QUARTERLY OF APPLIED MATHEMATHEMATICS

Vol. VIII January, 1951 No. 4

NEW MATRIX TRANSFORMATIONS FOR OBTAINING CHARACTERISTIC VECTORS*

By WILLIAM FELLER (*Princeton University*)

AND

GEORGE E. FORSYTHE (*National Bureau of Standards, Los Angeles*)

1. Summary of methods. Let A be a non-defective (see §2) square matrix of order n, symmetric or not, for which it is desired to determine some of the characteristic values ν and associated column-vectors X and row-vectors Y. In terms of matrix products these quantities are defined by the relations

$$AX = \nu X, \qquad YA = \nu Y. \tag{1}$$

A class of numerical procedures is based on iteration methods to obtain one characteristic value λ and the associated vectors C, R. Then A is transformed into a matrix A' and a new iteration is used to obtain a characteristic value ν' and characteristic vectors X', Y' of A', which can then be converted into corresponding quantities ν, X, Y for A. If more values are wanted, one can continue by transforming A' to A'', and so on. Vector iteration schemes for getting one characteristic value of a matrix were described in 1929 in [15]; these methods are explained and extended in [1, 13, 8, 9, 5].

In the present paper we are not interested in the iteration procedure as such, but wish to discuss a class of transformations whereby A' is obtained from A. The earliest of these known to us is "deflation," suggested by Hotelling [6, 7] for symmetric matrices and extended in Aitken's thorough study [1] to non-symmetric matrices, defective or not. In [3] and [4, p. 143] Duncan and Collar introduced a different transformation (see §3) for non-defective matrices; this was restated in [10] and [16]. It has the advantage that it reduces the order of the matrix, but it destroys the symmetry. In a relatively inaccessible paper [13] Semendiaev gave a careful exposition of Aitken's techniques, and extended them to cover the case of multiple characteristic values in full generality. His transformation is very general; in the simplest case it somewhat resembles that of Duncan and Collar. Semendiaev expressed his transformation in the form of a matrix relation $A' = UAU^{-1}$. Blanch has devised (unpublished) another modification of the Duncan-Collar reduction in the form UAU^{-1}.

In [14] Tucker published a related transformation yielding a matrix A' of order $n + 1$ which is defective with respect to a double characteristic value zero. Although the coefficients are obtained easily by bordering A, the increased order may be a disadvantage. Tucker's method is not directly a special case of our (5).

A distantly related matrix transformation is the "escalator method" of Morris and Head [12, 11]. It relates the complete set of characteristic values and vectors of A to

*Received May 20, 1950. The preparation of this paper was sponsored (in part) by the Office of Naval Research.

the complete set for a submatrix of order $n - 1$. For this reason the escalator method cannot be compared to the transformations considered in this paper, in which at each stage one deals with only two characteristic values of A.

In §2 we present in formulas (5) a four-parameter family of transformations from A to A'. This family is general enough to include deflation and the procedures of Duncan and Collar, Semendiaev, and Blanch as special cases; see §3. In §4 two subclasses of the transformation are discussed: *order-reducing* and *symmetry-preserving* transformations. Two new methods which are both order-reducing and symmetry-preserving appear promising for practical work with symmetric matrices A. Even for non-symmetric matrices we feel that our family of transformations offers a choice of procedures which may occasionally prove useful.

2. The general reduction method. The reduction formulas will be proved in all cases by means of the following lemma, which can be easily verified.

LEMMA. *Let the matrix A have the characteristic value ν with corresponding column-vector X and row-vector Y. Let U be a non-singular matrix. Then the matrix $A' = UAU^{-1}$ has the characteristic value ν with corresponding vectors*

$$X' = UX, \qquad Y' = YU^{-1}. \tag{2}$$

We assume for simplicity that A is not defective.* Let λ be a (known) characteristic value of A, with a corresponding column-vector $C = \{c_1, \cdots, c_n\}$ and row-vector $R = (r_1, \cdots, r_n)$. Let ν be some other (unknown) characteristic value of A, with corresponding column $X = \{x_1, \cdots, x_n\}$ and row $Y = (y_1, \cdots, y_n)$. Assume C, R, X, Y to be so normalized that

$$RC = YX = 1. \tag{3}$$

The characteristic values λ and ν are permitted to be equal, provided that X, Y satisfy the orthogonality conditions

$$YC = RX = 0, \tag{4}$$

which are automatic when $\lambda \neq \nu$.

