Two Generalizations of the Common Principal Component Model

Author(s): Bernhard K. Flury

Source: *Biometrika*, Mar., 1987, Vol. 74, No. 1 (Mar., 1987), pp. 59-69

Published by: Oxford University Press on behalf of Biometrika Trust

Stable URL: http://www.jstor.com/stable/2336021

Biometrika (1987), **74**, 1, *pp.* 59–69
Printed in Great Britain

Two generalizations of the common principal component model

BY BERNHARD K. FLURY
Department of Mathematics, Victoria University, Wellington, New Zealand

SUMMARY

Under the common principal component model the covariance matrices Ψ_i of k populations are assumed to have identical eigenvectors, that is, the same orthogonal matrix diagonalizes all Ψ_i simultaneously. This paper modifies the common principal component model by assuming that only q out of p eigenvectors are common to all Ψ_i, while the remaining $p-q$ eigenvectors are specific in each group. This is called a partial common principal component model. A related modification assumes that q eigenvectors of each matrix span the same subspace, a problem that was first considered by Krzanowski (1979). For both modifications this paper derives the normal theory maximum likelihood estimators. It is shown that approximate maximum likelihood estimates can easily be computed if estimates of the ordinary common principal component model are available. The methods are illustrated by numerical examples.

Some key words: Common principal components; Common space; Covariance matrix; Diagonalization; Eigenvalue; Eigenvector; Maximum likelihood.

1. INTRODUCTION: PARTIAL COMMON PRINCIPAL COMPONENTS

Common principal component analysis is a generalization of the well-known principal component analysis to several groups. The basic assumption is that the $p \times p$ covariance matrices $\Psi_1, \ldots, \Psi_k$ of k populations are simultaneously diagonalizable, i.e. that the hypothesis of common principal components

$$H_c: \quad B'\Psi_i B = \Lambda_i \quad (i = 1, \ldots, k),$$

where Λ_i is diagonal, holds for some orthogonal $p \times p$ matrix B. Estimation of the Ψ_i under H_c and an asymptotic chi-squared test for H_c under normality assumptions have been discussed by Flury (1984). An algorithm for computing exact maximum likelihood estimates of common principal components has been developed by Flury & Gautschi (1986) and programmed by Flury & Constantine (1985). Krzanowski (1984) has shown that a simpler method of estimation, namely taking the eigenvectors of a weighted sum of $S_1, \ldots, S_k$, is often sufficient for practical purposes.

In practical applications of principal component analysis, especially if the number of variables, p, is large, the investigator is often interested only in a subset of $q < p$ components, usually those associated with the largest eigenvalues. Similarly in the k-sample case we may wish to concentrate on q out of p components, provided that they recover most of the variability in each of the k groups simultaneously. This problem has been addressed by Flury (1986a) with a generalization of the criterion proposed by Anderson (1963, p. 135) for the one-sample case. This criterion is based on the asymptotic distribution of the maximum likelihood estimates of the Ψ_i under H_c. If we wish to discard $p-q$ components, however, then we should actually not care whether these components satisfy H_c or not. In practice it may occur that the common principal component model is rejected due to its inappropriateness for the components that are to be discarded anyway, while the components we are interested in may actually be common to all groups. A more appropriate model can therefore be defined by the hypothesis of partial common principal components

$$H_c(q): \quad B^{(i)\prime}\Psi_i B^{(i)} = \Lambda_i \quad (i=1,\dots,k),$$

with $B^{(i)} = (B_1 : B_2^{(i)})$, and Λ_i diagonal, where all $B^{(i)}$ are orthogonal $p \times p$ matrices, the common part B_1 has dimension $p \times q$ ($q \leqslant p-2$), and the $B_2^{(i)}$ are $p \times (p-q)$. Note that $H_c(p-1)$ implies H_c since all $B^{(i)}$ are orthogonal. Conversely, H_c implies $H_c(q)$ for all q between 1 and $p-1$; the partial model is thus more general than the ordinary common principal component model. Note also that no canonical order of the columns of the $B^{(i)}$ need be given, since the rank order of the diagonal elements of the Λ_i is not necessarily the same for all Λ_i. We assume, however, that some order of the q common components is defined such that we can speak unambiguously of a 'first' component, 'second' component, etc.

If we write

$$B^{(i)} = (\beta_1, \dots, \beta_q, \beta_{q+1}^{(i)}, \dots, \beta_p^{(i)}), \quad \Lambda_i = \operatorname{diag}(\lambda_{i1}, \dots, \lambda_{ip}) \quad (i=1,\dots,k),$$

the partial model can be written in spectral decomposition form as

$$\Psi_i = \sum_{j=1}^{q} \lambda_{ij}\beta_j\beta_j' + \sum_{j=q+1}^{p} \lambda_{ij}\beta_j^{(i)}\beta_j^{(i)\prime} \quad (i=1,\dots,k).$$

Here β_1 to β_q are the common eigenvectors of $\Psi_1, \dots, \Psi_k$, while $\beta_j^{(i)}$ $(j=q+1,\dots,p)$ are the specific eigenvectors of Ψ_i.

