PSYCHOMETRIKA—VOL. 49, NO. 3, 403–423
SEPTEMBER 1984

MATRIX CORRELATION

J. O. RAMSAY

MCGILL UNIVERSITY

JOS TEN BERGE

UNIVERSITY OF GRONINGEN

G. P. H. STYAN

MCGILL UNIVERSITY

A correlational measure for an n by p matrix X and an n by q matrix Y assesses their relation without specifying either as a fixed target. This paper discusses a number of useful measures of correlation, with emphasis on measures which are invariant with respect to rotations or changes in singular values of either matrix. The maximization of matrix correlation with respect to transformations XL and YM is discussed where one or both transformations are constrained to be orthogonal. Special attention is focussed on transformations which cause XL and YM to be n by s, where s may be any number between 1 and min (p, q). An efficient algorithm is described for maximizing the correlation between XL and YM where analytic solutions do not exist. A factor analytic example is presented illustrating the advantages of various coefficients and of varying the number of columns of the transformed matrices.

Introduction

Given two matrices X and Y with p and q columns respectively and both having n rows, the matrix correlation process typically requires two steps:

1. Apply transformations L and M to produce matrices XL and YM, respectively, where both transformed matrices have s columns, and
2. Use an appropriate measure of correlation $r(XL, YM)$ to summarize the match between the transformed matrices.

These two steps are usually connected in that the transformations will be required to maximize the correlation. It may be that there will be some constraints on the possible transformations.

The prototypical matrix correlation technique is undoubtably canonical correlation analysis, which provides a symmetric assessment of the congruence of two matrices X and Y having n rows and p and q columns, respectively ($n \geq p, q$). The two matrices are matched by transformations L and M so as to maximize the bilinear form $(XL, YM) :=$ tr $(L'X'YM)$ under the orthogonality constraints $L'X'XL = I$ and $M'Y'YM = I$. This process may be described as either stepwise with matrices L and M being p by 1 and q by 1, or global with L and M being p by s and q by s, $s = \min(p, q)$. In the stepwise case there are the additional conditions $L'X'XJ = 0$ and $M'Y'YK = 0$ where J and K are the transformations computed in previous steps. Canonical correlations have been defined in many ways: by classical analysis (Hotelling, 1936), by the theory of projectors (Rao &

This research was supported by grant APA 0320 from the Natural Sciences and Engineering Research Council of Canada. The authors wish to acknowledge valuable discussion of this problem with Jan de Leeuw, University of Leiden. Reprint requests should be sent to the first author at Department of Psychology, 1205 Dr. Penfield Ave., Montreal, Quebec, Canada H3A 1B1.

0033-3123/84/0900-5036$00.75/0

Yanai, 1979; Yanai, 1974), by singular value decompositions (Lancaster, 1969), by special matrix operators (Escoufier, 1973, 1977), and by functional analysis (Cailliez & Pages, 1976; Dauxois & Pousse, 1976). They have also been generalized in a number of ways: to more than two matrices (reviewed by Kettenring, 1971), to sets of random functions (Besse, 1979), to nonlinear transformations (Dauxois & Pousse, 1976), and to matrices containing indicator variables (Pages, Ramsay, & Tenenhaus, 1984.)

Implicit in all of this work has been the assumption that the rows and columns of X and Y have no internal structure that is not preserved by linear transformations, and thus that the convenient orthogonality constraints, $L'X'XL = I$ and $M'Y'YM = I$, are as valid as any other. It often arises in practice, however, that the only transformations that can be entertained for one or both matrices are orthogonal in some prespecified sense; that is, they satisfy the constraints $L'UL = I$ and/or $M'VM = I$ where U and V are symmetric positive definite real matrices. The most familiar of such situations arises when two factor pattern matrices derived by orthogonal factor analysis are to be compared, with the only possible linear transformations being orthogonal rotations. In multidimensional scaling, where a number of algorithms have been developed to use the individualized generalized Euclidean distance model $d_{ijr}^2 = (x_i - x_j)'U_r(x_i - x_j)$, it may be desirable to compare two solutions derived using different metric matrices U_1 and U_2. In such a case one must consider linear transformations which reduce the two configurations to a common metric. When this metric is the identity metric, this implies that one transformation must be U_1-orthogonal and the other U_2-orthogonal. As another example let X be subject only to rotations and let Y be subject to any linear transformation, as would be the case when a multidimensional scaling configuration is to be compared to a set of other measures on the stimuli. Such problems require the constraints $L'L = I$ and the fixing of tr $(M'Y'YM)$ to a specified value.

It may also be desirable to compare two matrices in a space of dimensionality s somewhere between one and min (p, q). A best representation of each matrix in a plane is especially useful for graphical displays of the relationship. While various functions of the first s canonical correlations permit a summary of congruence in s dimensions when any linear transformations are possible, it will be shown that the situation is not so simple with arbitrary orthogonality constraints.

In addition to the constraints on L and M and the dimensionality of comparison s, a third aspect of the matrix congruence problem is the choice of measure of correlation between the two transformed matrices. There are many possibilities, and certain advantages attach to certain choices. Within the context of canonical correlation a number of indices have been reviewed by Cramer and Nicewander (1979) and still others will be proposed here.

Implicit in any matrix correlation technique is a choice of metric for the row and column spaces associated with the two matrices. In almost all published work these are assumed to be the identity metrics, but there are often good practical reasons for applying differential weights to rows and/or columns. For example, in comparing two factor pattern matrices, where rows correspond to variables, it may be worth applying less weight to each variable within a block of variables known in advance to be very similar to one another. One may also wish to diminish the influence of variables known to be highly prone to measurement error. A choice of column metric is implied in comparing two multidimensional scaling solutions where one is primarily interested in how the first two dimensions match but one does not wish to ignore altogether the influence of other dimensions. Finally, a matrix correlation may be made robust against unusual or outlying rows or columns by appropriate choice of metrics. Thus, we assume that the metric for the column space is represented by symmetric positive definite matrix W of order n, and the metric for the row space by symmetric positive definite matrix N of order s.

Two matrices may also be related by minimizing a distance measure such as $\| YM - XL\|^2 := [\text{tr } (YM - XL)'(YM - XL)]$ with respect to L and M. This approach is usually referred to as Procrustes rotation. Although the emphasis in this paper is on correlational measures of relationship, some remarks will be made on this problem where appropriate. A fairly general treatment of Procrustes rotation in arbitrary row and column metrics is in *Rao and Yanai* (1979).

In this paper the problem of how to transform two matrices to congruence will be discussed separately from the problem of how to assess congruence after the transformations have taken place. Thus in the second section various indices of matrix correlation are presented which form a natural family with respect to their sensitivities to the orientations and degree of ellipticity of the two matrices. In practice two matrices are compared by choosing a measure of correlation and then transforming the matrices so as to maximize this measure. Results are presented in the next two sections on the optimization of these measures. Section 3 discusses the case in which both L and M are subject to arbitrary orthogonality constraints, and Section 4 discusses the case in which only L is constrained to be orthogonal. A practical example involving the comparison of two factor analysis results is given in the fifth section.

Notation and Preliminary Results A goal of this paper is to offer an exposition of the matrix correlation problem in the context of arbitrary metrics W, N, U, and V. However, for simplicity these will all be assumed to be identity matrices when correlational formulas and theorems are initially stated. The general results will only be stated without proof and then only when the generalization is less than obvious. However, in order to deal with arbitrary metrics, the following notation and lemma are essential.

The real numbers will be indicated by $\mathscr{R}$ and the fact that a real matrix X has n rows and p columns will be indicated by $X \in \mathscr{R}^{np}$, where $\mathscr{R}^{np}$ is the vector space of real n by p matrices. Real symmetric matrices of order s will be denoted by S_s and if also positive definite by S_s^+. Thus row metric $W \in S_n^+$ and column metric $N \in S_s^+$. The subset of $\mathscr{R}^{ps}$ consisting of matrices L satisfying $L'UL = I$ will be indicated by O_U^{ps}. The symbol $:=$ will be used to mean "is defined to be."

Two matrices A and B will be said to be column-orthogonal in the metric W when $A'WB = 0$ and row-orthogonal in the metric N when $ANB' = 0$. The inner product (A, B) in metrics W and N is tr $(A'WBN)$ with the associated norm $\| A \| := [\text{tr } (A'WAN)]^{1/2}$.

The following extension of the singular value decomposition theorem (svd) is fundamental to a general treatment.

