c \neq 0$.

Therefore

$$\frac{1}{k!}C^{(k)}(\lambda_0) = \frac{1}{k!}\Psi^{(k)}(\lambda_0 E, A), \tag{85}$$

where

$$\frac{1}{k!}\Psi^{(k)}(\lambda_0, \mu) = \frac{1}{k!}\left[\frac{\partial^k}{\partial\lambda^k}\Psi(\lambda, \mu)\right]_{\lambda=\lambda_0} = \frac{1}{k!}\left[\frac{\partial^k}{\partial\lambda^k}\frac{\psi(\mu)}{\mu - \lambda}\right]_{\lambda=\lambda_0} = \frac{\psi(\mu)}{(\mu - \lambda_0)^{k+1}} = h_{k+1}(\mu). \tag{86}$$

(83) follows from (82), (85), and (86).

2. *The matrices* (82) *have the ranks*

$$p,\ 2p + q,\ 3p + 2q + r,\ \ldots.$$

This property of the matrices (82) follows immediately from 1. and Theorem 8 (§ 7). if we equate the rank to $n - d$ and use formula (48) for the defect of a function on A (p. 154).

3. *In the sequence of matrices* (82) *every column of each matrix is a linear combination of the columns of every following matrix.*

Let us take two matrices $h_i(A)$ and $h_k(A)$ in (82) (see 1.). Suppose that $i < k$. Then it follows from (84) that:

$$h_i(A) = h_k(A)\,(A - \lambda_0 E)^{k-i}.$$

Hence the j-th column y_j $(j = 1, 2, \ldots, n)$ of $h_i(A)$ is expressed linearly by the columns $z_1, z_2, \ldots, z_n$ of $h_k(A)$:

$$y_j = \sum_{g=1}^{n} \alpha_g z_g,$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the elements of the j-th column of $(A - \lambda_0 E)^{k-i}$.

4. *Without changing the basic formulas* (71) *we may replace any column in* C *by an arbitrary linear combination of all the columns, provided we make the corresponding replacements in* $D, \ldots, K$.

We now proceed to the construction of the Jordan chains of columns for the elementary divisors

$$\underbrace{(\lambda - \lambda_0)^m, \ldots, (\lambda - \lambda_0)^m}_{p};\quad \underbrace{(\lambda - \lambda_0)^{m-1}, \ldots, (\lambda - \lambda_0)^{m-1}}_{q};\ \ldots.$$

Using the properties 2. and 4., we transform the matrix C into the form

$$C = (C_1, C_2, \ldots, C_p; o, o, \ldots, o); \tag{87}$$

where the columns $C_1, C_2, \ldots, C_p$ are linearly independent. Now

$$D=(D_1, D_2, \ldots, D_p;\ D_{p+1}, \ldots, D_n).$$

By 3., for every i $(1 \leqq i \leqq p)$ C_i is a linear combination of the columns $D_1, D_2, \ldots, D_n$:

$$C_i = \alpha_1 D_1 + \cdots + \alpha_p D_p + \alpha_{p+1} D_{p+1} + \cdots + \alpha_n D_n. \tag{88}$$

We multiply both sides of this equation by $A - \lambda_0 E$. Observing (see (73)) that

$$(A-\lambda_0 E)\, C_i = o \quad (i=1, 2, \ldots, p), \quad (A-\lambda_0 E)\, D_j = C_j \quad (j=1, 2, \ldots, n),$$

we obtain by (87)

$$o = \alpha_1 C_1 + \alpha_2 C_2 + \cdots + \alpha_p C_p;$$

hence in (88)

$$\alpha_1 = \cdots = \alpha_p = 0.$$

Therefore the columns $C_1, C_2, \ldots, C_p$ are linearly independent combinations of the columns $D_{p+1}, \ldots, D_n$. Therefore by 4. and 2., we can, without changing the matrix C, take the columns $C_1, \ldots, C_p$ instead of $D_{p+1}, \ldots, D_{2p}$ and zeros instead of $D_{2p+q+1}, \ldots, D_n$.

Then the matrix D assumes the form

$$D=(D_1, \ldots, D_p;\ C_1, C_2, \ldots, C_p;\ D_{2p+1}, \ldots, D_{2p+q};\ o, o, \ldots, o). \tag{89}$$

In the same way, preserving the forms (87) and (89) of the matrices C and D, we can represent the next matrix F in the form

$$\left.\begin{aligned} F=(&F_1, \ldots, F_p;\ D_1, \ldots, D_p;\ F_{2p+1}, F_{2p+q};\ C_1, \ldots, C_p;\\ &D_{2p+1}, \ldots, D_{2p+q}, F_{3p+2q+1}, \ldots, F_{3p+2q+r}, o, \ldots, o), \end{aligned}\right\} \tag{90}$$

etc.

Formulas (73) gives us the Jordan chains

$$\left.\begin{aligned} &\underbrace{\overbrace{(C_1, D_1, \ldots, K_1)}^{m}, \ldots, \overbrace{(C_p, D_p, \ldots, K_p)}^{m};}_{p} \\ &\underbrace{\overbrace{(D_{2p+1}, F_{2p+1} \ldots, K_{2p+1},)}^{m-1} \ldots, \overbrace{(D_{2p+q}, F_{2p+q}, \ldots, K_{2p+q})}^{m-1};\ \ldots}_{q} \end{aligned}\right\} \tag{91}$$

These Jordan chains are linearly independent. For all the columns C_i in (91) are linearly independent, because they form p linearly independent columns of C. All the columns C_i, D_j in (91) are independent, because they form $2p + q$ independent columns in D, etc.; finally, *all the columns in* (91)

are independent, because they form $n_0 = mp + (m-1)q + \ldots$ *independent columns in* K. The number of columns in (91) is equal to the sum of the exponents of the elementary divisors corresponding to the given characteristic valne λ_0.

Suppose that the matrix $A = \| a_{ik} \|_1^n$ has s distinct characteristic values λ_j

$$(j=1, 2, \ldots, s;$$

$$\Delta(\lambda) = (\lambda-\lambda_1)^{n_1} (\lambda-\lambda_2)^{n_2} \cdots (\lambda-\lambda_s)^{n_s}$$

$$\psi(\lambda) = (\lambda-\lambda_1)^{m_1} (\lambda-\lambda_2)^{m_2} \cdots (\lambda-\lambda_s)^{m_s}).$$

For each characteristic value λ_j we form its system of independent Jordan chains (91); the number of columns in this system is equal to n_j $(j=1, 2, 3, \ldots, s)$. All the chains so obtained contain $n = n_1 + n_2 + \ldots + n_s$ columns.

These n columns are linearly independent and form one of the required transforming matrices T.

The proof of the linear independence of these n columns proceeds as follows.

Every linear combination of these n columns can be represented in the form

$$\sum_{j=1}^{s} H_j = o, \tag{92}$$

where H_j is a linear combination of columns in the Jordan chains (91) corresponding to the characteristic valne λ_j $(j=1, 2, \ldots, s)$. But every column in the Jordan chain corresponding to the characteristic value λ_j satisfies the equation

$$(A-\lambda_j E)^{m_j} x = o.$$

Therefore

$$(A-\lambda_j E)^{m_j} H_j = o. \tag{93}$$

We take a fixed number j $(1 \leqq j \leqq s)$ and construct the Lagrange-Sylvester interpolation polynomial $r(\lambda)$ (See Chapter V, §§ 1, 2) with the following values on the spectrum of the matrix:

$$r(\lambda_i) = r'(\lambda_i) = \ldots = r^{(m_i-1)}(\lambda_i) = 0 \text{ for } i \neq j$$

and

$$r(\lambda_j) = 1,\ r'(\lambda_j) = \cdots = r^{(m_j-1)}(\lambda_j) = 0.$$

Then, for every $i \neq j$, $r(\lambda)$ is divisible by $(\lambda-\lambda_1)^{m_i}$ without remainder; therefore by (93),

$$r(A)\, H_i = o \qquad (i \neq j). \tag{94}$$

In exactly the same way, the difference $r(\lambda) - 1$ is divisible by $(\lambda - \lambda_j)^{m_j}$ without remainder; therefore

$$r(A) H_j = H_j. \tag{95}$$

Multiplying both sides of (92) by $r(A)$, we obtain from (94) and (95):

$$H_j = o.$$

This is valid for every $j = 1, 2, \ldots, s$. But H_j is a linear combination of independent columns corresponding to one and the same characteristic value λ_j $(j = 1, 2, \ldots, s)$. Therefore all the coefficients in the linear combination H_j $(j = 1, 2, \ldots, s)$, and hence all the coefficients in (92), are equal to zero.