2. Maximum likelihood estimation of partial common components

Estimation is based (Flury, 1984) on independent Wishart matrices S_i, typically sample covariance matrices from normal samples of size $N_i = n_i + 1$, such that $n_i S_i$ is distributed as $W_p(n_i, \Psi_i)$. The common likelihood function of $\Psi_1, \dots, \Psi_k$, given $S_1, \dots, S_k$, is

$$L(\Psi_1, \dots, \Psi) = C \times \prod_{i=1}^{k} \operatorname{etr}(-\tfrac{1}{2} n_i \Psi_i^{-1} S_i)|\Psi|^{-\frac{1}{2}n_i},$$

where C is a constant that does not depend on the Ψ_i. Equivalently to maximizing the likelihood, we can minimize

$$g(\Psi_1, \dots, \Psi_k) = -2\log L(\Psi_1, \dots, \Psi_k) + 2\log C = \sum_{i=1}^{k} n_i(\log|\Psi_i| + \operatorname{tr}\Psi_i^{-1}S_i).$$

Assuming that $H_c(q)$ holds for a fixed q, and using properties of the trace and the determinant, we get

$$g = \sum_{i=1}^{k} n_i\left(\sum_{j=1}^{p} \log\lambda_{ij} + \sum_{j=1}^{q} \beta_j' S_i \beta_j/\lambda_{ij} + \sum_{j=q+1}^{p} \beta_j^{(i)\prime} S_i \beta_j^{(i)}/\lambda_{ij}\right).$$

This function is to be minimized under the restriction of orthogonality of all $B^{(i)}$; that is

$$\beta'_h\beta_j = \begin{cases} 0 & (h \neq j) \\ 1 & (h = j) \end{cases} \quad (1 \leqslant h, j \leqslant q);$$

$$\beta_h^{(i)\prime}\beta_j^{(i)} = \begin{cases} 0 & (h \neq j) \\ 1 & (h = j) \end{cases} \quad (q < h, j \leqslant p;\ i = 1, \ldots, k);$$

$$\beta'_h\beta_j^{(i)} = 0 \quad (i = 1, \ldots, k;\ 1 \leqslant h \leqslant q < j \leqslant p). \tag{2·1}$$

Using appropriate Lagrange multipliers, the same technique as used by Flury (1984) yields the likelihood equations

$$\lambda_{ij} = \begin{cases} \beta'_j S_i \beta_j & (j = 1, \ldots, q), \\ \beta_j^{(i)\prime} S_i \beta_j^{(i)} & (j = q+1, \ldots, p), \end{cases} \quad (i = 1, \ldots, k), \tag{2·2}$$

$$\beta_h^{(i)\prime} S_i \beta_j^{(i)} = 0 \quad (h \neq j), \tag{2·3}$$

$$\beta'_l \left(\sum_{i=1}^{k} \frac{\lambda_{il} - \lambda_{ih}}{\lambda_{il}\lambda_{ih}} n_i S_i \right) \beta_h = 0 \quad (1 \leqslant l, h \leqslant q, l \neq h), \tag{2·4}$$

$$\left(\frac{1}{\lambda_{mj}} - \frac{1}{\lambda_{ml}} \right) n_m \beta'_l S_m \beta_j^{(m)} = \beta_j^{(m)\prime} \left\{ \sum_{i=1 \neq m}^{k} \left(n_i S_i \beta_l / \lambda_{il} - \sum_{h=q+1}^{p} \delta_{lh}^{(i)} \beta_h^{(i)} \right) \right\}$$

$$(m = 1, \ldots, k;\ 1 \leqslant l \leqslant q < h \leqslant p), \tag{2·5}$$

where

$$\delta_{lh}^{(i)} = n_i \beta'_l S_i \beta_h^{(i)} / \lambda_{ih} \quad (i = 1, \ldots, k;\ 1 \leqslant l \leqslant q < h \leqslant p) \tag{2·6}$$

are the $k\,q(p-q)$ Lagrange multipliers introduced with regard to the restrictions (2·1). By (2·3), the specific eigenvectors $\beta_h^{(i)}$ satisfy the same type of equations as if principal components were estimated in each group individually. Equation system (2·4) is exactly the same as the one occurring in ordinary common principal components (Flury, 1984, p. 893), but this time valid only for l, $h \leqslant q$, that is for the common components. The equations (2·5) link the common and the specific components.

Solving the likelihood equations exactly is extremely laborious, and no simple or elegant method has been found yet to accomplish this. Some remarks on the numerical methods used are given in Appendix 1.

An approximate solution, however, can easily be obtained, provided that maximum likelihood estimates for the ordinary common principal component model are available. The approximation is based on the observation that, if the partial model holds, the q common components are estimated almost correctly in the ordinary common principal component model, irrespective of the other components that may be specific in each population. This statement will be justified by Lemma 1 below, for which we need the following preliminaries. For a $p \times p$ matrix F, let us denote by diag (F) the diagonal matrix having the same diagonal elements as F. Then maximizing the likelihood under the ordinary common principal component model is the same as minimizing the function

$$\Phi(B) = \Phi(B; S_1, \ldots, S_k) = \prod_{i=1}^{k} |\mathrm{diag}\,(B' S_i B)|^{n_i} / |B' S_i B|^{n_i} \tag{2·7}$$

over $B \in O(p)$, the group of orthogonal $p \times p$ matrices (Flury & Gautschi, 1986). Now the lemma can be stated as follows.

LEMMA 1. *Assume that the positive-definite symmetric matrices S_i of dimension $p \times p$ have $q < p$ common eigenvectors. Let these be denoted by $a_1, \ldots, a_q$. Then the matrix $B_0 \in O(p)$ that minimizes the function $\Phi(B)$ has $a_1, \ldots, a_q$ as columns, or can be chosen so if B_0 is not uniquely defined.*

A proof of Lemma 1 is given in Appendix 2. See also § 5 of Flury & Gautschi (1986) for an example of dimension $p=6$ with $q=2$ common eigenvectors of two matrices.

In practical applications, when the S_i are sample covariance matrices, they will in general not have exactly identical eigenvectors. Under the hypothesis $H_c(q)$ and for large samples, however, we can expect q sets of 'almost identical' eigenvectors. Since Φ is a continuous function, the maximum likelihood estimates of the q common eigenvectors will be close, that is asymptotically converge, to the estimates obtained under the ordinary common principal component model. As a consequence of this, the following extremely simple procedure can be used to obtain approximate maximum likelihood estimates.