Lemma: Given matrices $U \in S_p^+$ and $V \in S_q^+$, and an arbitrary matrix $A \in \mathscr{R}^{pq}$, $p \geq q$, there exist matrices $P \in O_U^{pq}$, $Q \in O_V^{qq}$ and D such that

(a) D is diagonal with $d_{11} \geq \cdots \geq d_{qq} \geq 0$

(b) $A = PDQ'$

Proof Let U and V have the decompositions

$$U = Q_U D_U V'_U \qquad \text{and} \qquad V = Q_V D_V Q'_V,$$

respectively. Their symmetric square roots are given by

$$U^{1/2} = Q_U D_U^{1/2} Q'_U \qquad \text{and} \qquad V = Q_V D_V^{1/2} Q'_V.$$

Let the matrix $U^{1/2}AV^{1/2}$ have the conventional singular value decomposition $P^*D^*Q^{*\prime}$. Then the matrices $P := U^{-1/2}P^*$, $Q := V^{-1/2}Q^*$, and $D := D^*$ have the required properties. □

The singular value decomposition in metrics U and V, called here the (U, V)-orthogonal svd, corresponds to the definition of the eigenequation of a symmetric matrix

C in metric N as $CNz = \lambda z$, and the eigenvalues and eigenvectors of C are obtained by the (N, N)-orthogonal svd $C = QDQ'$. An arbitrary power C^m of a positive semidefinite matrix C in metric N is defined as $C^m := QD^mQ'$, where the power of a diagonal matrix will always be taken to be the result of taking the positive diagonal elements to that power. In particular, $C^{1/2}$ defined in this way satisfies the equation $C^{1/2}NC^{1/2} = C$. Unfortunately, the power C^{-1} corresponds to the usual definition of an inverse only in the metric I; although $C^{-1}NC = C^0$, $C^0 \neq I$ and $C^{-1}C \neq I$ in general. Wherever the power -1 is used the metric involved will be made explicit. The set of singular values of A and the diagonal matrix D will be referred to as the *spectrum* of A. The singular value decompositions of two matrices A and B will be denoted by $A = P_A D_A Q'_A$ and $B = P_B D_B Q'_B$, respectively.

The matrix correlation problem requires us to find matrices L and M subject to the necessary orthogonality constraints, and then to measure in some way the correlation $r(XL, YM)$ between the transformed matrices. We have essentially two choices to resolve: how shall we define the mappings L and M subject to the required orthogonality constraints, and what correlation function $r(\cdot, \cdot)$ shall we use? It may be that these problems are related; it is common to first choose a correlation measure and then optimize it with respect to the mappings. However other strategies are feasible, and therefore these two steps are treated independently in the next two sections.

Correlational Measures for Matrices

A function $r(\cdot, \cdot)$ must satisfy the following necessary conditions before it can be considered a useful measure of matrix correlation (Renyi, 1959):

Definition 1: The mapping $r\colon \mathscr{R}^{ns} \times \mathscr{R}^{ns} \rightarrow [0, 1]$ is called a correlation function in the metric W if for all nonzero scalars x and y, and when A and B are not both zero,

$$\begin{aligned} &\text{C1.} \quad r(xA, B) = r(A, yB) = r(A, B) \\ &\text{C2.} \quad r(A, B) = r(B, A) \\ &\text{C3.} \quad r(A, B) = 1 \quad \text{if} \quad A = yB \\ &\text{C4.} \quad r(A, B) = 0 \quad \text{iff} \quad A'WB = 0. \end{aligned} \tag{1}$$

In practice a correlational measure may be a mapping into $[-1, 1]$, and for such a measure these conditions are understood to apply to its absolute value or its square.

These conditions are necessary but not sufficient to ensure a useful measure of matrix correlation. In fact, matrices may be similar or dissimilar in a great many ways, and it is desirable in practice to capture some aspects of matrix relationships while ignoring others. This can take the form of specifying the converse of C3 by devising further situations leading to unit correlations.

Inner Product Correlation r_1

The most common measure of matrix correlation is based on the inner product $(A, B) := \operatorname{tr}(A'B)$. Because of the Cauchy-Schwarz inequality the absolute value of the correlation coefficient

$$\begin{aligned} r_1(A, B) &:= \operatorname{tr}(A'B)/[\operatorname{tr}(A'A) \operatorname{tr}(B'B)]^{1/2} \\ &= (A, B)/(\| A \| \, \| B \|) \end{aligned} \tag{2}$$

satisfies C1–C4. Expressed in terms of the singular value decompositions (svd's) of A and B,

$$r_1(A, B) = \mathrm{tr}\,(Q_A D_A P'_A P_B D_B Q'_B)/[\mathrm{tr}\,(D_A^2)\,\mathrm{tr}\,(D_B^2)]^{1/2} \tag{3}$$

where the powers of the diagonal matrices D_A and D_B are taken in the identity metric. When general metrics W and N are involved, $(A, B) = \mathrm{tr}\,(A'WBN)$ and the svd is (W, N)-orthogonal. Note that this measure is invariant with respect to postmultiplication of each matrix by a rotation matrix T only when $N = I$. The expression of r_1 in terms of the svd's of the two matrices helps to show that its value depends on three aspects of each matrix:

(i) the orthonormalized matrix P which is the most basic component,
(ii) the spectrum or scale component D, and
(iii) the orientation or correlational component Q.

This measure has been used explicitly or implicitly in almost all work on matrix correlation. It will be shown in the following section that a variety of other measures are equivalent to it under certain circumstances. Lingoes and Schönemann (1974) discuss r_1 in the context of assessing a comparison of X and Y by Procrustes rotation. It will be shown in the next section that this process does not in general maximize r_1, however.

Orientation-Independent Inner Product Correlation

The dependency of r_1 on the relative orientations of A and B can be removed by deleting Q_A and Q_B from the trace in the numerator:

$$\begin{aligned} r_2(A, B) &:= \mathrm{tr}\,(D_A P'_A P_B D_B)/[\mathrm{tr}\,(D_A^2)\,\mathrm{tr}\,(D_B^2)]^{1/2} \\ &= (P_A D_A, P_B D_B)/(\|A\|\,\|B\|) \\ &= r_1(P_A D_A, P_B D_B). \end{aligned} \tag{4}$$

This index is equivalent to computing r_1 if either both A and B have orthogonal columns or $Q_A = Q_B$. It reaches unity not only when the two matrices are proportional, but also when they can be made so by post-multiplication of either by a rotation matrix T. In the general metric case T is replaced by $N^{1/2}T$, where N-orthogonal and $N^{1/2}$ is taken in the identity metric.

Spectra-Independent Inner Product Correlations

The size of the inner product correlation r_1 will depend on whether the two sets of singular values or spectra, D_A and D_B, are similar or not. Two further indices can be derived from r_1 and r_2 by replacing the spectra by identity matrices:

$$\begin{aligned} r_3(A, B) &:= s^{-1}\,\mathrm{tr}\,(Q_A P'_A P_B Q'_B) \\ &= s^{-1}(P_A Q'_A, P_B Q'_B) \\ &= r_1(P_A Q'_A, P_B Q'_B) \end{aligned} \tag{5}$$

and

$$\begin{aligned} r_4(A, B) &:= s^{-1}\,\mathrm{tr}\,(P'_A P_B) \\ &= s^{-1}(P_A, P_B) \\ &= r_1(P_A, P_B). \end{aligned} \tag{6}$$

Coefficient r_3 can also be expressed as

$$r_3(A, B) = s^{-1} \text{ tr } [A'A)^{-1/2}(A'B)(B'B)^{-1/2}], \tag{7}$$

since $(A'A)^{-1/2} = (Q_A D_A P'_A P_A D_A Q'_A)^{-1/2} = Q_A D_A^{-1} Q_A^1$

Escoufier's RV Coefficient

Escoufier (1973, 1977; Robert & Escoufier, 1976) has proposed

$$\begin{aligned} \text{RV} &:= \text{tr } (B'AA'B)/\{\text{tr } [(A'A)^2] \text{ tr } [(B'B)^2]\}^{1/2} \\ &= r_1(P_A D_A^2 P'_A, P_B D_D^2 P'_B) \end{aligned} \tag{8}$$

This coefficient is closely related to r_2 in that all traces in RV are taken with respect to the squares of the matrices involved in the corresponding traces in r_2. Thus, this index is also orientation-independent and provides a technique for assessing correlation in the same sense as r_2. However, RV has the advantage of being expressible in terms of the original matrices rather than their singular value decompositions.

Yanai's GCD Measure

Yanai (1974) has proposed the trace of the product of two orthogonal projectors derived from A and B as a global measure of relationship. Converted to a correlation coefficient, it is

$$\begin{aligned} \text{GCD} &:= s^{-1} \text{ tr } (A(A'A)^{-1}A'B(B'B)^{-1}B') \\ &= r_1(P_A P'_A, P_B P'_B). \end{aligned} \tag{9}$$

This measure is also the average of the squared canonical correlations between A and B, and thus it is not surprising to discover that it is orientation-spectra-independent. In the general metric case for both RV and GCD we can have different column metrics, N_1 and N_2, for A and B, respectively.

