CHAPTER I

# The Algebra of Linear Transformations and Quadratic Forms

In the present volume we shall be concerned with many topics in mathematical analysis which are intimately related to the theory of linear transformations and quadratic forms. A brief résumé of pertinent aspects of this field will, therefore, be given in Chapter I. The reader is assumed to be familiar with the subject in general.

## §1. *Linear Equations and Linear Transformations*

**1. Vectors.** The results of the theory of linear equations can be expressed concisely by the notation of vector analysis. A system of $n$ real numbers $x_1, x_2, \cdots, x_n$ is called an *n-dimensional vector* or a vector in $n$-dimensional space and denoted by the bold face letter $\mathbf{x}$; the numbers $x_i$ $(i = 1, \cdots, n)$ are called the *components* of the vector $\mathbf{x}$. If all components vanish, the vector is said to be zero or the *null vector*; for $n = 2$ or $n = 3$ a vector can be interpreted geometrically as a "position vector" leading from the origin to the point with the rectangular coordinates $x_i$. For $n > 3$ geometrical visualization is no longer possible but geometrical terminology remains suitable.

Given two arbitrary real numbers $\lambda$ and $\mu$, the vector $\lambda\mathbf{x} + \mu\mathbf{y} = \mathbf{z}$ is defined as the vector whose components $z_i$ are given by $z_i = \lambda x_i + \mu y_i$. Thus. in particular, the sum and difference of two vectors are defined.

The number

$$(1) \qquad \mathbf{x}\cdot\mathbf{y} = x_1y_1 + \cdots + x_ny_n = y_1x_1 + \cdots + y_nx_n = \mathbf{y}\cdot\mathbf{x}$$

is called the "*inner product*" of the vectors $\mathbf{x}$ and $\mathbf{y}$.

Occasionally we shall call the inner product $\mathbf{x}\cdot\mathbf{y}$ the *component of the vector* $\mathbf{y}$ *with respect to* $\mathbf{x}$ or vice versa.

If the inner product $\mathbf{x}\cdot\mathbf{y}$ vanishes we say that the vectors $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal*; for $n = 2$ and $n = 3$ this terminology has an immediate geometrical meaning. The inner product $\mathbf{x}\cdot\mathbf{x} = \mathbf{x}^2$ of a vector with itself plays a special role; it is called the *norm* of the vector. The positive square root of $\mathbf{x}^2$ is called the *length* of the vector and denoted by $|\mathbf{x}| = \sqrt{\mathbf{x}^2}$. A vector whose length is unity is called a *normalized vector* or *unit vector*.

The following inequality is satisfied by the inner product of two vectors $\mathbf{a} = (a_1, \cdots, a_n)$ and $\mathbf{b} = (b_1, \cdots, b_n)$:

$$(\mathbf{a}\cdot\mathbf{b})^2 \leq \mathbf{a}^2\mathbf{b}^2$$

or, without using vector notation,

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leq \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right),$$

where the equality holds if and only if the $a_i$ and the $b_i$ are proportional, i.e. if a relation of the form $\lambda\mathbf{a} + \mu\mathbf{b} = 0$ with $\lambda^2 + \mu^2 \neq 0$ is satisfied.

The proof of this "*Schwarz inequality*"[1] follows from the fact that the roots of the quadratic equation

$$\sum_{i=1}^{n} (a_i x + b_i)^2 = x^2 \sum_{i=1}^{n} a_i^2 + 2x \sum_{i=1}^{n} a_i b_i + \sum_{i=1}^{n} b_i^2 = 0$$

for the unknown $x$ can never be real and distinct, but must be imaginary, unless the $a_i$ and $b_i$ are proportional. The Schwarz inequality is merely an expression of this fact in terms of the discriminant of the equation. Another proof of the Schwarz inequality follows immediately from the identity

$$\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2 - \left(\sum_{i=1}^{n} a_i b_i\right)^2 = \tfrac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} (a_j b_k - a_k b_j)^2.$$

Vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m$ are said to be *linearly dependent* if a set of numbers $\lambda_1, \lambda_2, \cdots, \lambda_m$ (not all equal to zero) exists such that the vector equation

$$\lambda_1\mathbf{x}_1 + \cdots + \lambda_m\mathbf{x}_m = 0$$

is satisfied, i.e. such that all the components of the vector on the left vanish. Otherwise the vectors are said to be *linearly independent*.

The $n$ vectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$ in $n$-dimensional space whose com-

[1] This relation was, as a matter of fact, used by Cauchy before Schwarz.

ponents are given, respectively, by the first, second, $\cdots$, and $n$-th rows of the array

$$\begin{matrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1, \end{matrix}$$

form a system of $n$ linearly independent vectors. For, if a relation $\lambda_1\mathbf{e}_1 + \cdots + \lambda_n\mathbf{e}_n = 0$ were satisfied, we could multiply[1] this relation by $\mathbf{e}_h$ and obtain $\lambda_h = 0$ for every $h$, since $\mathbf{e}_h^2 = 1$ and $\mathbf{e}_h \cdot \mathbf{e}_k = 0$ if $h \neq k$. Thus, systems of $n$ linearly independent vectors certainly exist. However, for any $n + 1$ vectors $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{n+1}$ (in $n$-dimensional space) there is at least one linear equation of the form

$$\mu_1\mathbf{u}_1 + \cdots + \mu_{n+1}\mathbf{u}_{n+1} = 0,$$

with coefficients that do not all vanish, since $n$ homogeneous linear equations

$$\sum_{i=1}^{n+1} u_{ik}\mu_i = 0 \qquad (k = 1, \cdots, n)$$

for the $n + 1$ unknowns $\mu_1, \mu_2, \cdots, \mu_{n+1}$ always have at least one nontrivial solution (cf. subsection 3).

**2. Orthogonal Systems of Vectors. Completeness.** The above "coordinate vectors" $\mathbf{e}_i$ form a particular system of *orthogonal unit vectors*. In general a system of $n$ orthogonal unit vectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$ is defined as a system of vectors of unit length satisfying the relations

$$\mathbf{e}_h^2 = 1, \qquad \mathbf{e}_h \cdot \mathbf{e}_k = 0 \qquad (h \neq k)$$

for $h, k, = 1, 2, \cdots, n$. As above, we see that the $n$ vectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$ are linearly independent.

If $\mathbf{x}$ is an arbitrary vector, a relation of the form

$$c_0\mathbf{x} - c_1\mathbf{e}_1 - \cdots - c_n\mathbf{e}_n = 0$$

with constants $c_i$ that do not all vanish must hold; for, as we have seen, any $n + 1$ vectors are linearly dependent. Since the $\mathbf{e}_i$ are linearly independent, $c_0$ cannot be zero; we may therefore, without

[1] To multiply two vectors is to take their inner product.

loss of generality, take it to be equal to unity. Every vector $\mathbf{x}$ can thus be expressed in terms of a system of orthogonal unit vectors in the form

$$\mathbf{x} = c_1\mathbf{e}_1 + \cdots + c_n\mathbf{e}_n. \tag{2}$$

The coefficients $c_i$, the *components of* $\mathbf{x}$ *with respect to the system* $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$, may be found by multiplying (2) by each of the vectors $\mathbf{e}_i$; they are

$$c_i = \mathbf{x}\cdot\mathbf{e}_i.$$

From any arbitrary system of $m$ linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$, we may, by the following *orthogonalization process* due to E. Schmidt, obtain a system of $m$ orthogonal unit vectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_m$: First set $\mathbf{e}_1 = \mathbf{v}_1/|\mathbf{v}_1|$. Then choose a number $c_1'$ in such a way that $\mathbf{v}_2 - c_1'\mathbf{e}_1$ is orthogonal to $\mathbf{e}_1$, i.e. set $c_1' = \mathbf{v}_2\cdot\mathbf{e}_1$. Since $\mathbf{v}_1$ and $\mathbf{v}_2$, and therefore $\mathbf{e}_1$ and $\mathbf{v}_2$, are linearly independent, the vector $\mathbf{v}_2 - c_1'\mathbf{e}_1$ is different from zero. We may then divide this vector by its length obtaining a unit vector $\mathbf{e}_2$ which is orthogonal to $\mathbf{e}_1$. We next find two numbers $c_1''$, $c_2''$ such that $\mathbf{v}_3 - c_1''\mathbf{e}_1 - c_2''\mathbf{e}_2$ is orthogonal to both $\mathbf{e}_1$ and $\mathbf{e}_2$, i.e. we set $c_1'' = \mathbf{v}_3\cdot\mathbf{e}_1$ and $c_2'' = \mathbf{v}_3\cdot\mathbf{e}_2$. This vector is again different from zero and can, therefore, be normalized; we divide it by its length and obtain the unit vector $\mathbf{e}_3$. By continuing this procedure we obtain the desired orthogonal system.

For $m < n$ the resulting orthogonal system is called *incomplete*, and if $m = n$ we speak of a *complete orthogonal system*. Let us denote the components of a vector $\mathbf{x}$ with respect to $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_m$ by $c_1, c_2, \cdots, c_m$ as before. The self-evident inequality

$$(\mathbf{x} - c_1\mathbf{e}_1 - \cdots - c_m\mathbf{e}_m)^2 \geq 0$$

is satisfied. Evaluating the left side term by term according to the usual rules of algebra (which hold for vectors if the inner product of two vectors is used whenever two vectors are multiplied), we find

$$\mathbf{x}^2 - 2\mathbf{x}\cdot\sum_{i=1}^{m} c_i\mathbf{e}_i + \sum_{i=1}^{m} c_i^2 = \mathbf{x}^2 - 2\sum_{i=1}^{m} c_i^2 + \sum_{i=1}^{m} c_i^2 \geq 0$$

or

$$\mathbf{x}^2 \geq \sum_{i=1}^{m} c_i^2, \tag{3}$$

where $m \leq n$ and $c_i = \mathbf{x} \cdot \mathbf{e}_i$; the following equality holds for $m = n$:

$$\mathbf{x}^2 = \sum_{i=1}^{m} c_i^2. \tag{4}$$

Relations (3) and (4)—(4) expresses the theorem of Pythagoras in vector notation—have an intuitive significance for $n \leq 3$; they are called, respectively, *Bessel's inequality* and the *completeness relation*. Relation (4), if it is satisfied for every vector $\mathbf{x}$, does in fact indicate that the given orthogonal system is complete since (4) could not be satisfied for a unit vector orthogonal to all vectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_m$, and such a vector necessarily exists if $m < n$.

The completeness relation may also be expressed in the more general form

$$\mathbf{x} \cdot \mathbf{x}' = \sum_{i=1}^{m} c_i c_i', \tag{5}$$

which follows from the orthogonality of the $\mathbf{e}_i$.

So far these algebraic relations are all purely formal. Their significance lies in the fact that they occur again in a similar manner in transcendental problems of analysis.

**3. Linear Transformations. Matrices.** A system of $n$ linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= y_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= y_2, \\ &\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= y_n, \end{aligned} \tag{6}$$

with coefficients $a_{ik}$, assigns a unique set of quantities $y_1, y_2, \cdots, y_n$ to every set of quantities $x_1, x_2, \cdots, x_n$. Such an assignment is called a *linear transformation* of the set $x_1, x_2, \cdots, x_n$ into the set $y_1, y_2, \cdots, y_n$, or, briefly, of the vector $\mathbf{x}$ into the vector $\mathbf{y}$. This transformation is clearly linear since the vector $\lambda_1\mathbf{y}_1 + \lambda_2\mathbf{y}_2$ corresponds to the vector $\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2$.

The most important problem in connection with linear transformations is the problem of inversion, the question, in other words, of the

existence of a solution of a system of linear equations. The answer is given by the following fundamental theorem of the theory of linear equations, whose proof we assume to be known:
*For the system of equations*

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

*or, briefly,*

$$\sum_{k=1}^{n} a_{ik}x_k = y_i \qquad (i = 1, \cdots, n), \tag{7}$$

*with given coefficients* $a_{ik}$, *the following alternative holds: Either it has one and only one solution* $\mathbf{x}$ *for each arbitrarily given vector* $\mathbf{y}$, *in particular the solution* $\mathbf{x} = 0$ *for* $\mathbf{y} = 0$; *or, alternatively, the homogeneous equations arising from* (7) *for* $\mathbf{y} = 0$ *have a positive number* $\rho$ *of nontrivial (not identically zero) linearly independent solutions* $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_\rho$, *which may be assumed to be normalized. In the latter case the "transposed" homogeneous system of equations*

$$\sum_{k=1}^{n} a'_{ik}x'_k = 0 \qquad (i = 1, \cdots, n), \tag{8}$$

*where* $a'_{ik} = a_{ki}$, *also has exactly* $\rho$ *linearly independent nontrivial solutions* $\mathbf{x}'_1, \mathbf{x}'_2, \cdots, \mathbf{x}'_\rho$. *The inhomogeneous system* (7) *then possesses solutions for just those vectors* $\mathbf{y}$ *which are orthogonal to* $\mathbf{x}'_1, \mathbf{x}'_2, \cdots, \mathbf{x}'_\rho$. *These solutions are determined only to within an additive term which is an arbitrary solution of the homogeneous system of equations, i.e. if* $\mathbf{x}$ *is a solution of the inhomogeneous system and* $\mathbf{x}_\sigma$ *is any solution of the homogeneous system, then* $\mathbf{x} + \mathbf{x}_\sigma$ *is also a solution of the inhomogeneous system.*

In this formulation of the fundamental theorem reference to the theory of determinants has been avoided. Later, to obtain explicit expressions for the solutions of the system of equations, determinants will be required.

