Common Principal Components in k Groups

BERNHARD N. FLURY*

This article generalizes the method of principal components to so-called "common principal components" as follows: Consider the hypothesis that the covariance matrices Σ_i for k populations are simultaneously diagonalizable. That is, there is an orthogonal matrix β such that $\beta' \Sigma_i \beta$ is diagonal for $i = 1, \ldots, k$. I derive the normal-theory maximum likelihood estimates of the common component Σ_i matrices and the log-likelihood-ratio statistics for testing this hypothesis. The solution has some favorable properties that do not depend on normality assumptions. Numerical examples illustrate the method. Applications to data reduction, multiple regression, and nonlinear discriminant analysis are sketched.

KEY WORDS: Maximum likelihood; Covariance matrices; Discriminant analysis.

1. INTRODUCTION

Principal Component Analysis (PCA) is a well-known and established technique of multivariate statistical analysis. It enjoys a solid theoretical foundation and possesses many optimal properties (see Okamoto 1969 and McCabe 1984). One important property of principal components (PC's) is that they can be considered as uncorrelated variables, obtained by an orthogonal rotation of the coordinate system. PCA is a one-sample method, and to my knowledge only two efforts have been made to generalize it to a two-sample method. Krzanowski (1979) compared the PC's of two different samples by computing the angles between the subspaces spanned by the first q PC's for each group. Flury (1983) used the characteristic vectors of $\Sigma_1^{-1}\Sigma_2$ to obtain uncorrelated variables in two groups simultaneously.

In practice we often deal with the situation of the same variables being measured on objects from different groups, and the covariance structure may vary from group to group. Examples for this range from three species of iris (Anderson 1935), male and female turtles (Jolicoeur and Mosimann 1960), and human and animal bones (Jolicoeur 1963) to real and forged bank notes (Flury and Riedwyl 1983). In all of these cases, tests of significance (not given in this article) suggest that the underlying population covariance matrices are not exactly identical in all groups. Yet sometimes the covariance matrices of different groups look somehow similar, and it seems reasonable to assume that the covariance matrices have a common basic structure. One such basic structure—proportionality of the covariance matrices—has been studied by Pillai, Al-Ani, and Jouris (1969) and by Rao (1982). Another basic structure, which is more general, is that some rotation diagonalizes the covariance matrices simultaneously in all populations. In Sections 4 and 5, this assumption is justified in certain practical applications. The mathematical formulation for k $(p \times p)$-covariance matrices $\Sigma_1, \ldots, \Sigma_k$ is expressed by the hypothesis

$$H_C\colon \beta'\Sigma_i\beta = \Lambda_i(\text{diagonal}), \qquad i = 1, \ldots, k, \quad (1.1)$$

where β is an orthogonal $p \times p$ matrix.

Since the columns of β can be viewed as direction cosines for rotated axes, let us refer to H_C as the *hypothesis of common PC's*, and the simultaneously transformed variables $\mathbf{U}_i = \beta'\mathbf{X}_i$ will be called *common principal components* (CPC's). Note that, in contrast to the one-sample case, no canonical ordering of the columns of β need be given, since the rank order of the diagonal elements of the Λ_i is not necessarily the same for all groups.

2. MAXIMUM LIKELIHOOD ESTIMATION OF COMMON PRINCIPAL AXES IN k NORMAL POPULATIONS

Let the p-variate random vectors $\mathbf{X}_i$ $(i = 1, \ldots, k)$ be independently distributed as $N_p(\mathbf{u}_i, \Sigma_i)$, where $\mu_i \in \mathcal{R}^p$ and the Σ_i are positive definite and symmetric (pds). For samples of size $N_i = n_i + 1$, denote, by $\mathbf{S}_i$ $(i = 1, \ldots, k)$, the usual unbiased sample covariance matrices. Assume $\min_{1 \le i \le k} n_i \ge p$. Then the matrices $n_i\mathbf{S}_i$ are independently distributed as $W_p(n_i, \Sigma_i)$ (Muirhead 1982, p. 85), and the common likelihood function of $\Sigma_1, \ldots, \Sigma_k$, given $\mathbf{S}_1, \ldots, \mathbf{S}_k$, is

$$L(\Sigma_1, \ldots, \Sigma_k) = C \times \prod_{i=1}^{k} \operatorname{etr}\left(-\frac{n_i}{2}\Sigma_i^{-1}\mathbf{S}_i\right) |\Sigma_i|^{-n_i/2}, \quad (2.1)$$

where C is a constant that does not depend on the Σ_i and etr denotes the exponential function of the trace. Instead of maximizing the likelihood function, minimize

$$g(\Sigma_1, \ldots, \Sigma_k) = -2 \log L(\Sigma_1, \ldots, \Sigma_k) + 2 \log C = \sum_{i=1}^{k} n_i (\log |\Sigma_i| + \operatorname{tr}\Sigma_i^{-1}\mathbf{S}_i). \quad (2.2)$$

Let us now assume that H_C holds for some orthogonal

* Bernhard N. Flury is Visiting Assistant Professor, Department of Mathematics, Indiana University, Bloomington, IN 47405. This work was done under Contract 82.008.0.82 of the Swiss National Science Foundation in the Department of Statistics, Purdue University. The author wishes to thank the editor, an associate editor, and two referees for their helpful comments.

