# LINEAR SYSTEMS IN SELF-ADJOINT FORM

C. LANCZOS, Dublin Institute for Advanced Studies

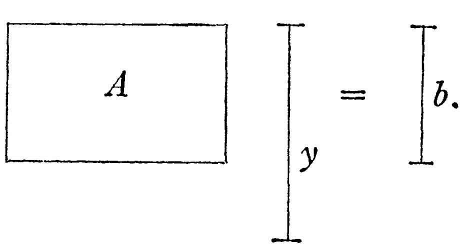
The principal axis problem of quadratic forms belongs to the best-investigated chapters of analysis. Much attention has been paid to the theory of matrices subjected to arbitrary linear transformations and the normal forms attainable by such transformations; (*cf.* [1],* p. 58). The relation of the general theory of linear equations to matrix calculus has also found exhaustive treatment; (*cf.* [2], Chapter IV). It seems, therefore, that the fundamental aspects of this field are practically exhausted.†

It is the purpose of the following discussions to approach the problem of general linear algebraic systems from a somewhat different viewpoint which throws new light on the nature of the principal axis problem by showing that the properties of symmetric matrices are extendable to arbitrary matrices to a surprisingly large degree, without demanding anything but *orthogonal* transformations. In this way the classical theory of linear algebraic forms, developed by Frobenius and Kronecker around the end of the last century (*cf.* [4], p. 268), which centers around the concept of the "rank" of a matrix, is elucidated from a totally different angle in which a certain *eigenvalue problem* plays the central role.

We formulate the given simultaneous system of $n$ equations in $m$ unknowns as the matrix equation

$$Ay = b, \tag{1}$$

in which $A$ denotes an arbitrary $n$-row, $m$-column matrix—briefly denoted as an $n \times m$ matrix—while $y$ is the unknown and $b$ the given right side, both column vectors of $m$, respectively, $n$ components. The matrix diagram associated with our system, (if we picture the case $n < m$), looks as follows:




We will replace this diagram by the following extended diagram:‡

* Numbers in square brackets refer to the references at the end of the paper.

† A very extensive literature is digested in [3].

‡ The symbol "tilde" ($\sim$) refers to a transposition of rows and columns. If $A$ has complex elements, the transposition shall include a change of $i$ to $-i$ in every element. This generalization is so obvious that we will assume the reality of $A$ and call $S$ a "symmetric" rather than "Hermitian" matrix.

$$\left[\begin{array}{c|c} 0 & A \\ \hline \tilde{A} & 0 \end{array}\right] \left[\begin{array}{c} 0 \\ \hline y \end{array}\right] = \left[\begin{array}{c} b \\ \hline 0 \end{array}\right]$$

This diagram belongs to the equation $Sw=g$, where the symmetric square matrix $S=\tilde{S}$ is defined as follows:

$$S = \left[\begin{array}{c|c} 0 & A \\ \hline \tilde{A} & 0 \end{array}\right], \tag{3}$$

while the column vectors $w$ and $g$ have $n+m$ components, which can be displayed more conveniently by transposing them into row vectors:

$$\begin{aligned} \tilde{w} &= \begin{array}{|c|c|} \multicolumn{1}{c}{(n)} & \multicolumn{1}{c}{(m)} \\ \hline 0 & y \end{array}, \\ \tilde{g} &= \begin{array}{|c|c|} \hline b & 0 \end{array}. \end{aligned} \tag{4}$$

On the other hand, if these two vectors have the following structure:

$$\begin{aligned} \tilde{w} &= \begin{array}{|c|c|} \multicolumn{1}{c}{(n)} & \multicolumn{1}{c}{(m)} \\ \hline x & 0 \end{array}, \\ \tilde{g} &= \begin{array}{|c|c|} \hline 0 & c \end{array}, \end{aligned} \tag{5}$$

we obtain the equation

$$Ax = c, \tag{6}$$

and the number of equations is now *greater* than the number of unknowns. Hence we can without loss of generality assume that $n \leqq m$ and put

$$m = n + r \qquad (r \geqq 0). \tag{7}$$

(In our final results all reference to a preestablished relation between $n$ and $m$ will disappear and the numbers $n$ and $m$ be left completely arbitrary.) A general vector $u$ of the $n+m$-dimensional space associated with the matrix $S$ will be introduced as follows:

$$\tilde{u} = |\underset{x}{\overset{(n)}{\rule{2cm}{0.4pt}}}|\underset{y}{\overset{(m)}{\rule{3cm}{0.4pt}}}|. \tag{8}$$