(i) Compute $B=(B_1:B_2) \in O(p)$, where B_1 has q columns and $O(p)$ denotes the group of orthogonal $p \times p$ matrices, as maximum likelihood estimates of the model with all p components common. Since the order of the columns of B is not uniquely defined, it may be necessary to apply some permutation to the columns of B in order to put the 'candidates' for common components in the first q columns.

(ii) Rotate the matrices $B_2' S_i B_2$ to diagonal form; i.e. find matrices $Q_i \in O(p-q)$ such that $Q_i' B_2' S_i B_2 Q_i$ is diagonal, and put $B_2^{(i)} = B_2 Q_i$ $(i=1, \ldots, k)$.

(iii) Mark the matrices B_1 and $B_2^{(i)}$ found in steps (i) and (ii) above by a tilde to indicate that they are approximate, and put

$$\tilde{B}^{(i)} = (\tilde{B}_1 : \tilde{B}_2^{(i)}), \quad \tilde{\Lambda}_i = \operatorname{diag}(\tilde{B}^{(i)\prime} S_i \tilde{B}^{(i)}) = \operatorname{diag}(\tilde{\lambda}_{i1}, \ldots, \tilde{\lambda}_{ip}),$$
$$\tilde{\Psi}_i = \tilde{B}^{(i)} \tilde{\Lambda}_i \tilde{B}^{(i)\prime} \quad (i=1, \ldots, k).$$

These approximate maximum likelihood estimates satisfy equations (2·2), (2·3) and (2·4), but not necessarily (2·5). If $H_c(q)$ holds, however, the $\delta_{lh}^{(i)}$ in equation (2·6) will be relatively small, and the approximate estimates will not violate (2·5) grossly, at least for large samples. We will also justify this claim by the numerical example in § 4.

Denoting the exact maximum likelihood estimates by $\hat{B}^{(i)}$, $\hat{\Lambda}_i$ and $\hat{\Psi}_i$, respectively, we can construct an exact and an approximate log likelihood ratio statistic for $H_c(q)$ versus the alternative of arbitrary Ψ_i by

$$X_c^2 = -2 \log \frac{L(\hat{\Psi}_1, \ldots, \hat{\Psi}_k)}{L(S_1, \ldots, S_k)} = \sum_{i=1}^{k} n_i \log \frac{|\hat{\Psi}_i|}{|S_i|}, \tag{2·8}$$

$$X_{c,\mathrm{APP}}^2 = -2 \log \frac{L(\tilde{\Psi}_1, \ldots, \tilde{\Psi}_k)}{L(S_1, \ldots, S_k)} = \sum_{i=1}^{k} n_i \log \frac{|\tilde{\Psi}_i|}{|S_i|} \tag{2·9}$$

respectively. The number of parameters estimated under the alternative is $\frac{1}{2}kp(p+1)$. Under $H_c(q)$, the number of parameters estimated is as follows: kp parameters for eigenvalues, $\frac{1}{2}p(p-1)$ parameters for one of the orthogonal matrices, say $B^{(1)}$, and $\frac{1}{2}(k-1)(p-q)(p-q-1)$ parameters for the specific eigenvectors of the other $k-1$ covariance matrices. By the general theory of likelihood ratio tests, the null distribution of X_c^2 is asymptotically chi-squared with $\frac{1}{2}(k-1)\{p(p-1)-(p-q)(p-q-1)\}$ degrees of freedom. Since X_c^2 is based on the exact maximum of the likelihood function, we have always $X_c^2 < X_{c,\mathrm{APP}}^2$, and the approximate statistic can therefore be used to accept $H_c(q)$, but not necessarily to reject it. The reader should also be cautioned that formal hypothesis testing may not be appropriate in many cases; see the remarks in § 5.

As for ordinary common components (Flury, 1984, § 3), we can compute covariance and correlation matrices between the estimated principal components. In the present case, the covariance matrices

$$F_i = \hat{B}^{(i)\prime} S_i \hat{B}^{(i)} = \begin{bmatrix} \hat{B}_1' S_i \hat{B}_1 & \hat{B}_1' S_i \hat{B}_2^{(i)} \\ \hat{B}_2^{(i)\prime} S_i \hat{B}_1 & \hat{B}_2^{(i)\prime} S_i \hat{B}_2^{(i)} \end{bmatrix} \tag{2·10}$$

have a submatrix $\hat{B}_2^{(i)\prime} S_i \hat{B}_2^{(i)}$ of dimension $(p-q)\times(p-q)$ which is exactly diagonal. Note that $\hat{\Lambda}_i = \operatorname{diag}(F_i)$. Similarly, the correlation matrices can be written as

$$R_i = \hat{\Lambda}_i^{-\frac{1}{2}} F_i \hat{\Lambda}_i^{-\frac{1}{2}} = \begin{bmatrix} R_{11}^{(i)} & R_{12}^{(i)} \\ R_{21}^{(i)} & I_{p-q} \end{bmatrix}. \tag{2·11}$$

The matrices R_i and F_i can also be computed using approximate rather than exact estimates.

3. Common space analysis

The first attempt to generalize principal components to a k-group method was by Krzanowski (1979), whose approach consists essentially of a descriptive comparison of the subspaces spanned by the first q components in each group. This is achieved by computing minimal angles between these subspaces, and by approximating the k q-dimensional subspaces by a single q-dimensional subspace. The idea behind Krzanowski's procedure is that one would like to say that the first q components in all groups contain the same information. Later Krzanowski (1982) gave critical values that can be used to assess the significance of differences between principal components in different samples.