The essential features of such a linear transformation are contained in the array of coefficients or *matrix* of the equations (7):

$$(9) \qquad A = (a_{ik}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

with the determinant

$$\mathsf{A} = | a_{ik} | = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

It is sometimes useful to denote the transformation itself (also called *tensor*[1] or *operator*) by a special letter **A**. The elements $a_{ik}$ of the matrix $A$ are called the components of the tensor. The linear transformation (7) may be regarded as a "*multiplication*" of the tensor **A** by the vector **x**, written symbolically in the form

$$\mathbf{Ax} = \mathbf{y}.$$

Many results in the algebra of linear transformations may be expressed concisely in terms of matrices or tensors, once certain simple rules and definitions known as *matrix algebra* have been introduced. First we define matrix multiplication; this concept arises if we suppose that the vector **x**, which is transformed in equations (7), is itself the product of a tensor **B** with components $b_{ik}$ and another vector **w**:

$$\sum_{j=1}^{n} b_{kj} w_j = x_k \qquad (k = 1, \cdots, n).$$

Multiplying **w** by a tensor **C** we obtain the vector **y**. The matrix $C$ which corresponds to the tensor **C** is obtained from $A$ and $B$ by the rule of *matrix multiplication*, $C = AB$, which states that the element $c_{ij}$ is the inner product of the $i$-th row of $A$ and the $j$-th column of $B$:

$$(10) \qquad c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \qquad (i, j = 1, \cdots, n).$$

[1] In modern usage the term "operator" is customary to denote linear transformations.

The tensor or transformation **C** is therefore called the inner product or simply the *product of the tensors or transformations* **A** and **B**. Henceforth tensors and the equivalent matrices will not be distinguished from each other. Note that matrix products obey the *associative law*

$$(AB)C = A(BC),$$

so that the product $A_1A_2 \cdots A_h$ of any number of matrices written in a fixed order has a unique meaning. For $A_1 = A_2 = \cdots = A_h = A$ we write this product as the $h$-th power $A^h$ of the matrix $A$. It is, on the other hand, essential to note that the *commutative law* of multiplication is in general *not* valid; $AB$, in other words, differs in general from $BA$. Finally the matrix $\lambda A + \mu B$ is defined as the matrix whose elements are $\lambda a_{ik} + \mu b_{ik}$; thus the null matrix is the matrix in which all components vanish.[1] The validity of the *distributive law*

$$(A + B)C = AC + BC$$

is immediately evident.

The *unit matrix* is defined by

$$E = (e_{ik}) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

It is characterized by the fact that the equation

$$AE = EA = A$$

holds for an arbitrary matrix $A$. The unit matrix corresponds to the *identity transformation*

$$x_i = y_i \qquad (i = 1, \cdots, n).$$

The zero-th power of every matrix $A$ is defined as the unit matrix:

$$A^0 = E.$$

[1] Note that in matrix algebra it does not necessarily follow from the matrix equation $AB = (0)$ that one of the two factors vanishes, as can be seen from the example $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$.

Since the powers $A^k$ of a matrix are defined, we can also define polynomials whose argument is a matrix. Thus, if

$$f(x) = a_0 + a_1x + \cdots + a_mx^m$$

is a polynomial of the $m$-th degree in the variable $x$, then $f(A)$ is defined by

$$f(A) = a_0E + a_1A + \cdots + a_mA^m$$

as a (symbolic) polynomial in the matrix $A$. This definition of a matrix as a function $f(A)$ of $A$ can even, on occasion, be extended to functions which are not polynomials but which can be expressed as power series. The matrix $e^A$, for example, may be defined by

$$B = e^A = E + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots = \sum_{\nu=0}^{\infty} \frac{A^\nu}{\nu!}.$$

Note that in such a series one first considers the sum of the first $N$ terms and then investigates whether each of the $n^2$ elements of the resulting matrix converges to a limit with increasing $N$; if this is the case, the matrix formed from the $n^2$ limiting values is considered to be the sum of the series. In the particular case of the matrix $e^A$ it turns out, as will be shown below, that the series always converges.

A particularly important relation is obtained for a matrix $S$ defined by a *geometric series* with partial sums $S_m$ given by

$$S_m = E + A + A^2 + \cdots + A^m.$$

Multiplying the equation which defines $S_m$ by $A$, we obtain the equation

$$S_mA + E = S_m + A^{m+1},$$

from which it follows that

$$S_m(E - A) = E - A^{m+1}.$$

Now if the matrix $S_m$ approaches a limit $S$ with increasing $m$, so that $A^{m+1}$ tends to zero, we obtain the relation

$$S(E - A) = E$$

for the matrix $S$ defined by the infinite geometric series

$$S = E + A + A^2 + \cdots = \sum_{\nu=0}^{\infty} A^\nu.$$

Under what circumstances an infinite geometric series of matrices or a *Neumann series*, as it is occasionally called, converges will be investigated in the next section.

Matrix polynomials may be handled very much like ordinary polynomials in $x$. For example, an identity between two polynomials in $x$ implies the corresponding identity for an arbitrary matrix $A$. Thus the identity

$$x^3 + 2x^2 + 3x + 4 \equiv (x^2 + 1)(x + 2) + (2x + 2)$$

corresponds to the relation

$$A^3 + 2A^2 + 3A + 4E \equiv (A^2 + E)(A + 2E) + (2A + 2E)$$

valid for every matrix $A$. The factorization

$$f(x) = a_0 + a_1 x + \cdots + a_m x^m = a_m(x - x_1)(x - x_2) \cdots (x - x_m),$$

where $x_1, x_2, \cdots, x_m$ are the zeros of the polynomial $f(x)$, leads to the matrix equation

$$\begin{aligned} f(A) &= a_0 E + a_1 A + \cdots + a_m A^m \\ &= a_m(A - x_1 E)(A - x_2 E) \cdots (A - x_m E) \end{aligned}$$

for every matrix $A$.

Every matrix $A$ with components $a_{ik}$, which may in general be complex, is associated with certain other matrices. If $\bar{a}_{ik}$ is the complex number conjugate to $a_{ik}$, then the matrix $\bar{A} = (\bar{a}_{ik})$ is called the *conjugate* matrix; the matrix $A' = (a_{ki})$ obtained by interchanging corresponding rows and columns of $A$ is called the *transposed* matrix or the *transpose* of $A$ and $A^* = \bar{A}' = (\bar{a}_{ki})$ the *conjugate transpose* of $A$. The conjugate transpose is thus obtained by replacing the elements by their complex conjugates and interchanging rows and columns.

The equation

$$(AB)' = B'A'$$

is immediately verifiable. A matrix for which $A = A'$ is called *symmetric*; a real matrix which satisfies

$$AA' = E$$

is called *orthogonal*. Finally, a complex matrix is called *unitary* if it satisfies

$$AA^* = E.$$

The inversion of the linear transformation (7) is possible for arbitrary $y_i$, as is known from the theory of determinants, if and only if the determinant $\mathsf{A} = | a_{ik} |$ does not vanish. In this case the solution is uniquely determined and is given by a corresponding transformation

$$x_i = \sum_{k=1}^{n} \check{a}_{ik} y_k \qquad (i = 1, \cdots, n). \tag{11}$$

The coefficients $\check{a}_{ik}$ are given by

$$\check{a}_{ik} = \frac{\mathsf{A}_{ki}}{\mathsf{A}} \tag{12}$$

where $\mathsf{A}_{ki}$ is the cofactor to the element $a_{ki}$ in the matrix $A$. The matrix $\check{A} = (\check{a}_{ik})$ is called the *reciprocal* or *inverse of* $A$ and is distinguished by the fact that it satisfies

$$A\check{A} = \check{A}A = E.$$

We denote this uniquely determined matrix by $A^{-1}$ instead of $\check{A}$; the determinant of $A^{-1}$ has the value $\mathsf{A}^{-1}$. Thus the solution of a system of equations whose matrix $A$ has a nonvanishing determinant is characterized, in the language of matrix algebra, by a matrix $B = A^{-1}$ which satisfies the relations

$$AB = BA = E.$$

**4. Bilinear, Quadratic, and Hermitian Forms.** To write the linear equations (7) concisely we may employ the *bilinear form* which corresponds to the matrix $A$. This bilinear form

$$A(u, x) = \sum_{i,k=1}^{n} a_{ik} u_i x_k \tag{13}$$

is obtained by multiplying the linear forms in $x_1, x_2, \cdots, x_n$ on the left-hand side in equation (7) by undetermined quantities $u_1, u_2, \cdots, u_n$ and adding. In this way we obtain from the system of equations (7) the single equation

$$A(u, x) = E(u, y) \tag{14}$$

valid for all $u$; here $E(u, y) = \sum_{i=1}^{n} u_i y_i$ is the bilinear form corresponding to the unit matrix, the *unit bilinear form.* The *symbolic product* of two bilinear forms $A(u, x)$ and $B(u, x)$ with matrices $A$ and $B$ is defined as the bilinear form $C(u, x)$ with the matrix $C = AB$; the $h$-th power $A^h(u, x)$ is often called the *h-fold iterated form.* The "*reciprocal bilinear form*" $A^{-1}(u, x)$ with the matrix $A^{-1}$ may, according to the theory of determinants, be written in the form

$$A^{-1}(u, x) = -\frac{\mathsf{A}(u, x)}{\mathsf{A}}, \tag{15}$$

where

$$\mathsf{A}(u, x) = \begin{vmatrix} 0 & u_1 & \cdots & u_n \\ x_1 & a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ x_n & a_{n1} & \cdots & a_{nn} \end{vmatrix} = -\sum_{i,k=1}^{n} \mathsf{A}_{ik} x_i u_k .$$

The *symmetric linear transformations*, characterized by the condition $a_{ik} = a_{ki}$, are of special interest. To investigate them it is sufficient to consider the *quadratic form*

$$A(x, x) = \sum_{i,k=1}^{n} a_{ik} x_i x_k \qquad (a_{ki} = a_{ik})$$

which is obtained from the bilinear form by putting $u_i = x_i$. For, from a quadratic form $A(x, x)$ one can obtain a symmetric bilinear form

$$\sum_{i,k=1}^{n} a_{ik} u_i x_k = \frac{1}{2} \sum_{i=1}^{n} u_i \frac{\partial A(x, x)}{\partial x_i}$$
$$= \frac{A(x + u, x + u) - A(x, x) - A(u, u)}{2},$$

which is called the *polar form* corresponding to the quadratic form $A(x, x)$.

If $A(u, x) = \sum_{i,k=1}^{n} a_{ik} u_i x_k$ is an arbitrary nonsymmetric bilinear form (with real coefficients), then $AA'(u, x)$ and $A'A(u, x)$ are always symmetric bilinear forms; specifically we have

$$AA'(u, x) = \sum_{k=1}^{n} \left( \sum_{i=1}^{n} a_{ik} x_i \sum_{j=1}^{n} a_{jk} u_j \right)$$

$$A'A(u, x) = \sum_{i=1}^{n} \left( \sum_{k=1}^{n} a_{ik} x_k \sum_{j=1}^{n} a_{ij} u_j \right).$$

The corresponding quadratic forms

$$AA'(x, x) = \sum_{k=1}^{n} \left( \sum_{i=1}^{n} a_{ik} x_i \right)^2,$$

$$A'A(x, x) = \sum_{i=1}^{n} \left( \sum_{k=1}^{n} a_{ik} x_k \right)^2,$$

which are sums of squares, assume only non-negative values. Forms of this kind are called *positive definite* quadratic forms.

