ANALYSIS OF LINEAR COMBINATIONS WITH EXTREME RATIOS OF VARIANCES

by

Bernhard N. Flury*
Purdue University and University of Berne

Technical Report #84-12

Department of Statistics
Purdue University

March 1984

*This work was done under grant no. 82.008.0.82 of the Swiss National Science Foundation. I wish to thank Henri Carnal, George P. McCabe, K.C.S. Pillai, Hans Riedwyl and David E. Tyler for their support.

ABSTRACT

If the covariance matrices $\underset{\sim}{\Sigma}_1$ and $\underset{\sim}{\Sigma}_2$ of two multivariate populations are not identical, insight into the differences between $\underset{\sim}{\Sigma}_1$ and $\underset{\sim}{\Sigma}_2$ can often be gained by analyzing the linear combinations with extreme ratios of variances, i.e. those defined by the eigenvectors associated with the extreme roots of $\underset{\sim}{\Sigma}_1^{-1}\underset{\sim}{\Sigma}_2$. This paper gives a descriptive method, similar to variable selection procedures in regression, for screening and simplifying these linear combinations. A hypothesis of redundancy of variables is defined, and a statistic for testing this hypothesis is derived. The method is illustrated by an example.

Keywords: covariance matrices; principal components; eigenvectors; fishing; redundancy of variables; patterned matrices; union-intersection test; largest and smallest roots; partitioned matrices; asymptotic chi-square test; multivariate normal distribution; elliptical distribution.

1. INTRODUCTION

Methods for comparing two or more covariance matrices are usually given rather little attention in applied statistical analysis. Most often, tests to compare covariance matrices are performed only to check assumptions for other multivariate methods such as MANOVA or linear discriminant analysis. The main purpose of this paper is to demonstrate that comparing two covariance matrices by analyzing certain linear combinations can be an interesting method itself, giving much more information than just the mere decision about equality or inequality.

In the one sample case, various hypotheses about a single covariance matrix such as sphericity, proportionality to a given matrix, zero correlation etc. (Morrison 1976, chapter 7; Srivastava and Carter 1983, chapter 12), and certain patterns (Anderson 1970, Szatrowski 1976) have been treated. The one-sample idea of elements of a matrix having a certain pattern can be extended to the idea that elements of different covariance matrices might somehow be related. Though there is an extensive knowledge about criteria for testing equality of covariance matrices (see e.g. Muirhead 1982, chapter 8), it seems that rather little work has been done on relationships between covariance matrices other than equality. Pillai et al (1969) and Rao (1983) treat proportionality of two covariance matrices. Flury (1983c) gives a test for equality of the principal component structure in k groups. The present paper treats still another kind of relationship between two covariance matrices $\underset{\sim}{\Sigma}_1$ and $\underset{\sim}{\Sigma}_2$: the hypothesis that certain eigenvectors of $\underset{\sim}{\Sigma}_1^{-1}\underset{\sim}{\Sigma}_2$ do not depend on some variables (i.e. have some zero coefficients).

A "natural" approach to this idea is as follows: many classical methods of multivariate analysis (e.g. linear discriminant function, multiple correlation,

canonical correlation) can be derived by applying Roy's (1957) union-intersection (UI) principle to linear combinations of the variables. In the case of two p×p covariance matrices $\underset{\sim}{\Sigma}_1$ and $\underset{\sim}{\Sigma}_2$, the UI-method is based on the fact that $\underset{\sim}{\Sigma}_1 = \underset{\sim}{\Sigma}_2$ exactly if $\min_{\underset{\sim}{a}\in\mathbb{R}^p} \underset{\sim}{a}'\underset{\sim}{\Sigma}_2\underset{\sim}{a}/\underset{\sim}{a}'\underset{\sim}{\Sigma}_1\underset{\sim}{a} = \max_{\underset{\sim}{a}\in\mathbb{R}^p} \underset{\sim}{a}'\underset{\sim}{\Sigma}_2\underset{\sim}{a}/\underset{\sim}{a}'\underset{\sim}{\Sigma}_1\underset{\sim}{a} = 1$. Let now $\underset{\sim}{S}_1$ and $\underset{\sim}{S}_2$ denote sample covariance matrices, then Roy's test (1957, chapter 6) is based on the maximization and minimization of

$$F(\underset{\sim}{a}) = \underset{\sim}{a}'\underset{\sim}{S}_2\underset{\sim}{a}/\underset{\sim}{a}'\underset{\sim}{S}_1\underset{\sim}{a} \tag{1.1}$$

over $\underset{\sim}{a}\in\mathbb{R}^p$, which leads to the problem of finding the extreme characteristic roots of $\underset{\sim}{S}_1^{-1},\underset{\sim}{S}_2$. With ℓ_1 and ℓ_p denoting the largest and smallest eigenvalue of $\underset{\sim}{S}_1^{-1}\underset{\sim}{S}_2$, Roy's test statistic is the pair (ℓ_1,ℓ_p), the so called "largest and smallest roots criterion". The hypothesis H_0: $\underset{\sim}{\Sigma}_1 = \underset{\sim}{\Sigma}_2$ is rejected if ℓ_1 is too large or ℓ_p is too small. For $n_1\underset{\sim}{S}_1$ and $n_2\underset{\sim}{S}_2$ being independently distributed as Wishart with n_i degrees of freedom and identical parameter matrix $\underset{\sim}{\Sigma}_1 = \underset{\sim}{\Sigma}_2 = \underset{\sim}{\Sigma}$, tables of Pillai (1967) or charts of Heck (1960), also published in Kres (1975) can be used. Alternative test criteria are available, but in this context it seems natural to use the criterion (ℓ_1,ℓ_p).

Let now $\underset{\sim}{X}^{(1)}$ and $\underset{\sim}{X}^{(2)}$ denote two independent p-dimensional random vectors with positive definite symmetric (p.d.s.) covariance matrices $\underset{\sim}{\Sigma}_1$ and $\underset{\sim}{\Sigma}_2$. By the spectral decomposition theorem (Basilevsky 1983, p. 235; see also section 2.1 of this paper), $\underset{\sim}{\Sigma}_1^{-1}\underset{\sim}{\Sigma}_2$ has p real positive eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$. Let the associated eigenvectors $\underset{\sim}{\beta}_1,\ldots,\underset{\sim}{\beta}_p$ be normalized such that $\underset{\sim}{\beta}_i'\underset{\sim}{\Sigma}_1\underset{\sim}{\beta}_i = 1$ (i=1,...,p). Define the variables $V_i^{(1)}$ and $V_i^{(2)}$ by $V_i^{(j)} = \underset{\sim}{\beta}_i'\underset{\sim}{X}^{(j)}$. Then the pairs of variables $(V_i^{(1)}, V_i^{(2)})$ have the ratios of variances

$$\frac{\text{var } V_i^{(2)}}{\text{var } V_i^{(1)}} = \frac{\underset{\sim}{\beta}_i' \underset{\sim}{\Sigma}_2 \underset{\sim}{\beta}_i}{\underset{\sim}{\beta}_i' \underset{\sim}{\Sigma}_1 \underset{\sim}{\beta}_i} = \lambda_i \qquad (i=1,\ldots,p). \tag{1.2}$$

Furthermore, the pairs $(V_i^{(1)}, V_i^{(2)})$ are mutually uncorrelated, and λ_i is the largest ratio of variances that can be obtained from linear combinations $\underset{\sim}{\beta}'\underset{\sim}{X}^{(1)}$ and $\underset{\sim}{\beta}'\underset{\sim}{X}^{(2)}$ uncorrelated with $(V_1^{(1)}, V_1^{(2)})$ to $(V_{i-1}^{(1)}, V_{i-1}^{(2)})$ (Flury 1983b). All information about differences in variability between the two random vectors can thus be condensed in the p ratios λ_i.

Now, for the sake of analyzing differences in variability, we are only interested in those λ_i which are far from 1, and discard those V_i which do not contribute much to the difference between $\underset{\sim}{\Sigma}_1$ and $\underset{\sim}{\Sigma}_2$, i.e. which have a variance ratio close to 1. (Similarly, in canonical correlation analysis, we are most interested in canonical variables with high correlation). In practical situations it can often be observed that the extreme roots ℓ_1 and ℓ_p of $\underset{\sim}{S}_1^{-1}\underset{\sim}{S}_2$ differ markedly from 1, while all other roots are close to 1. We will therefore concentrate on the linear combinations $V_1^{(j)} = \underset{\sim}{\beta}_1'\underset{\sim}{X}^{(j)}$ and $V_p^{(j)} = \underset{\sim}{\beta}_p'\underset{\sim}{X}^{(j)}$, the ones associated with the extreme ratios of variances. For terminological simplicity, we will call $\underset{\sim}{\beta}_1$ and $\underset{\sim}{\beta}_p$ the largest and smallest eigenvectors of $\underset{\sim}{\Sigma}_1^{-1}\underset{\sim}{\Sigma}_2$, respectively. The main aim of this paper is to give descriptive and confirmatory methods of simplifying these linear combinations with extreme ratios of variances, similar to methods used for screening regression and discriminant functions.

To motivate the methods to be presented in this paper, let us ask the following two questions: 1. Why should anyone (from a practical point of view) be interested in the eigenvectors $\underset{\sim}{\beta}_1$ and $\underset{\sim}{\beta}_p$ apart from the associated roots λ_1

and λ_p? 2. Given that someone is interested in $\underset{\sim}{\beta}_1$ (say), why should he want to simplify it?

To answer question 1, let us outline three potential applications.

Application 1: Suppose an educational experiment is conducted as follows: Two groups of students are being taught in the same p subjects. In group 1, each student is given extra lessons in the subjects in which he/she excels. In group 2, extra lessons are given to each student in the subject in which he/she shows poor achievement. This procedure can be expected to make the outcomes (test marks at the end of the experimental period) more homogeneous in group 2 than in group 1. We would therefore expect to find less variability in group 2 than in group 1 and could, hopefully, summarize this by the smallest eigenvector of $\underset{\sim}{S}_1^{-1}\underset{\sim}{S}_2$. Analyzing and interpreting the coefficients of this eigenvector could then identify variables or combinations of variables which are mainly responsible for the differences in variability.