Thus we have in r_1–r_4 a natural and complete system of coefficients completing the two-by-two table corresponding to orientation and spectrum independence. In each case the coefficient can be expressed in terms of the singular value decomposition, and in the case of r_2 and r_4 there associated coefficients, RV and GCD, respectively, which are expressible in terms of the original matrices. In the special case of $s = 1$, these coefficients are related as follows:

$$r_1 = r_2 = r_3 = r_4, \qquad r_4^2 = \text{RV} = \text{GCD}.$$

Many other families of correlation measures are possible which give varying sensitivities to the orientations and spectra of the two matrices. Since multiplication of a matrix by a scalar affects only its spectrum, one could in principal use any positive power of the singular values in r_1 or r_2 to provide a continuum of sensitivity to spectrum. Alternatively, if $X = PDQ'$, then we may use $X^* = PDHQ'$ as an argument in r_1, where H is a diagonal matrix containing positive weights to be applied to the singular values.

It should be noted that the components P_A and Q_A of the singular value decomposition of A are defined only to within mutual sign changes in any column even when the singular values are distinct. Thus, in computing coefficients r_2 and r_4 it will be necessary to ensure that corresponding columns of P_A and P_B are sign-compatible. This can be achieved by reversing the sign in any column of P_A when the inner product of that column with its mate in P_B is negative.

Some Orthogonal Transformations of X and Y

The complement to the problem of choosing a correlation index is the problem of mapping $\mathscr{R}^{np}$ and $\mathscr{R}^{nq}$ into the comparison space $\mathscr{R}^{ns}$ in which the index is to be applied, so that the correlation measure can be applied to XL and YM. We assume here that only linear transformations or orthogonal transformations are to be entertained, and in particular that the two matrices are comparable with respect to their column origins. This would automatically be the case for factor pattern matrices, but analyses involving multidimensional scaling configurations or arbitrary variables may require preliminary column centering.

In order to avoid trivial mappings, it is necessary to specify both normalization conditions and orthogonality conditions. If these are not determined a priori, then the choice is a matter of convenience. In general we shall wish to impose:

$$\text{D1.} \quad L'UL = I,$$

$$\text{D2.} \quad M'VM = I,$$

and possibly

$$\text{D3.} \quad L'UJ = 0, \qquad J \in O_U^{pt},$$

$$\text{D4.} \quad M'VK = 0. \qquad K \in O_V^{qt}.$$

where $t \leq \min(\text{rank}(X), \text{rank}(Y)) - s$, and U and V are metric matrices. Constraints D1 and D2 are internal orthogonality constraints while D3 and D4 are constraints with respect to previously computed or observed matrices J and K and thus are external. Matrices J and K can always be expressed in orthogonal form if they are not already thus. That they should have the same number of columns is not necessary in general, but will hold in most practical situations and is assumed for simplicity.

In order to keep the exposition as simple as possible the theorems in this section will be proved only for constraints D1 and D2, and assuming that W and N are identity matrices. However, in some situations where the extension to D1–D4 is not obvious, the more general results will also be stated. Associated with the external constraints are the projectors

$$H_J := I - JJ'U \text{ and } H_K := I - KK'V, \tag{10}$$

which satisfy $H_J L = L$, $H_J J = 0, H_K M = M$, and $H_K K = 0$.

Clipping Transformations

One simple class of transformations are those which map X and Y into subsets of their columns. These are valuable when X and Y are already defined with respect to rotation, perhaps by previous transformation to principal axis or by optimizing some analytic rotation criterion. It may be then that we simply wish to compare X and Y in terms of certain columns while ignoring differences on others. The matrix $C_s \in \mathscr{R}^{ps}$ defined as differences on others. The matrix $C_s \in \mathscr{R}^{ps}$ defined as the first s columns of the identity matrix will be called a clipping matrix, and is I-orthogonal. The number of rows of C_s will be obvious from context.

Maximizing the Inner Product

There are a variety of reasons why one might wish to transform so as to maximize the inner product $(XL, YM) := \text{tr}(L'X'WYMN)$. Under some circumstances this is equivalent to maximizing the correlation r_1, and in any case the transformations maximizing the inner product can provide useful starting values for iterative procedures for maximizing a number of correlational measures.

The following lemma permits a simple extension of the results for constraints D1 and D2 to those for D1–D4.

Lemma 1: The maximal value of (XL, YM) under constraints D1–D4 is equal to the maximal value of $(XH_J L, YH_K M)$ under constraints D1–D2.

Proof: Let L^* and M^* be the optimal values for $(XH_J L, YH_K M)$ under D1–D2. L^* and M^* can be decomposed into the sums $L^* = L_1^* + L_2^*$ and $M^* = M_1^* + M_2^*$ where L_2^* and M_2^* are in the subspace spanned by J and K, respectively, and L_1^* and M_1^* are in the respective orthogonal complements of these subspaces. Substitution of these sums into $(XH_J L^*, YH_K M^*)$ shows that $(XH_J L^*, YH_J M^*) = (XL_1^*, YM_1^*)$. Now suppose that there is a pair of transformations L^0 and M^0 satisfying D1–D4 such that $(XL^0, YM^0) > (XL_1^*, YM_1^*)$. Then $H_J L^0 = L^0$ and $H_K M^0 = M^0$ imply that $(XL^0, YM^0) = (XH_J L^0, YH_K M^0) > (XH_J L^*, YH_K M^*)$. But this contradicts the hypothesis made about L^* and M^* and thus the lemma follows. □

The following very important lemma states a result first proved by von Neumann (1937).

Lemma 2: For $A, B \in \mathscr{R}^{ns}$ with singular values $\alpha_i \geq \cdots \geq \alpha_s$ and $\beta_i \geq \cdots \geq \beta_s$, respectively, tr $(A'B) \leq \sum \alpha_i \beta_i$.

Proof: The argument proceeds by (i) proving Abel's Identity, (ii) proving the result for A and B symmetric, and (iii) extending the result to general matrices. Since the Lemma states the Cauchy-Schwarz inequality when $s = 1$, it will be assumed that $s > 1$.

(i) The following identity may be verified at once for $s = 2$:

$$\sum_{i=1}^{s} a_i b_i = \sum_{i=1}^{s-1} \left[(a_i - a_{i+1}) \sum_{j=1}^{i} b_j \right] + a_s \sum_{j=1}^{s} b_j. \tag{11}$$

When $s > 2$,

$$\begin{aligned}
\sum_{1}^{s} a_i b_i &= \sum_{i=2}^{s-1} a_i b_i + a_1 b_1 + a_s b_s \\
&= \sum_{i=2}^{s-1} a_i \left(\sum_{j=1}^{i} b_j - \sum_{j=1}^{i-1} b_j \right) + a_1 b_1 - a_s \sum_{j=1}^{s-1} + a_s \sum_{j=1}^{s} b_j \\
&= \sum_{i=1}^{s-1} \left(a_i \sum_{j=1}^{i} b_j \right) - \sum_{i=2}^{s} \left(a_i \sum_{j=1}^{i-1} b_j \right) + a_s \sum_{j=1}^{s} b_j \\
&= \sum_{i=1}^{s-1} \left(a_i \sum_{j=1}^{i} b_j \right) - \sum_{i=1}^{s-1} \left(a_{i+1} \sum_{j=1}^{i} b_j \right) + a_s \sum_{j=1}^{s} b_j \\
&= \sum_{i=1}^{s-1} \left[(a_i - a_{i+1}) \sum_{j=1}^{i} b_j \right] + a_s \sum_{j=1}^{s} b_j.
\end{aligned}$$

(ii) Now let A and B be symmetric, of order s, and with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_s$, $\mu_1 \geq \cdots \geq \mu_s$, λ_s and μ_s possibly negative. Then $A = P\Lambda P'$, $P \in O_I^{ss}$, and tr $A'B =$ tr $P\Lambda P'B =$ tr $\Lambda P'BP = \sum \lambda_i b_{ii}$, where b_{ii} is the ith diagonal element of PBP'. The majorization $\sum_i^k b_{ii} \leq \sum_1^k \mu_i$, $k \leq s$, results from the inequality

$$\text{tr } L'BL \leq \sum_1^k \mu_i, \qquad L'L = I$$

(Rao, 1973, p. 63). Thus, by Abel's Identity (i) and the fact that B and PBP' have the same eigenvalues

$$\text{tr } A'B = \sum_{i=1}^{s} \lambda_i b_{ii} = \sum_{i=1}^{s-1} \left[(\lambda_i - \lambda_{i+1}) \sum_{j=1}^{i} b_{jj} \right] + \lambda_s \sum_{j=1}^{s} b_{jj} \leq \sum_{i=1}^{s} \lambda_i \mu_i.$$