An important generalization of the quadratic form is the *Hermitian form.* A Hermitian form is a bilinear form

$$A(u, x) = \sum_{i,k=1}^{n} a_{ik} u_i x_k$$

whose coefficients $a_{ik}$ have complex values subject to the condition

$$a_{ik} = \bar{a}_{ki}\,.$$

Thus a Hermitian form assumes real values if the variables $u_i$ are taken to be the complex conjugates of $x_i$; it is usually written in the form

$$H(x, \bar{x}) = \sum_{i,k=1}^{n} a_{ik} x_i \bar{x}_k = \sum_{i,k=1}^{n} a_{ki} \bar{x}_i x_k\,.$$

To an arbitrary bilinear form

$$A(u, x) = \sum_{i,k=1}^{n} a_{ik} u_i x_k$$

with complex coefficients there correspond the Hermitian forms

$$AA^*(x, \bar{x}) = A\bar{A}'(x, \bar{x}) = \sum_{k=1}^{n} \left| \sum_{i=1}^{n} a_{ik} x_i \right|^2$$

and

$$A^*A(x, \bar{x}) = \bar{A}'A(x, \bar{x}) = \sum_{i=1}^{n} \left| \sum_{k=1}^{n} a_{ik} \bar{x}_k \right|^2.$$

If the variables of a bilinear form

$$A(x, y) = \sum_{i,k=1}^{n} a_{ik} x_i y_k$$

are subjected to the two transformations

$$x_i = \sum_{j=1}^{n} c_{ij}\zeta_j \qquad \text{and} \qquad y_k = \sum_{l=1}^{n} b_{kl}\eta_l$$

with matrices $C$ and $B$, respectively, we obtain

$$A(x, y) = \sum_{i,k=1}^{n} a_{ik}x_iy_k = \sum_{i,j,k,l=1}^{n} a_{ik}c_{ij}b_{kl}\zeta_j\eta_l$$

$$= \sum_{j,l=1}^{n} p_{jl}\zeta_j\eta_l; \qquad p_{jl} = \sum_{i,k=1}^{n} a_{ik}c_{ij}b_{kl}.$$

Thus $A$ is transformed into a bilinear form with the matrix

$$(p_{jl}) = C'AB,$$

whose determinant is, according to the theorem on the multiplication of determinants, equal to $\mathsf{AB\Gamma}$. In particular, if $A$ is a quadratic form

$$K(x, x) = \sum_{p,q=1}^{n} k_{pq}x_px_q$$

with the symmetric matrix $K = (k_{pq})$ and the determinant $\mathsf{K} = | k_{pq} |$, and if we set $C = B$, and transform the variables $x$ we obtain a quadratic form with the symmetric matrix $C'KC$ whose determinant is $\mathsf{K\Gamma}^2$.

**5. Orthogonal and Unitary Transformations.** We now consider the problem of finding "*orthogonal*" linear transformations $L$

$$x_p = \sum_{q=1}^{n} l_{pq}y_q = L_p(y) \qquad (p = 1, \cdots, n), \tag{16}$$

with the real matrix $L = (l_{pq})$ and the determinant $\Lambda = | l_{pq} |$, i.e. transformations which transform the *unit quadratic form*

$$E(x, x) = \sum_{p=1}^{n} x_p^2$$

into itself, thus satisfying the relation

$$E(x, x) = E(y, y) \tag{17}$$

for arbitrary $y$.

Applying our rules of transformation to the quadratic form

$A(x, x) = E(x, x)$, we find that requirement (17) yields the equations

$$L'EL = L'L = LL' = E; \qquad L' = L^{-1} \tag{18}$$

as a necessary and sufficient condition for the orthogonality of $L$. Thus the transposed matrix of an orthogonal transformation is identical with its reciprocal matrix; therefore the solution of equations (16) is given by the transformation

$$y_p = \sum_{q=1}^{n} l_{qp}x_q = L'_p(x), \tag{19}$$

which is likewise orthogonal. We see that an orthogonal transformation is one whose matrix is orthogonal as defined in subsection 3. Written out in detail, the orthogonality conditions become

$$\sum_{\nu=1}^{n} l_{\nu p}^2 = 1, \qquad \sum_{\nu=1}^{n} l_{\nu p}l_{\nu q} = 0 \qquad (p \neq q) \tag{20}$$

or, equivalently,

$$\sum_{\nu=1}^{n} l_{p\nu}^2 = 1, \qquad \sum_{\nu=1}^{n} l_{p\nu}l_{q\nu} = 0 \qquad (p \neq q). \tag{21}$$

To express an orthogonal transformation in vector notation we prescribe a system of $n$ orthogonal unit vectors $\mathbf{l}_1, \mathbf{l}_2, \cdots, \mathbf{l}_n$ into which the coordinate vectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$ are to be transformed. Then the vector $\mathbf{x}$ is represented by

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n = y_1\mathbf{l}_1 + y_2\mathbf{l}_2 + \cdots + y_n\mathbf{l}_n .$$

Multiplying by $\mathbf{e}_p$ we obtain $x_p = \sum_{q=1}^{n} y_q(\mathbf{e}_p\mathbf{l}_q)$; hence

$$l_{pq} = \mathbf{e}_p \cdot \mathbf{l}_q .$$

From (18) it follows that $\Lambda^2 = 1$, i.e. that the determinant of an orthogonal transformation is either $+1$ or $-1$. Therefore the determinant of an arbitrary quadratic form is invariant with respect to orthogonal transformations.

Furthermore, the relation $L'(AB)L = (L'AL)(L'BL)$ follows from (18) for the matrices $A$, $B$, and $L$ of any two bilinear forms and any orthogonal transformation. This means that the symbolic product of a number of bilinear forms may be transformed orthogonally by subjecting each factor to the same orthogonal transformation. In particular, it follows that the orthogonal transforms of two reciprocal forms are also reciprocal.

The generalization of these considerations to Hermitian forms

$$H(x, \bar{x}) = \sum_{p,q=1}^{n} h_{pq} x_p \bar{x}_q$$

leads to *unitary transformations*. A unitary transformation

$$x_p = \sum_{q=1}^{n} l_{pq} y_q \qquad (p = 1, \cdots n)$$

is defined as a transformation (with complex coefficients $l_{pq}$) which transforms the *unit Hermitian form*

$$\sum_{p=1}^{n} | x_p |^2 = \sum_{p=1}^{n} x_p \bar{x}_p$$

into itself, i.e. for which

$$\sum_{p=1}^{n} | x_p |^2 = \sum_{p=1}^{n} | y_p |^2.$$

In exactly the same way as above one obtains the matrix equation

$$LL^* = L^*L = E$$

as a necessary and sufficient condition for the unitary character of the transformation whose matrix is $L$. Here $L^* = \bar{L}'$ is the conjugate transpose of $L$. $L$ must therefore be a unitary matrix as defined in subsection 3. Specifically, a transformation is unitary if the following conditions hold:

$$(22) \qquad \sum_{\nu=1}^{n} | l_{\nu p} |^2 = 1, \qquad \sum_{\nu=1}^{n} l_{\nu p} \bar{l}_{\nu q} = 0 \qquad (p \neq q),$$

or, equivalently,

$$(23) \qquad \sum_{\nu=1}^{n} | l_{p\nu} |^2 = 1, \qquad \sum_{\nu=1}^{n} l_{p\nu} \bar{l}_{q\nu} = 0 \qquad (p \neq q).$$

The determinant of a unitary transformation has the absolute value 1, as follows immediately from the equation $LL^* = E$.

## §2. *Linear Transformations with a Linear Parameter*

In many problems the system of equations of a linear transformation takes the form

$$x_i - \lambda \sum_{k=1}^{n} t_{ik} x_k = y_i \qquad (i = 1, \cdots, n) \tag{24}$$

where $\lambda$ is a parameter (in general complex). The corresponding bilinear form is $E(u, x) - \lambda T(u, x)$, where $T(u, x)$ is the form whose matrix is $(t_{ik})$. As we have seen in the preceding section, the problem of solving the system of equations (24) is equivalent to the problem of finding the reciprocal bilinear form $R(u, y; \lambda)$ with the matrix $R$ which satisfies the equation $(E - \lambda T)R = E$. We know that this reciprocal matrix $R$ exists if and only if the determinant $|E - \lambda T|$ is different from zero.

Let us consider the zeros of the determinant $|E - \lambda T|$ or, equivalently, for $\kappa = 1/\lambda \neq 0$, the zeros of the determinant $|T - \kappa E|$. Clearly, $|T - \kappa E|$ is a polynomial in $\kappa$ of the $n$-th degree. Therefore there exist $n$ values of $\kappa$ (namely the zeros of the polynomial) for which the form $R(u, y; \lambda)$ fails to exist. These values $\kappa_i$ are known as the "*characteristic values*," "*proper values*," or "*eigenvalues*" of $T$ with respect to the unit matrix $E$; they form the so-called "*spectrum*" of the matrix $T$.[1]

The particular form of equations (24) suggests a solution by iteration: In the equation

$$x_i = y_i + \lambda \sum_{k=1}^{n} t_{ik} x_k$$

we substitute for the quantities $x_k$ on the right the expressions

$$y_k + \lambda \sum_{j=1}^{n} t_{kj} x_j,$$

and then again repeat this substitution. The procedure is conveniently described if we write $R = E + \lambda TR$ and continue:

$$\begin{aligned} R &= E + \lambda TR = E + \lambda T + \lambda^2 T^2 R \\ &= E + \lambda T + \lambda^2 T^2 + \lambda^3 T^3 R = \cdots . \end{aligned}$$

[1] Sometimes the set of values $\lambda_i = 1/\kappa_i$, for which no reciprocal of $E - \lambda T$ exists, is called the spectrum. We shall call this the "reciprocal spectrum" and the $\lambda_i$ the "reciprocal eigenvalues."

We thus obtain an expression for $R$ as an infinite series

$$R = E + \lambda T + \lambda^2 T^2 + \lambda^3 T^3 + \cdots,$$

which—assuming that it converges—represents the reciprocal matrix of $E - \lambda T$. To see this we simply multiply the series by $E - \lambda T$ and remember that symbolic multiplication may be carried out term by term provided the result converges. It is then immediately clear that the representation

$$R = (E - \lambda T)^{-1} = E + \lambda T + \lambda^2 T^2 + \lambda^3 T^3 + \cdots$$

is, formally, completely equivalent to the ordinary geometric series. (Compare the discussion of geometric series on page 9, where we need only put $A = \lambda T$ to obtain equivalence.)

Let us now represent our original system of equations using bilinear forms instead of the corresponding matrices:

$$E(u, x) - \lambda T(u, x) = E(u, y).$$

We may write the solution of this equation in the form

$$E(u, y) + \lambda \mathsf{T}(u, y; \lambda) = E(u, x),$$

which is completely symmetric to it; here

$$\mathsf{T}(u, y; \lambda) = T + \lambda T^2 + \lambda^2 T^3 + \cdots$$
$$= \frac{R(u, y; \lambda) - E(u, y)}{\lambda}.$$

The form $\mathsf{T}$ is called the *resolvent* of $T$.

The convergence of the above *Neumann series* for $R$ or $\mathsf{T}$ for sufficiently small $|\lambda|$ is easily proved: If $M$ is an upper bound of the absolute values of the numbers $t_{ik}$, it follows immediately that upper bounds for the absolute values of the coefficients of the forms $T^2$, $T^3, \cdots, T^h$ are given by $nM^2, n^2M^3, \cdots, n^{h-1}M^h$. Thus

$$(M + \lambda n M^2 + \lambda^2 n^2 M^3 + \cdots)$$
$$\cdot (|u_1| + |u_2| + \cdots + |u_n|)(|y_1| + \cdots + |y_n|)$$

is a majorant of the Neumann series for $\mathsf{T}(u, y; \lambda)$; it is certainly convergent for $|\lambda| < 1/nM$. Therefore our Neumann series also converges for sufficiently small $|\lambda|$ and actually represents the resolvent of $T(u, x)$.[1]

The above estimate proves, incidentally, that in any everywhere convergent power series $f(x) = \sum_{\nu=0}^{\infty} c_\nu x^\nu$ we may replace $x$ by an arbitrary matrix $A$ and obtain a new matrix $f(A) = \sum_{\nu=0}^{\infty} c_\nu A^\nu$. Thus, in particular, the matrix $e^A$ always exists.

While the above expression for $R$ or T converges only for sufficiently small $|\lambda|$, we may obtain from equation (15) of the previous section an expression for the reciprocal form or matrix $R = (E - \lambda T)^{-1}$ which retains its meaning even outside the region of convergence. For, if we identify the form $E - \lambda T$ with the form $A(u, x)$, we immediately obtain, for the reciprocal form,

$$R(u, y; \lambda) = -\frac{\Delta(u, y; \lambda)}{\Delta(\lambda)}$$

[1] The convergence of the majorant obtained above evidently becomes worse with increasing $n$. It may, however, be pointed out that, by slightly refining the argument, an upper bound for the coefficients of the form T can be obtained which is independent of $n$ and which, therefore, can be used for the generalization to infinitely many variables. We denote the elements of the matrix $T^\nu$ by $t_{ik}^{(\nu)}$ and set

$$\sum_{\alpha=1}^{n} | t_{p\alpha}^{(1)} | = z_p .$$

Then, if $\bar{M}$ is an upper bound for all the $n$ quantities $z_p$, it follows, as will be shown below by induction, that

$$\sum_{q=1}^{n} | t_{pq}^{(\nu)} | \leq \bar{M}^\nu;$$

therefore,

$$| t_{pq}^{(\nu)} | \leq \bar{M}^\nu$$

for $p, q = 1, 2, \cdots, n$ and every $\nu$. From this we see that our Neumann series converges for $|\lambda| < 1/\bar{M}$. We thus have a bound which does not depend on $n$ explicitly.