Application 2: Forging bank notes. Suppose a forger tries to produce notes which are as similar to real notes as possible. Since he has some basic knowledge of statistics, he doesn't compare his production with just a single real note, but rather takes a sample of real notes and measures p variables on this sample and on his own notes. First of all, he will probably compare the two mean vectors to make sure that the mean of his production does not deviate from the mean of the real notes. However, since he is a fairly sophisticated forger, he is also concerned about differences in variability, and he would certainly not like his notes to have too much variability, compared with the real notes. If $\underset{\sim}{S}_1$ and $\underset{\sim}{S}_2$ denote the covariance matrices of the real and forged samples, respectively, he will therefore be interested in the largest eigenvector of $\underset{\sim}{S}_1^{-1}\underset{\sim}{S}_2$, which shows him the "worst" aspect of his production.

If the associated eigenvalue ℓ_1 is much larger than 1, the forger will try to interpret the largest eigenvector in order to identify the combination of variables which caused too much variability. On the other hand, if $\ell_p < 1$, the forger can also please himself by studying the linear combination with respect to which he did a terrific job.

Application 3: Quality control. The basic idea of the above bank note example applies of course to situations in quality control, whenever two identical machines produce certain items, and the correct production is being supervised by taking samples of items from both machines. Apart from mean differences, bad adjustment or failure of a machine could again lead to differences in variability, and by studying the largest and/or smallest eigenvector of $S_1^{-1}S_2$ we would hopefully get some hints about the cause of the trouble. Alternatively, if S_1 comes from a very large "perfect" production, we can treat S_1 as a population matrix Σ and supervise the production of a single machine by analyzing the largest eigenvector of $\Sigma^{-1}S$. This one-sample procedure will be briefly discussed in section 5.

Now to question 2. The reasons for simplifying the extreme eigenvectors are the same as the reasons for screening regression- and discriminant functions. In the presence of (possibly high) correlations between the variables, some coefficients of the extreme eigenvectors may have very large variability. In order to get a more stable linear function it may therefore be reasonable to put some coefficients equal to zero (i.e. discard the associated variables), as long as the largest (smallest) ratio of variances doesn't change much.

For the regression analog, Breiman and Freedman (1983) have recently given a very sophisticated reason for not including too many variables in a regression equation, even if the population regression parameters of the omitted variables are not zero. The transfer of their argument to the method of this paper is of course on intuitive grounds, but nevertheless the general idea of rejecting variables which don't contribute much seems a reasonable principle.

A descriptive (exploratory) method of simplifying the linear combinations with extreme ratios of variances will be given in section 3. Section 4 treats the same problem in the framework of testing hypotheses about the eigenvectors of $\underset{\sim}{\Sigma}_1^{-1}\underset{\sim}{\Sigma}_2$.

2. SOME ALGEBRA OF EIGENVECTORS, AND HYPOTHESES ABOUT RELATIONSHIPS BETWEEN TWO COVARIANCE MATRICES

2.1. Simultaneous decomposition of two p.d.s. matrices

In this and the following sections we will use the simultaneous spectral decomposition of two p.d.s. matrices $\underset{\sim}{S}$ and $\underset{\sim}{T}$ of dimension pxp. The following theorem is repeated here because of its importance.

Theorem 1: Let $\underset{\sim}{S}$ and $\underset{\sim}{T}$ denote p.d.s. matrices of dimension pxp. Then

(i) There exists a real diagonal matrix $\underset{\sim}{\Lambda}$ and a real nonsingular matrix $\underset{\sim}{B}$ such that $\underset{\sim}{B}'\underset{\sim}{S}\underset{\sim}{B} = \underset{\sim}{I}_p$ and $\underset{\sim}{B}'\underset{\sim}{T}\underset{\sim}{B} = \underset{\sim}{\Lambda}$.

(ii) The columns of $\underset{\sim}{B} = (\underset{\sim}{b}_1,\ldots,\underset{\sim}{b}_p)$ are eigenvectors of $\underset{\sim}{S}^{-1}\underset{\sim}{T}$, and the diagonal elements of $\underset{\sim}{\Lambda}$ are the associated eigenvalues, i.e. $\underset{\sim}{S}^{-1}\underset{\sim}{T}\underset{\sim}{B} = \underset{\sim}{B}\underset{\sim}{\Lambda}$.

(iii) The columns of $\underset{\sim}{A}' = (\underset{\sim}{a}_1,\ldots,\underset{\sim}{a}_p) = (\underset{\sim}{B}')^{-1}$ are eigenvectors of $\underset{\sim}{T}\underset{\sim}{S}^{-1}$, and the diagonal elements of $\underset{\sim}{\Lambda}$ are the associated eigenvalues, i.e. $\underset{\sim}{T}\underset{\sim}{S}^{-1}\underset{\sim}{A}' = \underset{\sim}{A}'\underset{\sim}{\Lambda}$.

(iv) $\underset{\sim}{S} = \sum_{i=1}^{p} \underset{\sim}{a}_i\underset{\sim}{a}_i'$ and $\underset{\sim}{T} = \sum_{i=1}^{p} \lambda_i\underset{\sim}{a}_i\underset{\sim}{a}_i'$, where $\underset{\sim}{\Lambda} = \mathrm{diag}(\lambda_1,\ldots,\lambda_p)$ (spectral decomposition).

(v) $\underset{\sim}{S}^{-1}\underset{\sim}{T} = \sum_{i=1}^{p} \lambda_i\underset{\sim}{b}_i\underset{\sim}{b}_i'\underset{\sim}{S} = \underset{\sim}{B}\underset{\sim}{\Lambda}\underset{\sim}{A}$.

For a proof of (i), (ii) and (iv), see e.g. Basilevsky (1983, Theorem 5.19). (iii) and (v) follow then easily. Theorem 1 holds also if $\underset{\sim}{T}$ is symmetric but not necessarily positive definite. However, we will need it only for the positive definite case, which ensures that all λ_i are strictly positive. Note that $\underset{\sim}{A} = \underset{\sim}{B}'$ if $\underset{\sim}{S} = \underset{\sim}{I}_p$.

2.2. Eigenvectors with zero coefficients

In order to give conditions under which some coefficients of an eigenvector of $\underset{\sim}{S}^{-1}\underset{\sim}{T}$ are zero, we partition the matrices $\underset{\sim}{S}$ and $\underset{\sim}{T}$ as

$$\underset{\sim}{S} = \begin{pmatrix} \underset{\sim}{S}_{11} & \underset{\sim}{S}_{12} \\ \underset{\sim}{S}_{21} & \underset{\sim}{S}_{22} \end{pmatrix}; \quad \underset{\sim}{T} = \begin{pmatrix} \underset{\sim}{T}_{11} & \underset{\sim}{T}_{12} \\ \underset{\sim}{T}_{21} & \underset{\sim}{T}_{22} \end{pmatrix}, \tag{2.1}$$

where $\underset{\sim}{S}_{11}$ and $\underset{\sim}{T}_{11}$ are pxp, and $\underset{\sim}{S}_{22}$ and $\underset{\sim}{T}_{22}$ are (p-q)x(p-q). Let λ denote an eigenvalue of $\underset{\sim}{S}^{-1}\underset{\sim}{T}$ and $\underset{\sim}{b} = (\underset{\sim}{b}_1', \underset{\sim}{b}_2')'$ an associated eigenvector. Then the equation $\underset{\sim}{T}\underset{\sim}{b} = \lambda\underset{\sim}{S}\underset{\sim}{b}$ can be written in partitioned form as

$$\underset{\sim}{T}_{11}\underset{\sim}{b}_1 + \underset{\sim}{T}_{12}\underset{\sim}{b}_2 = \lambda\underset{\sim}{S}_{11}\underset{\sim}{b}_1 + \lambda\underset{\sim}{S}_{12}\underset{\sim}{b}_2 \tag{2.2}$$

$$\underset{\sim}{T}_{21}\underset{\sim}{b}_1 + \underset{\sim}{T}_{22}\underset{\sim}{b}_2 = \lambda\underset{\sim}{S}_{21}\underset{\sim}{b}_1 + \lambda\underset{\sim}{S}_{22}\underset{\sim}{b}_2. \tag{2.3}$$

This leads to

Theorem 2: Let $\underset{\sim}{S}$ and $\underset{\sim}{T}$ be defined as in the above text. Then

(i) If $\underset{\sim}{T}\begin{pmatrix} \underset{\sim}{b} \\ \underset{\sim}{0} \end{pmatrix} = \lambda\underset{\sim}{S}\begin{pmatrix} \underset{\sim}{b} \\ \underset{\sim}{0} \end{pmatrix}$ for some $\lambda \neq 0$, $\underset{\sim}{b} \neq \underset{\sim}{0} \in \mathbb{R}^q$, then $\underset{\sim}{T}_{11}\underset{\sim}{b} = \lambda\underset{\sim}{S}_{11}\underset{\sim}{b}$ and $\underset{\sim}{T}_{21}\underset{\sim}{b} = \lambda\underset{\sim}{S}_{21}\underset{\sim}{b}$.

(ii) If $\underset{\sim}{T}_{11}\underset{\sim}{b}^* = \lambda^*\underset{\sim}{S}_{11}\underset{\sim}{b}^*$ and $\underset{\sim}{T}_{21}\underset{\sim}{b}^* = \lambda^*\underset{\sim}{S}_{21}\underset{\sim}{b}^*$ for some $\lambda^* \neq 0$, $\underset{\sim}{b}^* \neq \underset{\sim}{0} \in \mathbb{R}^q$, then $\underset{\sim}{T}\begin{pmatrix} \underset{\sim}{b}^* \\ \underset{\sim}{0} \end{pmatrix} = \lambda^*\underset{\sim}{S}\begin{pmatrix} \underset{\sim}{b}^* \\ \underset{\sim}{0} \end{pmatrix}$.

