ELSEVIER Computational Statistics & Data Analysis 27 (1998) 445–459

COMPUTATIONAL STATISTICS & DATA ANALYSIS

Estimating correlation matrices that have common eigenvectors

James R. Schott

Department of Statistics, University of Central Florida, Orlando, FL 32816-2370, USA

Received 1 May 1997; received in revised form 1 December 1997

Abstract

In this paper we develop a method for obtaining estimators of the correlation matrices from k groups when these correlation matrices have the same set of eigenvectors. These estimators are obtained by utilizing the spectral decomposition of a symmetric matrix; that is, we obtain an estimate, say $\hat{P}$, of the matrix P containing the common normalized eigenvectors along with estimates of the eigenvalues for each of the k correlation matrices. It is shown that the rank of the Hadamard product, $\hat{P} \odot \hat{P}$, is a crucial factor in the estimation of these correlation matrices. Consequently, our procedure begins with an initial estimate of P which is then used to obtain an estimate $\hat{P}$ such that $\hat{P} \odot \hat{P}$ has its rank less than or equal to some specified value. Initial estimators of the eigenvalues of Ω_i, the correlation matrix for the ith group, are then used to obtain refined estimators which, when put in the diagonal matrix $\hat{D}_i$ as its diagonal elements, are such that $\hat{P}\hat{D}_i\hat{P}'$ has correlation-matrix structure. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Common principal components; Hadamard product; Principal components analysis

1. Introduction

Common principal components (CPC) in several covariance matrices is a topic that has received much attention in recent years. Various aspects of the CPC problem have been treated by Krzanowski (1979, 1982, 1984), Flury (1984, 1986, 1987, 1988), Keramidas et al. (1987), Schott (1988, 1991), Chen and Robinson (1989) and Fujioka (1993). One of the important applications of the CPC model is in the principal components analysis of the covariance matrices of several groups. More generally, in those situations in which group covariance matrices are not identical, the CPC model offers one alternative model for covariance structure that allows for differing covariance matrices while still retaining some common structure across groups.

0167-9473/98/$19.00 © 1998 Elsevier Science B.V. All rights reserved.
PII S0167-9473(98)00027-9

These applications of CPC models for covariance matrices extend as well to analyses involving correlation matrices and, in particular, most practical applications of principal components analysis utilize correlation matrices instead of covariance matrices. However, much less work has been done in this area. Some inferential methods have been developed by Schott (1991, 1997a) and Krzanowski (1993), but the more important problem of estimating the correlation matrices under the restrictions of the CPC model has yet to be solved.

2. The CPC model for correlation matrices

Suppose that the same m variables are being measured on subjects in k different groups, with the ith group having the $m\times m$ correlation matrix Ω_i. These k correlation matrices have common eigenvectors if there exists an $m\times m$ orthogonal matrix P such that $\Omega_i = PD_iP'$ for each i, where $D_i = \mathrm{diag}(d_{i1},\dots,d_{im})$ is a diagonal matrix with the eigenvalues of Ω_i as its diagonal elements. No assumption is made regarding the order of the diagonal elements of D_i so that the ordering for one particular group may differ from that of some of the other groups. The columns of P are the normalized common eigenvectors and these will be uniquely defined, except for sign, if for each $j = 1,\dots,m$, there is at least one group, say the ith group, such that d_{ij} is a simple eigenvalue of Ω_i. Since the normalized eigenvectors of a covariance or correlation matrix are used to construct its principal components, the model $\Omega_i = PD_iP'$ for $i = 1,\dots,k$ is sometimes referred to as the common principal components model.

Due to the complexity of the distribution of a correlation matrix, some of the usual methods of estimation, such as maximum likelihood estimation and minimum chi-squared estimation, become too complicated to be of any practical use. Consequently, an overriding concern in the development of estimators with common eigenvectors is that the procedure be simple enough so that the estimates can actually be computed. With this in mind, we develop a procedure which begins with initial estimators of P and D_i; these are then used to compute refined estimators which yield the required correlation-matrix structure.

In estimating Ω_i, we will first find an estimator of P and then find an estimator for each D_i. This process is much more difficult when Ω_i is a correlation matrix instead of a covariance matrix because the choice of an estimator of D_i may be affected by our choice of an estimator of P. For a given orthogonal matrix P, we will denote by $\rho(P)$ the set of all $m\times 1$ vectors $a = (a_1,\dots,a_m)'$ such that PD_aP' has correlation-matrix structure, where $D_a = \mathrm{diag}(a_1,\dots,a_m)$; that is, $\rho(P) = \{a : a_i \geq 0 \text{ and } e_i'PD_aP'e_i = 1, \text{ for } i = 1,\dots,m\}$, where e_i is the ith column of the $m\times m$ identity matrix. Thus, if we let $d_i = (d_{i1},\dots,d_{im})'$, then for some orthogonal matrix P, our CPC model implies that $d_i \in \rho(P)$ for $i = 1,\dots,k$.

The next result will illustrate that the estimation of each d_i critically depends upon the rank of the Hadamard product (see, for example, Schott 1997b, Section 7.6) $\hat{P}\odot\hat{P}$. Although $\hat{P}$ will always be nonsingular since it is an orthogonal matrix, in general, under the CPC model the matrix $\hat{P}\odot\hat{P}$ will be singular.

Theorem 1. *Let $\hat{P}$ be an $m \times m$ orthogonal matrix and let $\rho(\hat{P})$ be the set containing all $m \times 1$ vectors a for which $\hat{P}D_a\hat{P}'$ is a correlation matrix. If $\hat{P} \odot \hat{P}$ is nonsingular, then $\rho(\hat{P}) = \{1_m\}$, where 1_m is the $m \times 1$ vector of ones. On the other hand, if rank$(\hat{P} \odot \hat{P}) = m - r < m$, let the columns of the $m \times r$ matrix A be any set of orthonormal eigenvectors of $(\hat{P} \odot \hat{P})'(\hat{P} \odot \hat{P})$ corresponding to its zero eigenvalue. Then, in this case, $\rho(\hat{P}) = \{a : a = 1_m + Ac, c \in \Re^r, a_i \geq 0$ for $i = 1, \ldots, m\}$.*

Proof. Since $(\hat{P} \odot \hat{P})1_m = 1_m$, 1_m is a particular solution to the system of equations, $(\hat{P} \odot \hat{P})a = 1_m$, and so the general solution will be given by $1_m + Bc$, where is c an arbitrary $r \times 1$ vector and B is any $m \times r$ matrix whose columns form a basis for the null space of $\hat{P} \odot \hat{P}$. The result then follows since the columns of A do form a basis for the null space of $\hat{P} \odot \hat{P}$. □

