PSYCHOMETRIKA—VOL. 49, NO. 1, 49–55
MARCH 1984

ORTHOGONAL ROTATIONS TO MAXIMAL AGREEMENT FOR TWO OR MORE MATRICES OF DIFFERENT COLUMN ORDERS

JOS M. F. TEN BERGE AND DIRK L. KNOL

UNIVERSITY OF GRONINGEN

Methods for orthogonal Procrustes rotation and orthogonal rotation to a maximal sum of inner products are examined for the case when the matrices involved have different numbers of columns. An inner product solution offered by Cliff is generalized to the case of more than two matrices. A nonrandom start for a Procrustes solution suggested by Green and Gower is shown to give better results than a random start. The Green-Gower Procrustes solution (with nonrandom start) is generalized to the case of more than two matrices. Simulation studies indicate that both the generalized inner product solution and the generalized Procrustes solution tend to attain their global optima within acceptable computation times. A simple procedure is offered for approximating simple structure for the rotated matrices without affecting either the Procrustes or the inner product criterion.

Key words: Procrustes rotation, target rotation, congruence.

Procedures for rotating matrices to maximal agreement (often called matching procedures) have been offered by many authors and have found numerous applications both in the context of factor analysis and multidimensional scaling. Surveys of several methods have been given by Kettenring (1971) and, more extensively, by Ten Berge (Note 2). Ten Berge offered a taxonomy of matching procedures based on these following five dimensions.

First, each matching procedure is aimed at optimizing some particular criterion (loss function or agreement measure). Three criteria are especially popular: The Procrustes criterion, the inner product criterion, and the congruence criterion. Let A_i, $i = 1, 2, \ldots, m$, be given matrices of order $n \times k_i$, with $n \geq k_1 \geq k_2 \geq \cdots \geq k_m$, and let T_i be rotation matrices of order $k_i \times k_m$, $i = 1, 2, \ldots, m$. Then the Procrustes criterion is defined as

$$f(T_1, T_2, \ldots, T_m) = \sum_{i<j} \text{tr } (A_i T_i - A_j T_j)'(A_i T_i - A_j T_j) \tag{1}$$

which shows that 'Procrustes' is synonymous to 'least-squares' in this context. The inner product criterion is defined as

$$g(T_1, T_2, \ldots, T_m) = \sum_{i<j} \text{tr } T_i' A_i' A_j T_j. \tag{2}$$

The congruence criterion is defined as

$$h(T_1, T_2, \ldots, T_m) = \sum_{i<j} \sum_{p=1}^{k_m} \phi(A_i t_{ip}, A_j t_{jp}) \tag{3}$$

where ϕ is Tucker's coefficient of congruence (Tucker, Note 3), a measure of proportionality of vectors. The inner product criterion can be understood as a weighted congruence criterion, each congruence being weighted by the geometric mean of the squared lengths

The authors are obliged to Charles Lewis for helpful comments on a previous draft of this paper and to Frank Brokken for preparing a computer program that was used in this study.

Reprint requests should be addressed to Jos ten Berge and Dirk Knol, Subfakulteit Psychologie R.U. Groningen, Grote Markt 31/32, 9712 HV Groningen, The Netherlands.

0033-3123/84/0300-4131$00.75/0

of the two vectors involved. On the other hand, the inner product criterion coincides with the Procrustes criterion in some important special cases, as will be explained below.

A second dimension along which Ten Berge classified matching procedures is concerned with the use of *orthogonality* constraints on the rotation matrices $T_1, T_2, \ldots, T_m$. Very often these matrices are required to be columnwise orthogonal in some sense, but *unrestricted* solutions also exist.

A third distinction is that between *simultaneous* and *successive* solutions. In the latter, the columns of the rotation matrices are computed successively, subject to orthogonality constraints with respect to the columns already computed. This distinction is immaterial for unrestricted solutions.

Fourth, Ten Berge distinguished matching procedures for $m = 2$ matrices from *generalized* procedures involving three or more matrices. In the case $m = 2$ he added the distinction between *one-sided* procedures, where T_2 is an identity matrix, and *two-sided* procedures, where no such constraint is imposed.

Finally, Ten Berge distinguished procedures for the case $k_1 = k_2 = \cdots = k_m$, called *symmetric* in the present paper, from procedures which allow the matrices A_i to have unequal numbers of columns (the *asymmetric* case).