Note. Let us point out some transformations on the columns of the matrix T under which it is transformed into the same Jordan form (with the same arrangement of the Jordan diagonal blocks):

I. *Multiplication of all the columns of an arbitrary Jordan chain by a non-zero number.*

II. *Addition to each column (beginning with the second) of a Jordan chain of the preceding column of the same chain, multiplied by one and the same arbitrary number.*

III. *Addition to all the columns of a Jordan chain of the corresponding columns of another chain containing the same or a larger number of columns and corresponding to the same characteristic value.*

Example 1.

$$A = \begin{Vmatrix} 1 & 0 & 0 & 1 & -1 \\ 0 & 1 & -2 & 3 & -3 \\ 0 & 0 & -1 & 2 & -2 \\ 1 & -1 & 1 & 0 & 1 \\ 1 & -1 & 1 & -1 & 2 \end{Vmatrix}; \qquad \begin{gathered} \Delta(\lambda) = (\lambda-1)^4(\lambda+1), \\ \psi(\lambda) = (\lambda-1)^2(\lambda+1) = \lambda^3 - \lambda^2 - \lambda + 1. \\ \text{Elementary divisors of the matrix } A \\ (\lambda-1)^2,\ (\lambda-1)^2,\ \lambda+1. \end{gathered}$$

$$J = \begin{Vmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{Vmatrix},$$

$$\Psi(\lambda, \mu) = \frac{\psi(\mu) - \psi(\lambda)}{\mu - \lambda} = \mu^2 + (\lambda - 1)\mu + \lambda^2 - \lambda - 1,$$

$$C(\lambda) = \Psi(\lambda E, A) = A^2 + (\lambda - 1)A + (\lambda^2 - \lambda - 1)E.$$

Let us compute successively the column of A^2 and the corresponding columns of $C(\lambda)$, $C(1)$, $C'(\lambda)$, $C'(1)$, $C(-1)$. We must obtain two linearly independent columns of $C(1)$ and one non-zero column of $C(-1)$.

$$C(\lambda) = \begin{Vmatrix} 1 & 0 & 0 & 2* \\ 0 & 1 & 0 & 2* \\ 0 & 0 & 1 & 0* \\ 2 & -2 & 2 & -1* \\ 2 & -2 & 2 & -2* \end{Vmatrix} + (\lambda-1) \begin{Vmatrix} 1 & 0 & 0 & 1* \\ 0 & 1 & -2 & 3* \\ 0 & 0 & -1 & 2* \\ 1 & -1 & 1 & 0* \\ 1 & -1 & 1 & -1* \end{Vmatrix} + (\lambda^2-\lambda-1) \begin{Vmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{Vmatrix}$$

$$C(+1) = \begin{Vmatrix} 0 & 0 & 0 & 2* \\ 0 & 0 & 0 & 2* \\ 0 & 0 & 0 & 0* \\ 2 & -2 & 2 & -2* \\ 2 & -2 & 2 & -2* \end{Vmatrix}, \quad C'(\lambda) = \begin{Vmatrix} 1 & 0 & 0 & 1* \\ 0 & 1 & -2 & 3* \\ 0 & 0 & -1 & 2* \\ 1 & -1 & 1 & 0* \\ 1 & -1 & 1 & -1* \end{Vmatrix} + (2\lambda-1) \begin{Vmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{Vmatrix},$$

$$C'(+1) = \begin{Vmatrix} 2 & * & * & 1 & * \\ 0 & * & * & 3 & * \\ 0 & * & * & 2 & * \\ 1 & * & * & 1 & * \\ 1 & * & * & -1 & * \end{Vmatrix}, \quad C(-1) = \begin{Vmatrix} 0 & 0 & 0 & * & * \\ 0 & 0 & 4 & * & * \\ 0 & 0 & 4 & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & * & * \end{Vmatrix}.$$

Therefore[38]

$$T = (C_1(+1), C_1'(+1), C_4(+1), C_4'(+1), C_3(-1)) = \begin{Vmatrix} 0 & 2 & 2 & 1 & 0 \\ 0 & 0 & 2 & 3 & 4 \\ 0 & 0 & 0 & 2 & 4 \\ 2 & 1 & -2 & 1 & 0 \\ 2 & 1 & -2 & -1 & 0 \end{Vmatrix}.$$

The matrix T can be simplified a little. We

1) Divide the fifth column by 4;
2) Add the first column to the third and the second to the fourth;
3) Subtract the third column from the fourth;
4) Divide the first and second columns by 2;
5) Subtract the first column, multiplied by ½, from the second.

Then we obtain the matrix

$$T_1 = \begin{Vmatrix} 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 2 & 1 \\ 1 & 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{Vmatrix}.$$

We leave it to the reader to verify that $AT_1 = T_1 J$ and $|T_1| \neq 0$.

Example 2.

$$A = \begin{Vmatrix} 1 & -1 & 1 & -1 \\ -3 & 3 & -5 & 4 \\ 8 & -4 & 3 & -4 \\ 15 & -10 & 11 & -11 \end{Vmatrix}; \qquad \begin{gathered} \Delta(\lambda) = (\lambda+1)^4, \\ \psi(\lambda) = (\lambda+1)^3. \\ \text{Elementary divisors: } (\lambda+1)^3,\ \lambda+1. \end{gathered}$$

[38] Here the subscript denotes the number of the column; for example, $C_3(-1)$ denotes the third column of $C(-1)$.

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

We form the polynomials

$$h_1(\lambda)=\frac{\psi(\lambda)}{\lambda+1}=(\lambda+1)^3,\ h_2(\lambda)=\frac{\psi(\lambda)}{(\lambda+1)^2}=\lambda+1,\ h_3(\lambda)=\frac{\psi(\lambda)}{(\lambda+1)^3}=1$$

and the matrices[39]

$$C=h_1(A)=(A+E)^2,\ D=h_2(A)=A+E,\ F=E:$$

$$C=\begin{Vmatrix} 0 & 0 & 0 & 0 \\ 2 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ -2 & 1 & -1 & 1 \end{Vmatrix},\ D=\begin{Vmatrix} 2 & -1 & 1 & -1 \\ -3 & 4 & -5 & 4 \\ 8 & -4 & 4 & -4 \\ 15 & -10 & 11 & -10 \end{Vmatrix},\ F=\begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix}.$$

For the first three columns of T we take the third column of these matrices: $T=(C_3, D_3, F_3, *)$. In the matrices C, D, F, we subtract twice the third column from the first and we add the third column to the second and to the fourth. We obtain

$$\tilde{C}=\begin{Vmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{Vmatrix},\ \tilde{D}=\begin{Vmatrix} 0 & 0 & 1 & 0 \\ 7 & -1 & -5 & -1 \\ 0 & 0 & 4 & 0 \\ -7 & 1 & 11 & 1 \end{Vmatrix},\ \tilde{F}=\begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -2 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{Vmatrix}.$$

In the matrices $\tilde{D}$, $\tilde{F}$, we add the fourth column, multiplied by 7, to the first and subtract the fourth column from the second. We obtain

$$\tilde{\tilde{C}}=\begin{Vmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{Vmatrix},\ \tilde{\tilde{D}}=\begin{Vmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & -5 & -1 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 11 & 1 \end{Vmatrix},\ \tilde{\tilde{F}}=\begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 5 & 0 & 1 & 1 \\ 7 & -1 & 0 & 1 \end{Vmatrix}.$$

For the last column of T we take the first column of $\tilde{\tilde{F}}$. Then we have

$$T=(C_3, D_2, F_3, \tilde{\tilde{F}}_1)=\begin{Vmatrix} 0 & 1 & 0 & 1 \\ 1 & -5 & 0 & 0 \\ 0 & 4 & 1 & 5 \\ -1 & 11 & 0 & 7 \end{Vmatrix}.$$

As a check, we can verify that $AT=TJ$ and that $|T|\neq 0$.

