NATIONAL BUREAU OF STANDARDS REPORT

1644

COMPUTATIONAL METHODS

OF

LINEAR ALGEBRA

Chapter 1

Basic Material from Linear Algebra

V. N. Faddeeva

Translated from the Russian by Curtis D. Benster

Editor: G. E. Forsythe

NBS

U. S. DEPARTMENT OF COMMERCE

NATIONAL BUREAU OF STANDARDS

U. S. DEPARTMENT OF COMMERCE
Charles Sawyer, *Secretary*

NATIONAL BUREAU OF STANDARDS
A. V. Astin, [redacted] *Director*

THE NATIONAL BUREAU OF STANDARDS

The scope of activities of the National Bureau of Standards is suggested in the following listing of the divisions and sections engaged in technical work. In general, each section is engaged in specialized research, development, and engineering in the field indicated by its title. A brief description of the activities, and of the resultant reports and publications, appears on the inside of the back cover of this report.

1. **Electricity.** Resistance Measurements. Inductance and Capacitance. Electrical Instruments. Magnetic Measurements. Electrochemistry.

2. **Optics and Metrology.** Photometry and Colorimetry. Optical Instruments. Photographic Technology. Length. Gage.

3. **Heat and Power.** Temperature Measurements. Thermodynamics. Cryogenics. Engines and Lubrication. Engine Fuels.

4. **Atomic and Radiation Physics.** Spectroscopy. Radiometry. Mass Spectrometry. Physical Electronics. Electron Physics. Atomic Physics. Neutron Measurements. Nuclear Physics. Radioactivity. X-Rays. Betatron. Nucleonic Instrumentation. Radiological Equipment. Atomic Energy Commission Instruments Branch.

5. **Chemistry.** Organic Coatings. Surface Chemistry. Organic Chemistry. Analytical Chemistry. Inorganic Chemistry. Electrodeposition. Gas Chemistry. Physical Chemistry. Thermochemistry. Spectrochemistry. Pure Substances.

6. **Mechanics.** Sound. Mechanical Instruments. Aerodynamics. Engineering Mechanics. Hydraulics. Mass. Capacity, Density, and Fluid Meters.

7. **Organic and Fibrous Materials.** Rubber. Textiles. Paper. Leather. Testing and Specifications. Organic Plastics. Dental Research.

8. **Metallurgy.** Thermal Metallurgy. Chemical Metallurgy. Mechanical Metallurgy. Corrosion.

9. **Mineral Products.** Porcelain and Pottery. Glass. Refractories. Enameled Metals. Building Stone. Concreting Materials. Constitution and Microstructure. Chemistry of Mineral Products.

10. **Building Technology.** Structural Engineering. Fire Protection. Heating and Air Conditioning. Exterior and Interior Coverings. Codes and Specifications.

11. **Applied Mathematics.** Numerical Analysis. Computation. Statistical Engineering. Machine Development.

12. **Electronics.** Engineering Electronics. Electron Tubes. Electronic Computers. Electronic Instrumentation.

13. **Ordnance Development.** Mechanical Research and Development. Electromechanical Fuzes. Technical Services. Missile Fuzing Research. Missile Fuzing Development. Projectile Fuzes. Ordnance Components. Ordnance Tests. Ordnance Research.

14. **Radio Propagation.** Upper Atmosphere Research. Ionospheric Research. Regular Propagation Services. Frequency Utilization Research. Tropospheric Propagation Research. High Frequency Standards. Microwave Standards.

15. **Missile Development.** Missile Engineering. Missile Dynamics. Missile Intelligence. Missile Instrumentation. Technical Services. Combustion.

NBS PROJECT **NBS REPORT**

April 1, 1952

1101-11-5100 1644

COMPUTATIONAL METHODS OF LINEAR ALGEBRA*

Chapter 1
Basic Material from Linear Algebra

V. N. Faddeeva

Translated from the Russian by Curtis D. Benster**

Editor: G. E. Forsythe

*Published 1950 in Moscow and Leningrad.

**This translation was carried out under the auspices of the Department of Mathematics, University of California, Los Angeles. It was edited by George E. Forsythe, National Bureau of Standards, Los Angeles.

CONTENTS OF ENTIRE BOOK

(only Chapter I is translated here)

EDITOR'S INTRODUCTION

The first time he meets matrices, the student of numerical analysis is not usually ready for a systematic development of elementary divisor theory - and still less for a treatment over a general field. But he does need to learn the main facts about the algebra of finite matrices over the real or complex field, and he should know something about the proofs. These matters are supplied very well by Faddeeva in the chapter translated here, and the editor therefore believes that the chapter will be very popular.

Analogous statements could be made about the analytical aspects of finite matrix theory: norms, limit, etc.

In the two later chapters of her book, the author describes various known numerical processes for inverting matrices and for obtaining their latent roots and vectors, and gives many numerical examples. These chapters probably also merit translation, and we hope later to be able to provide one.

The first chapter was translated specifically to provide an introductory text for the editor's 1952 summer course on numerical matrix methods at the University of California, Los Angeles. With this in view, the translator has done more than create an interesting and faithful translation: he has improved the presentation in several respects. He has corrected misprints, and has helped the student by occasionally inserting phrases and by adding two paragraphs of recapitulation (pp. 44 and 75). He has changed a few notations to suit American taste. Finally, he has replaced the Russian bibliography by an English and American one, and has considerably supplemented the author's three or four references to the bibliography.

Only the more significant of the translator's additions have been credited to him in the text.

AUTHOR'S PREFACE

The numerical solution of the problems of mathematical physics is most frequently connected with the numerical solution of basic problems of linear algebra -- that of solving a system of linear equations, and that of the computation of the latent roots of a matrix. The present book is an endeavor at systematizing the most important numerical methods of linear algebra -- the classical ones and those elaborated quite recently as well.

The author does not pretend to an exhaustive completeness, having included an exposition of those methods only that have already been tested in practice. In the exposition the author has not strived for an irreproachable rigor, and has not analysed all conceivable cases and sub-cases arising in the application of this or that method, having limited herself to the most typical and practically important cases.

The book consists of three chapters. In the first chapter is given the material from linear algebra that is indispensable to what follows. The second chapter is devoted to the numerical solution of systems of linear equations and parallel questions. Lastly, the third chapter contains a description of numerical methods of computing the latent roots and latent vectors of a matrix.

For the interest manifested in the manuscript, and for a number of valuable suggestions, I express my sincere thanks to A. L. Brudno and G. P. Akilov.

CHAPTER I
BASIC MATERIAL FROM LINEAR ALGEBRA[1]

This chapter will be of an introductory nature. Without detailed proofs, it will impart material from linear algebra that will be indispensable to an understanding of the following chapters.

§1. MATRICES

1. An aggregate of numbers--which are, generally speaking, complex--arranged in the form of a rectangular table, is called a rectangular matrix. This array will have m rows and n columns, and may be set forth in the form:

$$(1) \qquad A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

the first subscript, then, designating the row, the second designating the column, in which the element in point is located.

This may be abbreviated to the form:

$$A = (a_{ij}) \qquad (i = 1, 2, \cdots, m; \quad j = 1, 2, \cdots, n) .$$

[1]This translation was carried out under the auspices of the Department of Mathematics, University of California, Los Angeles. It was edited by George E. Forsythe, National Bureau of Standards, Los Angeles.

Two matrices are <u>equal</u> if their corresponding elements are equal.

Matrices composed of a single row are called simply <u>rows</u> (or, as we shall approach them later, <u>row vectors</u>). Matrices composed of a single column are called <u>columns</u> (or <u>column vectors</u>).

If the number of rows of a matrix equals the number, n, of columns, it is called <u>square</u>, and of the <u>n-th order</u>.

Among square matrices, an important role is played by <u>diagonal</u> matrices, i.e., matrices of which only the elements along the <u>principal</u> (<u>leading</u>) <u>diagonal</u> are different from zero:

(2) $$\begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & \alpha_n \end{pmatrix} = [\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n] .$$

If all the numbers α_i of such a matrix are equal to each other, the matrix is said to be <u>scalar</u>:

(3) $$\begin{pmatrix} \alpha & 0 & \cdots & 0 \\ 0 & \alpha & \cdots & 0 \\ 0 & 0 & \cdots & \alpha \end{pmatrix} = [\alpha]$$

and, if $\alpha = 1$, the matrix is said to be the <u>unit</u> matrix:[1]

---

[1]Editor's note: Faddeeva and other continental mathematicians use the symbol E for a unit matrix.

(4)
$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 1 \end{pmatrix} = I \quad .$$

Lastly, a matrix all of whose elements are equal to zero is called a null matrix, or zero matrix. We shall designate it by the symbol 0.

The determinant whose elements are the elements of a square matrix (without disarrangement), is said to be the determinant of that matrix, and we write the determinant of the matrix A as $|A|$, or often as d(A).

2. Multiplication of a matrix by a number. The addition of matrices. A matrix whose elements are obtained by multiplying all the elements of the matrix A by a number α is called the product of the number α and the matrix A:

(5)
$$\alpha A = \begin{pmatrix} \alpha a_{11} & \alpha a_{12} & \cdots & \alpha a_{1n} \\ \alpha a_{21} & \alpha a_{22} & \cdots & \alpha a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \alpha a_{m1} & \alpha a_{m2} & \cdots & \alpha a_{mn} \end{pmatrix} .$$

A matrix C whose elements are the sums of the corresponding elements of A and B, matrices having like numbers of rows and columns, is called the sum of A and B:

$$(\cdot) \quad A + B = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{pmatrix}.$$

The operations introduced above have the following properties, as will be readily seen:

1. $A + (B + C) = (A + B) + C.$
2. $A + B = B + A.$
3. $A + 0 = A.$
4. $(\alpha + \beta)A = \alpha A + \beta A.$
5. $\alpha(A + B) = \alpha A + \alpha B.$

Here A, B, and C are matrices; α and β are numbers—generally speaking, complex.

3. The multiplication of matrices. Multiplication of the matrices A and B is defined only on the assumption that the number of columns of matrix A equals the number of rows of matrix[1] B. On this assumption, the elements of the product, C = AB, are defined in the following manner: the element in the i-th row and the j-th column of the matrix C is equal to the

[1]Translator's note: Indicating by superior letters the dimensions of the matrix A, i.e., the number, m, of rows, and the number, n, of columns, thus: $\overset{mxn}{A}$, the product condition may be given a more lucid expression notationally. Thus $\overset{mxn}{A} \cdot \overset{pxr}{B}$ is possible only if p = n. A and B are then conformable.

sum of the products of the elements of the i-th row of the matrix A by the corresponding elements of the j-th column of matrix[1] B. Thus:

$$(7)\quad AB = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \cdot & \cdot & \cdot & \cdot \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \cdot & \cdot & \cdot & \cdot \\ c_{m1} & c_{m2} & \cdots & c_{mp} \end{pmatrix} = C ,$$

where

$$(8)\quad \begin{gathered} c_{ij} = a_{i1} b_{1j} + a_{i2} b_{2j} + \cdots + a_{in} b_{nj} = \sum_{k=1}^{n} a_{ik} b_{kj} \\ (i = 1, 2, \cdots, m ; \quad j = 1, 2, \cdots, p) . \end{gathered}$$

It is to be noted that the product of two rectangular matrices is again a rectangular matrix, the number of rows of which is equal to the number of rows of the first matrix, and the number of columns of which is equal to the number of columns of the second matrix: $\overset{mxn}{A} \cdot \overset{nxp}{B} = \overset{mxp}{C}$. So, for instance, the product of a square matrix and a matrix composed of one column is a matrix of one column.

The commutative law for multiplication does not, generally speaking, hold. We shall make a few observations on this subject, however. The

---

[1]Translator's note: The bare definition of matrix multiplication given in the text provides a basis for the logically succeeding development. It may, however, appear arbitrary to the student who has not previously reconnoitred the ground. For a brief account of the source of this definition of matrix multiplication see, e.g., [1], §1-6. Abundant numerical illustrations will be found in [2], §1.4.

matrices AB and BA make sense simultaneously only if the number of rows of the first matrix is equal to the number of columns of the second, and the number of columns of the first is equal to the number of rows of the second. Given the fulfillment of these conditions, the matrices AB and BA will both be square, but of different orders, unless A and B be square. Thus even to put the question of the equality of the matrices AB and BA makes sense only for square matrices. But even in this case, generally speaking, AB $\neq$ BA.

In particular cases multiplication may be commutative, and in such cases the matrices are said to commute. Thus, for example, scalar matrices commute with any square matrix of the same order, for

$$\begin{pmatrix} \alpha & 0 & \cdots & 0 \\ 0 & \alpha & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & \alpha \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} \alpha & 0 & \cdots & 0 \\ 0 & \alpha & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & \alpha \end{pmatrix} =$$

(9)

$$= \begin{pmatrix} \alpha a_{11} & \alpha a_{12} & \cdots & \alpha a_{1n} \\ \alpha a_{21} & \alpha a_{22} & \cdots & \alpha a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \alpha a_{n1} & \alpha a_{n2} & \cdots & \alpha a_{nn} \end{pmatrix} .$$

Hence follows the special role of the unit matrix in the multiplication of matrices, to wit: amongst all square matrices of the same order, the unit matrix plays the same role as the number one does among numbers.

Indeed,

$$AI = IA = A \quad .$$

It can be shown that the multiplication of matrices is associative, viz., if AB and (AB)C make sense, so also do BC and A(BC), and

$$1. \quad A(BC) = (AB)C \quad .$$

The matrix product has also these properties:

$$2. \quad \alpha(AB) = (\alpha A)B = A(\alpha B) \quad ;$$

$$3. \quad (A + B)C = AC + BC \quad ;$$

$$4. \quad C(A + B) = CA + CB \quad ,$$

where A, B, C are matrices, α a number.

Let us interchange rows and columns in the matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = (a_{ij}) \quad ;$$

we obtain the transposed[1] matrix or transpose

---

[1]The word conjugate appears in the older literature.

(10) $$A^T = A' = (a_{ij})' = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \cdot & \cdot & \cdot & \cdot \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix} = (a_{ji}) \quad .$$

The following rule (the reversal rule) for a transposed product should be noted:

(10a) $$(AB)' = B'A' \quad .$$

In proof of this, note that the i-th row and j-th column of the matrix (AB)' is equal to the element of the j-th row and i-th column of the matrix AB, for this is merely the interchange of row with column, i.e., transposition; and that is equal to

(11) $$a_{j1}b_{1i} + a_{j2}b_{2i} + \cdots + a_{jn}b_{ni} \quad .$$

The last expression is obviously equal to the sum of the products of the elements of the i-th row of the matrix B' and the corresponding elements of the j-th row of the matrix A', i.e., is equal to the general (the i, j-th) element of the matrix B'A'.

In conclusion we shall remark that the determinant of a product of (square) matrices is equal to the product of the determinants of the multiplied matrices: $|AB| = |A| \; |B|$, which result is taken from determinant theory.

4. The partitioning of matrices. The handling of matrices of high orders requires, as a rule, a large number of operations. It is therefore often expeditious to reduce a computation involving matrices of high orders to computations upon matrices of lower orders. Such a reduction can be effected by partitioning the given matrices: each matrix may be conceived as composed of several matrices of lower orders, and this subdivision may be carried through in many ways, for example:

$$(11a)\quad \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{pmatrix} = \left(\begin{array}{c:ccc} a_{11} & a_{12} & a_{13} & a_{14} \\ \hdashline a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{array}\right) = \left(\begin{array}{cc:cc} a_{11} & a_{12} & a_{13} & a_{14} \\ \hdashline a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{array}\right).$$

The matrices into which the given matrix is partitioned are called its submatrices, or cells. In such a partition the horizontal and vertical subdividing lines are of course supposed to be carried across the whole matrix.

We shall not concern ourselves with the general case of the partitioning of a matrix[1], but shall here consider only a partition of square matrices in which the diagonal submatrices are square.

The basic operations on partitioned matrices whose diagonal matrices are of identical orders is connected in a quite natural way with operations upon the submatrices themselves. To wit, if we have

---

[1]See, e.g., [1], §11.