To prove the above inequality for arbitrary $\nu$ we assume it to be proved for the index $\nu - 1$; we then have

$$\sum_{q=1}^{n} | t_{pq}^{(\nu)} | = \sum_{q=1}^{n}{}' \left| \sum_{\alpha=1}^{n} t_{p\alpha}^{(1)} t_{\alpha q}^{(\nu-1)} \right| \leq \sum_{q=1}^{n} \sum_{\alpha=1}^{n} | t_{p\alpha}^{(1)} | \, | t_{\alpha q}^{(\nu-1)} |$$
$$= \sum_{\alpha=1}^{n} | t_{p\alpha}^{(1)} | \left( \sum_{q=1}^{n} | t_{\alpha q}^{(\nu-1)} | \right) \leq \bar{M}^{\nu-1} \cdot \sum_{\alpha=1}^{n} | t_{p\alpha}^{(1)} | \leq \bar{M}^\nu.$$

Since the inequality is valid for $\nu = 1$, it is proved for arbitrary $\nu$.

and, for the resolvent $\mathsf{T}$,

$$\mathsf{T}(u, y; \lambda) = -\frac{\Delta(u, y; \lambda)}{\lambda\Delta(\lambda)} - \frac{1}{\lambda} E(u, y),$$

where

$$\Delta(u, y; \lambda) = \begin{vmatrix} 0 & u_1 & \cdots & u_n \\ y_1 & 1 - \lambda t_{11} & \cdots & -\lambda t_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ y_n & -\lambda t_{n1} & \cdots & 1 - \lambda t_{nn} \end{vmatrix}$$

and

$$\Delta(\lambda) = \begin{vmatrix} 1 - \lambda t_{11} & -\lambda t_{12} & \cdots & -\lambda t_{1n} \\ -\lambda t_{21} & 1 - \lambda t_{22} & \cdots & -\lambda t_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -\lambda t_{n1} & -\lambda t_{n2} & \cdots & 1 - \lambda t_{nn} \end{vmatrix}$$

are polynomials in $\lambda$ of at most the $(n - 1)$-st and $n$-th degree. Thus the zeros of $\Delta(\lambda)$ form the reciprocal spectrum of the form $T$ as defined above, i.e. the totality of values of $\lambda$ for which the form $E - \lambda T$ has no reciprocal.

By means of the formula

$$T + \lambda T^2 + \lambda^2 T^3 + \cdots = -\frac{\Delta(u, y; \lambda)}{\lambda\Delta(\lambda)} - \frac{1}{\lambda} E(u, y)$$

the series on the left, which does not converge for all $\lambda$, is continued analytically into the entire $\lambda$-plane. The reciprocal form $R$, as well as the resolvent $\mathsf{T}$, is a rational function of $\lambda$ whose poles are given by the spectrum of the form $T$.

If we expand the determinants $\Delta(u, y; \lambda)$ and $\Delta(\lambda)$ in powers of $\lambda$, we obtain the expressions

$$\Delta(u, y; \lambda) = \Delta_1(u, y) - \lambda\Delta_2(u, y) + \lambda^2\Delta_3(u, y) - \cdots + (-1)^n\lambda^{n-1}\Delta_n(u, y),$$

$$\Delta(\lambda) = 1 - \lambda\Delta_1 + \lambda^2\Delta_2 - \cdots + (-1)^n\lambda^n\Delta_n ,$$

where

$$\Delta_h(u, y) = \sum \begin{vmatrix} 0 & u_{p_1} & \cdots & u_{p_h} \\ y_{p_1} & t_{p_1p_1} & \cdots & t_{p_1p_h} \\ \cdots & \cdots & \cdots & \cdots \\ y_{p_h} & t_{p_hp_1} & \cdots & t_{p_hp_h} \end{vmatrix}$$

and

$$\Delta_h = \sum \begin{vmatrix} t_{p_1p_1} & t_{p_1p_2} & \cdots & t_{p_1p_h} \\ t_{p_2p_1} & t_{p_2p_2} & \cdots & t_{p_2p_h} \\ \cdots & \cdots & \cdots & \cdots \\ t_{p_hp_1} & t_{p_hp_2} & \cdots & t_{p_hp_h} \end{vmatrix}.$$

The summations here are extended over all integers $p_1, p_2, \cdots, p_h$ from 1 to $n$ with $p_1 < p_2 < \cdots < p_h$.

It is often advantageous to consider the form $\kappa E - T$ with the determinant

$$\begin{vmatrix} \kappa - t_{11} & -t_{12} & \cdots & -t_{1n} \\ -t_{21} & \kappa - t_{22} & \cdots & -t_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -t_{n1} & -t_{n2} & \cdots & \kappa - t_{nn} \end{vmatrix} = \varphi(\kappa).$$

Its zeros $\kappa_1, \kappa_2, \cdots, \kappa_n$ (eigenvalues of $T$) are the reciprocals of the zeros of $\Delta(\lambda)$. For the reciprocal form $(\kappa E - T)^{-1}$, which exists for all values of $\kappa$ different from $\kappa_1, \kappa_2, \cdots, \kappa_n$, one obtains the Neumann series expansion

$$(\kappa E - T)^{-1} = \frac{E}{\kappa} + \frac{T}{\kappa^2} + \frac{T^2}{\kappa^3} + \cdots$$

which is valid for sufficiently large values of $|\kappa|$. A noteworthy conclusion can be drawn from this expansion. It is clear from the above discussion that the left side is a rational function of $\kappa$ with the denominator $\varphi(\kappa)$; therefore $\varphi(\kappa)(\kappa E - T)^{-1}$ must be a form which is integral and rational in $\kappa$ and its expansion in powers of $\kappa$ can contain no negative powers. Accordingly, if we multiply the above equation by $\varphi(\kappa) = \kappa^n + c_1\kappa^{n-1} + \cdots + c_n$, all the coefficients of negative powers of $\kappa$ in the resulting expression on the right must vanish. But the coefficient of $\kappa^{-1}$ is, as is seen immediately, the expression $T^n + c_1T^{n-1} + \cdots + c_n$, and we thus arrive at the following theorem, which is due to Cayley: *If the determinant of $\kappa E - T$ is denoted by $\varphi(\kappa)$, then the relation*

$$\varphi(T) = 0$$

*is satisfied by the matrix $T$.*

Another important aspect of the spectrum of the eigenvalues $\kappa_1, \kappa_2, \cdots, \kappa_n$ is expressed by the following theorem:

*If the eigenvalues of a matrix $T$ are $\kappa_1, \kappa_2, \cdots, \kappa_n$ and if $g(x)$ is any polynomial in $x$, then the eigenvalues of the matrix $g(T)$ are $g(\kappa_1)$, $g(\kappa_2), \cdots, g(\kappa_n)$.*

To prove this we start from the relation

$$| \kappa E - T | = \varphi(\kappa) = \prod_{\nu=1}^{n} (\kappa - \kappa_\nu),$$

which is an identity in $T$. We wish to obtain the relation

$$| \kappa E - g(T) | = \prod_{\nu=1}^{n} (\kappa - g(\kappa_\nu)).$$

Let $h(x)$ be an arbitrary polynomial of degree $r$ which may be written in terms of its zeros $x_1, x_2, \cdots, x_n$ in the form

$$h(x) = a \prod_{\iota=1}^{r} (x - x_\rho)$$

Then the identity

$$h(T) = a \prod_{\rho=1}^{r} (T - x_\rho E)$$

holds for an arbitrary matrix $T$. By considering the determinants of the matrices in this equation we obtain

$$\begin{aligned} | h(T) | &= a^n \prod_{\rho=1}^{r} | T - x_\rho E | = (-1)^{nr} a^n \prod_{\rho=1}^{r} | x_\rho E - T | \\ &= (-1)^{nr} a^n \prod_{\rho=1}^{r} \varphi(x_\rho) = (-1)^{nr} a^n \prod_{\rho=1}^{r} \left( \prod_{\nu=1}^{n} (x_\rho - \kappa_\nu) \right) \\ &= (-1)^{nr}(-1)^{nr} a^n \prod_{\nu=1}^{n} \left( \prod_{\rho=1}^{r} (\kappa_\nu - x_\rho) \right) = \prod_{\nu=1}^{n} h(\kappa_\nu). \end{aligned}$$

If we now let $h(T)$ be the function $\kappa E - g(T)$, the desired equation

$$| \kappa E - g(T) | = \prod_{\nu=1}^{n} (\kappa - g(\kappa_\nu))$$

follows immediately.

## §3. *Transformation to Principal Axes of Quadratic and Hermitian Forms*

Linear transformations $x = Z(y)$ which reduce a quadratic form

$$K(x, x) = \sum_{p,q=1}^{n} k_{pq} x_p x_q$$

to a linear combination of squares

$$K(x, x) = \sum_{p=1}^{n} \kappa_p y_p^2$$

are highly important in algebra. We are particularly interested in reducing $K(x, x)$ to this form by means of an orthogonal transformation

$$x_p = \sum_{q=1}^{n} l_{qp} y_q = L_p(y) \qquad (p = 1, \cdots, n).$$

Transformations of this kind are called *transformations to principal axes.*

**1. Transformation to Principal Axes on the Basis of a Maximum Principle.** Let us first convince ourselves that a transformation to principal axes is always possible for any given quadratic form $K(x, x)$. To do this we use the theorem that a continuous function of several variables (which are restricted to a finite closed domain) assumes a greatest value somewhere in this domain (Theorem of Weierstrass).[1]

[1] The transformation to principal axes may also be accomplished by direct algebraic methods. An orthogonal matrix $L$ is required, such that $L'KL = D$ is a diagonal matrix with diagonal elements $\kappa_1, \kappa_2, \cdots, \kappa_n$. From the relation $KL = LD$ we obtain the equations

$$\sum_{q=1}^{n} k_{pq} l_{qi} = l_{pi} \kappa_i$$

for the matrix elements $l_{qi}$, which yield the $\kappa_i$ as roots of equation (30), cf. p. 27. Then, on the basis of simple algebraic considerations we can construct an orthogonal system of $n^2$ quantities $l_{qi}$. The method used in the text is preferable to the algebraic method in that it may be generalized to a larger class of transcendental problems.

According to this theorem, there exists a unit vector $\mathbf{l}_1$ with components $l_{11}, l_{12}, \cdots, l_{1n}$ such that, for $x_1 = l_{11}, \cdots, x_n = l_{1n}$, $K(x, x)$ assumes its greatest value, say $\kappa_1$, subject to the subsidiary condition

$$\sum_{p=1}^{n} x_p^2 = 1. \tag{25}$$

Geometrically, the vector $\mathbf{l}_1$ represents on the "unit sphere" (25) a point $P$ so that the surface of the second degree $K(x, x) = \text{const.}$ containing $P$ touches the unit sphere at $P$.

There exists, moreover, a unit vector $\mathbf{l}_2$, orthogonal to $\mathbf{l}_1$, with components $l_{21}, \cdots, l_{2n}$ such that, for $x_1 = l_{21}, \cdots, x_n = l_{2n}$, $K(x, x)$ assumes the greatest possible value $\kappa_2$ subject to the condition

$$\sum_{p=1}^{n} l_{1p} x_p = 0 \tag{26}$$

in addition to condition (25). The problem solved by $\mathbf{l}_1$ for the whole unit sphere is solved by $\mathbf{l}_2$ for the manifold formed by the intersection of the unit sphere and the "plane" (26).

Furthermore, there exists a unit vector $\mathbf{l}_3$, orthogonal to $\mathbf{l}_1$ and $\mathbf{l}_2$, with components $l_{31}, l_{32}, \cdots, l_{3n}$ such that, for $x_i = l_{3i}$ $(i = l, \cdots, n)$, $K(x, x)$ takes on its greatest value $\kappa_3$, subject to the subsidiary conditions (25), (26), and

$$\sum_{p=1}^{n} l_{2p} x_p = 0. \tag{27}$$

Continuing in this manner we obtain a system of $n$ mutually orthogonal vectors $\mathbf{l}_1, \mathbf{l}_2, \cdots, \mathbf{l}_n$, which will be called the "*principal axis vectors*" or "*eigenvectors*." According to (21) their components $l_{qp}$ define an orthogonal transformation

$$x_p = \sum_{q=1}^{n} l_{qp} y_q \qquad (p = 1, \cdots, n); \tag{28}$$

this transformation, we assert, is the solution of our problem.

Since equations (28) are solved by

$$y_p = \sum_{q=1}^{n} l_{pq} x_q \qquad (p = 1, \cdots, n), \tag{29}$$

the equation $\mathbf{x} = \mathbf{l}_p$ is equivalent to the statement $y_p = 1$, $y_q = 0$

for $q \neq p$. Thus, in particular, the maximum $\kappa_1$ is attained for $y_1 = 1, y_2 = 0, \cdots, y_n = 0$; hence, in the transformed form

$$C(y, y) = \sum_{p,q=1}^{n} c_{pq} y_p y_q = K(x, x)$$

the first coefficient $c_{11}$ equals $\kappa_1$. The form

$$H(y, y) = \sum_{p,q=1}^{n} h_{pq}\, y_p\, y_q = C(y, y) - \kappa_1(y_1^2 + \cdots + y_n^2)$$

assumes, moreover, no positive values. For, by the maximum character of $\kappa_1$, $H(y, y)$ is nonpositive provided $\sum_{p=1}^{n} x_p^2 = \sum_{p=1}^{n} y_p^2 = 1$; hence it is nonpositive for all $y_i$ with $\sum_{p=1}^{n} y_p^2 \neq 0$. If $y_1$ should occur in the expression for $H(y, y)$, e.g. if $h_{12} = h_{21}$ were different from zero, we would obtain the value

$$2h_{12}\epsilon + h_{22}\epsilon^2 = \epsilon(2h_{12} + h_{22}\epsilon)$$

for $H(y, y)$ with

$$y_1 = 1, \qquad y_2 = \epsilon, \qquad y_3 = \cdots = y_n = 0.$$

This could be made positive by a suitable choice of $\epsilon$.