The present paper is concerned with only a small subset of the set of matching problems generated by Ten Berge (Note 2). Specifically, we shall examine the simultaneous orthogonal Procrustes and inner product rotation problems. The distinction between one-sided and two-sided rotations for $m = 2$ is irrelevant here because T_2 can always be incorporated in T_1: Postmultiplying both T_1 and T_2 by T_2' does not affect function (1) or (2). This leaves us with these four problems to consider:

Problem 1: Maximize $g(T_1) = \text{tr } T_1'A_1'A_2$ subject to the constraint $T_1'T_1 = I_{k_2}$;

Problem 2: Maximize (2) subject to the constraint

$$T_1'T_1 = T_2'T_2 = \cdots = T_m'T_m = I_{k_m}, \qquad m > 2;$$

Problem 3: Minimize $f(T_1) = \text{tr } (A_1T_1 - A_2)'(A_1T_1 - A_2)$ subject to the constraint $T_1'T_1 = I_{k_2}$; and

Problem 4: Minimize (1) subject to the constraint

$$T_1'T_1 = T_2'T_2 = \cdots = T_m'T_m = I_{k_m}, \qquad m > 2.$$

In the *symmetric* case Problems 1 and 3 coincide. Solutions are well-known, e.g., Green (1952) and Cliff (1966). The generalized Problems (2) and (4) also coincide in the symmetric case. A solution was derived by Ten Berge (1977). In the *asymmetric* case the four problems are distinct. A solution for Problem 1 was offered by Cliff (1966) and a solution for Problem 3 was suggested by Green and Gower (Note 1). The main purpose of the present paper is to generalize these two asymmetric solutions to the case $m > 2$.

Generalization of the Asymmetric Orthogonal Inner Product Rotation

The solution to Problem 1 can be obtained at once from the Eckart-Young or singular value decomposition

$$A_1'A_2 = PDQ' \tag{4}$$

with $P'P = Q'Q = QQ' = I_{k_2}$ and D $(k_2 \times k_2)$ diagonal and nonnegative. The solution for T_1 is

$$T_1 = PQ' \tag{5}$$

cf. Cliff (1966). For computational purposes it may be convenient to obtain this solution in a different way. Let A_2^+ be an $n \times k_1$ matrix obtained by adding $k_1 - k_2$ zero columns to A_2. Replacing A_2 in (4) by A_2^+ will yield a $k_1 \times k_1$ solution T_1^+ in (5). Deleting the last $k_1 - k_2$ columns from T_1^+ will yield T_1, cf. Ten Berge (Note 2, p. 29). This approach permits treatment of the asymmetric Problem (1) as a symmetric problem.

So far, no solutions for the generalized Problem 2 have been offered. Although Ten Berge (Note 2, p. 44) outlined an approximate solution this does not seem to be generally useful. However, a very simple and straightforward generalized solution can be derived. Consider the inner product function (2) as a function of T_i only, for fixed matrices T_j $(j \neq i)$:

$$g(T_i) = \text{tr } T_i' A_i' \sum_{j \neq i} A_j T_j + K_i \tag{6}$$

where K_i is a constant with respect to T_i. It is readily verified that maximizing (6) is simply a matter of rotating A_i to a maximal sum of inner products with the matrix $\sum_{j \neq i} A_j T_j$. Clearly, once starting values have been inserted in $T_1, T_2, \ldots, T_m$ each of these matrices can be replaced in turn by the optimal T_i, for fixed matrices T_j $(j \neq i)$. The optimal T_i can be obtained as T_1 from the solution of Problem 1, taking $A_1 = A_i$ and $A_2 = \sum_{j \neq i} A_j T_j$. Since each replacement increases (2) monotonely and (2) is bounded, this procedure must converge if it is terminated when no further significant increments are obtained. In fact this procedure is a very straightforward generalization of Ten Berge's solution for the *symmetric* generalized inner product rotation problem (Ten Berge, 1977).