It follows from Theorem 1 that the correlation matrices $\Omega_1, \ldots, \Omega_k$ will satisfy the CPC model with rank$(P \odot P) = m$ only if $\Omega_1 = \cdots = \Omega_k = I_m$. As a result, all practical applications of the CPC model will have rank$(P \odot P) \leq m - 1$. The lower bound for rank$(P \odot P)$ will be 1 or 2 depending upon the value of m. For instance, rank$(P \odot P) = 1$ is possible only if there exist Hadamard matrices of size $m \times m$ (see Hedayat and Wallis, 1978). If this is not the case, we can use Hadamard matrices to easily illustrate that this lower bound is 2. Let $m_* < m$ be the largest integer for which $m_* \times m_*$ Hadamard matrices exist and suppose that H is an $m_* \times m_*$ normalized Hadamard matrix so that each element in its first row is $+1$. Define H_* to be the $(m - m_*) \times m_*$ submatrix of H consisting of the last $(m - m_*)$ rows of H. Then it is easily verified that the matrix

$$P = \begin{pmatrix} m_*^{-1/2} H_* & (0) \\ (2m_* - m)^{1/2} m_*^{-1} 1_{m_*} 1'_{m_*} & m_*^{-1/2} H'_* \end{pmatrix}$$

is orthogonal and satisfies rank$(P \odot P) = 2$.

3. Estimation of Ω_i under the CPC model

3.1. Some simple initial estimators of P and D_i

Suppose that we have available independent random samples from the k groups, with a sample of size n_i from the ith group, and these are used to compute the usual sample correlation matrices, $R_1, \ldots, R_k$. It will not be difficult to use these sample data to construct individual estimators of P and D_i.

An initial estimator of P can be obtained by using the F–G diagonalization algorithm (Flury and Constantine, 1985; Flury and Gautschi, 1986). This procedure finds the orthogonal matrix, which we will denote as $\hat{P}_*$, that transforms the matrices $R_1, \ldots, R_k$ simultaneously as close as possible to diagonal matrices. Specifically, $\hat{P}_*$ is the orthogonal matrix that minimizes

$$\Phi(\hat{P}_*) = \prod_{i=1}^{k} \left\{ \frac{|\mathrm{diag}(\hat{P}_*' R_i \hat{P}_*)|}{|R_i|} \right\}^{n_i},$$

where $\mathrm{diag}(\hat{P}_*' R_i \hat{P}_*)$ is the diagonal matrix that has the same diagonal elements as $\hat{P}_*' R_i \hat{P}_*$. The matrix $\hat{P}_*$ can also be used to obtain an estimator for each d_i. For instance, d_i can be estimated using $\hat{d}_{*i}$ which has as its components, the diagonal elements of the matrix $\hat{P}_*' R_i \hat{P}_*$.

While the estimators $\hat{P}_*$ and $\hat{d}_{*i}$ may be reasonable individual estimators of P and d_i, they have not been constructed under the constraints imposed by a correlation matrix. Thus, in general, $\hat{P}_* D_{\hat{d}_{*i}} \hat{P}_*'$ will not have correlation-matrix structure. However, these estimators will prove to be useful since our method, given in the next section, for deriving estimators that do yield this correlation-matrix structure will require some initial estimators of P and d_i.

3.2. Refined estimators of P and D_i

It follows from Theorem 1 that as the rank of $\hat{P} \odot \hat{P}$ increases, we get a more limited collection of vectors in the set $\rho(\hat{P})$. Consequently, if the estimator, $\hat{P}$, of P is such that the rank of $\hat{P} \odot \hat{P}$ exceeds that of $P \odot P$, we may be unable to obtain reasonable estimators of the d_i vectors even when $\hat{P}$ is a very good estimator of P. Thus, our estimation process for P will first involve a method for obtaining an estimator $\hat{P}_r$ so that $\hat{P}_r \odot \hat{P}_r$ has rank no larger than $m-r$ for a prespecified value of r. We will then use some method for determining an appropriate choice for r, say r_*, and define $\hat{P} = \hat{P}_{r_*}$ as our final estimator of P.

We will obtain an estimator $\hat{P}_r$ satisfying $\mathrm{rank}(\hat{P}_r \odot \hat{P}_r) \leq m-r$ by utilizing an initial estimator of P such as the estimator $\hat{P}_*$ discussed in the previous section. In particular, we will define $\hat{P}_r$ so that

$$\mathrm{vec}(\hat{P}_r - \hat{P}_*)'\mathrm{vec}(\hat{P}_r - \hat{P}_*) = \min_{Q \in S_r} \mathrm{vec}(Q - \hat{P}_*)'\mathrm{vec}(Q - \hat{P}_*), \tag{3.1}$$

where $S_r = \{Q : Q \text{ is } m \times m,\ Q'Q = I_m,\ \mathrm{rank}(Q \odot Q) \leq m-r\}$; that is, $\hat{P}_r$ could be described as the closest orthogonal matrix to $\hat{P}_*$ satisfying $\mathrm{rank}(\hat{P}_r \odot \hat{P}_r) \leq m-r$.

Simplifying the notation in (3.1), we need to compute the orthogonal matrix Y which has $\mathrm{rank}(Y \odot Y) \leq m-r$ and minimizes $\mathrm{vec}(Y-X)'\mathrm{vec}(Y-X)$ for some given orthogonal matrix X. This solution can be obtained by finding the choice of Y, along with the choices of the $m \times (m-r)$ matrices F and H', which minimize the Lagrangian function

$$\begin{aligned} f(Y,F,H) &= \mathrm{vec}(Y-X)'\mathrm{vec}(Y-X) + \mathrm{tr}\{\Lambda(Y'Y - I_m)\} \\ &\quad + \mathrm{tr}\{\Delta(Y \odot Y - FH)\}, \end{aligned}$$

where the $m \times m$ symmetric matrix Λ and the $m \times m$ matrix Δ contain the Lagrange multipliers. Differentiating with respect to Y, F, and H, and then equating to zero, leads to the equations

$$Y - X + Y\Lambda + \Delta' \odot Y = (0), \tag{3.2}$$

$$H\Delta = (0), \tag{3.3}$$

$$\Delta F = (0), \tag{3.4}$$

which will be used, along with the constraints, $Y'Y - I_m = (0)$ and $Y \odot Y - FH = (0)$, to find the solution. Let $E_1 D_1 G_1'$ be the singular-value decomposition of the matrix $Y \odot Y$, where the $m \times m$ matrices $E = (E_1 \quad E_2)$ and $G = (G_1 \quad G_2)$ are orthogonal and the $(m-r) \times (m-r)$ matrix D_1 is diagonal. In this case, in view of Eqs. (3.3) and (3.4), we find that (3.2) may be written as

$$Y - X + YA + E_2 C G_2' \odot Y = (0),$$

where C is some $r \times r$ matrix. Solving this equation for A and then equating that expression to its transpose leads to