CHAPTER VII

THE STRUCTURE OF A LINEAR OPERATOR IN AN n-DIMENSIONAL SPACE

(*Geometrical Theory of Elementary Divisors*)

The analytic theory of elementary divisors expounded in the preceding chapter has enabled us to determine for every square matrix a similar matrix having 'normal' or 'canonical' form. On the other hand, we have seen in Chapter III that the behaviour of a linear operator in an n-dimensional space with respect to various bases is given by means of a class of similar matrices. The existence of a matrix of normal form in such a class is closely connected with important and deep properties of a linear operator in an n-dimensional space. The study of these properties is the object of the present chapter. The investigation of the structure of a linear operator will lead us, independently of the contents of the preceding chapter, to the theory of transformations of a matrix to a normal form. Therefore the contents of this chapter may be called *the geometrical theory of elementary divisors*.[1]

§ 1. The Minimal Polynomial of a Vector and a Space (with Respect to a Given Linear Operator)

1. We consider an n-dimensional vector space $\boldsymbol{R}$ over the field F and a linear operator $\boldsymbol{A}$ in this space.

Let $\boldsymbol{x}$ be an arbitrary vector of $\boldsymbol{R}$. We form the sequence of vectors

$$\boldsymbol{x},\ \boldsymbol{Ax},\ \boldsymbol{A^2x},\ \ldots. \tag{1}$$

Since the space is finite-dimensional, there is an integer p $(0\leqq p\leqq n)$ such that the vectors $\boldsymbol{x}$, $\boldsymbol{Ax}$, ..., $\boldsymbol{A^{p-1}x}$ are linearly independent, while $\boldsymbol{A^px}$ is a linear combination of these vectors with coefficients in F:

[1] The account of the geometric theory of elementary divisors to be given here is based on our paper [167]. For other geometrical constructions of the theory of elementary divisors, see [22], §§ 96-99 and also [53].

$$A^p x = -\gamma_1 A^{p-1} x - \gamma_2 A^{p-2} x - \cdots - \gamma_p x. \tag{2}$$

We form the monic polynomial $\varphi(\lambda) = \lambda^p + \gamma_1 \lambda^{p-1} + \cdots + \gamma_{p-1}\lambda + \gamma_p$. (A *monic* polynomial is a polynomial in which the coefficient of the highest power of the variable is unity.) Then (2) can be written:

$$\varphi(A)\, x = o. \tag{3}$$

Every polynomial $\varphi(\lambda)$ for which (3) holds will be called an *annihilating polynomial for the vector* x.[2] But it is easy to see that of all the monic annihilating polynomials of x the one we have constructed is of least degree. This polynomial will be called the *minimal annihilating polynomial of* x or simply the *minimal polynomial of* x.

Note that every annihilating polynomial $\tilde{\varphi}(\lambda)$ of x is divisible by the minimal polynomial $\varphi(\lambda)$.

For let

$$\tilde{\varphi}(\lambda) = \varphi(\lambda)\, \varkappa(\lambda) + \varrho(\lambda),$$

where $\varkappa(\lambda)$, $\varrho(\lambda)$ are quotient and remainder on dividing $\tilde{\varphi}(\lambda)$ by $\varphi(\lambda)$. Then

$$\tilde{\varphi}(A)\, x = \varkappa(A)\, \varphi(A)\, x + \varrho(A)\, x = \varrho(A)\, x$$

and therefore $\varrho(A)\, x = o$. But the degree of $\varrho(\lambda)$ is less than that of the minimal polynomial $\varphi(\lambda)$. Hence $\varrho(\lambda) \equiv 0$.

From what we have proved it follows, in particular, that every vector x has only one minimal polynomial.

2. We choose a basis $e_1, e_2, \ldots, e_n$ in R. We denote by $\varphi_1(\lambda), \varphi_2(\lambda), \ldots, \varphi_n(\lambda)$ the minimal polynomials of the basis vectors $e_1, e_2, \ldots, e_n$ and by $\psi(\lambda)$ the least common multiple of these polynomials ($\psi(\lambda)$ is taken with highest coefficient 1). Then $\psi(\lambda)$ is an annihilating polynomial for all the basis vectors $e_1, e_2, \ldots, e_n$. Since every vector $x \in R$ is representable in the form $x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n$, we have

$$\psi(A)\, x = x_1 \psi(A)\, e_1 + x_2 \psi(A)\, e_2 + \cdots + x_n \psi(A)\, e_n = o,$$

i.e.,

$$\psi(A) = O. \tag{4}$$

The polynomial $\psi(\lambda)$ is called *an annihilating polynomial for the whole space* R. Let $\tilde{\psi}(\lambda)$ be an arbitrary annihilating polynomial for the whole space R. Then $\tilde{\psi}(\lambda)$ is an annihilating polynomial for the basis vectors

[2] Of course, the phrase 'with respect to the given operator A' is tacitly understood. For the sake of brevity, this circumstance is not mentioned in the definition, because throughout this entire chapter we shall deal with a single operator A.

$e_1, e_2, \ldots, e_n$. Therefore $\tilde{\psi}(\lambda)$ must be a common multiple of the minimal polynomials $\varphi_1(\lambda), \varphi_2(\lambda), \ldots, \varphi_n(\lambda)$ of these vectors and must therefore be divisible without remainder by their *least* common multiple $\psi(\lambda)$. Hence it follows that, of all the annihilating polynomials for the whole space R, the one we have constructed, $\psi(\lambda)$, has the least degree and it is monic. This polynomial is uniquely determined by the space R and the operator A and is called the *minimal polynomial of the space* R.[3] The uniqueness of the minimal polynomial of the space R follows from the statement proved above: *every annihilating polynomial* $\tilde{\psi}(\lambda)$ *of the space* R *is divisible by the minimal polynomial* $\psi(\lambda)$. Although the construction of the minimal polynomial $\psi(\lambda)$ was associated with a definite basis $e_1, e_2, \ldots, e_n$, the polynomial $\psi(\lambda)$ itself does not depend on the choice of this basis (this follows from the uniqueness of the minimal polynomial for the space R).

Finally we mention that the minimal polynomial of the space R annihilates every vector x of R so that *the minimal polynomial of the space is divisible by the minimal polynomial of every vector in the space.*

§ 2. Decomposition into Invariant Subspaces with Co-Prime Minimal Polynomials

1. If some collection of vectors R' forming part of R has the property that the sum of any two vectors of R' and the product of any vector of R' by a number $\alpha \in \mathrm{F}$ always belongs to R', then that manifold R' is itself a vector space, a *subspace* of R.