Since $S$ is symmetric, we can transform it into a diagonal matrix $D$ with the help of an orthogonal transformation (the "principal-axis transformation"). The matrix $U$ of this orthogonal transformation is composed of the $n+m$ column vectors $u_i$ which satisfy the following eigenvalue problem:

$$Su_i = \lambda_i u_i \qquad (i = 1, 2, \cdots n + m) \tag{9}$$

with the added normalization

$$|u_i|^2 = 1. \tag{10}$$

As is well known, the eigenvalues $\lambda_i$ of our problem are all *real.*

In view of the special form of the matrix $S$ our eigenvalue problem separates into the two equations

$$Ay_i = \lambda_i x_i, \quad \tilde{A}x_i = \lambda_i y_i. \tag{11}$$

We will call this system of equations the "shifted eigenvalue problem" because on the right side the vectors $x_i$ and $y_i$ are in a "shifted" position, compared with the usual eigenvalue problem associated with $A$ and $\tilde{A}$.

The eigenvalue problem (11) has a number of interesting properties which we are now going to demonstrate.

1. If we premultiply the second equation by $A$ and substitute on the right side for $Ay_i$ its value from the first equation, we obtain

$$A\tilde{A}x_i = \lambda_i^2 x_i. \tag{12}$$

Similarly, premultiplying the first equation by $\tilde{A}$ and making use of the second equation we obtain

$$\tilde{A}Ay_i = \lambda_i^2 y_i. \tag{13}$$

We thus see that the vectors $x_i$ and $y_i$ can be defined *separately* in themselves. They are the eigenvectors (principal axes) of the nonnegative symmetric matrices $A\tilde{A}$, respectively, $\tilde{A}A$. The first matrix is an $n \times n$, the second an $m \times m$ symmetric matrix. Hence the vectors $x_i$ and $y_i$ belong to two completely different spaces, the one of $n$, the other of $m$ dimensions. The vectors $x_i$, put together columnwise, form an $n \times n$ complete orthogonal matrix $X$:

$$\tilde{X}X = X\tilde{X} = I. \tag{14}$$

The same can be done with the vectors $y_j$, resulting in the complete $m \times m$ orthogonal matrix $\overline{Y}$:

$$\tilde{\overline{Y}}\overline{Y} = \overline{Y}\tilde{\overline{Y}} = I. \tag{15}$$

2. The sequence of the columns of these two matrices is not arbitrary. The vectors $x_i$ and $y_i$, although defined independently by the equations (12) and (13), are in actual fact *paired*, since they are also in the relation (11) to each other. To every $y_i$ the corresponding $x_i$ can be found by

$$x_i = \frac{1}{\lambda_i} A y_i \tag{16}$$

and vice versa, provided that $\lambda_i$ is not zero. In fact, every solution of the system (11) with nonvanishing $\lambda_i$ can immediately be extended to a *second* pair of vectors since a simultaneous change of $\lambda_i$ to $-\lambda_i$ and $y_i$ to $-y_i$ leaves the equations (11) unchanged. Accordingly, all nonvanishing eigenvalues will appear in *pairs* $\pm\lambda_i$ and we can agree that what we will call $\lambda_i$, shall be a *positive* number, complemented by the negative eigenvalue $-\lambda_i$. For the time being we want to assume that all the $\lambda_i^2$ associated with the eigenvalue problem (12) are different from zero. This provides us with the $n$ positive numbers $\lambda_1, \cdots, \lambda_n$, complemented by the sequence $-\lambda_1, \cdots, -\lambda_n$ of negative eigenvalues.

3. An interesting property of the $x_i$, $y_i$ vectors is that their length is automatically *equal*. We see from (16) that we get

$$\tilde{x}_i x_i = \frac{1}{\lambda_i^2} \tilde{y}_i \tilde{A} A y_i = \tilde{y}_i y_i. \tag{17}$$

For this reason, if the length of the vector $x_i$ is normalized to 1, the length of the corresponding vector $y_i$ becomes automatically 1.