(11b)
$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1k} \\ A_{21} & A_{22} & \cdots & A_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ A_{k1} & A_{k2} & \cdots & A_{kk} \end{pmatrix}$$

and

(11c)
$$B = \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1k} \\ B_{21} & B_{22} & \cdots & B_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ B_{k1} & B_{k2} & \cdots & B_{kk} \end{pmatrix},$$

where A_{ii} and B_{ii} are square matrices of the same order, then

(12)
$$A + B = \begin{pmatrix} A_{11} + B_{11} & A_{12} + B_{12} & \cdots & A_{1k} + B_{1k} \\ A_{21} + B_{21} & A_{22} + B_{22} & \cdots & A_{2k} + B_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ A_{k1} + B_{k1} & A_{k2} + B_{k2} & \cdots & A_{kk} + B_{kk} \end{pmatrix}$$

and

(13)
$$AB = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1k} \\ C_{21} & C_{22} & \cdots & C_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ C_{k1} & C_{k2} & \cdots & C_{kk} \end{pmatrix},$$

where

$$(13a)\quad C_{ij} = A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{ik}B_{kj} \qquad i, j = 1, \cdots , k \quad .$$

We shall not stop for a proof of the last formula, but will note here only that the matrices A_{i1} and B_{1j} can indeed be multiplied, since the number of columns of matrix A_{i1} equals the number of rows of the matrix B_{1j}.

Formulas (12) and (13) show that operations with matrices partitioned in the manner indicated are to be conducted just as if in place of each submatrix there was a number.

An important special case of a partitioned matrix is the <u>bordered</u> matrix. Having a square matrix A_{n-1} of order n - 1:

$$(13b)\qquad A_{n-1} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} \\ a_{21} & a_{22} & \cdots & a_{2,n-1} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n-1,1} & a_{n-1,2} & \cdots & a_{n-1,n-1} \end{pmatrix} .$$

we form a square matrix of the n-th order, A_n, by appending to the matrix A_{n-1} a row: $v_{n-1} = (a_{n1}, \cdots , a_{n,n-1})$, a column: $u_{n-1} = (a_{1n}, \cdots , a_{n-1,n})$, and a number a_{nn}:

$$(14)\qquad A_n = \begin{pmatrix} & & & a_{1n} \\ & A_{n-1} & & a_{2n} \\ & & & \vdots \\ & & & a_{n-1,n} \\ a_{n1} & \cdots & a_{n,n-1} & a_{nn} \end{pmatrix} = \begin{pmatrix} A_{n-1} & u_{n-1} \\ v_{n-1} & a_{nn} \end{pmatrix} .$$

We shall say that the matrix A_n has been obtained by <u>bordering</u> the matrix A_{n-1}. The matrix A_n is naturally partitionable.

Operations upon a bordered matrix are conducted in accordance with the general rules for operations upon partitioned matrices. Letting

$$A = \begin{pmatrix} M & u \\ v & a \end{pmatrix}, \quad B = \begin{pmatrix} P & y \\ x & b \end{pmatrix} \tag{14a}$$

be two bordered matrices of order n, the meaning of m, v, u. a, and P, x, y, b, being those of the definition, the following statements are valid:

$$\left\{ \begin{aligned} \alpha A &= \begin{pmatrix} \alpha M & \alpha u \\ \alpha v & \alpha a \end{pmatrix}, \\ A + B &= \begin{pmatrix} M + P & u + y \\ v + x & a + b \end{pmatrix}, \\ AB &= \begin{pmatrix} MP + ux & My + ub \\ vP + ax & vy + ab \end{pmatrix}. \end{aligned} \right. \tag{15}$$

Here MP and ux are matrices of the (n - 1)-th order; My, ub are columns composed of n - 1 elements; vP and ax are analogous rows; and, lastly, vy + ab is a number.

5. Quasi-diagonal matrices. Let us consider still another particular case of partitioned matrices, namely the matrices called quasi-diagonal. These are square matrices along whose leading diagonals are arrayed square submatrices, the remainder of the elements being zero. An example would be the seventh-order quasi-diagonal matrix:

$$(15a)\qquad \left(\begin{array}{cc|ccc|cc} a_{11} & a_{12} & 0 & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & b_{11} & b_{12} & b_{13} & 0 & 0 \\ 0 & 0 & b_{21} & b_{22} & b_{23} & 0 & 0 \\ 0 & 0 & b_{31} & b_{32} & b_{33} & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & c_{11} & c_{12} \\ 0 & 0 & 0 & 0 & 0 & c_{21} & c_{22} \end{array}\right)$$

The cells of this matrix are obviously

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} ;$$

$$(15b)\qquad B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} ; \qquad C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$$

and the six null submatrices.

If two quasi-diagonal matrices are of like structure, the product of such matrices will also be a quasi-diagonal matrix of the same structure, the diagonal cells of which equal the products of the corresponding cells of the factor submatrices.

The determinant of a quasi-diagonal matrix is equal to the product of the determinants of the diagonal cells, on the strength of a notable theorem by Laplace[1].

6. The inverse and the adjoint matrices. A square matrix $A = (a_{ij})$ is said to be non-singular if its determinant is not equal to zero; in the contrary case it is of course singular.

The important concept of an inverse matrix is now introduced. A matrix B is called the inverse (or reciprocal) of the matrix A if

$$AB = I \quad . \tag{16}$$

We shall show that the necessary and sufficient condition for the existence of the inverse matrix is the non-singularity of the matrix A.

The necessity follows at once from the theorem concerning the determinant of a matrix product, for if $AB = I$, $|A||B| = 1$ and consequently $|A| \neq 0$.

Assume now that $|A| \neq 0$. In order to construct the inverse matrix we must give preliminary consideration to the adjoint (or adjugate) matrix, i.e., the matrix

---

[1]See, e.g. [1], §33.

(17) $$C = \begin{pmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \cdot & \cdot & \cdot & \cdot \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{pmatrix} .$$

Here A_{ij} is the algebraic complement (cofactor) of the element a_{ij} in the determinant of the matrix A, i.e., is the signed minor determinant of the element a_{ij}.

We shall show that the adjoint matrix has the following property:

(18) $$AC = |A|I .$$

In demonstration, reckoning the general element of the matrix AC by the rules for matrix multiplication, we find it to equal

(18a) $$a_{i1}A_{j1} + a_{i2}A_{j2} + \dots + a_{in}A_{jn} ,$$

i.e., zero for $i \neq j$, and $|A|$ for $i = j$, on the strength of a familiar theorem on the expansion of determinants[1].

The equality

(18') $$CA = |A|I$$

is established in like manner.

The adjoint matrix has meaning for any square matrix A. From the equality $AC = |A|I$ it follows that the matrix

---

[1]Translator's note: An expansion in terms of "alien cofactors" vanishes identically. See, e.g., [1], §19, 21.

(19) $$B = \frac{1}{|A|} C$$

is, for non-singular A, the sought inverse, for

(19a) $$AB = A \frac{1}{|A|} C = \frac{1}{|A|} AC = I \quad .$$

The constructed matrix has also the property

(20) $$BA = I \quad ,$$

which follows from equation (18').

We prove, lastly, the uniqueness of the inverse matrix. Assume that a matrix X exists such that $AX = I$. Multiplying this equation by B on the left, we have $X = B$. If it be assumed that $YA = I$, a multiplication on the right by B yields $Y = B$.

The matrix inverse to A is denoted by A^{-1}. It is obvious that $|A^{-1}| = |A|^{-1}$.

We note that the inverse of the product of two matrices also displays the reversal rule:

(21) $$(A_1A_2)^{-1} = A_2^{-1}A_1^{-1} \quad ,$$

since

(21a) $$A_1A_2A_2^{-1}A_1^{-1} = A_1A_1^{-1} = I \quad .$$

The determination of the inverse matrix is one of the fundamental problems of linear algebra. Equation (19) offers the possibility of computing the inverse matrix; however, the computation of the adjoint matrix is so labor-consuming that the cited equation is of importance only in theoretical relationships. Chapter II will be specially devoted to this problem of determining the inverse matrix.

7. Polynomials in a matrix. We now define the positive integral power of a square matrix, putting

$$\overbrace{A \cdot A \cdot \quad \cdots \quad \cdot A}^{n \text{ times}} = A^n \quad . \tag{22}$$

In view of the associative law, how the parentheses in this product are placed makes no difference, and we therefore omit them. It is evident from the definition that

$$\begin{cases} A^n A^m = A^{n+m} \\ (A^n)^m = A^{nm} \end{cases} \quad . \tag{23}$$

Hence it follows that powers of the same matrix are commutative.

We further put, by definition,

$$A^0 = I \quad , \tag{23a}$$

An expression of the form

$$\alpha_0 A^n + \alpha_1 A^{n-1} + \cdots + \alpha_n I \quad , \tag{23b}$$

where $\alpha_0, \alpha_1, \cdots, \alpha_n$ are complex numbers, is called a polynomial in a matrix, or matrix polynomial. This matrix polynomial may be regarded as the result of replacing the variable λ in an algebraic polynomial

$$\varphi(\lambda) = \alpha_0 \lambda^n + \alpha_1 \lambda^{n-1} + \cdots + \alpha_n \tag{24}$$

by the matrix A.

It is important to note that the rules for operation upon matrix polynomials do not differ from the rules for operation upon algebraic polynomials, viz.:

$$\left\{ \begin{array}{l} \text{given} \\ \qquad \varphi(\lambda) = \Psi(\lambda) \pm \chi(\lambda) \\ \qquad \omega(\lambda) = \Psi(\lambda)\chi(\lambda) \quad , \\ \\ \text{then} \\ \qquad \varphi(A) = \Psi(A) \pm \chi(A) \\ \qquad \omega(A) = \Psi(A)\chi(A) \quad . \end{array} \right. \tag{25}$$

This follows from the commutativity of the powers of a matrix.

8. The characteristic polynomial. The Cayley-Hamilton theorem. The minimum polynomial. The equation

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} = 0 \tag{26}$$

is called the characteristic equation of the matrix $A = (a_{ij})$. The left member of this equation, which may be written in the abbreviated form $|A - \lambda I|$, bears the name characteristic polynomial (or characteristic function) of the matrix. Characteristic equations are frequently encountered in applied mathematics[1].

The direct computation of the characteristic function presents considerable technical difficulties. If

$$\varphi(\lambda) = |A - \lambda I| = (-1)^n[\lambda^n - p_1\lambda^{n-1} - p_2\lambda^{n-2} - \cdots - p_n] \quad , \tag{27}$$

then

$$\begin{aligned} p_1 &= a_{11} + a_{22} + \cdots + a_{nn} \\ p_n &= (-1)^{n-1}\,|A| \quad ; \end{aligned} \tag{28}$$

and the remaining coefficients p_k are the sums, taken with the sign $(-1)^{k-1}$, of all the principal minors of the determinant of matrix A of order k, i.e., of the minors involving the principal diagonal[2]. The number of such minors equals the number of combinations of n things taken k at a time.

The roots of the characteristic equation are called the latent roots (characteristic numbers, proper values, eigenvalues) of the matrix A. From the well-known theorem of Vietà giving the connection between the roots of an equation and its coefficients, we have

---

[1]See, e.g., [2], §3.6.

[2]See, e.g., [1], §37.

$$(29)\quad \begin{cases} \lambda_1 + \lambda_2 + \cdots + \lambda_n = p_1 = a_{11} + a_{22} + \cdots + a_{nn} \\ \lambda_1\lambda_2 \cdots \lambda_n = (-1)^{n-1} p_n = |A| \ . \end{cases}$$

The quantity $p_1 = a_{11} + a_{22} + \cdots + a_{nn}$ is called the trace (or spur) of the matrix A, and is denoted by tr A.

Practically convenient methods for determining the coefficients and roots of the characteristic equation will be elaborated in Chapter III, which will be specially devoted to that group of questions. For the moment we leave them aside.

For any square matrix the following remarkable relation, known as the Cayley-Hamilton Theorem, obtains: if $\varphi(\lambda)$ is the characteristic polynomial of the matrix A, then $\varphi(A) = 0$, that is, speaking somewhat conditionally, the matrix is a root of its own characteristic equation.

For proof, let us consider the matrix B, the adjoint of the matrix $A - \lambda I$. Since each cofactor in the determinant $|A - \lambda I|$ is a polynomial in λ of degree not exceeding n - 1, the adjoint matrix may be represented as an algebraic polynomial with matrix coefficients[1], i.e., in the form

$$(29a)\qquad B = B_{n-1} + B_{n-2}\lambda + \cdots + B_o\lambda^{n-1} ,$$

where $B_{n-1}, \cdots, B_o$ are certain matrices not dependent on λ. On the strength of the fundamental property of the adjoint matrix, we have

$$(29b)\qquad \begin{aligned} (B_{n-1} + B_{n-2}\lambda + \cdots + B_o\lambda^{n-1})(A - \lambda I) &= |A - \lambda I| I = \\ &= (-1)^n(\lambda^n - p_1\lambda^{n-1} - \cdots - p_n)I \ . \end{aligned}$$

---

[1]Guard against confusion with a polynomial in a matrix. See, e.g., [3], Chap. III, § 4, p. 21 ff.

This equation is equivalent to the system of equations

$$
(29a) \quad \begin{cases} B_{n-1}A & = (-1)^{n+1} p_n I \\ B_{n-2}A - B_{n-1} & = (-1)^{n+1} p_{n-1} I \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ B_0 A - B_1 & = (-1)^{n+1} p_1 I \\ -B_0 A & = (-1)^n I \quad . \end{cases}
$$

Multiplying these equations on the right by I, A, A^2, ... , A^{n-1}, A^n, respectively, and adding, we obtain a null matrix on the left side, and on the right,

$$
(30) \qquad (-1)^n[-p_n I - p_{n-1}A - p_{n-2}A^2 - \cdots + A^n] = \varphi(A) \quad .
$$

Thus $\varphi(A) = 0$, which is what was required to be proved.

The Cayley-Hamilton relation shows that for a given square matrix a polynomial exists for which it is a root. Evidently such a polynomial is not unique, for if $\psi(\lambda)$ has such a property, so has any polynomial divisible by $\psi(\lambda)$. The polynomial of lowest degree having this property that the matrix A is a root of it, is called the minimum polynomial of the matrix A.

We shall prove that the characteristic polynomial is divisible by the minimum polynomial.

Let $q(\lambda)$ and $r(\lambda)$ be the quotient and remainder obtained upon dividing the characteristic polynomial $\varphi(\lambda)$ by the minimum polynomial $\psi(\lambda)$:

(30a) $$\varphi(\lambda) = \psi(\lambda)q(\lambda) + r(\lambda) \quad ,$$

the degree of $r(\lambda)$ being of course less than the degree of $\psi(\lambda)$.

Substituting A for λ in this equation, we have

$$r(A) = \varphi(A) - \psi(A)q(A) = 0 \quad .$$

Thus the matrix A proves to be a "root" of the polynomial $r(\lambda)$; it thence follows that $r(\lambda) \equiv 0$, since otherwise $\psi(\lambda)$ would not be the minimum function. Consequently $\psi(\lambda)$ divides $\varphi(\lambda)$.

9. Similar matrices. The matrix B is said to be similar to the matrix A if a non-singular matrix C exists such that $B = C^{-1}AC$. Matrix B is said to be obtained from matrix A by a similarity (or collineatory) transformation.

The similarity transformation has the following properties:

(30b) 1. $$C^{-1}A_1C + C^{-1}A_2C + \cdots + C^{-1}A_nC = C^{-1}(A_1 + A_2 + \cdots + A_n)C \quad .$$

(31) 2. $$C^{-1}A_1C \cdot C^{-1}A_2C \cdots C^{-1}A_nC = C^{-1}(A_1A_2 \cdots A_n)C \quad .$$

In particular, $(C^{-1}AC)^n = C^{-1}A^nC$.

Hence:

3. $f(C^{-1}AC) = C^{-1}f(A)C$ for any polynomial $f(\lambda)$.

From the last property it follows directly that similar matrices have the same minimum function.

We shall show that <u>similar matrices have also the same characteristic function</u>.

