

A GENERALIZED SOLUTION OF THE ORTHOGONAL PROCRUSTES PROBLEM*

PETER H. SCHÖNEMANN

PSYCHOMETRIC LABORATORY
UNIVERSITY OF NORTH CAROLINA †

A solution T of the least-squares problem $AT = B + E$, given A and B so that $\text{trace}(E'E) = \text{minimum}$ and $T'T = I$ is presented. It is compared with a less general solution of the same problem which was given by Green [5]. The present solution, in contrast to Green's, is applicable to matrices A and B which are of less than full column rank. Some technical suggestions for the numerical computation of T and an illustrative example are given.

1. Definition of the Problem and Solution

The least-squares problem of transforming a given matrix A into a given matrix B by an orthogonal transformation matrix T so that the sums of squares of the residual matrix $E = AT - B$ is a minimum will be called an "Orthogonal Procrustes problem" [8].

Mathematically this problem can be stated as follows.

$$(1.1) \quad AT = B + E,$$

$$(1.2) \quad TT' = T'T = I,$$

$$(1.3) \quad \text{tr}(E'E) = \text{min},$$

where the matrices A and E are both $n \times m$ and over the reals, but otherwise unrestricted, and both are assumed to be "known." In practical work A will usually be an observed matrix of an earlier study. Equation (1.1) states the model, (1.2) the side condition, (1.3) the criterion.

The latter can be written

$$(1.4) \quad g_1 = \text{tr}(E'E) = \text{tr}(T'A'AT - 2T'A'B + B'B).$$

*This paper is based on parts of a thesis submitted to the Graduate College of the University of Illinois in partial fulfillment of the requirements for a Ph.D. degree in Psychology.

†Now at the Ohio State University. The work reported here was carried out while the author was employed by the Statistical Service Unit Research, U. of Illinois. It is a pleasure to express my appreciation to Prof. K. W. Dickman, director of this unit, for his continuous support and encouragement in this and other work. I also gratefully acknowledge my debt to Prof. L. Humphreys for suggesting the problem and to Prof. L. R. Tucker, who derived (1.7) and (1.8) in summation notation, suggested an iterative solution (not reported here) and who provided generous help and direction at all stages of the project.

As a side condition $T'T = I$ will be chosen. The side condition

$$(1.5) \quad T'T = I$$

will be written

$$(1.6) \quad g_2 = \text{tr} [L(T'T - I)],$$

where the $(m \times m)$ matrix L is a matrix of (unknown) Lagrange multipliers. The function g to be differentiated partially with respect to the elements of the matrix T is then

$$(1.7) \quad g = g_1 + g_2.$$

Partial differentiation of g with respect to (the elements of) T [2] leads to the matrix of partial derivatives

$$(1.8) \quad \partial g / \partial T = (A'A + A'A)T - 2A'B + T(L + L')$$

which need be set to zero for an extremum of g_1 . Hence one has to solve

$$(1.9) \quad S = PT + TQ$$

where $A'A = P$, $A'B = S$, and $(L + L')2 = Q$ for convenience. (Equation (1.9) may be worthy of more general interest. For example, one observes that it includes the classic eigenproblem as a special case for $Q = \text{diagonal}$ and $S = \text{null}$.) In the present context one notes that both P and Q are symmetric so that

$$(1.10) \quad Q = T'S - T'PT = Q'.$$

But since $T'PT$ is symmetric if P is, $T'S$ must be symmetric or

$$(1.11) \quad T'S = S'T.$$

From (1.2) and (1.11) one finds $S = TS'T$, so that

$$(1.12) \quad SS' = TS'ST'.$$

One now deals with two known symmetric matrices $S'S$ and SS' , both of which must be diagonalizable by orthonormal matrices and both of which are known to have the same latent roots [13]. Now let

$$(1.13) \quad SS' = WD_*W' \quad \text{and} \quad S'S = VD_*V',$$

with

$$(1.14) \quad I = W'W = WW' = V'V = VV',$$

so that W and V together with D_* , the diagonal matrix of latent roots of SS' or $S'S$, give the canonical decomposition of both matrices. Then, from (1.12)

$$(1.15) \quad WD_*W' = TVD_*V'T'$$

so that

$$(1.16) \quad W = TV$$

or

$$(1.17) \quad T = WV'$$

will satisfy (1.15).

It is interesting to note that Gibson [4] and Johnson [10] obtain similar expressions as solutions of seemingly different least-squares problems.