4. The pairing of the vectors $x_i$, $y_i$ cannot occur unlimitedly, since we cannot have more than $n$ $x$-vectors while the number of $y$-vectors is $m$. Having obtained the $2n$ pairs $(x_i, y_i)$ and $(x_i, -y_i)$, the remaining $y_k$-vectors can have no $x_k$-associates, which is only possible if $\lambda_k = 0$. Hence the eigenvalue zero is always present among the eigenvalues if $r$ is not zero. Since the total number of eigenvalues is $n+m=2n+r$, we must have $r$ eigenvectors associated with the eigenvalue zero. The corresponding $x_k$-vectors vanish. We consider these additional $y$-vectors as columns of a matrix $Y_0$ which has $m$ rows and $r$ columns. The complete $m \times m$ $\overline{Y}$-matrix is thus composed of the $m \times n$ $Y$-matrix, associated with the matrix $X$, and the additional $m \times r$ matrix $Y_0$, associated with the zero matrix. The complete $U$-matrix which contains all the principal axes of the matrix $S$, in the sequence $\lambda = -\lambda_1, \cdots, -\lambda_n; \lambda_1, \cdots, \lambda_n; 0, \cdots, 0$; appears in the form (18) while the diagonal matrix $D$ into which $S$ is transformed if we rotate it into the reference system of the principal axes, becomes (19). Here $\Lambda$ denotes the $n \times n$ diagonal matrix whose diagonal elements are the $n$ positive numbers $\lambda_1, \cdots, \lambda_n$.

Let us observe that the columns of the matrix $U$, being composed of the two

$$
\text{(18)} \qquad U = \begin{array}{c} \\ n \\ m \end{array} \begin{array}{c} \begin{array}{ccc} n & n & m-n \end{array} \\ \left[\begin{array}{c|c|c} X & X & 0 \\ \hline -Y & Y & \sqrt{2}Y_0 \end{array}\right] \end{array} ;
$$

$$
\text{(19)} \qquad D = \begin{array}{c} \\ n \\ n \\ m-n \end{array} \begin{array}{c} \begin{array}{ccc} n & n & m-n \end{array} \\ \left[\begin{array}{c|c|c} -\Lambda & & \\ \hline & \Lambda & \\ \hline & & 0 \end{array}\right] \end{array}
$$

vectors $x_i$ and $y_i$, both of the length 1, are normalized to $\sqrt{2}$ instead of 1. For this reason the columns of the matrix $Y_0$ have to be multiplied by $\sqrt{2}$, in order to uniformize all lengths of the columns of $U$. The right side of (10) has to be changed accordingly from 1 to 2.

We will now drop the restricting condition that all the $\lambda_i$ of the eigenvalue problem (12) are positive. Generally only $p \leqq n$ eigenvalues need be different from zero. We thus introduce here a new number $p$ which is characteristic for the matrix $A$, in addition to the two numbers $n$ and $m$. This number, which in fact coincides with the "rank" of the matrix $A$, is here defined by the *number of independent eigensolutions of the system* (11) *which are possible if we demand that the eigenvalue* $\lambda_i$ *shall be a positive number.* This $p$ can take any value between 1 and the smaller of the two numbers $(n, m)$:

$$
\text{(20)} \qquad 1 \leqq p \leqq \min (n, m).
$$

The case $p=0$ is excluded if $A$ does not vanish identically because, if all eigenvalues of $A\tilde{A}$ are zero, the whole matrix $A\tilde{A}$ vanishes, which is only possible if $A=0$.

In view of this new number $p$ the previous picture of the matrix $U$ and the diagonal matrix $D$ changes to some extent. The multiplicity of the eigenvalue zero has now increased from $m-n$ to $m+n-2p$. The matrix $X$, associated with the nonzero eigenvalues, is no longer an $n \times n$ but an $n \times p$ matrix. Moreover, the diagonal matrix $\Lambda$ which appears in the construction of the matrix $D$, is no longer an $n \times n$ but a $p \times p$ matrix, composed of the positive numbers $\lambda_1, \cdots, \lambda_p$:

$$(21)\qquad U = \begin{array}{c|c|c|c|c|c|}
 & p & p & \multicolumn{2}{c}{n-p} & m-n \\ \hline
n & X & X & X_0' & X_0' & 0 \\ \hline
m & -Y & Y & -Y_0' & Y_0' & \sqrt{2}\,Y_0'' \\ \hline
\end{array}$$

$$(22)\qquad D = \begin{array}{c|c|c|c|}
 & p & p & m+n-2p \\ \hline
p & -\Lambda & & \\ \hline
p & & \Lambda & \\ \hline
m+n-2p & & & 0 \\ \hline
\end{array}$$

We have thus obtained a detailed picture of the structure of the matrices $U$ and $D$ which are associated with the principal axis transformation of the matrix $S$. The entire principal axis problem is included in the matrix equation*

$$SU = UD \tag{23}$$

with the added condition

$$\tilde{U}U = 2I. \tag{24}$$

If now we postmultiply (23) by $\tilde{U}$, we obtain the relation

$$2S = UD\tilde{U}. \tag{25}$$

We wish to construct $S$ on the basis of this relation, performing the matrix multiplication indicated on the right side of (25). First we obtain the matrix

* For this formulation of the principal axis problem *cf.* [5], p. 93.

