PSYCHOMETRIKA—VOL. 28, NO. 3
SEPTEMBER, 1963

ON A THEOREM STATED BY ECKART AND YOUNG

RICHARD M. JOHNSON

THE PROCTER & GAMBLE COMPANY

Proof is given for a theorem stated but not proved by Eckart and Young in 1936, which has assumed considerable importance in the theory of lower-rank approximations to matrices, particularly in factor analysis.

In a classic article in the first volume of *Psychometrika* [2] Eckart and Young stated a theorem giving a canonical form to which any real matrix can be reduced. That theorem underlay their solution to the problem of approximating one matrix by another of lower rank. During subsequent years the problem of lower-rank approximation has been one of the principal concerns of factor analysis. Consequently the Eckart-Young solution, and with it the theorem upon which it rests, have become cornerstones of several important multivariate techniques in psychometrics. The canonical form of the theorem has been called the "basic structure" of a matrix by Horst, who has discussed both its theoretical and practical aspects [3].

Curiously, Eckart and Young did not prove the theorem, instead directing their readers to two references. The present writer has not been able to find proof of the theorem in the first of these [1]. In the other reference MacDuffee [5] provides a particularly elegant and simple proof, but only for the special case where the matrix being reduced is square and nonsingular. A more complicated discussion of the square case is given by Sylvester [6]. Recently Keller [4] has rederived the least squares solution, but his paper contains no reference to the Eckart-Young contribution or to the original theorem. The following proof of the Eckart-Young theorem is therefore presented.

The Eckart-Young theorem is as follows:

> For any real matrix A, two orthogonal matrices P and Q can be found for which $P'AQ$ is a real diagonal matrix with no negative elements.

A diagonal matrix Λ (square or rectangular) was defined by Eckart and Young to be one for which $\lambda_{ij} = 0$ unless $i = j$. If such a matrix is rectangular, then there will be some rows or columns which consist entirely of zeros.

In the proof use is made of the fact that any real symmetric matrix can be transformed into a real (square) diagonal matrix through postmultiplication by an orthogonal matrix and premultiplication by the transpose of that orthogonal matrix. It is well known that the postmultiplier has as its columns the characteristic vectors of the symmetric matrix and the resulting diagonal matrix has as its diagonal elements the characteristic roots of the symmetric matrix.

Proof

Let A be of order $m \times n$. Without loss of generality, we may take $m \geq n$. The symmetric product AA' has the same rank as A, and consequently has at most n nonzero characteristic roots. Choosing P as an orthogonal matrix of order $m \times m$ with characteristic vectors of AA' in its columns, and Λ as an $n \times n$ diagonal matrix with all of the nonzero characteristic roots of AA' among its diagonal elements, we can write

$$P'AA'P = \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix}. \tag{1}$$

We also express the product

$$P'A = \begin{bmatrix} D \\ E \end{bmatrix}, \tag{2}$$

in which D is an unknown square matrix of order $n \times n$, and E, of order $(m - n) \times n$, is also unknown. The matrix D is augmented with E to allow the case where A and consequently $P'A$ are not square. For the special case where A is square E may be omitted from (2) and the rows and columns of zeros may be omitted from the right-hand side of (1).

Postmultiplying both sides of (2) by their transposes, we get

$$P'AA'P = \begin{bmatrix} D \\ E \end{bmatrix} [D' \quad E'] = \begin{bmatrix} DD' & DE' \\ ED' & EE' \end{bmatrix}. \tag{3}$$

The left-hand sides of (1) and (3) are then identical, giving

$$\begin{bmatrix} DD' & DE' \\ ED' & EE' \end{bmatrix} = \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} \tag{4}$$

and, in particular,

$$DD' = \Lambda \tag{5}$$

and

$$EE' = 0. \tag{6}$$

If all the elements of EE' are zero, so are all the elements of E, since the sum of diagonal elements of EE' is equal to the sum of squares of elements of E, which are real numbers. Therefore,

$$E = 0. \tag{7}$$

Since Λ is a diagonal matrix, (5) indicates that the rows of D are orthogonal to one another and that the sum of squared elements in each row is equal to the corresponding diagonal element in Λ. This means that the diagonal elements of Λ are nonnegative and also that some (unknown) orthogonal matrix Q' exists such that when the elements in each row are multiplied by the square root of the corresponding diagonal element in Λ, the resulting matrix is equal to D. Symbolically,

$$\Lambda^{1/2}Q' = D. \tag{8}$$

Substituting from (7) and (8) into (2), we get

$$P'A = \begin{bmatrix} \Lambda^{1/2}Q' \\ 0 \end{bmatrix}. \tag{9}$$

Equation (9) may be rewritten

$$P'A = \begin{bmatrix} \Lambda^{1/2} \\ 0 \end{bmatrix} Q', \tag{10}$$

which, upon postmultiplication by Q, yields (since Q' is orthogonal)

$$P'AQ = \begin{bmatrix} \Lambda^{1/2} \\ 0 \end{bmatrix}. \tag{11}$$

This shows that any real matrix A may be transformed into a real diagonal matrix upon pre- and postmultiplication by orthogonal matrices. When A is square the extra rows of zeros on the right-hand side of (11) will be absent. Finally, note that the elements of $\Lambda^{1/2}$ may always be taken as nonnegative; if any of these were negative the signs of all elements in those rows of P' or those columns of Q could be reversed, causing the negative values of $\Lambda^{1/2}$ to be made positive. Such changes would not destroy the orthogonality of either P or Q. This completes proof of the theorem.