It has thus been shown that, after the transformation, $K(x, x)$ is reduced to

$$C(y, y) = \kappa_1 y_1^2 + C_1(y, y),$$

where $C_1(y, y)$ is a quadratic form in the $n - 1$ variables $y_2, y_3, \cdots, y_n$. If the subsidiary condition $y_1 = 0$ is imposed the transformed form is equal to $C_1(y, y)$. In the same way we may now conclude that $C_1(y, y)$ is of the form $\kappa_2 y_2^2 + C_2(y, y)$, where $C_2(y, y)$ depends only on the $n - 2$ variables $y_3, y_4, \cdots, y_n$, and so forth.

Thus we have demonstrated the possibility of a transformation to principal axes so that

$$\sum_{p,q=1}^{n} k_{pq} x_p x_q = \sum_{p=1}^{n} \kappa_p y_p^2, \qquad \sum_{p=1}^{n} x_p^2 = \sum_{p=1}^{n} y_p^2.$$

We might note that the corresponding minimum problem would have served equally well as a starting point for the proof; i.e. we might have looked for the minimum of $K(x, x)$, subject to the auxiliary condition $E(x, x) = 1$. In that case we would have arrived at the quantities $\kappa_1, \kappa_2, \cdots, \kappa_n$ in the reverse order. One could also keep $K(x, x)$ constant and look for the maxima or minima of $E(x, x)$; then the minimum values $\lambda_i$ would be the reciprocals of the $\kappa_i$.

**2. Eigenvalues.** We shall now show that the values $\kappa_i$ defined in the previous subsection as successive maxima are identical with the eigenvalues as introduced in §2.

The equation

$$\varphi(\kappa) = (\kappa - \kappa_1)(\kappa - \kappa_2) \cdots (\kappa - \kappa_n) = 0$$

satisfied by the numbers $\kappa_i$, may be written in the form

$$\begin{vmatrix} \kappa - \kappa_1 & 0 & 0 & \cdots & 0 \\ 0 & \kappa - \kappa_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \kappa - \kappa_n \end{vmatrix} = 0.$$

But this determinant is just the determinant of the quadratic form

$$\kappa \sum_{p=1}^{n} y_p^2 - \sum_{p=1}^{n} \kappa_p y_p^2,$$

which is obtained by applying an orthogonal transformation to the form

$$\kappa \sum_{p=1}^{n} x_p^2 - K(x, x).$$

Therefore the relation

$$\begin{vmatrix} \kappa - \kappa_1 & 0 & \cdots & 0 \\ 0 & \kappa - \kappa_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \kappa - \kappa_n \end{vmatrix}$$

$$= \begin{vmatrix} \kappa - k_{11} & -k_{12} & \cdots & -k_{1n} \\ -k_{21} & \kappa - k_{22} & \cdots & -k_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -k_{n1} & -k_{n2} & \cdots & \kappa - k_{nn} \end{vmatrix}$$

is an identity in $\kappa$. Consequently the numbers $\kappa_i$ are the roots of the algebraic equation

$$(30) \qquad \begin{vmatrix} k_{11} - \kappa & k_{12} & \cdots & k_{1n} \\ k_{21} & k_{22} - \kappa & \cdots & k_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ k_{n1} & k_{n2} & \cdots & k_{nn} - \kappa \end{vmatrix} = 0$$

for the unknown $\kappa$; i.e. they are the eigenvalues introduced in §2.

Our method of derivation shows automatically that the roots of equation (30) are necessarily real if the $k_{pq}$ are arbitrary real quantities subject to the condition $k_{pq} = k_{qp}$.[1] We may also remark in passing that the absolute values of the reciprocals of the eigenvalues are geometrically significant as the squares of the lengths of the principal axes of the surface $K(x, x) = 1$ in $n$-dimensional space. If at least one eigenvalue is equal to zero the form is said to be "*degenerate*"; it can then be represented as a form of less than $n$ variables. It is clear from equation (30) that this is the case if and only if $| k_{pq} |$ vanishes. For $K(x, x)$ to be positive definite the condition $\kappa_p > 0$, $p = 1, 2, \cdots, n$ is necessary and sufficient.

Suppose the representation of a form $K(x, x)$ in terms of principal axes

$$K(x, x) = \sum_{p=1}^{n} \kappa_p y_p^2$$

is given. Then, using the properties of the orthogonal transformations of products discussed above, the expressions

$$K^2(x, x) = \sum_{p=1}^{n} \kappa_p^2 y_p^2, \qquad K^3(x, x) = \sum_{p=1}^{n} \kappa_p^3 y_p^2, \cdots$$

are easily obtained for the iterated forms. It follows that the eigenvalues of the $h$-fold iterated form $K^h(x, x)$ are the $h$-th powers of the eigenvalues of $K(x, x)$ (this also follows immediately from the theorem on page 22); moreover we see that, for even $h$, the form $K^h(x, x)$ is positive definite.

[1] Equation (30) is customarily called the secular equation because it occurs in the problem of secular perturbations of planetary orbits. For a direct proof that the eigenvalues are real, see Ch. III, §4, 2.

**3. Generalization to Hermitian Forms.** A transformation to principal axes can be carried out in exactly the same way for Hermitian forms. A Hermitian form

$$H(x, \bar{x}) = \sum_{p,q=1}^{n} h_{pq} x_p \bar{x}_q$$

with the matrix $H = \bar{H}'$ can always be transformed by a unitary transformation $L$, given by

$$x_p = \sum_{q=1}^{n} l_{qp} y_q ,$$

into the form

$$H(x, \bar{x}) = \sum_{p=1}^{n} \kappa_p y_p \bar{y}_p = \sum_{p=1}^{n} \kappa_p \mid y_p \mid^2 ,$$

where all the coefficients $\kappa_p$ are real. These eigenvalues $\kappa_m$ reappear as the maxima of the Hermitian form $H(x, \bar{x})$, subject to the auxiliary conditions

$$\sum_{p=1}^{n} \mid x_p \mid^2 = 1 \qquad \text{and} \qquad \sum_{p=1}^{n} l_{ip} \bar{x}_p = 0 \qquad (i = 1, \cdots, m - 1).$$

**4. Inertial Theorem for Quadratic Forms.** If we relinquish the requirement that the linear transformation be orthogonal, a quadratic form may be transformed into a sum of squares by many different transformations. In particular, after the above orthogonal transformation has been carried out, any transformation in which each variable is simply multiplied by a factor of proportionality leaves the character of the form as a sum of squares unaltered. Thus it is possible to transform the form in such a way that all the (real) coefficients have the value $+1$ or $-1$. The following theorem, known as the *inertial theorem for quadratic forms*, holds:

*The number of positive and negative coefficients, respectively, in a quadratic form reduced to an expression* $\sum c_p z_p^2$ *by means of a nonsingular real linear transformation does not depend on the particular transformation.*

Proof: The positive and negative coefficients may be made equal to $+1$ and $-1$, respectively. Suppose, now, that the quadratic form $K(x, x)$ is transformed by two different transformations into

$y_1^2 + \cdots + y_r^2 - y_{r+1}^2 - \cdots - y_n^2$ and $z_1^2 + \cdots + z_s^2 - z_{s+1}^2 - \cdots - z_n^2$ with $r < s$. We then have

$$y_1^2 + \cdots + y_r^2 + z_{s+1}^2 + \cdots + z_n^2 = y_{r+1}^2 + \cdots + y_n^2 + z_1^2 + \cdots + z_s^2.$$

Let us consider the conditions $y_1 = \cdots = y_r = z_{s+1} = \cdots = z_n = 0$, which imply that the remaining $y_i$ also vanish. By imagining the $z_i$ expressed in terms of the $y_i$ and regarding these conditions as a system of fewer than $n$ equations in the $y_i$, we obtain the contradiction that there exists a non-vanishing solution vector.

**5. Representation of the Resolvent of a Form.** The resolvent of the quadratic form $K(x, x)$ can be expressed in a suggestive way. According to §2 the resolvent may be defined by the symbolic equation

$$\mathsf{K}(x, x; \lambda) = \frac{[E(x, x) - \lambda K(x, x)]^{-1} - E(x, x)}{\lambda}.$$

We suppose that $K(x, x)$ has been brought into the form

$$K(x, x) = \sum_{p=1}^{n} \frac{y_p^2}{\lambda_p}.$$

The resolvent of $\sum_{p=1}^{n} y_p^2/\lambda_p$ must be identical with the resolvent of $K(x, x)$, since $[E(x, x) - \lambda K(x, x)]^{-1}$ goes over into

$$\left[E(y, y) - \lambda \sum_{p=1}^{n} \frac{y_p^2}{\lambda_p}\right]^{-1}$$

when the transformation is applied. Now the following relations hold:

$$\frac{1}{\lambda}\left[\left(\sum_{p=1}^{n} y_p^2 - \lambda \sum_{p=1}^{n} \frac{y_p^2}{\lambda_p}\right)^{-1} - E(y, y)\right]$$

$$= \frac{1}{\lambda}\left[\left(\sum_{p=1}^{n} \frac{\lambda_p - \lambda}{\lambda_p} y_p^2\right)^{-1} - E(y, y)\right] = \frac{1}{\lambda}\left[\sum_{p=1}^{n} \frac{\lambda_p}{\lambda_p - \lambda} y_p^2 - E(y, y)\right]$$

$$= \frac{1}{\lambda}\left[\sum_{p=1}^{n} \frac{\lambda_p}{\lambda_p - \lambda} y_p^2 - \sum_{p=1}^{n} y_p^2\right] = \sum_{p=1}^{n} \frac{y_p^2}{\lambda_p - \lambda}.$$

If we now transform back to the variables $x_p$, using the notation (19) we obtain the expression

$$\mathsf{K}(x, x; \lambda) = \sum_{p=1}^{n} \frac{[L'_p(x)]^2}{\lambda_p - \lambda} \tag{31}$$

for the resolvent of $K(x, x)$; thus, for the bilinear form, we have

$$\mathsf{K}(u, x; \lambda) = \sum_{p=1}^{n} \frac{L'_p(u)L'_p(x)}{\lambda_p - \lambda}. \tag{32}$$

From this representation it is evident, incidentally, that the residue of the rational function $\mathsf{K}(u, x; \lambda)$ of $\lambda$ at the point $\lambda_p$ is equal to $-L'_p(u)L'_p(x)$, assuming that $\lambda_p \neq \lambda_q$ for $p \neq q$.

**6. Solution of Systems of Linear Equations Associated with Forms.** In conclusion we shall present, with the help of the eigenvectors, the solution of the system of linear equations

$$x_p - \lambda \sum_{q=1}^{n} k_{pq} x_q = y_p \qquad (p = 1, \cdots, n) \tag{33}$$

associated with the quadratic form

$$K(x, x) = \sum_{p,q=1}^{n} k_{pq} x_p x_q .$$

If we apply the transformation to principal axes

$$x_p = \sum_{q=1}^{n} l_{qp} u_q , \qquad y_p = \sum_{q=1}^{n} l_{qp} v_q$$

to the variables $x_i$ and $y_i$, $K(x, x)$ goes over into

$$\sum_{q=1}^{n} \kappa_q u_q^2 ,$$

and the bilinear form $K(x, z)$ is similarly transformed. Hence, our system of equations (33) becomes

$$u_p - \lambda \kappa_p u_p = v_p \qquad (p = 1, \cdots, n). \tag{34}$$

the solution of which is

$$u_p = \frac{v_p}{1 - \lambda\kappa_p} = \frac{v_p}{1 - \dfrac{\lambda}{\lambda_p}} = \frac{\lambda_p}{\lambda_p - \lambda} v_p . \tag{35}$$

In terms of the original variables, we obtain the equivalent formula for the solution

$$\mathbf{x} = \sum_{p=1}^{n} \frac{\mathbf{y} \cdot \mathbf{l}_p}{1 - \dfrac{\lambda}{\lambda_p}} \mathbf{l}_p , \tag{36}$$

in which the solution appears as a development in terms of the eigenvectors $\mathbf{l}_1, \mathbf{l}_2, \cdots, \mathbf{l}_n$ of the form $K(x, x)$. We have here used the notation $\mathbf{y}\cdot\mathbf{l}_p = \sum_{q=1}^{n} l_{pq}y_q$.