We have

$$|B - \lambda I| = |C^{-1}AC - \lambda I| = |C^{-1}AC - \lambda C^{-1}IC| =$$

(31a)

$$= |C|^{-1}|A - \lambda I||C| = |A - \lambda I| \quad .$$

10. <u>Elementary transformations</u>. It is frequently necessary to effect the following operations upon matrices:

a) Multiplication of the elements of some row by a number;

b') Adding to the elements of some row numbers proportional to the elements of some preceding row.

b") Adding to the elements of some row numbers proportional to the elements of some following row.

Sometimes such transformations must be made upon the columns. Transformations of the type indicated are called <u>elementary transformations</u> of the matrix.

Any elementary transformation of the <u>rows</u> is equivalent to a <u>pre</u>multiplication of the matrix by a non-singular matrix of a special form, as will readily be verified. Operation a) is equivalent to a premultiplication of the matrix by the matrix

(32)
$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \alpha & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix};$$

operation b') is equivalent to a premultiplication by the matrix

(33)
$$\begin{pmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ & & \vdots & \ddots & & & & \\ & & \vdots & & \ddots & & & \\ & & \alpha & \cdots & \cdots & 1 & & \\ & & & & & & \ddots & \\ & & & & & & & 1 \end{pmatrix};$$

operation b") is equivalent to a premultiplication by the matrix

(34)
$$\begin{pmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & \cdots & \cdots & \alpha & & \\ & & & \ddots & & \vdots & & \\ & & & & \ddots & \vdots & & \\ & & & & & 1 & & \\ & & & & & & \ddots & \\ & & & & & & & 1 \end{pmatrix}.$$

Operations a), b') and b") are performed on columns by using just such elementary matrices[1] in postmultiplications of the matrix undergoing transformation, (33) now effecting b") upon columns, and (34) effecting b').

Examples of row operations:

$$(34a)\quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a & b & c \\ x & y & z \\ u & v & w \end{pmatrix} = \begin{pmatrix} a & b & c \\ \alpha x & \alpha y & \alpha z \\ u & v & w \end{pmatrix} ;$$

$$(34b)\quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \alpha & 1 \end{pmatrix} \begin{pmatrix} a & b & c \\ x & y & z \\ u & v & w \end{pmatrix} = \begin{pmatrix} a & b & c \\ x & y & z \\ u + \alpha x & v + \alpha y & w + \alpha z \end{pmatrix} ;$$

$$(34c)\quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \alpha \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a & b & c \\ x & y & z \\ u & v & w \end{pmatrix} = \begin{pmatrix} a & b & c \\ x + \alpha u & y + \alpha v & z + \alpha w \\ u & v & w \end{pmatrix} .$$

---

[1]Translator's note: Note that the elementary matrices may well be regarded as having been derived from the unit matrix by just such transformations, a), b') and b"), as it is proposed to make upon the rows/columns of A. Thus, for example, the elementary matrix of (34b) effects, by premultiplication, the following transformation of A: row 3 + α row 2, and is itself the result of such an operation upon the unit matrix. By postmultiplication it effects: column 2 + α column 3, and is this transformation of I.

In future work we will often have to conduct transformations of types a) and b') upon matrices. The result of several row-transformations of this type is equivalent to the premultiplication of the matrix by some <u>triangular</u> matrix, i.e., one of the form

$$(35)\qquad \begin{pmatrix} \gamma_{11} & 0 & \cdots & 0 \\ \gamma_{21} & \gamma_{22} & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn} \end{pmatrix},$$

with non-zero diagonal elements γ_{ii}. Indeed, each separate transformation of form a) or b') is equivalent to a premultiplication by a triangular matrix of the type indicated, and the product of two or more triangular matrices of like structure (i.e., both, say, <u>lower triangular</u>, as these) is again a like triangular matrix.

It should be further noted that the result of several column-transformations of the form b') and b"), such that to each, j-th, column is added a multiple, m_{ij}, of the elements of the i-th column (which itself remains unchanged), is equivalent to postmultiplication of the given matrix by a matrix of the form:

$$(36)\qquad \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ m_{i1} & m_{i2} \cdots m_{i,i-1} & 1 & m_{i,i+1} & \cdots & m_{in} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & \cdots & \cdots & 1 \end{pmatrix}.$$

11. <u>Decomposition of matrices into the product of two triangular matrices.</u> Triangular matrices, that is, matrices of the form

$$(37)\quad \begin{pmatrix} c_{11} & 0 & \cdots & 0 \\ c_{21} & c_{22} & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix} \text{ and } \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ 0 & b_{22} & \cdots & b_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & b_{nn} \end{pmatrix}, \quad \begin{matrix} (\text{no } c_{ii} = 0\ , \\ \text{no } b_{ii} = 0) \end{matrix}$$

have a number of convenient properties. For instance, the determinant of a triangular matrix equals the product of the elements of the principal diagonal; the product of two triangular matrices of like structure is again a triangular matrix of the same structure; a non-singular triangular matrix is easily inverted and its inverse is of like structure, etc.

The following theorems are therefore of interest.

THEOREM. <u>On condition that the leading submatrices of the matrix</u>

$$(37a)\qquad A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

<u>are non-singular, i.e., that</u>

$$(37b)\qquad a_{11} \neq 0\ , \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0, \ \cdots\ , \ |A| \neq 0\ ,$$

A may be represented as the product of a lower triangular matrix and an upper triangular matrix.

The proof will be carried through by the method of mathematical induction.

For n = 1, the statement is obvious: $(a_{11}) = (b_{11})(c_{11})$, and one of the factors may be taken arbitrarily. Let the theorem be true for a matrix of the (n - 1)-th order. We shall show it to be true for a matrix of the n-th order.

Partition the matrix A into a bordered matrix:

(37c) $$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot \quad \cdot \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} & & & a_{1n} \\ & A_{n-1} & & \vdots \\ & & & a_{n-1,n} \\ a_{n1} & a_{n2} & \cdots \; a_{n,n-1} & a_{nn} \end{pmatrix} = \begin{pmatrix} A_{n-1} & u \\ v & a_{nn} \end{pmatrix}$$

we shall seek a decomposition A = CB of the matrix A into the product of two matrices B and C of the required forms, first having partitioned these matrices into bordered form like that of A:

(37d) $$C = \begin{pmatrix} C_{n-1} & 0 \\ x & c_{nn} \end{pmatrix}, \quad B = \begin{pmatrix} B_{n-1} & y \\ 0 & b_{nn} \end{pmatrix}.$$

By the rule for multiplication of partitioned matrices:

$$CB = \begin{pmatrix} C_{n-1} & 0 \\ x & c_{nn} \end{pmatrix} \begin{pmatrix} B_{n-1} & y \\ 0 & b_{nn} \end{pmatrix} =$$

(37e)

$$= \begin{pmatrix} C_{n-1}B_{n-1} & C_{n-1}y \\ xB_{n-1} & xy + c_{nn}b_{nn} \end{pmatrix} = A \;;$$

whence we have

(37f) $$C_{n-1}B_{n-1} = A_{n-1} \quad .$$

Now such triangular matrices, C_{n-1} and B_{n-1}, exist, by the induction hypothesis. Furthermore, from the assumption that $|A_{n-1}| \neq 0$, it follows that $|C_{n-1}| \neq 0$ and $|B_{n-1}| \neq 0$.

Now x and y are found by the formulas

(37g) $$y = C_{n-1}^{-1}u \quad , \quad x = vB_{n-1}^{-1} \quad ,$$

wherewith they are determined uniquely in terms of u and v.

Thus it only remains for us to determine the diagonal elements c_{nn} and b_{nn} from the equation

(37h) $$c_{nn}b_{nn} = a_{nn} - xy \quad .$$

The last equation shows that one of the diagonal elements may be taken arbitrarily.

Thus the decomposition of a matrix into the product of two triangular matrices, lower and upper, will be unique only if we prescribe values for the diagonal elements of one of the triangular matrices.

It is convenient to consider, for example, that $b_{ii} = 1$, $i = 1, \cdots, n$. Then

(37i) $$c_{nn} = a_{nn} - xy$$

and accordingly matrix C will be uniquely determined.

12. Matrix notation for a system of linear equations. Let us consider the system of n linear equations in n unknowns:

(38) $$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n \end{cases}.$$

Utilizing the definition of matrix multiplication, the system may be written as a single equation in matrix notation:

(38') $$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

or simply as

(38") $$Ax = b \quad ,$$

A signifying the matrix of the coefficients of the system, b the column of free members, and x the column whose elements are the unknowns.

If the matrix of the system, A, is non-singular, we obtain at once the solution of system (38) by premultiplying (38") by A^{-1}:

(39) $$x = A^{-1}b = \frac{1}{|A|} Bb \quad ,$$

B being the adjoint matrix of A.

We shall show that the last formula is the matrix notation for the familiar Cramer's Rule:

(40) $$x_i = \frac{|A_i|}{|A|} \quad ,$$

where A_i is the matrix that is obtained from A by replacing the elements a_{ki} of the i-th column by the components b_i of b.

Indeed, the matrix equation (39) is equivalent to the n equations

(40a) $$x_i = \frac{A_{1i}b_1 + A_{2i}b_2 + \cdots + A_{ni}b_n}{|A|} \quad i = 1, \cdots, n \quad .$$

Since the A_{ki} are the cofactors of the element a_{ki} in the determinant of the matrix A, we obviously have

(40b) $$A_{1i}b_1 + A_{2i}b_2 + \cdots + A_{ni}b_n = |A_i| \quad ,$$

which proves our statement.

2. n-DIMENSIONAL VECTOR SPACE

In what is to follow an important role will be played by the so-called n-dimensional vector space R_n. A point X of such a space is an aggregate of n numbers, as a rule complex, arrayed in a definite order:

$$(1) \qquad X = (x_1, x_2, \cdots, x_n) \quad .$$

X is also called an n-dimensional vector. The numbers $x_1, x_2, \cdots, x_n$ are called the components of the vector. The number n is called the dimension of the space.[1]

Two vectors are said to be equal only if their corresponding components are equal. Fundamental operations on vectors are defined as follows: if $X = (x_1, x_2, \cdots, x_n)$ and $Y = (y_1, y_2, \cdots, y_n)$ are two n-dimensional vectors and a is an arbitrary complex number, we then put, by definition,

$$(2) \qquad \begin{cases} X + Y = (x_1 + y_1, x_2 + y_2, \cdots, x_n + y_n) \\ aX = (ax_1, ax_2, \cdots, ax_n) \quad . \end{cases}$$

The addition of vectors satisfies the commutative and associative laws:

---

[1] Translator's note: These are extensions, by a natural generalization, of the familiar vectors and components of vector analysis, where, relative to the basic vectors i, j, k, a vector $X = a_1 i + a_2 j + a_3 k$ has the components a_1, a_2, a_3. (See, e.g., [4], §5.) Here, abstracted from the relation with basic vectors and with more than three "components", and thus in a "space" of more than three "dimensions", the "vector" terminology is preserved as appropriate to the forms, operations, etc.

$$X + Y = Y + X$$

(2a)

$$(X + Y) + Z = X + (Y + Z) \quad .$$

The addition of vectors is connected with multiplication by numbers by the distributive laws

$$a(X + Y) = aX + aY$$

(3)

$$(a + b)X = aX + bY \quad .$$

The validity of all these laws follows directly from the definition of the operations.

For vectors of an n-dimensional space a scalar product[1] is introduced in accordance with the formula

(4)
$$(X, Y) = \sum_{k=1}^{n} x_k \bar{y}_k \quad ,$$

where $\bar{y}_k$ designates the complex conjugate[2] of y_k.

It is readily verified that the scalar product has the following properties:

---

[1] Cf. the scalar (dot) product of ordinary vectors, for which there is a convenient matrix notation: see [1], §8.9.

[2] See [5], p. 162, note.

1) $(X,X) > 0$ if $X \neq 0$; $(X,X) = 0$ if $X = 0$.

2) $(X,Y) = \overline{(Y,X)}$.

3) $(X_1 + X_2, Y) = (X_1,Y) + (X_2,Y)$.

4) $(aX,Y) = a(X,Y)$.

5) $(X,Y_1 + Y_2) = (X,Y_1) + (X,Y_2)$.

6) $(X,aY) = \bar{a}(X,Y)$.

In addition, $\sqrt{(X,X)}$ is called the length of the vector. In what follows we shall designate it by $||X||$.

Besides the n-dimensional complex space introduced above, it is useful to consider also an n-dimensional real space, i.e., the aggregate of vectors with real components.

In real space the scalar product is equal to the sum of the products of corresponding components of the vectors; the length of a vector equals the square root of the sum of the squares of its components.

We shall most often have to deal with real n-dimensional space, turning to complex space only as occasion requires.

1. Linear dependence. A vector $Y = c_1X_1 + c_2X_2 + \cdots + c_mX_m$ is said to be a linear combination of the vectors X_1, X_2, $\cdots$ X_m.

It is easily seen that if vectors $Y_1, \cdots, Y_k$ are linear combinations of the vectors $X_1, \cdots, X_m$, any linear combination $\gamma_1Y_1 + \cdots + \gamma_kY_k$ will also be a linear combination of the vectors $X_1, \cdots, X_m$.

Vectors $X_1, X_2, \cdots, X_m$ are called linearly dependent if constants $c_1, c_2, \cdots, c_m$ exist, not all zero, such that the equation

(5) $$c_1X_1 + c_2X_2 + \cdots + c_mX_m = 0$$

holds. If, however, this equation holds only when all the constants c_i are equal to zero, the vectors $X_1, X_2, \cdots, X_m$ are said to be linearly independent.

If the vectors $X_1, \cdots, X_m$ are linearly dependent, then at least one of them will be a linear combination of the rest. For if, for example, $c_m \neq 0$, we find from (5):

(6) $$X_m = -\frac{c_1}{c_m} X_1 - \cdots - \frac{c_{m-1}}{c_m} X_{m-1} \quad .$$

THEOREM 1. If the vectors $Y_1, \cdots, Y_k$ are linear combinations of the vectors $X_1, \cdots, X_m$, and $k > m$, the former set is linearly dependent.

The proof will be carried through by the method of mathematical induction. For $m = 1$, the theorem is obvious. Let the theorem be true on the assumption that the number of combined vectors be $m - 1$. Under the condition of the theorem, then,

(6a) $$\begin{aligned} Y_1 &= c_{11}X_1 + \cdots + c_{1m}X_m \\ &\cdots\cdots\cdots \\ Y_k &= c_{k1}X_1 + \cdots + c_{km}X_m \quad . \end{aligned}$$

Two cases are conceivable.

1. All the coefficients $c_{11}, \cdots, c_{k1}$ are equal to zero. Then $Y_1, \cdots, Y_k$ are in fact linear combinations of the $m - 1$ vectors

$X_2, \cdots, X_m$. On the strength of the induction hypothesis, $Y_1, \cdots, Y_k$ will be linearly dependent.

2. At least one coefficient of X_1 will be different from zero. Without violating the generality we may consider that $c_{11} \neq 0$.

Let us now consider the system of vectors

(6b)
$$
\begin{aligned}
Y_2' &= Y_2 - \frac{c_{21}}{c_{11}} Y_1 \\
&\cdots\cdots \\
Y_k' &= Y_k - \frac{c_{k1}}{c_{11}} Y_1 \quad .
\end{aligned}
$$

The vectors thus constructed are obviously linear combinations of the vectors $X_2, \cdots, X_m$, and the number of them is $k - 1 > m - 1$. On the strength of the induction hypothesis they are linearly dependent, i.e., constants $\gamma_2, \cdots, \gamma_k$ that are not simultaneously zero can be found such that

(6c)
$$\gamma_2 Y_2' + \cdots + \gamma_k Y_k' = 0 \quad .$$

Replacing $Y_2', \cdots, Y_k'$ by their expressions in terms of $Y_1, \cdots, Y_k$, we obtain

(6d)
$$\gamma_1 Y_1 + \gamma_2 Y_2 + \cdots + \gamma_k Y_k = 0 \quad ,$$

where $\gamma_1 = -\frac{c_{21}}{c_{11}} \gamma_2 - \cdots - \frac{c_{k1}}{c_{11}} \gamma_k$. The numbers $\gamma_1, \cdots, \gamma_k$ are not simultaneously equal to zero and accordingly $Y_1, \cdots, Y_k$ are linearly dependent. This proves Theorem 1.