$UD$. We know that multiplication by a diagonal matrix as a second factor means that the columns of the first matrix are in succession multiplied by the successive diagonal elements of the second matrix. This gives

$$UD = \begin{array}{c} \\ n \\ m \end{array} \begin{array}{ccc} p & p & n+m-2p \\ \hline -X\Lambda & X\Lambda & 0 \\ Y\Lambda & Y\Lambda & 0 \end{array} . \tag{26}$$

Now we should transpose $U$ and postmultiply by it. We can, however, leave $U$ in its original form (21), if we agree that "row by column multiplication" is changed to "row by row multiplication." We will indicate this kind of multiplication by a little circle $\circ$. First of all we multiply the first $n$ rows of (26) by the first $n$ rows of (21). This gives $-X\Lambda \circ X + X\Lambda \circ X + 0 = 0$. We continue by multiplying the first $n$ rows of (26) by the last $m$ rows of (21). This gives $X\Lambda \circ Y + X\Lambda \circ Y + 0 = 2X\Lambda \circ Y = 2X\Lambda\tilde{Y}$. We now come to the product of the last $m$ rows of (26) with the first $n$ rows of (21): $Y\Lambda \circ X + Y\Lambda \circ X + 0 = 2Y\Lambda \circ X = 2Y\Lambda\tilde{X}$. Finally the last $m$ rows of (26) multiplied by the last $m$ rows of (21) yield $-Y\Lambda \circ Y + Y\Lambda \circ Y + 0 = 0$. The complete result is the following matrix:

$$2S = \begin{array}{|c|c|} \hline 0 & 2X\Lambda\tilde{Y} \\ \hline 2Y\Lambda\tilde{X} & 0 \\ \hline \end{array} . \tag{27}$$

Comparison with the original form (3) of $S$ gives the following fundamental result:

$$A = X\Lambda\tilde{Y}, \tag{28}$$

$$\tilde{A} = Y\Lambda\tilde{X}. \tag{29}$$

The second equation contains no new statement since it merely repeats the first relation in transposed form. The equation (28) contains the following fundamental

DECOMPOSITION THEOREM. *An arbitrary nonzero matrix can be written as the product of the $n\times p$ orthogonal matrix $X$, $(\tilde{X}X = I)$, the $p\times p$ positive diagonal matrix $\Lambda$ and the transpose of the $m\times p$ orthogonal matrix $Y(\tilde{Y}Y = I)$.*

The remarkable fact about this theorem is that *the principal axes associated with the zero eigenvalue do not participate at all in the formation of the matrix $A$.* This has a profound effect on the solution problem of the equation $Ay=b$.

The matrices $X$, $Y$ and $\Lambda$ which appear in this theorem, are defined by the shifted eigenvalue problem

$$AY = X\Lambda, \qquad \tilde{A}X = Y\Lambda, \tag{30}$$

with the added condition that the diagonal elements of $\Lambda$ shall all be nonzero positive numbers.

*Example.* Consider the $n\times m$ matrix whose elements are all zero, except the single element $a_{ij}$ which may be given as the complex number $c$. Show that in this problem $p=1$, $\lambda_1=|c|$; all the $n$ elements of the vector $x_1$ vanish except the element $x_1^{(i)}$; all the $m$ elements of the vector $y_1$ vanish, except the single element $y_1^{(j)}$. Demonstrate the validity of the relation (28).

Before we continue with the further analysis of our problem, let us recall the two full $n\times n$ and $m\times m$ spaces associated with our eigenvalue problem. We can picture them as follows:

$$\begin{array}{c|c|c|} & p & n-p \\ \hline n & X & X_0 \\ \hline \end{array}\,, \qquad \begin{array}{c|c|c|} & p & m-p \\ \hline m & Y & Y_0 \\ \hline \end{array}\,.$$