Krzanowski's method is mathematically appealing and computationally simple, but suffers from two shortcomings. First, if, say, the qth and the $(q+1)$st eigenvalues in one of the groups are close, then the associated eigenvectors are unstable in the sense of high variances of the coefficients (Anderson, 1963, p. 130). This may lead to large angles between the q-dimensional subspaces which could wrongly be interpreted as evidence against the hypothesis of a common subspace for all k groups. Secondly, even if the hypothesis of a common subspace for the first q components is assumed to hold, Krzanowski's method does not indicate how the covariance matrices should be estimated under this hypothesis. We are now going to discuss a modification of partial common components that serves the same purpose as Krzanowski's method, but avoids these two shortcomings.

Since the technique used here is very similar to the one used in § 2, we do not give all details. Let $B^{(i)} = (B_1^{(i)}, B_2^{(i)}) \in O(p)$ such that $B_1^{(i)}$ has dimension $p \times q$, $B_2^{(i)}$ has dimension $p\times(p-q)$, and $\Lambda_i = B^{(i)}\Psi_i B^{(i)}$ is diagonal. Then the hypothesis of common space spanned by q principal components can, for all i, $m \leq k$, be written as

$$H_{\mathrm{CS}}(q): \quad B_1^{(i)\prime} B_2^{(m)} = 0.$$

Another way of putting $H_{\mathrm{CS}}(q)$ is $H_{\mathrm{CS}}^*(q)$: there exist matrices $H_1^{(i)} \in O(q)$, $H_2^{(i)} \in O(p-q)$ $(i = 2, \ldots, k)$ such that $B_1^{(i)} = B_1^{(1)} H_1^{(i)}$ and $B_2^{(i)} = B_2^{(1)} H_2^{(i)}$.

This expresses that under $H_{\mathrm{CS}}(q)$ the first q, or last $p-q$, characteristic vectors in group i span the same subspace as those in group 1.

As in § 2, we use maxium likelihood estimation based on sample covariance matrices S_i such that $n_i S_i \sim W_p(n_i, \Psi_i)$. Let us write

$$B_1^{(i)} = (\beta_1^{(i)}, \ldots, \beta_q^{(i)}), \quad B_2^{(i)} = (\beta_{q+1}^{(i)}, \ldots, \beta_p^{(i)}), \quad \Lambda_i = \mathrm{diag}\,(\lambda_{i1}, \ldots, \lambda_{ip}).$$

To restrict the parameter space to $H_{\mathrm{CS}}(q)$, we use appropriate Lagrange multipliers, for instance for the restrictions

$$\beta_h^{(i)\prime}\beta_j^{(i)} = \begin{cases} 0 & (h \neq j) \\ 1 & (h = j) \end{cases} \quad (i = 1, \ldots, k), \tag{3·1}$$

$$\beta_h^{(1)\prime}\beta_j^{(i)} = 0 \quad (i = 2, \ldots, k;\; h = 1, \ldots, q;\; j = q+1, \ldots, p).$$

The likelihood equations are then

$$\lambda_{ij} = \beta_j^{(i)\prime} S_i \beta_j^{(i)} \quad (i = 1, \ldots, k;\; j = 1, \ldots, p), \tag{3·2}$$

$$\beta_l^{(i)\prime} S_i \beta_h^{(i)} = 0 \quad (i = 1, \ldots, k;\; 1 \leq l < h \leq q \text{ or } q < l < h \leq p). \tag{3·3}$$

The equations involving simultaneously eigenvectors $\beta_l^{(i)}$ for $l \leq q$ and $\beta_h^{(i)}$ for $h > q$ are more complicated. Again, as for the partial common component model, finding the exact maximum likelihood estimates is rather laborious; see Appendix 1 for some details. Approximate maxium likelihood estimates, however, can easily be obtained provided that estimates for the ordinary common principal component model are available. The approximation is based on the following lemma.

LEMMA 2. *Assume that the positive-definite symmetric matrices S_i $(i = 1, \ldots, k)$ of dimension $p \times p$ have q eigenvectors each that span the same q-dimensional subspace $\mathscr{A}_q$. Then matrix $B_0 \in O(p)$ that minimizes the function $\Phi(B)$ has q columns which span $\mathscr{A}_q$.*

For a proof, see Appendix 2.

The sample covariance matrices S_i will, of course, in general not satisfy the assumption of Lemma 2 exactly, but under $H_{\mathrm{CS}}(q)$ approximately. Approximate maximum likelihood estimates are therefore obtained as in § 2, but in addition the matrices $B_1' S_i B_1$ are also rotated to diagonal form. That is, we find matrices $P_i \in O(q)$ such that $P_i' B_1' S_i B_1 P_i$ is diagonal, put $\tilde{B}_1^{(i)} = B_1 P_i$, and take $\tilde{B}^{(i)} = (\tilde{B}_1^{(i)} : \tilde{B}_2^{(i)})$. Again, these approximate estimates satisfy (3·2) and (3·3), but not necessarily the other likelihood equations. The exact and approximate log likelihood ratio statistics X^2_{CS} and $X^2_{\mathrm{CS,APP}}$ respectively for $H_{\mathrm{CS}}(q)$ are formally identical with (2·8) and (2·9), where $\hat{\Psi}_i$ and $\hat{\hat{\Psi}}_i$ are approximate and exact maximum likelihood estimates under $H_{\mathrm{CS}}(q)$. The number of parameters estimated under $H_{\mathrm{CS}}(q)$ can be found from the definition above as follows: kp parameters for eigenvalues, $\frac{1}{2}p(p-1)$ parameters for the orthogonal matrix $B^{(1)}$, and

$$(k-1)\{\tfrac{1}{2}q(q-1) + \tfrac{1}{2}(p-q)(p-q-1)\}$$

parameters for the orthogonal matrices $H_i^{(1)}$ and $H_i^{(2)}$. The number of degrees of freedom associated with the log likelihood ratio criterion is thus

$$\tfrac{1}{2}(k-1)\{p(p-1) - q(q-1) - (p-q)(p-q-1)\}.$$

The method proposed here does not solve exactly the same problem as Krzanowski's (1979) method. The main difference is that the subspaces compared by Krzanowski's method are defined by the eigenvectors associated with the largest roots, while no conditions or thresholds are placed upon eigenvalues in our method. The hypothesis $H_{\mathrm{CS}}(q)$ covers also models in which the subspace associated with the largest eigenvalues in one group may be associated with the smallest eigenvalues in another group, although this situation is unlikely to occur in practice. One might of course try to prescribe some canonical ordering to the eigenvalues, but this would make maximum likelihood estimation extremely complicated.

