J. Soc. Indust. Appl. Math.
Vol. 6, No. 1, March, 1958
Printed in U.S.A.

INVERSION OF MATRICES BY BIORTHOGONALIZATION AND RELATED RESULTS*[1, 2]

MAGNUS R. HESTENES

1. Introduction. The purpose of the present paper is to describe a method for inverting matrices together with the associated theory pertinent to the problem. The method is essentially a generalized elimination procedure. It is based on the fact that if $u_1, \cdots, u_n$ are the column vectors of a square matrix U, and $v_1, \cdots, v_n$ are the row vectors of a second square matrix V, then these two sets of vectors are biorthogonal if and only if $V = U^{-1}$. The problem of inversion of matrices is thereby reduced to that of constructing biorthogonal systems of vectors. This fact suggests at once a number of methods of inverting a matrix. One of these is described in detail below. It has been tested on computers and has been found to be very effective. The results here given are an outgrowth of the method of solving a system of linear equations proposed by T. Motzkin. A first report on this method was given by the author in 1955.[3]

An algorithm for constructing a biorthogonal system of vectors is found in Section 4, and its theoretical aspects are described in Section 6. The main ideas are illustrated in the examples given in Section 5. In Sections 7, 8 and 9 a discussion is given of various inversion routines and their properties. The connection with the elimination method is brought out here. After discussing the concept of principal vectors, heuristic error estimates are given in Section 11. The numerical experiments carried out by the author indicate that these estimates are easily obtained and are reliable. As a matter of fact the inversion routine described in Section 7 can be coded so that this error estimate can be computed during the routine at little or no expense.

The remainder of the paper is devoted to a description of methods of finding principal values and eigenvalues and the associated vectors. In the last section is found a method of decomposing a matrix into a linear combination of isometries.

---

* Received by the editors December 4, 1957.

[1] The preparation of this paper was sponsored by the Office of Ordnance Research, U. S. Army. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Presented at the Wayne State University Conference on Matrix Computations, Sept. 3–6, 1957.

[3] M. R. Hestenes, Iterative computational methods, *Communications on Pure and Applied Mathematics*, vol. VIII (1955), pp. 91–92.

A discussion of the related theory can be found in the excellent paper by G. E. Forsythe.†

2. Terminology and notation. Throughout the following pages it will be assumed that the reader is familiar with the elementary theory of matrices. Matrices will be denoted by capital letters. Thus, A, B, M, N, P, Q, etc. denote matrices. The elements of matrices are scalars. Unless otherwise expressly stated scalars will be taken to be real numbers. They will be denoted by the first letters of the alphabet, such as a, b, c, $\cdots$. Vectors will be denoted by the last letters of the alphabet, such as p, q, u, v, w, x, y, z. Some exceptions will be made to these conventions. For example in the notation $x = (x^1, \cdots, x^n)$ for a vector, the components $x^1, \cdots, x^n$ are scalars.

Given two vectors $x = (x^1, \cdots, x^n)$ and $(y^1, \cdots, y^n)$ we write

$$x + y = (x^1 + y^1, \cdots, x^n + y^n), \qquad ax = (ax^1, \cdots, ax^n).$$

The sum

$$(x, y) = x^1y^1 + \cdots + x^ny^n$$

is called the *inner product* of x and y. The *length* of x will be denoted by

$$\| x \| = (x, x)^{\frac{1}{2}}.$$

We have the relations

$$\| (x, y) \| \leqq \| x \| \, \| y \|, \qquad \| x + y \| \leqq \| x \| + \| y \|.$$

If A is an $n \times n$ dimensional matrix, then

$$(Ax, y) = (x, A^*y).$$

Here and elsewhere A^* denotes the transpose of A. If $A^* = A$, then A is said to be *symmetric*. If $(Ax, x) \geqq 0$, then A is *nonnegative*, if

$$(Ax, x) > 0$$

whenever $x \neq 0$, then A is *positive*. The terms "nonpositive" and "negative" are defined similarly. If $(x, y) = 0$, then x and y are said to be *orthogonal.*

If complex numbers are taken to be scalars, the inner product of two vectors x and y is given by the formula

$$(x, y) = x^1\bar{y}^1 + \cdots + x^n\bar{y}^n.$$

Here $\bar{a}$ denotes the complex conjugate of a. Moreover the symbol A^* de-

† G. E. Forsythe, *Solving linear equations can be interesting*, Bull. Amer. Math. Soc., 59 (1953), pp. 299–329.

notes the conjugate transpose of a matrix A. With these conventions in mind, all formulas will be written so that they hold both for the real and for the complex case unless otherwise expressly stated. However, in our discussions we shall restrict ourselves normally to the case when the scalars are real. The extensions to the complex case are easily made.

We shall have occasion to refer to the concept of the general reciprocal of a matrix. Its definition is given in Section 14 below. It is not needed for the understanding of the major portion of the paper.

3. Biorthogonal systems of vectors. As we shall see presently, there is a close connection between the problem of finding the inverse of a matrix and the problem of finding a biorthogonal system of vectors. Consider now vectors in an m-dimensional space. Two sets of n vectors

$$\begin{array}{c} u_1, \cdots, u_n \\ v_1, \cdots, v_n \end{array} \tag{3.1}$$

form a *biorthogonal system* if each vector of one set is orthogonal to all but one vector of the other set with order and normalization such that

$$(v_i, u_j) = \delta_{ij} \qquad (i, j = 1, \cdots, n; \delta_{ii} = 1; \delta_{ij} = 0, i \neq j). \tag{3.2}$$

For such a system $n \leqq m$ and both sets must be linearly independent.

The concept of biorthogonality just described can be expressed in matrix form as follows. The vectors $u_1, \cdots, u_n$ can be considered to be the column vectors of a matrix U and $v_1, \cdots, v_n$ the row vectors of a matrix V. The formula (3.2) is then equivalent to the equation

$$VU = I$$

where I is the n-dimensional identity matrix. If $n = m$ then V is the inverse of U. Hence, we see that a *square matrix V is the inverse of a square matrix U if and only if the row vectors of V are biorthogonal to the column vectors of U*. Thus any procedure for finding a set of vectors $v_1, \cdots, v_n$ biorthogonal to a set $u_1, \cdots, u_n$ is at the same time a method for inverting matrices. In the complex case, the vectors $u_1, \cdots, u_n$ are taken to be the conjugates of the column vectors of U. It is clear that the roles of row and column vectors may be interchanged. In case U is not a square matrix and V is of the form $V = BU^*$, then V is the general reciprocal of U in the sense of E. H. Moore.

In view of these remarks the problem with which we shall be concerned is the following:

PROBLEM I. *Given a set of vectors $u_1, \cdots, u_n$ to find vectors $v_1, \cdots, v_n$ so that the two sets form a biorthogonal system.*

Actually we shall modify this problem in a trivial way as follows:

PROBLEM II. *Given two sets of vectors*

$$\begin{matrix} u_1, \cdots, u_n \\ v_1^{(0)}, \cdots, v_n^{(0)} \end{matrix} \tag{3.3}$$

to find vectors $v_1, \cdots, v_n$ *of the form*

$$v_i = b_{i1}v_1^{(0)} + b_{i2}v_2^{(0)} + \cdots + b_{in}v_n^{(0)} \qquad (i = 1, \cdots, n) \tag{3.4}$$

so that the sets (3.1) *form a biorthogonal system.*

It is essential to our problem and its applications that the vectors $u_1, \cdots, u_n$ be unaltered.

Problem II can be rephrased in terms of matrices. Let

$$a_{ij} = (v_i^{(0)}, u_j) \qquad (i, j = 1, \cdots, n).$$

Then the matrix

$$A = (a_{ij})$$

is of the form

$$A = V^{(0)}U$$

where $u_1, \cdots, u_n$ are the column vectors of U and $v_1^{(0)} \cdots v_n^{(0)}$ are the row vectors of $V^{(0)}$. We seek a matrix

$$B = (b_{ij})$$

such that

$$BA = BV^{(0)}U = I.$$

The row vectors $v_1, \cdots, v_n$ of $V = BV^{(0)}$ are of the form (3.4) and yield a solution of Problem II. Since B is the inverse of A it follows that Problem II has a solution if and only if the matrix A has a non-zero determinant.

Of interest later in the text is the case in which U and $V^{(0)}$ are of the form

$$U = \begin{pmatrix} I \\ 0 \end{pmatrix} \qquad V^{(0)} = (A \ I).$$

Then $V^{(0)}U = A$ and the inverse B is exhibited in V as follows

$$V = BV^{(0)} = (I \ B).$$

This suggests the connection of our problem with that of inverting a matrix A by elimination. This connection will be discussed in Section 9 below.

4. Construction of biorthogonal systems. The purpose of this section is to recall one of the standard methods for constructing biorthogonal systems. This method will be used to devise routines for matrix inversion. Suppose now that we have given two sets

$$u_1, \cdots, u_n \qquad (4.1)$$
$$v_1, \cdots, v_n$$

of n vectors in an m-dimensional space ($m \geqq n$). The problem is to obtain a biorthogonal system by modifying the v's.

To this end let $v_1^{(0)}, \cdots, v_n^{(0)}$ be the initial choice of the vectors $v_1, \cdots, v_n$. We shall modify these vectors successively in n steps. Each step is similar to the preceding one and will be called a *cycle*. After n cycles have been completed the vectors $v_1^{(n)}, \cdots, v_n^{(n)}$ obtained will be a solution to our problem. In the k-th cycle the vectors $v_1^{(k-1)}, \cdots, v_n^{(k-1)}$ are transformed into a new set $v_1^{(k)}, \cdots, v_n^{(k)}$ by the following computations:

(α) Construct $v_k^{(k)}$ so that

$$(v_k^{(k)}, u_k) = 1 \qquad (4.2)$$

by using the formulas

$$c_{kk} = (v_k^{(k-1)}, u_k), \qquad c_k = 1/c_{kk}, \qquad v_k^{(k)} = c_k v_k^{(k-1)}. \qquad (4.3\alpha)$$

(β) Construct $v_j^{(k)} (j \neq k)$ so that

$$(v_j^{(k)}, u_k) = 0 \qquad (j \neq k; j = 1, \cdots, n) \qquad (4.4)$$

by using the formulas

$$c_{jk} = (v_j^{(k-1)}, u_k), \qquad v_j^{(k)} = v_j^{(k-1)} - c_{jk} v_k^{(k)}. \qquad (4.3\beta)$$

This procedure will be illustrated in the next section. A discussion of its properties will be given in §6. Before closing this section it is of interest to write our results in matrix form. Let $V^{(k)}$ be the matrix whose row vectors are $v_1^{(k)}, \cdots, v_n^{(k)}$ and set

$$A^{(k)} = V^{(k)} U. \qquad (4.5)$$

During the kth iteration the matrices $V^{(k-1)}, A^{(k-1)}$ are transformed into $V^{(k)}, A^{(k)}$ by a transformation

$$V^{(k)} = C^{(k)} V^{(k-1)}, \qquad A^{(k)} = C^{(k)} A^{(k-1)} \qquad (4.6)$$

where

$$C^{(k)} = (c_{ij}^{(k)}) \qquad (4.7\alpha)$$

is of the form

$$(4.7\beta) \qquad c_{ij}^{(k)} = \delta_{ij} \quad (j \neq k), \qquad c_{kk}^{(k)} = c_k\,, \qquad c_{ik}^{(k)} = -c_{ik}c_k \quad (i \neq k).$$

Here $\delta_{ii} = 1$, $\delta_{ij} = 0$ $(i \neq j)$. For example if $n = 4$ and $k = 3$, we have

$$C^{(3)} = \begin{pmatrix} 1 & 0 & -c_{13}c_3 & 0 \\ 0 & 1 & -c_{23}c_3 & 0 \\ 0 & 0 & c_3 & 0 \\ 0 & 0 & -c_{43}c_3 & 1 \end{pmatrix}.$$

The first relation (4.6) follows from the algorithm (4.3). The second relation (4.6) follows from the relations

$$A^{(k)} = V^{(k)}U = C^{(k)}V^{(k-1)}U = C^{(k)}A^{(k-1)}.$$

In carrying out the algorithm (4.3), the elements of the matrices $A^{(k)}$ and $C^{(k)}$ normally are not computed explicitly. An exception can be found in Illustration 2 given in the next section. These matrices will be discussed in detail in Section 6. Here it will be seen that $A^{(n)}$ is the identity matrix. The row vectors $v_1, \cdots, v_n$ of $V^{(n)}$ are accordingly the solution to our problem.

