PSYCHOMETRIKA—VOL. 1, NO. 3
SEPTEMBER, 1936

THE APPROXIMATION OF ONE MATRIX BY ANOTHER OF LOWER RANK

CARL ECKART AND GALE YOUNG
University of Chicago, Chicago, Illinois

The mathematical problem of approximating one matrix by another of lower rank is closely related to the fundamental postulate of factor-theory. When formulated as a least-squares problem, the normal equations cannot be immediately written down, since the elements of the approximate matrix are not independent of one another. The solution of the problem is simplified by first expressing the matrices in a canonic form. It is found that the problem always has a solution which is usually unique. Several conclusions can be drawn from the form of this solution.

A hypothetical interpretation of the canonic components of a score matrix is discussed.

Introduction

If N individuals are each subjected to n tests, it is a fundamental postulate of factor theory that the resulting $n \times N$ score matrix a can be adequately approximated by another matrix β whose rank r is less than the smaller of n or N. Closely associated to this postulate is the purely mathematical problem of finding that matrix β of rank r which most closely approximates a given matrix a of higher rank R. It will be shown that if the least-squares criterion of approximation be adopted, this problem has a general solution which is relatively simple in a theoretical sense, though the amount of numerical work involved in applications may be prohibitive. Certain conclusions can be drawn from the theoretical solution which may be of importance in practical work.

To formulate the problem precisely, it is convenient to define the "scalar product" of two $n \times N$ matrices as the following numerical function of their elements:

$$(a, \beta) = \sum_{i=1}^{n} \sum_{j=1}^{N} a_{ij} \beta_{ij} \tag{1}$$

where a_{ij} is the ij-element of the matrix a. This function has all the properties of the scalar product of vectors:

$$(a, \beta) = (\beta, a) \quad ; \tag{2}$$

$$(\alpha \pm \beta, \gamma) = (\alpha, \gamma) \pm (\beta, \gamma) \quad ; \tag{3}$$

$$(x\alpha, \beta) = x(\alpha, \beta) \quad \text{if } x \text{ is a number;} \tag{4}$$

$$(\alpha, \alpha) > 0 \quad \text{if } \alpha \neq 0 \ . \tag{5}$$

The positive number l, defined by $l^2 = (\alpha, \alpha)$ may be called the length of the matrix α; the number l^2 was called the span (Spannung) of α by Frobenius.* The length of the matrix $\alpha - \beta$ may be called the distance from α to β.

The problem can now be formulated in a definite manner: to find that matrix β of rank r, such that there is no other matrix of rank r whose distance from α is less than the distance from β to α. This amounts to requiring a least-squares solution of the approximation problem, every element of the given matrix being given equal weight.

Some preliminary theorems and remarks.

To simplify the following discussion, let $a, b, u, \cdots$ denote $n \times n$ matrices, $A, B, U, \cdots$ denote $N \times N$ matrices, and $\alpha, \beta, \cdots$ denote $n \times N$ matrices; the special case $n = N$ is not excluded. The $N \times n$ matrix which is obtained by writing the columns of α as rows is the transpose of α and will be denoted by α'. The products $a\alpha$, αA, $\alpha B'$, $\alpha' a, \cdots$ are defined in the usual way. It is then seen that the following equations are correct:

$$(\alpha, \beta) = (\alpha', \beta') \quad ; \tag{6}$$

$$(a\alpha, \beta) = (a, \alpha'\beta) = (a, \beta\alpha') \quad ; \tag{7}$$

$$(\alpha A, \beta) = (\alpha, \beta A') = (A, \alpha'\beta) \quad . \tag{8}$$

From Eq. (7) and (8) it follows that if u and U are orthogonal mattrices ($u\,u' = u'\,u = 1_n$, $U\,U' = U'\,U = 1_N$), then

$$(u\,\alpha\,U', u\,\beta\,U') = (\alpha, \beta) \ . \tag{9}$$

Another useful proposition is the following: if $(a, b) = 0$ for all symmetric (skew-symmetric) matrices, b, then a is skew-symmetric (symmetric).

The solution of the problem is much simplified by an appeal to two theorems which are generalizations of well-known theorems on square matrices.† They will not be proven here.