4. Applications

We now illustrate the foregoing methods. The computations were to four exact decimal digits of the orthogonal matrices, but all results will be given in two decimal digits only in order to save space and to ease reading the tables. Moreover, the symmetric matrices F_i and R_i in (2·10) and (2·11) will be given in combined form: the elements of F_i, variances and covariances, on and above the diagonal, and the elements of R_i, correlations, below the diagonal.

Airoldi & Hoffmann (1984) took various skull measurements on two different species of voles, small rodents. The four variables considered here are: condylo-incisive length, alveolar length of upper molar toothrow, zygomatic width, and interorbital width. The raw data were provided by J. P. Airoldi. We selected adult individuals, age exceeding 90 days, and assigned them to four different groups: (1) *Microtus californicus*, males; (2) *Microtus californicus*, females; (3) *Microtus ochrogaster*, males; and (4) *Microtus ochrogaster*, females. The sample sizes $N_i = n_i + 1$ were 82, 70, 58 and 54, respectively. Since the sample variances of the four variables were quite different, but the coefficients of variation were close to $\frac{1}{30}$ for all variables and all groups, we decided to take logs for the principal component analysis.

Table 1 (a), shows the sample covariance matrices S_i. The first principal component can be expected to be of the 'size'-type (Jolicoeur & Mosimann, 1960), i.e. to have positive coefficients of approximately the same magnitude for all variables; so let us estimate the parameters of a model with one common component. We start with the ordinary common principal component model; the maximum likelihood estimate of the orthogonal matrix B is given in Table 1 (b). In Table 1 (c) the exact estimates under $H_C(1)$ are listed. Note how little $\hat{\beta}_1$ differs from the first column of $\hat{B}$ under H_C. Also, the second to fourth components are remarkably similar in groups 1 and 2, although estimated individually. Table 1 (d) gives variances, covariances and correlations of the principal components obtained under $H_C(1)$; the $\hat{\lambda}_{ij}$ are on the main diagonal. For comparison the eigenvalues of the S_i are also listed. The agreement between these eigenvalues and the $\hat{\lambda}_{ij}$'s seems quite good. The value of the log likelihood ratio statistic for $H_C(1)$ is $X^2_C = 12·75$ on 9 degrees of freedom, so the partial model seems reasonable. The value of the approximate statistic (2·9) would be $X^2_{C,\mathrm{APP}} = 13·16$, which illustrates again that the approximate method is actually quite good. We conclude that the 'size'-component can be estimated identically for all four groups, the coefficients being 0·54, 0·55, 0·56 and 0·29.

In this example one may wonder whether it is actually necessary to estimate partial components; the S_i are so similar that one might assume equality of the population covariance matrices Ψ_i. This would then, of course, imply the ordinary common principal component model and any partial model. It turns out that the log likelihood ratio criterion for the hypothesis $\Psi_1 = \Psi_2 = \Psi_3 = \Psi_4$ against the alternative of arbitrary Ψ_i yields a chi squared of 71·9 on 30 degrees of freedom. What about the proportionality, which also would imply H_C? For the hypothesis $\Psi_i = c_i \Psi_1$ $(i = 2, 3, 4)$ a likelihood ratio test can be constructed (Flury, 1986b) which gives a chi squared of 69·1 on 27 degrees of freedom. Thus the partial component model is not just trivially correct because of identity or proportionality. The deviation from proportionality is quite apparent from the $\hat{\lambda}_{ij}$, the

Table 1. *Partial common components in four groups of voles*

(a) *Sample covariance matrices* ($\times 10^4$)

M. californicus, male ($n_1 = 81$), S_1

$$\begin{bmatrix} 13{\cdot}55 & 10{\cdot}00 & 10{\cdot}65 & -2{\cdot}05 \\ 10{\cdot}00 & 18{\cdot}19 & 8{\cdot}90 & 3{\cdot}58 \\ 10{\cdot}65 & 8{\cdot}90 & 15{\cdot}32 & 3{\cdot}00 \\ -2{\cdot}05 & 3{\cdot}58 & 3{\cdot}00 & 27{\cdot}10 \end{bmatrix}$$

M. californicus, female ($n_2 = 69$), S_2

$$\begin{bmatrix} 18{\cdot}74 & 14{\cdot}78 & 15{\cdot}62 & 2{\cdot}51 \\ 14{\cdot}78 & 19{\cdot}00 & 15{\cdot}51 & 7{\cdot}02 \\ 15{\cdot}62 & 15{\cdot}51 & 19{\cdot}45 & 5{\cdot}91 \\ 2{\cdot}51 & 7{\cdot}02 & 5{\cdot}91 & 23{\cdot}31 \end{bmatrix}$$

M. ochrogaster, male ($n_3 = 57$), S_3

$$\begin{bmatrix} 14{\cdot}99 & 7{\cdot}91 & 10{\cdot}99 & 7{\cdot}39 \\ 7{\cdot}91 & 14{\cdot}38 & 5{\cdot}87 & 4{\cdot}78 \\ 10{\cdot}99 & 5{\cdot}87 & 12{\cdot}96 & 6{\cdot}51 \\ 7{\cdot}39 & 4{\cdot}78 & 6{\cdot}51 & 14{\cdot}41 \end{bmatrix}$$