The matrices $X$ and $Y$ are composed of $p$ mutually orthogonal axes of the two respective spaces. They form a $p$-dimensional subspace within the full space. Hence it is generally not permissible to convert the relation $\tilde{X}X=I$ to $X\tilde{X}=I$, and the same holds for the matrix $Y$. The remaining axes, included in the $n\times(n-p)$ orthogonal matrix $X_0$, respectively the $m\times(m-p)$ orthogonal matrix $Y_0$, belong to the eigenvalue zero and are thus defined by the two noninterrelated equations

$$\tilde{A}X_0 = 0, \tag{31}$$

respectively,

$$AY_0 = 0. \tag{32}$$

The columns of these two matrices are thus composed of the solutions of the homogeneous equation

$$\tilde{A}\overset{0}{x}_i = 0 \qquad (i = 1, \cdots, n-p), \tag{33}$$

respectively,

$$A\overset{0}{y}_j = 0 \qquad (j = 1, \cdots, m-p). \tag{34}$$

In harmony with the orthogonal nature of principal axes we assume that the vectors $x_i^0$ are mutually orthogonal and their length is 1. The same can be said of the vectors $y_j^0$. This, however, is not self-evident since the principal axes belonging to a multiple eigenvalue (in this case the eigenvalue zero), are not orthogonal by nature, although they *can* be orthogonalized. We will have use for the vectors $\xi_i$, which are merely solutions of the homogeneous equation

$$\tilde{A}\xi_i = 0 \qquad (i = 1, \cdots, n - p) \tag{35}$$

*without* demanding their orthogonalization and the normalization of their length. We merely demand that they shall be linearly independent and that their number shall be $n-p$, in order to span the entire space $X_0$. Similarly we will consider the $m-p$ linearly independent solutions of the homogeneous equation

$$A\eta_j = 0 \qquad (j = 1, \cdots, m - p) \tag{36}$$

without demanding their orthogonalization and normalization. The vectors $\xi_i$, taken as column vectors, form the $n \times (n-p)$ matrix $\Xi_0$, the vectors $\eta_j$ the $m \times (m-p)$ matrix $H_0$. These matrices are no longer orthogonal but their orthogonality to the subspaces $X$ and $Y$ remains unchanged:

$$\tilde{\Xi}_0 X = 0, \qquad \tilde{H}_0 Y = 0. \tag{37}$$

With this preliminary information we return to the study of the linear equation (1) which could be done by diagonalization in the reference system of the principal axes of $S$, but we prefer to draw all our conclusions from the decomposition of our matrix $A$ into the product (28). The equation (1) can now be written in the form

$$X\Lambda\tilde{Y}y = b. \tag{38}$$

Premultiplication by the matrix $\tilde{\Xi}_0$ gives

$$0 = \tilde{\Xi}_0 b. \tag{39}$$

This equation, if written in the language of vectors, becomes

$$(\xi_i \cdot b) = 0 \qquad (i = 1, \cdots, n - p) \tag{40}$$

and we obtain the following (well-known)

COMPATIBILITY THEOREM. *The equation $Ay=b$ is solvable if and only if the given right side of the equation is orthogonal to every independent solution of the adjoint homogeneous equation $\tilde{A}\xi=0$.*

That the condition (40) is *necessary*, follows from (39). That it is also *sufficient* follows from the fact that if $b$ is perpendicular to the space $\Xi_0$ (or the equivalent space $X_0$), it must lie inside the space $X$, *i.e.* it must have the form

$$b = Xb'. \tag{41}$$

But in that case the equation (38)—premultiplying it by $\tilde{X}$—gives at once

$$\Lambda \tilde{Y} y = b'. \tag{42}$$

This equation is solvable by putting

$$y = Yy', \tag{43}$$

in which case we obtain

$$\Lambda y' = b', \; y' = \Lambda^{-1} b' \tag{44}$$

and finally

$$y = Y\Lambda^{-1} b' = Y\Lambda^{-1} \tilde{X} b. \tag{45}$$

The diagonal matrix $\Lambda$ contains only nonzero elements in the diagonal and is thus always invertible.

However, the solution (45) is not the *only* solution of our system. An arbitrary vector $y$, if analyzed in the reference system of the principal axes $(Y, Y_0)$, appears in the following form:

$$y = Yy' + Y_0 y_0'.$$

If we put this expression in (38), we observe that the term with $y_0'$ *drops out completely* from our equation. We thus obtain, as a counterpart of our previous Compatibility Theorem, the following

DEFICIENCY THEOREM: *The equation $Ay = b$ determines uniquely the projection of the vector $y$ into the space $Y$ but leaves its projection into the space $Y_0$ completely undetermined.*

We will now consider the solution of our system (1) under the following *auxiliary conditions*:

$$\tilde{X}_0 b = 0 \qquad \text{(by necessity)}, \tag{46}$$

$$\tilde{Y}_0 y = 0 \qquad \text{(by choice)}. \tag{47}$$

The second condition is not demanded by the original equation. By adding this condition we obtain a definite *particular solution* distinguished by the property that the solution finds its place in a *subspace of smallest capacity*, *viz.* the space $Y$.* This brings us back to the condition (43) and thus to the solution (45).