5. Illustrations. In the present section we shall illustrate three applications of the algorithm given in the last section. The first illustrates the inversion of a matrix by biorthogonalization. The second illustrates one of the many versions of inversion of a matrix by Gaussian eliminations. The third is concerned with the computation of the general reciprocal of a matrix in the sense of E. H. Moore.

Illustration 1. *Inversion by biorthogonalization.* It is desired to compute the inverse of the matrix

$$M = \begin{pmatrix} 0 & 2 & 1 \\ 1 & 0 & 2 \\ 2 & 1 & 0 \end{pmatrix}.$$

For this purpose choose $U = M$, $V^{(0)} = U^*$. Then

$$u_1 = v_1^{(0)} = (0, 1, 2)$$

$$u_2 = v_2^{(0)} = (2, 0, 1)$$

$$u_3 = v_3^{(0)} = (1, 2, 0).$$

Moreover

$$A^{(0)} = V^{(0)}U = \begin{pmatrix} 5 & 2 & 2 \\ 2 & 5 & 2 \\ 2 & 2 & 5 \end{pmatrix}.$$

The computations in each of the cycles are listed below:

(1) *First cycle.*

(α)
$$c_{11} = (v_1^{(0)}, u_1) = 5, \qquad c_1 = 1/5$$
$$v_1^{(1)} = c_1 v_1^{(0)} = (0, 1/5, 2/5)$$

(β)
$$c_{21} = (v_2^{(0)}, u_1) = 2$$
$$v_2^{(1)} = v_2^{(0)} - c_{21} v_1^{(0)} = (2, -2/5, 1/5)$$
$$c_{31} = (v_3^{(0)}, u_1) = 2$$
$$v_3^{(1)} = v_3^{(0)} - c_{31} v_1^{(0)} = (1, 8/5, -4/5).$$

Hence

$$V^{(1)} = \begin{pmatrix} 0 & \frac{1}{5} & \frac{2}{5} \\ 2 & -\frac{2}{5} & \frac{1}{5} \\ 1 & \frac{8}{5} & -\frac{4}{5} \end{pmatrix},$$

$$A^{(1)} = V^{(1)} U = \begin{pmatrix} 1 & \frac{2}{5} & \frac{2}{5} \\ 0 & \frac{21}{5} & \frac{6}{5} \\ 0 & \frac{6}{5} & \frac{21}{5} \end{pmatrix}.$$

(2) *Second cycle.*

(α)
$$c_{22} = (v_2^{(1)}, u_2) = 21/5, \qquad c_2 = 5/21$$
$$v_2^{(2)} = c_2 v_2^{(1)} = (10/21, -2/21, 1/21)$$

(β)
$$c_{32} = (v_3^{(1)}, u_2) = 6/5$$
$$v_3^{(2)} = v_3^{(1)} - c_{32} v_2^{(2)} = (3/7, 12/7, -6/7)$$
$$c_{12} = (v_1^{(1)}, u_2) = 2/5$$
$$v_1^{(2)} = v_1^{(1)} - c_{12} v_2^{(2)} = (-4/21, 5/21, 8/21).$$

Hence

$$V^{(2)} = \begin{pmatrix} -\frac{4}{21} & \frac{5}{21} & \frac{8}{21} \\ \frac{10}{21} & -\frac{2}{21} & \frac{1}{21} \\ \frac{3}{7} & \frac{12}{7} & -\frac{6}{7} \end{pmatrix},$$

$$A^{(2)} = V^{(2)} U = \begin{pmatrix} 1 & 0 & \frac{2}{7} \\ 0 & 1 & \frac{2}{7} \\ 0 & 0 & \frac{27}{7} \end{pmatrix}.$$

(3) *Third cycle.*

(α)
$$c_{33} = (v_3^{(2)}, u_3) = 27/7, \qquad c_3 = 7/27$$
$$v_3^{(3)} = c_3 v_3^{(2)} = (1/9, 4/9, -2/9)$$

$$(\beta) \qquad \begin{aligned} c_{13} &= (v_1^{(2)}, u_3) = 2/7 \\ v_1^{(3)} &= v_1^{(2)} - c_{13}v_3^{(3)} = (-2/9, 1/9, 4/9) \\ c_{23} &= (v_2^{(2)}, u_3) = 2/7 \\ v_2^{(3)} &= v_2^{(2)} - c_{23}v_3^{(3)} = (4/9, -2/9, 1/9). \end{aligned}$$

$$V^{(3)} = \begin{pmatrix} -\frac{2}{9} & \frac{1}{9} & \frac{4}{9} \\ \frac{4}{9} & -\frac{2}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{4}{9} & -\frac{2}{9} \end{pmatrix}, \qquad A^{(3)} = V^{(3)}U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The matrix $V^{(3)}$ is the inverse of $U = M$. It is of interest to observe that the determinants

$$d_1 = a_{11}^{(0)} = 5, \qquad d_2 = \begin{vmatrix} a_{11}^{(0)} & a_{12}^{(0)} \\ a_{21}^{(0)} & a_{22}^{(0)} \end{vmatrix} = 21, \qquad d_3 = \det A_0 = 81$$

of the indicated principal minors of $A_0 = (a_{ij}^{(0)})$ are related to the numbers c_1, c_2, c_3 in the above algorithm by the relations

$$c_1 = 1/d_1 = 1/5, \qquad c_2 = d_1/d_2 = 5/21, \qquad c_3 = d_2/d_3 = 7/27$$
$$c_1c_2 = 1/d_2 = 1/21, \qquad c_1c_2c_3 = 1/d_3 = 1/81.$$

As we shall see in the next section the nonvanishing of these determinants is a necessary and sufficient condition that the algorithm be applicable, that is, it can be carried out to its completion. The choice $V_0 = U^*$ was made to insure the nonvanishing of these determinants. If we had selected $V^{(0)} = I$, then $A^{(0)} = U$ and $d_1 = 0$, $d_2 = -2$, $d_2 = 9$. In this event the algorithm would fail since $c_{11} = d_1 = 0$ and further modifications would have to be made in order to apply the algorithm.

Illustration 2. It is desired to compute the inverse of

$$A = \begin{pmatrix} 2 & 1 & 2 \\ 4 & -3 & 1 \\ 3 & -6 & 0 \end{pmatrix}$$

by selecting

$$V^{(0)} = (A \ I), \qquad U = \begin{pmatrix} I \\ 0 \end{pmatrix}.$$

Then

$$\begin{aligned} u_1 &= (1, 0, 0, 0, 0, 0) \\ u_2 &= (0, 1, 0, 0, 0, 0) \\ u_3 &= (0, 0, 1, 0, 0, 0) \end{aligned}$$

$$\begin{aligned} v_1^{(0)} &= (2, 1, 2, 1, 0, 0) \\ v_2^{(0)} &= (4, -3, 1, 0, 1, 0) \\ v_3^{(0)} &= (3, -6, 0, 0, 0, 1). \end{aligned}$$

In this event $A^{(0)} = V^{(0)}U = A$ and the determinants d_1, d_2, d_3 described above are given by

$$d_1 = 2, \qquad d_2 = -10, \qquad d_3 = -15.$$

Since these numbers are different from zero the algorithm (4.3) can be carried out. The formation of the inner products of the form (v, u_i) is then equivalent to selecting the i-th component of v. The results after each cycle are stated in matrix forms follows.

(1) *First cycle.* A first application of the algorithm (4.3) yields the matrix

$$V^{(1)} = \begin{pmatrix} 1 & \frac{1}{2} & 1 & \frac{1}{2} & 0 & 0 \\ 0 & -5 & -3 & -2 & 1 & 0 \\ 0 & -\frac{15}{2} & -3 & -\frac{3}{2} & 0 & 1 \end{pmatrix}$$

whose row vectors are $v_1^{(1)}$, $v_2^{(1)}$, $v_3^{(1)}$. The matrix $A^{(1)} = V^{(1)}U$ is given by the first three columns of $V^{(1)}$.

(2) *Second cycle.* A second application of (4.3) yields the matrix

$$V^{(2)} = \begin{pmatrix} 1 & 0 & \frac{7}{10} & \frac{3}{10} & \frac{1}{10} & 0 \\ 0 & 1 & \frac{3}{5} & \frac{2}{5} & -\frac{1}{5} & 0 \\ 0 & 0 & \frac{3}{2} & \frac{3}{2} & -\frac{3}{2} & 1 \end{pmatrix}$$

whose row vectors are $v_1^{(2)}$, $v_2^{(2)}$, $v_3^{(2)}$. The matrix $A^{(2)} = V^{(2)}U$ is given by the first three columns of $V^{(2)}$.

(3) *Third cycle.* The final application of (4.3) yields the matrix

$$V^{(3)} = \begin{pmatrix} 1 & 0 & 0 & -\frac{2}{5} & \frac{4}{5} & -\frac{7}{15} \\ 0 & 1 & 0 & -\frac{1}{5} & \frac{2}{5} & -\frac{2}{5} \\ 0 & 0 & 1 & 1 & -1 & \frac{2}{3} \end{pmatrix}$$

whose row vectors are $v_1^{(3)}$, $v_2^{(3)}$, $v_3^{(3)}$. The matrix $A^{(3)} = V^{(3)}U = I$ is given by the first three columns of $V^{(3)}$. The last three columns give the inverse

$$A^{-1} = \begin{pmatrix} -\frac{2}{5} & \frac{4}{5} & -\frac{7}{15} \\ -\frac{1}{5} & \frac{2}{5} & -\frac{2}{5} \\ 1 & -1 & \frac{2}{3} \end{pmatrix}$$

of A.

In carrying out these computations it is seen that the matrix A has been inverted by one of the standard forms of Gaussian elimination. It should be observed that the constants c_{ij} appearing in the algorithm (4.3) determine a matrix

$$C = (c_{ij}) = \begin{pmatrix} 2 & \frac{1}{2} & \frac{7}{10} \\ 4 & -5 & \frac{3}{5} \\ 3 & -\frac{15}{2} & \frac{3}{2} \end{pmatrix}$$

composed of the first, second, and third columns of $V^{(0)}$, $V^{(1)}$, $V^{(2)}$ respectively.

Suppose next that we recompute the inverse of A by selecting $U = A$ and $V^{(0)} = I$. Then, as is easily seen, the matrices $V^{(1)}$, $V^{(2)}$, $V^{(3)}$ obtained after the first, second, third cycles are given by

$$V^{(1)} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ -2 & 1 & 0 \\ -\frac{3}{2} & 0 & 1 \end{pmatrix} \quad V^{(2)} = \begin{pmatrix} \frac{3}{10} & \frac{1}{10} & 0 \\ \frac{2}{5} & -\frac{1}{5} & 0 \\ \frac{3}{2} & -\frac{3}{2} & 1 \end{pmatrix} \quad V^{(3)} = \begin{pmatrix} -\frac{2}{5} & \frac{4}{5} & -\frac{7}{15} \\ -\frac{1}{5} & \frac{2}{5} & -\frac{2}{5} \\ 1 & -1 & \frac{2}{3} \end{pmatrix}.$$

These matrices appear as the last three columns of those computed earlier. Moreover, the constants c_{ij} are identical in these two cases. The two methods are accordingly equivalent.