M. ochrogaster, female ($n_4 = 53$), S_4

$$\begin{bmatrix} 13{\cdot}71 & 8{\cdot}82 & 11{\cdot}81 & 2{\cdot}80 \\ 8{\cdot}82 & 13{\cdot}70 & 7{\cdot}79 & 1{\cdot}47 \\ 11{\cdot}81 & 7{\cdot}79 & 19{\cdot}81 & 1{\cdot}08 \\ 2{\cdot}80 & 1{\cdot}47 & 1{\cdot}08 & 12{\cdot}44 \end{bmatrix}$$

(b) *Maximum likelihood estimate of the orthogonal matrix B under the common principal component model*; 4 *common components*

$$\hat{B} = \begin{bmatrix} 0{\cdot}53 & 0{\cdot}77 & 0{\cdot}19 & -0{\cdot}30 \\ 0{\cdot}55 & -0{\cdot}21 & -0{\cdot}81 & -0{\cdot}09 \\ 0{\cdot}56 & -0{\cdot}59 & 0{\cdot}55 & -0{\cdot}17 \\ 0{\cdot}32 & 0{\cdot}12 & 0{\cdot}08 & 0{\cdot}94 \end{bmatrix}$$

(c) *Exact maximum likelihood estimates of orthogonal matrices $B^{(i)}$ under the partial model with one common component*

$$\hat{B}^{(1)} = \begin{bmatrix} 0{\cdot}54 & 0{\cdot}76 & 0{\cdot}17 & -0{\cdot}30 \\ 0{\cdot}55 & -0{\cdot}25 & -0{\cdot}79 & -0{\cdot}08 \\ 0{\cdot}56 & -0{\cdot}57 & 0{\cdot}59 & -0{\cdot}11 \\ 0{\cdot}29 & 0{\cdot}16 & 0{\cdot}05 & 0{\cdot}94 \end{bmatrix} \qquad \hat{B}^{(2)} = \begin{bmatrix} 0{\cdot}54 & 0{\cdot}73 & 0{\cdot}28 & -0{\cdot}30 \\ 0{\cdot}55 & -0{\cdot}12 & -0{\cdot}82 & -0{\cdot}06 \\ 0{\cdot}56 & -0{\cdot}66 & 0{\cdot}48 & -0{\cdot}13 \\ 0{\cdot}29 & 0{\cdot}13 & 0{\cdot}11 & 0{\cdot}94 \end{bmatrix}$$

$$\hat{B}^{(3)} = \begin{bmatrix} 0{\cdot}54 & 0{\cdot}76 & 0{\cdot}35 & 0{\cdot}07 \\ 0{\cdot}55 & -0{\cdot}07 & -0{\cdot}59 & -0{\cdot}58 \\ 0{\cdot}56 & -0{\cdot}64 & 0{\cdot}51 & 0{\cdot}10 \\ 0{\cdot}29 & -0{\cdot}04 & -0{\cdot}52 & 0{\cdot}80 \end{bmatrix} \qquad \hat{B}^{(4)} = \begin{bmatrix} 0{\cdot}54 & 0{\cdot}84 & 0{\cdot}01 & -0{\cdot}06 \\ 0{\cdot}55 & -0{\cdot}35 & -0{\cdot}75 & 0{\cdot}00 \\ 0{\cdot}56 & -0{\cdot}40 & 0{\cdot}60 & -0{\cdot}41 \\ 0{\cdot}29 & -0{\cdot}13 & 0{\cdot}27 & 0{\cdot}91 \end{bmatrix}$$

(d) *Covariance and correlation matrices of partial common components*

$$R_1 \backslash F_1 = \begin{bmatrix} 36{\cdot}18 & -1{\cdot}36 & -1{\cdot}50 & 0{\cdot}25 \\ -0{\cdot}13 & 2{\cdot}84 & 0 & 0 \\ -0{\cdot}09 & 0 & 8{\cdot}13 & 0 \\ 0{\cdot}01 & 0 & 0 & 27{\cdot}02 \end{bmatrix} \qquad R_2 \backslash F_2 = \begin{bmatrix} 52{\cdot}43 & -0{\cdot}62 & -0{\cdot}25 & 0{\cdot}34 \\ -0{\cdot}05 & 3{\cdot}18 & 0 & 0 \\ -0{\cdot}02 & 0 & 3{\cdot}75 & 0 \\ 0{\cdot}01 & 0 & 0 & 21{\cdot}15 \end{bmatrix}$$

$$R_3 \backslash F_3 = \begin{bmatrix} 35{\cdot}17 & 2{\cdot}20 & -0{\cdot}89 & 4{\cdot}97 \\ 0{\cdot}22 & 2{\cdot}97 & 0 & 0 \\ -0{\cdot}05 & 0 & 8{\cdot}01 & 0 \\ 0{\cdot}26 & 0 & 0 & 10{\cdot}59 \end{bmatrix} \qquad R_4 \backslash F_4 = \begin{bmatrix} 34{\cdot}60 & 0{\cdot}85 & 2{\cdot}19 & -4{\cdot}30 \\ 0{\cdot}08 & 3{\cdot}41 & 0 & 0 \\ 0{\cdot}13 & 0 & 8{\cdot}53 & 0 \\ -0{\cdot}20 & 0 & 0 & 13{\cdot}14 \end{bmatrix}$$

On and above diagonal, variances and covariances; below diagonal, correlations
Eigenvalues of: S_1 are 36·32, 2·78, 8·05, 27·01; S_2 are 52·44, 3·17, 3·75, 21·15;
S_3 are 36·30, 2·80, 7·97, 9·66; S_4 are 35·62, 3·38, 8·32, 12·36.