A Solution for Problem 3

Above, Problem 3 was defined as the problem of minimizing

$$f(T_1) = \text{tr } (A_1 T_1 - A_2)'(A_1 T_1 - A_2) \tag{7}$$

subject to the constraint $T_1' T_1 = I_{k_2}$. Until very recently, no solution to this problem (with $k_1 > k_2$) was available, except for the special case of $k_2 = 1$, cf. Ten Berge & Nevels (1977). Ten Berge (Note 2, p. 29–30) derived as necessary conditions for a minimum of (7) that $T_1' A_1' A_2$ be symmetric and positive semidefinite. This result is stronger than that obtained by Green and Gower (Note 1, p. 4), who showed that $T_1' A_1' A_2$ must be symmetric. More importantly, Green and Gower offered an algorithm for the minimization of (7): Initially, the matrix A_2 is extended into an $n \times k_1$ matrix A_2^* by adding $k_1 - k_2$ arbitrary columns to it. Then the following two steps are taken iteratively:

Step 1. The symmetric orthogonal Procrustes problem for the matrices A_1 and A_2^* is solved by the conventional methods to yield a $k_1 \times k_1$ rotation matrix T_1^*.

Step 2. The matrix T_1^* is partitioned as $T_1^* = [T_1 \mid T_*]$ where T_1 is of order $k_1 \times k_2$ and T_* is of order $k_1 \times (k_1 - k_2)$. The last $k_1 - k_2$ columns of A_2^* are replaced by $A_1 T_*$.

Clearly, Step 2 has no effect on $f(T_1)$. Step 1 improves the fit between $A_1 T_1^*$ and A_2^*. This fit partitions into the fit between $A_1 T_1$ and A_2, cf. (7), and the fit between $A_1 T_*$ and the submatrix containing the last $k_1 - k_2$ columns of A_2^*. Since the latter fit can only be diminished by Step 1, the fit between $A_1 T_1$ and A_2 must be improved by Step 1. This explains why the procedure must improve on each cycle, as was claimed by Green and Gower. In addition, after convergence, the matrix $T_1^{*\prime} A_1' A_2^*$ must be symmetric and positive semidefinite by Step 1, cf. Ten Berge (1977, p. 269). Since $T_1' A_1' A_2$ is a submatrix of

$T_1^{*\prime}A_1^{\prime}A_2^*$ with diagonal elements that are also diagonal elements of the latter matrix, $T_1^{\prime}A_1^{\prime}A_2$ must, upon convergence, be symmetric and positive semidefinite, hence it must satisfy the necessary conditions mentioned above.

Avoiding Local Minima With the Green-Gower Method

Green and Gower (Note 1, p. 6) point out that their method need not converge to the global minimum. The following artificial example may demonstrate this. Let

$$A_1 = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad A_2 = \begin{pmatrix} 3 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \tag{8}$$

and let a zero column be added to A_2 to yield an initial A_2^*. Then Step 1 of the Green-Gower method will yield a rotation matrix

$$T_1^* = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & x \end{pmatrix} \tag{9}$$

with $x = 1$ or $x = -1$. The third column of T_1^* is T_*. Since $A_1 T_*$ is the zero vector, nothing changes by Step 2 of the Green-Gower method, which breaks down at this point, yielding a value $f(T_1) = 1$. This is not the global minimum of zero, which is attained for

$$T_1 = \begin{pmatrix} 1 & 0 \\ 0 & .5 \\ 0 & y \end{pmatrix} \tag{10}$$

where $y = (.75)^{1/2}$ or $-(.75)^{1/2}$.

In order to avoid local minima Green and Gower suggest computing several solutions, with different random starts for A_2^*. They report that "in our limited experience local optima are seldom found." It seems, however, that better starts can be chosen. For instance, a nonrandom start can be derived from a conditional solution to Problem 3 suggested by Ten Berge (Note 2, p. 30). This solution is based on the expansion of $f(T_1)$, cf. (7), as

$$f(T_1) = \text{tr } T_1^{\prime}A_1^{\prime}A_1 T_1 - 2 \text{ tr } T_1^{\prime}A_1^{\prime}A_2 + \text{tr } A_2^{\prime} A_2 \tag{11}$$

and takes $T_1 = KW$, where K contains the k_2 eigenvectors of $A_1^{\prime}A_1$, associated with the k_2 smallest eigenvalues, and W is a $k_2 \times k_2$ orthonormal matrix which maximizes tr $T_1^{\prime}A_1^{\prime}A_2$. Note that W does not affect tr $T_1^{\prime}A_1^{\prime}A_1 T_1$. It can be verified that this solution minimizes the second term of (11) conditional on minimizing the first term. This will not produce the global minimum of $f(T_1)$ but it will approximate this minimum. A nonrandom start associated with this conditional solution can be obtained by inserting the $k_2 - k_1$ eigenvectors of $A_1^{\prime}A_1$, associated with the $k_2 - k_1$ smallest eigenvalues, in T_* and adding the matrix $A_1 T_*$ to A_2. We shall refer to this start as 'the conditional start'.