*Quoted from MacDuffee, "Theory of Matrices", Ergebn. d. Mathem., v. 2, No. 5 p. 80 (1933).

†Courant and Hilbert, "Methoden der mathematischen Physik" Berlin, 1924; pp. 9 et seq., p. 25. MacDuffee, p. 78.

Theorem I. For any real matrix a, two orthogonal matrices u and U can be found so that $\lambda = uaU'$ is a real diagonal matrix with no negative elements.

A diagonal matrix λ (square or rectangular) is one for which $\lambda_{ij} = 0$ unless $i = j$. If a diagonal matrix is rectangular, then there will be some rows or columns which consist entirely of zeros. For the following, this remark is of some importance, as will be seen. The equation of the theorem may also be written

$$a = u' \lambda U \tag{10}$$

whose right side may be called the canonic resolution of a. If $n < N$, λ will have $N—n$ columns of zeros and a is seen to depend only on the first n rows of U. If u, λ and the first n rows of U are given, a is determined.

Let v be the diagonal $n \times n$ matrix which consists of the first n columns of λ, and ω the $n \times N$ matrix composed of the first n rows of U; then these remarks can be summarized by the equation

$$a = u' v \omega \tag{10.1}$$

where $\omega \omega' = 1_n$, but $\omega' \omega \neq 1_N$. For numerical work, Eq. (10.1) is preferable to Eq. (10); for formal manipulation, Eq. (10) is more convenient.

The numerical evaluation of u and v (or λ) can be accomplished from the consideration of the matrix $a = a\, a'$ alone. This matrix is closely related to the matrix of correlation coefficients of the tests. It is seen that

$$a = a\, a' = u' \lambda \lambda' u = u' v^2 u \ ,$$

and since $\lambda \lambda' = v^2$ is a diagonal matrix, it follows that u is one of the orthogonal matrices which transform the correlational matrix to diagonal form. The rows of u are unit vectors along the principal axes of a and the squares of the diagonal elements of v (or λ) are the characteristic values of a; this shows that the latter can never be negative numbers, a result which can also be obtained more directly.* The methods for determining the principal axes and characteristic values of a symmetric matrix are also known,† so that these remarks may be considered as indicating the method for calculating u and λ. If none of the characteristic values of a is zero (this will presumably be the case in the overwhelming proportion of actual calculations)

*Courant-Hilbert, p. 20.
†Courant-Hilbert, pp. 13, 16.

the matrix v will have a reciprocal, and ω can be obtained by solving Eq. (10.1):

$$\omega = v^{-1} u \alpha \ .$$

The numerical values of the elements in the remaining rows of the matrix U will not be needed, but could be found if necessary. For simplicity of manipulation, it is convenient to proceed as though this has been done.

The diagonal elements of λ were called the "canonical multipliers" of α by Sylvester.* The multipliers and characteristic values of a correlational matrix are identical; in the case of a symmetric matrix, there may be a difference in sign; for a general square matrix, there is no simple relation between the two; for a rectangular matrix, the characteristic values are not defined.

The correlational matrices, Sylvester called the "false squares" of α. In the foregoing (and in the usual treatment of factor theory) only the matrix $a = \alpha \alpha'$ has been considered. However, the matrix $A = \alpha' \alpha$ is related to the correlation coefficients of the individuals in the same manner as a is related to the correlation coefficients of the tests. There is complete mathematical symmetry between the two correlation matrices.

To every multiplier, there is associated a row of u and a row of U; this complex of $n + N + 1$ numbers may be called a canonic component of α.

Theorem II. If $\alpha \beta'$ and $\beta' \alpha$ are both symmetric matrices, then and only then can two orthogonal matrices u and U be found such that $\lambda = u \alpha U'$ and $\mu = u \beta U'$ are both real diagonal matrices.

Either one (but in general, not both) of the diagonal matrices may be further restricted to have no negative elements. This theorem is a generalization of the theorem that the principal axes of two symmetric matrices coincide if and only if $ab = ba$.