If two subspaces R' and R'' of R are given and if it is known that

1. R' and R'' have no vector in common except the null vector, and
2. every vector x of R can be represented in the form of a sum

$$x = x' + x'' \quad (x' \in R', x'' \in R''), \tag{5}$$

then we shall say that the space R is *decomposed* into the two subspaces R' and R'' and shall write:

$$R = R' + R'' \tag{6}$$

Note that the condition 1. implies the uniqueness of the representation (5). For if for a certain vector x we had two distinct representations in the form of a sum of terms from R' and R'', (5) and

$$x = \tilde{x}' + \tilde{x}'' \quad (\tilde{x}' \in R', \tilde{x}'' \in R'') \tag{7}$$

then, subtracting (7) from (5) term by term, we would obtain:

[3] If in some basis $e_1, e_2, \ldots, e_n$ a matrix $A = \| a_{ik} \|_1^n$ then the annihilating or minimal polynomial of the space R (with respect to A) is the annihilating or minimal polynomial of the matrix A, and vice versa. Compare with Chapter IV, § 6.

$$x' - \tilde{x}' = \tilde{x}'' - x'',$$

i.e., equality of the non-null vectors $x' - \tilde{x}' \in R'$ and $\tilde{x}'' - x'' \in R''$, which, by 1., is impossible.

Thus, condition 1. may be replaced by the requirement that the representation (5) be unique. In this form, the definition of decomposition immediately extends to an arbitrary number of subspaces.

Let

$$R = R' + R''$$

and let $e'_1, e'_2, \ldots, e'_{n'}$ and $e''_1, e''_2, \ldots, e''_{n''}$ be bases of R' and R'', respectively. Then the reader can easily prove that all these $n' + n''$ vectors are linearly independent and form a basis of R, so that a basis of the whole space is formed from bases of the subspaces. It follows, in particular, that $n = n' + n''$.

Example 1. Suppose that in a three-dimensional space three directions, not parallel to one and the same plane, are given. Since every vector in the space can be split, uniquely, into components in these three directions, we have

$$R = R' + R'' + R''',$$

where R is the set of all the vectors of one space, R' the set of all vectors parallel to the first direction, R'' to the second, and R''' to the third. In this case, $n = 3$ and $n' = n'' = n''' = 1$.

Example 2. Suppose that in a three-dimensional space a plane and a line intersecting the plane are given. Then

$$R = R' + R'',$$

where R is the set of all vectors of our space, R' the set of all vectors parallel to the given plane, and R'' the set of all vectors parallel to the given line. In this example, $n = 3$, $n' = 2$, $n'' = 1$.

2. A subspace $R' \subset R$ is called *invariant* with respect to the operator A if $AR' \subset R'$, i.e. if $x \in R'$ implies $Ax \in R'$. In other words, the operator A carries a vector of an invariant subspace into a vector of the same subspace.

In what follows we shall carry out a decomposition of the whole space into subspaces invariant with respect to A. The decomposition reduces the study of the behavior of an operator in the whole space to the study of its behavior in the various component subspaces.

We shall now prove the following theorem:

Theorem 1 (First Theorem on the Decomposition of a Space into Invariant Subspaces): *If for a given operator A the minimal polynomial $\psi(\lambda)$ of the space is represented over* F *in the form of a product of two co-prime polynomials $\psi_1(\lambda)$ and $\psi_2(\lambda)$ (with highest coefficients* 1)

$$\psi(\lambda) = \psi_1(\lambda)\,\psi_2(\lambda), \tag{8}$$

then the whole space R splits into two invariant subspaces I_1 and I_2

$$R = I_1 + I_2, \tag{9}$$

whose minimal polynomials are $\psi_1(\lambda)$ and $\psi_2(\lambda)$, respectively.

Proof. We denote by I_1 the set of all vectors $x \in R$ satisfying the equation $\psi_1(A)\,x = o$. I_2 is similarly defined by the equation $\psi_2(A)\,x = o$. I_1 and I_2 so defined are subspaces of R.

Since $\psi_1(\lambda)$ and $\psi_2(\lambda)$ are co-prime, it follows that there exist polynomials $\chi_1(\lambda)$ and $\chi_2(\lambda)$ (with coefficients in F) such that

$$1 = \psi_1(\lambda)\,\chi_1(\lambda) + \psi_2(\lambda)\,\chi_2(\lambda). \tag{10}$$

Now let x be an arbitrary vector of R. In (10) we replace λ by A and we apply both sides of the operator equation so obtained to the vector x:

$$x = \psi_1(A)\,\chi_1(A)\,x + \psi_2(A)\,\chi_2(A)\,x, \tag{11}$$

i.e.,

$$x = x' + x'', \tag{12}$$

where

$$x' = \psi_2(A)\,\chi_2(A)\,x, \quad x'' = \psi_1(A)\,\chi_1(A)\,x. \tag{13}$$

Furthermore,

$$\psi_1(A)\,x' = \psi(A)\,\chi_2(A)\,x = o, \quad \psi_2(A)\,x'' = \psi(A)\,\chi_1(A)\,x = o,$$

i.e.,

$$x' \in I_1, \text{ and } x'' \in I_2.$$

I_1 and I_2 have only the null vector in common. For if $x_0 \in I_1$ and $x_0 \in I_2$, i.e., $\psi_1(A)\,x_0 = o$ and $\psi_2(A)\,x_0 = o$, then by (11)

$$x_0 = \chi_1(A)\,\psi_1(A)\,x_0 + \chi_2(A)\,\psi_2(A)\,x_0 = o.$$

Thus we have proved that $R = I_1 + I_2$.

Now suppose that $\boldsymbol{x} \in \boldsymbol{I}_1$. Then $\psi_1(\boldsymbol{A})\boldsymbol{x} = \boldsymbol{o}$. Multiplying both sides of this equation by $\boldsymbol{A}$ and reversing the order of $\boldsymbol{A}$ and $\psi_1(\boldsymbol{A})$, we obtain $\psi_1(\boldsymbol{A})\boldsymbol{A}\boldsymbol{x} = \boldsymbol{o}$, i.e., $\boldsymbol{A}\boldsymbol{x} \in \boldsymbol{I}_1$. This proves that the subspace $\boldsymbol{I}_1$ is invariant with respect to $\boldsymbol{A}$. The invariance of the subspace $\boldsymbol{I}_2$ is proved similarly.

We shall now show that $\psi_1(\lambda)$ is the minimal polynomial of $\boldsymbol{I}_1$. Let $\tilde{\psi}_1(\lambda)$ be an arbitrary annihilating polynomial for $\boldsymbol{I}_1$, and $\boldsymbol{x}$ an arbitrary vector of $\boldsymbol{R}$. Using the decomposition (12) already established, we write:

$$\tilde{\psi}_1(\boldsymbol{A})\psi_2(\boldsymbol{A})\boldsymbol{x} = \psi_2(\boldsymbol{A})\tilde{\psi}_1(\boldsymbol{A})\boldsymbol{x}' + \tilde{\psi}_1(\boldsymbol{A})\psi_2(\boldsymbol{A})\boldsymbol{x}'' = \boldsymbol{o}.$$

Since $\boldsymbol{x}$ is an arbitrary vector of $\boldsymbol{R}$, it follows that the product $\tilde{\psi}_1(\lambda)\psi_2(\lambda)$ is an annihilating polynomial for $\boldsymbol{R}$ and is therefore divisible by $\psi(\lambda) = \psi_1(\lambda)\psi_2(\lambda)$ without remainder; in other words, $\psi_1(\lambda)$ is divisible by $\tilde{\psi}_1(\lambda)$. But $\tilde{\psi}_1(\lambda)$ is an arbitrary annihilating polynomial for $\boldsymbol{I}_1$ and $\psi_1(\lambda)$ is a particular one of the annihilating polynomials (by the definition of $\boldsymbol{I}_1$). Hence $\psi_1(\lambda)$ is the minimal polynomial of $\boldsymbol{I}_1$. In exactly the same way it is shown that $\psi_2(\lambda)$ is the minimal polynomial for the invariant subspace $\boldsymbol{I}_2$.