(iii) Let λ denote a simple characteristic root of $\underset{\sim}{S}^{-1}\underset{\sim}{T}$ and $\begin{pmatrix} \underset{\sim}{b}_1 \\ \underset{\sim}{b}_2 \end{pmatrix}$ the associated eigenvector. If $\underset{\sim}{T}_{11}\underset{\sim}{a} = \lambda\underset{\sim}{S}_{11}\underset{\sim}{a}$ and $\underset{\sim}{T}_{21}\underset{\sim}{a} = \lambda\underset{\sim}{S}_{21}\underset{\sim}{a}$ for some $\underset{\sim}{a} \neq \underset{\sim}{0}$, then $\underset{\sim}{b}_1 = \underset{\sim}{a}$ (up to multiplication with a scalar $\neq 0$), and $\underset{\sim}{b}_2 = \underset{\sim}{0}$.

Proof: (i) and (ii) follow immediately from (2.2) and (2.3). (iii) follows from (ii) and the simplicity of λ.

Thus, if an eigenvector of $\underset{\sim}{S}^{-1}\underset{\sim}{T}$ contains zeros in the last p-q positions, it can be found from the submatrices $\underset{\sim}{S}_{11}$ and $\underset{\sim}{T}_{11}$, but $\underset{\sim}{T}_{11}\underset{\sim}{b}^* = \lambda^*\underset{\sim}{S}_{11}\underset{\sim}{b}^*$ is only a necessary condition for $\underset{\sim}{T}\begin{pmatrix}\underset{\sim}{b}^*\\ \underset{\sim}{0}\end{pmatrix} = \lambda^*\underset{\sim}{S}\begin{pmatrix}\underset{\sim}{b}^*\\ \underset{\sim}{0}\end{pmatrix}$, even if λ^* is a characteristic root of $\underset{\sim}{S}^{-1}\underset{\sim}{T}$. However, things get somewhat simpler for the eigenvectors associated with the extreme roots.

Lemma 1: Let $\lambda_1^{(p)}$ and $\lambda_1^{(q)}$ denote the largest eigenvalues of $\underset{\sim}{S}^{-1}\underset{\sim}{T}$ and $\underset{\sim}{S}_{11}^{-1}\underset{\sim}{T}_{11}$, respectively. Then $\lambda_1^{(p)} \geq \lambda_1^{(q)}$.

Proof: With $\underset{\sim}{a} = (\underset{\sim}{a}_1', \underset{\sim}{a}_2')' \in \mathbb{R}^p$,

$$\lambda_1^{(p)} = \max_{\underset{\sim}{a}\in\mathbb{R}^p} \underset{\sim}{a}'\underset{\sim}{T}\underset{\sim}{a}/\underset{\sim}{a}'\underset{\sim}{S}\underset{\sim}{a} \geq \max_{\substack{\underset{\sim}{a}\in\mathbb{R}^p \\ \underset{\sim}{a}_2=\underset{\sim}{0}}} \underset{\sim}{a}'\underset{\sim}{T}\underset{\sim}{a}/\underset{\sim}{a}'\underset{\sim}{S}\underset{\sim}{a} = \lambda_1^{(q)}.$$

Theorem 3: Let λ denote the largest characteristic root of $\underset{\sim}{S}^{-1}\underset{\sim}{T}$. Then

(i) If $\underset{\sim}{T}_{11}\underset{\sim}{b}^* = \lambda\underset{\sim}{S}_{11}\underset{\sim}{b}^*$ for some $\underset{\sim}{b}^* \neq \underset{\sim}{0} \in \mathbb{R}^q$, then $\underset{\sim}{T}\begin{pmatrix}\underset{\sim}{b}^*\\ \underset{\sim}{0}\end{pmatrix} = \lambda\underset{\sim}{S}\begin{pmatrix}\underset{\sim}{b}^*\\ \underset{\sim}{0}\end{pmatrix}$.

(ii) Suppose λ is simple and $\underset{\sim}{b} = \begin{pmatrix}\underset{\sim}{b}_1\\ \underset{\sim}{b}_2\end{pmatrix}$ is the associated eigenvector. If $\underset{\sim}{T}_{11}\underset{\sim}{b}^* = \lambda\underset{\sim}{S}_{11}\underset{\sim}{b}^*$ for some $\underset{\sim}{b}^* \neq \underset{\sim}{0} \in \mathbb{R}^q$, then $\underset{\sim}{b}_1 = \underset{\sim}{b}^*$ and $\underset{\sim}{b}_2 = \underset{\sim}{0}$.

Proof: (ii) follows immediately from (i) and the simplicity of λ. To show (i), assume that $\underset{\sim}{b}^*$ is normalized such that $\underset{\sim}{b}^{*\prime}\underset{\sim}{S}_{11}\underset{\sim}{b}^* = 1$ and consider the function

$$f(\underset{\sim}{x}) = \frac{\begin{pmatrix}\underset{\sim}{b}^*\\ \underset{\sim}{x}\end{pmatrix}'\underset{\sim}{T}\begin{pmatrix}\underset{\sim}{b}^*\\ \underset{\sim}{x}\end{pmatrix}}{\begin{pmatrix}\underset{\sim}{b}^*\\ \underset{\sim}{x}\end{pmatrix}'\underset{\sim}{S}\begin{pmatrix}\underset{\sim}{b}^*\\ \underset{\sim}{x}\end{pmatrix}} = \frac{\lambda + 2\underset{\sim}{b}^{*\prime}\underset{\sim}{T}_{12}\underset{\sim}{x} + \underset{\sim}{x}'\underset{\sim}{T}_{22}\underset{\sim}{x}}{1 + 2\underset{\sim}{b}^{*\prime}\underset{\sim}{S}_{12}\underset{\sim}{x} + \underset{\sim}{x}'\underset{\sim}{S}_{22}\underset{\sim}{x}} \qquad (2.4)$$

for $\underset{\sim}{x} \in \mathbb{R}^{p-q}$. The vector of first derivatives of f with regard to the elements of $\underset{\sim}{x}$, evaluated at $\underset{\sim}{x} = \underset{\sim}{0}$, is

$$\left.\frac{\partial f}{\partial \underset{\sim}{x}}\right|_{\underset{\sim}{x}=\underset{\sim}{0}} = 2(\underset{\sim}{T}_{21}\underset{\sim}{b}^* - \lambda \underset{\sim}{S}_{21}\underset{\sim}{b}^*). \tag{2.5}$$

If $\underset{\sim}{T}_{21}\underset{\sim}{b}^* \neq \lambda \underset{\sim}{S}_{21}\underset{\sim}{b}^*$, there exists $\underset{\sim}{x}_0 \in \mathbb{R}^{p-q}$, $\underset{\sim}{x}_0 \neq \underset{\sim}{0}$, such that $f(\underset{\sim}{x}_0) = \lambda_0 > \lambda = f(\underset{\sim}{0})$. By Lemma 1 this is impossible. Therefore $\underset{\sim}{T}_{21}\underset{\sim}{b}^* = \lambda \underset{\sim}{S}_{21}\underset{\sim}{b}^*$, and (i) follows now from Theorem 2.

In statistical applications we will often have coefficients of eigenvectors which are close to zero, but not exactly zero. Intuitively we would expect that ignoring such a coefficient or replacing it by zero should not affect the associated ratio of variances too much. This is confirmed by the following theorem.

Theorem 4: Let $\underset{\sim}{S}$ and $\underset{\sim}{T}$ be defined and partitioned as above, and let $q = p-1$, i.e. $\underset{\sim}{S}_{22} = s_{22}$ and $\underset{\sim}{T}_{22} = t_{22}$ are scalars. Let $\underset{\sim}{b}$ denote an eigenvector of $\underset{\sim}{S}^{-1}\underset{\sim}{T}$ and λ the associated eigenvalue, and let $\underset{\sim}{b}$ be normalized such that $\underset{\sim}{b}'\underset{\sim}{S}\underset{\sim}{b} = 1$. Let $\underset{\sim}{b}_x = \underset{\sim}{b} + \begin{pmatrix} \underset{\sim}{0} \\ x \end{pmatrix}$, where $x \in \mathbb{R}$, and let $f(x) = \underset{\sim}{b}_x'\underset{\sim}{T}\underset{\sim}{b}_x / \underset{\sim}{b}_x'\underset{\sim}{S}\underset{\sim}{b}_x$. Then, in a neighborhood of zero,

$$f(x) = \lambda + (t_{22} - \lambda s_{22})x^2 + \mathcal{O}(x^3). \tag{2.6}$$

Proof: Expand $f(x)$ in a Taylor series about zero.

Thus, if the p-th coefficient b_p of eigenvector b is close to zero, neglecting b_p will change the ratio of the two quadratic forms from $f(0) = \lambda$ to approximately $f(-b_p) \approx \lambda + (t_{22} - \lambda s_{22})b_p^2$.

Lemma 1 and Theorem 3 can of course be formulated analogously for the smallest eigenvalue of $\underset{\sim}{S}^{-1}\underset{\sim}{T}$ and its associated eigenvector(s).

2.3. Statistical hypotheses about eigenvectors

If two covariance matrices $\underset{\sim}{\Sigma}_1$ and $\underset{\sim}{\Sigma}_2$ are not identical, three different kinds of questions might be interesting to ask.