© Journal of the American Statistical Association
December 1984, Volume 79, Number 388
Theory and Methods Section

matrix β. Let $\Lambda_i = \text{diag}(\lambda_{i1}, \ldots, \lambda_{ip})$, then

$$\log |\Sigma_i| = \sum_{j=1}^{p} \log \lambda_{ij}, \qquad i = 1, \ldots, k, \quad (2.3)$$

and

$$\text{tr}\, \Sigma_i^{-1} S_i = \text{tr}(\beta \Lambda_i^{-1} \beta' S_i) = \text{tr}(\Lambda_i^{-1} \beta' S_i \beta)$$
$$= \sum_{j=1}^{p} \beta_j' S_i \beta_j / \lambda_{ij}, \qquad i = 1, \ldots, k, \quad (2.4)$$

where β_j is the jth column of β. Therefore,

$$g(\Sigma_1, \ldots, \Sigma_k)$$
$$= g(\beta_1, \ldots, \beta_p, \lambda_{11}, \ldots, \lambda_{1p}, \lambda_{21}, \ldots, \lambda_{kp})$$
$$= \sum_{i=1}^{k} n_i \left[\sum_{j=1}^{p} (\log \lambda_{ij} + \beta_j' S_i \beta_j / \lambda_{ij}) \right]. \quad (2.5)$$

The function g is to be minimized under the restrictions

$$\beta_h' \beta_j = 0 \quad \text{if } h \neq j$$
$$= 1 \quad \text{if } h = j. \quad (2.6)$$

Thus we wish to minimize the function

$$G(\Sigma_1, \ldots, \Sigma_k)$$
$$= g(\Sigma_1, \ldots, \Sigma_k) - \sum_{h=1}^{p} \gamma_h (\beta_h' \beta_h - 1)$$
$$- 2 \sum_{h<j}^{p} \gamma_{hj} \beta_h' \beta_j, \quad (2.7)$$

where the $\gamma_h (1 \leq h \leq p)$ and $\gamma_{hj} (1 \leq h < j \leq p)$ are $p(p+1)/2$ Lagrange multipliers. Taking partial derivatives with respect to λ_{ij} and setting them equal to zero yields

$$\lambda_{ij} = \beta_j' S_i \beta_j, \quad i = 1, \ldots, k, \quad j = 1, \ldots, p, \quad (2.8)$$

and it follows from (2.4) that

$$\text{tr}\, \Sigma_i^{-1} S_i = p, \qquad i = 1, \ldots, k. \quad (2.9)$$

The vector of partial derivatives of G with respect to β_j, set equal to zero, yields

$$\sum_{i=1}^{k} n_i S_i \beta_j / \lambda_{ij} - \sum_{\substack{h=1 \\ h \neq j}}^{p} \gamma_{jh} \beta_h - \gamma_j \beta_j = 0,$$
$$j = 1, \ldots, p, \quad (2.10)$$

where we put $\gamma_{jh} = \gamma_{hj}$ if $j < h$. Multiplying (2.10) from the left by β_j' gives

$$\gamma_j = \sum_{i=1}^{k} n_i, \qquad j = 1, \ldots, p, \quad (2.11)$$

and thus

$$\sum_{i=1}^{k} n_i S_i \beta_j / \lambda_{ij} - \left(\sum_{i=1}^{k} n_i \right) \beta_j$$
$$- \sum_{\substack{h=1 \\ h \neq j}}^{p} \gamma_{jh} \beta_h = 0, \qquad j = 1, \ldots, p. \quad (2.12)$$

Multiplying (2.12) from the left by β_l' $(l \neq j)$ implies

$$\sum_{i=1}^{k} n_i \beta_l' S_i \beta_j / \lambda_{ij} = \gamma_{jl}, \quad j = 1, \ldots, p, \; l \neq j. \quad (2.13)$$

Interchanging the indexes j and l in (2.13) and noting that $\beta_j' S_i \beta_l = \beta_l' S_i \beta_j$ and $\gamma_{jl} = \gamma_{lj}$, it follows that

$$\sum_{i=1}^{k} n_i \beta_l' S_i \beta_j / \lambda_{il} = \gamma_{jl}, \quad l = 1, \ldots, p, \; j \neq l, \quad (2.14)$$

and, therefore, comparing (2.13) and (2.14),

$$\beta_l' \left(\sum_{i=1}^{k} n_i \frac{\lambda_{il} - \lambda_{ij}}{\lambda_{il} \lambda_{ij}} S_i \right) \beta_j = 0,$$
$$l, j = 1, \ldots, p, \; l \neq j. \quad (2.15)$$

Since the (l, j)th equation of (2.15) is the same as the (j, l)th equation, we have actually only $p(p-1)/2$ equations, say, for $1 \leq l < j \leq p$. These have to be solved under the orthonormality conditions $\beta' \beta = I_p$ of (2.6) and the restriction (2.8). Flury and Gautschi (1984) developed an efficient procedure, called the FG algorithm, to accomplish this. Note that since the group of orthogonal $p \times p$ matrices is compact, there is always a solution of (2.15) that minimizes (2.7) and one that maximizes it. In certain extreme cases where some S_i are close to singularity and far from simultaneous diagonalizability, there may also exist several local minima (for details, see Flury and Gautschi 1984).