Let γ, ρ, β, t be complex parameters, and put for abbreviation

$$\Gamma = 1 - \gamma c_n, \qquad \mathrm{P} = 1 - \rho r_n, \qquad f = \Gamma\mathrm{P} - \gamma\rho.$$

We now introduce a new matrix $A' = A'(\gamma, \rho, \beta; t)$, defined for all values of the parameters except when $\beta = 0$ and at the same time $f \neq 0$:

$$\left.\begin{aligned}
a'_{ij} &= a_{ij} - \gamma c_i a_{nj} - \rho r_j a_{in} + \gamma\rho c_i r_j (a_{nn} - \lambda) - f t c_i r_j, \\
a'_{in} &= -\beta[a_{in} - \gamma c_i a_{nn} - t c_i r_n \Gamma + (\lambda - t)\gamma c_i], \\
a'_{nj} &= \begin{cases} -f\beta^{-1}[a_{nj} - \rho r_j a_{nn} - t c_n r_j \mathrm{P} + (\lambda - t)\rho r_j & (\text{if } \beta \neq 0) \\ 0 \ (\text{if } f = \beta = 0) \qquad (1 \le i \le n - 1,\ 1 \le j \le n - 1), \end{cases} \\
a'_{nn} &= f(a_{nn} - t c_n r_n) + (\lambda - t)(1 - f).
\end{aligned}\right\} \tag{5}$$

*In the terminology of [9] a *defective* matrix A is one for which no transform PAP^{-1} is a diagonal matrix. An equivalent definition is that A has one or more non-linear elementary divisors; see [2].

For all values of the parameters a matrix $U = U(\gamma, \rho, \beta)$ will be defined below, with the property that

$$A'(\gamma, \rho, \beta; t) = U(A - tCR)U^{-1}. \tag{6}$$

By the lemma A' has the characteristic values of $A - tCR$, i.e., those of A except that the single characteristic value λ is changed to $\lambda - t$. In particular, ν is a characteristic value of A', and the vectors X, Y are transformed into characteristic vectors $X' = UX$ and $Y' = YU^{-1}$ of A' corresponding to ν. By use of the formulas for U given below, it is easily shown in each case that

$$\left.\begin{aligned} x_i' &= x_i - \gamma c_i x_n \qquad (1 \le i \le n-1), \\ x_n' &= \begin{cases} -f x_n \beta^{-1} & (\text{if } \beta \ne 0) \\ 0 & (\text{if } f = \beta = 0), \end{cases} \\ y_j' &= y_j - \rho r_j y_n \qquad (1 \le j \le n-1), \qquad y_n' = -\beta y_n . \end{aligned}\right\} \tag{7}$$

The vectors C, R are transformed into characteristic vectors $C' = UC$, $R' = RU^{-1}$ of A' corresponding to $\lambda - t$. The formulas for the components of C', R' are given for each case below, as are the definitions of U, U^{-1}.

An arbitrary choice of the parameters γ, ρ, β, t may be used to analyze the matrix A. Knowing λ, C, R, one computes A' from (5). By iteration or otherwise one next determines a new characteristic value of A'. By the lemma ν is also a characteristic value of A, and the vectors of A corresponding to ν may be calculated from the relations (7); these give x_k and y_k in terms of x_k' and y_k' except when $f = 0$. (The case $\beta = 0$ is included in the exceptional case $f = 0$, as was stated before (5).) When $f = 0$, one first gets x_n, y_n from formulas (8), which are derived from (7), (4), and (3):

$$\begin{aligned} x_n &= -(\gamma + r_n \Gamma)^{-1} \sum_{i=1}^{n-1} r_i x_i' , \\ y_n &= -(\rho + c_n \mathrm{P})^{-1} \sum_{i=1}^{n-1} y_i' c_i . \end{aligned} \tag{8}$$

If still more characteristic values of A are desired, one can use ν, X', Y' in (5) to transform A' into A'', and so on. The procedure is useful to the extent that it is easier to find ν, X', Y' from A' than it is to find ν, X, Y from A.