$$(G_2 C' E_2' \odot Y' - X')Y - Y'(E_2 C G_2' \odot Y - X) = (0).$$

If we denote the expression on the left-hand side of this equation by Z_C, it can be easily shown that for fixed X, Y, E_2, and G_2, the minumum value of $\text{vec}(Z_C)'\text{vec}(Z_C)$ is given by

$$f_1(Y) = \text{vec}(Y'X - X'Y)'\{I - L(L'L)^+ L'\}\text{vec}(Y'X - X'Y),$$

where $L = (I - K_{mm})(I \otimes Y')D_{\text{vec}(Y)}(G_2 \otimes E_2)$, K_{mm} is a commutation matrix, and $D_{\text{vec}(Y)}$ is the $m^2 \times m^2$ diagonal matrix satisfying $D_{\text{vec}(Y)} 1_{m^2} = \text{vec}(Y)$. Thus, a matrix Y will be a solution if it is an orthogonal matrix and $f_2(Y) = f_1(Y) + \lambda_{m-r+1}\{(Y \odot Y)'(Y \odot Y)\} = 0$, where $\lambda_{m-r+1}\{(Y \odot Y)'(Y \odot Y)\}$ denotes the $(m-r+1)$th largest eigenvalue of $(Y \odot Y)'(Y \odot Y)$. The function f_2 can then be used to find Y numerically. For instance, the solutions for the examples discussed in Section 5 were computed using the downhill simplex method (see, for example, Press et al., 1992). In order to implement this procedure the function f_2 was written in terms of $\frac{1}{2}m(m+1)$ variables, say the nonzero elements in an $m \times m$ lower triangular matrix U. Each U matrix was used to compute an orthogonal Y matrix by using the required orthogonality conditions to fill in the zeroes in the upper triangular portion of U, and then the columns of this matrix were normalized. This numerical method performed satisfactorily for our examples. An alternative approach for obtaining a solution to (3.1), one suggested by a referee, would apply sequential quadratic programming techniques (see, for example, Fletcher, 1987).

As with the estimation of P, the estimation of the d_i vectors will make use of initial estimators along with the refined estimator $\hat{P}_r$. If we use the initial estimators $\hat{d}_{*i}$ described in Section 3.1, then for each i we define the refined estimator $\hat{d}_{ri}$ so that

$$(\hat{d}_{ri} - \hat{d}_{*i})'(\hat{d}_{ri} - \hat{d}_{*i}) = \min_{u \in \rho(\hat{P}_r)} (u - \hat{d}_{*i})'(u - \hat{d}_{*i}).$$

In other words, $\hat{d}_{ri}$ is the vector in $\rho(\hat{P}_r)$ that is closest to $\hat{d}_{*i}$.

Let the columns of the $m \times r$ matrix A_r be any set of orthonormal eigenvectors of $(\hat{P}_r \odot \hat{P}_r)'(\hat{P}_r \odot \hat{P}_r)$ corresponding to its zero eigenvalue, so that from Theorem 1 any $a \in \rho(\hat{P}_r)$ can be expressed as $a = 1_m + A_r c$ for some $r \times 1$ vector c. Since $A_r A_r'$ is the projection matrix for the space spanned by the r columns of A_r, for any $m \times 1$ vector γ, $A_r A_r' \gamma$ will be the vector in this space closest to γ. From this and the fact that $A_r' 1_m = 0$, it immediately follows that $\tilde{\gamma} = 1_m + A_r A_r' \gamma$ is the vector in

$\rho(\hat{P}_r)$ closest to γ, as long as all of its components are nonnegative. If some of these components are negative, we must use the following result. The proof of this and all subsequent theorems can be found in the appendix.

Theorem 2. *Suppose that γ is an $m \times 1$ vector and denote by $\tilde{\gamma}$ the vector in $\rho(\hat{P}_r)$ closest to γ. Then $\tilde{\gamma} = 1_m + A_r A_r' \gamma$ if all of the components of $1_m + A_r A_r' \gamma$ are nonnegative. Otherwise $\tilde{\gamma}$ will be of the form*

$$\tilde{\gamma} = 1_m - A_r B_2^+ 1_s + Q_1' B_1 (I_r - B_2^+ B_2) B_1' \gamma_1,$$

where Q is some $m \times m$ permutation matrix, $(Q_1' \quad Q_2')$ represents some partitioning of the matrix Q', and $B_1 = Q_1 A_r$, $B_2 = Q_2 A_r$, and $\gamma_1 = Q_1 \gamma$.

It is important to note that Theorem 2 does not indicate the sizes of the submatrices Q_1 and Q_2 or even which $m \times m$ permutation matrix Q should be used. Clearly, $\tilde{\gamma}$ will be calculated using choices of Q_1 and Q_2 which produce the smallest value of $(\tilde{\gamma} - \gamma)'(\tilde{\gamma} - \gamma)$ over all choices for Q_1 and Q_2, but in most situations it will not be necessary to evaluate this sum of squares for all choices of Q_1 and Q_2.

Now if $\hat{P}_r$ is obtained from (3.1) and $\hat{d}_{ri}$ is computed using Theorem 2, that is, $\hat{d}_{ri} = \tilde{\gamma}$ with $\gamma = \hat{d}_{*i}$, then the estimators $\hat{\Omega}_{ri} = \hat{P}_r D_{\hat{d}_{ri}} \hat{P}_r'$, $i = 1, \ldots, k$, will have correlation-matrix structure and satisfy the conditions of the CPC model. A measure of the closeness of these estimators to the corresponding sample correlation matrices is given by

$$\tau_r = \sum_{i=1}^{k} \mathrm{tr}\{(\hat{\Omega}_{ri} - R_i)^2\}.$$

In estimating the Ω_i matrices, we will employ a procedure that begins with $r = m - 1$ and then continually reduces r by one until τ_r increases; that is, if $\tau_{m-1} > \cdots > \tau_s$ and $\tau_s < \tau_{s-1}$, then we will use the estimators based on $r* = s$. We will denote the final estimator of Ω_i as $\hat{\Omega}_i = \hat{P} D_{\hat{d}_i} \hat{P}'$, so that $\hat{P} = \hat{P}_{r*}$, $\hat{d}_i = \hat{d}_{r*i}$, and $\hat{\Omega}_i = \hat{\Omega}_{r*i}$.

We end this section with the following important result which establishes the consistency of the estimators, $\hat{\Omega}_1, \ldots, \hat{\Omega}_k$ when $\mathrm{rank}(P \odot P) = m - r*$.