Let us denote the solution that maximizes the likelihood by $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)$ and $\hat{\lambda}_{ij}$ $(i = 1, \ldots, k, j = 1, \ldots, p)$ and put $\hat{\Lambda}_i = \text{diag}(\hat{\lambda}_{i1}, \ldots, \hat{\lambda}_{ip})$ and $\hat{\Sigma}_i = \hat{\beta} \hat{\Lambda}_i \hat{\beta}'$ $(i = 1, \ldots, k)$. Then, using (2.9), the maximum of (2.1) is obtained as

$$L(\hat{\Sigma}_1, \ldots, \hat{\Sigma}_k) = C \prod_{i=1}^{k} \exp(-pn_i/2) \, |\hat{\Sigma}_i|^{-n_i/2}, \quad (2.16)$$

whereas the unrestricted maximum can easily be seen to be equal to

$$L(S_1, \ldots, S_k) = C \prod_{i=1}^{k} \exp(-pn_i/2) \, |S_i|^{-n_i/2}. \quad (2.17)$$

The log-likelihood-ratio statistic for testing H_C is, therefore,

$$X^2 = -2 \log \frac{L(\hat{\Sigma}_1, \ldots, \hat{\Sigma}_k)}{L(S_1, \ldots, S_k)}$$
$$= \sum_{i=1}^{k} n_i \log \frac{|\hat{\Sigma}_i|}{|S_i|}, \quad (2.18)$$

and by the general theory of likelihood ratio tests (see Rao 1973, chap. 6) it follows that the null distribution of X_2 is asymptotically ($\min_{1 \leq i \leq k} n_i \to \infty$) chi squared with $(k-1)p(p-1)/2$ degrees of freedom.

3. PROPERTIES OF COMMON PRINCIPAL COMPONENTS

Before we apply the method derived in Section 2 to numerical examples, it may be useful to state some simple properties of the common principal components, which

will illustrate their meaning and facilitate their correct application. Note that from now on, I will refer to the *sample common principal components*

$$\mathbf{U}_i = \hat{\boldsymbol{\beta}}'\mathbf{X}_i, \qquad i = 1, \ldots, k, \tag{3.1}$$

as CPC's, suppressing the prefix *sample*.

For the ith group, let us define the matrix $\mathbf{F}_i$ as

$$\mathbf{F}_i = \hat{\boldsymbol{\beta}}'\mathbf{S}_i\hat{\boldsymbol{\beta}}, \qquad i = 1, \ldots, k. \tag{3.2}$$

The $\mathbf{F}_i$ are the sample covariance matrices of the $\mathbf{U}_i$. Since

$$\hat{\boldsymbol{\Lambda}}_i = \operatorname{diag}(\mathbf{F}_i), \qquad i = 1, \ldots, k, \tag{3.3}$$

the statistic (2.18) can be written as a function of the $\mathbf{F}_i$ alone, namely,

$$X^2 = \sum_{i=1}^{k} n_i \log \frac{|\operatorname{diag}\mathbf{F}_i|}{|\mathbf{F}_i|} = \sum_{i=1}^{k} n_i \log \frac{\prod_{j=1}^{p} f_{jj}^{(i)}}{\prod_{j=1}^{p} l_{ij}}, \tag{3.4}$$

where $f_{jj}^{(i)}$ is the (j, j) element of $\mathbf{F}_i$ and the l_{ij} $(j = 1, \ldots, p)$ are the eigenvalues of $\mathbf{F}_i$. (Incidentally, these are identical with those of $\mathbf{S}_i$). The likelihood ratio criterion is, therefore, a measure of simultaneous diagonalizability of k pds matrices $\mathbf{S}_i$. The CPC's can be viewed as obtained by a simultaneous transformation, yielding variables that are as uncorrelated as possible.

To stress the importance of the $\mathbf{F}_i$ matrices, we will compute them in the examples of Section 4. Sometimes it may be even more convienient to interpret the correlation matrices

$$\mathbf{R}_i = \hat{\boldsymbol{\Lambda}}_i^{-1/2}\,\mathbf{F}_i\,\hat{\boldsymbol{\Lambda}}_i^{-1/2}, \qquad i = 1, \ldots, k. \tag{3.5}$$

These can easily provide information about departures from the hypothesis of common principal axes, which requires that $\mathbf{R}_i$ is close to $\mathbf{I}_p$.

From the equation system (2.15) we can still gain a better understanding of CPC's. First, let us write it as

$$\boldsymbol{\beta}_l'\,(n_1\mathbf{A}_{lj}^{(1)} + \cdots + n_k\mathbf{A}_{lj}^{(k)})\,\boldsymbol{\beta}_j = 0, \qquad 1 \le l < j \le p, \tag{3.6}$$

where

$$\mathbf{A}_{lj}^{(i)} = \frac{\lambda_{il} - \lambda_{ij}}{\lambda_{il}\,\lambda_{ij}}\,\mathbf{S}_i, \qquad i = 1, \ldots, k. \tag{3.7}$$

Equation (3.6) remains unchanged if we replace one of the $\mathbf{S}_i$ by a proportional matrix $c\mathbf{S}_i$ $(c > 0)$. CPC's are invariant under proportionality, where different constants of proportionality are admitted for the k groups.