We know from the general theory of linear operators that the general solution of a linear system of equations is obtainable by adding to any particular solution an arbitrary solution of the homogeneous equation. Applying this principle to our problem we get

$$y = y_p + Y_0 \eta, \tag{48}$$

where $\eta$ is an arbitrary column vector of $m-p$ elements while for $y_p$ we can choose the particular solution (45), obtained under the auxiliary condition (47).

* This "normalization condition," which makes the solution unique, is equivalent to putting $y = \tilde{A} v$, where the vector $v$ is unrestricted.

We can interpret the deficiency of the given system in the following terms. The equation (1) is not sufficient for the determination of $y$ but it *may* become sufficient by *added information.* The information needed is a statement concerning the projection of $y$ into the space $Y_0$. Hence we can conceive the equation (1) as *part of a more elaborate system,* the addition taking the form

$$\tilde{Y}_0 y = b_0, \tag{49}$$

where $b_0$ is a free column vector of $m-p$ components. In this case the previous solution $y=y_p$ is not more than a *preliminary result,* while the complete solution takes the form (48). But now premultiplication by $\tilde{Y}_0$ yields

$$\eta = b_0 \tag{50}$$

and the complete solution—obtained after complementing the original system (1) by the additional system (49)—becomes

$$y = y_p + Y_0 b_0 = Y\Lambda^{-1}\tilde{X}b + Y_0 b_0. \tag{51}$$

The result of this analysis may be summarized as follows: The zero-fields $X_0$, $Y_0$, associated with the solutions of the homogeneous equations $\tilde{A}x=0$, $Ay=0$, do not participate directly in the solution of the linear system (1) but merely decide the *compatibility* and the *deficiency* of the system. The compatibility conditions (46) have to be assumed in order to have a solution at all. The deficiency of the system can be removed by putting the solution in the space $Y$, assuming that the added information (49) will later provide the missing $Y_0$-portion of the solution. By this procedure an arbitrarily over-determined (although compatible) or under-determined (and thus deficient) linear system permits a unique solution.

Our solution can be put in the form

$$y = Gb, \tag{52}$$

where the $m\times n$ matrix $G$ is defined as follows:

$$G = Y\Lambda^{-1}\tilde{X}. \tag{53}$$

In view of the form (52) of the solution we can conceive the matrix $G$ as the "inverse" of the matrix $A$. We should thus expect that the product $GA$ has the property of the unit matrix $I$. It would be a mistake, however, to assume that the product $GA$ must come out as the unit matrix. The product $GA$ does not operate on an *arbitrary* vector $b$ but on a vector which is subject to the condition (46). This means that $b$ is inside the space $X$ and has thus the form (41). Now

$$AG = X\Lambda\tilde{Y}Y\Lambda^{-1}\tilde{X} = X\tilde{X} \tag{54}$$

and $AG$ operating on $b$ becomes

$$AGb = X\tilde{X}b = X\tilde{X}Xb' = Xb' = b, \tag{55}$$

which shows that $AG$ has in fact the property of the unit matrix with respect to all "permissible" vectors $b$.

On the other hand, let us premultiply (1) by $G$:

$$GAy = Gb = y. \tag{56}$$

This shows that the product $GA$ must also have the property of the unit matrix, but again operating on a special class of vectors, subject to the condition (47). This condition puts $y$ into the space $Y$ which means that $y$ can be put in the form (43). Now

$$GA = Y\Lambda^{-1}\tilde{X}X\Lambda\tilde{Y} = Y\tilde{Y} \tag{57}$$

and therefore

$$GAy = Y\tilde{Y}y = Y\tilde{Y}Yy' = Yy' = y. \tag{58}$$

Once again the product $GA$ has the property of the unit matrix $I$ with respect to all permissible vectors $y$.

The matrix (53) has all the properties demanded by E. H. Moore in his "general analysis," (1906); (*cf.* [6]), establishing the "generalized inverse" of a matrix in abstract terms. We can likewise demonstrate that the conditions demanded by R. Penrose (*cf.* [7]; see also R. Rado, [8]) concerning the generalized inverse of a matrix are fulfilled. In our analysis the inverse matrix did not come about by any definitions in terms of matrix equations but by an *explicit method* of solving the linear system (1), based on the properties of an eigenvalue problem.