Theorem 3. *Suppose that $\Omega_1, \ldots, \Omega_k$ are positive-definite correlation matrices which satisfy the CPC model and have normalized common eigenvectors that are uniquely defined except for sign. If $\mathrm{rank}(P \odot P) = \mathrm{rank}(\hat{P} \odot \hat{P}) = m - r$, then for each i, $\hat{\Omega}_i$ converges in probability to Ω_i.*

An implication of Theorem 3 and disadvantage of this estimation procedure is that the consistent estimation of $\Omega_1, \ldots, \Omega_k$ depends on the correct identification of r.

4. Validity of the CPC model

Before using the estimators developed for the CPC model in the previous section, we will need to be able to determine if the population correlation matrices satisfy this CPC model. The ideal approach here would be to make use of the estimators themselves in assessing the validity of the CPC model. For instance, if the CPC model holds with $\mathrm{rank}(P \odot P) = m - s$, then the asymptotic mean of each off-diagonal element of the matrix $\hat{P}'R_i\hat{P}$ would be zero as long as we have computed $\hat{P}$ with $\mathrm{rank}(\hat{P} \odot \hat{P}) \leq m - r$, where $r \leq s$. Consequently, we could use the off-diagonal elements of $\hat{P}_r'R_i\hat{P}_r$ to construct Wald-type statistics to test sequentially the hypotheses

$$\mathrm{H}_{0r} : \Omega_i = PD_iP', \quad i = 1, \ldots, k; \quad \mathrm{rank}(P \odot P) \leq m - r.$$

We would first test the hypothesis with $r = 1$, then with $r = 2$, and continue until either H_{0r} is rejected for some r or it is not rejected for either $r = m-1$ or $m-2$ depending on the value of m. Since we are assuming that the k population correlation matrices are not identical, the first test will determine whether or not the CPC model holds. The subsequent testing will determine the appropriate value to use for $\mathrm{rank}(P \odot P)$, and so this would offer an alternative method to the one described in Section 3.2 for determining the value of r_*.

An asymptotically chi-squared Wald-type statistic could be constructed if we were able to show that the asymptotic null distribution of $n^{\frac{1}{2}}\,\mathrm{vec}(\hat{P}_r)$ is normal and obtain its asymptotic covariance matrix, where n here is defined to be $n = n_1 + \cdots + n_k$. We will not attempt to solve this difficult problem here. Instead, we propose a fairly simple alternative test for the null hypothesis that the CPC model holds, that is, for H_{01}. The construction of our statistic is based on the property that (see, for example, Schott, 1997b, Section 4.7) $\Omega_1, \ldots, \Omega_k$ satisfy the CPC model if and only if $\Omega_i\Omega_j = \Omega_j\Omega_i$ for every $i \neq j$. Our statistic utilizes the $\frac{1}{2}m(m-1) \times 1$ vectors $v_{ij} = \tilde{\mathrm{v}}(R_iR_j - R_jR_i)$ containing the columns in the strictly lower triangular portion of $R_iR_j - R_jR_i$, stacked one underneath the other. It also involves the basis matrix for skew-symmetry, $\tilde{D}_m$ (see Magnus, 1988), which is the $m^2 \times \frac{1}{2}m(m-1)$ matrix satisfying $\tilde{D}_m\tilde{\mathrm{v}}(A) = \mathrm{vec}(A)$ for every skew-symmetric $m \times m$ matrix A.

Let $m_* = \frac{1}{2}m(m-1)$, $k_* = \frac{1}{2}k(k-1)$ and define the $m_*k_* \times 1$ vector v as

$$v = n^{1/2}\,(\,\eta_2\eta_1v_{21}', \ldots, \eta_m\eta_1v_{m1}', \eta_3\eta_2v_{32}', \ldots, \eta_m\eta_2v_{m2}', \ldots, \eta_m\eta_{m-1}v_{mm-1}'\,)',$$

where $\eta_i = (n_i/n)^{1/2}$. The asymptotic mean of v is the 0 vector if and only if the CPC model holds. Further, it is not difficult to show that the asymptotic normality of v follows from the fact that $n_i^{1/2}\mathrm{vec}(R_i - \Omega_i)$ is asymptotically normal for each i. Thus, if $\hat{\Theta}^+$ is a consistent estimator of the Moore–Penrose generalized inverse of the asymptotic covariance matrix, Θ, of v then we could use the Wald statistic $v'\hat{\Theta}^+v$ to test the null hypothesis H_{01}. Unfortunately, it will not be possible, in general, to find such an estimator $\hat{\Theta}^+$ since the rank of Θ depends on the matrices $\Omega_1, \ldots \Omega_k$ which are not completely specified by the null hypothesis. However, we can still construct a test based on v through its sum of squares. Specifically, we will use the statistic

$$T = 2v'v = \sum_{i>j} n\eta_i^2\eta_j^2\,\mathrm{vec}(R_iR_j - R_jR_i)'\,\mathrm{vec}(R_iR_j - R_jR_i).$$

It follows from the asymptotic normality of v that T will be asymptotically a quadratic form in normal variates. Consequently, its asymptotic distribution should be adequately approximated by a distribution of the form $c\chi_d^2$. To obtain values for c and d, we will use the asymptotic mean and variance of T given in the next theorem.

Theorem 4. *Let Ψ_i denote the asymptotic covariance matrix of $n_i^{1/2}\{\mathrm{vec}(R_i) - \mathrm{vec}(\Omega_i)\}$ and define $\Upsilon_{ijk} = \tilde{D}_m'(\Omega_i \otimes I_m)\Psi_j(\Omega_k \otimes I_m)\tilde{D}_m$. Then, asymptotically,*

$$\mu_T = 2\sum_{j=1}^{k}\sum_{i\neq j}\eta_i^2\,\mathrm{tr}(\Upsilon_{iji}), \tag{4.1}$$

$$\sigma_T^2 = 8\sum_{j=1}^{k}\left\{\sum_{i\neq j}\sum_{l\neq j}\eta_i^2\eta_l^2\,\mathrm{tr}(\Upsilon_{ijl}\Upsilon_{ijl}') + \sum_{i\neq j}\eta_i^2\eta_j^2\,\mathrm{tr}(\Upsilon_{jij}\Upsilon_{iji})\right\}. \tag{4.2}$$

A general formula for the asymptotic covariance matrix, Ψ_i, can be found in Magnus (1988). When sampling from a normal distribution, this covariance matrix simplifies to

$$\begin{aligned}\Psi_i = \tfrac{1}{2}(I_m + K_{mm})\{&\Omega_i\otimes\Omega_i - (I_m\otimes\Omega_i)\Lambda_m(\Omega_i\otimes\Omega_i) - (\Omega_i\otimes\Omega_i)\Lambda_m(I_m\otimes\Omega_i)\\ &+(I_m\otimes\Omega_i)\Lambda_m(\Omega_i\otimes\Omega_i)\Lambda_m(I_m\otimes\Omega_i)\}(I_m + K_{mm}),\end{aligned} \tag{4.3}$$

where $\Lambda_m = \sum_{i=1}^{m}(e_ie_i' \otimes e_ie_i')$. Thus, in this case, a simple consistent estimator of Υ_{ijl} can be formed by replacing each Ω_h in Υ_{ijl} by R_h. If these estimators are used in (4.1) and (4.2) to obtain the estimators $\hat{\mu}_T$ and $\hat{\sigma}_T^2$, then the values of c and d in the $c\chi_d^2$ approximation of T are given by $c = \frac{1}{2}\hat{\sigma}_T^2/\hat{\mu}_T$ and $d = 2\hat{\mu}^2/\hat{\sigma}_T^2$.