1. Any eigenvector of $\underset{\sim}{\Sigma}_1^{-1}\underset{\sim}{\Sigma}_2$ associated with a root $\lambda = 1$ gives no information about differences in variability. It might therefore be interesting to know how many (and which) eigenvalues are different from 1. Rao (1983) has treated a similar problem in the context of "familial correlations", but his approach applies as well to the comparison of covariance matrices. If p-k eigenvalues of $\underset{\sim}{\Sigma}_1^{-1}\underset{\sim}{\Sigma}_2$ are unity, then the simultaneous spectral decomposition of $\underset{\sim}{\Sigma}_1$ and $\underset{\sim}{\Sigma}_2$ (cf. Theorem 1) can be written as

$$\underset{\sim}{\Sigma}_1 = \underset{\sim}{\beta}_1\underset{\sim}{\beta}_1' + \ldots + \underset{\sim}{\beta}_p\underset{\sim}{\beta}_p' \tag{2.7}$$

$$\begin{aligned}\underset{\sim}{\Sigma}_2 &= \lambda_1\underset{\sim}{\beta}_1\underset{\sim}{\beta}_1' + \ldots + \lambda_k\underset{\sim}{\beta}_k\underset{\sim}{\beta}_k' + \underset{\sim}{\beta}_{k+1}\underset{\sim}{\beta}_{k+1}' + \ldots + \underset{\sim}{\beta}_p\underset{\sim}{\beta}_p' \\ &= \underset{\sim}{\Sigma}_1 + \sum_{j=1}^{k}(\lambda_j - 1)\underset{\sim}{\beta}_j\underset{\sim}{\beta}_j' \\ &= \underset{\sim}{\Sigma}_1 + \underset{\sim}{\Gamma}, \text{ say,}\end{aligned} \tag{2.8}$$

where the λ_j are not necessarily greater than 1, and $\underset{\sim}{\Gamma}$ is a symmetric matrix of rank k. For more details the reader is referred to Rao (1983, paragraph 5).

2. If we are merely interested in finding a subset of variables which would be sufficient to reject the null hypothesis $\underset{\sim}{\Sigma}_1 = \underset{\sim}{\Sigma}_2$, then simultaneous confidence intervals, based on the largest and smallest roots' criterion, might be a useful tool. Given quantiles c_1 and c_p from the null distribution of the criterion such that

$$P(c_1 \geq \ell_1 \geq \ell_p \geq c_p \mid \underset{\sim}{\Sigma}_1 = \underset{\sim}{\Sigma}_2) = 1-\alpha \tag{2.9}$$

(where ℓ_1 and ℓ_p are the sample extreme roots), we can construct any linear combination $(\underset{\sim}{a}'\underset{\sim}{X}^{(1)}, \underset{\sim}{a}'\underset{\sim}{X}^{(2)})$ and compare the associated sample ratio of variances $F(\underset{\sim}{a}) = \underset{\sim}{a}'\underset{\sim}{S}_2\underset{\sim}{a}/\underset{\sim}{a}'\underset{\sim}{S}_1\underset{\sim}{a}$ with c_1 and c_p. The acceptance region for H_0 is the set $\{\underset{\sim}{a} \in \mathbb{R}^q: \; c_p \leq \underset{\sim}{a}'\underset{\sim}{S}_2\underset{\sim}{a}/\underset{\sim}{a}'\underset{\sim}{S}_1\underset{\sim}{a} \leq c_1\}$, and it might be interesting to

identify vectors $\underset{\sim}{a} \in \mathbb{R}^p$ which are in a sense "simpler" than the extreme eigenvectors of $\underset{\sim}{S}_1^{-1}\underset{\sim}{S}_2$, but which are still outside the acceptance region.

3. The next two sections of this paper will be devoted to the problem of simplifying the linear combinations associated with the extreme ratios of variances by discarding redundant variables. Some motivation for this has been given in Section 1. More generally, we may wish to know whether some variables have zero coefficients in several eigenvectors of $\underset{\sim}{\Sigma}_1^{-1}\underset{\sim}{\Sigma}_2$ simultaneously.

Let $\underset{\sim}{\beta}_1,\ldots,\underset{\sim}{\beta}_p$ denote the eigenvectors of $\underset{\sim}{\Sigma}_1^{-1}\underset{\sim}{\Sigma}_2$ and partition them into the first q and the last p-q coefficients, i.e. write $\underset{\sim}{\beta}_j = (\underset{\sim}{\beta}_{j1}', \underset{\sim}{\beta}_{j2}')'$, where $\underset{\sim}{\beta}_{j1}$ has dimension q. Let v denote a subset of m integers between 1 and p. Then we define the <u>hypothesis of simultaneous redundancy of (the last) p-q variables for m eigenvectors</u> as

$$H_v(p,q): \quad \underset{\sim}{\beta}_{j2} = \underset{\sim}{0} \quad \text{for all } j \in v. \tag{2.10}$$

Of course $H_v(p,q)$ makes only sense if m, the number of elements in v, does not exceed q.

Strictly speaking $H_v(p,q)$ is only defined if $\lambda_i \neq \lambda_j$ for all pairs (i,j) such that $i \in v$, $j \notin v$. Otherwise, if $\lambda_i = \lambda_j$ for some $i \in v$, $j \notin v$, the associated eigenvectors are not uniquely defined. However, it might still make sense to try to simplify by taking a linear combination of all eigenvectors associated with the multiple root such that as many coefficients as possible vanish.

As a numerical example to illustrate this, let

$$\underset{\sim}{\Sigma}_1 = \begin{pmatrix} 1 & -.5 & -.5 \\ -.5 & .75 & .5 \\ -.5 & .5 & .75 \end{pmatrix} \text{ and } \underset{\sim}{\Sigma}_2 = \begin{pmatrix} 4 & -2 & -2 \\ -2 & 2.25 & .25 \\ -2 & .25 & .25 \end{pmatrix}.$$

The eigenvalues of $\underset{\sim}{\Sigma}_1^{-1}\underset{\sim}{\Sigma}_2$ are $\lambda_1 = \lambda_2 = 4$, $\lambda_3 = 1$. Suppose we wish to find an eigenvector associated with the largest root which has as many zeros as possible. Then taking $\underset{\sim}{\beta}_1 = (1, 0, 0)'$ would be a good choice.

3. EXPLORATORY ANALYSIS OF THE EXTREME EIGENVECTORS: FISHING FOR REDUCED SETS OF VARIABLES

Quite often in applied multivariate analysis (as well as in multiple regression) it is unknown in advance which variables should be discarded or which coefficients of a linear equation should be tested. Subset selection procedures such as "best subset" regression, stepwise regression and backward elimination are widely used. In this section we are going to propose a similar procedure for analyzing the extreme eigenvectors of $\underset{\sim}{S}_1^{-1}\underset{\sim}{S}_2$.

The basic idea behind this procedure is taken from theorem 3: If coefficients of the largest (smallest) eigenvector are zero, the eigenvector can as well be found from the reduced set of variables, and if we suspect that a coefficient deviates from zero only by sampling error, the associated variable should be discarded. By theorem 4, annihilating a small coefficient b (without altering the other coefficients) will change the associated ratio of variances by an amount proportional to b^2, but by maximizing again over the (p-1)-dimensional subset the actual loss is most often smaller than this. It makes therefore sense to judge the influence of a variable on the linear combination with largest (smallest) ratio of variances by the change in the ratio of variances which occurs when the variable is omitted. If the variables are correlated, it can also happen that even variables with "large" coefficients can be eliminated without affecting the largest (smallest) ratio of variances much. In such cases, the elimination will be associated with major changes in the coefficients of other variables.

Although the partial statistics to be defined could as well be used in other selection procedures, we will illustrate it for simplicity by a backward

elimination method. In any case, an overall test of significance for H_0: $\underset{\sim}{\Sigma}_1 = \underset{\sim}{\Sigma}_2$ should precede the analysis of the extreme eigenvectors in order to prevent ourselves from the ridiculous results which may occur in "blind" subset selection. (Freedman 1983, Rencher and Pun 1980). Analyzing the largest (smallest) eigenvector will in most cases only make sense if the largest (smallest) root of $\underset{\sim}{\Sigma}_1^{-1}\underset{\sim}{\Sigma}_2$ is actually larger (smaller) than 1. Roy's largest and smallest roots' criterion is therefore a natural candidate for the overall test of H_0: $\underset{\sim}{\Sigma}_1 = \underset{\sim}{\Sigma}_2$ (although other tests may have better power in some situations, see Pillai and Chu (1979) and references therein).

Let now $\underset{\sim}{b}_1,\ldots,\underset{\sim}{b}_p$ denote the characteristic vectors of $\underset{\sim}{S}_1^{-1}\underset{\sim}{S}_2$, where $\underset{\sim}{S}_1$ and $\underset{\sim}{S}_2$ are p.d.s. sample covariance matrices, and $\ell_1 \geq \ldots \geq \ell_p$ the associated characteristic roots. Write

$$\begin{aligned} Y_{max}^{(p)} &= \underset{\sim}{b}_1'\underset{\sim}{X} \\ Y_{min}^{(p)} &= \underset{\sim}{b}_p'\underset{\sim}{X} \end{aligned} \tag{3.1}$$

for the two linear combinations with extreme ratios of variances ℓ_1 and ℓ_p respectively. (In order to avoid too many indices, we do not distinguish notationally between the variables in the first group and those in the second group). We will also write $F_{max}^{(p)}$ and $F_{min}^{(p)}$ instead of ℓ_1 and ℓ_p. Let us now introduce partial statistics for the analysis of Y_{max}. Denote by $F_{max}^{(p-1)}(i)$ the largest ratio of variances that can be obtained <u>without</u> using X_i. By theorem 3, $F_{max}^{(p-1)} \leq F_{max}^{(p)}$, with equality exactly if variable X_i has a zero coefficient in $Y_{max}^{(p)}$ (or the coefficient can be chosen to be zero if the largest root is not simple). The ratio

$$PCF(i) = F_{max}^{(p-1)}(i)/F_{max}^{(p)}(i) \qquad (3.2)$$

measures the Partial Change of the F-ratio due to elimination of X_i. Clearly, $0 < PCF(i) \leq 1$ holds. In addition, it is useful to compute $LPCF(i) = \log PCF(i)$ - the reason for this will be given a little later.