(e) *Maximum likelihood estimates of population covariance matrices* Ψ_i

$$\hat{\Psi}_1=\begin{bmatrix} 15\cdot04 & 9\cdot94 & 11\cdot49 & -1\cdot63 \\ 9\cdot94 & 16\cdot52 & 8\cdot13 & 3\cdot18 \\ 11\cdot49 & 8\cdot13 & 15\cdot47 & 3\cdot07 \\ -1\cdot63 & 3\cdot18 & 3\cdot07 & 27\cdot13 \end{bmatrix} \qquad \hat{\Psi}_2=\begin{bmatrix} 19\cdot43 & 14\cdot99 & 15\cdot84 & 2\cdot57 \\ 14\cdot99 & 18\cdot72 & 15\cdot23 & 6\cdot83 \\ 15\cdot84 & 15\cdot23 & 19\cdot18 & 5\cdot72 \\ 2\cdot57 & 6\cdot83 & 5\cdot72 & 23\cdot19 \end{bmatrix}$$

$$\hat{\Psi}_3=\begin{bmatrix} 13\cdot13 & 8\cdot35 & 10\cdot78 & 4\cdot53 \\ 8\cdot35 & 17\cdot19 & 8\cdot06 & 3\cdot09 \\ 10\cdot78 & 8\cdot06 & 14\cdot52 & 4\cdot46 \\ 4\cdot53 & 3\cdot09 & 4\cdot46 & 11\cdot90 \end{bmatrix} \qquad \hat{\Psi}_4=\begin{bmatrix} 12\cdot65 & 9\cdot35 & 9\cdot78 & 4\cdot38 \\ 9\cdot35 & 15\cdot88 & 7\cdot38 & 3\cdot93 \\ 9\cdot78 & 7\cdot38 & 16\cdot73 & 2\cdot21 \\ 4\cdot38 & 3\cdot93 & 2\cdot21 & 14\cdot40 \end{bmatrix}$$

diagonal elements of the F_i-matrices: the species *M. californicus*, especially the males, has a relatively large second-largest eigenvalue, compared with the species *M. ochrogaster.*

Finally, note that the maxium likelihood estimates $\hat{\Psi}_i$, given in Table 1(e), agree closely with the S_i, thus again confirming the partial common component model. The $\hat{\Psi}_i$ have, by their construction, one exactly identical eigenvector, namely $\hat{\beta}_1$.

Because of pressure on space, a second example can regrettably be reported only in outline. Full details are available from the author. The data, kindly provided by W. J. Krzanowski, concerned examination scores obtained in eight subjects by two groups of students, the sample sizes being 19 and 15. For an earlier analysis, see Krzanowski (1979). A common space model is considered with four components, but there is some suggestion of model inadequacy, although with the small sample sizes involved here the analysis has to be regarded as exploratory.

5. Conclusions

A word of caution is in order concerning applications of the methods proposed here. Although the likelihood ratio tests for $H_C(q)$ and $H_{CS}(q)$ are formally correct, it is in most practical cases not possible to fix q in advance, nor is it possible to decide a priori which components will be those to be considered as common. The flavour of these methods is much more towards model building than towards hypothesis testing, and the asymptotic chi-squared statistics should therefore be used with caution. Interpreting the chi squareds of different models, as in the example in § 4, as a descriptive measure of goodness of fit, however, can clearly be useful. Another limitation of the methods is that they are based on covariance matrices, whereas principal component analysis is often done on correlation matrices.

As the examples show, there may also be a need for more complicated models. For instance in the example in § 4, it might be worth estimating common principal components under the hypothesis $\Psi_1=\Psi_2$, $\Psi_3=\Psi_4$. In a three group problem one might wish to estimate the parameters of a model with q components common to all groups and r components common to groups 1 and 2 only. Or we could estimate three invariant subspaces of dimensions q, r and $p-q-r$ instead of just two, an obvious generalization of the common space model. Or one might mix the two models: q common components and two invariant subspaces of dimensions r and $p-q-r$, respectively. Approximate maximum likelihood estimates for such models can always easily be obtained once the parameters of the ordinary common principal component model have been estimated.

Acknowledgements

This research was supported by the Swiss National Science Foundation. The author wishes to thank a referee for very careful and helpful comments on an earlier draft.

APPENDIX 1

Computational methods for exact maximum likelihood estimates

Let B denote the orthogonal matrix obtained for the common principal component model. For a small angle $\alpha > 0$ and for all pairs (l, j) $(1 \leqslant l \leqslant q < j \leqslant p)$ do the following.

(i) Rotate the lth and the jth column of B by an angle α, or $-\alpha$, and call the new matrix $B^* = (B_1^*, B_2^*)$.
(ii) Adjust the q first columns of B^* by applying the FG-algorithm (Flury & Gautschi, 1986) to the matrices $B_1^{*\prime} S_i B_1^*$; that is replace B_1^* by $B_1^* Q$, where $Q \in O(q)$ is the orthogonal matrix returned by the algorithm.
(iii) Diagonalize the matrices $B_2^{*\prime} S_i B_2^*$ to get a $p \times (p-q)$ matrix $B_2^{(i)*}$ in each group.

If the value of the likelihood function increases by this procedure, replace B by B^*. Otherwise apply the same procedure to the next pair (l, j).

The angle α was decreased in five steps from 8 degrees to 0·08 degrees. At the end of the computations the matrix B was such that any rotation of two columns by an angle of 0·08 degrees would decrease the likelihood. In addition, the derivatives (2·5) were computed as an error check.

For the common space model the computational method was the same, except for step (ii), which was replaced by groupwise diagonalizations of the matrices $B_1^{*\prime} S_i B_1^*$, analogously to step (iii). More details concerning the computational method may be obtained from the author upon request.

APPENDIX 2

Proofs of Lemmas 1 *and* 2

Proof of Lemma 1. First notice that, equivalently to minimizing Φ as defined in (2·7), we can minimize the simpler function

$$\Phi^*(B; S_1, \ldots, S_k) = \prod_{i=1}^{k} |\operatorname{diag}(B' S_i B)|^{n_i},$$

since $|B' S_i B| = |S_i|$ for any orthogonal matrix B.