We can write out the solving matrix more explicitly by substituting for $X$ and $Y$ the constituting column vectors $x_i$ and $y_i$ which appeared in the solution of the eigenvalue problem (11). Let us denote the components of the $p$ vectors $x_\alpha$ by $x_\alpha^{(i)}$ $(i=1, \cdots, n)$, the components of the conjugate vectors $y_\alpha$ by $y_\alpha^{(j)}$ $(j=1, \cdots, m)$. Then the element $g_{ij}$ of the matrix $G$ comes out as follows:

$$g_{ij} = \sum_{\alpha=1}^{p} \frac{\overset{(i)}{y_\alpha}\ \overset{(j)}{x_\alpha}}{\lambda_\alpha}, \tag{59}$$

while the element $a_{ij}$ of the original matrix $A$ becomes:

$$a_{ij} = \sum_{\alpha=1}^{p} \overset{(i)}{x_\alpha}\lambda_\alpha \overset{(j)}{y_\alpha}. \tag{60}$$

More important, however, is another interpretation of the solution (45). We know that the vector $y$ lies inside the space $Y$ which is composed of the $p$ orthogonal vectors $y_1, \cdots, y_p$. Hence $y$ can be analyzed in terms of these vectors:

$$y = \eta_1 y_1 + \eta_2 y_2 + \cdots + \eta_p y_p. \tag{61}$$

On the other hand, the right side $b$ lies inside the space $X$ and can be analyzed in terms of the conjugate orthogonal vectors $x_1, \cdots, x_p$:

(62) $$b = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

where the coefficients $\beta_i$ are obtainable by projecting $b$ on the axes $X_i$:

(63) $$\beta_i = (b \cdot x_i).$$

We will call the two conjugate expansions (61) and (62) "co-orthogonal" since they involve two sets of orthogonal vectors which are in a one-to-one correspondence to each other.* Then the linear system (1), under the added auxiliary conditions (46) and (47), establishes the following relation between the coefficients $\beta_i$ and $\eta_i$:

(64) $$\eta_i = \beta_i / \lambda_i.$$

We see that the eigenvalues $\lambda_i$ play the role of a "transfer function" in going from the right to the left, or from the left to the right.

Since the coefficients $\beta_i$ are available on the basis of (63), the coefficients $\eta_i$ become determined on the basis of (64) and the unknown vector $y$ appears in the form of an orthogonal expansion (61), with given coefficients.

*Operations in function space.* The field of continuous linear operators—*i.e.*, the domain of linear differential or integral equations—can be handled on the basis of matrix operations if we introduce the infinite-dimensional "function space" and the matrices associated with this space; (*cf.* [9], p. 57). The results obtained in the theory of solving arbitrary linear algebraic systems can thus be extended to the realm of linear differential and integral equations. The characteristic feature of our investigation was that we have dealt with an arbitrarily over-determined or under-determined system and yet arrived at a unique solution under the proper auxiliary conditions.

The usual type of boundary-value problems considered in classical analysis are of the so-called "well-posed" type. This means that the given data—the differential equation with a given right side plus the boundary conditions—suffice for a unique solution and that the data can be prescribed freely, without the danger of incompatibility. Such problems realize in the language of matrices the case $n = m = p$: the number of equations is equal to the number of unknowns and the eigenvalue zero is not present (the matrix $A$ is nonsingular).

Our investigation has shown that we can expect a valid and unique solution of a linear system under much more general conditions. The given operator (including the boundary conditions) may or may not comprise all the dimensions of the function space. The classical case usually considered belongs to those operators which comprise the *entire* function space, in both $X$ and $Y$ relations, *i.e.* in relation to the given right side as well as in relation to the unknown function. If we have a problem which is not "well-posed," this merely means that

* The expression "bi-orthogonal" would be misleading since it usually refers to two sets of *mutually orthogonal* vectors, while the two expansions (61) and (62) involve two sets of vectors which are orthogonal *within themselves*.

the given operator *omits* certain dimensions of the function space, either with respect to the given right side, or with respect to the unknown, or with respect to both. This, however, is no reason to reject the given boundary value problem. If the omission occurs with respect to the $X$-space, this means that the given data cannot be given freely but must be contained in a certain subspace of the function space. This condition can be met if we replace the word "prescribed data" by "observed data" because no matter how many surplus data we observe (in addition to the minimum number which would have sufficed for a unique solution), these data cannot be inconsistent since the prescribed mathematical law was in operation throughout our observations. Our operator is restricted to certain dimensions of the function space and cannot lead out of this space, no matter how many observations we perform.