5. Some examples

In this section, we illustrate the methods of this paper by looking at some examples that have appeared in the literature.

Example 1. Here we consider the iris data first analyzed by Anderson (1935) and Fisher (1936). Random samples of 50 plants from each of the three species of iris, versicolor, virginica, and setosa were used. From each plant, the four measurements, sepal length, sepal width, petal length and petal width were obtained. Flury (1984) analyzed these data and found that the CPC model is not appropriate for the three covariance matrices corresponding to the three species. We will consider the CPC model applied to the correlation matrices instead of the covariance matrices. Before proceeding to this analysis, we first check to see if the three correlation matrices are identical. This can be done by using the Wald statistic developed in Schott (1996). The computed value of this statistic using the sample correlation matrices of this example, which are given in Table 1a, is $W = 58.71$, and this can be compared to the quantiles of the chi-squared distribution with 12 degrees of freedom. Thus, the

Table 1
CPC correlation matrices for Iris data

(a) *Sample correlation matrices*:

$$R_1 = \begin{pmatrix} 1.00 & 0.53 & 0.75 & 0.55 \\ 0.53 & 1.00 & 0.56 & 0.66 \\ 0.75 & 0.56 & 1.00 & 0.79 \\ 0.55 & 0.66 & 0.79 & 1.00 \end{pmatrix}, \quad R_2 = \begin{pmatrix} 1.00 & 0.46 & 0.86 & 0.28 \\ 0.46 & 1.00 & 0.40 & 0.54 \\ 0.86 & 0.40 & 1.00 & 0.32 \\ 0.28 & 0.54 & 0.32 & 1.00 \end{pmatrix}, \quad R_3 = \begin{pmatrix} 1.00 & 0.74 & 0.27 & 0.28 \\ 0.74 & 1.00 & 0.18 & 0.23 \\ 0.27 & 0.18 & 1.00 & 0.33 \\ 0.28 & 0.23 & 0.33 & 1.00 \end{pmatrix}$$

(b) *Initial estimates*:

$$\hat{P}_* = \begin{pmatrix} 0.51 & -0.53 & -0.64 & -0.24 \\ 0.49 & 0.52 & 0.21 & -0.67 \\ 0.52 & -0.44 & 0.69 & 0.25 \\ 0.48 & 0.50 & -0.28 & 0.66 \end{pmatrix}, \quad \begin{aligned} \hat{d}'_{*1} &= (2.92 \quad 0.51 \quad 0.18 \quad 0.39), \\ \hat{d}'_{*2} &= (2.44 \quad 0.96 \quad 0.15 \quad 0.45), \\ \hat{d}'_{*3} &= (2.01 \quad 0.46 \quad 0.57 \quad 0.96). \end{aligned}$$

(c) *Refined estimates*:

$$\hat{P} = \begin{pmatrix} 0.50 & -0.50 & -0.50 & -0.50 \\ 0.50 & 0.50 & 0.50 & -0.50 \\ 0.50 & -0.50 & 0.50 & 0.50 \\ 0.50 & 0.50 & -0.50 & 0.50 \end{pmatrix}, \quad \begin{aligned} \hat{d}'_1 &= (2.92 \quad 0.51 \quad 0.18 \quad 0.39), \\ \hat{d}'_2 &= (2.44 \quad 0.96 \quad 0.15 \quad 0.45), \\ \hat{d}'_3 &= (2.01 \quad 0.46 \quad 0.57 \quad 0.96). \end{aligned}$$

(d) *Estimated CPC correlation matrices*:

$$\hat{\Omega}_1 = \begin{pmatrix} 1.00 & 0.66 & 0.72 & 0.55 \\ 0.66 & 1.00 & 0.55 & 0.72 \\ 0.72 & 0.55 & 1.00 & 0.65 \\ 0.55 & 0.72 & 0.65 & 1.00 \end{pmatrix}, \quad \hat{\Omega}_2 = \begin{pmatrix} 1.00 & 0.45 & 0.70 & 0.30 \\ 0.45 & 1.00 & 0.29 & 0.70 \\ 0.70 & 0.29 & 1.00 & 0.45 \\ 0.30 & 0.70 & 0.45 & 1.00 \end{pmatrix}, \quad \hat{\Omega}_3 = \begin{pmatrix} 1.00 & 0.49 & 0.24 & 0.29 \\ 0.49 & 1.00 & 0.29 & 0.24 \\ 0.24 & 0.29 & 1.00 & 0.49 \\ 0.29 & 0.24 & 0.49 & 1.00 \end{pmatrix}$$

Note: 1 = Versicolor, 2 = Virginica, 3 = Setosa.

hypothesis of equal correlation matrices would be rejected at any reasonable significance level. Next, we will check the adequacy of the CPC model. Using Theorem 4 and the covariance structure given in (4.3), we find that the estimated mean and variance of T are $\hat{\mu}_T = 12.01$ and $\hat{\sigma}_T^2 = 58.87$ and these lead to $c = 2.45$ and $d = 5$, where d has been rounded to the nearest integer. Thus, $T/c = 14.11$ will be compared to the quantiles of the chi-squared distribution with 5 degrees of freedom. This yields a p-value between .01 and 0.025, so that the CPC model does fit to some degree although the fit is not particularly great.