Furthermore, we note that the weight of matrix $\mathbf{S}_i$ in (2.15) is $n_i(\lambda_{il} - \lambda_{ij})/\lambda_{il}\,\lambda_{ij}$. Now, by (2.8), λ_{ij} is the variance of the linear combination $\boldsymbol{\beta}_j'\,\mathbf{X}_i$, and it becomes clear that the weight of matrix $\mathbf{S}_i$ in the (l, j)th equation is smaller when the two variances λ_{ij} and λ_{il} are closer. If $\lambda_{ij} = \lambda_{il}$, then $\mathbf{S}_i$ disappears from the (l, j)th equation. Of course this makes perfect sense, because it corresponds to sphericity of $\mathbf{X}_i$ in the plane spanned by $\boldsymbol{\beta}_j$ and $\boldsymbol{\beta}_l$. In the extreme case $\mathbf{S}_i = c\mathbf{I}_p$, the ith matrix will disappear completely from (2.15). Therefore, we can say that the overall influence of matrix $\mathbf{S}_i$ is proportional not only to its degrees of freedom n_i but also to its deviation from sphericity.

Finally, it is worth noting that in the case $k = 1$, CPCA reduces to the well known PCA, and the FG algorithm computes the eigenvectors of the single pds matrix $\mathbf{S} = \mathbf{S}_1$ in this case.

4. APPLICATIONS

The practical use of CPCA can be demonstrated best if we apply it to some well-known examples in the multivariate literature.

Example 1: Fisher's iris data. These famous and most-abused data were first published by Anderson (1935) and were used by Fisher (1936) as an example of discriminant analysis. The four variables were (a) sepal length, (b) sepal width, (c) petal length, and (d) petal width, measured on three species of iris: versicolor, virginica, and setosa. The sample covariance matrices, each based on 49 degrees of freedom, are shown in part a of Table 1. Part b of the same table displays the ML estimates $\hat{\boldsymbol{\Sigma}}_i$ under the restriction of common principal axes. Part c shows the estimates of the common principal axes—each column of $\hat{\boldsymbol{\beta}}$ representing the coefficients for one component. The CPC's do not seem to have an obvious interpretation. Part d of Table 1 shows the variances $\hat{\lambda}_{ij}$ in all three groups, ordered according to the columns of $\hat{\boldsymbol{\beta}}$. (Of course, this order is irrelevant and depends only on the initial approximation used in the FG algorithm.) To compare PC's and CPC's, we also give the eigenvalues of the $\mathbf{S}_i$ (or $\mathbf{F}_i$), again ordered in the same sense. If the hypothesis of common principal axes in all three populations is true, we would expect all of the $\hat{\lambda}_{ij}$ to be close to the eigenvalues of $\mathbf{S}_i$. This is the case for sample 1 (versicolor), whereas the differences are larger in sample 2 (virginica) and even worse in sample 3 (setosa). This impression is confirmed by the value $X^2 = 63.91$ of the statistic (2.18), which can be compared with quantiles of the chi squared distribution with 12 degrees of freedom. At any reasonable level of significance, we would conclude that the assumption of common principal axes does not hold, though the chi squared approximation might still be rather poor for sample sizes of 50, and nonnormality might affect the exact significance level.

To learn more about the deviation of the data from the model of common principal axes, we can look at the $\mathbf{F}_i$ and $\mathbf{R}_i$ matrices, given in part e of Table 1. Note that the diagonals of the $\mathbf{F}_i$ matrices contain the variances of the CPC's. There is obviously a high correlation between the first and third CPC's in group 3 (setosa), which might explain the inadequacy of the assumption of common principal axes.

The results of a second CPCA, based on groups 1 (versicolor) and 2 (virginica) only, are displayed in Table 2. The value of the statistic (2.18) is $X^2 = 13.46$, which lies

Table 1. Estimation of CPC's in Fisher's 1936 Iris Data

(a) *Sample Covariance Matrices*

Versicolor ($N_1 = 50$)

$$\mathbf{S}_1 = \begin{pmatrix} 26.6433 & 8.5184 & 18.2898 & 5.5780 \\ 8.5184 & 9.8469 & 8.2653 & 4.1204 \\ 18.2898 & 8.2653 & 22.0816 & 7.3102 \\ 5.5780 & 4.1204 & 7.3102 & 3.9106 \end{pmatrix}$$

Virginica ($N_2 = 50$)

$$\mathbf{S}_2 = \begin{pmatrix} 40.4343 & 9.3763 & 30.3290 & 4.9094 \\ 9.3763 & 10.4004 & 7.1380 & 4.7629 \\ 30.3290 & 7.1380 & 30.4588 & 4.8824 \\ 4.9094 & 4.7629 & 4.8824 & 7.5433 \end{pmatrix}$$

Setosa ($N_3 = 50$)

$$\mathbf{S}_3 = \begin{pmatrix} 12.4249 & 9.9216 & 1.6355 & 1.0331 \\ 9.9216 & 14.3690 & 1.1698 & .9298 \\ 1.6355 & 1.1698 & 3.0159 & .6069 \\ 1.0331 & .9298 & .6069 & 1.1106 \end{pmatrix}$$