Gibson [4] wishes to approximate a given transformation matrix B by an orthogonal matrix T in a least-squares sense. His solution is $T = WV'$, where W and V are the matrices of latent vectors corresponding to the Eckart-Young decomposition of B . This solution follows from (1.1) by simply setting $A = I$ so that

$$(1.18) \quad T = B + E$$

$$(1.19) \quad S = A'B = B = WD_s^{1/2}V'$$

replace (1.1) and (1.9) of the general case.

The square case of Johnson's problem can be treated as a further specialization of Gibson's. Johnson [10] solves (1.18) above, but not for T directly. Instead, the problem is to find a matrix Q which maps B into an orthogonal matrix T where

$$(1.20) \quad T = BQ,$$

and Q is square. From Gibson's problem one knows that an orthogonal T which minimizes $\text{tr}(E'E)$ in (1.18) is given by $T = WV'$ as in (1.17) and since B must be assumed to have full column rank (to admit $T'T = I$) one may simply solve (1.20) for the unknown matrix Q .

$$(1.21) \quad \begin{aligned} Q &= (B'B)^{-1}B'T = (VD_s^{-1}V')(VD_s^{1/2}W')(WV') \\ &= VD_s^{-1/2}V', \end{aligned}$$

which is indeed the result Johnson obtains by a different route.

2. Sufficiency and Uniqueness of T^*

Since (1.9) states a necessary condition on T for $\text{tr}(E'E)$ to be a minimum, and (1.17) is an algebraic consequence of (1.9), it also states a necessary condition for T , now explicitly. But this condition is also necessary for $\text{tr}(E'E)$ to be a maximum and since the orientation of the eigenvectors in V and W is arbitrary (even if all roots are distinct) one has to select a particular T

*After completion of an earlier draft of this paper it has come to my attention that some of the algebra in this section is treated in Ch. VII of P. Horst, *Matrix algebra for social scientists*. New York: Holt, Rinehart and Winston, 1963.

which is not only necessary, but also sufficient for a minimum. To find such a T assume for a moment that all roots of $S'S$ are distinct.

From (1.4) and (1.17) one has

$$(2.1) \quad \begin{aligned} g_1 &= \text{tr}(E'E) = \text{tr}(T'PT - 2T'S + B'B) \\ &= \text{tr}(P + B'B) - 2 \text{tr}(T'S), \end{aligned}$$

since $\text{tr}(T'PT) = \text{tr}(P)$, as T is orthogonal. Hence it follows that

$$(2.2) \quad \theta = \text{tr}(T'S) = \text{tr}(VW'S)$$

has to be a maximum if g_1 is to be a minimum. But

$$(2.3) \quad \theta = \text{tr}(T'S) = \text{tr}(VW'WD_s^{1/2}V')V = \text{tr}(WW'D_s^{1/2}V'V)$$

by cyclic permutation which leaves the trace unchanged, so that, finally

$$(2.4) \quad \theta = \text{tr}(D_s^{1/2}).$$

For θ to be a maximum, so that the criterion g_1 is a minimum, one has to choose all diagonal elements in $D_s^{1/2}$ nonnegative. Once they have been so chosen the orientation of W , given that of V , is determined by the condition that

$$(2.5) \quad S = WD_s^{1/2}V'$$

which was used in (2.3). This so-called "Eckart-Young decomposition" of S [3] was discussed more recently in [9, 13].

The above argument also guarantees, for the case of distinct roots, the uniqueness of T . In this case the vectors in V and W are determined up to orientation. Now let

$$(2.6) \quad S = W^*K_w D_s^{1/2} K_w V^{*'} \quad \text{so that} \quad D_s^{1/2} = K_w W^{*'} S V^* K_w,$$

where V^* and W^* are arbitrarily orientated latent vectors of $S'S$ and SS' and where the K 's are diagonal matrices with $+1$ or -1 , in arbitrary distribution, as diagonal elements. If one fixes K_w and therewith $W = W^*K_w$, then K_w will be uniquely determined by the requirement that $D_s^{1/2}$ be nonnegative, so that θ be a maximum (or g_1 be a minimum).