The initial estimates of the common eigenvectors computed by the F–G algorithm, along with the associated eigenvalue estimates, can be found in Table 1b. The CPC model was fitted using $r = 1, 2$, and 3, and this yielded the values $\tau_1 = 0.62$, $\tau_2 = 0.50$, and $\tau_3 = 0.47$ for the measures of closeness of the resulting estimators to the original sample correlation matrices. Thus, we will estimate the common correlation matrices using $\hat{P}_3$. The matrix $\hat{P} = \hat{P}_3$ and the refined eigenvalues $\hat{d}_i$ computed from it, are given in Table 1c, while the corresponding estimated CPC correlation matrices can be found in Table 1d. For this particular example, the common eigenvectors have very simple structure, a consequence of the fact that $\text{rank}(\hat{P} \odot \hat{P}) = 1$. For all three species, the first principal component corresponds to the average of the four standarized variables, and so it is an overall measure of size. The remaining principal components

Table 2
CPC correlation matrices for bankruptcy data

(a) *Sample correlation matrices*:

$$R_1 = \begin{pmatrix} 1.00 & 0.60 & -0.13 & 0.12 & 0.33 \\ 0.60 & 1.00 & 0.45 & 0.05 & -0.19 \\ -0.13 & 0.45 & 1.00 & 0.06 & -0.78 \\ 0.12 & 0.05 & 0.06 & 1.00 & 0.03 \\ 0.33 & -0.19 & -0.78 & 0.03 & 1.00 \end{pmatrix}, \quad R_2 = \begin{pmatrix} 1.00 & 0.50 & 0.11 & 0.33 & 0.15 \\ 0.50 & 1.00 & 0.29 & 0.48 & 0.06 \\ 0.11 & 0.29 & 1.00 & 0.36 & 0.27 \\ 0.33 & 0.48 & 0.36 & 1.00 & -0.08 \\ 0.15 & 0.06 & 0.27 & -0.08 & 1.00 \end{pmatrix}.$$

(b) *Initial estimates*:

$$\hat{P}_* = \begin{pmatrix} 0.54 & -0.65 & -0.49 & -0.14 & -0.15 \\ 0.59 & 0.68 & -0.10 & -0.42 & -0.07 \\ 0.26 & -0.28 & 0.54 & -0.32 & 0.68 \\ 0.53 & 0.06 & 0.30 & 0.79 & -0.05 \\ -0.03 & 0.20 & -0.61 & 0.29 & 0.71 \end{pmatrix}, \quad \begin{aligned} \hat{d}'_{*1} &= (1.61 \quad 0.20 \quad 1.74 \quad 1.26 \quad 0.18), \\ \hat{d}'_{*2} &= (2.03 \quad 0.44 \quad 0.89 \quad 0.47 \quad 1.17). \end{aligned}$$

(c) *Refined estimates*:

$$\hat{P} = \begin{pmatrix} 0.54 & -0.71 & -0.31 & -0.17 & -0.28 \\ 0.57 & 0.66 & -0.24 & -0.43 & -0.04 \\ 0.27 & -0.22 & 0.59 & -0.36 & 0.64 \\ 0.55 & 0.12 & 0.28 & 0.78 & -0.01 \\ 0.01 & -0.04 & -0.65 & 0.24 & 0.72 \end{pmatrix}, \quad \begin{aligned} \hat{d}'_1 &= (1.54 \quad 0.64 \quad 1.83 \quad 0.63 \quad 0.36), \\ \hat{d}'_2 &= (2.04 \quad 0.42 \quad 0.88 \quad 0.50 \quad 1.15). \end{aligned}$$

(d) *Estimated CPC correlation matrices*:

$$\hat{\Omega}_1 = \begin{pmatrix} 1.00 & 0.36 & -0.04 & 0.17 & 0.30 \\ 0.36 & 1.00 & -0.02 & 0.21 & 0.20 \\ -0.04 & -0.02 & 1.00 & 0.33 & -0.58 \\ 0.17 & 0.21 & 0.33 & 1.00 & -0.21 \\ 0.30 & 0.20 & -0.58 & -0.21 & 1.00 \end{pmatrix}, \quad \hat{\Omega}_2 = \begin{pmatrix} 1.00 & 0.55 & 0.03 & 0.44 & -0.05 \\ 0.55 & 1.00 & 0.18 & 0.46 & 0.05 \\ 0.03 & 0.18 & 1.00 & 0.29 & 0.16 \\ 0.44 & 0.46 & 0.29 & 1.00 & -0.06 \\ -0.05 & 0.05 & 0.16 & -0.06 & 1.00 \end{pmatrix}.$$

Note: 1 = Bankrupt, 2 = Solvent.

are simple contrasts in the four standardized variables; the associated eigenvalues are ordered the same for the versicolor and virginica species but differently for setosa. It should be noted that the refined eigenvalues are identical to the initial eigenvalues, and this will be the case whenever $\operatorname{rank}(\hat{P} \odot \hat{P}) = 1$ since we will then have $A_{m-1}A'_{m-1} = (I_m - m^{-1}1_m1'_m)$, and consequently, $A_{m-1}A'_{m-1}\gamma = \gamma - 1_m$ for any vector γ satisfying $\gamma' 1_m = m$.

Example 2. The data for this example can be found in Morrison (1990) and originates from a study by Altman (1968). Five financial ratios were obtained for each of 33 bankrupt firms and 33 solvent firms. These variables were (working capital)/(total assets), (retained earnings)/(total assets), (earnings before interest and taxes)/(total assets), (market value equity)/(book value of total liabilities), and sales/(total assets). The two sample correlation matrices can be found in Table 2a. The test for equal correlation matrices yields $W = 31.90$, which exceeds the 99.5th quantile of the chi-squared distribution with 10 degrees of freedom. The test for CPC structure produces the statistic $T = 12.81$, and the associated quantities $\hat{\mu}_T = 12.78$, $\hat{\sigma}^2_T = 76.39$, $c = 2.99$, and $d = 4$. Upon comparing $T/c = 4.29$ to the quantiles of the chi-squared distribution with 4 degrees of freedom, we find that the CPC model fits at any reasonable significance level.

Since $m = 5$, $\text{rank}(P \odot P)$ must be either 4, 3, or 2, and these correspond to 1, 2, and 3 for values of r. In estimating the correlation matrices, we will use $\hat{P} = \hat{P}_2$ since $\tau_1 = 1.05$, $\tau_2 = 0.85$, while $\tau_3 = 1.19$. Some of the relevant computations can be found in the remaining parts of Table 2. The common eigenvectors are not as easy to interpret as those of the previous example. However, one of the principal components, the one corresponding to the first refined eigenvector, could be roughly described as an overall measure of the size of the first four standardized variables. Note that this is the first principal component for the solvent group, but the second principal component for the bankrupt group. The principal components corresponding to the third and fifth refined eigenvectors may be useful in contrasting the two groups since the corresponding eigenvalues are quite different. The eigenvalue corresponding to the third refined eigenvector is much larger in the bankrupt group than the solvent group, while the opposite is true for the fifth refined eigenvector.