(b) *MLE's of Population Covariance Matrices*

$$\hat{\mathbf{\Sigma}}_1 = \begin{pmatrix} 29.5860 & 7.3004 & 18.6600 & 4.6667 \\ 7.3004 & 7.4546 & 6.6121 & 2.8309 \\ 18.6600 & 6.6121 & 21.2145 & 6.2692 \\ 4.6667 & 2.8309 & 6.2692 & 3.2273 \end{pmatrix}$$

$$\hat{\mathbf{\Sigma}}_2 = \begin{pmatrix} 40.6417 & 11.5005 & 27.8263 & 7.9275 \\ 11.5005 & 11.0588 & 8.8976 & 2.8603 \\ 27.8263 & 8.8976 & 29.6478 & 7.0677 \\ 7.9275 & 2.8603 & 7.0677 & 7.4885 \end{pmatrix}$$

$$\hat{\mathbf{\Sigma}}_3 = \begin{pmatrix} 9.4477 & 3.5268 & 4.5255 & 1.2613 \\ 3.5268 & 10.2264 & -2.5687 & .2601 \\ 4.5255 & -2.5687 & 9.5669 & 2.1149 \\ 1.2613 & .2601 & 2.1149 & 1.6793 \end{pmatrix}$$

(c) *Coefficients of Common Principal Components*

$$\hat{\mathbf{\beta}} = \begin{pmatrix} .7367 & -.6471 & -.1640 & .1084 \\ .2468 & .4655 & -.8346 & -.1607 \\ .6047 & .5003 & .5221 & -.3338 \\ .1753 & .3382 & .0628 & .9225 \end{pmatrix}$$

(d) *Variances $\hat{\lambda}_{ij}$ of CPC's and Eigenvalues of $\mathbf{S}_i$*

Versicolor				
$\hat{\lambda}_{1j}$ =	48.4602	7.4689	5.5394	1.0139
Eigenvalues =	48.7874	7.2384	5.4776	.9790
Virginica				
$\hat{\lambda}_{2j}$ =	69.2235	6.7124	7.5367	5.3642
Eigenvalues =	69.5255	5.2295	10.6552	3.4266
Setosa				
$\hat{\lambda}_{3j}$ =	14.6444	2.7526	12.5065	1.0169
Eigenvalues =	23.6456	2.6796	3.6919	.9033

(e) *Covariance and Correlation Matrices of CPC's*

$$\mathbf{F}_1 = \begin{pmatrix} 48.4602 & 3.4072 & -1.1931 & .7172 \\ 3.4072 & 7.4689 & -.3776 & .2049 \\ -1.1931 & -.3776 & 5.5394 & -.3278 \\ .7172 & .2049 & -.3278 & 1.0139 \end{pmatrix}$$

$$\mathbf{R}_1 = \begin{pmatrix} 1.000 & .1791 & -.0728 & .1023 \\ .1791 & 1.0000 & -.0587 & .0745 \\ -.0728 & -.0587 & 1.0000 & -.1383 \\ .1023 & .0745 & -.1383 & 1.0000 \end{pmatrix}$$

$$\mathbf{F}_2 = \begin{pmatrix} 69.2235 & -1.6211 & 2.6003 & -2.9062 \\ -1.6211 & 6.7124 & -1.9278 & 2.3514 \\ 2.6003 & -1.9278 & 7.5367 & -2.2054 \\ -2.9062 & 2.3514 & -2.2054 & 5.3642 \end{pmatrix}$$

$$\mathbf{R}_2 = \begin{pmatrix} 1.0000 & -.0752 & .1138 & -.1508 \\ -.0752 & 1.0000 & -.2710 & .3919 \\ .1138 & -.2710 & 1.0000 & -.3468 \\ -.1508 & .3919 & -.3468 & 1.0000 \end{pmatrix}$$

$$\mathbf{F}_3 = \begin{pmatrix} 14.6444 & -.5682 & -9.9950 & -.2106 \\ -.5682 & 2.7526 & .0487 & -.4236 \\ -9.9950 & .0487 & 12.5065 & .4235 \\ -.2106 & -.4236 & .4235 & 1.0169 \end{pmatrix}$$

$$\mathbf{F}_3 = \begin{pmatrix} 1.000 & -.0895 & -.7385 & -.0546 \\ -.0895 & 1.0000 & .0083 & -.2532 \\ -.7385 & .0083 & 1.0000 & .1188 \\ -.0546 & -.2532 & .1188 & 1.0000 \end{pmatrix}$$

NOTE: The sample covariance matrices reported here were multiplied with 10^2.

between the 95th and the 99th percentile of the chi squared distribution with six degrees of freedom. Although it seems doubtful whether the hypothesis of common principal axes should be accepted, we can note that the variances of the CPC's (diagonal of the $\mathbf{F}_i$ matrices) are now much closer to the eigenvalues of the $\mathbf{S}_i$. Since the two $\mathbf{F}_i$ matrices are based on the same linear transformation, we can also look at the ratios of variances of the four CPC's, and we note that the largest ratio is 6.2186/5.1354 = 1.21 (for the third CPC), whereas the smallest ratio is obtained from the fourth CPC as 1.0119/4.5813 = .221. These two ratios are close to the extreme characteristic roots of $\mathbf{S}_2^{-1}\mathbf{S}_1$, as can be expected from corollary 1 of Flury (1983).

Example 2: Bone dimensions of the North American Marten. Jolicoeur (1963) measured the variables (a) log (length of the humerus), (b) log (width of the humerus), (c) log (length of the femur), and (d) log (width of the femur) on $N_1 = 92$ male and $N_2 = 47$ female individuals of the species *Martes americana*. Principal component analyses performed separately on each group showed a similar pattern of the two orthogonal matrices. The CPCA of these data is summarized in Table 3. The value of the log-likelihood ratio statistic is $X^2 = 8.34$, with six degrees of freedom. The hypothesis of common principal axes, therefore, seems quite plausible. Comparing the coefficients of the CPC's to those given by Jolicoeur for each sex separately, we can see that his interpretation applies as well to our analysis, but it becomes simpler, since transformations are the same in both groups. Table 3 (part d) shows that the diagonal elements and eigenvalues of the $\mathbf{F}_i$ are very close. The correlations between the CPC's, displayed in Table 3 (part e), confirm the impression that CPC's are justified in this example. Note that correlations in group 2 (females) tend to be larger in absolute value than in group 1—possibly an effect of the unequality of sample sizes.