If the given operator omits certain dimensions in the $Y$ relation, this makes our solution incomplete since we obtain no information concerning the missing dimensions. We do get, however, a unique solution in those dimensions which are represented in the operator. We can then add later observations in order to complete our solution with respect to the missing dimensions.

We thus obtain a method for the solution of boundary value problems which can be arbitrarily over-determined (although consistent) or under-determined, and thus far from that "well-posed" type of problems that we expect under the customary conditions. Under these relaxed conditions the "inverse" of the operator does not exist any more in the ordinary sense. But even the "generalized inverse" in the sense of the matrix $G$ which omits the zero-field and avoids the division by zero, need not necessarily exist. In the case of *finite* matrices it cannot happen that the matrix $G$, defined by (53) and more specifically by (59), should not exist. But in the case of continuous operators the corresponding expansion—called under simplified conditions the "bilinear expansion of the Green's function" (*cf.* [**9**], p. 360)—becomes an *infinite series* which may or may not converge. In many problems of an unconventional type to which the present theory is applicable, the bilinear expansion becomes in fact meaningless since it has no tendency to converge. Nor does the "inverse bilinear expansion" which corresponds to (60), converge and represent the given operator. This does *not*, however, interfere with the solution of our problem in terms of the two "co-orthogonal expansions" (61) and (62) which remain uniformly convergent even if the expansion of the inverse operator fails, provided that the right side is taken from that restricted subspace of the function space which is allotted to it by the nature of the given operator.

Although the application of the general theory to the field of continuous operators will be discussed in more detail in a separate paper*, it may not be without interest to give an example of the type of unconventional boundary value problems which become solvable by the method here presented.

* Proceedings of the International Congress of Mathematicians, Edinburgh, 1958 (to be published in 1960).

*Problem.* Given a simply-connected domain $C$ of the complex plane, enclosed by a smooth boundary, let it be known that the function $f(z)$ of the complex variable $z=x+iy$ is analytical throughout $C$ including the boundary, and let the value of $f(z)$ be given along the arbitrarily small arc $S$ of the boundary. Find $f(z)$ inside the domain $C$.

The method of analytical continuation shows that this problem has a unique solution but the theory of analytical functions gives no clue toward a solution which would obtain the value of $f(z)$ at a distant point $z$ directly in terms of the given boundary values along $S$.

The shifted eigenvalue problem associated with the present problem gives the solution in the following form. Associated with the domain $C$ and the arc $S$ we can define an infinite set of functions

$$f_1(z, z^*), \cdots, f_i(z, z^*), \cdots \tag{65}$$

which exist inside and on the boundary $S$, $S'$, together with a corresponding set of functions defined along the arc $S$:

$$g_1(s), \cdots, g_i(s), \cdots. \tag{66}$$

The desired function $f(z)$ can be expanded into the infinite sum

$$f(z) = \sum_{\alpha=1}^{\infty} \gamma_\alpha f_\alpha(z, z^*) \tag{67}$$

where the expansion coefficients $\gamma_i$ are obtained as follows:

$$\gamma_i = \int_S f(s) g_i(s) ds. \tag{68}$$

None of the functions $f_\alpha(z, z^*)$ are analytical (in view of the dependence on $z^*$, the "complex conjugate" of $z$). Nor do these $f_\alpha(z, z^*)$ satisfy the given boundary conditions. In fact, all the $f_\alpha(z, z^*)$ *vanish* along $S$. And yet, the infinite sum (67) *converges uniformly* to the correct value of $f(z)$ at every point of the domain $C$ (including the outer boundary $S'$) which excludes the arc $S$.

## References


1. H. W. Turnbull and A. C. Aitken, Introduction to the Theory of Canonical Matrices, London, 1932.
2. R. A. Frazer, W. J. Duncan, and A. R. Collar, Elementary Matrices, Cambridge, 1950.
3. C. C. MacDuffee, The Theory of Matrices, New York, 1946.
4. Encyklop. der Math. Wissenschaften, vol. I, Leipzig, 1904.
5. C. Lanczos, Applied Analysis, Englewood Cliffs, N. J., 1956.
6. E. H. Moore, General Analysis, vol. 1, Amer. Philos. Soc., Philadelphia, 1935.
7. R. Penrose, A generalized inverse for matrices, Proc. Cambridge Philos. Soc., vol. 51, 1955, p. 406.
8. R. Rado, Note on generalized inverses of matrices, Proc. Cambridge Philos. Soc., vol. 52, 1956, p. 600.
9. R. Courant and D. Hilbert, Methods of Mathematical Physics, vol. 1, New York, 1953.