(iii) For $A, B \in \mathscr{R}^{ns}$, $n \geq s$, with singular values as stated in the lemma, the corresponding symmetric matrices of order $n + s$,

$$A^* := \begin{bmatrix} 0 & A' \\ A & 0 \end{bmatrix} \quad \text{and} \quad B^* := \begin{bmatrix} 0 & B' \\ B & 0 \end{bmatrix},$$

are symmetric, of order $n + s$, and have nonzero eigenvalues $\pm\alpha_i$, $\pm\beta_i$, $i = 1, \ldots, s$. The inequality then follows from the fact that

$$\text{tr } A'B = \tfrac{1}{2} \text{ tr } A^*B^* \leq \tfrac{1}{2} \sum [\alpha_i \beta_i + (-\alpha_i)(-\beta_i)] = \sum \alpha_i \beta_i. \qquad \square$$

Lemma 2 gives a upper bound for (A, B) which is much tighter than $\| A \| \; \| B \|$ given by the Cauchy-Schwarz inequality when $s > 1$, and which is attainable under orthonormal transformations. It has many important consequences. The first proof was by von Neumann (1937) with later proofs by Richter (1958) for symmetric matrices, Mirsky (1959), and Theobald (1975). Fan (1951) and Kristof (1970) extended the result to more than two matrices, with the latter applying the result to a number of multivariate data analytic problems. ten Berge (1983) extended Kristof's results and presented a number of applications. This proof, which appears in print for the first time, seems to the authors to be much shorter and simpler than previous arguments and is due to Styan (1976).

The following theorem subsumes a number of well known results:

Theorem 1. The following values of L and M maximize (XL, YM) and satisfy constraints D1 and D2:

$$L = PC_s \quad \text{and} \quad M = QC_s, \tag{12}$$

where P and Q are defined by the singular value decomposition

$$X'Y = PDQ' \tag{13}$$

Proof: From Lemma 2 we have

$$\max_{L,\, M} \{\text{tr } L'X'YM\} = \max_{L^0, M^0} \{\text{tr } [C_s' L^{0\prime} X'YC_q' M^0 C_s]\}$$
$$\leq \text{tr } (C_s' DC_s),$$

where L^0, $M^0 \in O_{pp}$. It remains to note that this bound is attained for L and M chosen as indicated. $\square$

Corollary: When M is fixed, $L = PQ'$.

In the more general case, $L = PC_s Q_N'$ and $M = QC_s Q_N'$ where P and Q are defined by the (U, V)-orthogonal svd of $U^{-1}H_J' X'WYH_K V^{-1}$ and the decomposition of N and the inverses of U and V are taken in the identity metric. The simpler theorem proven above has been discussed by many authors including Green (1952), Cliff (1966), and Schönemann (1966). ten Berge (1983) has extended Theorem 1 to more than two matrices and given a number of applications.

Maximizing the Inner Product Correlation r_1

It is natural to search for those transformations which maximize the measure of correlation to be used. The problem for coefficient r_1 in its complete generality does not have a closed-form solution. However, r_1 is maximized by maximizing (XL, YM) in the special cases specified in the following theorem:

Theorem 2. If one of the following conditions hold then r_1 is maximized by maximizing (XL, YM):

(a) constraints D1–D4 apply and $U = X'WY$ and $V = Y'WY$,
(b) constraints D1 and D2 apply and $s = 1$,
(c) $p = q = s$ and $N = U = V = I$.

Proof: Norms $\| XL \|$ and $\| YM \|$ are invariant with respect to L and M under the metric defined in case (a) or in case (c). In case (b) the constraints D1 and D2 impose only a scale constraint on L and M. The correlation r_1 is invariant under changes of scale and hence the appropriate L and M can be computed using the metrics $X'WX$ and $Y'WY$ and then rescaling. □

Case (c) has been discussed by many authors, especially in the context of Procrustes matching. This literature is reviewed in ten Berge (1977).

The general problem requires the use of numerical optimization. The Appendix details an algorithm for the optimization of r_1 or the minimization of $\| XL - YM \|$ which exploits Theorem 1 and has proved satisfactory in applications.

Maximizing Orientation-Independent Correlations

In the case of $p = q = s$ and $U = V = I$ coefficients r_2, r_4, RV, and GCD are attractive because they are invariant with respect to rotations. In other situations, however, their behavior with respect to various transformations is more complex. The following theorem states that among the class of optimal transformations is a pair resulting from clipping after some permutation of the columns of the principal axis oriented matrices.

Maximizing Spectra-Independent Correlation r_3

Coefficient r_3 can be readily optimized in the case $p = q = s$ by noting that the singular values of XL and YM will be the same as those of X and Y, respectively. Since r_3 does not depend on the singular values, maximizing it reduces to maximizing $(P_X Q_X' L, P_Y Q_Y' M)$ and therefore Theorem 1 applies.

Some Orthogonal-Linear Transformations

Some situations impose orthogonality conditions on only one of the transformations, which will be assumed to be X. In this case any linear transformation M may be applied to Y. For example, X may be the consequence of a p-dimensional Euclidean multidimensional scaling analysis while Y may contain a set of q measurements on the stimuli obtained in some other way and without any prior conditions on the correlations among them.

Some overall scale constraint must be imposed on M in order to achieve identifiability. The following is used here:

$$\text{D5.}\quad \operatorname{tr}(M'Y'WYMN) = \| YM \|^2 = s.$$

The inner product (XL, YM) is maximized as specified in the following theorem.

Theorem 3. If constraints D1 and D5 hold and $Y'Y$ is nonsingular, then the maximum of (XL, YM) is given by

$$M \propto (Y'Y)^{-1}Y'XL,$$

and

$$L = I \quad \text{if} \quad p = s \quad \text{or} \quad L = QC_s \quad \text{if} \quad p > s,$$

where Q is given by the svd

$$X'Y(Y'Y)^{-1}Y'X = QDQ'. \tag{15}$$

Proof: For any L, M as defined can easily be shown to be a maximizing value of M which yields a finite maximum. Thus $(XL, YM) \leq (XL, Y(Y'Y)^{-1}Y'L)$ and it remains to apply Lemma 2 to show that $(XL, Y(Y'Y)^{-1}Y'L) \leq \text{tr}\,(D)$ and that this bound is attained when L is as defined. □

Corollary: When $q = s = 1$, L is proportional to $X'Y$ and $M = 1$.

The extension of Theorem 3 to the general case is along the same lines as the extension of Theorem 1.

The general problem of maximizing correlation r_1 in the orthogonal-linear case must also be dealt with numerically. The algorithm described in the Appendix can also deal with this situation.

An Example: Evans' Factor Patterns

Evans (1971) in his extensive discussion of matrix comparison procedures provides a number of factor pattern matrices resulting from the extraction of six factors from observation on 18 variables. One set of observations was taken from Canadian sixth grade children, and this factor matrix is to be compared from data on Filipino children at the same level. The two matrices are displayed in inside-out format (Ramsay, 1980; Wainer & Thissen, 1981) in Table 1. Evans' paper can be consulted for the original numbers.

Comparing Correlational Measures for $s = 6$

The first analysis involves calculating all of the coefficients discussed in the section on orthogonal-orthogonal correlation for a six-dimensional comparison. In this case $p = q = s$ and the optima of all measures can be obtained without recourse to numerical methods. Table 2 displays these correlations. The maximal r_1 value (.80) and Escoufier's RV (.82) give similar results and indicate a high degree of agreement between the two matrices. Coefficient r_2 (.61), however, which is equivalent to r_1 when the two matrices have been rotated to principal axis orientation, is somewhat lower suggesting that this orientation does not make the matrices as comparable. Similarly, removing the effect of the singular values by computing coefficient r_3 (.71) also lowers the correlation. Coefficient r_4 (.36) compares the matrices in principal axis orientation and with spectrum removed and is thus much lower, as is coefficient GCD (.58) This suggests that there is strong agreement in the directions corresponding to dominant singular values since the spectra play an important role, but that agreement associated with subdominant singular values is poor giving rise to low correlations when the matrices are in effect put in principal axis form.