Is the CPC model appropriate for this data? Does it make sense from a biological point of view? The author

Table 2. CPCA of Iris Versicolor and Iris Virginica

(a) *MLE's of Population Covariance Matrices*

$$\hat{\Sigma}_1 = \begin{pmatrix} 28.1693 & 7.8487 & 19.1036 & 4.9047 \\ 7.8487 & 8.8832 & 6.1772 & 3.4523 \\ 19.1036 & 6.1772 & 21.9980 & 6.1257 \\ 4.9047 & 3.4523 & 6.1257 & 3.4319 \end{pmatrix}$$

$$\hat{\Sigma}_2 = \begin{pmatrix} 38.7745 & 10.4893 & 28.7304 & 7.8987 \\ 10.4893 & 13.1731 & 9.5169 & 4.3989 \\ 28.7304 & 9.5169 & 29.6661 & 7.1734 \\ 7.8987 & 4.3989 & 7.1734 & 7.2231 \end{pmatrix}$$

(b) *Coefficients of CPC's*

$$\hat{\beta} = \begin{pmatrix} .7206 & -.2914 & -.6159 & .1286 \\ .2545 & .9019 & -.1900 & -.2927 \\ .6188 & -.1186 & .7188 & -.2939 \\ .1817 & .2960 & .2607 & .9008 \end{pmatrix}$$

(c) *Covariance Matrices of CPC's*

$$\mathbf{F}_1 = \begin{pmatrix} 48.5836 & 2.7247 & .9370 & .4711 \\ 2.7247 & 6.6683 & .9278 & -.3659 \\ .9370 & .9287 & 6.2186 & -.0118 \\ .4711 & -.3659 & -.0118 & 1.0119 \end{pmatrix}$$

$$\mathbf{F}_2 = \begin{pmatrix} 69.1434 & -4.1101 & -.7288 & -2.2362 \\ -4.1101 & 9.9766 & -.1066 & 2.5981 \\ -.7288 & -.1066 & 5.1354 & .4135 \\ -2.2362 & 2.5981 & .4135 & 4.5813 \end{pmatrix}$$

has analyzed several biometrical examples, and, indeed, in most cases the hypothesis of CPC's seemed reasonable from a statistical point of view. However, if we think of the frequent interpretation of PC's as independent factors of growth or independent sources of variation (see the examples in Morrison 1976), then it is quite plausible to assume that the same growth factors should be found in relates species, with possibly differing variability. Besides its mathematical appeal, the CPC model can, therefore, be given a biological meaning.

Example 3: Real and forged bank notes. From Flury and Riedwyl (1983) we take the following variables, measured on real and forged Swiss bank notes:

1. width of the bank note, measured on the left side
2. width of the bank note, measured on the right side
3. width of the lower margin
4. width of the upper margin.

All measurements were in millimeters. The two samples consisted of $N_1 = 100$ real and $N_2 = 85$ forged notes, which had probably all been produced by the same forger. The results of a CPCA on these data are displayed in Table 4. The CPC's can roughly be interpreted as (a) slant of the cut, (b) width of the print, (c) vertical position of the print, and (d) size. These interpretations can be brought into relation with two independent phases of the production process (printing and cutting), which makes the assumption of common principal axes seem reasonable. The value of (2.18) is $X^2 = 12.04$, which is close to the 95th percentile of the chi squared distribution with six degrees of freedom. Note that the diagonal elements and eigenvalues of the $\mathbf{F}_i$ matrices agree very closely, so the assumption of common principal axes seems at least a good approximation to reality.

This example also shows a desirable aspect of the method: Since two eigenvalues of $\mathbf{S}_2$ are rather close (.1024 and .1305, respectively), the associated principal components might be unstable in the sense of a nearly spherical distribution in the plane spanned by the two eigenvectors. This problem, however, is taken care of by sample 1, where the corresponding variances differ much more. This illustrates a property of CPCA that was already mentioned at the end of Section 3.

Table 3. CPCA of Bone Measurements on Martes Americana

(a) *Sample Covariance Matrices*

Males ($N_1 = 92$)

$$\mathbf{S}_1 = \begin{pmatrix} 1.1544 & .9109 & 1.0330 & .7993 \\ .9109 & 2.0381 & .7056 & 1.4083 \\ 1.0330 & .7056 & 1.2100 & .7958 \\ .7993 & 1.4083 & .7958 & 2.0277 \end{pmatrix}$$

Females ($N_2 = 47$)

$$\mathbf{S}_2 = \begin{pmatrix} .9617 & .2806 & .9841 & .6775 \\ .2806 & 1.8475 & .3129 & 1.2960 \\ .9841 & .3129 & 1.2804 & .7923 \\ .6775 & 1.2960 & .7923 & 1.7819 \end{pmatrix}$$