It remains only to exhibit $U = U(\gamma, \rho, \beta)$, so that the reader may verify equation (5). For completeness C' and R' are also given. There will be three cases. All matrices are exhibited in a partitioned form, with a square matrix of $(n - 1)$-th order at the upper left.

Case 1. $\Gamma \ne 0$, $\beta \ne 0$. Here we define

$$U = U(\gamma, \rho, \beta) = \begin{pmatrix} \delta_{ij} & \vdots & -\gamma c_i \\ \cdots\cdots & \vdots & \cdots\cdots \\ -\rho r_j \beta^{-1} & \vdots & -(f + \rho r_n)\beta^{-1} \end{pmatrix},$$

where $\delta_{ij} = 0$ $(i \neq j)$ and $\delta_{ij} = 1$ $(i = j)$. It can be verified that

$$U^{-1} = \begin{pmatrix} \delta_{ij} - \gamma\rho\Gamma^{-1}c_i r_j & \vdots & -\beta\Gamma^{-1}\gamma c_i \\ \cdots\cdots & \vdots & \cdots\cdots \\ -\rho r_j \Gamma^{-1} & \vdots & -\beta\Gamma^{-1} \end{pmatrix}.$$

It is found that

$$c_i' = \Gamma c_i (1 \leq i \leq n - 1), \qquad c_n' = -\rho\beta^{-1} - fc_n\beta^{-1};$$

$$r_j' = f\Gamma^{-1}r_j(1 \leq j \leq n - 1), \qquad r_n' = -\beta\gamma\Gamma^{-1} - \beta r_n .$$

Case 2. $\Gamma = 0$. $\beta \neq 0$. Here $\gamma = c_n^{-1}$, and we define

$$U = U(c_n^{-1}, \rho, \beta) = \begin{pmatrix} \delta_{ij} + c_i r_j & \vdots & c_i c_n^{-1}(r_n c_n - 1) \\ \cdots\cdots & \vdots & \cdots\cdots \\ -(2\rho + c_n \mathrm{P})\beta^{-1} r_j & \vdots & \rho\beta^{-1}c_n^{-1} - (2\rho + c_n\mathrm{P})\beta^{-1}r_n \end{pmatrix},$$

$$U^{-1} = \begin{pmatrix} \delta_{ij} - c_i r_j + \rho c_n^{-1}(r_n c_n - 1)c_i r_j & \vdots & \beta c_i c_n^{-1}(r_n c_n - 1) \\ \cdots\cdots & \vdots & \cdots\cdots \\ -(2\rho + c_n\mathrm{P})r_j & \vdots & -\beta + \beta(r_n c_n - 1) \end{pmatrix}.$$

It is found that

$$c_i' = c_i(1 \leq i \leq n - 1), \qquad c_n' = -(\rho + c_n\mathrm{P})\beta^{-1},$$

$$r_j' = -\rho r_j c_n^{-1}(1 \leq j \leq n - 1), \qquad r_n' = -\beta c_n^{-1}.$$

Case 3. $f = \beta = 0$. Here γ, ρ are restricted to such values that $f \equiv (1 - \gamma c_n)(1 - \rho r_n) - \gamma\rho = 0$. We define

$$U = U(\gamma, \rho, 0) = \begin{pmatrix} \delta_{ij} - \Gamma c_i r_j & \vdots & -\gamma c_i - \Gamma c_i r_n \\ \cdots\cdots & \vdots & \cdots\cdots \\ r_j & \vdots & r_n \end{pmatrix},$$

$$U^{-1} = \begin{pmatrix} \delta_{ij} - \mathrm{P}c_i r_j & \vdots & c_i \\ \cdots\cdots & \vdots & \cdots \\ -\rho r_j - \mathrm{P}c_n r_j & \vdots & c_n \end{pmatrix}.$$