A system of linearly independent vectors is said to constitute a basis for a space if any vector of the space is a linear combination of the vectors of the system.

An example of a basis is the set of vectors

(7)
$$\begin{cases} e_1 = (1, 0, \cdots, 0) \\ e_2 = (0, 1, \cdots, 0) \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ e_n = (0, 0, \cdots, 1) \end{cases},$$

for it is obvious that for any vector $X = (x_1, x_2, \cdots, x_n)$ we have

(7a)
$$X = x_1e_1 + x_2e_2 + \cdots + x_ne_n \quad .$$

This we shall call the initial basis of the space. Such a basis is not the only one possible -- quite the contrary: in the choice of a basis one may be arbitrary within wide limits. Despite this, the number of vectors forming a basis does not depend on its selection. In proof of this, let $Y_1, \cdots, Y_k$ and $Z_1, \cdots Z_m$ be two bases, and assume further that $k > m$. The vectors $Y_1, \cdots, Y_k$ are linear combinations of the vectors $Z_1, \cdots, Z_m$. In the light of Theorem 1, $Y_1, \cdots, Y_k$ are linearly dependent, which contradicts the definition of basis. So $k = m$. Furthermore, since the initial basis is constituted by n vectors, any other basis will also consist of n vectors. The number of vectors forming a basis thus coincides with the dimension of the space.

Let $U_1, \cdots, U_n$ form the basis of a space. Any vector X will then be a linear combination of $U_1, \cdots U_n$:

(8) $$X = \xi_1 U_1 + \xi_2 U_2 + \cdots + \xi_n U_n \quad .$$

The coefficients of this resolution uniquely define the vector X, for if

(8a) $$X = \xi_1 U_1 + \cdots + \xi_n U_n = \xi_1' U_1 + \cdots + \xi_n' U_n \quad ,$$

then $(\xi_1 - \xi_1')U_1 + \cdots + (\xi_n - \xi_n')U_n = 0$, and accordingly $\xi_1 - \xi_1' = 0, \cdots, \xi_n - \xi_n' = 0$, in view of the linear independence of the vectors $U_1, \cdots, U_n$.

The coefficients $\xi_1, \cdots, \xi_n$ are called the coordinates of the vector X with respect to the basis $U_1, \cdots, U_n$. Note that the components of a vector $x_1, \cdots, x_n$ are the coordinates of the vector X with respect to the initial basis.

2. Orthogonal systems of vectors. Two non-zero vectors of a space are said to be orthogonal if their scalar product equals zero. A system of vectors is said to be orthogonal if any two vectors of the system are orthogonal to one another. In speaking of an orthogonal system, we shall henceforth assume that all the vectors of this system are different from zero.

THEOREM 2. The vectors forming an orthogonal system are linearly independent.

Proof. Let $X_1, \cdots, X_k$ be an orthogonal system, and let

(8b) $$c_1 X_1 + c_2 X_2 + \cdots + c_k X_k = 0 \quad .$$

In view of the properties of the scalar product we have:

$$(8c)\qquad 0 = (c_1X_1 + \cdots + c_kX_k, X_i) = c_1(X_1,X_i) + \\ + \cdots + c_i(X_i,X_i) + \cdots + c_k(X_k,X_i) = c_i||X_i||^2 ,$$

and, since $||X_i||^2 > 0$, $c_i = 0$ for any $i = 1, 2, \cdots, n$. Thus the sole possible values for $c_1, c_2, \cdots, c_n$ in the equation $c_1X_1 + c_2X_2 + \cdots + c_nX_n = 0$ are $c_1 = c_2 = \cdots = c_n = 0$, i.e., the vectors $X_1, X_2, \cdots, X_n$ are linearly independent. It thence follows, first, that the number of vectors forming an orthogonal system does not exceed n, and, second, that any orthogonal system of n vectors forms a basis of the space. Such a basis is called orthogonal. If we have, in addition, $||X_i|| = 1$, the basis is said to be orthonormal. An example of an orthonormal basis is the initial basis.

From any system of linearly independent vectors $X_1, \cdots, X_k$, it is possible to go over to an orthogonal system of vectors $X_1', \cdots X_k'$ by means of the process spoken of as orthogonalization. The following theorem describes this process.

THEOREM 3. Let $X_1, \cdots, X_n$ be linearly independent. An orthogonal system of vectors $X_1', \cdots, X_n'$ may be constructed that is connected with the original set by the relations:

$$(9)\qquad \begin{cases} X_1' = X_1 \\ X_2' = X_2 + \alpha_{21}X_1 \\ \cdots\cdots\cdots\cdots\cdots \\ X_k' = X_k + \alpha_{k1}X_1 + \cdots + \alpha_{k,k-1}X_{k-1} \end{cases} .$$

The proof will be carried through by induction.

Let $X_1', \dots, X_{m-1}'$ be already constructed and different from zero. We seek X_m' in the form

$$X_m' = X_m + \gamma_1 X_1' + \dots + \gamma_{m-1} X_{m-1}' \quad . \tag{9'}$$

Choose the coefficients $\gamma_1, \dots, \gamma_{m-1}$ so that $(X_m', X_j') = 0$ for $j = 1, \dots, m-1$. This is easily done, for

$$(X_m', X_j') = (X_m, X_j') + \gamma_j (X_j', X_j') \quad . \tag{9'a}$$

Now $(X_j', X_j') \neq 0$, since $X_j' \neq 0$ by the induction hypothesis, and it is accordingly sufficient to take

$$\gamma_j = - \frac{(X_m, X_j')}{(X_j', X_j')} \quad . \tag{9'b}$$

Replacing the $X_1', \dots, X_{m-1}'$ in equation (9') by their expressions in terms of $X_1, \dots, X_{m-1}$, we obtain finally

$$X_m' = X_m + \alpha_{m1} X_1 + \dots + \alpha_{mm-1} X_{m-1} \quad . \tag{9'c}$$

It remains to be proved that $X_m' \neq 0$. But this is obvious, for otherwise the vector X_m would be a linear combination of the vectors $X_1, \dots, X_{m-1}$, which contradicts the condition of the theorem. The basis of the induction exists, since for $m = 1$ the theorem is trivial.

One may pass from any orthogonal system to the corresponding orthonormal system by dividing each vector by its length.

The process described permits of great latitude in the choice of an orthonormal basis, for one may pass from any basis to an orthonormal one by orthogonalization and normalization.

The scalar product of two vectors is very simply expressible in terms of the coordinates of these vectors with respect to any orthonormal basis, for, if $U_1, \dots, U_n$ is an orthonormal basis, and

(9'd) $$X = \xi_1 U_1 + \dots + \xi_n U_n \; ; \qquad Y = \eta_1 U_1 + \dots + \eta_n U_n \; ,$$

then

(9'e) $$\begin{aligned}(X,Y) &= (\xi_1 U_1 + \dots + \xi_n U_n \; , \quad \eta_1 U_1 + \dots + \eta_n U_n) = \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} (\xi_i U_i \; , \; \eta_j U_j) = \sum_{i=1}^{n} \sum_{j=1}^{n} \xi_i \bar{\eta}_j (U_i, U_j) = \sum_{i=1}^{n} \xi_i \bar{\eta}_i \; .\end{aligned}$$

Thus the expression of the scalar product in terms of the coordinates of the vectors with respect to any orthonormal basis coincides with its expression in terms of the components of the vectors, i.e., in terms of the coordinates with respect to the initial basis.

3. Transformation of coordinates. Let us elucidate the change in the coordinates of a vector that accompanies a change of basis.

Let $e_1, e_2, \dots, e_n$ and $e'_1, e'_2, \dots e'_n$ be two bases, and let

(10) $$\begin{cases} e'_1 = a_{11} e_1 + a_{21} e_2 + \dots + a_{n1} e_n \\ e'_2 = a_{12} e_1 + a_{22} e_2 + \dots + a_{n2} e_n \\ \dots\dots\dots\dots\dots\dots \\ e'_n = a_{1n} e_1 + a_{2n} e_2 + \dots + a_{nn} e_n \end{cases} \quad .$$

We connect with the transformation of coordinates a matrix A, the columns of which consist of the coordinates of the vectors e_1', e_2', ..., e_n' with respect to the basis e_1, e_2, ..., e_n, i.e., the matrix

(11) $$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} .$$

The matrix A is non-singular, for it has an inverse, by means of which the vectors e_1, e_2, ..., e_n are expressible in terms of the vectors e_1', e_2', ..., e_n'.

Now designate by x_1, ..., x_n the coordinates of a vector X with respect to the basis e_1, e_2, ..., e_n, and by x_1', x_2', ..., x_n' its coordinates with respect to the basis e_1', e_2', ..., e_n'. Let us determine the relation of dependence between the old and the new coordinates. We have:

(11a)
$$\begin{aligned} X &= x_1e_1 + x_2e_2 + \cdots + x_ne_n = x_1'e_1' + x_2'e_2' + \cdots + x_n'e_n' = \\ &= x_1'(a_{11}e_1 + a_{21}e_2 + \cdots + a_{n1}e_n) + \\ &+ x_2'(a_{12}e_1 + a_{22}e_2 + \cdots + a_{n2}e_n) + \\ &+ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad + \\ &+ x_n'(a_{1n}e_1 + a_{2n}e_2 + \cdots + a_{nn}e_n) = \\ &= (a_{11}x_1' + a_{12}x_2' + \cdots + a_{1n}x_n')e_1 + \\ &+ (a_{21}x_1' + a_{22}x_2' + \cdots + a_{2n}x_n')e_2 + \\ &+ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad + \\ &+ (a_{n1}x_1' + a_{n2}x_2' + \cdots + a_{nn}x_n')e_n \quad , \end{aligned}$$

whence, on the strength of the linear independence of the vectors $e_1, e_2, \cdots, e_n$:

$$(12)\qquad \begin{cases} x_1 = a_{11}x_1' + a_{12}x_2' + \cdots + a_{1n}x_n' \\ x_2 = a_{21}x_1' + a_{22}x_2' + \cdots + a_{2n}x_n' \\ \cdots\cdots\cdots\cdots\cdots \\ x_n = a_{n1}x_1' + a_{n2}x_2' + \cdots + a_{nn}x_n' \end{cases}.$$

The last equations may be written in matrix form. Of course the set of a vector's coordinates can be considered either as a column or as a row. In view of the definition of matrix multiplication, we can postmultiply a square matrix only by a column, not by a row. In future we shall, therefore (except by special stipulation), take the coordinates of a vector as a column. Often in arguments where the basis is fixed (for instance, when the vector is given in terms of its coordinates with respect to the initial basis), we shall identify the vector with the column of its coordinates.

Equation (12) may be written in the form

$$(13)\qquad x = Ax'$$

where

(14)
$$x = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} \quad \text{and} \quad x' = \begin{pmatrix} x_1' \\ x_2' \\ \cdot \\ \cdot \\ \cdot \\ x_n' \end{pmatrix}$$

are the coordinate columns of the vector X with respect to the bases $e_1, \dots, e_n$ and $e_1', \dots, e_n'$ respectively.

This section may be recapitulated in the more succinct matrix notation:

$$A = \left(a_{ij}\right) .$$

$$E' = \begin{pmatrix} e_1' \\ e_2' \\ \cdot \\ \cdot \\ \cdot \\ e_n' \end{pmatrix} ,$$

$$E = \begin{pmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{pmatrix} .$$

where each of these elements is a row-vector and E' is in fact a partitioned matrix. (The "prime" does not here indicate transposition, but merely identifies the basis.)

Whereupon (10) may be written as:

$$(15)\qquad \underset{n\times n}{E'} = \underset{n\times n}{A^T}\ \underset{n\times n}{E}$$

$$X = x_1e_1 + x_2e_2 + \cdots x_ne_n \qquad = x_1'e_1' + x_2'e_2' + \cdots + x_n'e_n'$$

$$= [x_1\ x_2\ \cdots\ x_n]\begin{pmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{pmatrix} \qquad = [x_1'\ x_2'\ \cdots\ x_n']\begin{pmatrix} e_1' \\ e_2' \\ \cdot \\ \cdot \\ \cdot \\ e_n' \end{pmatrix}$$

i.e.,

$$(16)\qquad \underset{1\times n}{X} = \underset{1\times n}{(x^T)}\ \underset{n\times n}{E} \qquad = \underset{1\times n}{(x'^T)}\ \underset{n\times n}{E'}\ ,$$

So $x^TE = x'^TE'$, and, using (15) in (16)

$= x'^TA^TE.$

Postmultiply by E^{-1}, which exists, since the e_i are linearly independent, i.e., $|E| \neq 0$:

$$x^T = x'^TA^T \ ,\ \text{and, by the reversal rule:}$$

$$x = Ax'\ ,\ \text{i.e., (13) holds.}$$

4. Subspaces. A set of vectors $X \subset R_n$ such that any linear combination of the vectors of this set is itself a vector of the same set, is said to be a (linear) subspace of the space R_n. If a group of vectors

$U_1, \cdots, U_m$ that are linearly independent—or even linearly dependent—be given, then the set of all possible linear combinations of them will obviously constitute a subspace. A subspace constructible in this manner is said to be the subspace <u>spanned</u> by the system of vectors $U_1, \cdots, U_m$.

We shall show that a basis exists in every subspace, i.e., that there is a set of linearly independent vectors, by linear combinations of which one may exhaust the entire subspace. Let us construct the basis in the following manner. Take first an arbitrary vector X_1, different from zero, and consider all its linear combinations, i.e., all vectors of the form cX_1. If these exhaust the entire subspace, X_1 then forms its basis. If the contrary is true, a vector X_2 will be found, linearly independent of X_1. Consider the set of linear combinations of X_1 and X_2: if they do not exhaust the subspace, a vector will be found linearly independent of them, and so forth. The process cannot go on interminably, for in the space R_n there cannot be more than n linearly independent vectors. We shall thus have constructed a finite system of vectors $X_1, \cdots, X_k$ such that their linear combinations exhaust our subspace, i.e., we shall have constructed our basis.

It will be remarked that the reasoning set forth indicates much latitude in the choice of a basis. However the <u>number</u> of vectors forming a basis will not depend on the manner of its selection, in the light of Theorem 1. That number is called the <u>dimension</u> of the subspace.

Note that the set composed solely of the null vector, as also the set composed of the entire space, will each be subspaces in the sense of our definition. We shall regard them as trivial subspaces.

5. The connection between the dimension of a subspace and the rank of a matrix. We introduce the important concept of rank, appropriate to any rectangular matrix A,

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

Any determinant whose rows and columns "fit" the rows and columns of a matrix is called a minor of this matrix. More exactly, a minor of order k of the matrix A is a determinant of the k-th order composed of the elements situated at the intersections of any k rows and any k columns of the matrix A, in their natural arrangement.

The order of whatever non-vanishing minor is of largest order is called the rank of the matrix A. In other words, the rank of a matrix is a number r such that among the minors of the matrix there exists a non-zero minor of order r, but all minors of order $r + 1$ and higher are equal to zero or are not composable (as in the case, for instance, of a rectangular $m \times r$ matrix, $m > r$).

The following important theorem is valid:

THEOREM. The maximum number of linearly independent rows of a matrix, as also the maximum number of linearly independent columns, coincides with the rank of the matrix[1].

---

[1]See, e.g., [1], pp. 62-3, No.4 ff.

From this theorem it follows directly that the dimension of the subspace spanned by the vectors $U_1, \cdots, U_m$ equals the rank of the matrix composed of the elements of these vectors.

In proof, if the rank of a matrix whose columns are the components of the vectors $U_1, \cdots, U_m$ equals r, then of these m vectors r will be linearly independent, and these will correspond to the linearly independent columns of the matrix; all the rest of the columns will be linear combinations of them. Any vector subspace is a linear combination of the vectors $U_1, \cdots, U_m$, which are themselves linear combinations of but r selected linearly independent vectors. Consequently any vector is a linear combination of r vectors, and therefore the rank r of the matrix in question coincides with the dimension of the subspace.

§3. LINEAR TRANSFORMATIONS

1. Let us associate with each vector X of a space a certain vector Y of the same space. Such an association we shall call a <u>transformation</u> of the space. We shall designate the result of the application of transformation $\underline{A}$ to the vector X by $\underline{A}X$.