The case of multiple nonzero roots is chiefly of theoretical, rather than practical, interest. In [13] it is shown that in this case the decomposition (2.5) is again determined by the requirement that D_s be diagonal and positive definite. As in the preceding section this requirement determines V once W has been fixed, or vice versa. In particular, the product $T = WV'$ is unique in this case. In contrast to the foregoing case, however, some additional effort is needed to compute V , say, once W has been fixed. In [13] it is further shown that in the case of multiple nonzero roots (1.13) yield matrices W^* and V^* which in general will not diagonalize S but rather transform S into a non-

diagonal matrix D^* which has square blocks D_i^* of order $n_i \times n_i$ along the main diagonal and which is zero elsewhere. Each block in the diagonal is proportional to an arbitrary orthogonal matrix Q_i , the scalar factor corresponding to a latent root of multiplicity n_i . For an arbitrarily chosen pair of matrices W and V^* , satisfying (1.13), one would have

$$(2.7) \quad D_i^{*1/2} = D_i^{1/2} Q' = W' S V^*$$

whence it is seen that

$$(2.8) \quad Q = D_i^{*-1/2} D_i^{1/2}$$

will accomplish the desired transformation of V^* into V which will diagonalize S , if used in conjunction with W :

$$(2.9) \quad W' S V = W' S V^* Q = D_i^{1/2} Q' Q = D_i^{1/2}.$$

The case of distinct roots is evidently a special case of this. The matrices K_w of the preceding section geometrically represent reflections, which are a subgroup of the more general rotations within subspaces of dimensions n_i represented by the partitions Q_i of Q .

Finally suppose there is a multiple zero root. Then the foregoing argument applies to those latent vectors in V and W which belong to nonzero roots, because those alone suffice to reproduce the given S , hence those vectors are uniquely determined in the above sense. Those in the nullspace of either SS' or $S'S$ are not determined in this manner, but rather can be chosen arbitrarily as long as one makes sure that $WW' = VV' = I$ so that (1.14) is satisfied and can be used in (1.16). Hence it appears that in the case of multiple zeros, as with Schmid-Leiman type A and B [12], the transformation matrix T is not unique, since the vectors in the nullspace do not add into the criterion, which is only a function of the nonzero roots, which in turn are only functions of S and the latent vectors not in the nullspace, as (2.6) shows.

3. Comparison with Green's Results

Green [5] presented a solution to a somewhat less general formulation of the Orthogonal Procrustes problem. Green's solution is of interest in its own right and will be discussed in this section in comparison with the present solution.

Green's model

$$(3.1) \quad AT = B + E$$

and his criterion function

$$(3.2) \quad \text{tr}(E'E) = \min$$

are identical with the model and criterion employed in section (1.1) and (1.3), but Green places more stringent restrictions on the matrices A and B in

requiring that both be of full column rank, which implies, for example, that the matrix $S'S$ be positive definite. This condition will not be satisfied for Schmid-Leiman type A and B matrices [12]. Green also selects an alternate formulation of the side condition for differentiation, i.e.,

$$(3.3) \quad TT' = I,$$

which choice affects the normal equations obtained after differentiation,

$$(3.4) \quad S = PT + QT$$

where S , P , and Q are defined as in (1.9). Equation (3.4) differs from equation (1.9) in that the (unknown) matrix T appears twice as a right-multiplier, and hence T can be factored out

$$(3.5) \quad S = (P + Q)T.$$

Since Green assumed A and B of full column rank, the matrix

$$(3.6) \quad SS' = (P + Q)^2$$

must be positive definite. Green rewrites (3.6) as

$$(3.7) \quad P + Q = (SS')^{1/2} = WD_*^{1/2}W;$$

which, together with (3.7) leads to

$$(3.8) \quad \begin{aligned} T &= (SS')^{-1/2}S \\ &= WD_*^{-1/2}W'S, \end{aligned}$$

which is Green's solution of a special case of the Orthogonal Procrustes problem.

As Green points out the square roots in (3.7) must be taken positive to ensure a minimum of $f = \text{tr}(E'E)$. The reasoning, though not presented in Green's paper, could be similar to the argument presented in sec. 2.

To see that Green's solution is indeed a special case of the more general solution presented in sec. 1, consider the latter, as given in

$$(1.17) \quad T = WV'.$$

Assuming, as Green does, that all latent roots in D_* are positive, one may write the identity

$$(3.9) \quad D_*^{-1/2}W'WD_*^{1/2} = I.$$

Hence

$$(3.10) \quad \begin{aligned} T &= WV' = W(D_*^{-1/2}W'WD_*^{1/2})V' \\ &= (WD_*^{-1/2}W')(WD_*^{1/2}V') \end{aligned}$$

$$(3.11) \quad = (SS')^{-1/2}S,$$

which is Green's solution in (3.8).