Illustration 3. It is desired to compute the general reciprocal of the matrix

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

This matrix is of rank 2 and the method described above cannot be applied without modification. However, we adjoin two additional rows *orthogonal* to those of M so as to obtain a matrix

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

of rank 4 to which the algorithm is applicable with $V^{(0)} = U^*$. Carrying out the four cycles required by the algorithm one obtains successively the matrices

$$V^{(0)} = U^* = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & -1 & 0 & 0 & 1 \\ 1 & 0 & 1 & -1 & -1 \end{pmatrix} \quad A = V^{(0)}U = \begin{pmatrix} 3 & 1 & 1 & 1 \\ 1 & 3 & 0 & 0 \\ 1 & 0 & 3 & 0 \\ 1 & 0 & 0 & 4 \end{pmatrix}$$

$$V^{(1)} = \begin{pmatrix} \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ -\frac{1}{3} & 1 & \frac{2}{3} & -\frac{1}{3} & 1 \\ \frac{2}{3} & -1 & -\frac{1}{3} & -\frac{1}{3} & 1 \\ \frac{2}{3} & 0 & \frac{2}{3} & -\frac{4}{3} & -1 \end{pmatrix} \quad V^{(2)} = \begin{pmatrix} \frac{3}{8} & -\frac{1}{8} & \frac{2}{8} & \frac{3}{8} & -\frac{1}{8} \\ -\frac{1}{8} & \frac{3}{8} & \frac{2}{8} & -\frac{1}{8} & \frac{3}{8} \\ \frac{5}{8} & -\frac{7}{8} & -\frac{2}{8} & -\frac{3}{8} & \frac{9}{8} \\ \frac{5}{8} & \frac{1}{8} & \frac{6}{8} & -\frac{11}{8} & -\frac{7}{8} \end{pmatrix}$$

$$V^{(3)} = \begin{pmatrix} \frac{6}{21} & 0 & \frac{6}{21} & \frac{9}{21} & -\frac{6}{21} \\ -\frac{2}{21} & \frac{7}{21} & \frac{5}{21} & -\frac{3}{21} & \frac{9}{21} \\ \frac{5}{21} & -\frac{7}{21} & -\frac{2}{21} & -\frac{3}{21} & \frac{9}{21} \\ \frac{15}{21} & 0 & \frac{15}{21} & -\frac{30}{21} & -\frac{15}{21} \end{pmatrix}$$

$$V^{(4)} = \begin{pmatrix} \frac{1}{5} & 0 & \frac{1}{5} & \frac{3}{5} & -\frac{1}{5} \\ -\frac{1}{15} & \frac{1}{3} & \frac{4}{15} & -\frac{1}{5} & \frac{2}{5} \\ \frac{4}{15} & -\frac{1}{3} & -\frac{1}{15} & -\frac{1}{5} & \frac{2}{5} \\ \frac{1}{5} & 0 & \frac{1}{5} & -\frac{2}{5} & -\frac{1}{5} \end{pmatrix}.$$

The matrix $V^{(4)}$ is the general reciprocal of U. Moreover the matrix

$$N = \begin{pmatrix} \frac{1}{5} & 0 & \frac{1}{5} \\ -\frac{1}{15} & \frac{1}{3} & \frac{4}{15} \\ \frac{4}{15} & -\frac{1}{3} & -\frac{1}{15} \\ \frac{1}{5} & 0 & \frac{1}{5} \end{pmatrix}$$

obtained from $V^{(4)}$ by deleting the last two columns is the general reciprocal of M. This follows from our construction of U, as one readily verifies.

6. Properties of the biorthogonalization algorithm. We shall now return to a theoretical consideration of the algorithm (4.3). Using the notations described in §4 we have as our first result:

THEOREM 6.1. *Having completed the first* $(k - 1)$ *cycles, the kth cycle* (4.3) *can be carried out if and only if* $c_{kk} \neq 0$. *At the end of the kth cycle the matrix*

$$A^{(k)} = (a_{ij}^{(k)}) = V^{(k)}U, \tag{6.1a}$$

where

$$a_{ij}^{(k)} = (v_i^{(k)}, u_j), \tag{6.1b}$$

has the property that

$$a_{ij}^{(k)} = \delta_{ij} \qquad (j \leqq k, i = 1, \cdots, n) \tag{6.2}$$

where $\delta_{ii} = 1$, $\delta_{ij} = 0 (i \neq j)$. *The numbers* c_{jk} *appearing in the kth cycle* (4.3) *form the kth column of* $A^{(k-1)}$. *Accordingly the column vectors of the matrix*

$$C = (c_{jk}) \tag{6.3a}$$

are given by the first, second, third, $\cdots$ *column vector of* $A^{(0)}, A^{(1)}, A^{(2)}, \cdots$ *respectively, that is,*

$$c_{jk} = a_{jk}^{(k-1)}. \tag{6.3b}$$

The first statement follows from the fact that division by c_{kk} is used in (4.3). The second statement will be proved by induction. Observe first that algorithm (4.3) was designed to require the relations (6.2) to hold when $j = k$. Suppose now that the relations (6.2) hold when $k = h - 1$, where $1 < h \leqq n$. Then, by (4.3α) with $k = h$ we have

$$(v_h^{(h)}, u_j) = c_h(v_h^{(h-1)}, u_j) = 0 \qquad (j < h).$$

Similarly, by (4.3β)

$$(v_i^{(h)}, u_j) = (v_i^{(h-1)}, u_j) - c_{ih}(v_h^{(h)}, u_j) = \delta_{ij} \qquad (j < h).$$

Since (6.2) holds for $k = 1$, it holds for every integer $h \leqq n$, as was to be proved.

The last two statements in the theorem follow from the definition of the numbers c_{jk} and of the matrices $V^{(k)}$.

COROLLARY: *If the numbers* $c_{11}, \cdots, c_{nn}$ *are different from zero, then* $A^{(n)} = V^{(n)}U = I$ *and the row vectors* $v_1, \cdots, v_n$ *of* $V^{(n)}$ *are biorthogonal to the column vectors* $u_1, \cdots, u_n$ *of* U.

The nonvanishing of the numbers $c_{11}, \cdots, c_{nn}$ can be expressed in terms of the nonvanishing of certain principal minors of the matrix

$$A^{(0)} = V^{(0)}U.$$

It will be convenient for the moment to drop the superscript and set $A = A^{(0)}$, except when we are concerned with the sequence $A^{(0)}, A^{(1)}, \cdots$. Thus

$$A = (a_{ij}), \qquad a_{ij} = (v_i^{(0)}, u_j) \qquad (i, j, = 1, \cdots, n). \tag{6.4}$$

With these notations in mind we have

THEOREM 6.2. *Let* $d_0 = 1$ *and let* d_k $(k = 1, 2, \cdots, n)$ *be the determinant of the k-dimensional principal minor*

$$A_k = (a_{pq}) \qquad (p, q = 1, \cdots, k) \tag{6.5}$$

of A. *Suppose that* $d_i \neq 0$ $(i < k)$. *Then the number* c_{kk} *appearing in* (4.3) *is given by the formula*

$$c_{kk} = d_k/d_{k-1} \tag{6.6a}$$

or equivalently

$$d_k = c_{11}c_{22} \cdots c_{kk}. \tag{6.6b}$$

Hence $c_{kk} \neq 0$ *if and only if* $d_k \neq 0$. *Let* d_{kp} *be the cofactor of* a_{pk} *in* A_k. *The numbers* c_{jk} *appearing in the kth cycle* (4.3) *are given by the formulas*

$$c_{jk} = -\frac{d_{jk}}{d_{k-1}} \qquad (j < k)$$

$$(6.7) \qquad c_{kk} = \frac{d_k}{d_{k-1}}$$

$$c_{jk} = a_{jk} - \sum_{p=1}^{k-1} a_{jp} c_{pk} \qquad (j > k).$$

In order to prove this result recall the relation

$$A^{(k)} = C^{(k)} A^{(k-1)} \qquad (k > 0)$$

where $C^{(k)}$ differs from the identity only in the kth column, the elements in the kth column being

$$c_{jk}^{(k)} = -c_{jk} c_k \; (j \neq k), \qquad c_{kk}^{(k)} = c_k = 1/c_{kk}.$$

The determinant of $C^{(k)}$ is c_k. Let

$$B^{(k)} = C^{(k)} C^{(k-1)} \cdots C^{(1)}.$$

Then $B^{(k)}$ is of the form

$$B^{(k)} = \begin{pmatrix} B_k & 0 \\ B_k' & I \end{pmatrix}$$

where B_k is a square matrix of dimension k. The kth column of $B^{(k)}$ coincides with that of $C^{(k)}$, as one readily verifies. We have

$$B^{(k)} A = A^{(k)},$$

where $A = A^{(0)} = (a_{ij})$. If we write A in the form

$$A = \begin{pmatrix} A_k & A_k'' \\ A_k' & A_k''' \end{pmatrix}$$

we have, by (6.2),

$$B_k A_k = I, \qquad B_k' A_k + A_k' = 0.$$

Hence

$$B_k = A_k^{-1}, \qquad B_k' = -A_k' B_k.$$

The jth $(j < k)$ and kth elements of the last column of A_k^{-1} are

$$d_{jk}/d_k \; (j < k) \qquad d_{k-1}/d_k$$

where d_{jk} is the cofactor of a_{kj} in A_k. The corresponding elements of B_k are

$$-c_{jk}c_k, \qquad c_k,$$

by construction. Hence

$$-c_{jk}c_k = d_{jk}/d_k \; (j < k), \qquad c_k = d_{k-1}/d_k.$$

Since $c_k = 1/c_{kk}$ the first two relations (6.7) follow. In order to prove the last equation recall that the last columns of B_k' are given by

$$-c_{jk}c_k \qquad\qquad (j = k + 1, \cdots, n).$$

Computing the last column of B_k' from the product $B_k' = -A_k'B_k$ it is seen that the remaining equations in (6.7) hold. In this proof we have tacitly assumed that $d_k \neq 0$. By continuity considerations these relations hold even if $d_k = 0$, provided that $d_h \neq 0$ $(h < k)$. This proves Theorem 6.2.

THEOREM 6.3. *A necessary and sufficient condition that the algorithm* (4.3) *can be used to compute a set of vectors*

$$v_1 = v_1^{(n)}, \cdots, v_n = v_n^{(n)}$$

biorthogonal to $u_1, \cdots, u_n$ *is that the numbers* $d_1, \cdots, d_n$, *described in Theorem* 6.2, *be different from zero.*

This result follows from Theorem 6.2 and the corollary to Theorem 6.1.

THEOREM 6.4. *The numbers* c_{jk} *in the algorithm* (4.3) *are determined by the elements of the matrix*

$$A = V^{(0)}U.$$

This result is a consequence of the relations (6.7). It follows that two distinct pairs of matrices $V^{(0)}$ and U having the same product matrix A generate the same number c_{jk}. This situation arose in Illustration 2 in the last section.

COROLLARY. *If* W *is an* $n \times m$ *dimensional matrix such that*

$$WU = 0, \tag{6.8}$$

then the numbers c_{jk} *appearing in the algorithm* (4.3) *are unchanged by replacing* $V^{(0)}$ *by* $V^{(0)} - W$.

THEOREM 6.5. *Suppose that the matrix* U *has rank* n *and that* $V^{(0)} = U^*$, *that is, suppose that the vectors* $u_1, \cdots, u_n$ *are linearly independent and that*

$$v_1^{(0)} = u_1, \qquad v_2^{(0)} = u_2, \cdots, v_n^{(1)} = u_n,$$

then the determinants $d_1, \cdots, d_n$ *of the principal minors* $A_1, \cdots, A_n$ *of* A *described in Theorem* 6.2 *are positive and the algorithm* (4.3) *can be applied.*

In this event

$$a_{ij} = (u_i, u_j) \qquad\qquad (i, j = 1, \cdots, n).$$

Setting

$$u = u_i \pi_i \qquad (i \text{ summed})$$

we have

$$\| u \|^2 = a_{ij} \pi_i \pi_j > 0 \tag{6.9}$$

unless

$$u = u_i \pi_i = 0.$$

Since $u_1, \cdots, u_n$ are linearly independent, the last equation holds only in case $\pi_i = 0$ $(i = 1, \cdots, n)$. The quadratic form (6.9) is accordingly positive definite. The determinants $d_1, \cdots, d_n$ are therefore positive. Theorem 6.5 now follows from Theorem 6.3.