Let us now illustrate the use of these partial statistics in a backward elimination procedure. The first step of the analysis of Y_{max} consists of a table of coefficients of $Y_{max}^{(p)}$ together with the partial statistics $F_{max}^{(p-1)}(i)$, PCF(i) and LPCF(i) for all variables. In the second step, eliminate variable X_i if $|LPCF(i)| \leq |LPCF(j)|$ for j=1,...,p. (Normally this variable will be unique). Replace p by p-1 and compute again, as in step 1, a list of coefficients of Y_{max} and partial statistics.

This procedure can be continued until a specified stopping criterion is satisfied or until (in step p) only one variable remains.

For the analysis of Y_{min} (based on the "smallest root version" of lemma 1 and theorem 3) we define $F_{min}^{(p-1)}(i)$ as the smallest ratio of variances that can be obtained without variable X_i, and put $PCF(i) = F_{min}^{(p-1)}(i)/F_{min}^{(p)}$, $LPCF(i) = \log PCF(i)$. Clearly, $1 \leq LPCF(i) < \infty$, and the variable to be eliminated is again X_i if $|LPCF(i)| \leq |LPCF(j)|$ for j=1,...,p.

The reasons for computing LPCF(i) can now be given as follows:

- The elimination rule can be formulated identically for Y_{max} and Y_{min} in terms of |LPCF|.
- The values of |LPCF| are in the "familiar" range from zero to infinity, with a value close to zero indicating that the corresponding variable can be eliminated without much loss of information.

- The values of |LPCF| do not depend on the numbering of the groups, i.e. |LPCF(i)| is the same whether we analyze Y_{max} based on $S_1^{-1}S_2$ or Y_{min} based on $S_2^{-1}S_1$. It is therefore convenient to use the same stopping criterion for the analyses of both Y_{max} and Y_{min}: e.g., stop the procedure if $\min_{i \leq p} |LPCF(i)| > c$ for a given constant c.

Numerical example: Flury and Riedwyl (1983) measured the following six variables on 100 real and 100 forged swiss bank notes:

LENGTH = length of the bank note

LEFT = width of the bank note, measured on the left side

RIGHT = width of the bank note, measured on the right side

BOTTOM = width of the lower margin

TOP = width of the upper margin

DIAGONAL = length of the print diagonal.

The two sample covariance matrices are given in table 1. Under $H_0: \Sigma_1 = \Sigma_2$, assuming multivariate normality in both groups,

$$p(.43 \leq \text{smallest root} \leq \text{largest root} \leq 2.31) = .95 \qquad (3.3)$$

approximately. (These quantiles are taken from an unpublished simulation study). Table 1c gives the eigenvalues and associated eigenvectors of $S_1^{-1}S_2$. Since both F_{max} and F_{min} exceed the respective limits, it may be worthwhile to analyze Y_{max} as well as Y_{min}. We give here a detailed stepwise analysis of Y_{min} to illustrate the elimination procedure.

Step 1: all six variables ($F_{min}^{(6)} = .284$)

variable	coefficient	PCF	LPCF	$F_{min}^{(5)}$
LENGTH	-.396	1.034	.033	.293
LEFT	-1.174	1.092	.088	.310
RIGHT	-.374	1.014	.014	.288
BOTTOM	-.512	1.515	.416	.430
TOP	-.842	1.539	.431	.437
DIAGONAL	.587	1.216	.195	.345

The first three variables have rather small values of LPCF, and the elimination of each of them wouldn't affect F_{min} much. The first candidate is RIGHT, whose elimination increases F_{min} by mere 1.4 percent (PCF(RIGHT) = 1.014). Thus we get

Step 2: elimination of RIGHT ($F_{min}^{(5)} = .288$)

variable	coefficient	PCF	LPCF	$F_{min}^{(4)}$
LENGTH	-.347	1.026	.026	.295
LEFT	-1.522	1.423	.353	.410
BOTTOM	-.498	1.519	.418	.437
TOP	-.849	1.536	.429	.442
DIAGONAL	.525	1.214	.194	.350

Compared with step 1, variable LEFT has now more weight (LPCF(LEFT) = .353 instead of .088), and the coefficient of LEFT has changed more than all other

coefficients, which remain essentially constant. A similar effect occurs if we eliminate LEFT instead of RIGHT. This means that we can remove either LEFT or RIGHT without much loss of information, but not both of them. The next variable to be eliminated is now LENGTH.

Step 3: elimination of LENGTH ($F_{min}^{(4)} = .295$)

variable	coefficient	PCF	LPCF	$F_{min}^{(3)}$
LEFT	1.734	1.826	.602	.539
BOTTOM	.462	1.500	.406	.443
TOP	.845	1.512	.413	.446
DIAGONAL	-.491	1.188	.172	.351

This table is very similar to the one of step 2, except perhaps for the fact that LEFT has again gained some weight. All LPCF-values are now rather large (the minimum increase of F_{min} by eliminating one more variable is 18.8 percent), which might be taken as a criterion for stopping the procedure. What happens if we proceed nevertheless? Well, let's see.

Step 4: elimination of DIAGONAL ($F_{min}^{(3)} = .351$)

variable	coefficient	PCF	LPCF	$F_{min}^{(2)}$
LEFT	2.097	2.143	.762	.752
BOTTOM	.359	1.283	.249	.450
TOP	.794	1.348	.299	.473

Step 5: elimination of BOTTOM ($F_{min}^{(2)} = .450$)

variable	coefficient	PCF	LPCF	$F_{min}^{(1)}$
LEFT	2.511	2.134	.758	.960
TOP	.397	1.090	.086	.491

Finally, in step 6, we are left with variable LEFT and its univariate ratio of variances.

In step 4 we can see that BOTTOM and TOP have lost some of their importance. This can again be taken as a hint that the elimination of DIAGONAL was not justified. The same effect occurs even more distinctly in step 5, after the elimination of BOTTOM: variable TOP, which is highly correlated with BOTTOM in both groups, loses its importance almost entirely. The importance of BOTTOM and TOP for Y_{min} lies therefore in their joint contribution - which shows another advantage of the elimination procedure, for a forward selection algorithm would not discover the joint importance of two variables.

If we decide (arbitrarily) that eliminating a single variable should not increase F_{min} by more than 10 percent, we have to stop after step 3, thus getting

$$Y_{min}^{(4)} = 1.73 \text{ LEFT} + .46 \text{ BOTTOM} + .84 \text{ TOP} - .49 \text{ DIAGONAL}$$

as a reduced solution. The associated $F_{min}^{(4)}$ is .295, which is not much larger than $F_{min}^{(6)}$. Furthermore, $Y_{min}^{(4)}$ is still clearly in the rejection region for H_0: $\underset{\sim}{\Sigma}_1 = \underset{\sim}{\Sigma}_2$. Since no other eigenvector of $\underset{\sim}{S}_1^{-1}\underset{\sim}{S}_2$ is in the rejection region, it is reasonable to summarize the "too small variability" of the forged notes, compared with the real ones, in the single linear combination $Y_{min}^{(4)}$.

What is the interpretation of $Y_{min}^{(4)}$? Since we do not know much about how bank notes are produced, it is not obvious. However, the importance of LEFT (or RIGHT), BOTTOM and TOP suggests that the forger's mechanism for cutting the notes and putting the print on the paper is very precise. After all, it is rather surprising (and speaks for the quality of the forger's work) that there is a linear combination with smaller variability in the forged notes than in the real notes!

Finally, some remarks.

1. In the above example, the six univariate ratios of variances are: LENGTH .83, LEFT .49, RIGHT .70, BOTTOM 3.10, TOP .96, DIAGONAL 1.56. Comparing these ratios with the stepwise analysis of Y_{min}, we see that variables with a univariate F-ratio smaller than 1 may be redundant for Y_{min} (e.g. LENGTH, RIGHT). On the other hand there may be variables playing an important role in Y_{min}, but having a univariate F-ratio larger than 1 (BOTTOM, DIAGONAL). This shows that a previous selection of variables based on univariate F-ratios might be badly misleading.

2. The example shows also that the partial statistic |LPCF| behaves much like partial statistics used in regression or discriminant analysis. Actually, similar phenomena as those described in the above analysis of Y_{min} occur quite often in these two methods, especially when some variables are nearly collinear.

3. If, say, the two largest eigenvalues of $S_1^{-1}S_2$ are close, there may be different or even disjoint subsets of less than p variables with approximately the same F_{max}. Automatic elimination according to the principle of rejecting the variable with smallest |LPCF| will of course only find a hierarchical

sequence of subsets, but nevertheless the partial statistics introduced here can be used in a search for alternative solutions.

4. It might be argued that Y_{min} and Y_{max} should be analyzed simultaneously. However, this would make things much more complicated as soon as different variables are removed from Y_{max} and Y_{min} respectively. Furthermore, under normal theory assumptions, these linear combinations are asymptotically independent, and in an exploratory context it is therefore reasonable to analyze Y_{max} and Y_{min} separately.

5. The method proposed in this section is clearly a fishing trip (Selvin and Stuart 1966). Many users of multiple regression are not aware of the fact that the null distribution of the smallest partial F-statistic is not an F-distribution, and that test statistics should therefore be used with utmost care, if at all, in an exploratory context - a "drawback to be concerned about" (Draper and Smith 1981, p. 311). To avoid a similar confusion in the method of this paper I recommend therefore strongly not to use the asymptotic chi square statistic (section 4.3) for the stepwise procedure, but rather the statistics PCF and LPCF.

6. Instead of setting coefficients equal to zero, we might also wish to simplify Y_{min} and Y_{max} in the sense of equating some coefficients or forcing them to be in a certain relationship. Suppose, e.g., that we wish to try a solution where the coefficients of variables X_1 and X_2 are in the ratio a/b. This can easily be done by computing the extreme eigenvectors using variables $aX_1 + bX_2, X_3, \ldots, X_p$. If simplification is done by rounding coefficients to a fixed number of decimal places, it might be possible to give bounds for the maximum change in the ratio of variances induced by the rounding. Bibby (1980)

has obtained such bounds for the principal component case.