Hunter [7] observed that the symmetry argument which led from (1.10) to the more general solution in (1.17) might as well have been used on Green's (3.5). Hence it appears that the alternate choice of the side condition is of no mathematical consequence. But Green's derivation of T suggests itself naturally once (3.4) was obtained as a consequence of using $TT' = I$ as a side condition. On the other hand, our choice of $T'T = I$ which led to (1.9) precluded a solution by standard algebraic techniques, since T appears once as a left-multiplier and once as a right-multiplier.

4. Programming Suggestions and an Illustrative Example

The basic algebra of the generalized Orthogonal Procrustes solution as presented in sec. 1 is rather straightforward and should not present many programming difficulties. A schematic flowchart is given in Table 1.

TABLE 1
Flowchart for Orthogonal Procrustes
(General Solution)

(1) Read A	(4) $T = WV'$
Read B	(5) $B^* = AT$
(2) Compute	(6) $E = B - B^*$
$S = A'B$	(7) Output T, AT, $E = B - AT$
(3) Diagonalize	(8) End.
$S'S = VD_s V'$	
$SS' = WD_s W'$	

Two minor difficulties arise at steps (3) and (4). At step (3) the diagonalization is to be accomplished by orthogonal V and W . Certain eigenvalue subroutines, e.g., HOW [6], yield orthogonal eigenvectors only for distinct roots. For such routines the occurrence of multiple zero roots could be handled in the following manner.

Let

$$(4.1) \quad C = C' \quad (m \times m)$$

be any symmetric matrix of rank $r < m$ and let it be desired to find a full set of eigenvectors T ($m \times m$) such that

$$(4.2) \quad T'CT = D = \text{diagonal},$$

$$T'T = TT' = I.$$

- (1) Find the eigenvectors T_r of C which correspond to the r nonzero

roots of C so that

$$(4.3) \quad C = T_r D_r T_r', \quad D_r = (d_{ii}) = \text{diagonal}, \quad T_r' T_r = I.$$

(2) Find the nullspace N of T_r , i.e., solve

$$(4.4) \quad T_r' N = 0$$

for N , which will be $m \times (m - r)$, so that the columns of N give $m - r$ independent solutions of the homogeneous system $T_r' x = 0$.

(3) Orthogonalize N by G , e.g., by a Gram-Schmidt process, so that

$$(4.5) \quad NG = T_0, \quad T_0' T_0 = I, \quad T_r' T_0 = 0.$$

(4) Assemble

$$(4.6) \quad T = [T_r \quad T_0]$$

which will contain a full set of n orthogonal vectors which diagonalize C

$$(4.7) \quad C = [T_r \quad T_0] \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} T_r' \\ T_0' \end{bmatrix}$$

and $T' T = T T' = I$.

None of these complications arise if eigenproblem subroutines are available which are based on the Jacobi method provided storage was set aside for accumulating the successive 2×2 rotations.

It has been pointed out [15] that the double factorization at step (8) could have been avoided by utilizing W_r in (2.5), i.e., that part of W which corresponds to the nonzero roots in SS' . Since W_r and the corresponding set in V , say V_r , together with D_r suffice to reproduce S in (2.5), one might compute V_r as $V_r = S' W_r D_r^{-1/2}$, if so desired. But the speed of HOW is such as to make it unlikely that this approach would lead to a marked gain in program efficiency, and one still is left with the task of completing W and V by the vectors in their null-spaces, which can be tackled in a variety of ways. There probably are many different technical solutions to the problem of finding the T of (1.17), and further experience will tell which is to be preferred and under which circumstances. The problem of speed, in any case, does not appear to be a crucial one for an Orthogonal Procrustes routine.

The second programming difficulty concerns the proper orientation of the (orthogonalized) eigenvectors in V and W of step (4). The algebra presented in sec. 2 may be applied for this purpose: compute $W^* S V^* = D_s^{*1/2}$ for arbitrarily orientated columns of W^* and V^* and reflect the columns of W^* , say, until all diagonal elements in $D_s^{*1/2}$ are nonnegative. Call the reflected matrix of eigenvectors W , then $T = W V^*$ will minimize $\text{tr}(E'E)$ as shown in sec. 2.

The general solution of the Orthogonal Procrustes problem as described

in sec. 1 was programmed, in FORTRAN, for the IBM 7094. The program was tried on a number of numerical examples taken from the literature. Here only a small and artificial example will be presented to illustrate some of the geometry involved.