(b) *MLE's of Population Covariance Matrices*

$$\hat{\Sigma}_1 = \begin{pmatrix} 1.0709 & .8014 & .9883 & .8461 \\ .8014 & 2.0413 & .6642 & 1.3798 \\ .9883 & .6642 & 1.1938 & .9137 \\ .8461 & 1.3798 & .9137 & 2.1241 \end{pmatrix}$$

$$\hat{\Sigma}_2 = \begin{pmatrix} 1.0566 & .4326 & 1.0137 & .6820 \\ .4326 & 1.8615 & .2895 & 1.3034 \\ 1.0137 & .2895 & 1.2826 & .6850 \\ .6820 & 1.3034 & .6850 & 1.6708 \end{pmatrix}$$

(c) *Coefficients of CPC's*

$$\hat{\beta} = \begin{pmatrix} .7288 & .3914 & .4864 & -.2811 \\ -.1408 & .5662 & -.5757 & -.5729 \\ -.6637 & .3941 & .6306 & -.0810 \\ .0920 & .6090 & -.1855 & .7656 \end{pmatrix}$$

(d) *Variances of the CPC's and Eigenvalues of* $\mathbf{S}_i$

Males				
$\hat{\lambda}_{1j}$ =	.1228	4.5419	1.0811	.6844
Eigenvalues =	.1209	4.5482	1.1163	.6447
Females				
$\hat{\lambda}_{2j}$ =	.1359	3.7641	1.5987	.3727
Eigenvalues =	.1240	3.7749	1.6047	.3679

(e) *Correlation Matrices of CPC's*

$$\mathbf{R}_1 = \begin{pmatrix} 1.0000 & .0762 & -.0792 & -.0394 \\ .0762 & 1.0000 & .0010 & -.0831 \\ -.0792 & .0010 & 1.0000 & -.1432 \\ -.0394 & -.0831 & -.1432 & 1.0000 \end{pmatrix}$$

$$\mathbf{R}_2 = \begin{pmatrix} 1.000 & -.1758 & .1314 & .1437 \\ -.1758 & 1.0000 & -.0034 & .1257 \\ .1314 & -.0034 & 1.0000 & .0823 \\ .1437 & .1257 & .0823 & 1.0000 \end{pmatrix}$$

NOTE: Data transformed logarithmically.

Table 4. CPCA of Real and Forged Bank Notes

(a) *Sample Covariance Matrices*

Real Notes ($N_1 = 100$)

$$\mathbf{S}_1 = \begin{pmatrix} .1326 & .0859 & .0567 & .0491 \\ .0859 & .1263 & .0582 & .0306 \\ .0567 & .0582 & .4132 & -.2635 \\ .0491 & .0306 & -.2635 & .4212 \end{pmatrix}$$

Forged Notes ($N_2 = 85$)

$$\mathbf{S}_2 = \begin{pmatrix} .0641 & .0489 & .0289 & -.0130 \\ .0489 & .0940 & -.0109 & .0071 \\ .0289 & -.0109 & .7242 & -.4330 \\ -.0130 & .0071 & -.4330 & .4039 \end{pmatrix}$$

(b) *MLE's of Population Covariance Matrices*

$$\hat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} .1253 & .0849 & .0640 & .0425 \\ .0849 & .1329 & .0507 & .0435 \\ .0640 & .0507 & .4674 & -.2512 \\ .0425 & .0435 & -.2512 & .3677 \end{pmatrix}$$

$$\hat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} .0677 & .0461 & .0404 & -.0189 \\ .0461 & .0836 & .0170 & -.0202 \\ .0404 & .0170 & .6641 & -.4399 \\ -.0189 & -.0202 & -.4399 & .4708 \end{pmatrix}$$

(c) *Coefficients of CPC's*

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} .7664 & .3140 & .0469 & .5585 \\ -.6297 & .5390 & .0299 & .5586 \\ -.0921 & -.5133 & .7783 & .3497 \\ -.0874 & -.5895 & -.6254 & .5037 \end{pmatrix}$$

(d) *Variances of CPC's and Eigenvalues of $\mathbf{S}_i$*

Real Notes				
$\hat{\lambda}_{1j}$ =	.0431	.0865	.6750	.2887
Eigenvalues =	.0426	.0827	.6815	.2865
Forged Notes				
$\hat{\lambda}_{2j}$ =	.0272	.1163	1.0207	.1220
Eigenvalues =	.0265	.1024	1.0269	.1305

(e) *Correlation Matrices of CPC's*

$$\mathbf{R}_1 = \begin{pmatrix} 1.0000 & .0165 & -.0524 & .0820 \\ .0165 & 1.0000 & .1971 & -.0032 \\ -.0524 & .1971 & 1.0000 & -.0711 \\ .0820 & -.0032 & -.0711 & 1.0000 \end{pmatrix}$$

$$\mathbf{R}_2 = \begin{pmatrix} 1.000 & -.0087 & .0384 & -.1292 \\ -.0087 & 1.0000 & -.2155 & .0999 \\ .0384 & -.2155 & 1.0000 & .0288 \\ -.1292 & .0999 & .0288 & 1.0000 \end{pmatrix}$$

5. CONCLUSIONS

As the preceding examples show, CPCA may have quite useful applications. Some reasons for performing CPCA can be summarized as follows:

1. *Reduction of Parameter Space.* In many examples, especially biological ones, it makes sense to assume that the hypothesis of common principal components holds (see the discussion at the end of example 2). If H_C holds, it is obviously better to reduce the parameter space (ignoring mean vectors) from $kp + kp(p-1)/2$ to $kp + p(p-1)/2$ elements, and the estimates can be expected to have smaller variance than those of the unrestricted parameter space.