The principal axis vector or eigenvector $\mathbf{l}_p$ is itself the normalized solution of the homogeneous equations

$$x_q - \lambda_p \sum_{r=1}^{n} k_{qr}x_r = 0$$

or

$$u_q - \lambda_p\kappa_q u_q = 0 \qquad (q = 1, \cdots, n).$$

If, for $q \neq p$, all the $\kappa_q$ are different from $\kappa_p = 1/\lambda_p$, there exists only one normalized solution,

$$u_p = 1,$$

$$u_q = 0 \qquad (q \neq p)$$

or

$$\mathbf{x} = \mathbf{l}_p.$$

If several characteristic numbers coincide the principal axis vectors are not uniquely determined.

## §4. *Minimum-Maximum Property of Eigenvalues*

**1. Characterization of Eigenvalues by a Minimum-Maximum Problem.** In the above discussion we have obtained the eigenvalues by solving a series of maximum problems, each one of which depended on the solutions of the previous problems of the series. We shall now show that each eigenvalue can be directly characterized as the solution of a somewhat different problem in which all reference to the solutions of previous problems is avoided.

The problem is to maximize the form

$$K(x, x) = \sum_{p,q=1}^{n} k_{pq}x_p x_q$$

if the condition (25)

$$\sum_{p=1}^{n} x_p^2 = 1$$

is imposed and if the $h - 1$ equations

$$(37) \qquad \sum_{p=1}^{n} \alpha_{\nu p}x_p = 0 \qquad (\nu = 1, \cdots, h-1;\ h \leq n)$$

must be satisfied. This maximum value of $K(x, x)$ is of course a function of the parameters $\alpha_{\nu p}$. We now choose the $\alpha_{\nu p}$ in such a way as to give this maximum its least possible value. We assert that this minimum value of the maximum is just the $h$-th eigenvalue $\kappa_h$ of $K(x, x)$, provided the eigenvalues are ordered in a sequence of decreasing values, $\kappa_1$ being the greatest eigenvalue, $\kappa_2$ the next, and so on.

The transformation to principal axes changes $K(x, x)$ into

$$\sum_{p=1}^{n} \kappa_p y_p^2 \qquad (\kappa_1 \geq \cdots \geq \kappa_n),$$

condition (25) into

$$\sum_{p=1}^{n} y_p^2 = 1, \tag{38}$$

and equations (37) into

$$\sum_{p=1}^{n} \beta_{\nu p} y_p = 0, \qquad (\nu = 1, \cdots h - 1; h \leq n) \tag{39}$$

where the $\beta_{\nu p}$ are new parameters. If we set

$$y_{h+1} = \cdots = y_n = 0$$

equations (39) become $h - 1$ equations in $h$ unknowns $y_1, y_2, \cdots, y_h$, which can certainly be satisfied for a set of values $y_i$ also satisfying (38). For these values we have

$$K(x, x) = \kappa_1 y_1^2 + \cdots + \kappa_h y_h^2 \geq \kappa_h(y_1^2 + \cdots + y_n^2) = \kappa_h .$$

Thus the required maximum of $K(x, x)$ for any set of values $\beta_{\nu p}$ is not less than $\kappa_h$; but it is just equal to $\kappa_h$ if we take for (39) the equations

$$y_1 = \cdots = y_{h-1} = 0.$$

It follows therefore that:

*The $h$-th eigenvalue $\kappa_h$ of the quadratic form $K(x, x)$ is the least value which the maximum of $K(x, x)$ can assume if, in addition to the condition*

$$\sum_{p=1}^{n} x_p^2 = 1,$$

*$h - 1$ arbitrary linear homogeneous equations connecting the $x_p$ are prescribed.*

**2. Applications. Constraints.** This independent minimum-maximum property of the eigenvalues shows how the eigenvalues are changed if $j$ independent constraints

$$\sum_{p=1}^{n} \gamma_{sp} x_p \qquad (s = 1, \cdots, j) \tag{40}$$

are imposed on the variables, so that $K(x, x)$ reduces to a quadratic form $\tilde{K}(x, x)$ of $n - j$ independent variables. The $h$-th eigenvalue $\tilde{\kappa}_h$ is obtained from the same minimum-maximum problem as $\kappa_h$, in which the totality of sets of admissible values $x_i$ has been narrowed down by (40). Therefore the maximum, and thus the eigenvalue of $\tilde{K}(x, x)$, certainly does not exceed the corresponding quantity for $K(x, x)$.

Furthermore, $\kappa_{j+h}$ is the least maximum which $K(x, x)$ can possess if, in addition to (25), $h + j - 1$ linear homogeneous conditions are imposed on the $x_p$; $\kappa_{j+h}$ is therefore certainly not greater than $\tilde{\kappa}_h$, for which $j$ of these conditions are given by the fixed equations (40).

We have thus the theorem: *If a quadratic form $K(x, x)$ of $n$ variables is reduced by $j$ linear homogeneous constraints to a quadratic form $\tilde{K}(x, x)$ of $n - j$ variables, then the eigenvalues $\tilde{\kappa}_1, \tilde{\kappa}_2, \cdots, \tilde{\kappa}_{n-j}$ of $\tilde{K}(x, x)$ are not greater than the corresponding numbers of the sequence $\kappa_1, \kappa_2, \cdots, \kappa_{n-j}$ and not less than the corresponding numbers of the sequence $\kappa_{j+1}, \kappa_{j+2}, \cdots, \kappa_n$.*[1]

If, in particular, we let $j = 1$ and take for our constraint the condition $x_n = 0$, then the quadratic form $K$ goes over into its $(n - 1)$-st "*section*," and we obtain the theorem: *The h-th eigenvalue of the $(n - 1)$-st section is at most equal to the h-th eigenvalue of the original quadratic form, and at least equal to the $(h + 1)$-st eigenvalue.*

If this theorem is applied to the $(n - 1)$-st section of the quadratic form, there results a corresponding theorem for the $(n - 2)$-nd section, and so forth. In general we note that the eigenvalues of any two successive sections of a quadratic form are ordered in the indicated manner.

Moreover, we may conclude: *If a positive definite form is added*

[1] This may be illustrated geometrically: Let us consider the ellipse formed by the intersection of an ellipsoid and a plane passing through its center. The length of the major axis of this ellipse is between the lengths of the longest and the second axes of the ellipsoid, while the length of the minor axis of the ellipse is between those of the second and the shortest axes of the ellipsoid.

*to $K(x, x)$, the eigenvalues of the sum are not less than the corresponding eigenvalues of $K(x, x)$.*

Instead of utilizing a minimum-maximum problem to characterize the eigenvalues we may use a *maximum-minimum problem*. In this case the eigenvalues will appear in the opposite order.

It may be left to the reader to formulate and prove the minimum-maximum character of the eigenvalues of Hermitian forms.

## §5. *Supplement and Problems*

**1. Linear Independence and the Gram Determinant.** The question of the linear dependence of $m$ given vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$ may be very simply decided in the following way without explicity determining the rank of the component matrix: We consider the quadratic form

$$G(x, x) = (x_1\mathbf{v}_1 + \cdots + x_m\mathbf{v}_m)^2 = \sum_{i,k=1}^{m} (\mathbf{v}_i \cdot \mathbf{v}_k)x_i x_k .$$

Clearly $G(x, x) \geq 0$, and the vectors $\mathbf{v}_i$ are linearly dependent if and only if there exists a set of values $x_1, x_2, \cdots, x_m$ with (25′)

$$\sum_{i=1}^{m} x_i^2 = 1,$$

for which $G(x, x) = 0$. Thus if the vectors $\mathbf{v}_i$ are linearly dependent the minimum of the form $G(x, x)$ subject to condition (25′) must be equal to zero. But this minimum is just the smallest eigenvalue of the quadratic form $G(x, x)$, i.e. the least root of the equation

$$\text{(41)} \qquad \begin{vmatrix} \mathbf{v}_1^2 - \kappa & (\mathbf{v}_1 \cdot \mathbf{v}_2) & \cdots & (\mathbf{v}_1 \cdot \mathbf{v}_m) \\ (\mathbf{v}_2 \cdot \mathbf{v}_1) & \mathbf{v}_2^2 - \kappa & \cdots & (\mathbf{v}_2 \cdot \mathbf{v}_m) \\ \cdots & \cdots & \cdots & \cdots \\ (\mathbf{v}_m \cdot \mathbf{v}_1) & (\mathbf{v}_m \cdot \mathbf{v}_2) & \cdots & \mathbf{v}_m^2 - \kappa \end{vmatrix} = 0.$$

The theorem follows:

*A necessary and sufficient condition for the linear dependence of the vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$ is the vanishing of the "Gram determinant"*

$$\text{(42)} \qquad \Gamma = \begin{vmatrix} \mathbf{v}_1^2 & (\mathbf{v}_1 \cdot \mathbf{v}_2) & \cdots & (\mathbf{v}_1 \cdot \mathbf{v}_m) \\ (\mathbf{v}_2 \cdot \mathbf{v}_1) & \mathbf{v}_2^2 & \cdots & (\mathbf{v}_2 \cdot \mathbf{v}_m) \\ \cdots & \cdots & \cdots & \cdots \\ (\mathbf{v}_m \cdot \mathbf{v}_1) & (\mathbf{v}_m \cdot \mathbf{v}_2) & \cdots & \mathbf{v}_m^2 \end{vmatrix}.$$

An alternate expression for $\Gamma$ follows from (41). If the left side of equation (41), which is satisfied by the (all non-negative) eigenvalues $\kappa_1, \kappa_2, \cdots, \kappa_m$ of $G(x, x)$, is developed in powers of $\kappa$, then the term independent of $\kappa$ is equal to $\Gamma$, while the coefficient of $\kappa^m$ is equal to $(-1)^m$. According to a well-known theorem of algebra it follows that

$$\Gamma = \kappa_1\kappa_2 \cdots \kappa_m . \tag{43}$$

Consequently *the Gram determinant of an arbitrary system of vectors is never negative.* Relation

$$\Gamma = |(\mathbf{v}_i \cdot \mathbf{v}_k)| \geq 0 \qquad (i, k = 1, \cdots, m), \tag{44}$$

in which the equality holds only for linearly dependent vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$, is a generalization of the Schwarz inequality (see page 2)

$$\mathbf{v}_1^2\mathbf{v}_2^2 - (\mathbf{v}_1 \cdot \mathbf{v}_2)^2 = \begin{vmatrix} \mathbf{v}_1^2 & (\mathbf{v}_1 \cdot \mathbf{v}_2) \\ (\mathbf{v}_2 \cdot \mathbf{v}_1) & \mathbf{v}_2^2 \end{vmatrix} \geq 0.$$

The value of the Gram determinant or, alternatively, the lowest eigenvalue $\kappa_m$ of the form $G(x, x)$ represents a measure of the linear independence of the vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$. The smaller this number, the "flatter" is the $m$-dimensional polyhedron defined by vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$; if it is equal to zero the polyhedron collapses into one of at most $m - 1$ dimensions. In this connection the Gram determinant has a simple geometrical significance. It is equal to the square of the $m!$-fold volume of the $m$-dimensional polyhedron defined by the vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$. Thus, for $m = 2$, it is the square of twice the area of the triangle formed from $\mathbf{v}_1$ and $\mathbf{v}_2$.

Gram's criterion for linear dependence must of course be equivalent to the usual one. The latter states that vectors are linearly dependent if and only if all determinants formed with $m$ columns of the rectangular component array

$$\begin{matrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{matrix}$$

are equal to zero. And indeed, according to a well-known theorem of the theory of determinants,

$$\Gamma = \sum \begin{vmatrix} v_{1s_1} & v_{1s_2} & \cdots & v_{1s_m} \\ v_{2s_1} & v_{2s_2} & \cdots & v_{2s_m} \\ \cdots & \cdots & \cdots & \cdots \\ v_{ms_1} & v_{ms_2} & \cdots & v_{ms_m} \end{vmatrix}^2, \tag{45}$$

where the summation is extended over all integers $s_1, s_2, \cdots, s_m$ from 1 to $n$ with $s_1 < s_2 < \cdots < s_m$.

**2. Hadamard's Inequality for Determinants.** Every determinant

$$A = | a_{ik} | = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{2n} & \cdots & a_{nn} \end{vmatrix}$$

with real elements $a_{ik}$ satisfies the inequality

$$A^2 \leq \prod_{i=1}^{n} \sum_{k=1}^{n} a_{ik}^2. \tag{46}$$

Proof: Let the elements $a_{ik}$ vary, keeping the sums of squares

$$\sum_{k=1}^{n} a_{ik}^2 = c_i^2 \qquad (i = 1, \cdots, n)$$

fixed. If $A_{\max}^2$ is the greatest value of the function $A^2$ of the elements $a_{ik}$ under these $n$ conditions—the existence of such a maximum follows immediately from Weierstrass's theorem (see page 23)—then the elements of $A_{\max}$ in each row must be proportional to the corresponding cofactors. For, if $h$ is fixed, we have

$$A = a_{h1}A_{h1} + \cdots + a_{hn}A_{hn} ;$$

thus, by the Schwarz inequality,

$$A^2 \leq \sum_{k=1}^{n} a_{hk}^2 \sum_{k=1}^{n} A_{hk}^2 = c_h^2 \sum_{k=1}^{n} A_{hk}^2 .$$

If the $a_{hk}$ are not proportional to the $A_{hk}$ the inequality holds, and $A^2$ certainly can not have its maximum value. For, in this case, by suitably changing the $n$ quantities $a_{hk}$ $(k = 1, \cdots, n)$, with $c_h^2$ and the $A_{hk}$ held constant, the square of the determinant can be made equal to the right-hand side.