Solution of the problem

The distance of β from α is given by x, where

$$x^2 = (\alpha, \alpha) - 2(\alpha, \beta) + (\beta, \beta) \quad ; \qquad (11)$$

x is a function of all the elements of β, and these are to be determined so that its value is a minimum. The elements of β are not all independent, however, because of the requirement that its rank be less than the number of its rows or columns. Theorem I makes it possible

*Messeng. Math., *19* p. 45 (1889).

to eliminate some of the interdependence: suppose β to have been resolved into canonic form:

$$\beta = u' \mu U \tag{12}$$

with μ diagonal, and u and U orthogonal matrices. Then the rank of β will be r if and only if μ has this rank i.e., if just r of the diagonal elements of μ are different from zero; the non-vanishing elements of μ will be independent. However, the elements of u or U will not be independent, since these matrices must be orthogonal. It is not necessary to express these matrices in terms of independent parameters because of the following proposition:* if u is any orthogonal matrix and the independent variables that determine it are given any infinitesimal increments, the resulting increment of u is

$$\delta u = u s \ , \tag{13}$$

where s is a skew-symmetric matrix whose elements are infinitesimal, but otherwise arbitrary.

The Eq. (11) becomes, because of Eq. (12) and (9),

$$x^2 = (\alpha, \alpha) - 2(\alpha, u' \mu U) + (\mu, \mu) \ . \tag{14}$$

Since x is to be a minimum, it follows that $\delta x^2 = 0$ when u is given the increment δu (Eq. (13)).
Hence

$$0 = (\alpha, -s u' \mu U) = -(\alpha, s \beta) = -(\alpha \beta', s) \ . \tag{15}$$

Since s is an arbitrary skew-symmetric matrix, it follows that $\alpha \beta'$ must be symmetric. Discussing the increment of U in the same manner, it will be found that $\beta' \alpha$ must also be symmetric, and hence, by Theorem II, the orthogonal matrices can be found so that Eq. (12) and

$$\alpha = u' \lambda U \tag{12.1}$$

(with λ the diagonal matrix of the multipliers of α) are both valid. Then Eq. (11) becomes

$$\begin{aligned} x^2 &= (\lambda - \mu, \lambda - \mu) \\ &= \textstyle\sum_i (\lambda_i - \mu_i)^2 \end{aligned} \tag{11.1}$$

λ_i and μ_i being the diagonal elements of the corresponding matrices.

It remains to determine the matrix μ so that this expression has its minimum value, subject to the condition that just r of the μ_i shall

*Courant-Hilbert, p. 27.

be different from zero. It may be supposed that $\lambda_1 \geqq \lambda_2 \geqq \lambda_3 \geqq \cdots$; then an obvious solution of the problem is

$$\mu_i = \lambda_i \ , \quad i \leqq r \ ;$$
$$= 0 \ , \quad i > r \ . \tag{17}$$

This is also the only solution unless $\lambda_r = \lambda_{r+1}$.

The procedure of finding β may be summarized: (1.) express a in the canonic form of Eq. (12.1) ; (2.) replace all but r of the diagonal elements of λ by zeros, beginning with the smallest and continuing in order of increasing magnitude. (3.) The resulting matrix is μ, and β is given by Eq. (12). The solution is unique unless $\lambda_r = \lambda_{r+1}$.

The minimum value of x^2 is

$$x^2 = \sum_{i=r+1}^{R} \lambda_i^2$$

(R being the rank of a and the multipliers being numbered as before). This leads readily to the following conclusion: if l is the length of a, the smallest upper bound for x is $l(1-r/R)^{1/2}$. This can be utilized in estimating the significance of an approximation to a obtained by some other method than the present one.

Another important conclusion is that if β is the best approximation (of rank r) to a, then $b = \beta\beta'$ ($B = \beta'\beta$) will also be the best approximation (of rank r) to $a = a\alpha'$ ($A = \alpha' a$). Thus, if the best approximation to the score matrix is found, then the correlational matrices calculated from it will automatically be the best approximations to the correlational matrices calculated from the original score matrix. This is not surprising, but requires proof; the proof is readily supplied from the foregoing results.