It is found that

$$c_i' = 0 \qquad (1 \leq i \leq n - 1), \qquad c_n' = 1,$$

$$r_j' = 0 \qquad (1 \leq j \leq n - 1), \qquad r_n' = 1.$$

3. Known special cases of our transformation. For certain values of the parameters γ, ρ, β, t the matrices A' defined by (5) have previously been used to analyze matrices A. We know of the following special cases:

(a) Duncan and Collar [3] and [4, p. 143]: Case 2, with $\gamma = c_n^{-1}$, $\rho = 0$, $\beta = -1$, $t = \lambda$. (Rows and columns have been interchanged in our presentation.) The matrix A' is the result of subtracting from each of the other rows c_n^{-1} times the matrix product of C by the last row of A.

(b) Hotelling [6] (deflation): Case 1, with $\gamma = \rho = 0$, $\beta = -1$, $t = \lambda$. Here U is the identity matrix.

(c) Semendiaev [13, p. 212]: Case 3, with $\gamma = c_n^{-1}$, $\rho = \beta = t = 0$. This is only a special case of Semendiaev's general reduction.

(d) Blanch (unpublished procedure used at the National Bureau of Standards, Los Angeles): Case 1, with $\gamma = 0$, $\rho = r_n^{-1}$, $\beta = 1$, $t = 0$. To compare method (d) with (a), to which it is closely related, one should interchange rows and columns.

4. New special cases of our transformation. One useful class of matrix transformations consists of those in §2 for which $f = 0$; here β and t remain unrestricted. For these it is seen that $a'_{nj} = 0$ $(j = 1, 2, \cdots, n-1)$, so that A' is essentially reduced to order $n - 1$. We call these transformations *order-reducing*; by their use subsequent iterations become shorter. The methods (a), (c), (d) in §3 are order-reducing.

Another special class of matrix transformations consists of those in §2 for which $\gamma = \rho$, $\beta^2 = f$, with t unrestricted. When A is symmetric it is reasonable to pick $c_i \equiv r_i$, and then A' is also symmetric; hence this class of transformations is called *symmetry-preserving*. In §3 only method (b) is symmetry-preserving.

New transformations which are both symmetry-preserving and order-reducing are those in Case 3 of §2 for which $\rho = \gamma$. Except for the unessential freedom allowed t, there are commonly two of these transformations. When $c_i \equiv r_i$ these may be defined by

$$\gamma = \rho = (c_n + 1)^{-1} = (r_n + 1)^{-1}, \tag{9}$$

$$\gamma = \rho = (c_n - 1)^{-1} = (r_n - 1)^{-1}. \tag{10}$$

Symmetric matrices are more convenient to deal with than non-symmetric ones, in that by their use the storage requirement is approximately halved and the round-off errors are more easily estimated. For dealing with symmetric matrices A, therefore, the transformations defined by (9) and (10) look promising. For non-symmetric matrices A the method (d) of §3 seems quite satisfactory, but it may occasionally be useful to have other subcases of (5) available.

5. Numerical example. From [8, p. 327] we obtain the symmetric matrix*

$$A = \begin{pmatrix} -2 & -2 & 0 & 3 & -1 \\ -2 & 0 & -3 & 5 & 0 \\ 0 & -3 & -5 & 1 & 1 \\ 3 & 5 & 1 & -3 & -1 \\ -1 & 0 & 1 & -1 & -1 \end{pmatrix}.$$

By an iteration one can obtain the dominant characteristic value $\lambda = -9.88649$ and corresponding normalized row-vector

$$R = (-.35616, -.52348, -.46374, .61437, .08124).$$

Since A is symmetric, the column-vector C has the same components.

*We have corrected a misprint in [8]. Mr. William Paine of the National Bureau of Standards, Los Angeles, assisted with the calculations.