Since $a_1, \ldots, a_q$ are eigenvectors of all matrices S_i $(i = 1, \ldots, k)$, there exists a matrix $A \in O(p)$, $A = (a_1, \ldots, a_q, a_{q+1}, \ldots, a_p)$, such that

$$T_i = A' S_i A = \begin{bmatrix} D_i & 0 \\ 0' & E_i \end{bmatrix} \quad (i = 1, \ldots, k),$$

where 0 is the $q \times (p-q)$ matrix with zeros in all positions, D_i is a diagonal matrix of dimension $q \times q$, and E_i has dimension $(p-q) \times (p-q)$. We are first going to show that the matrix $M_0 \in O(p)$ that minimizes

$$\Phi^*(M; T_1, \ldots, T_k) = \prod_{i=1}^{k} |\operatorname{diag}(M' T_i M)|^{n_i}$$

has the form

$$M_0 = \begin{bmatrix} I_q & 0 \\ 0' & M_2 \end{bmatrix},$$

where $M_2 \in O(p-q)$. To show this, note that the matrix $M_0 = (m_1, \ldots, m_p)$ must satisfy the equation system (Flury & Gautschi, 1986) which characterizes the stationary points of the function Φ:

$$m_l'\left(\sum_{i=1}^{k} n_i \frac{\theta_{il}-\theta_{ij}}{\theta_{il}\theta_{ij}} T_i\right) m_j = 0 \quad (l \neq j), \tag{A·1}$$

where $\theta_{ih} = m_h' T_i m_h$. Since the matrix in brackets of (A·1) is block-diagonal, a stationary point of the function Φ is attained if we choose

$$M = \begin{bmatrix} M_1 & 0 \\ 0' & M_2 \end{bmatrix}, \tag{A·2}$$

where $M_1 \in O(q)$ and $M_2 \in O(p-q)$ are such that M_1 minimizes $\Phi_q^*(H) = \Pi\, |\mathrm{diag}\,(H'D_iH)|^{n_i}$ and M_2 minimizes $\Phi_{p-q}^*(H) = \Pi\, |\mathrm{diag}\,(H'E_iH)|^{n_i}$. To see that this is a minimum rather than just a stationary point, take any pair of vectors $(v_1' : 0')'$ and $(0' : v_2')'$ where v_1 is a column of M_1 and v_2 is a column of M_2. If we rotate these two vectors by an angle α and put $c = \cos\alpha$, $s = \sin\alpha$, we get two new orthogonal vectors $(cv_1' : sv_2')'$ and $(-sv_1' : cv_2')'$, with

$$\begin{bmatrix} cv_1 & -sv_1 \\ sv_2 & cv_2 \end{bmatrix}' T_i \begin{bmatrix} cv_1 & -sv_1 \\ sv_2 & cv_2 \end{bmatrix} = \begin{bmatrix} c^2 v_1' D_i v_1 + s^2 v_2' E_i v_2 & \\ & s^2 v_1' D_i v_1 + c^2 v_2' E_i v_2 \end{bmatrix}. \tag{A·3}$$

The product of the diagonal elements of (A·3), which occurs as a factor in the function Φ^*, is a minimum if $c^2 = 1$ or $c^2 = 0$. Therefore any rotation of two vectors of M away from the form (A·2) will increase the value of Φ^*, unless the two diagonal elements of (A·3) are exactly identical in each group. In this case any rotation of the two vectors $(v_1' : 0')'$ and $(0' : v_2')'$ will leave Φ^* unchanged, and for the sake of uniqueness we can require M to be of block diagonal form. To complete the proof, we note that all D_i are exactly diagonal. This implies that m_1 to m_q can be chosen to be the first q unit vectors. Therefore $B_0 = AM_0$ has the vectors $a_1, \ldots, a_q$ as columns.

Proof of Lemma 2. Let $a_1, \ldots, a_q$ denote orthonormal vectors spanning $\mathscr{A}_q$, and $a_{q+1}, \ldots, a_p$ orthonormal vectors spanning the $(p-q)$-dimensional subspace orthogonal to $\mathscr{A}_q$. Put $A = (a_1, \ldots, a_p)$ and proceed as in the proof of Lemma 1, where the D_i are now not necessarily diagonal.

References

AIROLDI, J. P. & HOFFMANN, R. S. (1984). Age variation in voles (Microtus californicus, M. ochrogaster) and its significance for systematic studies. *Occasional Papers of the Museum of Natural History*, Lawrence, Kansas: The University of Kansas, no. 111, 1–45.

ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Statist.* **34**, 122–48.

FLURY, B. N. (1984). Common principal components in k groups, *J. Am. Statist. Assoc.* **79**, 892–8.

FLURY, B. N. (1986a). Asymptotic theory for common principal component analysis. *Ann. Statist.* **14**, 418–30.

FLURY, B. N. (1986b). Proportionality of k covariance matrices. *Statist. Prob. Letters* **4**, 29–33.

FLURY, B. N. & CONSTANTINE, G. (1985). Algorithm AS 211: The FG 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. Comp.* **7**, 169–84.

JOLICOEUR, P. & MOSIMANN, J. E. (1960). Size and shape variation in the painted turtle: A principal component analysis. *Growth* **24**, 339–54.

KRZANOWSKI, W. J. (1979). Between-groups comparison of principal components. *J. Am. Statist. Assoc.* **74**, 703–7. Correction (1981), **76**, 1022.

KRZANOWSKI, W. J. (1982). Between-group comparison of principal components—Some sampling results. *J. Statist. Comp. Simul.* **15**, 141–54.

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

[*Received September* 1985. *Revised July* 1986]