This completes the proof of the theorem.

Let us decompose $\psi(\lambda)$ into irreducible factors over F:

$$\psi(\lambda) = [\varphi_1(\lambda)]^{e_1}[\varphi_2(\lambda)]^{e_2}\cdots[\varphi_s(\lambda)]^{e_s} \tag{14}$$

(here $\varphi_1(\lambda), \varphi_2(\lambda), \ldots, \varphi_s(\lambda)$ are distinct irreducible polynomials over F with highest coefficient 1). Then by the theorem we have

$$\boldsymbol{R} = \boldsymbol{I}_1 + \boldsymbol{I}_2 + \ldots + \boldsymbol{I}_s, \tag{15}$$

where $\boldsymbol{I}_k$ is an invariant subspace with the minimal polynomial $[\varphi_k(\lambda)]^{e_k}$ $(k = 1, 2, \ldots, s)$.

Thus, the theorem reduces the study of the behaviour of a linear operator in an arbitrary space to the study of the behaviour of this operator in a space where the minimal polynomial is a power of an irreducible polynomial over F. We shall take advantage of this to prove the following important theorem:

THEOREM 2: *In a vector space there always exists a vector whose minimal polynomial coincides with the minimal polynomial of the whole space.*

We consider first the special case where the minimal polynomial of the space $\boldsymbol{R}$ is a power of an irreducible polynomial $\varphi(\lambda)$:

$$\psi(\lambda) = [\varphi(\lambda)]^l.$$

In $\boldsymbol{R}$ we choose a basis $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$. The minimal polynomial of $\boldsymbol{e}_i$ is a divisor of $\psi(\lambda)$ and is therefore representable in the form $[\varphi(\lambda)]^{l_i}$, where $l_i \leqq l$ $(i = 1, 2, \ldots, n)$.

But the minimal polynomial of the space is the least common multiple of the minimal polynomials of the basis vectors, so that $\psi(\lambda)$ is the largest of the powers $[\varphi(\lambda)]^{l_i}$ $(i = 1, 2, \ldots, n)$. In other words, $\psi(\lambda)$ coincides with the minimal polynomial of one of the basis vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$.

Turning now to the general case, we prove the following preliminary lemma:

LEMMA: *If the minimal polynomials of the vectors $\boldsymbol{e}'$ and $\boldsymbol{e}''$ are co-prime, then the minimal polynomial of the sum vector $\boldsymbol{e}' + \boldsymbol{e}''$ is equal to the product of the minimal polynomials of the constituent vectors.*

Proof. Let $\chi_1(\lambda)$ and $\chi_2(\lambda)$ be the minimal polynomials of the vectors $\boldsymbol{e}'$ and $\boldsymbol{e}''$. By assumption, $\chi_1(\lambda)$ and $\chi_2(\lambda)$ are co-prime. Let $\chi(\lambda)$ be an arbitrary annihilating polynomial of the vector $\boldsymbol{e} = \boldsymbol{e}' + \boldsymbol{e}''$. Then

$$\chi_2(\boldsymbol{A})\chi(\boldsymbol{A})\boldsymbol{e}' = \chi_2(\boldsymbol{A})\chi(\boldsymbol{A})\boldsymbol{e} - \chi(\boldsymbol{A})\chi_2(\boldsymbol{A})\boldsymbol{e}'' = \boldsymbol{o},$$

i.e., $\chi_2(\lambda)\chi(\lambda)$ is an annihilating polynomial of $\boldsymbol{e}'$. Therefore $\chi_2(\lambda)\chi(\lambda)$ is divisible by $\chi_1(\lambda)$, and since $\chi_1(\lambda)$ and $\chi_2(\lambda)$ are co-prime, $\chi(\lambda)$ is divisible by $\chi_1(\lambda)$. It is proved similarly that $\chi(\lambda)$ is divisible by $\chi_2(\lambda)$. But $\chi_1(\lambda)$ and $\chi_2(\lambda)$ are co-prime. Therefore $\chi(\lambda)$ is divisible by the product $\chi_1(\lambda)\chi_2(\lambda)$. Thus, every annihilating polynomial of the vector $\boldsymbol{e}$ is divisible by $\chi_1(\lambda)\chi_2(\lambda)$. Therefore $\chi_1(\lambda)\chi_2(\lambda)$ is the minimal polynomial of the vector $\boldsymbol{e} = \boldsymbol{e}' + \boldsymbol{e}''$.

We now return to Theorem 2. For the proof in the general case we use the decomposition (15). Since the minimal polynomials of the subspaces $\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots, \boldsymbol{I}_s$ are powers of irreducible polynomials, our assertion is already proved for these subspaces. Therefore there exist vectors $\boldsymbol{e}' \in \boldsymbol{I}_1, \boldsymbol{e}'' \in \boldsymbol{I}_2, \ldots, \boldsymbol{e}^{(s)} \in \boldsymbol{I}_s$ whose minimal polynomials are $[\varphi_1(\lambda)]^{e_1}$, $[\varphi_2(\lambda)]^{e_2}, \ldots, [\varphi_s(\lambda)]^{e_s}$, respectively. By the lemma, the minimal polynomial of the vector $\boldsymbol{e} = \boldsymbol{e}' + \boldsymbol{e}'' + \cdots + \boldsymbol{e}^{(s)}$ is equal to the product

$$[\varphi_1(\lambda)]^{e_1}\,[\varphi_2(\lambda)]^{e_2}\cdots[\varphi_s(\lambda)]^{e_s},$$

i.e., to the minimal polynomial of the space $\boldsymbol{R}$.

§ 3. Congruence. Factor Space

1. Suppose given a subspace $\boldsymbol{I} \subset \boldsymbol{R}$. We shall say that two vectors $\boldsymbol{x}, \boldsymbol{y}$ of $\boldsymbol{R}$ are *congruent* modulo $\boldsymbol{I}$ and shall write $\boldsymbol{x} \equiv \boldsymbol{y} \pmod{\boldsymbol{I}}$ if and only if $\boldsymbol{y} - \boldsymbol{x} \in \boldsymbol{I}$. It is easy to verify that the concept of congruence so introduced has the following properties:

For all $x, y, z \in R$

1. $x \equiv x \pmod{I}$ (reflexivity of congruence).
2. From $x \equiv y \pmod{I}$ it follows that $y \equiv x \pmod{I}$ (symmetry of congruence).
3. From $x \equiv y \pmod{I}$ and $y \equiv z \pmod{I}$ it follows that $x \equiv z \pmod{I}$ (transitivity of congruence).

The presence of these three properties enables us to make use of congruence to divide all the vectors of the space into classes, by assigning vectors that are pairwise congruent (mod I) to the same class (vectors of distinct classes are incongruent (mod I)). The class containing the vector x will be denoted by $\bar{x}$.[5] The subspace I is one of these classes, namely $\bar{o}$. Note that to every congruence $x \equiv y \pmod{I}$ there corresponds the equality[6] of the associated classes: $\bar{x} = \bar{y}$.