In order to identify more clearly the matrices P, Q, and Λ, premultiply both sides of (11) by their transposes, We get

$$Q'A'PP'AQ = [\Lambda^{1/2}\ 0'] \begin{bmatrix} \Lambda^{1/2} \\ 0 \end{bmatrix} = \Lambda, \tag{12}$$

which, upon substitution of $PP' = I$, yields

$$Q'A'AQ = \Lambda. \tag{13}$$

The columns of P were specified initially to contain characteristic vectors of AA' and the diagonal positions of Λ to contain nonzero characteristic roots of AA'. Equation (13) shows that the columns of Q contain characteristic vectors of $A'A$ and that the characteristic roots of $A'A$ are also contained in the diagonal positions of Λ. As an interesting by-product, this shows that the characteristic roots of $A'A$ and AA' are identical except that, when A is of order $m \times n$ with $m > n$, AA' has $m - n$ additional characteristic roots of zero.

The important application of the theorem is derived from the fact that when(11) is premultiplied by P and postmultiplied by Q', we have

$$(14) \qquad A = P \begin{bmatrix} \Lambda^{1/2} \\ 0 \end{bmatrix} Q'.$$

It is convenient to order the diagonal elements of $\Lambda^{1/2}$ in descending order of magnitude and to arrange the columns of P and the rows of Q' to correspond. We then define

$\Lambda_r^{1/2}$ as a diagonal submatrix containing the first r diagonal elements of $\Lambda^{1/2}$,

P_r as a submatrix containing the r corresponding columns of P,

Q_r' as a submatrix containing the r corresponding rows of Q'.

In the product indicated by (14), if only r diagonal elements of $\Lambda^{1/2}$ happen to be nonzero, then the last $m - r$ columns of P and the last $n - r$ rows of Q' are multiplied by zero and consequently are of no importance. In that case,

$$(15) \qquad A = P_r \Lambda_r^{1/2} Q_r' .$$

If the rank of A is greater than r, then Eckart and Young [2] have shown that the product on the right-hand side of (15) furnishes the least squares approximation to A of rank r. Keller [4] has shown that the least squares approximation is given by the product AQ_rQ_r' , which can easily be seen to be the same thing.

To show this, denote the remaining columns of P as P_s , the remaining rows of Q' as Q_s' , and the diagonal matrix containing the remaining diagonal elements of Λ as Λ_s . Then

$$(16) \qquad A = [P_r \; P_s] \begin{bmatrix} \Lambda_r^{1/2} & 0 \\ 0 & \Lambda_s^{1/2} \end{bmatrix} \begin{bmatrix} Q_r' \\ Q_s' \end{bmatrix},$$

which can be written as

$$(17) \qquad A = P_r \Lambda_r^{1/2} Q_r' + P_s \Lambda_s^{1/2} Q_s' .$$

Since $Q_s'Q_r = 0$, postmultiplication of (17) by Q_rQ_r' gives

$$(18) \qquad AQ_rQ_r' = P_r \Lambda_r^{1/2} Q_r'Q_rQ_r' = P_r \Lambda_r^{1/2} Q_r' .$$

To determine the sum of squared errors of approximation, define the "error" matrix

$$F = A - P_r \Lambda_r^{1/2} Q_r' \,. \tag{19}$$

Substituting from (17) into (19) we get

$$F = P_s \Lambda_s^{1/2} Q_s' \,. \tag{20}$$

The sum of squared elements of F is given by the trace of the product $F'F$,

$$\mathrm{tr}\,(F'F) = \mathrm{tr}\,(Q_s \Lambda_s^{1/2} P_s' P_s \Lambda_s^{1/2} Q_s'). \tag{21}$$

Since $P_s'P_s = Q_s'Q_s = I$, we get upon simplification and permutation under the trace sign

$$\mathrm{tr}\,(F'F) = \mathrm{tr}\,(Q_s \Lambda_s Q_s') = \mathrm{tr}\,(\Lambda_s Q_s' Q_s) = \mathrm{tr}\,(\Lambda_s). \tag{22}$$

The sum of squared errors of approximation is therefore the trace of the matrix Λ_s , which is simply the sum of those characteristic roots of $A'A$ and AA' which are not used in the approximation. If A is of rank r, then all but the first r roots of $A'A$ and AA' are zero, and the "approximation" is exact.

REFERENCES

[1] Courant, R. and Hilbert, D. *Methoden der Mathematischen Physik.* Berlin: Julius Springer, 1924.

[2] Eckart, C. and Young, G. The approximation of one matrix by another of lower rank. *Psychometrika,* 1936, **1**, 211–218.

[3] Horst, P. *Matrix algebra for social scientists.* New York: Holt, Rinehart and Winston, Inc., 1963.

[4] Keller, J. B. Factorization of matrices by least squares. *Biometrika,* 1962, **49**, 239–242.

[5] MacDuffee, C. C. *The theory of matrices.* (Corrected reprint of first edition.) New York: Chelsea Publishing Co., 1946.

[6] Sylvester, J. J. On the reduction of a bilinear quantic of the nth order to the form of a sum of n products by a double orthogonal substitution. *Messenger of Mathematics,* 1890, **19**, 42–46.

Manuscript received 1/22/63
Revised manuscript received 4/2/63