Appendix: proofs

Proof of Theorem 2. When at least one of the components of $1_m + A_r A_r' \gamma$ is negative, we can obtain the vector $\tilde{\gamma}$ that minimizes the distance to γ by using the method of Lagrange multipliers. Since $\tilde{\gamma} = 1_m + A_r c$ for some $r \times 1$ vector c, we will consider the function of c and an $m \times 1$ vector z

$$f(c, z) = (1_m + A_r b - \gamma)'(1_m + A_r c - \gamma) + \alpha'\{z \odot z - (1_m + A_r c)\},$$

where α is the $m \times 1$ vector of Lagrange multipliers. Note that the introduction of the vector z is used to guarantee the nonnegativity of $1_m + A_r c$. Differentiation of f with respect to c and with respect to z yields the two equations

$$2(1_m + A_r c - \gamma)' A_r - \alpha' A_r = 0', \tag{A.1}$$

$$\alpha \odot z = 0, \tag{A.2}$$

which will be used along with the constraint

$$z \odot z - (1_m + A_r c) = 0 \tag{A.3}$$

to obtain the solution for c. It follows from (A.2) that for each i, $\alpha_i = 0$ or $z_i = 0$. Let $Q = (\,Q_1' \quad Q_2'\,)'$ be any permutation matrix for which $\alpha_* = Q\alpha$ has the form

$$\alpha_* = \begin{pmatrix} Q_1 \alpha \\ Q_2 \alpha \end{pmatrix} = \begin{pmatrix} 0 \\ \alpha_2 \end{pmatrix},$$

where each component of α_2 is nonzero. Similarly, define

$$B = QA_r = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \qquad z_* = Qz = \begin{pmatrix} z_1 \\ 0 \end{pmatrix}, \qquad \gamma_* = Q\gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}.$$

Now, since $A_r'A_r = I_r$ and $1_m'A_r = 1_m'(\hat{P} \odot \hat{P})'A_r = 1_m'(0) = 0'$, we find by using (A.1) that

$$c = A_r'\left(\tfrac{1}{2}\alpha + \gamma\right) = B'\left(\tfrac{1}{2}\alpha_* + \gamma_*\right),$$

and so it follows that

$$\tilde{\gamma} = 1_m + A_r c = 1_m + Q'BB'\left(\tfrac{1}{2}\alpha_* + \gamma_*\right) = 1_m + \tfrac{1}{2}Q'BB_2'\alpha_2 + Q'BB'\gamma_*. \qquad \text{(A.4)}$$

Using (A.3) and the fact that z_* has the form $(z_1' \quad 0')'$, we find that

$$1_s + B_2 c = 1_s + \tfrac{1}{2}B_2B_2'\alpha_2 + B_2B'\gamma_* = 0,$$

where s corresponds to the number of rows in B_2. Solving this system of linear equations for α_2, we obtain

$$\alpha_2 = -2(B_2B_2')^+(1_s + B_2B'\gamma_*) + \{I_s - (B_2B_2')^+B_2B_2'\}u,$$

where u is an arbitrary $s \times 1$ vector. Consequently,

$$B_2'\alpha_2 = -2B_2^+(1_s + B_2B'\gamma_*).$$

Substituting this in (A.4) and then simplifying, we obtain

$$\tilde{\gamma} = 1_m - A_rB_2^+1_s + Q_1'B_1(I_r - B_2^+B_2)B_1'\gamma_1,$$

and so the result follows.

Before proving Theorem 3, we will need to establish the consistency of the estimator $\hat{P}_*$.

Lemma A.1. *Suppose that the CPC model holds and that for each j there is one of the correlation matrices, say Ω_i, such that d_{ij} is a simple eigenvalue of Ω_i. In addition, suppose that the columns of $\hat{P}_*$ have been signed so that the first nonzero element in each of the columns of P has the same sign as the corresponding element in $\hat{P}_*$. Then $\hat{P}_*$ converges in probability to P.*

Proof of Lemma 1. We will prove the result by using the implicit function theorem to show that $\hat{P}_*$ is a continuous function of $R_1, \ldots, R_k$ at $R_1 = \Omega_1, \ldots, R_k = \Omega_k$, where the matrices $\Omega_1, \ldots, \Omega_k$ satisfy the CPC model. The consistency of $\hat{P}_*$ will then follow from the consistency of the sample correlation matrices $R_1, \ldots, R_k$. Note that $\hat{P}_*$ can be found by minimizing the function

$$f(P, \Delta) = \log \Phi(P) + \operatorname{tr}\{\Delta(P'P - I_m)\},$$

where Δ is a symmetric matrix containing Lagrange multipliers. Thus, $\hat{P}_*$ can be found as the solution for P when solving the system of equations

$$u_1 = \sum_{i=1}^{k}\sum_{j=1}^{m} n_i(e_j'P'R_iPe_j)^{-1}\operatorname{vec}(R_iPe_je_j') + \operatorname{vec}(P\Delta) = 0 \qquad \text{(A.5)}$$

and $u_2 = D_m^+ \operatorname{vec}(P'P - I_m) = 0$ for P and Δ; here D_m is the $m^2 \times \frac{1}{2}m(m+1)$ duplication matrix (see Magnus, 1988, Section 4.3). Differentiation of $u = (u_1', u_2')'$ with respect to $(\operatorname{vec}(P)', \{D_m^+ \operatorname{vec}(\Delta)\}')'$ yields the Jacobian matrix $\Gamma J \Gamma'$, where

$$J = \begin{pmatrix} J_{11} & D_m \\ 2D_m^+ & (0) \end{pmatrix}, \qquad \Gamma = \begin{pmatrix} I_m \otimes P & (0) \\ (0) & I_m \end{pmatrix}$$

and

$$J_{11} = (\Delta \otimes I_m) + \sum_{i=1}^{k} \sum_{j=1}^{m} n_i \{(e_j' P' R_i P e_j)^{-1} (e_j e_j' \otimes P' R_i P)$$
$$-2(e_j' P' R_i P e_j)^{-2} (e_j e_j' \otimes P' R_i P e_j e_j' P' R_i P)\}.$$

When $R_i = \Omega_i = P D_i P'$, we find from (A.5) that $\Delta = -(\sum n_i) I_m$ and

$$J_{11} = \sum_{i=1}^{k} n_i \left\{ (D_i^{-1} \otimes D_i) - (I_m \otimes I_m) - 2 \sum_{j=1}^{m} (e_j e_j' \otimes e_j e_j') \right\}.$$

This diagonal matrix J_{11} has its $\{(i-1)m+i\}$th diagonal element nonzero, and since the jth diagonal element of D_i is distinct for some i, it is easily shown by using the arithmetic–geometric mean inequality that for every $j \neq l$, the $\{(j-1)m+l\}$th and $\{(l-1)m+j\}$th diagonal elements of J_{11} cannot both be 0. This along with the structure of D_m guarantees that the first m^2 rows of J are linearly independent. Since D_m^+ has full row rank it then follows that J is nonsingular when evaluated at the CPC Ω_i and, consequently, so is $\Gamma J \Gamma'$. Thus, the implicit function theorem applies and so the proof is complete.