THEOREM 6.6. *Let c_{jk} be the numbers appearing in the algorithm* (4.3). *The initial vectors $v_1^{(0)}, \cdots, v_n^{(0)}$ are biorthogonal to $u_1, \cdots, u_n$ if and only if the matrix*

$$C = (c_{jk}) \qquad (j, k = 1, \cdots, n)$$

is the identity matrix.

In view of this result the deviation of C from the identity can be taken as a measure of the deviation of the system

$$u_1, \cdots, u_n$$
$$v_1^{(0)}, \cdots, v_n^{(0)}$$

from being a biorthogonal system. This measure will be useful as a check in computations.

7. Inversion by biorthogonalization. The algorithm (4.3) suggests the following method of inverting a square matrix U of rank n. Let $V^{(0)}$ be an initial estimate of the inverse of U. Select $u_1, \cdots, u_n$ to be the column vectors of U (of the conjugate of U in the complex case) and let $v_1^{(0)}, \cdots, v_n^{(0)}$ be the row vectors of $V^{(0)}$. Then the algorithm (4.3), if completed, will produce vectors $v_1, \cdots, v_n$ that are the row vectors of the inverse $V = U^{-1}$ of U. It remains to select $V^{(0)}$ so that the algorithm can be applied. There are two choices that will insure the nonvanishing of the numbers $d_1, \cdots, d_n$. The first choice is $V^{(0)} = U^*$ in which case $v_i^{(0)} = u_i$. The second is any reasonable approximation of the inverse of U. The first choice will be used initially, the second will be used to improve the approximate solution obtained. Such an improvement may be necessary because of rounding off errors.

The inversion routine determined by the algorithm (4.3) with an arbitrary initial choice of $V^{(0)}$ will be called an *inversion by biorthogonalization*. Its properties when $V^{(0)} = U^*$ can be described as follows:

(1) Apart from rounding off errors the computations are independent of the coordinate system used.

(2) If the matrix U is invertible, the algorithm (4.3) can be applied. This is because the numbers $d_1, \cdots, d_n$ described in Theorems 6.2 and 6.3 are positive.

(3) If the vectors $u_1, \cdots, u_n$ are orthonormal, that is, if U^* is the inverse of U, then no corrections are made. The algorithm simply verifies that U^* is the inverse of U.

(4) If the matrix is well conditioned, that is, if the vectors $u_1, \cdots, u_n$ are almost mutually orthogonal and of about the same length, then only small corrections are made.

(5) If U is ill-conditioned, then the initial choice $V^{(0)} = U^*$ is a poor estimate of the inverse of U. In this event there will be considerable rounding off error, so that the resultant matrix $V = V^{(n)}$ may not be an adequate estimate of U^{-1}. However, the algorithm can be repeated with $V^{(0)} = V$ so as to yield an improved version of the inverse.

(6) From the remark just made the algorithm can be looked upon as an iterative process, to correct rounding off errors. In theory the algorithm should yield the inverse of U after one application. In practice, an approximate inverse V_1 is obtained. A repetition of the algorithm with $V^{(0)} = V_1$ will yield an improved estimate V_2. Selecting $V^{(0)} = V_2$ we obtain a third estimate V_3, and so on. Thus a sequence of estimates $V_0 = U^*, V_1, V_2, \cdots,$ of U^{-1} is obtained. Recall that each step of the algorithm can be looked upon as asking whether or not a matrix $V^{(0)}$ has a certain property of the inverse and correcting the matrix $V^{(0)}$ if it fails to have this property. The particular question asked is, of course, whether or not a particular element c_{ij} of the matrix C described in Theorem 6.6 is the element δ_{ij} of the identity matrix I. Let r be the largest of the numerical values of the differences $c_{ij} - \delta_{ij}$ $(i, j = 1, \cdots, n)$. Then r can be taken as a measure of the accuracy with which $V^{(0)}$ approximates U^{-1}. If this measure is computed during each iteration one obtains a measure r_i of $V^{(0)} = V_i$ as an estimate of U^{-1} together with an improved estimate V_{i+1} of U^{-1}. The number r_i can be used to determine when the iteration should terminate. In actual practice the maximum r of the absolute values of the off diagonal elements of (c_{ij}) is more convenient to compute and is adequate for the purpose of estimating the accuracy of $V^{(0)}$ as an estimate of U^{-1}.

(7) If the choice $V^{(0)} = U^*$ is made, then the product $c_1 \cdots c_n$ of the numbers c_i appearing in (4.3) is the reciprocal of the square of the determinant of U.

(8) The number of arithmetic operations used in the algorithm is given as follows:

$$n \text{ divisions}$$

$$2n^3 \text{ multiplications}$$

$$2n^2(n - 1) \text{ additions.}$$

To this number must be added additional arithmetic operations which enter into a code for carrying out these computations. The code can be written so as to significantly reduce the number of operations when a large percentage of the elements of M are zero.

The author has carried out experiments on the SWAC using the procedure described in (6) above. In computing V_{i+1} from V_i the scale ρ_i of r_i was kept in place of r_i since one is interested only in the order of magnitude of r_i. The method was very effective. For well-conditioned matrices the estimate V_2 was not significantly better than V_1. For ill-conditioned matrices there was no significant improvement of V_3, V_4, $\cdots$ over V_2 as an estimate of the inverse. The code can be written so that one can test V_i without obtaining a new estimate v_{i+1} of U^{-1} by carrying out the first steps in (4.3α) and (4.3β) and bypassing the remaining steps. It was found that the accuracy to which the inverse could be computed agreed favorably with the accuracy that was predicted by the condition number described in §11 below. The following techniques were used:

(1) Vectors were represented as "floating vectors," that is, an n-dimensional vector x was represented by $n + 1$ scalars x_0, x_1, $\cdots$, x_n. Here x_0 is a scale and the ith component of x is $2^{x_0}x_i$. The integer x_0 was chosen so that

$$\tfrac{1}{2} \leqq \max_{i=1,\cdots,n} |x_i| < 1.$$

(2) Each number c_{jk} appearing in the computation was represented in the form $2^{a_0}a_1$, where $\frac{1}{2} \leqq |a_1| < 1$ and written in the form (a_0, a_1).

(3) Inner products were computed with double precision, the final result being rounded off in the form (a_0, a_1) described in (2) above. When floating vectors are used, this can be carried out at only a small expense timewise over single precision.

8. Inversion by biorthogonalization, alternate procedure. In the last section it was pointed out that if U has rank n, the routine for biorthogonalization can be carried out if we select $V^{(0)} = U^*$ or if $V^{(0)}$ is a reasonable estimate of U^{-1}. It is a simple matter to construct examples for which the algorithm will fail if we select $V^{(0)} = I$ or some other matrix independent of the choice of U. However the procedure can be modified in such a way that it can be applied whenever the product $V^{(0)}U$ is non-singular. The modification consists of modifying the step (α) in each cycle. Using the notations described in Section 4 a modified step (α') of (α) in the kth cycle, can be stated as follows:

(α') Compute the inner products

$$(v_k^{(k-1)}, u_h) \qquad (h = k, k+1, \cdots, n) \tag{8.1}$$

and interchange u_k with a vector u_j ($j \geqq k$) for which this inner product has a maximum absolute value (or is greater in absolute value than some suitably selected threshold value). After this interchange has been made carry out step (α) as described in Section 4.

The modified algorithm in which steps (α'), (β) are used in place of (α), (β) can be successfully applied whenever the product

$$A^{(0)} = V^{(0)} U$$

has rank n. For in this event there is a rearrangement of the column vectors $u_1, \cdots, u_n$ of U so that the numbers $d_1, \cdots, d_n$, described in Theorem 6.2, are different from zero. The modified algorithm is a method of carrying out such a rearrangement. If this rearrangement had been made beforehand, no rearrangement would be made in step (α') and the computations would proceed as in the original algorithm.

It is of interest to observe that if $A^{(0)}$ is singular, then for some integer k the numbers (8.1) will all be zero. By step (β) in the ($k-1$)st cycle, we have

$$(v_k^{(k-1)}, u_j) = 0 \tag{8.2}$$

when $j < k$. It follows that (8.2) holds for $j = 1, \cdots, n$. Thus, the vector $x = v_k^{(k-1)}$ is a solution of the equation

$$U^*x = 0. \tag{8.3}$$

If $V^{(0)}$ has rank n, as we shall suppose, then $x \neq 0$. Thus, the modified algorithm yields a method of solving the homogeneous equation (8.3) when U is a singular matrix.

There are a number of other ways in which the step (α) can be modified. For example, in the k-cycle one could replace step (α) by

(α'') Select $v_i^{(k-1)}$ ($i \geqq k$) such that the absolute value of

$$(v_i^{(k-1)}, u_k)$$

is a maximum. Interchange $v_i^{(k-1)}$ and $v_k^{(k-1)}$ and carry out step (α) as originally stated.

A third modification is to replace step (α) in the kth cycle by

(α''') Select $v_i^{(k-1)}$($i \geqq k$) and u_j($j \geqq k$) such that the absolute value of

$$(v_i^{(k-1)}, u_j)$$

is a maximum. Interchange $v_i^{(k-1)}$ and $v_k^{(k-1)}$ and interchange u_j and u_k. Then carry out step (α) as originally stated.

This last modification requires considerable computation and in many cases would be impractical.

9. Connections with the Gauss elimination method. In Illustration 2 discussed in Section 5 it was seen that in the case considered the algorithm (4.3) was equivalent to one form of the Gauss elimination method. This relationship in the general case will be discussed more fully in the present section.

As a first step recall that having given two matrices U and $V^{(0)}$ the algorithm (4.3) generates a sequence of n matrices $V^{(1)}, \cdots, V^{(n)}$ with the property that

$$V^{(n)}U = I. \tag{9.1}$$

As was seen in Section 4, the matrices $V^{(k-1)}$ and $V^{(k)}$ are connected by the formula

$$V^{(k)} = C^{(k)}V^{(k-1)} \tag{9.2a}$$

where $C^{(k)}$ is defined by (4.7). Writing

$$B^{(0)} = I, \qquad B^{(k)} = C^{(k)}B^{(k-1)} \tag{9.3}$$

it is seen that

$$V^{(k)} = B^{(k)}V^{(0)}. \tag{9.2b}$$

If, as before, we set

$$V^{(k)}U = A^{(k)} = (a_{ij}^{(k)})$$

then we have the relations

$$A^{(k)} = C^{(k)}A^{(k-1)} = B^{(k)}A^{(0)}. \tag{9.4}$$

Since $A^{(n)} = I$ it follows that $B^{(n)}$ is the inverse of $A^{(0)}$. According to (6.3b) the numbers c_{jk} appearing in the formulas (4.3) in the kth cycle of our algorithm are of the form

$$c_{jk} = a_{jk}^{(k-1)} \qquad (j = 1, \cdots, n).$$

It follows that the relation $A^{(k)} = C^{(k)}A^{(k-1)}$ can be put in the form

$$a_{kj}^{(k)} = a_{kj}^{(k-1)}/a_{kk}^{(k-1)} \tag{9.5α}$$

$$a_{ij}^{(k)} = a_{ij}^{(k-1)} - a_{ik}^{(k-1)}a_{kj}^{(k-1)}/a_{kk}^{(k-1)} \qquad (i \neq k). \tag{9.5β}$$

These are the formulas used in order to invert the matrix $A^{(0)}$ by the Gauss elimination method. The kth cycle of the algorithm (4.3) is therefore equivalent to applying the elimination algorithm (9.5) to $A^{(k-1)}$.