Comparing Values of r_1 for Different Values of s

The next step is to consider the correlation between the two matrices for each possible dimension of comparison. Figure 1 displays the maximized conventional inner product correlation r_1 along with the correlation resulting from maximizing (XL, YM) for

Table 1. Two Varimax-Rotated Factor Patterns from Evans (1971) Displayed in Inside-Out Format

Factor Loading	Canadian 1	2	3	4	5	6	Filipino 1	2	2	4	5	6
1.00												
.97												
.93												
.90								E				
.87												
.83	B	FE	D		M		B			G	K	
.80								F				
.77							A				L	
.73	A											
.70				KJ								
.67						RQ						
.63					O							
.60				L							M	
.57							N					
.53			C				J				O	R
.50									C			
.47		PH					P	M				Q
.43	N							J			J	
.40								P				
.37								NO		H	C	
.33		L		I	J		DC	H	H	N	P	
.30		J		O	PG		Q	R				
.27	H					P		KC	D	Q	DN	
.23	CJ	I	G	HPC		J		AD		IOM	R	
.20	OD	BG	A	Q	NHE	C	E			RD	IFE	I
.17		O				N		Q		P		
.13	FE	RADM	BJ	FA	A	E	I		EK	C		KGB
.10	L	Q	IOEK	REGB	I	K		B	L	A		EOJ
.07	P	N	RNF	D	D			L	AB		HG	CF
.03	KM		MLP	NM	----	OLF	OG		RGNF	LF	A	
.00	R	K	Q		B	AGDM				JKE	QB	AL
-.03	QI				K	B		G	Q			
-.07	G	C										DPH
-.10									J	B		M
-.13									OP			
-.17								I				
-.20			H			I			M			
-.23												N
-.27						H						
-.30												
-.33									I			
-.37												
-.40												

The value of a factor loading for a variable is given by the row in which the letter corresponding to that variable is located. Dashes indicate more symbols than can fit in a particular position. The correspondence between letters and variables is as follows:

A Verbal meaning	G Identical pictures	M Letter series
B Vocabulary	H Maze tracing	N Word grouping
C Word endings	I Finding A's	O Number series
D Word beginnings	J Arithmetic	P Raven's matrices
E Spatial relations	K Subtraction, multiplication	Q Picture-number
F Card relations	L Division	R Object-number

Table 2. Correlational Measures for Factor Patterns in Table 1 Assessed for the Optimal r Transformations and s = 6

Correlation	r_1	RV	r_2	r_3	r_4	GCD
Original Estimate	.80	.82	.61	.71	.36	.58
Jacknifed Estimate	.67	.79	.45	.70	-.10	.48

each possible comparison dimension s. Here one notices that the correlation is very high in up to three dimensions and then begins to fall off fairly rapidly, indicating again that the two factor solutions are very similar in their dominant modes of variation but have significant differences in their minor modes of variation. Note, too, that the maximizing of the inner product falls well short of providing the maximum inner product correlation. Figure 2 displays the factor solutions matched in two dimensions by maximizing r_1.

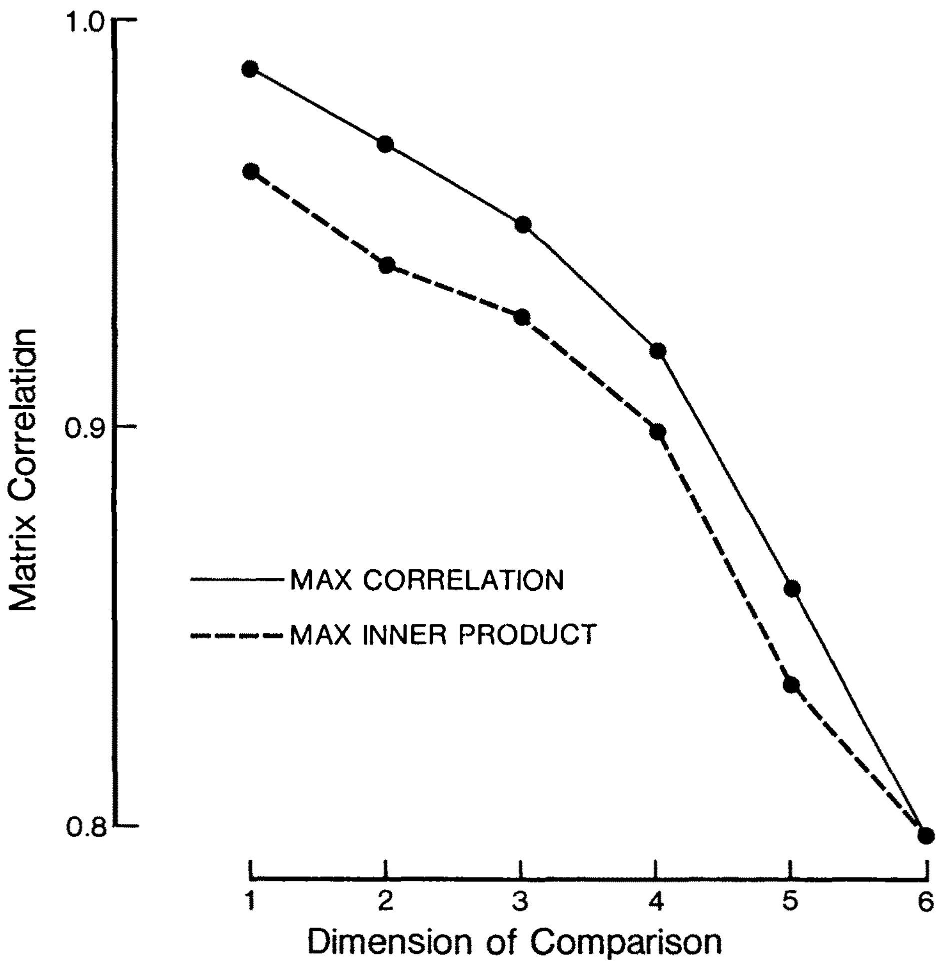

FIGURE 1
The points joined by a solid line indicate the size of the inner product correlation r_1 when maximized in each possible dimensionality for the Evans data. The points joined by a dashed line indicate the values of r_1 obtained by maximizing the inner product.

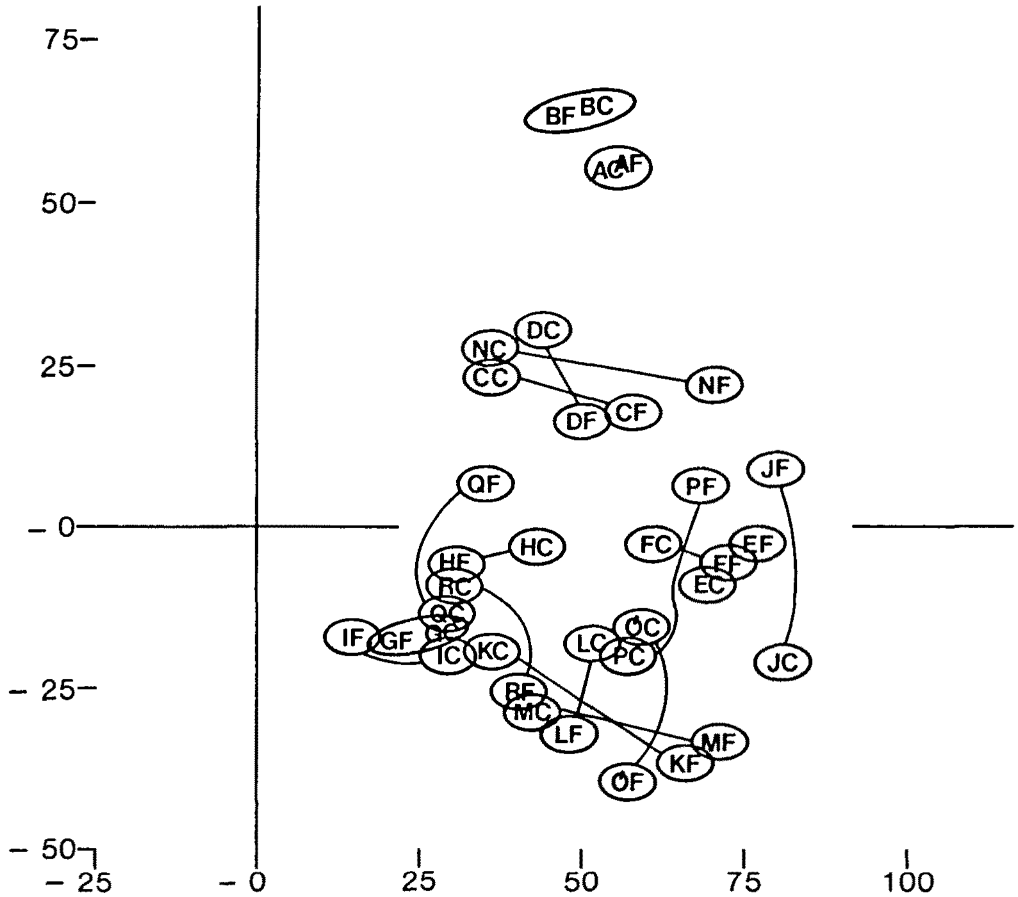

FIGURE 2
A comparison of the factor loadings for the Canadian (C) and Filipino (F) data after rotation to maximize the inner product correlation r_1 in two dimensions. The first letter in each point label identifies the variable as specified in Table 1. Decimal points have been suppressed in the axis labels.