If we now multiply $A_{\max}$ by itself, we obtain, according to the multiplication theorem for determinants,

$$A_{\max}^2 = \prod_{i=1}^{n} c_i^2,$$

since the inner products of different rows of $A_{\max}$ vanish as a result of the proportionality just demonstrated and of elementary theorems on determinants. Therefore the original determinant satisfies Hadamard's inequality

$$A^2 \leq \prod_{i=1}^{n} c_i^2 = \prod_{i=1}^{n} \sum_{k=1}^{n} a_{ik}^2.$$

The geometrical meaning of Hadamard's inequality is that the volume of the polyhedron formed from $n$ vectors of given lengths in $n$-dimensional space is greatest if the vectors are mutually orthogonal.

Hadamard's inequality is also valid for complex $a_{ik}$ if $A$ and $a_{ik}$ are replaced by their absolute values.

**3. Generalized Treatment of Canonical Transformations.** For generalizations and applications to many problems of analysis the following concise treatment of the simultaneous canonical transformation of two quadratic forms is most appropriate. Again we consider two quadratic forms in an $n$-dimensional vector space of vectors $\mathbf{x}, \mathbf{y}, \cdots$:

$$\text{(a)} \qquad H(x, x) = \sum_{p,q=1}^{n} h_{pq} x_p x_q,$$

which we assume positive definite, and

$$\text{(b)} \qquad K(x, x) = \sum_{p,q=1}^{n} k_{pq} x_p x_q,$$

which is not necessarily definite. By definition we interpret $H(x, x)$ as the square of the length of the vector $\mathbf{x}$, and the polar form

$$H(x, y) = (\mathbf{x}, \mathbf{y}) = \sum_{p,q=1}^{n} h_{pq} x_p y_q$$

as the inner product of $\mathbf{x}$ and $\mathbf{y}$. The problem is to find a linear transformation

$$x_p = \sum_{q=1}^{n} l_{pq} y_q \qquad (p = 1, \cdots, n)$$

which transforms $K$ and $H$ into the sums

$$K(x, x) = \sum_{p=1}^{n} \rho_p y_p^2, \qquad H(x, x) = \sum_{p=1}^{n} y_p^2.$$

To obtain this transformation explicit expressions for the forms $K$ and $H$ are not required; our proof is based merely on the properties that $H$ and $K$ are continuous functions of the vector $\mathbf{x}$, that with arbitrary constants $\lambda$ and $\mu$ equations of the form

$$(47) \quad H(\lambda x + \mu y, \lambda x + \mu y) = \lambda^2 H(x, x) + 2\lambda\mu H(x, y) + \mu^2 H(y, y)$$

$$(48) \quad K(\lambda x + \mu y, \lambda x + \mu y) = \lambda^2 K(x, x) + 2\lambda\mu K(x, y) + \mu^2 K(y, y)$$

hold, and that $H$ is positive definite, vanishing only for $\mathbf{x} = 0$.

We consider a sequence of maximum problems: First we define a vector $\mathbf{x} = \mathbf{x}^1$ for which the quotient

$$K(x, x)/H(x, x)$$

attains its maximum value $\rho_1$. Without affecting the value of this quotient the vector $\mathbf{x}$ may be normalized, i.e. subjected to the condition $H(x, x) = 1$.

Then we define another normalized vector $\mathbf{x}^2$ for which the quotient $K(x, x)/H(x, x)$ attains its maximum value $\rho_2$ under the orthogonality condition $H(x, x^1) = 0$. Proceeding in this way, we define a sequence of normalized vectors $\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^k$, such that for $\mathbf{x} = \mathbf{x}^k$ the quotient $K(x, x)/H(x, x)$ attains its maximum value $\rho_k$ under the orthogonality conditions

$$H(x, x^\nu) = 0 \qquad (\nu = 1, \cdots, k - 1).$$

After $n$ steps we obtain a complete system of vectors $\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^n$ for which the relations

$$(49) \quad H(x^i, x^k) = 1, \quad i = k; \qquad H(x^i, x^k) = 0, \quad i < k;$$

and

$$(50) \quad K(x^i, x^k) = \rho_k, \quad i = k; \qquad K(x^i, x^k) = 0, \quad i < k$$

hold. Relations (49) are merely the orthogonality relations stipulated in our maximum problems. To prove relations (50) we consider first $\mathbf{x}^1$. The maximum property of $\mathbf{x}^1$ is expressed by the inequality

$$K(x^1 + \epsilon\zeta, x^1 + \epsilon\zeta) - \rho_1 H(x^1 + \epsilon\zeta, x^1 + \epsilon\zeta) \leq 0$$

valid for an arbitrary constant $\epsilon$ and an arbitrary vector $\boldsymbol{\zeta}$. Because of (47) and (48), it yields

$$2\epsilon A + \epsilon^2 B \leq 0$$

where

$$A = K(x^1, \zeta) - \rho_1 H(x^1, \zeta), \qquad B = K(\zeta, \zeta) - \rho_1 H(\zeta, \zeta).$$

Since this inequality is valid for arbitrarily small positive or negative $\epsilon$ it implies that $A = 0$ or that

$$K(x^1, \zeta) - \rho_1 H(x^1, \zeta) = 0 \tag{51}$$

for arbitrary $\boldsymbol{\zeta}$. The maximum problem for $\mathbf{x}^h$ yields as above

$$K(x^h, \zeta) - \rho_h H(x^h, \zeta) = 0$$

for an arbitrary vector $\boldsymbol{\zeta}$ satisfying the relations

$$H(\zeta, x^\nu) = 0 \qquad (\nu = 1, \cdots, h - 1).$$

Now, for $h < k$, we may take $\boldsymbol{\zeta} = \mathbf{x}^k$. Since $H(x^h, x^k) = 0$, we may conclude that $K(x^h, x^k) = 0$ for $h < k$, while by definition $K(x^h, x^h) = \rho_h$.

Since the $n$ orthogonal vectors $\mathbf{x}^\nu$ form a complete system in our vector space, an arbitrary vector $\mathbf{x}$ can be expressed in the form

$$\mathbf{x} = \sum_{\nu=1}^{n} y_\nu \mathbf{x}^\nu$$

where $y_\nu = H(x, x^\nu)$. We substitute these expressions in $H$ and $K$ and use the expansions corresponding to (47), (48) for $n$ summands; because of (49), (50) it follows immediately that

$$H(x, x) = \sum_{\nu=1}^{n} y_\nu^2,$$

$$K(x, x) = \sum_{\nu=1}^{n} \rho_\nu y_\nu^2.$$

Thus we have accomplished the required transformation.

Exactly as before the values $\rho_h$ are shown to have the following minimum-maximum property.

*Under the auxiliary conditions*

$$\sum_{p=1}^{n} \alpha_{\nu p} x_p = 0 \qquad (\nu = 1, \cdots, h-1),$$

*$\rho_h$ (with $\rho_1 \geq \cdots \geq \rho_n$) is the least value which the maximum of $K(x, x)/H(x, x)$ can assume—this maximum is regarded as a function of the parameters $\alpha_{\nu p}$.*

To construct the transformation of which we have proved the existence we first show that for all integers $h$ the "variational equation"

$$K(x^h, \zeta) - \rho_h H(x^h, \zeta) = 0$$

holds with an arbitrary vector $\boldsymbol{\zeta}$. So far the relation has been proved only under the restriction $(\boldsymbol{\zeta}, \mathbf{x}^\nu) = 0$ for $\nu < h$. However, if $\zeta$ is arbitrary the vector $\boldsymbol{\eta} = \boldsymbol{\zeta} - c_1\mathbf{x}^1 - \cdots - c_{h-1}\mathbf{x}^{h-1}$ with $c_\nu = (\boldsymbol{\zeta}, \mathbf{x}^\nu)$ satisfies the orthogonality condition $H(\eta, x^\nu) = 0$, $\nu < h$, hence $0 = K(x^h, \eta) - \rho_h H(x^h, \eta) = K(x^h, \zeta) - \rho_h H(x^h, \zeta)$; here the final equality sign follows from (49) and (50).

Writing the variational equation for $\mathbf{x}^h = \mathbf{x}$, $\rho_h = \rho$ we obtain for the components $x_j$ of $\mathbf{x} = \mathbf{x}^h$ the system of linear homogeneous equations

$$\sum_{j=1}^{n} (k_{ij} - \rho h_{ij}) x_j = 0 \qquad (i = 1, \cdots, n);$$

hence the values $\rho_h$ satisfy the determinant equation $\| k_{ij} - \rho h_{ij} \| = 0$ and the vectors $\mathbf{x}^h$ are obtained from the linear equations after the quantities $\rho = \rho_h$ have been found. Clearly, these considerations characterize the numbers $\rho_h$ and the vectors $\mathbf{x}^h$ as the eigenvalues and eigenvectors of the matrix $(\mathbf{k}_{pq})$ with respect to the matrix $(\mathbf{h}_{pq})$.

Thus for each eigenvalue $\rho_h$ there exists a solution in the form of a vector $\mathbf{x}^h$. The solutions for different eigenvalues are orthogonal; if two eigenvalues are equal the corresponding solutions are not necessarily orthogonal but may be made so by the orthogonalization process of page 4. These mutually orthogonal solutions may be normalized to unit length; the resulting vectors are the eigenvectors of the problem and their components are the coefficients of the required transformation.

These coefficients $l_{pq}$ are obtained from $\mathbf{x} = \sum_{q=1}^{n} y_q \mathbf{x}^q$ if we multiply by the vector $\mathbf{e}^p$ which defines the original coordinate system. Thus $x_p = (\mathbf{x}, \mathbf{e}^p) = \sum_{q=1}^{n} y_q(\mathbf{x}^q, \mathbf{e}^p)$; hence $l_{pq} = (\mathbf{x}^q, \mathbf{e}^p)$.

**4. Bilinear and Quadratic Forms of Infinitely Many Variables.** Under suitable conditions our theory remains valid if the number of variables increases beyond all bounds. For example, this is the case if both the sum of the squares of the coefficients of the bilinear or quadratic forms and the sum of the squares of the variables converge. This theory of forms of infinitely many variables, developed by Hilbert, may then be applied to numerous problems of analysis. However, the theory of forms in vector spaces of infinitely many dimensions can be more adequately developed on the basis of abstract concepts as indicated in subsection 3. As we shall see, many topics in analysis can be illuminated from the viewpoint of such a generalized theory of quadratic forms.

**5. Infinitesimal Linear Transformations.** An infinitesimal linear transformation is defined as a transformation whose matrix is

$$A = E + (\epsilon\alpha_{ik}) = \begin{pmatrix} 1+\epsilon\alpha_{11} & \epsilon\alpha_{12} & \cdots & \epsilon\alpha_{1n} \\ \epsilon\alpha_{21} & 1+\epsilon\alpha_{22} & \cdots & \epsilon\alpha_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \epsilon\alpha_{n1} & \epsilon\alpha_{n2} & \cdots & 1+\epsilon\alpha_{nn} \end{pmatrix},$$

where $\epsilon$ denotes an infinitesimal quantity of the first order, i.e. a quantity whose higher powers are, for the problem at hand, negligible in comparison with lower powers of $\epsilon$. The product of two such infinitesimal transformations with the matrices $A = E + (\epsilon\alpha_{ik})$ and $B = E + (\epsilon\beta_{ik})$ has the matrix $C = E + (\epsilon\alpha_{ik} + \epsilon\beta_{ik})$. Thus the product does not depend on the order of the factors; in other words, *infinitesimal transformations commute with each other.*

Furthermore, *the reciprocal matrix of $A = E + (\epsilon\alpha_{ik})$ is $A^{-1} = E - (\epsilon\alpha_{ik})$, and the determinant of the matrix $A$ is equal to*

$$1 + \epsilon(\alpha_{11} + \alpha_{22} + \cdots + \alpha_{nn}).$$

If the infinitesimal transformation is to be orthogonal, we have the condition $A'A = E$, where $A'$ is the transposed matrix. We must therefore have $\alpha_{ik} + \alpha_{ki} = 0$, or, in other words:

*A necessary and sufficient condition for the orthogonality of an in-*

*finitesimal transformation is that the difference between its matrix and the unit matrix be skew-symmetric.*