It is elementary to prove that congruences may be added term by term and multiplied by a number of F:

1. From

$$x \equiv x' \text{ and } y \equiv y' \pmod{I}$$

it follows that

$$x + y \equiv x' + y' \pmod{I}.$$

2. From

$$x \equiv x' \pmod{I}$$

it follows that

$$\alpha x \equiv \alpha x' \pmod{I} \ (\alpha \in F).$$

These properties of congruence show that the operations of addition and multiplication by a number of F do not 'break up' the classes. If we take two classes $\bar{x}$ and $\bar{y}$ and add elements $x, x', \ldots$ of the first class to arbitrary elements $y, y', \ldots$ of the second class, then all the sums so obtained belong to one and the same class, which we call the *sum* of the classes $\bar{x}$ and $\bar{y}$ and denote by $\bar{x} + \bar{y}$. Similarly, if all the vectors $x, x' \ldots$ of the class x are multiplied by a number $\alpha \in F$, then the products belong to one class, which we denote by $\alpha\bar{x}$.

Thus, in the manifold $\bar{R}$ of all classes $\bar{x}, \bar{y}, \ldots$ two operations are introduced: 'addition' and 'multiplication by a number of F.' It is easy to verify that these operations have the properties set forth in the definition of a vector space (Chapter III, § 1). Therefore $\bar{R}$, as well as R, is a vector

[5] Since each class contains an infinite set of vectors, there is, by this condition, an infinite number of ways of designating the class.

[6] That is, identity.

space over the field F. We shall say that $\bar{R}$ is a *factor space* of R. If $n, m, \bar{n}$ are the dimensions of the spaces R, I, $\bar{R}$, respectively, then $\bar{n} = n - m$.

2. All the concepts introduced in this section can be illustrated very well by the following example.

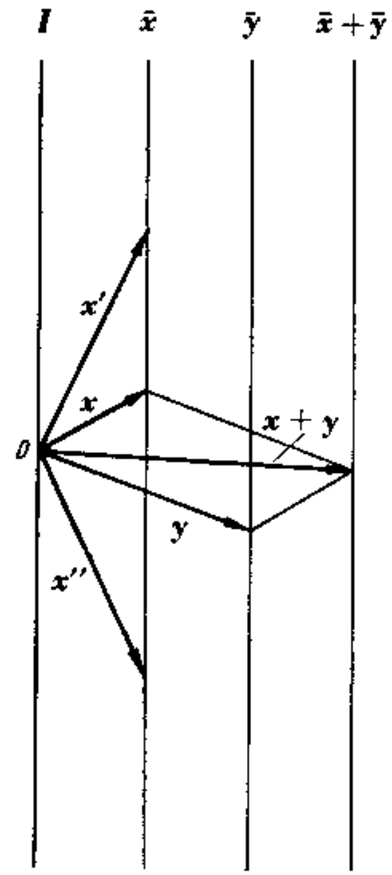

Fig. 4

Example. Let R be the set of all vectors of a three-dimensional space and F the field of real numbers. For greater clarity, we shall represent vectors in the form of directed segments beginning at a point O. Let I be a straight line passing through O (more accurately: the set of vectors that lie along some line passing through O; Fig. 4.).

The congruence $x \equiv x' \pmod{I}$ signifies that the vectors x and x' differ by a vector of I, i.e., the segment containing the end-points of x and x' is parallel to I. Therefore the class $\bar{x}$ is represented by the line passing through the end-point of x and parallel to I (more accurately: by the 'bundle' of vectors starting from O whose end-points lie on that line). 'Bundles' may be added and multiplied by a real number (by adding and multiplying the vectors that occur in the bundles). These 'bundles' are also the elements of the factor space $\bar{R}$. In this example, $n = 3$, $m = 1$, $\bar{n} = 2$.

We obtain another example by taking for I a plane passing through O. In this example, $n = 3$, $m = 2$, $\bar{n} = 1$.

Now let A be a linear operator in R. Let us assume that I *is an invariant subspace with respect to* A. The reader will easily prove that from $x \equiv x' \pmod{I}$ it follows that $Ax \equiv Ax' \pmod{I}$, so that the operator A can be applied to both sides of a congruence. In other words, if the operator A is applied to all vectors $x, x', \ldots$ of a class $\bar{x}$, then the vectors $Ax, Ax', \ldots$ also belong to one class, which we denote by $A\bar{x}$. The linear operator A carries classes into classes and is, thus, a linear operator in $\bar{R}$.

We shall say that the vectors $x_1, x_2, \ldots, x_p$ are *linearly dependent modulo* I if there exist numbers $\alpha_1, \alpha_2, \ldots, \alpha_p$ in F, not all equal to zero, such that

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_p x_p \equiv o \pmod{I}. \qquad (16)$$

Note that not only the concept of linear dependence of vectors, but also all the concepts, statements, and reasonings, in the preceding sections of this chapter can be repeated word for word with the symbol '$=$' replaced throughout by the symbol '$\equiv \pmod{\boldsymbol{I}}$,' where $\boldsymbol{I}$ is some fixed subspace invariant with respect to $\boldsymbol{A}$.

Thus, we can introduce the concepts of an annihilating polynomial and of the minimal polynomial of a vector or a space $(\mathrm{mod}\ \boldsymbol{I})$. All these concepts will be called 'relative,' in contrast to the 'absolute' concepts that were introduced earlier (and that hold for the symbol '$=$').

The reader should observe that *the relative minimal polynomial (of a vector or a space) is a divisor of the absolute one*. For example, let $\sigma_1(\lambda)$ be the relative minimal polynomial of a vector $\boldsymbol{x}$ and $\sigma(\lambda)$ the corresponding absolute minimal polynomial.

Then

$$\sigma(\boldsymbol{A})\,\boldsymbol{x} = \boldsymbol{o}\,,$$

and hence it follows that also

$$\sigma(\boldsymbol{A})\,\boldsymbol{x} \equiv \boldsymbol{o} \pmod{\boldsymbol{I}}.$$

Therefore $\sigma(\lambda)$ is a relative annihilating polynomial of $\boldsymbol{x}$ and as such is divisible by the relative minimal polynomial $\sigma_1(\lambda)$.

Side by side with the 'absolute' statements of the preceding sections we have 'relative' statements. For example, we have the statement: 'In every space there always exists a vector whose relative minimal polynomial coincides with the relative minimal polynomial of the whole space.'

The truth of all 'relative' statements depends on the fact that by operating with congruences modulo $\boldsymbol{I}$ we deal essentially with equalities—however not in the space $\boldsymbol{R}$, but in the space $\bar{\boldsymbol{R}}$.

§ 4. Decomposition of a Space into Cyclic Invariant Subspaces

1. Let $\sigma(\lambda) = \lambda^p + \alpha_1 \lambda^{p-1} + \cdots + \alpha_{p-1}\lambda + \alpha_p$ be the minimal polynomial of a vector $\boldsymbol{e}$. Then the vectors

$$\boldsymbol{e},\ \boldsymbol{A}\boldsymbol{e},\ \ldots,\ \boldsymbol{A}^{p-1}\boldsymbol{e} \tag{17}$$

are linearly independent, and

$$\boldsymbol{A}^p\boldsymbol{e} = -\alpha_p\boldsymbol{e} - \alpha_{p-1}\boldsymbol{A}\boldsymbol{e} - \cdots - \alpha_1\boldsymbol{A}^{p-1}\boldsymbol{e}\,. \tag{18}$$

The vectors (17) form a basis of a p-dimensional subspace $\boldsymbol{I}$. We shall call this subspace *cyclic* in view of the special character of the basis (17) and of (18).[7] The operator $\boldsymbol{A}$ carries the first vector of (17) into the second, the second into the third, etc. The last basis vector is carried by $\boldsymbol{A}$ into a linear combination of the basis vectors in accordance with (18). Thus, $\boldsymbol{A}$ carries every basis vector into a vector of $\boldsymbol{I}$ and hence an arbitrary vector of $\boldsymbol{I}$ into another vector of $\boldsymbol{I}$. In other words, *a cyclic subspace is always invariant with respect to* $\boldsymbol{A}$.