We shall call the transformation $\underline{A}$ <u>linear</u> if

1. $\underline{A}(\alpha X) = \alpha \underline{A}X$, for any complex number α;
2. $\underline{A}(X_1 + X_2) = \underline{A}X_1 + \underline{A}X_2$.

We shall define, furthermore, operations upon linear transformations. The <u>product</u> of the linear transformations $\underline{A}$ and $\underline{B}$, $\underline{AB} = \underline{C}$, will be a transformation constituted by the transformations $\underline{B}$ and $\underline{A}$ in turn, $\underline{B}$ being completed first, and then $\underline{A}$.

The product of linear transformations is a linear transformation, as is readily seen, since

$$\underline{C}(X_1 + X_2) = \underline{A}(\underline{B}(X_1 + X_2)) = \underline{A}(\underline{B}X_1 + \underline{B}X_2) =$$

(1)
$$= \underline{AB}X_1 + \underline{AB}X_2 = \underline{C}X_1 + \underline{C}X_2 \quad ,$$

$$\underline{C}\alpha X = \underline{AB}\alpha X = \underline{A}\alpha\underline{B}X = \alpha\underline{AB}X = \alpha\underline{C}X \quad .$$

The sum of the linear transformations $\underline{A}$ and $\underline{B}$ will be a transformation $\underline{C}$ which associates the vector X with the vector $\underline{A}X + \underline{B}X$. This sum of linear transformations is obviously itself a linear transformation.

2. Representation of a linear transformation by a matrix. Let us choose, in the space R_n, some basis $e_1, e_2, \cdots, e_n$. A linear transformation relates to the vectors of the basis the vectors $\underline{A}e_1, \underline{A}e_2, \cdots, \underline{A}e_n$.

Let $\underline{A}e_1, \cdots, \underline{A}e_n$ be given in terms of their coordinates with respect to the basis $e_1, e_2, \cdots e_n$, i.e., let

(2)
$$\begin{cases} \underline{A}e_1 = a_{11}e_1 + a_{21}e_2 + \cdots + a_{n1}e_n \\ \underline{A}e_2 = a_{12}e_1 + a_{22}e_2 + \cdots + a_{n2}e_n \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \underline{A}e_n = a_{1n}e_1 + a_{2n}e_2 + \cdots + a_{nn}e_n \end{cases} .$$

Consider the matrix A, its columns composed of the coordinates of the vectors $\underline{A}e_1, \underline{A}e_2, \cdots, \underline{A}e_n$:

(3) $$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} .$$

We shall show that the matrix A uniquely defines the linear transformation[1].

Indeed, if the matrix A is known for the linear transformation, i.e., if $\underline{A}e_1, \underline{A}e_2, \cdots \underline{A}e_n$ are determined, this is sufficient to find the transformation of any vector, for if

(3a) $$X = x_1e_1 + \cdots + x_ne_n ,$$

then

(3b) $$\underline{A}X = x_1\underline{A}e_1 + \cdots + x_n\underline{A}e_n .$$

Hence the coordinates of the transformed vector are easily found, for we have

(3c) $$Y = \underline{A}X = \sum_{k=1}^{n} y_ke_k = \sum_{i=1}^{n} x_i\underline{A}e_i = \sum_{i=1}^{n}\sum_{k=1}^{n} a_{ki}x_ie_k ,$$

whence

(3d) $$y_k = \sum_{i=1}^{n} a_{ki}x_i ,$$

---

[1] Author's remark:
Note, however, that the matrix of the coefficients in the relations (2) forms a matrix which is the transpose of that that we connect with the linear transformation.

or, in matrix notation,

(4) $$y = Ax ,$$

where y and x are columns of the coordinates of vectors Y and X.

Conversely, an arbitrary matrix A may be connected with a certain linear transformation. Indeed, the transformation given by the formula

(4a) $$y = Ax ,$$

where y and x are, as above, the columns of coordinates of the vectors Y and X, is linear for any matrix A.

The established one-to-one correspondence between transformations and matrices is preserved when operations are performed upon transformations, for the matrix of the sum of transformations equals the sum of the matrices of the summand transformations, and the matrix of a product of transformations equals the product of the matrices corresponding to the factor transformations.

3. The connection between the matrices of a linear transformation with respect to different bases. We will now elucidate how the matrix of a linear transformation changes with a change of the basis of the space.

Assume that from the basis $e_1, \dots, e_n$ we have passed to the basis $e'_1, \dots, e'_n$, and let

(4b)
$$\begin{aligned} e'_1 &= c_{11}e_1 + c_{21}e_2 + \dots + c_{n1}e_n \\ e'_2 &= c_{12}e_1 + c_{22}e_2 + \dots + c_{n2}e_n \\ &\dots\dots\dots\dots \\ e'_n &= c_{1n}e_1 + c_{2n}e_2 + \dots + c_{nn}e_n , \end{aligned}$$

that is,

$$E' = C^T E$$

The coordinates of any vector of the space will have changed accordingly by the formula

(4c) $$x = Cx' ,$$

where

(4d) $$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix} , \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{pmatrix} , \quad x' = \begin{pmatrix} x'_1 \\ x'_2 \\ \cdot \\ x'_n \end{pmatrix} .$$

The matrix of the transition, C, is evidently non-singular. It will coincide with the matrix of the linear transformation sending the basis $e_1, e_2, \cdots, e_n$ into the basis $e'_1, e'_2, \cdots, e'_n$.

Let us now consider a linear transformation $\underline{A}$, and let the matrix A correspond to it with respect to the basis $e_1, e_2, \cdots, e_n$, and the matrix B with respect to the basis $e'_1, e'_2, \cdots, e'_n$.

If x is the column of the coordinates of the vector X with respect to the basis $e_1, \cdots, e_n$, and x'--with respect to the basis $e'_1, \cdots, e'_n$, y and y' being the analogous columns for vector Y, we have

(4e) $$\begin{aligned} y &= Ax \\ y' &= Bx' . \end{aligned}$$

But $x = Cx'$, $y = Cy'$, and therefore

$$Cy' = ACx' \tag{4f}$$

or

$$y' = Bx' = C^{-1}ACx' \quad . \tag{4g}$$

Thus similar matrices correspond to the same linear transformation with respect to different bases. Furthermore, the matrix by means of which the similarity transformation is effected coincides with the matrix of transformation of coordinates.

4. The transfer rule for a matrix in a scalar product. Let X and Y be two vectors given by their components with respect to the initial basis: $X = (x_1, \ldots, x_n)$, $Y = (y_1, \ldots, y_n)$, and let $\underline{A}$ be a linear transformation with matrix $A = (a_{ik})$. Designate by $\underline{A}^*$ the linear transformation with matrix A^*, the elements of which are the complex-conjugates of their counterparts in A, and which are placed in transposed position: $A^* = (a_{ik})^* = (\overline{a}_{ik}^T) = (\overline{a_{ki}})$. We shall prove the following formula:

$$(\underline{A}X, Y) = (X, \underline{A}^*Y) \quad . \tag{5}$$

In demonstration, we have

$$(\underline{A}X, Y) = \sum_{i=1}^{n} \sum_{k=1}^{n} a_{ik} x_k \overline{y}_i = \sum_{k=1}^{n} x_k \overline{\sum_{i=1}^{n} \overline{a}_{ik} y_i} = (X, \underline{A}^*Y) \quad . \tag{5a}$$

5. The rank of a linear transformation. Let $\underline{A}$ be a certain linear transformation. The set of vectors $\underline{A}X$ will obviously constitute a subspace, which we shall denote by $\underline{A}R_n$.

The dimension of this subspace is said to be the rank of the transformation $\underline{A}$.

We shall show the rank of a transformation to be equal to the rank of the matrix corresponding to this transformation on any basis whatever, $e_1, e_2, \cdots, e_n$. Obviously the subspace $\underline{A}R_n$ is spanned by the vectors $\underline{A}e_1, \underline{A}e_2, \cdots, \underline{A}e_n$. The dimension of $\underline{A}R_n$ is accordingly equal to the rank of a matrix whose columns are composed of the coordinates of the vectors $\underline{A}e_1, \underline{A}e_2, \cdots, \underline{A}e_n$, i.e., to the rank of a matrix corresponding to the transformation.

Since the dimension of a subspace does not depend upon the selection of the basis, it follows from the foregoing that the ranks of similar matrices are equal.

6. The latent vectors of a linear transformation. By a latent vector (characteristic vector, proper vector or eigenvector) of a linear transformation $\underline{A}$ is meant any non-zero vector X such that

$$(6) \qquad \underline{A}X = \lambda X \quad ,$$

where λ is any complex number.

The number λ is called the latent root (characteristic value, proper value, or eigenvalue) of the transformation. The spectrum of the transformation is the aggregate of its latent roots.

The latent roots and latent vectors of the transformation may be determined in the following manner. Let the transformation $\underline{A}$ be connected with the matrix $A = (a_{ik})$ with respect to some basis; let the coordinates of the latent vector X, with respect to this basis, be $x_1, \cdots, x_n$.

The coordinates of the vector $\underline{A}X$ will then be:

(6a)
$$\left(\sum_{k=1}^{n} a_{1k}x_k, \cdots, \sum_{k=1}^{n} a_{nk}x_k \right),$$

and thus for the determination of $x_1, x_2, \cdots, x_n$ and the latent root λ we will have the system of equations:

(7)
$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = \lambda x_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = \lambda x_2 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = \lambda x_n \end{cases}$$

or

(7')
$$\begin{cases} (a_{11} - \lambda)\, x_1 + a_{12}x_2 + \cdots + a_{1n}\, x_n = 0 \\ a_{21}x_1 + (a_{22} - \lambda)\, x_2 + \cdots + a_{2n}\, x_n = 0 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + (a_{nn} - \lambda)\, x_n = 0 \end{cases} .$$

This system of homogeneous equations in $x_1, \cdots, x_n$ will have a non-zero solution only in case

$$\text{(7'a)} \qquad \begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{vmatrix} = 0 \quad ,$$

i.e., if λ is a root of the characteristic polynomial of the matrix. Thus the following is valid:

THEOREM. The latent roots of a transformation coincide with the roots of the characteristic polynomial of the matrix that is connected with this transformation with respect to an arbitrary basis.

From the theorem often called the fundamental theorem of higher algebra, we know that every polynomial has at least one root. A linear transformation will consequently have at least one latent root, which may be complex even though the matrix of the transformation be real. In view of the theory of linear homogeneous systems of equations, there will be a non-zero solution of system (7) for each latent root, i.e., with each latent root at least one latent vector is associated.

Obviously if X is a latent vector of the transformation A, then, for all $c \neq 0$, cX will also be a latent vector of transformation A corresponding to the same latent root. Furthermore if several latent vectors correspond to some one latent root, then any linear combination of them will be a latent vector of the transformation associated with the same root. The set of latent vectors corresponding to a single latent root forms a linear

subspace. We shall establish that its dimension, γ, does not exceed the multiplicity of the latent root. Indeed, let $X_1, \cdots, X_\gamma$ be linearly independent latent vectors corresponding to the single latent root λ_1. Construct a basis of the space $X_1, \cdots, X_n$, having taken as the first γ vectors the vectors $X_1, \cdots, X_\gamma$. With respect to this basis the linear transformation under consideration is connected with a matrix whose first γ columns have the form

(7'b)
$$\begin{array}{cccc} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & \lambda_1 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 0 \end{array},$$

for $\underline{A}X_1 = \lambda_1 X_1, \cdots, \underline{A}X_\gamma = \lambda_1 X_\gamma$. Now $(\lambda - \lambda_1)^\gamma$ is a factor of the characteristic polynomial of this matrix, and accordingly λ_1 is of multiplicity k not less than γ, i.e., $\gamma \leq k$. It would naturally be supposed that $\gamma = k$, i.e., that to a k-multiple root of the characteristic polynomial there correspond k linearly independent latent vectors. But this is in fact not true. In reality, the number of linearly independent vectors may be less than the multiplicity of the latent root.

Let us confirm the preceding statement with an example. Consider the linear transformation with the matrix

(7'c)
$$A = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix} .$$

Then $|A - \lambda I| = (\lambda - 3)^2$, and thus $\lambda = 3$ is a double root of the characteristic polynomial.

The system of equations for determining the coordinates of the latent vector of the transformation $\underline{A}$ will be:

(7'd)
$$3x_1 + x_2 = 3x_1$$
$$3x_2 = 3x_2$$

whence $x_2 = 0$, and thus all the latent roots of the transformation in question will be $(x_1, 0) = x_1(1, 0)$. So in this instance only one linearly independent vector is associated with a double root.

Generally speaking, the coordinates of a latent vector on the chosen basis are to be determined from the system (7) of linear equations, in which for λ the latent root λ_i is substituted. But as is known from the theory of systems of linear equations, the number of linearly independent solutions of a homogeneous system equals $n - r$, where r is the rank of the matrix composed of the coefficients of the system. Therefore if r denote the rank of the matrix $A - \lambda I$, $\gamma = n - r$. Thus $n - r \leq k$, and the equality does not always hold.

In case the basis does not change in the course of the argument, we shall often identify the linear transformation with the matrix of the

linear transformation with respect to this basis, and any vector space with the columns of its coordinates.

On this agreement, it makes sense to speak of a latent vector of a matrix, understanding by this a column x satisfying the condition

$$Ax = \lambda x \quad . \tag{7'e}$$

We remark that if a latent root of a real matrix is complex, the coordinates of an associated latent vector will be complex. A vector whose coordinates are the complex-conjugates of those of a given latent vector of a real matrix is also a latent vector of that matrix, and is associated with the complex-conjugate latent root. To convince oneself of this, it is enough to change all numbers in the equation $Ax = \lambda x$ into their complex conjugates.

7. Properties of the latent roots and vectors of a matrix. We shall establish several properties of the latent roots and vectors of a real matrix.

First of all we note that a matrix and its transpose have identical characteristic polynomials and consequently identical spectra. This is evident since $|A^T - \lambda I| = |A - \lambda I|$, on the strength of the fact that a determinant is not altered when its rows and columns are interchanged.

Now let λ_r and λ_s denote distinct latent roots of the real matrix A, and $\bar{\lambda}_s$ the complex conjugate of λ_s. As we saw above, $\bar{\lambda}_s$ is also a latent root of matrix A, and thus also of the transposed matrix A'. Let

X_r be a latent vector of the matrix A belonging to the latent root λ_r, and X_s' the latent vector of the matrix A' belonging to the latent root $\bar{\lambda}_s$. We shall show that X_r and X_s' are orthogonal.

With this object in view, let us form the scalar product (AX_r, X_s') and reckon it by two methods.

By one method the reckoning is

$$(7'f) \qquad (AX_r, X_s') = (\lambda_r X_r, X_s') = \lambda_r(X_r, X_s') \quad .$$

On the other hand, since matrix A is real, we have $A^* = A'$, and therefore

$$(7'g) \qquad (AX_r, X_s') = (X_r, A'X_s') = (X_r, \bar{\lambda}_s X_s') = \lambda_s(X_r, X_s') \quad .$$

Thus $\lambda_r(X_r, X_s') = \lambda_s(X_r, X_s')$. But the condition was that $\lambda_r \neq \lambda_s$, and therefore $(X_r, X_s') = 0$, which is what was required to be proved.

In case all latent roots are distinct, the demonstrated property gives $n^2 - n$ relations of orthogonality between the latent vectors of matrices A and A'. We shall later return to these properties in more detail.

For a real symmetric matrix, the properties of orthogonality are considerably simplified thanks to the fact that all its latent roots are real. In proof, letting λ and X be respectively latent root and vector, we have $(AX, X) = \lambda(X, X)$; $(AX, X) = (X, A^T X) = (X, AX) = (X, \lambda X) = \bar{\lambda}(X, X)$. Thus $(\lambda - \bar{\lambda})(X, X) = 0$, and as $(X, X) > 0$, $\lambda = \bar{\lambda}$, i.e., λ is real.

From the reality of the latent roots of a real symmetric matrix it follows that vectors with real components may be taken as the latent vectors

belonging to those roots; the components will indeed be found by solving the linear homogeneous system with real coefficients.