In order to investigate the liability of either starting option to producing local minima the options were compared in a number of actual computations. Data sets were selected or constructed in such a way that the global minimum could be determined or was known in advance.

For the case $k_2 = 1$ the global minimum of $f(T_1)$ can be found by a method offered by Ten Berge and Nevels (1977). The authors computed both the global minimum and the minima attained by the Green-Gower method with the conditional and the random starts, for 21 sets of data with $k_2 = 1$ and k_1 varying between 4 and 8. In all 21 cases the conditional start yielded the global minimum. The random start yielded the global minimum at once in 19 cases. In two cases the random start had to be repeated twice before the global minimum was attained.

A similar study was conducted for 13 sets of data where perfect fit was known to exist by construction. In these sets k_1 varied from 3 to 8 and k_2 varied from 2 to 4. Again the conditional start yielded the global minimum in all cases. The random start produced the global minimum at once in eleven cases. In two cases, four or five restarts were needed.

In addition, 15 artificial data sets were constructed in which $A_1'A_1$ was a scalar matrix. For these data minimizing $f(T_1)$ coincides with maximizing $g(T_1)$, tr $T_1'A_1'A_1T_1$ being constant. The global minima were derived from Cliff's solution to Problem 1 (see above). In all 15 cases both starting options yielded the global minimum.

Finally, the number of iterations required by the two starting options were compared. The 15 artificial data sets with scalar $A_1'A_1$ and the four data sets for which the random option produced local minima were left out of consideration. For the conditional start the initial truncated eigendecomposition of $A_1'A_1$ was counted as one additional iteration. Over the 30 data sets considered, both options required the same number of iterations in 5 cases; the conditional option required more iterations in 12 cases and less iterations in 13 cases.

The first three studies indicate that the conditional start is superior to the random start in terms of sensitivity to local minima. Both starts seem to entail similar computation times but since the random start needs to be repeated several times the conditional start approach saves time as well. It can be concluded that the Green-Gower method with the conditional start is the most attractive solution to the asymmetric orthogonal Procrustes problem for two matrices currently available.

A Solution for the Generalized Orthogonal Asymmetric Procrustes Problem

Above, the solution to Problem 1 was generalized to a solution of Problem 2, by rotating each matrix in turn to a maximal sum of inner products with the sum of the remaining rotated matrices, cf. (6). In a parallel fashion the solution to Problem 3 discussed in the previous section can be generalized to the case of more than two matrices (Problem 4). Consider the generalized Procrustes function (1) as a function of T_i only, for fixed matrices T_j $(j \neq i)$:

$$f(T_i) = (m-1) \text{ tr } T_i'A_i'A_iT_i - 2 \text{ tr } T_i'A_i'\sum_{j\neq i} A_jT_j + K_i \tag{12}$$

where K_i is a constant with respect to T_i. Completing the square yields

$$\begin{aligned} f(T_i) &= (m-1) \text{ tr } T_i'A_i'A_iT_i - 2 \text{ tr } \left[T_i'A_i'\sum_{j\neq i} A_jT_j \right] \\ &\quad + (m-1)^{-1} \text{ tr } \left[\left(\sum_{j\neq i} A_jT_j \right)' \left(\sum_{j\neq i} A_jT_j \right) \right] + K_i^+ \\ &= (m-1) \text{ tr } \left[\left(A_iT_i - \sum_{j\neq i} \frac{A_jT_j}{m-1} \right)' \left(A_iT_i - \sum_{j\neq i} \frac{A_jT_j}{m-1} \right) \right] + K_i^+ \end{aligned} \tag{13}$$

where K_i^+ is another constant with respect to T_i. From (13) it is obvious that minimizing (12) is simply a matter of rotating A_i to a best least squares fit with $(m-1)^{-1} \sum_{j\neq i} A_jT_j$, the average of the remaining rotated matrices. Again, after starting values have been inserted in $T_1, T_2, \ldots, T_m$ each of the matrices T_i $(i = 1, 2, \ldots, m)$ can be replaced in turn by the optimal T_i for fixed matrices T_j $(j \neq i)$. The optimal T_i can be obtained by solving Problem 3 for $A_1 = A_i$ and $A_2 = (m-1)^{-1} \sum_{j\neq i} A_jT_j$. Each replacement reduces (1) monotonely and (1) is bounded. Therefore, the procedure must converge if it is terminated when no further significant reduction of (1) is obtained.