In applying the algorithm (4.3) the matrices $A^{(0)}$, $A^{(1)}$, $\cdots$, $A^{(n)}$ and $B^{(0)}$, $B^{(1)}$, $\cdots$, $B^{(n)}$ normally are not computed. Consequently even though $B^{(n)}$ is the inverse of $A^{(0)}$ it is not recorded. There is a simple modification of the method that will record the matrices $B^{(0)}$, $\cdots$, $B^{(n)}$. To this end let $V_1^{(0)}$ and U_1 be the augmented matrices

$$V_1^{(0)} = (V^{(0)}\ I), \qquad U_1 = \begin{pmatrix} U \\ 0 \end{pmatrix}.$$

Then

$$V_1^{(0)} U_1 = V^{(0)} U = A^{(0)}.$$

It follows that if the algorithm (4.3) is applied to the matrices $V_1^{(0)}$ and U_1 in place of $V^{(0)}$ and U, the numbers c_{jk} and hence the matrix $C^{(k)}$ will be unaltered. Consequently

$$V_1^{(k)} = B^{(k)} V_1^{(0)} = (V^{(k)}\ B^{(k)})$$

by virtue of (8.2) and (8.3). The final matrix

$$V_1^{(n)} = (V^{(n)}\ B^{(n)})$$

has the property that

$$V^{(n)} U = I, \qquad B^{(n)} A^{(0)} = I.$$

Hence, if U is a square matrix, $V^{(n)}$ is the inverse of U and $B^{(n)}$ is the inverse of A.

Two special cases are of interest. In the first case we select $U = I$, $V^{(0)} = A$. Then

$$V_1^{(0)} = (A\ I), \qquad U_1 = \begin{pmatrix} I \\ 0 \end{pmatrix},$$

hence

$$V_1^{(0)} U_1 = A^{(0)} = A$$

and

$$V_1^{(k)} = (A^{(k)}\ B^{(k)}),$$
$$V_1^{(n)} = (I\ A^{-1}).$$

In this event the matrices $A^{(k)}$ as well as $B^{(k)}$ are recorded, and the matrix A has been inverted by the formulas (8.5) extended to compute $B^{(k)}$ as well as $A^{(k)}$. This is one of the standard forms of the Gauss elimination method. In the second case we select $V^{(0)} = I$ and $U = A$. We then have

$$V^{(0)} U = A$$

and

$$V^{(k)} = B^{(k)}, \qquad B^{(k)}A = A^{(k)}$$

$$V_1^{(k)} = (B^{(k)}\ B^{(k)}), \qquad U_1 = \begin{pmatrix} A \\ 0 \end{pmatrix}.$$

The numbers c_{jk} resulting from this are the same as in the preceding case. They are however, computed in a different manner. Since $V^{(k)} = B^{(k)}$ there is no point in exhibiting the augmented matrix $V_1^{(k)}$.

As is well known, an effective elimination code cannot be written without introducing pivoting. The alternate procedure described in the previous section includes an effective pivoting device.

10. Principal values and principal directions of a system of vectors.* The success with which the algorithm (4.3) can be applied depends upon the situation of the vectors $u_1, \cdots, u_n$ relative to each other. If they are mutually orthogonal or very nearly so, then the rounding off error will be small. On the other extreme, if these vectors point more or less in the same direction the rounding off error will be large. In order to describe this situation more fully it will be convenient to introduce the concept of principal vectors and principal values of a set of n vectors $u_1, \cdots, u_n$ in an m-dimensional euclidean space.

Intuitively we shall define the first principal direction to be that direction in which the vectors $u_1, \cdots, u_n$ point more than in any other direction. In order to make this concept precise let x be a unit vector. Let z be the vector defined by the components

$$(x, u_1), \qquad (x, u_2), \cdots, (x, u_n)$$

of the vectors $u_1, \cdots, u_n$ on x. If U is the matrix whose column vectors are $u_1, \cdots, u_n$, then

$$z = U^*x.$$

The length of z is a measure of how much the vectors $u_1, \cdots, u_n$ point in the direction x. Select a unit vector x_1 such that the corresponding vector $z_1 = U^*x_1$ has maximum length. The length

$$\lambda_1 = \| z_1 \|$$

will be called the *first principal value* and x_1 will be called a corresponding *principal direction*. Clearly, λ_1 is the norm of U^* and hence also of U.

* The terms "singular values" and "singular directions" are also used for these concepts.

Let x_2 be a unit vector orthogonal to x_1 such that $z_2 = U^*x_2$ has maximum length. The length

$$\lambda_2 = \| z_2 \|$$

will be called the *second principal value* and x_2 a corresponding *principal direction*. Having chosen unit vectors $x_1, \cdots, x_{i-1}$ ($i \leqq n$) select a unit vector x_i orthogonal to $x_1, \cdots, x_{i-1}$ such that the length of $z_i = U^*x_i$ is a maximum. The length

$$\lambda_i = \| z_i \|,$$

if non-null, will be called the *i*th *principal value* and x_i a corresponding *principal direction*. A non-null vector proportional to x_i will be called a *principal vector corresponding to* λ_i.

Since

$$\| z \|^2 = \| U^*x \|^2 = (U^*x, U^*x) = (UU^*x, x),$$

it follows that $\lambda_1^2, \lambda_2^2, \cdots, \lambda_n^2$ are the eigenvalues of UU^* and $x_1, x_2, \cdots, x_n$ are corresponding eigenvectors. Consequently

$$(U^*x_i, U^*x_j) = (UU^*x_i, x_j) = 0 \qquad (i \neq j),$$

that is,

$$(z_i, z_j) = 0.$$

The vectors $z_1, z_2, \cdots,$ are therefore mutually orthogonal vectors of lengths $\lambda_1, \lambda_2, \cdots$. If r is the rank of U, then $\lambda_1, \cdots, \lambda_r$ will be different from zero, and $U^*x = 0$ for all vectors x orthogonal to $x_1, \cdots, x_r$. It is clear from our construction that each vector x_i ($i \leqq r$) lies in the space spanned by the column vectors $u_1, \cdots, u_n$ of U. The vectors $z_1, \cdots, z_r$ lie in the space spanned by the row vectors of U.

Our results will be summed up in the following:

THEOREM 10.1. *Let r be the rank oj the matrix U whose column vectors are $u_1, \cdots, u_n$. Then the system $u_1, \cdots, u_n$ has r linearly independent principal vectors in a maximal set. These vectors lie in the space spanned by $u_1, \cdots, u_n$. Let x_1 and x_2 be principal vectors corresponding to distinct principal values λ_1 and λ_2. Then x_1 and x_2 are orthogonal. So also are $z_1 = U^*x_1$ and $z_2 = U^*x_2$. If $x_1, \cdots, x_r$ are r mutually orthogonal principal vectors of unit length corresponding to principal values $\lambda_1, \cdots, \lambda_r$, then the vectors*

$$z_1 = U^*x_1, \cdots, z_r = U^*x_r$$

are mutually orthogonal vectors of lengths $\lambda_1, \cdots, \lambda_r$ respectively.

As a further result we have

THEOREM 10.2. *Let U be a matrix of rank r and let $x_1, \cdots, x_r$ be r mutually* *orthogonal non-zero vectors lying in the space spanned by the column vectors of U. If the vectors*

$$z_1 = U^*x_1, \cdots, z_r = U^*x_r \tag{10.1}$$

are mutually orthogonal, then $x_1, \cdots, x_r$ are principal vectors of the column vectors of U and $z_1, \cdots, z_r$ are principal vectors of the row vectors of U. The corresponding principal values are

$$\lambda_1 = \frac{\| z_1 \|}{\| x_1 \|}, \cdots, \lambda_r = \frac{\| z_r \|}{\| x_r \|}.$$

Let X and Z be the matrices whose column vectors are $x_1, \cdots, x_r$ and $z_1, \cdots, z_r$ respectively. Then

$$U^*X = Z. \tag{10.2}$$

By virtue of our hypotheses

$$X^*UU^*X = Z^*Z$$

is a diagonal matrix. Since the vector x_i is in the space spanned by the column vectors of U, it is orthogonal to the eigenvectors of UU^* corresponding to the eigenvalue $\lambda = 0$. From these facts it follows that x_i is an eigenvector of UU^* and, since $z_i \neq 0$, is accordingly a principal vector of the column vectors of U.

It remains to prove that z_i is a principal vector of the row vectors of U. To this end we may suppose that the vectors $x_1, \cdots, x_r$ are unit vectors. The lengths $\lambda_1, \cdots, \lambda_r$ of $z_1, \cdots, z_r$ are then the principal values of U corresponding to $x_1, \cdots, x_r$. Select vectors $y_1, \cdots, y_r$ such that

$$z_i = \lambda_i y_i .$$

Then we may write

$$Z = Y\Lambda$$

where $y_1, \cdots, y_r$ are the column vectors of Y and Λ is a diagonal matrix having $\lambda_1, \cdots, \lambda_r$ as its diagonal elements. We then have, by equation (10.2)

$$U^*X = Y\Lambda, \qquad X^*U = \Lambda Y^*. \tag{10.3}$$

Suppose, for the moment, that $m = n = r$. Then X and Y are orthogonal matrices and X^* and Y^* are their inverses. Consequently, if we multiply the second equation in (10.3) on the left by X and on the right by Y it is seen that

$$UY = X\Lambda. \tag{10.4}$$

This result is valid even if m, n, r are not equal. The column vectors of UY are therefore orthogonal. It follows from the results described in the last paragraph that the vectors $y_1, \cdots, y_r$ (and hence also $z_1, \cdots, z_r$) are principal vectors of the column vectors U^* and hence of the row vectors of U. The corresponding principal values are again $\lambda_1, \cdots, \lambda_r$. This completes the proof of Theorem 10.2.

Multiplying (10.4) on the right by Λ it is seen that

$$UZ = X\Lambda^2.$$

Hence, analogous to (9.1), we have

$$x_1 = \lambda_1^{-2}Uz_1, \cdots, x_r = \lambda_r^{-2}Uz_r. \tag{10.5}$$

This relation holds even if $x_1, \cdots, x_r$ are not unit vectors, but are any vectors satisfying the hypotheses of our theorem.

COROLLARY 1. *The row vectors and the column vectors have the same principal values. Accordingly these values also will be called the principal values of U.*

COROLLARY 2. *If U is symmetric its principal values are the absolute values of the nonzero eigenvalues of U. A principal vector of U is an eigenvector of U if the corresponding principal value is not the absolute value of two eigenvalues μ and $-\mu$.*

THEOREM 10.3. *Suppose that $m = n = r$. The matrix U is expressible in the form*

$$U = X\Lambda Y^*,$$

where X, Y, Λ have the properties described above. Moreover, the matrix

$$V = Y\Lambda^{-1}X^*$$

is the inverse U^{-1} of U so the row vectors $v_1, \cdots, v_r$ of V are biorthogonal to the column vectors $u_1, \cdots, u_n$ of U.

Since $m = n = r$ the matrices X and Y are orthogonal matrices so that

$$XX^* = X^*X = I, \qquad YY^* = Y^*Y = I.$$

Hence by (10.4) we have

$$U = UYY^* = X\Lambda Y^*.$$

Moreover

$$VU = Y\Lambda^{-1}X^*X\Lambda Y^* = I.$$

This proves the theorem.

We state, without proof, the following:

THEOREM 10.4. *The principal values of the general reciprocal V of U are the reciprocals of the principal values of U.*

If $m = n = r$ then $V = U^{-1}$ and this result follows from Theorem 10.3.

THEOREM 10.5. *Let*

$$h_i = \| u_i \|, \qquad h = \max_i h_i .$$

The first principal value λ_1 *of the system* $u_1, \cdots, u_n$ *(and hence of the matrix* U *with columns* $u_1, \cdots, u_n$*) lies on the interval*

$$h \leqq \lambda_1 \leqq (h_1^2 + \cdots + h_n^2)^{\frac{1}{2}} \leqq hn^{\frac{1}{2}}. \tag{10.6}$$

Let α_i *be an integer so that the vector*

$$p_i = 2^{-\alpha_i} u_i$$

has the property that the maximum of the absolute values of its components on the coordinate axes lie on the interval $\frac{1}{2} \leqq t < 1$. *Set*

$$\alpha = \max \alpha_i .$$

Then

$$2^{\alpha - 1} \leqq \lambda_1 \leqq 2^{\alpha} n. \tag{10.7}$$

The integer α_i *will be called the binary scale of* u_i.