The orthogonality property of the latent vectors of a real symmetric matrix is very simply formulated in view of the coincidence of the matrix with its transpose and the reality of the latent roots, viz.: latent vectors belonging to distinct latent roots are orthogonal.

8. The latent roots of a positive definite quadratic form. A homogeneous polynomial of the second degree in several variables $x_1, \cdots, x_n$ is called a quadratic form. We shall consider only those with real coefficients. Any quadratic form may be written as

(7'h) $$\Phi(x_1, x_2, \cdots, x_n) = \sum_{i,k=1}^{n} a_{ik} x_i x_k ,$$

where[1] $a_{ik} = a_{ki}$.

A quadratic form is said to be positive definite if its values are positive for any real values of $x_1, \cdots, x_n$, not all zero.

It is evident that the diagonal coefficients of a positive definite form are positive, for

(7'i) $$a_{11} = \Phi(1, 0, \cdots, 0), \quad a_{22} = \Phi(0, 1, \cdots, 0), \cdots$$
$$a_{nn} = \Phi(0, 0, \cdots, 1) .$$

Denoting by X the vector with components $(x_1, \cdots, x_n)$, we may write a quadratic form as (AX, X), A being the matrix composed of the coefficients

[1] See, e.g., [1], §10.5 ff.

of the form. This matrix is symmetric, on the strength of the definition. The latent roots of the matrix are called the latent roots of the quadratic form. In view of the previous results, <u>all latent roots of a quadratic form are real</u>.

We shall show that <u>if a quadratic form is positive definite, its latent roots are positive</u>.

In demonstration, let X be a real latent vector, belonging to λ, a latent root of the matrix of the form. Then, since the form is positive definite, $(AX, X) > 0$. On the other hand, $(AX, X) = \lambda(X, X)$. Thus

$$\lambda = \frac{(AX,X)}{(X,X)} \quad . \tag{7'j}$$

But both numerator and denominator of this fraction are positive, and consequently $\lambda > 0$, which is what was required to be proved.

Let there now be given any real, non-singular matrix A. Obviously $B = A'A$ is a symmetric matrix, since $B' = (A'A)' = A'A'' = A'A = B$.

We shall show that a quadratic form with matrix B is positive definite. We have, indeed,

$$(BX, X) = (A'AX, X) = (AX, AX) > 0 \tag{7'k}$$

for any real vector X.

We shall establish, lastly, that if A is the matrix of a positive definite quadratic form, $(AX, X) > 0$ even for a complex vector X.

In proof, let $X = Y + iZ$, where Y and Z are vectors with real components. Then

$$(AX, X) = (AY + iAZ, Y + iZ) =$$

(7'l) $$= (AY, Y) + i(AZ, Y) - i(AY, Z) + (AZ, Z) =$$

$$= (AY, Y) + (AZ, Z) > 0$$

because $(AZ, Y) = (Z, AY) = (AY, Z)$.

In complex space, instead of the quadratic form one deals with an Hermitian form, an expression of the type

(7'm) $$\sum_{i,k=1}^{n} a_{ik}x_i\bar{x}_k \quad ,$$

under the condition that $a_{ki} = \overline{a_{ik}}$.

The matrix of an Hermitian form is called Hermitian (or Hermitian symmetric); a linear transformation with an Hermitian matrix relative to an orthonormal base is called self-conjugate. It is obvious that

(7'n) $$\Sigma\, a_{ik}x_i\bar{x}_k = (AX, X) \quad .$$

To show that all the values of an Hermitian form are real, we have only to note that

(7'o) $$(AX, X) = (X, A^{*}X) = \overline{(AX, X)} \quad .$$

If all the values of an Hermitian form are positive, it is called positive definite.

It can be shown that the latent roots of an Hermitian matrix are real.[1] The latent roots of a positive definite Hermitian form are positive.

---

[1]See, e.g., [1], §30.9

9. The reduction of a matrix to diagonal form. Let us consider the matrix A all of whose latent roots, $\lambda_1, \cdots, \lambda_n$, are distinct, and the transformation $\underset{\sim}{A}$ connected with it with respect to the initial basis. It will have n distinct latent vectors $X_1, \cdots, X_n$. We shall show that the vectors $X_1, \cdots, X_n$ are linearly independent.

Assume the contrary: let the vectors $X_1, \cdots, X_n$ be linearly dependent. Without detriment to the generality we may assume that the vectors $X_1, \cdots, X_k$, where $k < n$, are linearly independent, and thus that the vectors $X_{k+1}, \cdots, X_n$ are linear combinations of them. In particular, let

(8) $$X_n = \sum_{i=1}^{k} c_i X_i \quad ;$$

then

(8a) $$\underset{\sim}{A} X_n = \underset{\sim}{A} \sum_{i=1}^{k} c_i X_i = \sum_{i=1}^{k} \lambda_i c_i X_i \quad .$$

On the other hand,

(8b) $$\underset{\sim}{A} X_n = \lambda_n X_n = \sum_{i=1}^{k} \lambda_n c_i X_i \quad ,$$

whence

(9) $$\sum_{i=1}^{k} (\lambda_n - \lambda_i)\, c_i X_i = 0 \quad .$$

But $\lambda_n \neq \lambda_i$, on assumption. Thus, since the vectors X_i are linearly independent, all the coefficients c_i equal zero, and therefore $X_n = 0$,

which contradicts the definition of a latent vector. So the vectors $X_1, X_2, \cdots, X_n$ are linearly independent. Let us adopt them as a new basis of the space. With respect to the new basis the linear transformation $\underset{=}{A}$ will be connected with a matrix whose columns are composed of the coordinates of the vectors $\underset{=}{A}X_1, \underset{=}{A}X_2, \cdots, \underset{=}{A}X_n$ with respect to the basis $X_1, X_2, \cdots, X_n$.

But

(9a) $$AX_k = \lambda_k X_k \quad ,$$

and the matrix of the transformation of the new basis will consequently be diagonal: $\lceil \lambda_1, \lambda_2, \cdots, \lambda_n \rfloor$.

So the linear transformation $\underset{=}{A}$ has, with respect to the initial basis, the matrix A, and with respect to the basis of the latent vectors, the diagonal matrix $\lceil \lambda_1, \lambda_2, \cdots \lambda_n \rfloor$. Accordingly, on the strength of what has been noted above,

(10) $$V^{-1}AV = \lceil \lambda_1, \lambda_2, \cdots, \lambda_n \rfloor \quad ,$$

where V is the matrix whose columns are the coordinates (with respect to the initial basis) of the latent vectors.

Observation. If the latent roots of a matrix are of multiplicity greater than one, but to each latent root there correspond as many latent vectors as it has multiplicity, the matrix may also be reduced to diagonal form. This will be the case, for example, with symmetric matrices: it can be proved that to each latent root of a symmetric matrix there

correspond as many linearly independent latent vectors as the multiplicity of the latent root. Moreover, the linearly independent latent vectors belonging to a single latent root may be subjected to the orthogonalizing process. We have seen, too, that the latent vectors of a symmetric matrix that belong to distinct latent roots are mutually orthogonal. Thus for a symmetric matrix it is possible to construct an orthogonal system of latent vectors forming a basis for the whole space.

The question of the transformation of a symmetric matrix to diagonal form is closely connected with the theory of quadratic forms.

10. The latent roots and latent vectors of similar matrices. It has been established that similar matrices have identical characteristic polynomials, and consequently identical spectra of latent roots.

We have explained the geometrical cause of this circumstance, viz.: similar matrices may be regarded as matrices of one and the same transformation, referred to different bases. Therefore the latent vectors of similar matrices are columns of the coordinates of the latent vectors of the transformation under consideration, with respect to different bases, and are thus connected by the relation $x' = C^{-1}x$, C being the matrix of transformation of coordinates. This circumstance may be verified formally: if $Ax = \lambda x$, $(C^{-1}AC)(C^{-1}x) = \lambda(C^{-1}x)$.

11. The latent roots of a polynomial in a matrix. Let A be a matrix with latent roots $\lambda_1, \ldots, \lambda_n$, and let $\varphi(x) = a_0 + a_1 x + \cdots + a_m x^m$ be the given polynomial. Then the latent roots of the matrix will be $\varphi(\lambda_1), \varphi(\lambda_2), \cdots, \varphi(\lambda_n)$.

This is readily established for a matrix all of whose latent roots are distinct. Indeed, such a matrix can be reduced to diagonal form by a similarity transformation:

(10.a) $$A = C^{-1} \lceil \lambda_1, \lambda_2, \cdots, \lambda_n \rfloor C \quad .$$

Accordingly,

(10.b) $$\varphi(A) = C^{-1} \varphi \lceil \lambda_1, \lambda_2, \cdots, \lambda_n \rfloor C \quad .$$

But

(10.c) $$\varphi(\lceil \lambda_1, \cdots, \lambda_n \rfloor) = \lceil \varphi(\lambda_1), \varphi(\lambda_2), \cdots, \varphi(\lambda_n) \rfloor \quad ,$$

which follows from the fact that

(10.d) $$\lceil \lambda_1, \cdots, \lambda_n \rfloor^k = \lceil \lambda_1^k, \cdots, \lambda_n^k \rfloor \quad .$$

Consequently

(10.e) $$\varphi(\lceil \lambda_1, \cdots, \lambda_n \rfloor) = \sum_{k=o}^{m} a_k \lceil \lambda_1^k, \cdots, \lambda_n^k \rfloor =$$
$$= \lceil \sum_{k=o}^{m} a_k \lambda_1^k, \cdots, \sum_{k=o}^{m} a_k \lambda_n^k \rfloor = \lceil \varphi(\lambda_1), \varphi(\lambda_2), \cdots, \varphi(\lambda_n) \rfloor \, .$$

Thus the matrix $\varphi(A)$ is similar to the matrix $\lceil \varphi(\lambda_1), \cdots, \varphi(\lambda_n) \rfloor$ and accordingly its latent roots are $\varphi(\lambda_1), \varphi(\lambda_2), \cdots, \varphi(\lambda_n)$, Q.E.D..

This result remains true for any matrix, of which one may readily convince oneself, for example, by considerations of continuity.

We particularly note that the latent roots of the matrix A^k are λ^k.

12. <u>The normalization of the latent vectors of a matrix. The second group of orthogonality relations.</u> Let A be a real matrix whose latent roots, $\lambda_1, \lambda_2, \cdots, \lambda_n$, are distinct, and let $X_1, X_2, \cdots, X_n$ be the latent vectors corresponding to them. As we saw, the transposed matrix, A', has the same latent roots. Let $X_1', X_2', \cdots, X_n'$ be the latent vectors of the matrix A', and their enumeration be so chosen that X_i and X_i' belong to complex conjugate roots. We established above that the following orthogonality relation holds: $(X_i', X_j) = 0$ for $i \neq j$. We shall show now that, having chosen $X_1', X_2', \cdots, X_n'$ in any manner (they are determined but for a numerical multiplier), we may norm the vectors $X_1, \cdots, X_n$ so that $(X_i', X_i) = 1$.

In demonstration of this, the vectors $X_1, X_2, \cdots, X_n$ are known to be linearly independent, and they accordingly form a basis of the space. Resolve X_i' in terms of this basis:

$$X_i' = \gamma_1 X_1 + \gamma_2 X_2 + \cdots + \gamma_n X_n \quad . \tag{10.f}$$

Forming the scalar product (X_i', X_i'), we obtain

$$(X_i', X_i') =$$

$$\gamma_1(X_i', X_1) + \cdots + \gamma_i(X_i', X_i) + \cdots + \gamma_n(X_i', X_n) = \gamma_i(X_i', X_i) \quad , \tag{10.g}$$

whence we conclude that $(X_i', X_i) = \alpha_i \neq 0$, for $(X_i', X_i') > 0$.

Adopting instead of the vectors $X_1, \cdots, X_n$ the vectors $\frac{1}{\alpha_1} X_1, \cdots, \frac{1}{\alpha_n} X_n$, we arrive at the required normalization, since

$$(10.h) \qquad \left(X_i', \frac{1}{\alpha_i} X_i\right) = \frac{1}{\alpha_i} (X_i', X_i) = 1 \quad .$$

From the relations of orthogonality and normality set forth above, we may extract another group of relations between the components of the latent vectors of the matrix A and its transpose.

Form the matrices

$$(10.i) \qquad X' = \begin{pmatrix} x_{11}' & x_{21}' & \cdots & x_{n1}' \\ x_{12}' & x_{22}' & \cdots & x_{n2}' \\ \cdot & \cdot & \cdot & \cdot \\ x_{1n}' & x_{2n}' & \cdots & x_{nn}' \end{pmatrix} \text{ and } X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix} .$$

The columns of the matrix X are composed of the components of the vectors $X_1, \cdots, X_n$. The rows of the matrix X' are composed of the numbers complex-conjugate with the components of the vectors $X_1', \cdots, X_n'$. (We observe that the numbers that are the complex conjugates of the components of the vectors $X_1', \cdots, X_n'$ are the components of the vectors $\overline{X_1'}, \cdots, \overline{X_n'}$, and will also be the latent vectors of the matrix A' that belong to the latent roots $\lambda_1, \cdots, \lambda_n$. Thus the i-th row of the matrix X' and the i-th column of the matrix X are composed of the components of the latent vectors

of the matrix A' and A belonging to the same latent root λ_i, and not to latent roots that are complex conjugates of each other.)

It is readily seen that:

$$X'X = I \quad , \tag{11}$$

for we have the element of the i-th row and j-th column of the matrix X'X equalling $\sum_{k=1}^{n} x'_{ki}x_{kj} = (X_j, X'_j) = \delta_{ij}$, where δ_{ij} is Kronecker's delta: $\left(\delta_{ij} = \begin{matrix} 0, & i \neq j; \\ 1, & i = j. \end{matrix}\right)$. Thus X' and X are mutually inverse matrices, and accordingly XX' is likewise equal to I. This gives a second group of orthogonality and normality relations between the latent vectors of the matrices A and A', viz.:

$$\begin{cases} x_{i1}x'_{j1} + x_{i2}x'_{j2} + \cdots + x_{in}x'_{jn} = 0 \quad i \neq j \\ x_{i1}x'_{i1} + x_{i2}x'_{i2} + \cdots + x_{in}x'_{in} = 1 \end{cases} \quad . \tag{12}$$

Thus for a matrix of the second order the ordinary conditions of orthogonality may be written in the form (we preserve only one index of the components of the latent vectors, designating the first components by x, the second by y):

$$\begin{aligned} x_1x'_2 + y_1y'_2 &= 0 \\ x_2x'_1 + y_2y'_1 &= 0 \quad , \end{aligned} \tag{12.a}$$

and the normality conditions:

$$x_1'x_1 + y_1'y_1 = 1 \qquad (12.b)$$
$$x_2'x_2 + y_2'y_2 = 1 \; .$$

The new relations will be

$$x_1y_1' + x_2y_2' = 0 \quad x_1x_1' + x_2x_2' = 1 \qquad (12.c)$$
$$y_1x_1' + y_2x_2' = 0 \quad y_1y_1' + y_2y_2' = 1 \; .$$

Observation. In case the latent roots of the matrix are multiple, and to each latent root there correspond as many linearly independent latent vectors as the multiplicity of the root, these indicated properties of the latent vectors hold as before.

§4. THE JORDAN CANONICAL FORM

We have proved above that if a matrix has n distinct latent roots, it may be brought into diagonal form by a similarity transformation. Given the presence of multiple roots, however, such a transformation may not always be possible. Nonetheless the question whether the form may possibly be rendered more simple via a similarity transformation can well be put. The problem is equivalent to discovering a basis with respect to which the linear transformation connected with the given matrix would have a matrix of simplest form, and the latter proves to be the Jordan canonical form.

Proof of a fundamental theorem to the effect that any matrix may be brought into the Jordan canonical form by a similarity transformation is

rather complicated and we will not dwell on it here. The gist of it is stated, for instance, in [2], §3.16; a scholarly presentation is available in, e.g., [3], Chap. V-VI. We shall limit ourselves to a description of this canonical form.