Every vector $\boldsymbol{x} \in \boldsymbol{I}$ is representable in the form of a linear combination of the basis vectors (17), i.e., in the form

$$\boldsymbol{x} = \chi(\boldsymbol{A})\,\boldsymbol{e}\,, \tag{19}$$

where $\chi(\lambda)$ is a polynomial in λ of degree $\leqq p-1$ with coefficients in F. By forming all possible polynomials $\chi(\lambda)$ of degree $\leqq p-1$ with coefficients in F we obtain all the vectors of $\boldsymbol{I}$, each once only, i.e., for only one polynomial $\chi(\lambda)$. In view of the basis (17) or the formula (19) we shall say that the vector $\boldsymbol{e}$ *generates* the subspace.

Note that *the minimal polynomial of the generating vector* $\boldsymbol{e}$ *is also the minimal polynomial of the whole subspace* $\boldsymbol{I}$.

2. We are now ready to establish the fundamental proposition of the whole theory, according to which the space $\boldsymbol{R}$ splits into cyclic subspaces.

Let $\psi_1(\lambda) = \psi(\lambda) = \lambda^m + \alpha_1\lambda^{m-1} + \cdots + \alpha_m$ be the minimal polynomial of the space $\boldsymbol{R}$. Then there exists a vector $\boldsymbol{e}$ in the space for which this polynomial is minimal (Theorem 2, p. 180). Let $\boldsymbol{I}_1$ denote the cyclic subspace with the basis

$$\boldsymbol{e},\ \boldsymbol{A}\boldsymbol{e},\ \ldots,\ \boldsymbol{A}^{m-1}\boldsymbol{e}\,. \tag{20}$$

If $n = m$, then $\boldsymbol{R} = \boldsymbol{I}_1$. Suppose that $n > m$ and that the polynomial

$$\psi_2(\lambda) = \lambda^p + \beta_1\lambda^{p-1} + \cdots + \beta_p$$

is the minimal polynomial of $\boldsymbol{R} \pmod{\boldsymbol{I}_1}$. By the remark at the end of § 3, $\psi_2(\lambda)$ is a divisor of $\psi_1(\lambda)$, i.e., there exists a polynomial $\varkappa(\lambda)$ such that

$$\psi_1(\lambda) = \psi_2(\lambda)\,\varkappa(\lambda)\,. \tag{21}$$

---

[7] It would be more accurate to call this subspace: cyclic with respect to the linear operator $\boldsymbol{A}$. But since the whole theory is built up with reference to a single operator $\boldsymbol{A}$, the words 'with respect to the linear operator $\boldsymbol{A}$' are omitted for the sake of brevity (see the similar remark in footnote 2, p. 176).

Moreover, in $\boldsymbol{R}$ there exists a vector $\boldsymbol{g}^*$ whose relative minimal polynomial is $\psi_2(\lambda)$. Then

$$\psi_2(\boldsymbol{A})\,\boldsymbol{g}^* \equiv \boldsymbol{o} \pmod{\boldsymbol{I}_1}, \tag{22}$$

i.e., there exists a polynomial $\chi(\lambda)$ of degree $\leqq m-1$ such that

$$\psi_2(\boldsymbol{A})\,\boldsymbol{g}^* = \chi(\boldsymbol{A})\,\boldsymbol{e}. \tag{23}$$

We apply the operator $\varkappa(\boldsymbol{A})$ to both sides of the equation. Then by (21) we obtain on the left $\psi_1(\boldsymbol{A})\boldsymbol{g}^*$, i.e. zero, because $\psi_1(\lambda)$ is the absolute minimal polynomial of the space; therefore

$$\varkappa(\boldsymbol{A})\,\chi(\boldsymbol{A})\,\boldsymbol{e} = \boldsymbol{o}.$$

This equation shows that the product $\varkappa(\lambda)\,\chi(\lambda)$ is an annihilating polynomial of the vector $\boldsymbol{e}$ and is therefore divisible by the minimal polynomial $\psi_1(\lambda) = \varkappa(\lambda)\,\psi_2(\lambda)$, so that $\chi(\lambda)$ is divisible by $\psi_2(\lambda)$:

$$\chi(\lambda) = \varkappa_1(\lambda)\,\psi_2(\lambda), \tag{24}$$

where $\varkappa_1(\lambda)$ is a polynomial. Using this decomposition of $\chi(\lambda)$, we may rewrite (23) as follows:

$$\psi_2(\boldsymbol{A})\,[\boldsymbol{g}^* - \varkappa_1(\boldsymbol{A})\,\boldsymbol{e}] = \boldsymbol{o}. \tag{25}$$

We now introduce the vector

$$\boldsymbol{g} = \boldsymbol{g}^* - \varkappa_1(\boldsymbol{A})\,\boldsymbol{e}. \tag{26}$$

Then (25) can be written as follows:

$$\psi_2(\boldsymbol{A})\,\boldsymbol{g} = \boldsymbol{o}. \tag{27}$$

The last equation shows that $\psi_2(\lambda)$ is an absolute annihilating polynomial of the vector $\boldsymbol{g}$ and is therefore divisible by the absolute minimal polynomial of $\boldsymbol{g}$. On the other hand, we have from (26):

$$\boldsymbol{g} \equiv \boldsymbol{g}^* \pmod{\boldsymbol{I}_1}. \tag{28}$$

Hence $\psi_2(\lambda)$, being the relative minimal polynomial of $\boldsymbol{g}^*$, is the same for $\boldsymbol{g}$ as well. Comparing the last two statements, we deduce that $\psi_2(\lambda)$ is simultaneously the relative and the absolute minimal polynomial of $\boldsymbol{g}$.

From the fact that $\psi_2(\lambda)$ is the absolute minimal polynomial of $\boldsymbol{g}$ it follows that the subspace $\boldsymbol{I}_2$ with the basis

$$\boldsymbol{g},\ \boldsymbol{A}\boldsymbol{g},\ \ldots,\ \boldsymbol{A}^{p-1}\boldsymbol{g} \tag{29}$$

is cyclic.

From the fact that $\psi_2(\lambda)$ is the relative minimal polynomial of $\boldsymbol{g} \pmod{\boldsymbol{I}_1}$ it follows that the vectors (29) are linearly independent $(\mathrm{mod}\,\boldsymbol{I}_1)$, i.e., no linear combination with coefficients not all zero can be equal to a linear combination of the vectors (20). Since the latter are themselves linearly independent, our last statement asserts the linear independence of the $m+p$ vectors

$$\boldsymbol{e},\ \boldsymbol{A}\boldsymbol{e},\ \ldots,\ \boldsymbol{A}^{m-1}\boldsymbol{e};\ \boldsymbol{g},\ \boldsymbol{A}\boldsymbol{g},\ \ldots,\ \boldsymbol{A}^{p-1}\boldsymbol{g}. \tag{30}$$

The vectors (30) form a basis of the invariant subspace $\boldsymbol{I}_1 + \boldsymbol{I}_2$ of dimension $m+p$.

If $n = m+p$, then $\boldsymbol{R} = \boldsymbol{I}_1 + \boldsymbol{I}_2$. If $n > m+p$, we consider $\boldsymbol{R} \pmod{\boldsymbol{I}_1 + \boldsymbol{I}_2}$ and continue our process of separating cyclic subspaces. Since the whole space $\boldsymbol{R}$ is of finite dimension n, this process must come to an end with some subspace $\boldsymbol{I}_t$, where $t \leqq n$.