7. The partial statistics defined in this section can of course readily be generalized to measure the influence of two or more variables on Y_{max} (Y_{min}) simultaneously.

4. TESTING FOR REDUNDANCY OF VARIABLES

4.1. Notation and terminology

In order to establish notation, we indicate here briefly the matrix techniques to be used in the following two subsections. The notation used here parallels closely the notation of Tyler (1981).

Let $\underset{\sim}{M}$ denote a real p×p-matrix which is symmetric in the metric of a p.d.s. matrix $\underset{\sim}{\Gamma}$, i.e. $\underset{\sim}{\Gamma}\underset{\sim}{M}$ is symmetric. The p eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p$ of $\underset{\sim}{M}$ are real; they form the spectral set of $\underset{\sim}{M}$. The eigenvectors $\underset{\sim}{\beta}_i$ (i = 1,...,p) of $\underset{\sim}{M}$ can be chosen and normalized such that $\underset{\sim}{\beta}_i'\underset{\sim}{\Gamma}\underset{\sim}{\beta}_j = \delta_{ij}$, where δ_{ij} is the Kronecker delta. We will always assume in the sequel that the eigenvectors satisfy this condition, even if there are multiple eigenvalues. The projection operator associated with the i-th eigenvalue is denoted by $\underset{\sim}{P}_i = \underset{\sim}{\beta}_i\underset{\sim}{\beta}_i'\underset{\sim}{\Gamma}$. By the spectral decomposition theorem these p projection matrices add to the identity matrix, i.e. $\underset{\sim}{P}_1 + \ldots + \underset{\sim}{P}_p = \underset{\sim}{I}_p$, and $\underset{\sim}{M} = \sum_{i=1}^{p} \lambda_i \underset{\sim}{P}_i$.

The Moore-Penrose inverse of $\underset{\sim}{M}$, denoted by $\underset{\sim}{M}^+$, is a uniquely defined generalized inverse and can be written as $\underset{\sim}{M}^+ = \sum_{\lambda_i \neq 0} \lambda_i^{-1} \underset{\sim}{P}_i$. Note that $\underset{\sim}{P}_i^+ = \underset{\sim}{P}_i$.

We will also use the "vec"-transformation (which transforms a r×s-matrix to a rs×1-vector by stacking its columns), the Kronecker matrix product (denoted by the symbol "⊗"), and the commutation matrix of order $p^2 \times p^2$ (denoted by $\underset{\sim}{I}_{(p,p)}$).

The reader who is unfamiliar with the above matrix techniques will find the following references useful: Nerring (1970); Rao (1973, Chapter 1); Searle (1982, Chapter 11A); Mardia, Kent and Bibby (1979, Appendix A); Muirhead (1982, p. 17, 73ff, 90); Magnus and Neudecker (1979); Neudecker (1969); Basilevsky (1983, Chapters 5 and 6), Rao and Mitra (1971).

However, the test statistic $R_V(p,q)$ derived in section 4.3 is written in terms of the usual, more familiar matrix operations. The reader who is not interested in the technical details of the proof can understand and compute it with a basic knowledge of matrix algebra.

4.2. Tyler's asymptotic test on eigenvectors

Tyler (1981) has considered the following rather general situation: Let $\underset{\sim}{M}$ be a p×p-matrix symmetric in the metric of the p.d.s. matrix $\underset{\sim}{\Gamma}$, with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p$. Let w denote a subset of m integers from $\{1,2,\ldots,p\}$ $(1 \leq m < p)$. Let $\underset{\sim}{A}$ denote a fixed p×r-matrix with $\text{rank}(\underset{\sim}{A}) = r$. Under the assumption $\min_{\substack{i\in w \\ j\notin w}} |\lambda_i - \lambda_j| > 0$, the following two hypotheses on $\underset{\sim}{M}$ are treated:

- For $r \leq m$, the hypothesis

  H_0: The columns of $\underset{\sim}{A}$ lie in the subspace generated by the set of eigenvectors of $\underset{\sim}{M}$ associated with the m roots λ_i for $i \in w$.

- For $r \geq m$, the hypothesis

  H_0^*: The eigenvectors of $\underset{\sim}{M}$ associated with the roots λ_i for $i \in w$ lie in the subspace generated by the columns of $\underset{\sim}{A}$.

Putting $\underset{\sim}{M} = \underset{\sim}{\Sigma}_1^{-1}\underset{\sim}{\Sigma}_2$ and $\underset{\sim}{A} = \begin{pmatrix} \underset{\sim}{I}_q \\ \underset{\sim}{0} \end{pmatrix}$, our hypothesis of redundancy $H_V(p,q)$ can be written in the form H_0^*.

Tyler's test is based on a sequence of estimates $\underset{\sim}{M}_n$ of $\underset{\sim}{M}$ such that $a_n(\underset{\sim}{M}_n - \underset{\sim}{M})$ converges in distribution to a multivariate normal distribution, where a_n is an increasing sequence of real numbers. Moreover, we need a sequence $\underset{\sim}{\Gamma}_n$ of p.d.s. matrices converging to $\underset{\sim}{\Gamma}$ in probability, such that $\underset{\sim}{M}_n$ is symmetric in the metric of $\underset{\sim}{\Gamma}_n$. Denote the eigenvectors of $\underset{\sim}{M}$ by $\underset{\sim}{\beta}_i$ $(i = 1,\ldots,p)$, and the associated eigenprojections by $\underset{\sim}{P}_i = \underset{\sim}{\beta}_i\underset{\sim}{\beta}_i'\underset{\sim}{\Gamma}$. Analogously, denote the eigenvectors of $\underset{\sim}{M}_n$ by $\underset{\sim}{b}_i$ and let $\hat{\underset{\sim}{P}}_i = \underset{\sim}{b}_i\underset{\sim}{b}_i'\underset{\sim}{\Gamma}_n$.

Tyler's test is based on the asymptotic distribution of $\hat{P}_W = \sum_{i \in W} \hat{P}_i$. He shows that, under fairly general conditions (which hold in our case), the random vector

$$V_n = \mathrm{vec}[a_n(I_p - \hat{P}_W)A] \tag{4.1}$$

converges in distribution to Normal $(0, \psi_0(A))$, where

$$\psi_0(A) = (A' \otimes I_p) C_W' \psi C_W (A \otimes I_p), \tag{4.2}$$

$$C_W = \sum_{i \in W} \sum_{j \notin W} (\lambda_i - \lambda_j)^{-1} P_i \otimes P_j', \tag{4.3}$$

and ψ is the asymptotic covariance matrix of $\mathrm{vec}[a_n(M_n - M)]$.

Suppose that we have a sequence of p.d.s. estimates ψ_n of ψ, converging to ψ in probability. Then an estimate of $\psi_0(A)$ which is consistent under H_0 is obtained by defining

$$\hat{\psi}_0(A) = (A' \otimes I_p) \hat{C}_W' \psi_n \hat{C}_W (A \otimes I_p), \tag{4.4}$$

where

$$\hat{C}_W = \sum_{i \in W} \sum_{j \notin W} (\ell_i - \ell_j)^{-1} \hat{P}_i \otimes \hat{P}_j', \tag{4.5}$$

and ℓ_i are the eigenvalues of M_n.

The test statistic proposed by Tyler is

$$T_n(A) = a_n^2 \{\mathrm{vec}[(I_p - \hat{P}_W)A]\}' [\hat{\psi}_0(A)]^+ \mathrm{vec}[(I_p - \hat{P}_W)A]. \tag{4.6}$$

Under H_0, T_n is asymptotically distributed as chi square with $r(p-m)$ degrees of freedom. Moreover, $T_n(A)$ is invariant under the transformation $A \to AB$ for any nonsingular $r \times r$-matrix B.

The problem of testing H_0^* can be approached as follows: Let B denote a fixed $p \times (p-r)$-matrix whose columns are orthogonal to those of A, i.e. $A'B = 0$.

Then H_0^* can be rephrased as (see also Theorem 1 (iii) of this paper)

H_0: The columns of $\underset{\sim}{B}$ lie in the subspace generated by the set of eigenvectors or $\underset{\sim}{M}'$ associated with the p-m roots λ_i for $i \notin w$.

4.3. Derivation of the test of redundancy

Our hypothesis $H_v(p,q)$ of simultaneous redundancy of p-q variables for the m eigenvectors $\underset{\sim}{\beta}_i$ $(i \in v)$ of $\underset{\sim}{\Sigma}_1^{-1}\,\underset{\sim}{\Sigma}_2$ can now be formulated in the form H_0^* of Tyler's approach by putting $\underset{\sim}{A} = \begin{pmatrix} \underset{\sim}{I}_q \\ \underset{\sim}{0} \end{pmatrix}$, where $\underset{\sim}{0}$ is a $(p-q)\times q$ - matrix of zeros. Putting $\underset{\sim}{B} = \begin{pmatrix} \underset{\sim}{0}' \\ \underset{\sim}{I}_{p-q} \end{pmatrix}$, this is equivalent to

H_0: The columns of $\underset{\sim}{B}$ lie in the subspace generated by the eigenvectors $\underset{\sim}{\alpha}_i$ $(i \notin v)$ of $\underset{\sim}{M}' = \underset{\sim}{\Sigma}_2\,\underset{\sim}{\Sigma}_1^{-1}$.

For convenience we will from now on write w for the complement of v, that is, w contains the p-m indices not in v. We are now going to derive the test for H_0 in terms of the eigenvectors $\underset{\sim}{a}_i$ of $\underset{\sim}{S}_2\,\underset{\sim}{S}_1^{-1}$ and then relate the resulting test statistic to the eigenvectors $\underset{\sim}{b}_i$ of $\underset{\sim}{S}_1^{-1}\,\underset{\sim}{S}_2$.