Simulation Studies for the Two Generalized Procedures

The solutions to Problem 2 and 4, briefly labelled 'inner product method' and 'Procrustes method' in this section, were implemented in a Pascal computer program. The program was run both on randomly generated factor loading matrices and on real data.

First, the inner product method was run on 21 sets of random matrices. Each set contained at least three and at most 10 matrices, with 10, 15 or 20 rows, and numbers of columns varying between 1 and 12. The values of (2) obtained were compared with two upper bounds that can be derived for (2) as generalizations of upper bounds for the symmetric case derived by Ten Berge (1977, p. 273). The ratios of obtained values of (2) to the best upper bound in each set varied from .879 to .993, with an average of .936. This implies that the obtained values of (2) cannot be far below their maxima. Computing times on a CDC Cyber 170/760 were reasonable. In only one data set, with 10 matrices, more than half a minute CP-time was needed.

Second, the Procrustes method was run on 22 sets of random matrices where by construction perfect fit was known to exist. Again, each set contained between 3 and 10 matrices with 10, 15 or 20 rows and numbers of columns varying between 1 and 12. In 21 cases perfect fit was obtained. A minor departure from perfect fit was obtained in one case. Computing times were higher than for the inner products method. In only two cases, with 8 and 10 matrices, respectively, more than two minutes CP-time were needed.

Third, 20 sets of random matrices were constructed such that $A_i'A_i$ was a scalar matrix, thus yielding the Procrustes criterion and the inner product criterion identical up to a constant. Each set contained between three and eight matrices with 10, 15 or 20 rows and numbers of columns varying between 1 and 12. For the inner product method, the ratios of obtained values of (2) to the best upper bound in each set varied from .927 to .999, with an average of .971. The Procrustes method yielded values of (2) which were equal or differed trivially from the values obtained by the inner product method.

Finally, the inner product method was run on 11 sets of real matrices. Each set contained between three and six matrices with numbers of rows ranging from 8 to 18 and numbers of columns ranging from 1 to 8. The ratios of obtained values of (2) to the best upper bound for each set now had an average of .999. In addition, the computations never took more than two seconds CP-time.

These studies show that there is no severe local optimum problem for the two generalized methods discussed in this section. Also, computing times on a high speed computer tend to be quite acceptable.

A description of the computing program used in this study is available upon request from the authors.

Rotation to Simple Structure in the Matched Space

For $m > 1$ both the Procrustes criterion (1) and the inner product criterion (2) are insensitive to a joint rotation in the matched space. That is, postmultiplying the matrices $T_1, T_2, \ldots, T_m$ by an arbitrary orthonormal $k_m \times k_m$ matrix N does not affect the match. This property was discussed by Cliff (1966, p. 40) for the case $m = 2$ and readily generalizes to the case $m > 2$. It permits approximating 'simple structure' for the rotated matrices without any loss of agreement between them. An approximation to simple structure can be obtained conveniently by constructing the supermatrix

$$A^* = \begin{pmatrix} A_1 T_1 \\ A_2 T_2 \\ \vdots \\ A_m T_m \end{pmatrix} \tag{14}$$

and subjecting it to a varimax rotation. This yields a common rotation N for the matrices T_i, $i = 1, 2, \ldots, m$ which approximates simple structure for all matrices $A_i T_i N$ simultaneously.

REFERENCE NOTES

1. Green, B. F. & Gower, J. C. (1979). *A problem with congruence.* Paper presented at the Annual Meeting of the Psychometric Society, Monterey, California.
2. Ten Berge, J. M. F. (1977). *Optimizing factorial invariance.* Unpublished doctoral dissertation, University of Groningen.
3. Tucker, L. R. (1951). *A method of synthesis of factor analysis studies* (Personnel Research Section Report No. 984). Washington, D.C.: Department of the Army.

REFERENCES

Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika, 31*, 33–42.

Green, B. F. (1952). The orthogonal approximation of an oblique structure in factor analysis. *Psychometrika, 17*, 429–440.

Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika, 58*, 433–451.

Ten Berge, J. M. F. (1977). Orthogonal Procrustes rotation for two or more matrices. *Psychometrika, 42*, 267–276.

Ten Berge, J. M. F. & Nevels, K. (1977). A general solution for Mosier's oblique Procrustes problem. *Psychometrika, 42*, 593–600.

Manuscript received 6/21/82
Final version received 9/22/83