We have arrived at the following theorem:

Theorem 3 (Second Theorem on the Decomposition of a Space into Invariant Subspaces): *Relative to a given linear operator $\boldsymbol{A}$ the space can always be split into cyclic subspaces $\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots, \boldsymbol{I}_t$ with minimal polynomials $\psi_1(\lambda), \psi_2(\lambda), \ldots, \psi_t(\lambda)$*

$$\boldsymbol{R} = \boldsymbol{I}_1 + \boldsymbol{I}_2 + \ldots + \boldsymbol{I}_t \tag{31}$$

such that $\psi_1(\lambda)$ coincides with the minimal polynomial $\psi(\lambda)$ of the whole space and that each $\psi_i(\lambda)$ is divisible by $\psi_{i-1}(\lambda)$ $(i = 2, 3, \ldots, t)$.

3. We now mention some properties of cyclic spaces. Let $\boldsymbol{R}$ be a cyclic n-dimensional space and $\psi(\lambda) = \lambda^m + \ldots$ its minimal polynomial. Then it follows from the definition of a cyclic space that $m = n$.[1] Conversely, suppose that $\boldsymbol{R}$ is an arbitrary space and that it is known that $m = n$. Applying the proof of the decomposition theorem, we represent $\boldsymbol{R}$ in the form (31). But the dimension of the cyclic subspace $\boldsymbol{I}_1$ is m, because its minimal polynomial coincides with the minimal polynomial of the whole space. Since $m = n$ by assumption, we have $\boldsymbol{R} = \boldsymbol{I}_1$, i.e., $\boldsymbol{R}$ is a cyclic space.

Thus we have established the following *criterion for cyclicity of a space*:

Theorem 4: *A space is cyclic if and only if its dimension is equal to the degree of its minimal polynomial.*

Next, suppose that we have a decomposition of a cyclic space $\boldsymbol{R}$ into two invariant subspaces $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$:

$$\boldsymbol{R} = \boldsymbol{I}_1 + \boldsymbol{I}_2. \tag{32}$$

We denote the dimensions of $\boldsymbol{R}$, $\boldsymbol{I}_1$, and $\boldsymbol{I}_2$ by n, n_1, and n_2, their minimal polynomials by $\psi(\lambda)$, $\psi_1(\lambda)$, and $\psi_2(\lambda)$, and the degrees of these minimal polynomials by m, m_1, and m_2, respectively. Then

$$m_1 \leqq n_1, \quad m_2 \leqq n_2. \tag{33}$$

We add these inequalities term by term:

$$m_1 + m_2 \leqq n_1 + n_2. \tag{34}$$

Since $\psi(\lambda)$ is the least common multiple of $\psi_1(\lambda)$ and $\psi_2(\lambda)$, we have

$$m \leqq m_1 + m_2. \tag{35}$$

Moreover, it follows from (32) that

$$n = n_1 + n_2. \tag{36}$$

(34), (35), and (36) give us a chain of relations

$$m \leqq m_1 + m_2 \leqq n_1 + n_2 = n. \tag{37}$$

But since the space $\boldsymbol{R}$ is cyclic, the extreme numbers of this chain, m and n, are equal. Therefore we have equality in the middle terms, i.e.,

$$m = m_1 + m_2 = n_1 + n_2.$$

From the fact that $m = m_1 + m_2$ we deduce that $\psi_1(\lambda)$ and $\psi_2(\lambda)$ are co-prime.

Bearing (33) in mind, we find from $m_1 + m_2 = n_1 + n_2$ that

$$m_1 = n_1 \quad \text{and} \quad m_2 = n_2. \tag{38}$$

These equations mean that the subspaces $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ are cyclic.

Thus we have arrived at the following proposition:

Theorem 5: *A cyclic space can only split into invariant subspaces that* 1. *are also cyclic and* 2. *have co-prime minimal polynomials.*

The same arguments (in the opposite order) show that Theorem 5 has a converse:

Theorem 6: *If a space is split into invariant subspaces that* 1. *are cyclic and* 2. *have co-prime minimal polynomials, then the space itself is cyclic.*

Suppose now that $\boldsymbol{R}$ is a cyclic space and that its minimal polynomial is a power of an irreducible polynomial over F: $\psi(\lambda) = [\varphi(\lambda)]^e$. In this case, the minimal polynomial of every invariant subspace of $\boldsymbol{R}$ must also be a power of this irreducible polynomial $\varphi(\lambda)$. Therefore the minimal polynomials of any two invariant subspaces cannot be co-prime. But then, by what we have proved, $\boldsymbol{R}$ cannot split into invariant subspaces.

Suppose, conversely, that some space $\boldsymbol{R}$ is known not to split into invariant subspaces. Then $\boldsymbol{R}$ is a cyclic space, for otherwise, by the second decomposition theorem, it could be split into cyclic subspaces; moreover, the minimal polynomial of $\boldsymbol{R}$ must be a power of an irreducible polynomial, because otherwise $\boldsymbol{R}$ could be split into invariant subspaces, by the first decomposition theorem.

Thus we have reached the following conclusion:

Theorem 7: *A space does not split into invariant subspaces if and only if* 1. *it is cyclic and* 2. *its minimal polynomial is a power of an irreducible polynomial over* F.

We now return to the decomposition (31) and split the minimal polynomials $\psi_1(\lambda), \psi_2(\lambda), \ldots, \psi_t(\lambda)$ of the cyclic subspaces $\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots, \boldsymbol{I}_t$ into irreducible factors over F:

$$\left.\begin{aligned} \psi_1(\lambda) &= [\varphi_1(\lambda)]^{c_1}[\varphi_2(\lambda)]^{c_2}\cdots[\varphi_s(\lambda)]^{c_s}, \\ \psi_2(\lambda) &= [\varphi_1(\lambda)]^{d_1}[\varphi_2(\lambda)]^{d_2}\cdots[\varphi_s(\lambda)]^{d_s}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \psi_t(\lambda) &= [\varphi_1(\lambda)]^{l_1}[\varphi_2(\lambda)]^{l_2}\cdots[\varphi_s(\gamma)]^{l_s} \end{aligned}\right\} \tag{39}$$
$$(c_k \geqq d_k \geqq \ldots \geqq l_k \geqq 0;^{8}\ k = 1, , 2, \ldots, s).$$

To $\boldsymbol{I}_1$ we apply the first decomposition theorem. Then we obtain

$$\boldsymbol{I}_1 = \boldsymbol{I}_1' + \boldsymbol{I}_1'' + \cdots + \boldsymbol{I}_1^{(s)};$$

where $\boldsymbol{I}_1', \boldsymbol{I}_1'', \ldots, \boldsymbol{I}_1^{(s)}$ are cyclic subspaces with the minimal polynomials $[\varphi_1(\lambda)]^{c_1}, [\varphi_2(\lambda)]^{c_2}, \ldots, [\varphi_s(\lambda)]^{c_s}$. Similarly we decompose the spaces $\boldsymbol{I}_2, \ldots, \boldsymbol{I}_t$. In this way we obtain a decomposition of the whole space $\boldsymbol{R}$ into cyclic subspaces with the minimal polynomials $[\varphi_k(\lambda)]^{c_k}, [\varphi_k(\lambda)]^{d_k}, \ldots, [\varphi_k(\lambda)]^{l_k}$ $(k = 1, 2, \ldots, s)$. (Here we neglect the powers whose exponents are zero.) From Theorem 7 it follows that these cyclic subspaces are indecomposable (into invariant subspaces). We have thus arrived at the following theorem:

Theorem 8 (Third Theorem on the Decomposition of a Space into Invariant Subspaces): *A space can always be split into cyclic invariant subspaces*

$$\boldsymbol{R} = \boldsymbol{I}' + \boldsymbol{I}'' + \cdots + \boldsymbol{I}^{(u)}. \tag{40}$$

such that the minimal polynomial of each of these cyclic subspaces is a power of an irreducible polynomial.

This theorem gives the decomposition of a space into indecomposable invariant subspaces.

[8] Some of the exponents $d_k, \ldots, l_k$ for $k > 1$ may be equal to zero.