In general 2^α is a better estimate of λ_1 than is indicated by (10.7). In the proof we can assume that $h = h_1$. If x is a unit vector, then the length of

$$z = U^* x$$

satisfies the inequality

$$\| z \| = (\Sigma \mid (x, u_i) \mid^2)^{\frac{1}{2}} \leqq (h_1^2 + \cdots + h_n^2)^{\frac{1}{2}}.$$

Hence

$$\lambda_1 \leqq (h_1^2 + \cdots + h_n^2)^{\frac{1}{2}} \leqq h\sqrt{n}.$$

Taking $x = h_i^{-1} u_i$ we see that

$$h_i \leqq \| z \| \leqq \lambda_1 .$$

Hence (10.6) holds. In order to prove (10.7) observe that

$$\tfrac{1}{2} \leqq \| p_i \| < \sqrt{n}, \qquad h_i = 2^{\alpha_i} \| p_i \|.$$

Hence

$$2^{\alpha_i - 1} \leqq h_i < 2^{\alpha_i} \sqrt{n}.$$

Combining this result with those just obtained we see that (10.7) holds.

THEOREM 10.6. *Let* V *be the general reciprocal of* U *and let* $v_1, \cdots, v_n$ *be the row vectors of* V. *Set*

$$k_i = \| v_i \|, \qquad k = \max k_i .$$

The least principal value λ_r *of* U *satisfies the inequalities*

$$k \leqq \lambda_r^{-1} \leqq (k_1^2 + \cdots + k_n^2)^{\frac{1}{2}} \leqq k\sqrt{n}. \tag{10.8}$$

Moreover, if β_i *is the binary scale of* v_i *and* $\beta = \max_i \beta_i$, *then*

$$2^{\beta-1} \leqq \lambda_r^{-1} \leqq 2^{\beta} n. \tag{10.9}$$

This result follows from Theorems 10.4 and 10.5. If $m = n = r$, then $V = U^{-1}$. Consequently, if β is the maximum scale of the row vectors and the column vectors of U^{-1}, then $2^{-\beta}$ is a reasonable estimate of the least principal value of U. If U is symmetric then $2^{-\beta}$ is an estimate of the distance from the origin to the eigenvalue that is closest to the origin.

11. Rounding off errors. Let U be a nonsingular matrix of dimension n. The purpose of the present section is to obtain a heuristic estimate of the accuracy to which the inverse V or U can be computed.

As before, let $u_1, \cdots, u_n$ be the column vectors of U and let $v_1, \cdots, v_n$ be the row vectors of its inverse V. We shall suppose that u_i and v_i are given in the scaled form

$$u_i = 2^{\alpha_i} p_i, \qquad v_i = 2^{\beta_i} q_i, \tag{11.1}$$

where the *binary scales* α_i and β_i have been chosen so that each of the vectors p_i and q_i has the property that the maximum of the absolute values of its projections on the coordinate axes lie on the interval $\frac{1}{2} \leqq t < 1$. This representation is convenient for computational purposes.

Suppose now that each of the components of p_i is representable by a binary number composed of γ binary digits. Let q_i' be the vector obtained from q_i by rounding each of its components to γ binary digits. Then

$$q_i = q_i' + 2^{-\gamma} q_i''$$

where the binary scale of q_i'' is at most zero. Set

$$v_i' = 2^{\beta_i} q_i'$$

and let V' be the matrix having $v_1', \cdots, v_n'$ as its row vectors. We shall be concerned with the error induced by the substitution of V' for the inverse V of U. The matrix

$$E = I - V'U$$

will be taken as a measure of this error. Its elements ϵ_{ij} are

$$\epsilon_{ij} = \delta_{ij} - (v_i', u_j) = 2^{\beta_i + \alpha_j - \gamma}(q_i'', p_j). \tag{11.2}$$

Let ρ be the maximum of the binary scales of these elements. Then ρ is the greatest integer such that

$$|\epsilon_{ij}| < 2^{-\rho}.$$

Moreover,

$$\rho = \max (\gamma - \beta_i - \alpha_j - \mu_{ij}) \tag{11.3}$$

where μ_{ij} is the scale of (q_i'', p_j). We take

$$\rho_0 = \gamma - \beta - \alpha \tag{11.4}$$

with

$$\alpha = \max_i \alpha_i, \qquad \beta = \max_i \beta_i$$

as an estimate of ρ. This choice is suggested by the relations (11.2). As a further justification, observe that because the binary scales of q_i'' and p_j are at most zero, we have

$$|(q_i'', p_j)| < n.$$

Consequently

$$|\epsilon_{ij}| < n2^{-\rho_0}$$

and

$$\rho \geqq \rho_0 - \log_2 n. \tag{11.5}$$

In computational work one frequently alters the matrix U so that the binary scales $\alpha_1, \cdots, \alpha_n$ are equal to zero. This change adds α_i to β_i but does not alter q_i''. The scale μ_{ij} of the product (q_i'', p_j) is unaltered. However, the scale ρ may be altered since this change is equivalent to replacing α_j by α_i in (11.3). Suppose now that $\alpha_j = 0$ and that the index i has been selected so that $\beta_i = \beta$. Then

$$\rho \leqq \rho_0 + \nu$$

where ν is the maximum of the binary scales of the inner products

$$(q_i'', p_1), \cdots, (q_i'', p_n).$$

The vectors $p_1, \cdots, p_n$ have zero as their binary scale. Since the vectors $p_1, \cdots, p_n$ are linearly independent and q_i'' is a random vector of scale at most zero, it will usually happen that ν is numerically small and hence that ρ_0 is a good estimate of ρ. In the numerical experiments carried out by the author the numbers ρ and ρ_0 were in close agreement. In view of these results one can reasonably expect to lose about $\alpha + \beta$ binary digits in computing the inverse of a matrix.

Let $\lambda_1, \cdots, \lambda_n$ be the principal values of U. The ratio

$$\sigma = \lambda_n/\lambda_1$$

is sometimes called the *condition number* of U. Using the inequalities

$$2^{\alpha-1} \leqq \lambda_1 \leqq 2^{\alpha}n$$

$$2^{\beta-1} \leqq \lambda_n^{-1} \leqq 2^{\beta} n$$

given by (10.7) and (10.9) it is seen that the $\alpha + \beta$ and σ are related by the condition

$$n^{-2} 2^{-(\alpha+\beta)} \leqq \sigma \leqq 2^{2-(\alpha+\beta)}.$$

The inequality on the left is normally a gross underevaluation of σ. In the few cases checked by the author the quantity $2^{-(\alpha+\beta)}$ was a reasonable estimate of σ. Thus, $-\log_2 \sigma$ is also an estimate of the significant figures lost in inverting a matrix. Or conversely one can estimate σ by $2^{-\delta}$, where δ is the number of significant figures lost.

12. Computation of principal values and principal vectors by orthogonalization. The results given in Theorem 10.2 suggests a method of computing the principal values and the principal vectors of a matrix U. Recall that if U is symmetric, then the principal values are absolute values of eigenvalues. Moreover, the principal vectors are eigenvectors if there is no pair of eigenvalues of U which differ only in sign. The method therefore can be used to compute eigenvalues and eigenvectors of symmetric (or hermitian) matrices.

The method consists of generating an orthogonal matrix Y such that the column vectors of

$$W = UY$$

are mutually orthogonal. According to Theorm 10.2 the nonnull column vectors of W are principal vectors of the column vectors of U and their lengths are the corresponding principal values. If U is nonsingular, then the column vectors can be normalized so as to obtain an orthogonal matrix X. The column vectors of

$$Z = U^*X$$

are mutually orthogonal and are the principal vectors of the row vectors of U.

Let $u_1, \cdots, u_n$ be the column vectors of U and let T_{ij} $(i < j,\ i, j = 1, \cdots, n)$ be the transformation defined to perform the following steps.

(i) Interchange u_i and u_j if $\| u_i \| < \| u_j \|$.

(ii) Replace u_i and u_j respectively by u_i' and u_j', where

$$\begin{aligned} u_i' &= u_i \cos \varphi + u_j \sin \varphi, \\ u_j' &= -u_i \sin \varphi + u_j \cos \varphi, \end{aligned} \tag{12.1}$$

where φ is an angle between $-\pi/2$ and $\pi/2$ chosen so that u_i' and u_j' are orthogonal, the angle φ being zero if u_i and u_j are orthogonal.

(iii) Leave the remaining vectors unaltered.

The transformation T_{ij} is an orthogonal transformation. There are $n(n - 1)/2$ transformations of this type. Order these transformations so that T_{hi} precedes T_{jk} if $h < j$ or if $h = j$ and $i < k$. Let T be the transformation obtained by applying these $n(n - 1)/2$ transformations T_{ij} successively (on the right) as prescribed by their order. Thus, symbolically

$$T = T_{12}T_{13} \cdots T_{1n}T_{23} \cdots T_{2n}T_{34} \cdots T_{n-1,n} .$$

The application of this transformation to the vectors $u_1, \cdots, u_n$ is equivalent to replacing the matrix U by

$$U_1 = UR_1$$

where R_1 is an orthogonal matrix. Applying the transformation to U_1 results in

$$U_2 = U_1R_2$$

and so on. Thus we obtain matrices

$$U_k = U_{k-1}R_k = UY_k \qquad (k = 1, 2, 3, \cdots)$$

where R_k and

$$Y_k = R_1R_2 \cdots R_k$$

are orthogonal matrices. As will be seen presently, the limit

$$W = \lim_{k=\infty} U_k = \lim_{k=\infty} UY_k \tag{12.2}$$

will normally exist and be of the form

$$W = UY$$

where Y is an orthogonal matrix. Moreover its column vectors $w_1, \cdots, w_n$ are mutually orthogonal and

$$\lambda_i = \| w_i \| \geqq \lambda_j = \| w_j \| \qquad (i < j;\, i, j = 1, \cdots, n).$$

The non-zero vectors $w_1, \cdots, w_r$ in this set are principal vectors of $u_1, \cdots, u_n$ and $\lambda_1, \cdots, \lambda_r$ are the corresponding principal values.

The method just described is equivalent to the modified Jacobi method for finding the eigenvalues of U^*U, whose convergence has been established by Forsythe and Henrici.[4]

In order to see this connection let

$$a_{pq} = (u_p, u_q) \qquad (p, q = 1, \cdots, n)$$

[4] G. E. Forsythe and P. Henrici, *The cyclic Jacobi method for computing the principal values of a complex matrix.* (Submitted for publication.)

and

$$a'_{pq} = (u_p', u_q')$$

where $u_1', \cdots, u_n'$ are the vectors obtained from $u_1, \cdots, u_n$ by the transformation T_{ij}. In view of (12.1) and the relation $u_p' = u_p$ if $p \neq i$ and $p \neq j$ we have

$$\begin{aligned}
a'_{iq} &= a_{iq} \cos \theta + a_{jq} \sin \theta && (q \neq i, q \neq j) \\
a'_{jq} &= -a_{iq} \sin \theta + a_{jq} \cos \theta \\
a'_{pi} &= a_{pi} \cos \theta + a_{pj} \sin \theta && (p \neq i, p \neq j) \\
a'_{pj} &= -a_{pi} \sin \theta + a_{pj} \cos \theta \\
a_{ii} &= a_{ii} \cos^2 \theta + 2a_{ij} \cos \theta \sin \theta + a_{jj} \sin^2 \theta \\
0 = 2a'_{ij} &= -(a_{ii} - a_{jj}) \sin 2\theta + 2a_{ij} \cos 2\theta \\
a'_{jj} &= a_{ii} \sin^2 \theta - 2a_{ij} \sin \theta \cos \theta + a_{jj} \cos^2 \theta.
\end{aligned}$$

These formulas are those of a single step in the application of the Jacobi method for diagonalizing a matrix $A = U^*U$. It follows from the results given by Forsythe and Henrici that

$$D = \lim_{k=\infty} U_k^* V_k \tag{12.3}$$

exists and is a diagonal matrix. If U has distinct principal values and is of rank $r \geqq n - 1$, then the limit (12.2) exists also, as one readily verifies. The existence of the limit (12.3) as a diagonal matrix is sufficient for the purposes of numerical analysis. Consequently we shall not pursue the question of existence of the limit (12.2) further at this time.