2. *Data Reduction.* If the main purpose of a PCA consists of data reduction, and if in all k groups the smallest variances appear in the same CPC's, it might be useful to transform the data of all groups simultaneously to CPC's and to discard the CPC's that have small variances in all k groups simultaneously.

3. *Regression on PC's.* A frequent application of PCA is to use the PC's instead of the measured variables as regressors in multiple regression. The obvious extension of this is to use CPC's if the regression is to be calculated on grouped objects. In contrast to the one-group situation, however, the regressors will not be exactly uncorrelated in this case.

4. *Nonlinear Discriminant Analysis.* Under the assumption of common principal axes, the computation of Mahalanobis distances for classification of observations could be much simplified: Let $\mathbf{u} = (u_1, \ldots, u_p)'$ denote an observation and $\bar{\mathbf{u}}^{(i)} = (\bar{u}_1^{(i)}, \ldots, \bar{u}_p^{(i)})'$ denote the mean vector of the ith sample in the space of common principal axes. The formula for computing the Mahalanobis distance between $\mathbf{u}$ and $\bar{\mathbf{u}}_{(i)}$ becomes simply

$$D^2(\mathbf{u}, \bar{\mathbf{u}}^{(i)}) = \mathbf{d}_i' \hat{\boldsymbol{\Lambda}}_i^{-1} \mathbf{d}_i = \sum_{j=1}^{p} d_{ij}^2 / \hat{\lambda}_{ij}, \qquad i = 1, \ldots, k, \quad (5.1)$$

where $\mathbf{d}_i = (u_1 - \bar{u}_1^{(i)}, \ldots, u_p - \bar{u}_p^{(i)})'$ and $\hat{\boldsymbol{\Lambda}}_i = \operatorname{diag} \mathbf{F}_i$ as defined in Section 3. If H_C holds, this procedure can be expected to improve the rate of correct classification thanks to the greater stability of estimates mentioned before.

Further Developments

This article focuses attention on uncorrelatedness of PC's rather than on the aspect of maximizing the amount of variability accounted for by a given number q ($<p$) of PC's. The most useful applications of the CPC model, however, especially for large p, would probably be those in which some relatively small number q of rotated axes are sufficient to recover most of the variability in each of the k groups. It might, therefore, be useful to attempt to impose further restrictions on the model in the sense of (partial) orderings of variances of the PC's, or by placing thresholds upon cumulative sums of recovered variances.

At the present state of knowledge it does not seem that the methodology of this article would allow easy modifications to require such restrictions. In addition, it is not guaranteed that the resulting further reduction in the maximum restricted likelihood would again yield asymptotic chi squared statistics (boundary problems). Furthermore, the FG algorithm offers no obvious modification to this situation. However, some progress in the direction of the questions raised is expected from a study of the asymptotic distribution of CPC's, which is planned for the near future.

REFERENCES

ANDERSON, E. (1935), "The Irises of the Gaspe Peninsula," *Bulletin of the American Iris Society*, 59, 2–5.

FISHER, R.A. (1936), "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, 7, 179–188.

FLURY, B. (1983), "Some Relations Between the Comparison of Covariance Matrices and Principal Component Analysis," *Computational Statistics and Data Analysis*, 1, 97–109.

FLURY, B., and GAUTSCHI, W. (1984), "An Algorithm for Simultaneous Orthogonal Transformation of Several Positive Definite Matrices to Nearly Diagonal Form," Technical Report 84-1, Purdue University, Dept. of Statistics. (Accepted for publication in *SIAM Journal on Scientific and Statistical Computing*.)

FLURY, B., and RIEDWYL, H. (1983), *Angewandte Multivariate Statistik*, Stuttgart: Verlag Gustav Fischer.

JOLICOEUR, P. (1963), "The Degree of Generality of Robustness in Martes Americana," *Growth*, 27, 1–27.

JOLICOEUR, P., and MOSIMANN, J.E. (1960), "Size and Shape Variation in the Painted Turtle: A Principal Component Analysis," *Growth*, 24, 339–354.

KRZANOWSKI, W.J. (1979), "Between-Group Comparison of Principal Components," *Journal of the American Statistical Association*, 74, 703–707.

McCABE, G.P. (1984), "Principal Variables," *Technometrics*, 26, 137–144.

MORRISON, D.F. (1976), *Multivariate Statistical Methods*, New York: McGraw-Hill.

MUIRHEAD, R.J. (1982), *Aspects of Multivariate Statistical Theory*, New York: John Wiley.

OKAMOTO, M. (1969), "Optimality of Principal Components," in *Multivariate Analysis—II*, ed. P.R. Krishnaiah, New York: Academic Press, 673–685.

PILLAI, K.C.S., AL-ANI, S., and JOURIS, G.M. (1969), "On the Distribution of the Ratios of the Roots of a Covariance Matrix and Wilk's Criterion for Tests of Three Hypotheses," *Annals of Mathematical Statistics*, 40, 2033–2040.

RAO, C.R. (1973), *Linear Statistical Inference and Its Applications* (2nd ed.), New York: John Wiley.

——— (1982), "Likelihood Ratio Tests for Relationships Between Two Covariance Matrices," Technical Report 82-36, University of Pittsburgh, Center for Multivariate Analysis.