Proof of Theorem 3. The proof will be complete if we can show that $\hat{P}$ and $\hat{d}_i$ are consistent estimators of P and d_i since $\hat{\Omega}_i$ is a continuous function of these quantities. Now, since both P and $\hat{P}$ are in S_r, and $\hat{P}$ is the choice of $Q \in S_r$ so that $\operatorname{vec}(Q - \hat{P}_*)' \operatorname{vec}(Q - \hat{P}_*)$ is minimized, it follows that

$$\operatorname{vec}(\hat{P} - P)' \operatorname{vec}(\hat{P} - P) \leq \operatorname{vec}(\hat{P}_* - P)' \operatorname{vec}(\hat{P}_* - P) + \operatorname{vec}(\hat{P} - \hat{P}_*)' \operatorname{vec}(\hat{P} - \hat{P}_*)$$
$$\leq 2 \operatorname{vec}(\hat{P}_* - P)' \operatorname{vec}(\hat{P}_* - P).$$

Thus, we see that $\hat{P}$ must converge in probability to P since $\hat{P}_*$ converges in probability to P. Due to the consistency of $\hat{P}$, $(\hat{P} \odot \hat{P})'(\hat{P} \odot \hat{P})$ is also consistent for $(P \odot P)'(P \odot P)$. If the columns of the $m \times r$ matrix B_r are orthonormal eigenvectors of $(P \odot P)'(P \odot P)$ corresponding to its zero eigenvalue, and the columns of A_r are orthonormal eigenvectors of $(\hat{P} \odot \hat{P})'(\hat{P} \odot \hat{P})$ corresponding to its r smallest eigenvalues, then it follows from the continuity of eigenprojections (see Kato, 1982) that $A_r A_r'$ is a consistent estimator of $B_r B_r'$. In addition, the consistency of $\hat{d}_{*i}$ follows from that of $\hat{P}_*$ and R_i. Consequently, $1_m + A_r A_r' \hat{d}_{*i}$ converges in probability to $1_m + B_r B_r' d_i = d_i$, and this is sufficient to guarantee that $\hat{d}_i$ converges in probability to d_i since each component of d_i is positive. This completes the proof. □

Proof of Theorem 4. Let $A_i = R_i - \Omega_i$, so it follows that, asymptotically, $n_i^{1/2}\text{vec}(A_i) \sim N_{m^2}(0, \Psi_i)$. If the CPC model holds, then we find that

$$R_iR_j - R_jR_i = A_i\Omega_j + \Omega_iA_j - \Omega_jA_i - A_j\Omega_i.$$

Using this and the fact that $\text{vec}(\Omega_jA_i) = K_{mm}\text{vec}(A_i\Omega_j)$ and $(I_m - K_{mm}) = \tilde{D}_m\tilde{D}_m'$, we also find that

$$n^{1/2}\eta_i\eta_j\,\text{vec}(R_iR_j - R_jR_i) = \tilde{D}_m\tilde{D}_m'\{\eta_j(\Omega_j \otimes I_m)n_i^{1/2}\,\text{vec}(A_i) - \eta_i(\Omega_i \otimes I_m)n_j^{1/2}\,\text{vec}(A_j)\}. \tag{A.6}$$

Using (A.6), we can obtain the matrix H for which we have $T = a'Ha$, where $a = (n_1^{1/2}\,\text{vec}(A_1)', \ldots, n_k^{1/2}\,\text{vec}(A_k)')'$. Thus, using standard results on quadratic forms in normal variables, we have, asymptotically, $\mu_T = \text{tr}(H\Psi)$ and $\sigma_T^2 = 2\,\text{tr}\{(H\Psi)^2\}$, where Ψ is the asymptotic covariance matrix of a, that is, the block diagonal matrix $\text{diag}(\Psi_1, \ldots, \Psi_k)$. The result then follows after some straightforward simplification.

References

Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J. Finance 23, 589–609.

Anderson, E., 1935. The irises of the gaspe peninsula. Bull. Amer. Iris Soc. 59, 2–5.

Chen, K.H., Robinson, J., 1989. Comparison of factor spaces of two related populations. J. Multivariate Anal. 28, 190–203.

Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Ann. Eugenics 7, 179–88.

Fletcher, R., 1987. Practical Methods of Optimization. Wiley, New York.

Flury, B.N., 1984. Common principal components in k groups. J. Amer. Statist. Assoc. 79, 892–8.

Flury, B.N., 1986. Asymptotic theory for common principal component analysis. Ann. Statist. 14, 418–30.

Flury, B.N., 1987. Two generalizations of the common principal component model. Biometrika, 74, 59–69.

Flury, B.N., 1988. Common Principal Components and Related Multivariate Models. Wiley, New York.

Flury, B.N., Constatine, G., 1985. Algorithm AS 211, the F–G diagonalization algorithm. Appl. Statist. 34, 177–83.

Flury, B.N., Gautschi, W., 1986. An algorithm for simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly diagonal form. SIAM J. Sci. Statist. Comput. 7, 169–84.

Fujioka, T., 1993. An approximate test for common principal component subspaces in two groups. Ann. Inst. Statist. Math. 45, 147–58.

Hedayat, A., Wallis, W.D., 1978. Hadamard matrices and their applications. Ann. Statist. 6, 1184–238.

Kato, T., 1982. A Short Introduction to Perturbation Theory for Linear Operators. Springer, New York.

Keramidas, E.M., Devlin, S.J., Gnanadesikan, R., 1987. A graphical procedure for comparing the principal components of several covariance matrices. Commun. Statist., Part B–Comput. Simulation 16, 161–91.

Krzanowski, W.J., 1979. Between-group comparison of principal components. J. Amer. Statist. Assoc. 74, 703–7.

Krzanowski, W.J., 1982. Between-group comparison of principal components–some sampling results. J. Statist. Comput. Simulation. 15, 141–54.

Krzanowski, W.J., 1984. Principal component analysis in the presence of group structure. Appl. Statist. 33, 164–8.

Krzanowski, W.J., 1993. Permutational tests for correlation matrices. Statist. Comput. 3, 37–44.

Magnus, J.R., 1988. Linear Structures. Oxford University Press, New York.

Morrison, D.F., 1990. Multivariate Statistical Methods. McGraw-Hill, New York

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numer. Recipes in Fortran, 2nd ed.. Cambridge University Press, Cambridge, UK.

Schott, J.R., 1988. Common principal component subspaces in two groups. Biometrika 75, 229–36.

Schott, J.R., 1991. Some tests for common principal component subspaces in several groups. Biometrika, 78, 771–7.

Schott, J.R., 1996. Testing for the equality of several correlation matrices. Statist. Probab. Lett. 27, 85–9.

Schott, J.R., 1997a. Asymptotics of Eigenprojections of Correlation Matrices with Some Applications in Principal Components Analysis. Biometrika 84, 327–37.

Schott, J.R., 1997b. Matrix Analysis for Statistics. Wiley, New York.