We shall apply the reduction defined by (9) to obtain the subdominant characteristic value of A. One finds that $\gamma = \rho = (1.08124)^{-1} = .924864$. Putting these values of ρ, γ into (5) with $t = 0, \beta = 0$, one finds that

$$a'_{ij} = a_{ij} - .924864(c_i a_{nj} + r_j a_{in}) + 7.60127 c_i r_j, \qquad a'_{in} = a'_{ni} = 0$$

$$(1 \leq i \leq n - 1, 1 \leq j \leq n - 1),$$

$$a'_{nn} = -9.88649.$$

Hence

$$A' = \begin{pmatrix} -1.69458 & -1.06695 & 1.15597 & 1.57555 & 0 \\ -1.06695 & 2.08298 & -.67057 & 2.07121 & 0 \\ 1.15597 & -.67057 & -2.50753 & -2.16277 & 0 \\ 1.57555 & 2.07121 & -2.16277 & 1.00552 & 0 \\ 0 & 0 & 0 & 0 & -9.88649 \end{pmatrix}.$$

By a separate iteration of the non-trivial fourth-order minor of A' one can determine that $\nu = -4.75772$ is a characteristic value of A', with corresponding row-vector

$$Y' = (-.54888, -.15891, .67842, .46175, 0).$$

By the lemma of §2, -4.75772 is a characteristic value of A, and it remains only to determine the corresponding row-vector Y. By (8),

$$y_n = -\sum_{i=1}^{n-1} y'_i c_i = -.24775. \tag{11}$$

From (11) and (7) one can then compute the first four components of Y. One finds that

$$Y = (-.46727, -.03896, .78468, .32098, -.24775).$$

If more characteristic values and vectors of A were desired, one would start by using ν and the first four components of Y' to reduce the first four rows and columns of A' to a matrix A'' of order four which would be bordered with zeros.

Bibliography

[1] Aitken, A. C., *Studies in practical mathematics II. The evaluation of the latent roots and latent vectors of a matrix*, Proc. Roy. Soc. Edinburgh **57**, 269-304 (1937).
[2] Bôcher, M., *Introduction to Higher Algebra*, New York, 1907, Chapter XX.
[3] Duncan, W. J. and Collar, A. R., *A method for the solution of oscillation problems by matrices*, Phil. Mag. [7], **17**, 865-909 (1934).
[4] Frazer, R. A., Duncan, W. J., and Collar, A. R., *Elementary Matrices*, Cambridge, England, 1938.
[5] Hestenes, M. R. and Karush, W., *A method of gradients for the calculation of the characteristic roots and vectors of a real symmetric matrix*, multilithed typescript at the National Bureau of Standards, Los Angeles, 1950.
[6] Hotelling, Harold, *Analysis of a complex of statistical variables into principal components*, J. Educational Psychology **24**, 417-441, 498-520 (1933).

[7] Hotelling, Harold, *Simplified calculation of principal components*, Psychometrika **1**, 27-35 (1936).
[8] Kincaid, W. M., *Numerical methods for finding characteristic roots and vectors of matrices*, Q. Appl. Math. **5**, 320-345 (1947).
[9] Lanczos, Cornelius, *An iteration method for the solution of the eigen-value problem of linear differential and integral operators*, multilithed typescript at the National Bureau of Standards, Los Angeles, 1949.
[10] Morris, J., *Frequency equations*, Aircraft Engineering **14**, 108-110 (1942).
[11] Morris, Joseph, *The Escalator Method*, New York, 1947, Chap. 10. especially pp. 115-118.
[12] Morris, J. and Head, J. W., *Lagrangian frequency equations*, Aircraft Engineering **14**, 312-316 (1942). An appendix by G. Temple summarizes the escalator method in matrix notation.
[13] Semendiaev, K. A., *The determination of latent roots and invariant manifolds of matrices by means of iterations*, Prikladnaya matematika i mehanika (Leningrad) **7**, 193-222 (1943). Russian, with brief English summary.
[14] Tucker, Ledyard R., *Determination of the successive principal components without computation of tables of residual correlation coefficients*, Psychometrika **9**, 149-153 (1944).
[15] von Mises, R., and Pollaczek-Geiringer, H., *Praktische Verfahren der Gleichungsaufloesung*, ZS. angew. Math. Mech. **9**, 152-164 (1929).
[16] Wielandt, Helmut, *Das Iterationsverfahren bei nichtselbstadjungierten linearen Eigenwertaufgaben*, Math. ZS. **50**, 93-143 (1944), p. 96.