The data for this 4×2 example are given in Table 2. Suppose one considers the four rows in each matrix A and B as giving the (Cartesian) coordinates of two sets of four points each, and the object is to rotate the first set (v_k) out of the subspace in which it is contained (viz., the x -axis) into the whole two-dimensional space so as to minimize the distances between the transformed points (v_k^*) and the corresponding points (v_k') described by the matrix B . The rotated points would be described by a matrix AT and the matrix T would have to be chosen orthogonal if the angular relations between the points (v_k^*), and in particular their distances from the origin are to remain invariant under the transformation. But even if T is chosen optimally, the fit will not be perfect, because some of the points described in A are further away from the origin than the corresponding target points described in B . For example v_1' , if treated as the endpoint of a vector emanating from the origin, is of length .90, whereas its corresponding target point v_1 is only of length .85.

The points were so chosen as to maximize the fit if the first set of points is rotated by 45° . This is indeed the angle of rotation corresponding to the matrix T in Table 2. A matrix $B - AT = E$ would then give the coordinates of four residual vectors corresponding to the distances between the rotated points (v_k^*) and the target points (v_k). In this particular example, all four error vectors lie in the direction of the line of best fit, i.e., the 45° line, because both sets of points can be made collinear, on this line.

TABLE 2

Orthogonal Procrustes - Hypothetical Example

Matrix A		Matrix B	
	<u>1</u> <u>2</u>		<u>1</u> <u>2</u>
1	900 000	1	600 600
2	600 000	2	400 400
3	-600 000	3	-400 -400
4	-900 000	4	-600 -600

Matrix T		Matrix AT	
	<u>1</u> <u>2</u>		<u>1</u> <u>2</u>
1	707 707	1	636 636
2	707 -707	2	424 424
		3	-424 -424
		4	-636 -636

5. *Summary*

The least-squares problem of transforming a given matrix A into a given matrix B by an orthogonal transformation matrix T so that the sum of squares of the residual matrix $E = B - AT$ is a minimum is called an "Orthogonal Procrustes problem." Green presented a solution of this problem in 1952. His solution exists if and only if both A and B (in $S = A'B$) are of full column rank.

The presently proposed solution of the Orthogonal Procrustes problem is not subject to this constraint. Some algebraic properties of T are discussed and a number of technical suggestions for the numerical computation of T , including a flowchart are presented.

REFERENCES

- [1] Bellman, R. *Introduction to matrix analysis*. New York: McGraw-Hill, 1960.
- [2] Dwyer, P. S. and McPhail, M. S. Symbolic matrix derivatives. *Ann. math. Statist.*, 1948, **19**, 517-534.
- [3] Eckart, C. and Young, G. The approximation of one matrix by another of lower rank. *Psychometrika*, 1936, **1**, 211-218.
- [4] Gibson, W. A. On the least-squares orthogonalization of an oblique transformation. *Psychometrika*, 1962, **27**, 193-196.
- [5] Green, B. F. The orthogonal approximation of an oblique structure in factor analysis. *Psychometrika*, 1952, **17**, 429-440.
- [6] HOW. FORTRAN Subroutine, using non-iterative methods of Householder, Ortega, and Wilkinson, solves for eigenvalues and corresponding eigenvectors of a real symmetric matrix. Program writeup F2 BC HOW. Berkeley Division, Univ. Calif., 1962.
- [7] Hunter, J. Personal communication, 1964.
- [8] Hurley, J. R. and Cattell, R. B. Producing direct rotation to test a hypothesized factor structure. *Behav. Sci.*, 1962, **7**, 258-262.
- [9] Johnson, R. M. On a theorem by Eckart and Young. *Psychometrika*, 1963, **28**, 259-263.
- [10] Johnson, R. M. The minimal transformation to orthonormality. Paper read at the joint meeting of the Psychonomic Society and the Psychometric Society, Niagara Falls, Ontario, Canada, 1964.
- [11] Mosier, C. I. Determining a simple structure when loadings for certain tests are known. *Psychometrika*, 1939, **4**, 149-162.
- [12] Schmid, J. and Leiman, J. M. The development of hierarchical factor solutions. *Psychometrika*, 1957, **22**, 53-61.
- [13] Schönemann, P. H., Bock, R. D., and Tucker, L. R. Some notes on a theorem by Eckart and Young. Univ. of North Carolina Psychometric Laboratory Research Memorandum No. 25, 1965.
- [14] Thurstone, L. L. *Multiple-factor analysis*. Chicago: Univ. Chicago Press, 1947.
- [15] Tucker, L. R. Personal communication, 1964.

Manuscript received 11/9/64

Revised manuscript received 6/8/65