Concerning a hypothetical interpretation of the canonic components of β

It is reasonable to inquire if, when a is a score matrix, those of its canonic components that enter into β may be interpreted as the independent factors that determine the differences in the performance of various individuals. In order to discuss this question let $p = 1, \cdots n$ number the tests, $i = 1, \cdots N$ number the individuals, and ϱ or $\sigma = 1, \cdots r$ number the components of β that have non-vanishing multipliers. Then the component ϱ consists of the $n + N + 1$ numbers

$$u_{\rho p} \ , \quad \lambda_\rho > 0 \ , \quad U_{\rho i}$$

which satisfy the equations

$$\textstyle\sum_p u_{\rho p}\, u_{\sigma p} = \delta_{\rho\sigma} \quad ; \quad \sum_i U_{\rho i}\, U_{\sigma i} = \delta_{\rho\sigma} \tag{18}$$

$$\beta_{pi} = \textstyle\sum_\rho u_{\rho p}\, \lambda_\rho\, U_{\rho i} \ . \tag{19}$$

These equations suggest that if these numbers have any empirical significance, then $N^{1/2}\, U_{\rho i}$ will be the standard score of individual i for the ability to exercise the factor ϱ, and $n^{1/2}\, u_{\rho p}$ the standard score of the test p for its demand for the exercise of the factor. The appearance of the multipliers in Eq. (19) does not correspond to the accepted postulates of factor theory, and their interpretation is thus not immediate.

To clarify this point, one may inquire after those empirical circumstances, which if they are actually realized, will lead one to accept the foregoing interpretation. One such set of circumstances would be the following: suppose that two sets of individuals, $N_{\rm I}$ and $N_{\rm II}$ in number, have been tested by the same battery of n tests. The scores of the $N_{\rm I}$ individuals can then be arranged in an $n \times N_{\rm I}$ matrix $a_{\rm I}$, those of the $N_{\rm II}$ individuals into an $n \times N_{\rm II}$ matrix $a_{\rm II}$; or, the scores of all $N_{\rm I} + N_{\rm II}$ individuals can be arranged into an $n \times (N_{\rm I} + N_{\rm II})$ matrix a obtained by adjoining $a_{\rm I}$ and $a_{\rm II}$. First consider $a_{\rm I}$ and $a_{\rm II}$; if these can be approximated by matrices $\beta_{\rm I}$ and $\beta_{\rm II}$ having the same u-matrices* (within reasonable limits of error), then the foregoing interpretation would be appropriate. In this case the u-matrix obtained by approximating a would also be the same as the others. The multipliers $\lambda_{\rho,{\rm I}}$, $\lambda_{\rho,{\rm II}}$ and λ_ρ, and of course also the $U_{\rho i,{\rm I}}$, $U_{\rho i,{\rm II}}$ and $U_{\rho i}$ obtained from the three score-matrices would not be identical. In order to discuss their interrelations, let the index i now run from 1 to $N_{\rm I} + N_{\rm II}$, the first $N_{\rm I}$ values designating individuals of the first set, etc. Then it can be shown that

$$\lambda_\rho^2 = \lambda_{\rho,{\rm I}}^2 + \lambda_{\rho,{\rm II}}^2 \tag{20}$$

and

$$\begin{aligned} \lambda_\rho\, U_{\rho i} &= \lambda_{\rho,{\rm I}}\, U_{\rho i,{\rm I}}\,, \ \text{if } i \leqq N_{\rm I} \ ; \\ &= \lambda_{\rho,{\rm II}}\, U_{\rho i,{\rm II}}\,, \ \text{if } i > N_{\rm I} \ . \end{aligned} \tag{21}$$

These are precisely the relations that would be expected according to the suggested interpretation if $\lambda_{\rho,{\rm I}}^2/N_{\rm I}$ and $\lambda_{\rho,{\rm II}}^2/N_{\rm II}$ were the variances of the two sets of individuals. This is not a tenable interpretation, however, because of the symmetric manner in which the

*The sufficient condition for this equality is that $a_{\rm I}\, a_{\rm II}$ be a symmetric matrix, where $a_{\rm I} = \alpha_{\rm I}\alpha_{\rm I}'$, etc. This criterion can be applied even when $u_{\rm I}$ and $u_{\rm II}$ are not known.

individuals and the tests enter into the calculations. It may, however, be supposed that $\lambda_{\rho,\mathrm{I}}^2/nN_\mathrm{I}$ is the product of the variance of the set of individuals with the variance of the set of tests. It should be noted that it has been implicitly assumed that the scores are not measured from the averages of the different sets of individuals, but from a fiducial zero which is the same for all sets.