Any infinitesimal transformation with the matrix $C = E + (\epsilon\gamma_{ik})$ may be represented as the product of an orthogonal transformation $A = E + (\epsilon\alpha_{ik})$ and a symmetric transformation $B = E + (\epsilon\beta_{ik})$, where

$$\alpha_{ik} = \tfrac{1}{2}(\gamma_{ik} - \gamma_{ki}),$$

$$\beta_{ik} = \tfrac{1}{2}(\gamma_{ik} + \gamma_{ki}).$$

Consider a symmetric transformation $y_i = \sum_k s_{ik}x_k$ whose matrix is $S = (s_{ik})$, not necessarily infinitesimal. Its geometrical significance is that of a dilatation in $n$ mutually orthogonal directions. To see this let us transform the quadratic form $S(x, x)$ to principal axes, transforming the $x_i$ into $u_i$ and the $y_i$ into $v_i$. We then have

$$\sum_{i,k=1}^{n} s_{ik}x_ix_k = \sum_{i=1}^{n} \kappa_i u_i^2,$$

and the equations $y_i = \sum_k s_{ik}x_k$ become

$$v_i = \kappa_i u_i .$$

These equations evidently represent a dilatation by the factor $\kappa_i$ in the direction of the $i$-th principal axis. The ratio of the increase of volume to the initial volume, known as the volume dilatation, is evidently given by the difference $\kappa_1\kappa_2 \cdots \kappa_n - 1 = | s_{ik} | - 1$. If, in particular, the transformation is infinitesimal, i.e. $(s_{ik}) = E + (\epsilon\beta_{ik})$, we have

$$\kappa_1 \cdots \kappa_n - 1 = \epsilon(\beta_{11} + \cdots + \beta_{nn}).$$

Since an orthogonal transformation represents a rotation we may summarize by stating:

*An infinitesimal transformation whose matrix is $E + (\epsilon\gamma_{ik})$ may be represented as the product of a rotation and a dilatation; the volume dilatation is $\epsilon \sum_{i=1}^{n} \gamma_{ii}$.*

**6. Perturbations.** In the theory of small vibrations and in many problems of quantum mechanics it is important to determine how the eigenvalues and eigenvectors of a quadratic form $K(x, x) = \sum_{i,k=1}^{n} b_{ik}x_ix_k$ are changed if both the form $K(x, x)$ and the unit form $E(x, x)$ are altered. Suppose $E(x, x)$ is replaced by $E(x, x) + \epsilon\mathsf{A}(x, x)$ and $K(x, x)$ by $K(x, x) + \epsilon\mathsf{B}(x, x)$, where

$$\mathsf{A}(x, x) = \sum_{i,k=1}^{n} \alpha_{ik}x_ix_k, \qquad \mathsf{B}(x, x) = \sum_{i,k=1}^{n} \beta_{ik}x_ix_k,$$

and $\epsilon$ is a parameter. The problem is then to transform $E + \epsilon\mathsf{A}$ and $K + \epsilon\mathsf{B}$ simultaneously into canonical form. If we put

$$K(x, x) + \epsilon\mathsf{B}(x, x) = \sum_{i,k=1}^{n} b'_{ik}x_ix_k,$$

$$E(x, x) + \epsilon\mathsf{A}(x, x) = \sum_{i,k=1}^{n} a'_{ik}x_ix_k,$$

the equations for the components of the eigenvectors become

$$\sum_{k=1}^{n} (b'_{ik} - \rho' a'_{ik})x'_k = 0 \qquad (i = 1, \cdots, n),$$

where $\rho'$ may be obtained from the condition that the determinant of this system of equations must vanish. Let us denote the eigenvalues of $K(x, x)$ by $\rho_1, \rho_2, \cdots, \rho_n$ and assume that they are all different; let the corresponding values for the varied system be denoted by $\rho'_1, \rho'_2, \cdots, \rho'_n$. The original form $K(x, x)$ may be assumed to be a sum of squares:

$$K(x, x) = \rho_1x_1^2 + \rho_2x_2^2 + \cdots + \rho_nx_n^2.$$

The quantities $\rho'_i$, being simple roots of an algebraic equation, are single-valued analytic functions of $\epsilon$ in the neighborhood of $\epsilon = 0$; the same is, therefore, true of the components $x'_{hk}$ of the varied eigenvectors belonging to the eigenvalues $\rho'_h$. Thus the quantities $\rho'_h$ and $x'_{hk}$ may be expressed as power series in $\epsilon$, the constant terms of which are, of course, the original eigenvalues $\rho_h$ and the components of the original eigenvectors $x_{hk}$, respectively. In order to compute successively the coefficients of $\epsilon$, $\epsilon^2$, $\cdots$ we must substitute these power series in the equations

$$\sum_{k=1}^{n} (b'_{ik} - \rho'_h a'_{ik})x'_{hk} = 0 \qquad (i, h = 1, \cdots, n)$$

in which we have $b'_{ik} = \rho_{ik} + \epsilon\beta_{ik}$, $a'_{ik} = \delta_{ik} + \epsilon\alpha_{ik}$, with $\rho_{ii} = \rho_i$, $\rho_{ik} = 0$ $(i \neq k)$, $\delta_{ii} = 1$, $\delta_{ik} = 0$ $(i \neq k)$. By collecting the terms in each power of $\epsilon$ in these equations and then setting the coefficient of each power of $\epsilon$ equal to zero we obtain an infinite sequence of new equations. An equivalent procedure, which is often somewhat more convenient, is given by the following considerations of orders of magnitude, in which $\epsilon$ is regarded as an infinitesimal quantity. We first consider the equation with $i = h$. By setting the coefficient of the first power of $\epsilon$ equal to zero we obtain

$$\rho_h' = \frac{\rho_h + \epsilon\beta_{hh}}{1 + \epsilon\alpha_{hh}} = \rho_h - \epsilon\rho_h\alpha_{hh} + \epsilon\beta_{hh},$$

except for terms of the second or higher orders in $\epsilon$. The same procedure applied to the equations with $i \neq h$ yields the result

$$x_{hh}' = 1, \qquad x_{hi}' = -\epsilon\,\frac{\alpha_{ih}\rho_h - \beta_{ih}}{\rho_h - \rho_i},$$

except for infinitesimal quantities of the second order in $\epsilon$.

By using these values of the components of the eigenvectors we may easily obtain the eigenvalues up to and including the second order in $\epsilon$. Again we consider the $h$-th equation for the components of the $h$-th eigenvector:

$$\sum_{k=1}^{n} (b_{hk}' - \rho_h' a_{hk}')x_{hk}' = 0.$$

If we neglect quantities of the third order in $\epsilon$ on the left-hand side and write the term with $h = k$ separately, we obtain

$$\begin{aligned} b_{hh}' - \rho_h' a_{hh}' &= \sum_{k=1}^{n}{}' \epsilon(b_{hk}' - \rho_h' a_{hk}')\,\frac{\alpha_{kh}\rho_h - \beta_{kh}}{\rho_h - \rho_k} \\ &= -\epsilon^2 \sum_{k=1}^{n}{}' \frac{(\alpha_{kh}\rho_h - \beta_{kh})^2}{\rho_h - \rho_k}. \end{aligned}$$

It follows that

$$\rho_h' = \rho_h - \epsilon(\rho_h\alpha_{hh} - \beta_{hh}) - \epsilon^2\alpha_{hh}(\beta_{hh} - \rho_h\alpha_{hh}) + \epsilon^2 \sum_{k=1}^{n}{}' \frac{(\alpha_{kh}\rho_h - \beta_{kh})^2}{\rho_h - \rho_k}.$$

Here we have used the symbol $\sum_k'$ to denote summation over all values of $k$ from 1 to $n$ except for $k = h$.

**7. Constraints.** Constraints expressed by linear conditions

$$\gamma_1 x_1 + \cdots + \gamma_n x_n = 0,$$

and the resulting diminution of the number of independent variables of the quadratic form $K(x, x) = \sum_{p,q=1}^{n} k_{pq}x_p x_q$, may be regarded as the end result of a continuous process. Consider the quadratic form $K(x, x) + t(\gamma_1 x_1 + \cdots + \gamma_n x_n)^2$, where $t$ is a positive parameter. If $t$ increases beyond all bounds, each eigenvalue increases monotonically. The greatest eigenvalue increases beyond all bounds, while the others approach the eigenvalues of the quadratic form which is obtained from $K(x, x)$ by elimination of one variable in accordance with the given constraint.

**8. Elementary Divisors of a Matrix or a Bilinear Form.** Let **A** be a tensor and $A = (a_{ik})$ the corresponding matrix. Then the polynomial

$$|\kappa E - A| = \begin{vmatrix} \kappa - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \kappa - a_{22} & \cdots & -a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -a_{n1} & -a_{n2} & \cdots & \kappa - a_{nn} \end{vmatrix}$$

may be decomposed according to certain well-known rules into the product of its "elementary divisors"

$$(\kappa - r_1)^{e_1}, (\kappa - r_2)^{e_2}, \cdots, (\kappa - r_h)^{e_h},$$

where some of the numbers $r_1, r_2, \cdots, r_h$ may be equal. For each divisor $(\kappa - r_\nu)^{e_\nu}$ there is a system of $e_\nu$ vectors $\mathbf{f}_1^{(\nu)}, \mathbf{f}_2^{(\nu)}, \cdots, \mathbf{f}_{e_\nu}^{(\nu)}$ such that the equations

$$\mathbf{A}\mathbf{f}_1^{(\nu)} = r_\nu \mathbf{f}_1^{(\nu)}, \quad \mathbf{A}\mathbf{f}_2^{(\nu)} = r_\nu \mathbf{f}_2^{(\nu)} + \mathbf{f}_1^{(\nu)}, \cdots, \mathbf{A}\mathbf{f}_{e_\nu}^{(\nu)} = r_\nu \mathbf{f}_{e_\nu}^{(\nu)} + \mathbf{f}_{e_\nu-1}^{(\nu)}$$

are valid. Here the $n$ vectors

$$\mathbf{f}_1^{(1)}, \cdots, \mathbf{f}_{e_1}^{(1)}; \quad \mathbf{f}_1^{(2)}, \cdots, \mathbf{f}_{e_2}^{(2)}; \cdots; \mathbf{f}_1^{(h)}, \cdots, \mathbf{f}_{e_h}^{(h)}$$

are linearly independent. If they are introduced as new variables $x_1^{(1)}, x_2^{(2)}, \cdots, x_{e_h}^{(h)}$, the matrix $A$ is transformed into the matrix

$$\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & A_h \end{pmatrix}$$

in which $A_1, A_2, \cdots, A_h$ are themselves matrices; $A_\nu$ is a matrix of order $e_\nu$:

$$A_\nu = \begin{pmatrix} r_\nu & 0 & 0 & \cdots & 0 & 0 \\ 1 & r_\nu & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & r_\nu \end{pmatrix}.$$

**9. Spectrum of a Unitary Matrix.** We shall now show that the spectrum of a unitary matrix lies on the unit circle, i.e. that all of its eigenvalues have the absolute value 1.

We note that the elements of a unitary matrix cannot exceed unity in absolute value. Therefore the absolute values of the coefficients of the characteristic equations of all unitary matrices of the $n$-th degree must lie below a certain bound which is independent of the particular matrix considered. Since the absolute values of the first and last coefficients of the characteristic equation are equal to 1, this means that the absolute values of the eigenvalues must lie between certain positive upper and lower bounds which are independent of the particular matrix. On the other hand all powers $A^m$ of a unitary matrix $A$ are also unitary, and their eigenvalues are the $m$-th powers of the corresponding eigenvalues of $A$. But the absolute values of these powers and their reciprocals can remain below a bound which is independent of $m$ only if the absolute value of each eigenvalue (and all of its powers) is 1.

Another proof, which can be used for infinite matrices as well, follows from the convergence of the Neumann series for $(E - \lambda A)^{-1}$. The series

$$(E - \lambda A)^{-1} = E + \lambda A + \lambda^2 A^2 + \cdots,$$

where $A$ is a unitary matrix, certainly converges if $|\lambda| < 1$. For the elements of the matrices $A^m$ all have absolute values of at most 1, and thus the geometric series is a dominating series for the matrix elements. Thus no zeros of $|E - \lambda A|$ can lie inside the unit circle. On the other hand we have, in virtue of the relation $A\bar{A}' = E$,

$$(E - \lambda A)^{-1} = -\frac{1}{\lambda}\overline{A'}\left(E + \frac{1}{\lambda}\overline{A'} + \frac{1}{\lambda^2}\overline{A'^2} + \cdots\right).$$

Here the geometric series on the right converges for $| 1/\lambda | < 1$ since $\bar{A}'$ is also a unitary matrix. Thus no zero of $| E - \lambda A |$ can lie outside the unit circle. Therefore all these zeros lie on the unit circle, and our assertion is proved.

### *References*

*Textbooks*

Bôcher, M., Introduction to Higher Algebra. Macmillan, New York, 1907.

Kowalewski, G., Einführung in die Determinantentheorie. Veit, Leipzig, 1909.

Wintner, A., Spektraltheorie der unendlichen Matrizen. S. Hirzel, Leipzig, 1929.

*Monographs and Articles*

Courant, R., Zur Theorie der kleinen Schwingungen. Zts. f. angew. Math. u. Mech., Vol. 2, 1922, pp. 278–285.

Fischer, E., Über quadratische Formen mit reelen Koeffizienten. Monatsh. f. Math. u. Phys., Vol. 16, 1905, pp. 234–249. The maximum-minimum character of the eigenvalues probably was first mentioned in this paper.

Hilbert, D., Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen, especially sections 1 and 4. Fortschritte der mathematischen Wissenschaften in Monographien, Heft 3. B. G. Teubner, Leipzig and Berlin, 1912.