Jacknifing to Assess Stability

The stability of these coefficients is naturally an important question. The techniques of jacknifing and bootstrapping suggest themselves in this context. If one had access to the original data, these procedures could be applied in the usual way. However, one may also jacknife by eliminating each pair of corresponding rows in turn. The results cannot be interpreted in a statistical sense, since the results in one row certainly would be different if any other row (variable) were eliminated from the original design. However, this process should nonetheless be able to give valuable clues about the extent to which correlations depend on the values in a particular row and some overall indication of the relative stability of the measures. Moreover the jacknifed estimates of the correlational measures will to some extent correct for the natural bias in these measures resulting from their optimization. The pseudovalues for measure r are given by $r_i = nr_0 - (n-1)r_{-i}$, $i = 1, \ldots, n$, where r_0 is the measure for all rows present and r_{-i} is the measure for the ith row removed. Table 3 displays the pseudo-values for each measure for the full six-dimensional comparison in inside-out format. From this it is clear that the rows corresponding to the tests "maze tracing" and "finding A's" exert a strong influence on measures r_1, r_3, and GCD. The fact that their pseudo-values are low for these measures indicates removing

Table 3. Pseudo-values for Correlational Measures

Pseudo-value	r_1	RV	r_2	r_3	r_4	GCD
3.0						
2.9						
2.8					A	
2.7						
2.6			A			
2.5						
2.4						
2.3						
2.2						
2.1						
2.0						
1.9						
1.8						
1.7			B			
1.6						
1.5					B	
1.4			J			
1.3		J				
1.2				Q		
1.1		EA		RE		
1.0	E	PO		KDF		
.9	FAB	FB		CAG	J	QER
.8	JRKL	DLRNC	P	BL		KDFA
.7	QOPCDN	QGM	R	J	R	C
.6	MG	KI	NE	OPM		GB
.5		H	I	N		M
.4			DFG		G	OLJ
.3	I		QC			P
.2	H		L		IP	N
.1					N	
.0				I		
-.1				H	DQCE	
-.2			H		F	I
-.3			O			
-.4			K		L	H
-.5						
-.6						
-.7						
-.8						
-.9			M			
-1.0						
-1.1					H	
-1.2						
-1.3						
-1.4						
-1.5					K	
-1.6						
-1.7					O	
-1.8						
-1.9						
-2.0						
-2.1						
-2.2						
-2.3						
-2.4					M	
-2.5						

Table 4. Row Weights (x100) for Robust Analyses of r_1

Variable	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Original Data	98	95	89	90	98	97	88	76	79	84	87	84	84	84	89	84	90	95
Canadian A Loadings x 10	00	93	89	90	98	98	88	79	79	98	86	94	82	84	92	86	96	96

these rows increases the measures sharply and thus that these rows tend to depress the overall correlation. Similarly, the test "verbal meaning" has a strong but opposite impact on measures r_2 and r_4; that is, removing this row decreases the correlation indicating that the overall correlation is strongly dependent on this test. Table 3 also permits a visual inspection of the variability in the pseudo-values for each measure and it is clear that both r_1 and RV are highly stable while r_2 and r_4 are very unstable. The average pseudo-values give the jacknifed estimate of the correlation and are given in Table 2. Jacknifing in this way obviously shrinks all correlations substantially.

Using Metric W to Achieve Robustness

One of the considerations that may lead one to use something other than the identity matrix for W is the desire to weight rows less heavily which are associated with markedly poor congruences. In order to illustrate how this can be done, a robust analysis of the correlation r_1 between these matrices in six dimensions was used along the lines suggested by Ramsay & Novick (1980). It proceeded as follows:

1. Compute L and M optimizing r_1, $E = XL - YM$, and assess r_1.
2. For each column of E compute
$$s_j^2 = \underline{(n-4)}^{-1} \sum_{i=3}^{n-2} e_{(i)j}^2,$$
where $e_{(i)j}$ is the ith order statistic in column j. This gives a trimmed estimate of variance for each column of E.
3. Compute $d_i^2 = \sum_i e_{ij}^2/s_j^2$ and $w_{ii} = \exp(-.02 d_i^2)$, $i = 1, \ldots, n$. This defines the diagonal metric matrix W.
4. Assess r_1 in the metric W. Check to see if it differs substantially from the previously assessed value. If so, recompute E and return to step 3.

This algorithm will converge quickly to a metric which will apply weights of nearly unity to rows of XL and YM which have a reasonable congruence as defined by the standardized distance measure d_i, and will apply reduced weight to rows having substantially poorer congruences. The constant .02 controls the rate at which rows are deweighted for large distances. In general, an appropriate value for this constant is $1/(2\chi^2_{.001})$, where $\chi^2_{.001}$ is the 0.001 critical value of chi squared with p degrees of freedom (Ramsay, 1980).

The algorithm converged in 3 iterations using a criterion of a change of .0001 in r_1 from one iteration to the next, and produced a final value of 0.81. The final weights associated with each row are given in Table 3. The somewhat lower weights associated with the variables H and I are one more indication that these variables are tending to reduce the value of r_1. Also in Table 3 are the weights associated with the robust analysis after multiplying the variable A loadings for the Canadian factors by 10. Again the algorithm converged in three iterations and moved from an initial correlation of .50 to a final value of .80.

Conclusions

The problem of correlating two matrices presents many aspects. One must decide what invariances the correlation is to have, with invariances with respect to rotation, spectrum change, and linear transformation all being possibly desirable. In general a particular kind of invariance can be achieved in one of two ways. The first is the use of a correlational measure which is intrinsically invariant in the desired way. For example, if rotational invariance is required, then one might choose RV, r_2, or r_4. The second procedure is to optimize a correlation over the class of transformations with respect to which invariance is desired. The optimized correlation is then automatically invariant, ignoring possible local optimum problems. Thus, the problem of optimizing r_1 with respect to rotations and linear transformations has been considered. This second approach has the advantage that optimizing transformations themselves may be of considerable interest, but is obviously more expensive in general. It appears that coefficient RV should be better known as a generally useful way of comparing two matrices defined to within either rotations (when it is automatically invariant) or to within linear transformations (when it can be optimized simply).

An aspect of matrix correlation that also deserves more attention in the authors' opinion is the possibility of comparing the matrices in a subspace of dimension s. Two-dimensional comparisons have obvious value from a graphical perspective. As the example illustrated, it may also be desirable to try out a range of subspaces to determine a dimensionality for the comparison which is close to that for $s = 1$ while being substantially less than $s = \min(p, q)$. The incorporation of the external orthogonality constraints D3 and D4 into the theorems on optima makes it possible to carry out a series of comparisons for a fixed value of s, such as 2, which are mutually orthogonal.

Finally, the results in this paper have depended heavily on the singular value decomposition. This essential tool has a number of generalizations which can be used to widen the scope of the results in this paper. For example, its counterpart for continuous-time stochastic processes is known as the Karhunen-Loeve decomposition, while versions also exist for the more general class of compact continuous linear mappings from one Hilbert space into another.

Appendix

The problem of optimization of a function with respect to orthogonal matrices arises very often in multivariate data analysis and frequently cannot be solved analytically. Numerical methods have been developed for many special cases, especially in the psychometric and factor analytic literature. The algorithm described here is in the spirit of most of these approaches in that the function is optimized in each possible plane in turn.

When the orthogonal matrix L is p by s, $p > s$, these rotations fall into three subsets. The first contains rotations within the space spanned by L, which can be taken to be within planes defined by two different columns of L. The second contains rotations in planes defined by a column of L and one in its orthogonal complement L^* defined by the relation $L : L^* \in O_I^{pp}$. The third contains rotations in planes entirely within the space spanned by L^*. Since these latter do not affect the function being optimized, only the first two sets are of importance. The first set is defined by the class of transformations LH, $H \in O_I^{ss}$, and optimization with respect to such rotations reduces to optimization with respect to H. Fortunately, it is often possible to find the optimal H in a single step; this is the case for r_1 and for the Procrustes problem, for example. For such problems, planar rotations are required only for planes involved in the second set; that is, for rotations of vectors with images in the column spaces of both L and L^*. It is worthwhile, therefore, to consider the sets separately.