A matrix--most commonly a submatrix--of the following form is called a canonical box:

$$(1) \qquad \begin{pmatrix} \lambda_i & 0 & 0 & \cdots & 0 & 0 \\ 1 & \lambda_i & 0 & \cdots & 0 & 0 \\ 0 & 1 & \lambda_i & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 1 & \lambda_i \end{pmatrix} .$$

On its principal diagonal the single number λ_i is everywhere to be found; directly under the diagonal (in the subdiagonal) are disposed elements that are all units; all the rest of the elements are zero.

A canonical box cannot be simplified by utilizing a similarity transformation. It is obvious that a canonical box has the sole multiple latent root λ_i. It may be easily verified that a canonical box has only one latent vector. The minimum polynomial of a box coincides with its characteristic polynomial, viz., it equals $(\lambda - \lambda_i)^{m_i}$ where m_i is the order of the box. The Jordan (classical) canonical form is a quasi-diagonal matrix composed of canonical boxes:

$$
(2) \qquad \begin{bmatrix}
\begin{matrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 1 & \lambda_1 & 0 & \cdots & 0 \\ 0 & 1 & \lambda_1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & \lambda_1 \end{matrix} & & \\
& \ddots & \\
& & \begin{matrix} \lambda_s & 0 & 0 & \cdots & 0 \\ 1 & \lambda_s & 0 & \cdots & 0 \\ 0 & 1 & \lambda_s & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & \lambda_s \end{matrix}
\end{bmatrix}.
$$

It is admissible that the same number λ_i appear in several canonical boxes.

All the numbers λ_i appearing in the different boxes are latent roots of the canonical matrix, and the multiplicity of the latent roots equals the sum of the orders of the boxes in which it figures as diagonal element. In proof of this, by the theorem concerning the determinant of a quasi-diagonal matrix, the characteristic polynomial of the canonical matrix equals the product of the characteristic polynomials of the separate boxes, each of them equal to $(\lambda - \lambda_i)^{m_i}$, where λ_i is the latent root and m_i is the order of the i-th box. Hence follows directly the statement under proof.

The determination of the Jordan boxes for a given matrix A presents certain difficulties. The characteristic polynomial $\prod (\lambda - \lambda_i)^{m_i}$ coincides with the characteristic polynomial of the original matrix, and

it is consequently possible to find it without knowing the canonical matrix itself. Nonetheless, a knowledge of the characteristic polynomial still does not make possible the complete determination of the canonical form, for to a latent root λ_i of multiplicity k_i there may correspond several Jordan boxes containing this number as a diagonal element, and regarding them only the sum of their orders will be known, not the order of each box in particular. If the canonical form is to be fully determined, a knowledge of the "elementary divisors" of the matrix must be drawn upon.

Designate by $D_i(\lambda)$ the greatest common divisor of all the minors of the i-th order of the determinant $|A - \lambda I|$. In particular, $D_n(\lambda)$ coincides with the characteristic polynomial. It can be proved that all $D_i(\lambda)$, as $D_n(\lambda)$, are general for the class of similar matrices. It can be proved, moreover, that $D_{i-1}(\lambda)$ divides $D_i(\lambda)$.[1]

Put

$$\frac{D_i(\lambda)}{D_{i-1}(\lambda)} = E_i(\lambda) \quad ; \tag{3}$$

obviously

$$D_n(\lambda) = \prod_{i=1}^{n} E_i(\lambda) \quad . \tag{4}$$

It turns out, moreover, that $E_n(\lambda) = \frac{D_n(\lambda)}{D_{n-1}(\lambda)}$ is the minimum polynomial of the matrix.

Resolve $E_i(\lambda)$ into linear factors. Then

---

[1]See [3], p. 23 ff., [5] 50 ff.

(5) $$E_i(\lambda) = \prod_{j=1}^{s} (\lambda_j - \lambda)^{m_{ij}} ,$$

Here s denotes the number of distinct latent roots, $\sum_{i=1}^{n} m_{ij} = k_j$; $\sum_{j=1}^{s} \sum_{i=1}^{n} m_{ij} = n$. It is obvious that among the exponents m_{ij} only some will different from zero.

<u>Translator's</u> <u>Note</u>. It may be helpful to the student to have <u>in extenso</u> a synopsis of the relations of these important entities:

$$\frac{D_n(\lambda)}{D_{n-1}(\lambda)} = E_n(\lambda);\quad \begin{cases} D_i \text{ is the (i-th) \underline{determinantal} \underline{divisor} of } D_n; \\ E_i \text{ is the (i-th) \underline{invariant} \underline{factor} of } D_n. \end{cases}$$

Thus

$$D_n(\lambda) = D_{n-1}(\lambda) \cdot E_n(\lambda) ,$$

and, in sequence,

$$D_{n-1}(\lambda) = D_{n-2}(\lambda) \cdot E_{n-1}(\lambda) ,$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$D_3(\lambda) = D_2(\lambda) \cdot E_3(\lambda) ,$$

$$D_2(\lambda) = D_1(\lambda) \cdot E_2(\lambda) ,$$

$$D_1(\lambda) = E_1(\lambda) .$$

Reversing our view of the development:

$$D_1(\lambda) = E_1(\lambda) \quad ,$$

$$D_2(\lambda) = D_1(\lambda)\cdot E_2(\lambda) = E_1(\lambda)\cdot E_2(\lambda) \quad ,$$

$$D_3(\lambda) = D_2(\lambda)\cdot E_3(\lambda) = E_1(\lambda)\cdot E_2(\lambda)\cdot E_3(\lambda) = \prod_{i=1}^{3} E_i(\lambda) \quad ,$$

. . . .

$$(6) \quad D_n(\lambda) = D_{n-1}(\lambda)\cdot E_n(\lambda) = E_1(\lambda)\cdot \cdots \cdot E_n(\lambda) = \prod_{i=1}^{n} E_i(\lambda) \ .$$

Consider now that each E_i is factorable into linear factors holding the s distinct latent roots λ_j. These factors may be multiple, or may not appear in the given E_i (multiplicity = 0):

$$(7) \quad E_i(\lambda) = (\lambda - \lambda_1)^{m_{i1}}\cdot(\lambda - \lambda_2)^{m_{i2}}\cdot \cdots \cdot(\lambda - \lambda_s)^{m_{is}} = \prod_{j=1}^{s} (\lambda - \lambda_j)^{m_{ij}} \ .$$

Here m_{ij} is the power (perhaps zero) of the j-th distinct linear factor of E_i; there are s of such powered prime factors, each of which latter is called an <u>elementary</u> <u>divisor</u> of $D_n(\lambda)$, since, using (7) in (6), we have

$$D_n(\lambda) = \prod_{i=1}^{n} E_i(\lambda) = \prod_{i=1}^{n}\prod_{j=1}^{s} (\lambda - \lambda_j)^{m_{ij}} \ .$$

Since there are n latent roots of $D_n(\lambda)$, some perhaps multiple, $\sum_{i=1}^{n}\sum_{j=1}^{s} m_{ij} = n$.

Since the linear factor for the j-th latent root, $(\lambda - \lambda_j)$, may appear in more than one E_i—a total of k_j times, in fact, we have $\sum_{i=1}^{n} m_{ij} = k_j$.

It is usual, on behalf of greater generality, not to limit one's concern to $D_n(\lambda) = |A - \lambda I|$, but to conceive $D_n(\lambda)$ as a matrix all of whose elements are polynomials--of whatsoever degree--with coefficients in some specified field, rather than, as here, a matrix whose diagonal elements only are polynomials, and those of the first degree. The above treatment, otherwise unaltered, then has as its subject such a generalized $D_n(\lambda)$, a "lambda matrix", as it has come to be spoken of.

The binomials $(\lambda - \lambda_j)^{m_{ij}}$ are known as the elementary divisors of $|A - \lambda I|$, and, by extension, as those of the matrix A. A knowledge of the elementary divisors permits us to construct the canonical form, viz.: the Jordan boxes are constructed by starting from the number λ_j, and the orders of these boxes are equal to the exponents m_{ij}. The number of boxes containing λ_j equals the number of exponents m_{ij} not equal to zero.

In case the elementary divisors are linear, i.e., if all the non-zero exponents m_{ij} are equal to one, the Jordan boxes degenerate into diagonal elements, and the canonical form turns out to be simply a diagonal form, wherein, of course, a single latent root will appear as often as a diagonal element as it has multiplicity as a root of the characteristic equation.

The converse is also obvious, for it is clear that if a matrix can be brought to diagonal form, its elementary divisors are linear. Therefore matrices with distinct latent roots, as also symmetric matrices, have linear elementary divisors.

If all the elementary divisors $(\lambda - \lambda_j)^{m_{ij}}$ are relatively prime (which occurs only in case $D_{n-1}(\lambda) = 1$), each latent root appears in only one canonical box, and the order of the box equals the multiplicity of the corresponding latent root. Only in this case does the minimum polynomial coincide with the characteristic polynomial.

Let us now consider the matrix transforming the given matrix into canonical form. With this object in view we introduce into the discussion a linear transformation connected with the matrix A with respect to the initial basis. Then the columns of the transforming matrix will be components of the vectors of that basis in which the linear transformation in question is described as a canonical matrix.

Let this canonical matrix have the form

$$(8) \qquad \begin{pmatrix} \Lambda_1 & & & & \\ & \Lambda_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \Lambda_s \end{pmatrix},$$

where

$$(9) \qquad \Lambda_r = \begin{pmatrix} \lambda_r & 0 & \cdots & 0 \\ 1 & \lambda_r & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & & \lambda_r \end{pmatrix}.$$

and the order of A_r equal to m_r; let $U_1^{(1)}, U_2^{(1)}, \dots, U_{m_1}^{(1)}, \dots, U_1^{(s)}, \dots, U_{m_s}^{(s)}$ be the corresponding basis. Then the following formulas for the transformation hold:

$$
(10) \qquad
\begin{aligned}
AU_1^{(r)} &= \lambda_r U_1^{(r)} + U_2^{(r)} \\
AU_2^{(r)} &= \lambda_r U_2^{(r)} + U_3^{(r)} \\
&\cdots\cdots\cdots \\
AU_{m_r-1}^{(r)} &= \lambda_r U_{m_r-1}^{(r)} + U_{m_r}^{(r)} \\
AU_{m_r}^{(r)} &= \lambda_r U_{m_r}^{(r)}
\end{aligned}
$$

for all $r = 1, \dots, s$.

We see that among the vectors of the canonical basis are the latent vectors of the given matrix, one per box. It can be proved that with this all linearly independent latent vectors of the matrix are exhausted, and consequently the number of linearly independent latent vectors of the given matrix equals the number of canonical boxes in its canonical form. In particular, the number of linearly independent latent vectors belonging to the given latent root equals the number of canonical boxes containing this root. It is not of greater multiplicity than the latent root, and is equal to its multiplicity in case, and only in case, all boxes containing the given latent root are of order 1, i.e., when the corresponding elementary divisors are linear.

§5. THE CONCEPT OF LIMIT FOR VECTORS AND MATRICES

Let a sequence of vectors $X^{(1)}, X^{(2)}, \dots, X^{(k)}, \dots$ with components $(x_1^{(1)}, \dots, x_n^{(1)}), \dots, (x_1^{(k)}, \dots, x_n^{(k)}), \dots$ be given. If a limit exists for each component: $\lim_{k\to\infty} x_i^{(k)} = x_i$, the vector X with components $x_1, \dots, x_n$, is called the _limit_ of the sequence $X^{(1)}, X^{(2)}, \dots, X^{(k)}, \dots$ and the sequence itself is said to be _convergent_ to the vector X. This is written in the form $X^{(k)} \to X$ or $\lim_{k\to\infty} X^{(k)} = X$.

In the same fashion, given a sequence of square matrices $A^{(1)}, A^{(2)}, \dots, A^{(k)}, \dots$ with elements $(a_{ik}^{(1)}), (a_{ij}^{(2)}), \dots, (a_{ij}^{(k)}), \dots$ the matrix A with elements $a_{ij} = \lim_{k\to\infty} a_{ij}^{(k)}$ is called the _limit_ of the sequence, if all these limits exist.

In accordance with such a definition of a limit, an _infinite series of vectors_ $X^{(1)} + X^{(2)} + \dots + X^{(k)} + \dots$ is said to be _convergent_ if $\lim_{k\to\infty} (X^{(1)} + X^{(2)} + \dots + X^{(k)})$ exists; this limit is called the _sum_ of the given series. Obviously it is necessary and sufficient for the convergence of a series of vectors that all the series composed of their corresponding components, i.e., components bearing the same indices, converge; the sums of these series are the components of the sum of the series of vectors.

The concept of the convergence of a series of matrices is defined analogously.

In applied questions it is usually important to judge not only the convergence of a sequence or series, but also to judge the _rapidity_ of

this convergence. With this object in view, the introduction of the norms of vectors and matrices is quite useful. A norm may be introduced in different ways, and in different cases one or other norm will prove to be most convenient.

Generally, the <u>norm of a vector</u> X is an associated non-negative number $\|X\|$ satisfying the following requirements:

1) $\|X\| > 0$ for $X \neq 0$ and $\|0\| = 0$;

2) $\|cX\| = |c|\,\|X\|$ for any numerical multiplier c ;

3) $\|X + Y\| \leq \|X\| + \|Y\|$ (the "triangle inequality") .

From requirements 2) and 3) it is readily deduced that

(1) $$\|X - Y\| \geq \Big|\, \|X\| - \|Y\| \,\Big| \ .$$

Indeed, we have

(2) $$\|X\| = \|X - Y + Y\| \leq \|X - Y\| + \|Y\|$$

and therefore

(3) $$\|X - Y\| \geq \|X\| - \|Y\| \ .$$

But

(4) $$\|X - Y\| = \|Y - X\| \geq \|Y\| - \|X\| \ .$$

Consequently

$$||X - Y|| \geq \Big| ||X|| - ||Y|| \Big| \quad . \tag{5}$$

We shall henceforth make use of the following three ways of assigning a norm: if $X = (x_1, x_2, \ldots, x_n)$,

$$\begin{cases} \text{I.} & ||X||_{I} = \max\limits_{i} |x_i| \\ \text{II.} & ||X||_{II} = |x_1| + |X_2| + \cdots + |x_n| \\ \text{III.} & ||X||_{III} = \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2} \quad . \end{cases} \tag{6}$$

It is obvious that for all three norms all the requirements 1) - 3) are fulfilled.

The concept of the norm of a vector generalizes the concept of the length of a vector, since for length all the requirements 1) - 3) are fulfilled. The third norm introduced by us is indeed none other than the length of the vector.

Furthermore, it is easily established that a necessary and sufficient condition that the sequence of vectors $X^{(k)}$ converge to the vector X is that $||X^{(k)} - X|| \to 0$ for each of the three norms indicated. For the first norm this is obvious. For the second and third norm this follows from the obvious inequalities

$$\begin{aligned} ||X||_I &\leq ||X||_{II} \leq n\, ||X||_I \\ ||X||_I &\leq ||X||_{III} \leq \sqrt{n}\, ||X||_I \quad . \end{aligned} \tag{7}$$

It is easily shown that for convergence of a sequence of vectors $X^{(k)}$ to a vector X it is necessary and sufficient that $||X - X^{(k)}|| \longrightarrow 0$, whatever norm satisfying conditions 1) - 3) we may choose. Here, if $X(k) \longrightarrow X$, $||X^{(k)}|| \longrightarrow ||X||$, for $\left| ||X|| - ||X^{(k)}|| \right| \leq ||X - X^{(k)}|| \longrightarrow 0$.

In an analogous fashion, the norm of a square matrix A is a non-negative number A satisfying the conditions

$$(8) \quad \begin{cases} 1) \ ||A|| > 0 \text{ if } A \neq 0 \text{ and } ||0|| = 0 \ ; \\ 2) \ ||cA|| = |c| \, ||A|| \ ; \\ 3) \ ||A + B|| \leq ||A|| + ||B|| \ ; \\ 4) \ ||AB|| \leq ||A|| \, ||B|| \ . \end{cases}$$

Just as in the case of the norms of vectors, the condition $||A^{(k)} - A|| \longrightarrow 0$ is necessary and sufficient in order that $A^{(k)} \longrightarrow A$, and just as in the case of the norms of vectors, it follows from $A^{(k)} \longrightarrow A$ that $||A^{(k)}|| \longrightarrow ||A||$.