Assume now that $\underset{\sim}{S}_1$ and $\underset{\sim}{S}_2$ are independent sample covariance matrices from normal samples of size $n_1 + 1$ and $n_2 + 1$, i.e.

$$n_i \underset{\sim}{S}_i \sim W_p(n_i, \underset{\sim}{\Sigma}_i) \qquad (i = 1,2)\ . \tag{4.7}$$

Then, for $n_i \to \infty$, $\sqrt{n_i}\,\underset{\sim}{S}_i$ converges in distribution to a random matrix with mean $\underset{\sim}{\Sigma}_i$. The asymptotic covariance matrix of vec $(\sqrt{n_i}\underset{\sim}{S}_i)$ is (Muirhead 1982, p. 113)

$$(\underset{\sim}{I}_{p^2} + \underset{\sim}{I}_{(p,p)})\ (\underset{\sim}{\Sigma}_i \otimes \underset{\sim}{\Sigma}_i) \qquad (i = 1,2). \tag{4.8}$$

Put $n = n_1 + n_2$ and suppose that n_1 and n_2 go to infinity such that the limits

$$k_i = \lim_{n_i \to \infty} n/n_i \qquad (i = 1,2) \tag{4.9}$$

are bounded away from 0 and ∞ . By expanding $\underset{\sim}{S}_2\underset{\sim}{S}_1^{-1}$ in a Taylor series about $\underset{\sim}{M} = \underset{\sim}{\Sigma}_2\underset{\sim}{\Sigma}_1^{-1}$ we get the approximation

$$\sqrt{n}\,(\underset{\sim}{S}_2\underset{\sim}{S}_1^{-1} - \underset{\sim}{M}) \sim \sqrt{n}\,[(\underset{\sim}{S}_2 - \underset{\sim}{\Sigma}_2)\,\underset{\sim}{\Sigma}_1^{-1} - \underset{\sim}{M}(\underset{\sim}{S}_1 - \underset{\sim}{\Sigma}_1)\,\underset{\sim}{\Sigma}_1^{-1})$$

$$= \sqrt{k_2 n_2}\;\underset{\sim}{S}_2\,\underset{\sim}{\Sigma}_1^{-1} - \sqrt{k_1 n_1}\;\underset{\sim}{M}\underset{\sim}{S}_1\,\underset{\sim}{\Sigma}_1^{-1} \quad . \tag{4.10}$$

As n goes to infinity, this converges in distribution to a normal matrix

$$\underset{\sim}{N} = \sqrt{k_2}\;\underset{\sim}{N}_2 - \sqrt{k_1}\;\underset{\sim}{N}_1 \tag{4.11}$$

with mean zero, where $\underset{\sim}{N}_1$ and $\underset{\sim}{N}_2$ are the asymptotic distributions of $\sqrt{n_1}\;\underset{\sim}{M}\underset{\sim}{S}_1\,\underset{\sim}{\Sigma}_1^{-1}$ and $\sqrt{n_2}\;\underset{\sim}{S}_2\,\underset{\sim}{\Sigma}_1^{-1}$, respectively. The covariance matrix of $\text{vec}(\underset{\sim}{N})$ is

$$\underset{\sim}{\Psi} = k_1\,\text{cov}(\text{vec}(\underset{\sim}{N}_1)) + k_2\,\text{cov}(\text{vec}(\underset{\sim}{N}_2))$$

$$= k_1\,\underset{\sim}{\Sigma}_1^{-1} \otimes \underset{\sim}{M}\underset{\sim}{\Sigma}_2 + k_2\underset{\sim}{\Sigma}_1^{-1}\underset{\sim}{M} \otimes \underset{\sim}{\Sigma}_2$$

$$+ k_1(\underset{\sim}{\Sigma}_1^{-1} \otimes \underset{\sim}{M})\underset{\sim}{I}_{(p,p)}\,(\underset{\sim}{I}_p \otimes \underset{\sim}{\Sigma}_2)$$

$$+ k_2(\underset{\sim}{\Sigma}_1^{-1} \otimes I_p)\underset{\sim}{I}_{(p,p)}(\underset{\sim}{M} \otimes \underset{\sim}{\Sigma}_2). \tag{4.12}$$

Let the eigenvectors $\underset{\sim}{\alpha}_i$ of $\underset{\sim}{M}$ be normalized such that $\underset{\sim}{\alpha}_i'\,\underset{\sim}{\Sigma}_2^{-1}\,\underset{\sim}{\alpha}_i = 1$ $(i = 1,\ldots,p)$. Noting that $\underset{\sim}{M}$ is symmetric in the metric of $\underset{\sim}{\Sigma}_2^{-1}$, the eigenprojections are

$$\underset{\sim}{P}_i = \underset{\sim}{\alpha}_i\underset{\sim}{\alpha}_i'\,\underset{\sim}{\Sigma}_2^{-1} \qquad (i = 1,\ldots,p) \quad . \tag{4.13}$$

Using the orthogonality of the $\underset{\sim}{P}_i$, we get

$$\underset{\sim}{C}_w'(\underset{\sim}{\Sigma}_1^{-1} \otimes \underset{\sim}{M}\,\underset{\sim}{\Sigma}_2)\underset{\sim}{C}_w = \sum_{i\in w}\sum_{j\in v} \frac{\lambda_i\lambda_j}{(\lambda_i-\lambda_j)^2}\,\underset{\sim}{\Sigma}_2^{-1}\,\underset{\sim}{P}_i \otimes \underset{\sim}{\Sigma}_2\,\underset{\sim}{P}_j' \qquad (4.14)$$

and

$$\underset{\sim}{C}_w'(\underset{\sim}{\Sigma}_1^{-1}\underset{\sim}{M}\otimes\underset{\sim}{\Sigma}_2)\underset{\sim}{C}_w = \sum_{i\in w}\sum_{j\in v} \frac{\lambda_i^2}{(\lambda_i-\lambda_j)^2}\,\underset{\sim}{\Sigma}_2^{-1}\,\underset{\sim}{P}_i \otimes \underset{\sim}{\Sigma}_2\underset{\sim}{P}_j' \qquad (4.15)$$

and therefore

$$\underset{\sim}{\psi}_0(\underset{\sim}{B}) = \sum_{i\in w}\sum_{j\in v} \frac{k_1\lambda_i\lambda_j + k_2\lambda_i^2}{(\lambda_i-\lambda_j)^2}\,\underset{\sim}{B}\underset{\sim}{\Sigma}_2^{-1}\,\underset{\sim}{P}_i\underset{\sim}{B} \otimes \underset{\sim}{\Sigma}_2\underset{\sim}{P}_j'$$

$$= (\underset{\sim}{I}_{p-q} \otimes \underset{\sim}{\Sigma}_2)\,(\sum_{j\in v}\underset{\sim}{G}_j \otimes \underset{\sim}{P}_j') \quad , \qquad (4.16)$$

where

$$\underset{\sim}{G}_j = \underset{\sim}{B}\,\underset{\sim}{\Sigma}_2^{-1}\Big(\sum_{i\in w} \frac{k_1\lambda_i\lambda_j + k_2\lambda_i^2}{(\lambda_i-\lambda_j)^2}\,\underset{\sim}{\alpha}_i\underset{\sim}{\alpha}_i'\Big)\,\underset{\sim}{\Sigma}_2^{-1}\underset{\sim}{B} \quad . \qquad (4.17)$$

$\underset{\sim}{G}_j$ has rank p-q. (Note that the last two terms of (4.12) vanish thanks to the commutation property of $\underset{\sim}{I}_{(p_1p)}$). Using the fact that $\underset{\sim}{P}_i^+ = \underset{\sim}{P}_i$ (i=1,...,p), we get therefore

$$[\underset{\sim}{\psi}_0(\underset{\sim}{B})]^+ = \sum_{j\in v}\underset{\sim}{G}_j^{-1} \otimes \underset{\sim}{\Sigma}_2^{-1}\underset{\sim}{P}_j' \quad . \qquad (4.18)$$

Replacing $\underset{\sim}{\Sigma}_i$, $\underset{\sim}{\alpha}_j$ and λ_j by the corresponding sample quantities $\underset{\sim}{S}_i$, $\underset{\sim}{a}_j$ and $\underset{\sim}{\ell}_j$, we get a consistent estimate $\hat{\underset{\sim}{\psi}}_0(B)$ and its Moore-Penrose inverse

$$[\hat{\underset{\sim}{\psi}}_0(B)]^+ = \sum_{j\in v}\hat{\underset{\sim}{G}}_j^{-1} \otimes \underset{\sim}{S}_2^{-1}\hat{\underset{\sim}{P}}_j' \qquad (4.19)$$