An algorithm which has worked well in practice proceeds as follows for a function $F(L, M)$, L subject to constraint D1 and M subject to either D2 or D5:

0. Compute an appropriate pair of initial values $L^{(0)}$ and $M^{(0)}$.
1. Modify L on iteration v as follows:
   1.1. Optimize F with respect to rotations of the first kind.
   1.2. For each pair (i, j), $1 \le i \le s$, $1 \le j \le p - s$, let $L^{(v+1)}$ be defined as $\cos\theta\, L_i^{(v)} + \sin\theta\, L_j^{*(v)}$ and carry out a single optimization step regarding F as a function of θ. This might employ Newton's method from an initial value of $\theta = 0$ with a possible reduction in step size in order to ensure an improvement in F.
2. Modify M as was done for L if M is subject to D2 or as appropriate if D5 applies.
3. Test for convergence. If successful, exit; otherwise return to 1.

This algorithm is expressed in Figures 3 and 4 using the PROC MATRIX language of the Statistical Analysis Systems (SAS). Since the syntax of this matrix-oriented language is more or less self-explanatory, this procedure can be used as a pattern for developing algorithms in other languages as well as being directly executable in SAS. For further details on PROC MATRIX the manual should be consulted (SAS Institute, 1982).

The procedure is written as a SAS MACRO so that the essential parameters can be passed to the procedure through a call to the macro. It should be noted, however, that the single quote in the program below is not processed properly by version 82.4 or previous versions of SAS, and must be replaced by % STR (%') prior to execution. The DQUOTE option is also required.

REFERENCES

Berge, J. M. F. ten (1977). Optimizing factorial invariance. Doctoral dissertation, University of Groningen, Netherlands.

Berge, J. M. F. ten (1983). A generalization of Kristof's theorem on the trace of certain matrix products, *Psychometrika, 48*, 519–523.

Besse, P. (1979). Etude descriptive d'un processus. Thèse de troisième cycle, l'Université Paul-Sabatier de Toulouse, France.

Cailliez, F. & Pages, JP. (1976). *Introduction à l'Analyse des Données.* Paris: Société de Mathématiques Appliquées et de Sciences Humaines, 9 rue Duban, 75016 Paris.

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

Cramer, E. & Nicewander, W. A. (1979). Some symmetric, invariant measures of multivariate association, *Psychometrika, 44*, 43–54.

Dauxois, J. & Pousse, A. (1976). Les analyses factorielles en calcul des probabilités et en statistique: Essai d'étude synthètique. Thèse d'état, l'Université Paul-Sabatier de Toulouse, France.

Escoufier, Y. (1973). Le traitment des variables vectorielles, *Biometrics, 29*, 751–760.

Escoufier, Y. (1977). Operators related to a data matrix. In J. R. Barra, et al, (Eds.) *Recent Developments in Statistics.* Amsterdam: North-Holland.

Evans, G. T. (1971). Transformation of factor matrices to achieve congruence, *British Journal of Mathematical and Statistical Psychology, 24*, 22–48.

Fan, Ky (1951). Maximum properties and inequalities for the eigenvalues of completely continuous operators. *Proc. Nat. Acad. Sci., 37*, 760–766.

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

Hotelling, H. (1936). Relations between two sets of variates, *Biometrika, 28*, 321–377.

Jaffrenou, P. A. (1978). Sur l'analyse des familles finies de variables vectorielles: Bases algébrique et application à la description statistique. Thèse de troisième cycle, l'Université de Saint-Etienne, France.

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

Kristof, W. (1970). A theorem on the trace of certain matrix products and some applications. *Journal of Mathematical Psychology, 7*, 515–530.

Lancaster, H. O. (1969). *The chi square distribution.* New York: Wiley.

```
%MACRO MATCOR(X,Y,XROW,YROW,TF=0,CF=0,SX=MINPQ,SM=1);

  *  THE PARAMETERS FOR MATCOR ARE AS FOLLOWS:
       X     ... FIRST  MATRIX IN THE COMPARISON
       Y     ... SECOND MATRIX IN THE COMPARISON
       XROW  ... VARIABLE CONTAINING ROW LABELS FOR X
       YROW  ... VARIABLE CONTAINING ROW LABELS FOR Y
       TF    ... 0 IF BOTH MATRICES ARE TO BE ROTATED,
                 NONZERO IF Y IS LINEARLY TRANSFORMED
                 AND X ROTATED
       CF    ... 0 IF CORRELATION IS TO BE MAXIMIZED,
                 NONZERO IF DISTANCE IS TO BE MINIMIZED,
       SX    ... MAXIMUM DIMENSIONALITY FOR COMPARISON
       SM    ... MINIMUM DIMENSIONALITY FOR COMPARISON;

PROC MATRIX;

  * INPUT X AND Y;

  FETCH X DATA=&X COLNAME=XCOLLABS ROWNAME=&XROW;
  FETCH Y DATA=&Y COLNAME=YCOLLABS ROWNAME=&YROW;

  * SET VARIOUS VALUES;

  P=NCOL(X);   Q=NCOL(Y);
  IF P>Q THEN MINPQ=Q;  ELSE MINPQ=P;
  TOL=1E-4;   PIOVER4=.785398;   BLANK=" ";
  XTY=X'*Y;   XTX=X'*X;   YTY=Y'*Y;

  * INITIALIZE L AND M;

  TRANFLAG=&TF;  CRITFLAG=&CF;
  IF TRANFLAG=0 THEN DO;
    * INITIALIZATION FOR ORTHOGONAL-ORTHOGONAL CASE;
    SVD U D V XTY;
    RNK=Q+1-RANK(D);  L=U(,RNK);  M=V(,RNK);
    SINGVALS=D(RNK,)';
  END;
  ELSE DO;
    * INITIALIZATION FOR ORTHOGONAL-LINEAR CASE;
    YTYINV=INV(YTY);
    A=XTY*YTYINV*XTY';   L=EIGVEC(A);
    BETA=YTYINV*XTY';    M=BETA*L;
  END;

  *  LOOP THROUGH POSSIBLE DIMENSIONALITIES OF COMPARISON;

  SMAX=&SX;                               SMIN=&SM;
  IF SMAX>MINPQ THEN SMAX=MINPQ;
  IF SMIN>MINPQ THEN SMIN=MINPQ;
  DO S=SMAX TO SMIN BY -1;                K=1:S;

    *  COMPUTE INITIAL VALUE OF CRITERION;

    XTRANSFM=L(,K);   B11=TRACE(XTRANSFM'*XTX*XTRANSFM);
    YTRANSFM=M(,K);   B22=TRACE(YTRANSFM'*YTY*YTRANSFM);
                      B12=TRACE(XTRANSFM'*XTY*YTRANSFM);
    IF CRITFLAG=0 THEN CRIT=B12#/SQRT(B11*B22);
    ELSE DO;  CRIT=SQRT(B11+B22-2#B12);  TOL=CRIT*1E-4;  END;

    * OUTPUT INITIAL RESULTS;

    NOTE PAGE DIMENSIONALITY OF COMPARISON IS NOW;
    PRINT S ROWNAME=BLANK COLNAME=BLANK;
    NOTE INITIAL TRANSFORMATIONS;
    PRINT XTRANSFM ROWNAME=XCOLLABS FORMAT=8.3;
    PRINT YTRANSFM ROWNAME=XCOLLABS FORMAT=8.3;
    MAINITER=0;  NPHASE1=1;  NPHASE2=2;
    MATRIX=MAINITER||MAINITER||CRIT;  HISRLAB=BLANK;
    IF CRITFLAG=0 THEN NOTE INITIAL CORRELATION COEFFICIENT;
                  ELSE NOTE INITIAL DISTANCE;
    PRINT CRIT ROWNAME=BLANK COLNAME=BLANK FORMAT=10.3;

    * MAIN ITERATION LOOP;

    CRITOLD=CRIT-1;
    IF S<P | (S<Q & TRANFLAG=0) THEN
    DO MAINITER = 1 TO 10 WHILE(ABS(CRIT-CRITOLD)>TOL);
      CRITOLD=CRIT;

      * OPTIMIZATION WITH RESPECT TO L;

      PV=P;  B=B22;  TV=L;  LABV="L";  XIT=(XTY*M(,K))';  YIT=XTX;
      LINK ROTATE;
      L=TV;  CRIT=CRITV;  B11=BV;

      * OPTIMIZATION WITH RESPECT TO M;

      IF TRANFLAG=0 THEN DO;      * ORTHOGONAL-ORTHOGONAL CASE;
          PV=Q;  B=B11;  TV=M;  LABV="M";  XIT=L(,K)'*XTY;  YIT=YTY;
          LINK ROTATE;
          M=TV;  CRIT=CRITV;  B22=BV;
        END;
                    ELSE DO;      * ORTHOGONAL-LINEAR CASE;
          M=BETA*L;  MC=M(,K);
          B12=TRACE(L(,K)'*XTY*MC);  B22=TRACE(MC'*YTY*MC);
          IF CRITFLAG=0 THEN CRIT=B12#/SQRT(B11*B22);
                        ELSE CRIT=SQRT(B11+B22-2#B12);
        END;

    END;                                 * END OF MAIN ITERATION LOOP;
```