13. Eigenvalues by inversion. The method of inverting matrices here given is particularly suitable for obtaining eigenvalues and eigenvectors by the method of differential corrections. This method is frequently called Newton's method. It can also be looked upon as a power method, in which a power of the inverse of a matrix instead of the matrix itself is used.

We shall consider the general case in which the scalars are complex numbers and A is an arbitrary matrix. We seek a value λ_0 and two vectors x_0, y_0 such that

$$\begin{aligned}
Ax_0 &= \lambda_0 x_0 \\
y_0^* A &= \lambda_0 y_0^*.
\end{aligned}$$

If λ_0 is a simple eigenvalue no generality is lost if we add the restriction

$$y_0^* x_0 = 1.$$

Here x_0, y_0 are considered to be matrices composed of one column. We shall assume that λ_0 is a simple eigenvalue. Then the determinant of the matrix

$$U_0 = \begin{pmatrix} A - \lambda_0 I & x_0 \\ y_0^* & 0 \end{pmatrix}$$

is different from zero, as one readily verifies.

The method of finding eigenvalues and eigenvectors by inversion of matrices can be described as follows: Select an initial estimate x_1, y_1, λ_1 of x_0, y_0, λ_0. Having obtained the kth estimate x_k, y_k, λ_k form the matrix

$$U_k = \begin{pmatrix} A - \lambda_k I & x_k \\ y_k^* & 0 \end{pmatrix}. \tag{13.1a}$$

Compute its inverse

$$U_k^{-1} = \begin{pmatrix} B_k & x_{k+1} \\ y_{k+1}^* & -\rho_{k+1} \end{pmatrix} \tag{13.1b}$$

and select x_{k+1}, y_{k+1} from the inverse as indicated. Compute

$$\sigma_{k+1} = y_{k+1}^* x_{k+1}, \qquad \lambda_{k+1} = \lambda_k + \rho_{k+1}/\sigma_{k+1}. \tag{13.1c}$$

The system x_{k+1}, y_{k+1}, λ_{k+1} is the $(k + 1)$-estimate of the solution. An alternate choice of λ_{k+1} is given by the formula

$$\sigma_k = y_k^* x_k, \qquad \lambda_{k+1} = \lambda_k + \rho_k/\sigma_k. \tag{13.2}$$

With this choice the method is a method of differential corrections, as we shall see presently, and is commonly called Newton's method. We shall also see that if we modify the method further and select $\lambda_{k+1} = \lambda_k$ at each step then x_{k+1} and y_{k+1}^* are proportional to

$$(A - \lambda_1 I)^{-k} x_1, \qquad y_1^*(A - \lambda_1 I)^{-k}. \tag{13.3}$$

Consequently, the method can also be considered to be an extension of the inverse power method. As a matter of fact if one desires to find the eigenvalues nearest to λ_1 one should select $\lambda_{k+1} = \lambda_k$ until ρ_k becomes stable and then proceed with (13.1c) or (13.2).

In order to see more clearly the connections between successive iterations observe that since $U_k U_k^{-1} = I$ one obtains, by computing the last column of $U_k U_k^{-1}$, the relation

$$(A - \lambda_k I) x_{k+1} = \rho_{k+1} x_k, \qquad y_k^* x_{k+1} = 1. \tag{13.4a}$$

Similarly, from the identity $U_k^{-1} U_k = I$, we have

$$y_{k+1}^*(A - \lambda_k I) = \rho_{k+1} y_k^*, \qquad y_{k+1}^* x_k = 1, \tag{13.4b}$$

consequently

$$\rho_{k+1} = y_{k+1}^*(A - \lambda_k I)x_{k+1} . \tag{13.5}$$

If we select λ_{k+1} by the use of formula (13.1c) we have

$$\rho_{k+1} = (\lambda_{k+1} - \lambda_k)y_{k+1}^* x_{k+1} \tag{13.6a}$$

and hence also

$$\lambda_{k+1} = \frac{y_{k+1}^* A x_{k+1}}{y_{k+1}^* x_{k+1}} . \tag{13.6b}$$

The selection (13.1c) of λ_{k+1} was made so that λ_{k+1} would be given by the generalized Rayleigh quotient (13.6b). In the case of hermitian matrices the selection $x_1 = y_1$ is made. We then have $x_k = y_k$ for all values of k and λ_k is given by the usual Rayleigh quotient.

The formulas (13.4) hold independently of the method of deriving λ_{k+1} from λ_k . Inasmuch as

$$x_{k+1} = \rho_{k+1}(A - \lambda_k I)^{-1} x_k , \qquad y_{k+1}^* = \rho_{k+1} y_k^* (A - \lambda_k I)^{-1}$$

it is clear that if we select $\lambda_{k+1} = \lambda_k$ at each step, then x_{k+1} and y_{k+1}^* will be proportional to the vectors (13.3), as stated above. This fact insures the convergence of the method if we precondition our vectors by selecting $\lambda_{k+1} = \lambda_k$ in the initial stages of the computation.

In order to see the connection with the method of differential corrections let us first recall Newton's method in simpler form. The problem at hand is to solve the equation

$$f(z) = 0.$$

Let z_1 be an estimate of the solution and rewrite the equation in the form

$$f(z_1 + \delta z) = 0.$$

In order to obtain an estimate δz_1 of the solution δz let δz_1 be the solution of the linear equation

$$f(z_1) + \delta f = 0$$

where δf is the variation or differential of f at $z = z_1$. The new estimate of the original equation is $z_2 = z_1 + \delta z_1$. The algorithm is then repeated with z_1 replaced by z_2 and so on. In this manner a sequence $z_1 , z_2 , \cdots$ of estimates are obtained which will converge quadratically to the solution z_0 if the linearized equation is nonsingular at $z = z_0$ and if z_1 is a sufficiently accurate estimate of z_0 .

Consider now the problem of solving the system of equations

$$(A - \lambda I)x = 0, \qquad y^*(A - \lambda I) = 0, \qquad y^* x = 1,$$

by Newton's method. The corresponding linearized equation can be written in the form (assuming that $y^*x = 1$)

$$(13.7)\qquad \begin{aligned} (A - \lambda I)(x + \delta x) &= \delta\lambda x, \\ (y^* + \delta y)(A - \lambda I) &= \delta\lambda y^*, \\ y^*(x + \delta x) + (y^* + \delta y)x &= 2. \end{aligned}$$

From the first two of these equations with

$$x_1 = x + \delta x, \qquad y_1 = y + \delta y$$

it is seen that

$$y_1^*(A - \lambda I)x_1 = \delta\lambda y_1^* x = \delta\lambda y^* x_1 .$$

Hence

$$y_1^* x = y^* x_1 ,$$

provided that $\delta\lambda \neq 0$, as we shall assume. It follows that the equations (13.7) are equivalent to the equations

$$(13.8)\qquad \begin{aligned} (A - \lambda I)x_1 &= \delta\lambda x, \qquad & y^* x_1 &= 1 \\ y_1^*(A - \lambda I) &= \delta\lambda y^*, \qquad & y_1^* x &= 1. \end{aligned}$$

These equations are of the form (13.4). The new value of λ is

$$\lambda_1 = \lambda + \delta\lambda.$$

If in (13.8) we drop the condition that $y^*x = 1$ then the formula for λ_1 is given by the equation

$$(13.9)\qquad \sigma = y^* x, \qquad \lambda_1 = \lambda + \delta\lambda/\sigma$$

and we obtain a new set x_1, y_1, λ_1 that is equivalent to the one obtained under the assumption that $\sigma = 1$, as one readily verifies. The equations (13.8) and (13.9) are of the form (13.4) and (13.6) and the connection between the two methods is established.

The method of inversion by biorthogonalization is particularly well adapted to the method of finding eigenvalues and eigenvectors by inversion. This is because U_k^{-1} is normally a very good estimate of U_{r+1}^{-1}. Hence in order to compute U_{k+1}^{-1} only small corrections need to be made.

The assumption that U_0 be nonsingular was made so that the theorems on Newton's method would be immediately applicable. However, it is not essential to the problem as stated. If U_0 is singular, one can expect to have considerable difficulties with rounding off errors.

Ostrowski* has shown that if (13.5) is used in the case of an hermitian matrix A, the convergence is cubic in character. He has also shown that under certain conditions the same will be true when A is nonhermitian.

14. The general reciprocal of a matrix. The purpose of this section is to define the concept of the general reciprocal of a matrix U. This concept was first defined by E. H. Moore.[5] It has been rediscovered recently by R. Penrose.[6] The definition here given is geometrical in character. It is hoped that this description of the general reciprocal will give one a better understanding of this useful concept.

Consider an $m \times n$ dimensional matrix U. The transformation

$$x = Uy \tag{14.1}$$

transforms a vector y in an n-dimensional vector space $\mathfrak{E}_n$ into a vector x in an m-dimensional vector space $\mathfrak{E}_m$. Let $\mathfrak{N}$ be the set of vectors y which are annihilated by U, that is, they satisfy the equation

$$Uy = 0.$$

The orthogonal complement $\mathfrak{R}$ of $\mathfrak{N}$ will be called the *carrier of* U *in* $\mathfrak{E}_n$. It is spanned by the row vectors of U. Under the transformation (14.1) the carrier $\mathfrak{R}$ is mapped into a subspace $\mathfrak{R}^*$ of $\mathfrak{E}_m$. It is easily seen that $\mathfrak{R}^*$ is the carrier of U^* in $\mathfrak{E}_m$. It is spanned by the column vectors of U. The mapping (14.1) of $\mathfrak{R}$ into $\mathfrak{R}^*$ is one to one. Let

$$y = U^{-1}x \tag{14.2}$$

be the inverse mapping of $\mathfrak{R}^*$ into $\mathfrak{R}$. Extend the definition of U^{-1} over $\mathfrak{E}_m$ so that U^{-1} annihilates the orthogonal complement of $\mathfrak{R}^*$. The linear transformation so defined determines an $n \times m$ matrix, also denoted by U^{-1}, whose row vectors span $\mathfrak{R}^*$ and column vectors span $\mathfrak{R}$. It is called the *general reciprocal* of U. It is clear that $(U^{-1})^* = U^{*-1}$ is the general reciprocal of U^*.

Every matrix U accordingly uniquely determines three further matrices U^*, U^{-1}, U^{*-1}. If $U^* = U^{-1}$ then $U = U^{*-1}$. In this event U will be called

* A. Ostrowski, *On iterative methods for computing eigenvalues*, to be published in Z. Angew. Math. Phys. See also S. H. Crandall, *Iterative procedures related to relaxation methods for eigenvalue problems*, Proc. Roy. Soc. London. Ser. A, 207 (1951), pp. 416–423; R. von Holdt, *An Iterative procedure for the calculation of eigenvalues and eigenvectors of a real symmetric matrix*, J. Assoc. Comput. Mach., 3 (1956), pp. 223–238; H. Wielandt, *Das Iterationsverfahren bei nichtselbstadjungierten linearen Eigenwertaufgaben*, Math. Z., 50 (1944), pp. 93–143.

[5] E. H. Moore, *General analysis, Part I*, Mem. Amer. Philos. Soc. 1 (1935), p. 197.

[6] R. Penrose, *A generalized inverse for matrices*, Proc. Cambridge Philos. Soc., 51 (1953), pp. 406–413.

an *isometry*,† because its carrier $\mathfrak{R}$ is mapped isometrically onto the carrier $\mathfrak{R}^*$ of U^*. This definition is at variance with that normally used in the theory of Hilbert space, in that we do not require $\mathfrak{R}$ to coincide with $\mathfrak{E}_n$. If $\mathfrak{E}_m = \mathfrak{E}_n$ and $U = U^*$, then U is *symmetric* (or hermitian). If $\mathfrak{E}_m = \mathfrak{E}_n$ and $U = U^* = U^{-1} = U^{*-1}$ then U is a projection. Thus a projection is a symmetric isometry.

It should be observed that the product

$$F = U^{-1}U$$

is the projection in $\mathfrak{E}_n$ associated with the subspace $\mathfrak{R}$, the carrier of U. Similarly

$$E = UU^{-1}$$

is the projection in $\mathfrak{E}_m$ associated with the carrier $\mathfrak{R}^*$ of U^*.