The norm of a matrix may be introduced in an infinite variety of ways. Because in the majority of problems connected with estimates both matrices and vectors appear simultaneously in the reasoning, it is expeditious to introduce the norm of a matrix in such a way that it will be rationally connected with the vector norms employed in the argument in hand. We shall say that the norm of a matrix is compatible with a given norm of vectors if for any matrix A and any vector X the following inequality is satisfied:

$$(9) \qquad ||AX|| \leq ||A|| \, ||X|| \ .$$

We will now indicate a device making it possible to construct the matrix norm so as to render it compatible with a given vector norm, to wit: we shall adopt for the norm of the matrix A the maximum of the norms of the vectors AX on the assumption that the vector X runs over the set of all vectors whose norm equals unity:

$$\|A\| = \max_{\|X\|=1} \|AX\| . \tag{10}$$

In consequence of the continuity of a norm, for each matrix A this maximum is attainable, i.e., a vector X_0 can be found such that $\|X_0\| = 1$ and $\|AX_0\| = \|A\|$.

We shall prove that a norm constructed in such a manner satisfies requirements 1) - 4), set previously, and the compatibility condition.

Let us begin with the verification of the <u>first requirement</u>.

Let $A \neq 0$. Then a vector X, $\|X\| = 1$, can be found such that $AX \neq 0$, and accordingly $\|AX\| \neq 0$. Therefore $\|A\| = \max_{\|X\|=1} \|AX\| \neq 0$. If, however, $A = 0$, $\|A\| = \max_{\|X\|=1} \|0X\| = 0$.

<u>Second requirement</u>. On the strength of the definition, $\|cA\| = \max \|cAX\|$. Obviously $\|cAX\| = |c|\ \|AX\|$ and thus $\|cA\| = \max_{\|X\|=1} |c|\ \|AX\| = |c| \max \|AX\| = |c|\ \|A\|$.

Let us verify, furthermore, the compatibility condition.

Let $Y \neq 0$ be any vector; then $X = \frac{1}{\|Y\|} Y$ will satisfy the condition that $\|X\| = 1$. Consequently $\|AY\| = \|A(\|Y\|X)\| = \|Y\|\ \|AX\| \leq \|Y\|\ \|A\|$.

Third requirement. For the matrix A + B find a vector X_o such that $||A + B|| = ||(A + B)X_o||$ and $||X_o|| = 1$. Then $||A + B|| = ||(A + B)X_o|| =$
$= ||AX_o + BX_o|| \leq ||AX_o|| + ||BX_o|| \leq ||A|| \, ||X_o|| + ||B|| \, ||X_o|| =$
$= ||A|| + ||B||$.

Lastly, the fourth requirement. For the matrix AB find a vector X_o such that $||X_o|| = 1$ and $||ABX_o|| = ||AB||$. Then $||AB|| = ||ABX_o|| =$
$= ||A(BX_o)|| \leq ||A|| \, ||BX_o|| \leq ||A|| \, ||B|| \, ||X_o|| = ||A|| \, ||B||$.

We have verified the satisfaction of all four requirements and the compatibility condition. A matrix norm constructed in this manner we shall speak of as subordinate to the given norm of vectors. It is obvious that for any matrix norm, subordinate to whatsoever vector norm, $||I|| = 1$.

Let us now construct matrix norms subordinate to the three norms of vectors introduced above.

(11) I. $$||X||_1 = \max_i |x_i| .$$

The matrix norm subordinate to this vector norm is

(12) $$||A||_1 = \max_i \sum_{k=1}^{n} |a_{ik}| .$$

In proof, let $||X|| = 1$. Then

(13) $$||AX|| = \max_i \left| \sum_{k=1}^{n} a_{ik} x_k \right| \leq \max_i \sum_{k=1}^{n} |a_{ik}| \, |x_k| \leq \max_i \sum_{k=1}^{n} |a_{ik}| .$$

Consequently

$$(14) \qquad \max_{\|X\|=1} \|AX\| \leq \max_i \sum_{k=1}^{n} |a_{ik}| \quad .$$

We shall now prove that $\max_{\|X\|=1} \|AX\|$ is in fact equal to $\max_i \sum_{k=1}^{n} |a_{ik}|$. For this we shall construct a vector X_o such that $\|X_o\| = 1$ and $\|AX_o\| = \max_i \sum_{k=1}^{n} |a_{ik}|$. Letting $\sum_{k=1}^{n} |a_{ik}|$ attain its greatest value for $i = j$, and then taking as the component $x_k^{(o)}$ of the vector X_o: $x_k^{(o)} = \frac{|a_{jk}|}{a_{jk}}$, if $a_{jk} \neq 0$, and $x_k^{(o)} = 1$ if $a_{jk} = 0$, we have, obviously, $\|X_o\| = 1$. Furthermore,

$$(15) \qquad \left| \sum_{k=1}^{n} a_{ik} x_k^{(o)} \right| \leq \sum_{k=1}^{n} \left| a_{ik} \right| \leq \sum_{k=1}^{n} \left| a_{jk} \right| \qquad \text{for } i \neq j$$

and

$$(16) \qquad \left| \sum_{k=1}^{n} a_{jk} x_k^{(o)} \right| = \sum_{k=1}^{n} |a_{jk}| \quad .$$

Consequently

$$(17) \qquad \max_i \left| \sum_{k=1}^{n} a_{ik} x_k^{(o)} \right| = \sum_{k=1}^{n} |a_{jk}| = \max_i \sum_{k=1}^{n} |a_{ik}| \quad .$$

Thus $\|AX_o\| = \max_i \sum_{k=1}^{n} |a_{ik}|$, Q.E.D.

$$(18) \quad \text{II.} \qquad \|X\|_{II} = \sum_{i=1}^{n} |x_i| \quad .$$

The matrix norm subordinate to this vector norm is

$$(19) \qquad ||A||_{II} = \max_k \sum_{i=1}^{n} |a_{ik}| \ .$$

In proof thereof, let $||X|| = 1$, then

$$||AX|| = \sum_{i=1}^{n} \left| \sum_{k=1}^{n} a_{ik} x_k \right| \leq \sum_{i=1}^{n} \sum_{k=1}^{n} |a_{ik}| \; |x_k| \leq$$

$$(20) \qquad \leq \sum_{k=1}^{n} |x_k| \left(\sum_{i=1}^{n} |a_{ik}| \right) \leq \left(\max_k \sum_{i=1}^{n} |a_{ik}| \right) \sum_{k=1}^{n} |x_k| \leq$$

$$\leq \max_k \sum_{i=1}^{n} |a_{ik}| \ .$$

Now let us take a vector X_o of the following form: let $\sum_{i=1}^{n} |a_{ik}|$ attain its greatest value for the column numbered j. Put $x_k^{(o)} = 0$ for $k \neq j$ and $x_j^{(o)} = 1$. Obviously a vector constructed in this manner has its norm equal to unity. Furthermore

$$(21) \qquad ||AX_o|| = \sum_{i=1}^{n} \left| \sum_{k=1}^{n} a_{ik} x_k^{(o)} \right| = \sum_{i=1}^{n} |a_{ij}| = \max_k \sum_{k=1}^{n} |a_{ik}| \ .$$

Thus

$$(22) \qquad \max ||AX_o|| = \max_k \sum_{i=1}^{n} |a_{ik}| \ ;$$

Q.E.D.

$$(23) \qquad \text{III.} \qquad ||X||_{III}^2 = \sum_{k=1}^{n} |x_k|^2 = (X, X) \ .$$

The matrix norm subordinate to this vector norm is

(24) $$||A||_{III} = \sqrt{\lambda_1} \quad ,$$

where λ_1 is the largest latent root of the matrix A'A.

In proof, we have

(25) $$||A|| = \max_{||X||=1} ||AX|| \quad ;$$

but

(26) $$||AX||_{III}^2 = (AX, AX) = (X, A'AX) \quad .$$

The matrix A'A is symmetric. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be its latent roots and $X_1, X_2, \cdots, X_n$ be the orthonormal system of latent vectors belonging to these latent roots.

Now take any vector X with its norm equal to unity and resolve it in terms of the latent vectors:

(27) $$X = c_1X_1 + c_2X_2 + \cdots + c_nX_n \quad .$$

Then

(28) $$(X, X) = c_1^2 + c_2^2 + \cdots + c_n^2 = 1 \quad .$$

Moreover,

$$||AX||^2 = (X, A'AX) =$$

(29) $$= (c_1X_1 + \cdots + c_nX_n \ , \ c_1\lambda_1X_1 + \cdots + c_n\lambda_nX_n) =$$

$$= \lambda_1c_1^2 + \cdots + \lambda_nc_n^2 \leq \lambda_1(c_1^2 + \cdots + c_n^2) = \lambda_1 \ .$$

For the vector $X = X_1$:

(30) $$||AX_1||^2 = (X_1, A'AX_1) = (X_1, \lambda_1X_1) = \lambda_1 \ .$$

Thus

(31) $$\max_{||X||=1} ||AX|| = \sqrt{\lambda_1} \ ,$$

Q.E.D.

We shall now prove several theorems connected with the concept of limit.

THEOREM 1. <u>In order that $A^m \longrightarrow 0$, it is necessary and sufficient that all the latent roots of the matrix A have modulus less than unity.</u>

Proof. Assume for simplicity that the matrix A can be brought into diagonal form: $A = C \Lambda C^{-1}$, where $\Lambda = \lceil \lambda_1, \lambda_2, \cdots, \lambda_n \rfloor$ and $\lambda_1, \lambda_2, \cdots, \lambda_n$ are the latent roots of matrix A. Then $A^m = C \Lambda^m C^{-1}$. It is obvious that $\Lambda^m = \lceil \lambda_1^m, \lambda_2^m, \cdots, \lambda_n^m \rfloor$. In order that $A^m \longrightarrow 0$, it is necessary and sufficient that $\Lambda^m \longrightarrow 0$, for the which it is in turn necessary and sufficient that the latent roots $\lambda_1, \lambda_2, \cdots, \lambda_n$ have modulus less than unity.

In case the matrix A cannot be brought into diagonal form, the theorem is proved either with the aid of considerations of continuity or by passing to the Jordan canonical form. We shall not dwell on the details of this proof.

The conditions given in Theorem 1 are inconvenient for checks, inasmuch as they require foreknowledge of the latent roots of the matrix A. We shall therefore establish some simpler sufficient conditions rendering $\lim_{m \to \infty} A^m = 0$.

THEOREM 2. _In order that $A^m \to 0$, it is sufficient that any one of the norms of A be less than unity._

Proof. On the strength of the fourth requirement of a norm, we have

$$(32) \qquad ||A^m|| \leq ||A^{m-1}|| \; ||A|| \leq ||A^{m-2}|| \; ||A||^2 \leq \dots \leq ||A||^m \; .$$

Therefore $||A^m|| \to 0$ if $||A|| < 1$, and thus, in view of the foregoing, $A^m \to 0$.

Combining Theorems 1 and 2, we arrive at the following result:

THEOREM 3. _The modulus of no latent root of a matrix exceeds any of its norms._

Proof. Let $||A|| = a$. Consider a matrix $B = \frac{1}{a+\epsilon} A$, where ϵ is any positive number. We have

$$(33) \qquad ||B|| = \frac{a}{a+\epsilon} < 1 \; ,$$

amd accordingly $B^m \to 0$ as $m \to \infty$. On the strength of Theorem 1 its latent roots have modulus less than unity. But it is obvious that the

latent roots of the matrix B equal $\frac{1}{a+\epsilon}\lambda_i$, where λ_i are the latent roots of the matrix A. Thus $\frac{|\lambda_i|}{a+\epsilon} \leq 1$, i.e., $|\lambda_i| < a + \epsilon$. Since ϵ may be taken arbitrarily small, $|\lambda_i| \leq a$.

THEOREM 4. In order that the series

$$I + A + \dots + A^m + \dots \tag{34}$$

converge, it is necessary and sufficient that $A^m \to 0$ as $m \to \infty$. In such a case the sum of series (34) equals $(I - A)^{-1}$.

Proof. The necessity of this condition is obvious. We shall show that it is sufficient.

On the strength of Theorem 1, all latent roots of the matrix A are less than of unit modulus.

Accordingly

$$|I - A| \neq 0 \ , \tag{35}$$

and therefore $(I - A)^{-1}$ exists.

Consider the identity

$$(I + A + A^2 + \dots + A^k)(I - A) = I - A^{k+1} \ . \tag{36}$$

Postmultiplying it by $(I - A)^{-1}$, we obtain

$$I + A + A^2 + \dots + A^k = (I - A)^{-1} - A^{k+1}(I - A)^{-1} \ , \tag{37}$$

whence it follows that, as $k \to \infty$,

(38) $$I + A + \cdots + A^k \longrightarrow (I - A)^{-1}$$

since $A^{k+1} \longrightarrow 0$.

Thus

(39) $$I + A + \cdots + A^k + \cdots = (I - A)^{-1} \quad ,$$

which is what was required to be proved.

In the light of Theorem 1, the necessary and sufficient condition for the convergence of the series (34) is the inequality $|\lambda_i| < 1$ for all latent roots of the matrix A. A sufficient token of convergence, in view of Theorem 2, is the inequality $||A|| < 1$, whatever one of the norms be employed. Given that this condition is satisfied, it is easy to give the following estimate of the rapidity of convergence of the series (34):

THEOREM 5. *If* $||A|| < 1$,

(40) $$||(I - A)^{-1} - (I + A + \cdots + A^k)|| = \frac{||A||^{k+1}}{1 - ||A||} \quad .$$

Proof. We have:

(41) $$(I - A)^{-1} - (I + A + \cdots + A^k) = A^{k+1} + A^{k+2} + \cdots$$

whence

(42) $$\begin{aligned} &||(I - A)^{-1} - (I + A + \cdots + A^k)|| \leq \\ &\leq ||A||^{k+1} + ||A||^{k+2} + \cdots = \frac{||A||^{k+1}}{1 - ||A||} \end{aligned}$$

and the theorem is proved.

5/2/52

in finding a proof for the general case.

We consider ourselves obligated to note in conclusion that in the method here proposed we have utilized the ideas of a method of successive approximations for the solution of systems of linear equations expounded in an unpublished work of A. M. Lopshits.

August 14, 1952

THE NATIONAL BUREAU OF STANDARDS

Functions and Activities

The functions of the National Bureau of Standards are set forth in the Act of Congress, March 3, 1901, as amended by Congress in Public Law 619, 1950. These include the development and maintenance of the national standards of measurement and the provision of means and methods for making measurements consistent with these standards; the determination of physical constants and properties of materials; the development of methods and instruments for testing materials, devices, and structures; advisory services to Government Agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; and the development of standard practices, codes, and specifications. The work includes basic and applied research, development, engineering, instrumentation, testing, evaluation, calibration services and various consultation and information services. A major portion of the Bureau's work is performed for other Government Agencies, particularly the Department of Defense and the Atomic Energy Commission. The scope of activities is suggested by the listing of divisions and sections on the inside of the front cover.

Reports and Publications

The results of the Bureau's work take the form of either actual equipment and devices or published papers and reports. Reports are issued to the sponsoring agency of a particular project or program. Published papers appear either in the Bureau's own series of publications or in the journals of professional and scientific societies. The Bureau itself publishes three monthly periodicals, available from the Government Printing Office: The Journal of Research, which presents complete papers reporting technical investigations; the Technical News Bulletin, which presents summary and preliminary reports on work in progress; and Basic Radio Propagation Predictions, which provides data for determining the best frequencies to use for radio communications throughout the world. There are also five series of nonperiodical publications: The Applied Mathematics Series, Circulars, Handbooks, Building Materials and Structures Reports, and Miscellaneous Publications.

Information on the Bureau's publications can be found in NBS Circular 460, Publications of the National Bureau of Standards ($1.00). Information on calibration services and fees can be found in NBS Circular 483, Testing by the National Bureau of Standards (25 cents). Both are available from the Government Printing Office. Inquiries regarding the Bureau's reports and publications should be addressed to the Office of Scientific Publications, National Bureau of Standards, Washington 25, D. C.

NBS