FIGURE 3
SAS PROC MATRIX Code for the Algorithm.

```
  * PRINT FINAL RESULTS;

  IF MAINITER>0 THEN DO;
    IF ABS(CRIT-CRITOLD)<TOL THEN NOTE CONVERGENCE ACHIEVED;
                              ELSE NOTE MAX. NO. ITERATIONS TAKEN;
    IF CRITFLAG=0 THEN NOTE FINAL CORRELATION COEFFICIENT;
                       ELSE NOTE FINAL DISTANCE;
    PRINT CRIT ROWNAME=BLANK COLNAME=BLANK FORMAT=10.3;
    NOTE SKIP=3 HISTORY OF ITERATIONS;
    IF CRITFLAG=0 THEN HISCLAB="ITER" "PHASE" "R";
                  ELSE HISCLAB="ITER" "PHASE" "DISTANCE";
    PRINT MATRIX ROWNAME=HISRLAB COLNAME=HISCLAB;
  END;
  NOTE SKIP=5 FINAL TRANSFORMATIONS;
  XTRANSFM=L(,K);  YTRANSFM=M(,K);
  *  SWITCH SIGNS IF SUM OF COEFFICIENTS IS NEGATIVE IN A COLUMN OF L;
  DO I=1 TO K;
    IF XTRANSFM(+,I)<0 THEN DO;
      XTRANSFM(,I)=-XTRANSFM(,I);  YTRANSFM(,I)=-YTRANSFM(,I);
  END;  END;
  PRINT XTRANSFM ROWNAME=XCOLLABS FORMAT=8.3;
  PRINT YTRANSFM ROWNAME=YCOLLABS FORMAT=8.3;
  XIMAGE=X*XTRANSFM;  YIMAGE=Y*YTRANSFM;
  NOTE SKIP=5 TRANSFORMED MATRICES;
  PRINT XIMAGE ROWNAME=&XROW FORMAT=8.3;
  PRINT YIMAGE ROWNAME=&YROW FORMAT=8.3;

END;
STOP;

* OPTIMIZE CRITV BY ROTATING INTO SUBSPACE FOR EACH PAIR OF
  DIMENSIONS (PHASE 1) AND THEN ROTATING WITHIN SUBSPACE (PHASE 2);

ROTATE: LINK COMPUT2;

  * ROTATING TV INTO SUBSPACE;

  PHASE1: DO I2=S+1 TO PV;  DO I1=1 TO S;
    CRIT0=CRITV*CRITV;  IF CRITFLAG=0 THEN CRIT0=-CRIT0;
    U=TV(,I1);  V=TV(,I2);
    LINK DERIV;
    TRY=TV(,K);  ANGLE=-D1F#/D2F;
    ANGLE=MIN(ANGLE||PIOVER4);  ANGLE=MAX(ANGLE||-PIOVER4);
    TEST: LINK COMPUT1;
    IF ANGLE<1E-3     THEN GOTO NEXT;
    IF CRITTRY<CRIT0 THEN GOTO NEXT;
    ANGLE=ANGLE#/2;  GOTO TEST;
    NEXT: BV=BTRY;  B12=B12TRY;
    IF CRITFLAG=0 THEN CRITV=SQRT(-CRITTRY);
                  ELSE CRITV=SQRT(CRITTRY);
    TV(,I1)=CS#U+SN#V;  TV(,I2)=CS#V-SN#U;  TVC=TV(,K);
  END;  END;
  MATRIX=MATRIX//(MAINITER||NPHASE1||CRITV);  HISRLAB=HISRLAB//LABV;

  * ROTATING WITHIN SUBSPACE;

    PHASE2: IF S>1 THEN DO; TEMP=XIT*TVC;
      SVD U D V TEMP;    RNK=S+1-RANK(D);
      TV(,K)=TVC*V(,RNK)*U(,RNK)`;
      LINK COMPUT2;
    END;
    MATRIX=MATRIX//(MAINITER||NPHASE2||CRITV);  HISRLAB=HISRLAB//LABV;
  RETURN;

  COMPUT1: SN=SIN(ANGLE);  CS=SQRT(1-SN*SN);
    TRY(,I1)=CS#U+SN#V;
    BTRY=TRACE(TRY`*YIT*TRY);
    B12TRY=TRACE(XIT*TRY);
    IF CRITFLAG=0 THEN CRITTRY=-((B12TRY)##2)#/(BTRY*B);
                  ELSE CRITTRY=B+BTRY-2#B12TRY;
  RETURN;

  COMPUT2: TVC=TV(,K);
    BV=TRACE(TVC`*YIT*TVC);  B12=TRACE(XIT*TVC);
    IF CRITFLAG=0 THEN CRITV=B12#/SQRT(BV*B);
                  ELSE CRITV=SQRT(BV+B-2#B12);
  RETURN;

  DERIV: T1=U`*YIT*U;    T2=V`*YIT*V,     T3=U`*YIT*V;
         T4=XIT(I1,)*V;  T5=-XIT(I1,)*U;
    D1B12= (XIT(I1,)*V)#/B12;  D1B11=T3#/BV;
    D2B12=-(XIT(I1,)*U)#/B12;  D2B11=(T2-T1)#/BV;
    IF CRITFLAG=0 THEN DO;
      D1F=2*CRIT0*(T4#/B12-T3#/BV);
      D2F=D1F##2#/CRIT0+2*CRIT0*(T5#/B12-(T2-T1)#/BV-(T4#/B12)##2
                                  +(T3#/BV)##2);
    END;
    ELSE DO;  D1F=2#(T3-T4);  D2F=2#(T2-T1-T5);  END;
  RETURN;

%MEND MATCOR;
```

FIGURE 4
Continuation of the SAS PROC MATRIX Code

Lingoes, J. C. & Schönemann, P. H. (1974). Alternative measures of fit for the Schönemann-Carroll matrix fitting algorithm, *Psychometrika, 39*, 423–427.

Pages, J. P., Ramsay, J. O., & Tenenhaus, M. (1984). A geometrical approach to multivariate analysis, Unpublished manuscript submitted for publication.

Ramsay, J. O. (1980). Exploring multivariate data with inside-out plots. Presented at the Psychometric Society Meetings, Iowa City.

Ramsay, J. O. & Novick, M. R. (1980) PLU robust Bayesian decision theory: Point estimation, *Journal of American Statistical Association, 75*, 901–907.

Rao, C. R. and Yanai, H. (1979). General definition and decomposition of projectors and some applications to statistical problems, *Journal of Statistical Planning and Inference, 3*, 1–17.

Rao, C. R. (1979) Separation theorems for singular values of matrices and their applications in multivariate analysis, *Journal of Multivariate Analysis, 9*, 362–377.

Renyi, A. (1959). On measures of dependence, *Acta Mathematica of the Academy of Science of Hungary, 10*, 441–451.

Richter, H. (1958). Zur Abschatzung von Matrizennormen. *Dieze Nachr., 18*, 178–187.

Robert, P. & Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: The RV-coefficient. *Applied Statistics, 25*, 257–265.

SAS Institute Inc. (1982). *SAS User's Guide: Statistics, 1982 Edition.* Cary, NC: SAS Institute Inc.

Schönemann, P. H. (1966). A generalized solution of the orthogonal Procrustes problem, *Psychometrika, 31*, 1–10.

Styan, G. P. H. (1976). Separation theorems. Unpublished manuscript, McGill University.

Theobald, C. M. (1975). An inequality with application to multivariate analysis, *Biometrika, 62*, 461–466.

von Neumann, J. (1937). Some matrix-inequalities and metrization of matric-space. *Tomask Univ. Rev., 1*, 286–300. (Reprinted in *Collected Works*, Pergamon, Oxford, Vol. IV, 205–218.)

Wainer, H. & Thissen, D. (1981). Graphical data analysis. *Annual Review of Psychology, 32*, 191–241.

Yanai, H. (1974). Unification of various techniques of multivariate analysis by means of generalized coefficient of determination (G.C.D.), *J. of Behaviormetrics, 1*, 45–54.

Yanai, H. (1980). A proposition of generalized method for forward selection of variables, *Behaviormetrika, 7*, 95–107.

Manuscript received 10/19/82
Final version received 5/30/84