In the next section it will be seen that every matrix U has associated with it an isometry R sharing the same carrier as U and which is connected with the principal vectors of U.

15. Isometries associated with matrices. Consider now an $m \times n$ dimensional matrix U of rank r. As was seen in Section 10, it has associated with it three matrices X, Y, Λ such that

$$X^*U = \Lambda Y^*. \tag{15.1}$$

The matrix X is an $m \times r$ dimensional matrix whose column vectors $x_1, \cdots, x_r$ form an orthonormal basis for the carrier $\mathfrak{R}^*$ of U^* and are the principal vectors of the column vectors of U. We have

$$X^*X = I_r, \qquad XX^* = E, \tag{15.2a}$$

where I_r is the r-dimensional identity and E is the $m \times m$ dimensional projection matrix determined by $\mathfrak{R}^*$. Hence X is an isometry and

$$EX = X, \qquad EU = U. \tag{15.2b}$$

Similarly, the column vectors $y_1, \cdots, y_r$ of the $n \times r$ dimensional matrix Y form an orthonormal basis of the carrier $\mathfrak{R}$ of U and are the principal vectors of the column vectors of U^*. Moreover

$$Y^*Y = I_r, \qquad YY^* = F \tag{15.3a}$$

$$FY = Y, \qquad FU^* = U^*. \tag{15.3b}$$

Finally the matrix Λ is an $r \times r$ dimensional diagonal matrix whose diagonal elements are the principal values $\lambda_1, \cdots \lambda_r$ of U. Since

$$U = EU = XX^*U = X\Lambda Y^*,$$

† The term "partial isometry" is also used for this concept.

the formula

$$U = X\Lambda Y^* \tag{15.4}$$

holds in general, as was stated in Section 10.

The $m \times n$ dimensional matrix

$$R = XY^*$$

satisfies the conditions

$$RR^* = E, \qquad R^*R = F \tag{15.5}$$

and hence is an isometry having the same carrier as U, the carrier of R^* coinciding with that of U^*. This can be seen from the computations

$$RR^* = XY^*YX^* = XI_rX^* = XX^* = E,$$

$$R^*R = YX^*XY^* = YI_rY^* = YY^* = F.$$

The square roots P and Q of UU^* and U^*U respectively are connected with U and R by the formulas

$$P = UR^* = RU^*, \qquad Q = U^*R = R^*U \tag{15.6a}$$

$$U = PR = RQ, \qquad U^* = QR^* = R^*P. \tag{15.6b}$$

This follows because

$$UR^* = X\Lambda Y^*YX^* = X\Lambda X^* = P$$

$$R^*U = YX^*X\Lambda Y^* = Y\Lambda Y^* = Q$$

$$P = P^* = RU^*, \qquad Q = Q^* = U^*R$$

$$U = EU = RR^*U = RQ, \qquad U^* = QR^*$$

$$U = UF = UR^*R = PR, \qquad U^* = R^*P.$$

It should be noted that

$$ER = R = RF = ERF.$$

This last statement is equivalent to the statement that the carrier of R is $\mathfrak{R}$ and that of R^* is $\mathfrak{R}^*$. It is interesting to observe that

$$X = RY, \qquad Y = R^*X$$

or equivalently that

$$x_i = Ry_i\,, \qquad y_i = R^*x_i \qquad (i = 1, \cdots, n).$$

Thus R is the isometry that maps the principal vectors of the row vectors of U into the principal vectors of the column vectors of U.

The matrix R is uniquely determined by the matrix U. If U is a positive definite matrix A, then $R = I$. We shall see that in the general case R plays a role relative to U much the same as that played by the identity I relative to a positive definite matrix A. For example, a number λ is a principal value of U if and only if there is a vector $y \neq 0$ in $\mathfrak{R}$ such that

$$Uy = \lambda Ry \tag{15.7a}$$

or equivalently if and only if there is a vector $x \neq 0$ in $\mathfrak{R}^*$ such that

$$x^*U = \lambda x^*R. \tag{15.7b}$$

The vectors x and y are corresponding principal vectors of U in the sense described in Section 10 above. The equations (15.7) are analogous to the equation

$$Az = \lambda Iz$$

that determines the eigenvalues and eigenvectors of A. In order to carry out this analogy further recall that corresponding to the eigenvalues $\lambda_1, \cdots, \lambda_n$ of A there exist projections $E_1, \cdots, E_n$ such that

$$\begin{aligned} E_iE_j &= 0 \qquad\qquad (i \neq j)\\ E_iU &= UE_i \\ I &= E_1 + \cdots + E_n \\ A &= \lambda_1E_1 + \cdots + \lambda_nE_n \\ A^{-1} &= \lambda_1^{-1}E_1 + \cdots + \lambda_1^{-1}E_n . \end{aligned}$$

Similarly, corresponding to the principal values $\lambda_1, \cdots, \lambda_r$ of U there exist isometries $R_1, \cdots, R_r$ such that

$$R_iR_j^* = 0, \qquad R_i^*R_j = 0 \qquad (i \neq j) \tag{15.8a}$$

$$R_iR^*U = UR^*R_i \tag{15.8b}$$

$$R = R_1 + \cdots + R_r \tag{15.8c}$$

$$U = \lambda_1R_1 + \cdots + \lambda_rR_r \tag{15.8d}$$

$$U^{-1} = \lambda_1^{-1}R_1^* + \cdots + \lambda_r^{-1}R_r^* \tag{15.8e}$$

where U^{-1} is the general reciprocal of U. The outer product

$$R_i = x_iy_i^* \qquad (i = 1, \cdots, r) \tag{15.9}$$

of the principal vectors x_i and y_i (considered as one rowed matrices) have this property, as we shall see presently. If $\lambda_1, \cdots, \lambda_r$ are distinct, then the isometries $R_1, \cdots, R_r$ are uniquely determined. In general, the sum of the isometries R_i for which the corresponding principal values λ_i are equal is unique. The decomposition (15.8d) and related results were given by J. W. Gibbs* in 1884 and more recently by Penrose in the paper cited above.

In order to establish the relations (15.8) it is convenient to write the formula for R_i given in (15.9) in a somewhat different form. To this end let X_i be the $m \times r$ dimensional matrix that agrees with X in the ith column and is zero elsewhere. Similarly, let Y_i be the $n \times r$ dimensional matrix that agrees with Y in the ith column and is zero elsewhere. Then we have the relations

$$\text{(15.10a)} \qquad X = X_1 + \cdots + X_r, \qquad Y = Y_1 + \cdots + Y_r$$

$$\text{(15.10b)} \qquad X_iX_j^* = 0, \qquad X_iY_j^* = 0, \qquad Y_iY_j^* = 0 \qquad (i \neq j)$$

$$\text{(15.10c)} \qquad X_iX_i^* = X_iX^* = XX_i^*, \qquad Y_iY_i^* = Y_iY^* = YY_i^*$$

$$\text{(15.10d)} \qquad X_iY_i^* = X_iY^* = XY_i^*,$$

as one verifies from the definitions of X_i and Y_i. The matrices

$$\text{(15.11)} \qquad E_i = X_iX_i^*, \qquad R_i = X_iY_i^*, \qquad F_i = Y_iY_i^*$$

have the properties

$$\begin{aligned} E_1 + \cdots + E_r &= XX^* = E \\ R_1 + \cdots + R_r &= XY^* = R \\ F_1 + \cdots + F_r &= YY^* = F \end{aligned}$$

$$\text{(15.12)} \qquad \begin{aligned} &E_iE_j = 0, \qquad R_i^*R_j = 0, \qquad R_iR_j^* = 0, \qquad F_iF_j = 0 \quad (i \neq j) \\ &E_i^2 = E_i = E_i^*, \qquad F_i^2 = F_i = F_i^* \\ &E_i = R_iR_i^* = R_iR^* = RR_i^*, \\ &\qquad\qquad F_i = R_i^*R_i = R_i^*R = R^*R_i \\ &R_i = E_iR_i = E_iR = RF_i = R_iF_i. \end{aligned}$$

The properties are easily established with the help of (15.10). The matrices E_i and F_i are projections as well as isometries. The matrix R_i is an isometry and is identical with that given by formula (15.9). Inasmuch as

$$X\Lambda = \lambda_1X_1 + \cdots + \lambda_rX_r$$

* J. W. Gibbs., *The collected works of J. Willard Gibbs*, Yale Univ. Press, vol. II, part 2, pp. 61–65.

it follows, from (15.4) and (15.10), that

$$U = \lambda_1 R_1 + \cdots + \lambda_r R_r .$$

The matrix

$$U^{-1} = \lambda_1^{-1} R_1^* + \cdots + \lambda_r^{-1} R_r^*$$

has the property that

$$UU^{-1} = E_1 + \cdots + E_r = E$$

$$U^{-1}U = F_1 + \cdots + F_r = F$$

and hence is the general reciprocal of U. A simple computation shows that

$$R_i^* R^* U = UR^* R_i = \lambda_i R_i .$$

This completes the proof of the relations (15.8).

Let $\mathfrak{A}$ be the class of all $m \times n$ dimensional matrices A such that

$$AR^*U = UR^*A, \qquad EA = AF = A.$$

Let $\mathfrak{B}$ be the set of all matrices B in $\mathfrak{A}$ such that

$$BR^*A = AR^*B$$

for every A in $\mathfrak{A}$. We shall show that every matrix B in $\mathfrak{B}$ is of the form

$$B = \mu_1 R_1 + \cdots + \mu_r R_r \tag{15.13}$$

where $\mu_1, \cdots, \mu_r$ are real numbers and $\mu_i = \mu_j$ whenever $\lambda_i = \lambda_j$. In other words, there is a function $f(\lambda)$ such that

$$B = f(U) = f(\lambda_1)R_1 + \cdots + f(\lambda_r)R_r .$$

In order to prove this result observe first that, by (15.8b), the matrices R_i are in $\mathfrak{A}$. Given a matrix B in $\mathfrak{B}$ we have accordingly

$$R_i R^* B = E_i B = BR^* R_i = BF_i .$$

Consider the matrix

$$B_i = E_i B = BF_i = E_i B F_i . \tag{15.14}$$

Given a vector y orthogonal to y_i and a vector x orthogonal to x_i we have

$$B_i y = 0, \qquad x^* B_i = 0$$

and it follows that B_i is of the form

$$B_i = \mu_i R_i .$$

Consequently, by (15.14)

$$B = EB = \Sigma E_i B = \Sigma \mu_i R_i .$$

It remains to show that $\mu_i = \mu_j$ if $\lambda_i = \lambda_j$. To this end suppose that $\lambda_i = \lambda_j$ $(i \neq j)$. Let

$$R_{ij} = x_i y_j^*.$$

Then

$$R_i R^* R_{ij} = E_i R_{ij} = R_{ij}$$

$$R_h R^* R_{ij} = E_h R_{ij} = 0 \qquad (h \neq i)$$

$$R_{ij} R^* R_k = R_{ij} F_k = 0 \qquad (k \neq j)$$

$$R_{ij} R^* R_j = R_{ij} F_j = R_{ij} .$$

From these relations it follows that

$$E R_{ij} = R_{ij} F = R_{ij}$$

$$R_{ij} R^* U - U R^* R_{ij} = (\lambda_j - \lambda_i) R_{ij} = 0.$$

Hence R_{ij} is in $\mathfrak{A}$ and

$$0 = R_{ij} R^* B - B R^* R_{ij} = (\mu_j - \mu_i) R_{ij} .$$

This is possible only in case $\mu_i = \mu_j$. The result described in the previous section is therefore established.

Setting

$$U^{(0)} = U, \qquad U^{(k)} = U^{(k-1)} R^* U = U R^* U^{(k-1)}$$

it is readily seen that $\lambda_1^k, \cdots, \lambda_r^k$ are the principal values of $U^{(k)}$ and that R is its associated isometry. The matrix $U^{(k)}$ belongs to the class $\mathfrak{B}$ and corresponds to the function $f(\lambda) = \lambda^k$. The principal vectors of $U^{(k)}$ are also principal vectors of U.

UNIVERSITY OF CALIFORNIA, LOS ANGELES