Using lemma 2.2.3(iii) in Muirhead (1982) and the fact that $\hat{\underset{\sim}{P}}_w = \sum_{i\in w}\underset{\sim}{a}_i\underset{\sim}{a}_i'\underset{\sim}{S}_2^{-1} = \underset{\sim}{I}_p - \sum_{j\in v}\underset{\sim}{a}_j\underset{\sim}{a}_j'\underset{\sim}{S}_2^{-1} = \underset{\sim}{I}_p - \hat{\underset{\sim}{P}}_v$ yields

$$T_n(\underset{\sim}{B}) = n[\mathrm{vec}(\hat{\underset{\sim}{P}}_{v}\hat{\underset{\sim}{B}})]' \; [\hat{\underset{\sim}{\psi}}_0(\underset{\sim}{B})]^{+} \; \mathrm{vec}(\hat{\underset{\sim}{P}}_{v}\underset{\sim}{B})$$

$$= n \sum_{j\in v} \mathrm{trace}\; (\hat{\underset{\sim}{G}}_j^{-1}\underset{\sim}{B}'\hat{\underset{\sim}{P}}_{v}'\hat{\underset{\sim}{S}}_2^{-1}\hat{\underset{\sim}{P}}_{v}\underset{\sim}{B})$$

$$= n \sum_{j\in v} \mathrm{trace}\; (\hat{\underset{\sim}{G}}_j^{-1}\underset{\sim}{B}'\underset{\sim}{S}_2^{-1}\hat{\underset{\sim}{P}}_{j}\underset{\sim}{B}) \qquad (4.20)$$

where the last equality follows from $\hat{\underset{\sim}{P}}_{v}'\underset{\sim}{S}_2^{-1}\hat{\underset{\sim}{P}}_{j}'\hat{\underset{\sim}{P}}_{v} = \underset{\sim}{S}_2^{-1}\hat{\underset{\sim}{P}}_{j}'$ for $j\in v$. Noting that $\underset{\sim}{S}_2^{-1}\hat{\underset{\sim}{P}}_{j} = \underset{\sim}{S}_2^{-1}\underset{\sim}{a}_j\underset{\sim}{a}_j'\underset{\sim}{S}_2$ $(j = 1,\ldots,p)$, we see from Theorem 1 that

$$\underset{\sim}{S}_2^{-1}\underset{\sim}{a}_j = \ell_j^{-1/2}\underset{\sim}{b}_j \qquad (j = 1,\ldots,p) \quad , \qquad (4.21)$$

where $\underset{\sim}{b}_j$ is the j-th eigenvector of $\underset{\sim}{S}_1^{-1}\underset{\sim}{S}_2$, normalized such that $\underset{\sim}{b}_j'\underset{\sim}{S}_1\underset{\sim}{b}_j = 1$. Thus we get

$$T_n(\underset{\sim}{B}) = n \sum_{j\in v} \underset{\sim}{b}_j'\underset{\sim}{B}[\underset{\sim}{B}'\Big(\sum_{i\in w} \frac{k_1\ell_j^2 + k_2\ell_i\ell_j}{(\ell_i - \ell_j)^2}\; \underset{\sim}{b}_i\underset{\sim}{b}_i'\Big)\underset{\sim}{B}]^{-1}\underset{\sim}{B}'\underset{\sim}{b}_j \qquad (4.22)$$

We are now going to use the special structure of B to simplify (4.22), and, to mark the fact that $T_n(B)$ is a statistic for testing the Redundancy of p-q variables for the eigenvectors $\underset{\sim}{\beta}_i (i\in v)$, we will call it $R_v(p,q)$ from now on. Partition $\underset{\sim}{b}_i$ as

$$\underset{\sim}{b}_i = \begin{pmatrix} \underset{\sim}{b}_{i1} \\ \underset{\sim}{b}_{i2} \end{pmatrix} \qquad (4.23)$$

in q and p-q components, then

$$R_v(p,q) = n \sum_{j\in v} \underset{\sim}{b}_{j2}' \quad \Big(\sum_{i\in w} \frac{k_1\ell_j^2 + k_2\ell_i\ell_j}{(\ell_i - \ell_j)^2}\; \underset{\sim}{b}_{i2}\underset{\sim}{b}_{i2}'\Big)^{-1}\; \underset{\sim}{b}_{j2}. \qquad (4.24)$$

$R_v(p,q)$ is asymptotically distributed as chi square with m(p-q) degrees of freedom under H_0.

If the redundancy of only one variable is to be tested, (4.24) simplifies

to

$$R_v(p,1) = n \sum_{j \in v} [b_{jp}^2 \Big/ \sum_{i \in w} \frac{k_1 \ell_j^2 + k_2 \ell_i \ell_j}{(\ell_i - \ell_j)^2} \, b_{ip}^2] \quad , \tag{4.25}$$

where b_{hp} is the p-th (last) coefficient of b_h. Furthermore, if only one eigenvector, say the first one, is under consideration, we get

$$R_1(p,1) = n \, b_{1p}^2 \Big/ \sum_{i=2}^{p} \frac{k_1 \ell_1^2 + k_2 \ell_1 \ell_i}{(\ell_1 - \ell_i)^2} \, b_{ip}^2 \tag{4.26}$$

with one degree of freedom. This can be used to test the significance of the partial statistics defined in section 3.

Note that (4.24) thru (4.26) depend on the correct normalization of the eigenvectors $\underset{\sim}{b}_i$. They must be normalized such that $\underset{\sim}{b}_i' \underset{\sim}{S}_1 \underset{\sim}{b}_i = c$ $(i = 1,\ldots,p)$ for some $c > 0$. The most convenient way of doing this is of course to use the standard convention $\underset{\sim}{b}_i' \underset{\sim}{S}_1 \underset{\sim}{b}_i = 1$.

4.4. Remarks and Applications

1. Non-normality. It is well known that tests on variances depend, even asymptotically, much more on the usual normality assumptions than tests on means. Considerable attention has been given to the case of samples from elliptical distributions (Muirhead 1982, p. 32-40, 329-331, 352, and references therein). Tyler (1981) has specialized his test statistic T_n to hypotheses in principal component analysis and canonical correlation analysis, using samples from an elliptical distribution with finite fourth moments. Analogously, the test statistic $R_v(p,q)$ can be generalized to the elliptical situation: Let κ_1 and κ_2 denote the kurtosis parameters of two elliptical populations, defined such that $3\kappa_i$ is the kurtosis of any marginal distribution (Muirhead 1982, p. 41).

It can be shown that the only change in $R_v(p,q)$ is that the constants $k_i = n/n_i$ have to be replaced by $k_i^* = k_i(1+\kappa_i)$. If $\kappa_1 = \kappa_2 = \kappa$ (say), then the correct statistic can be written as

$$R_v^*(p,q) = (1 + \kappa)^{-1} R_v(p,q) \ . \tag{4.27}$$

The asymptotic null distribution is still chi square with $m(p-q)$ degrees of freedom. In practice, κ_1 and κ_2 can be replaced by consistent estimators. This does not affect the validity of the asymptotic chi square approximation, but it does affect the rate of convergence. Under the multivariate normal model we have $\kappa = 0$, and no correction is necessary.

Formula (4.27) shows that testing H_0 from an elliptical sample of size n and kurtosis parameter κ is essentially the same as testing H_0 from a normal sample of size $n(1 + \kappa)^{-1}$. If κ is negative (e.g. the uniform distribution within an ellipsoid has parameter $\kappa = -.6$), then the use of the normal theory procedure leads to a conservative test. In practice, however, one is rather concerned about heavy-tailed distributions, which have positive values of κ. In such cases, the normal theory test rejects H_0 too readily, thus exceeding the nominal α - level.

While the above correction for kurtosis works fine in theory, its application has some flaws. First, consistent estimators of κ can be defined (Srivastava and Carter 1983, p. 66), but their convergence to κ may be very slow, especially if κ is large. Second, the elliptical model might not hold, and skewness might further affect the correctness of the asymptotic approximation.

If the assumption of elliptical populations is unreasonable, a feasible alternative approach might be to sample from distributions with the same covariance matrices $\underset{\sim}{\Sigma}_1$ and $\underset{\sim}{\Sigma}_2$, but which are normal or at least close to normal, so that the normal theory procedure can be applied. This can be achieved for

instance by replacing pairs of observations by their sum or difference times $1/\sqrt{2}$, thus reducing the sample size by a factor 2. By the multivariate central limit theorem, the new data will be closer to normality, and the sample covariance matrices closer to wishartness. Taking the sum or difference of two observations reduces κ by a factor 2.

More generally, integers $r_i > 1$ $(i = 1,2)$ can be chosen and groups of r_i observations can be used to get $n_i^* = n_i/r_i$ new observations which are closer to normality than the original data. Instead of acturally carrying out the computations for the reduced samples, we can also approximate the result by simply replacing n_i by $n_i^* = n_i/r_i$ and using $n^* = n_1^* + n_2^*$, $k_i^* = n^*/n_i^*$ to replace n and k_i in (4.24). If $r_1 = r_2 = r$, this amounts to the same as dividing $R_V(p,q)$ by a factor r. Thus the effect is very similar to the effect of the correction for kurtosis. This procedure results in a loss of power if the underlying distributions are actually normal or elliptical with small κ. However, it has two advantages: First, if the distributions are not elliptical, inference on eigenvectors can still be done. Second, taking differences (or as an approximation, dividing $R_V(p,q)$ by 2) removes any effect of skewness!

Similar methods can of course be applied to the overall tests for identity of covariance matrices. For these tests, some work on the effect of non-normality has been done (Davis 1982; Pillai and Sudjana 1975).

If one is completely unwilling to rely on parametric theory, it would probably be worthwhile to try a bootstrap method (Efron 1982).

2. Multiple eigenvalues. The validity of the asymptotic distribution of $R_v(p,q)$ depends on the assumption that $d(v,w) = \min_{\substack{j\in v \\ i\in w}} |\lambda_i - \lambda_j| > 0$, and, as Tyler (1981, p. 732) remarks, the sample size n necessary to insure that the chi square approximation is "good" is in general inversely related to the above quantity. In practice, of course, we never know whether d(v,w) is zero or not, but we may notice that some eigenvalues of $\underset{\sim}{S}_1^{-1}\underset{\sim}{S}_2$ are very close. Suppose for simplicity, that we wish to analyze the largest eigenvector $\underset{\sim}{b}_1$, but the roots ℓ_1 and ℓ_2 are close. Then it might be reasonable to test the redundancy of some variables for $\underset{\sim}{b}_1$ and $\underset{\sim}{b}_2$ simultaneously. Otherwise, as can be seen from (4.24), the matrix $\hat{\underset{\sim}{G}}_2$ will be blown up, and the test statistic will tend to be small. This reflects, of course, the fact that if λ_1 and λ_2 are close, the associated sample eigenvectors $\underset{\sim}{b}_1$ and $\underset{\sim}{b}_2$ have a relatively large variablility.

3. Properties of $R_v(p,q)$. It is obvious from (4.24) that changing the signs of eigenvectors $\underset{\sim}{b}_i$ does not affect $R_v(p,q)$. It is less obvious, but can easily by shown, that $R_v(p,q)$ does not depend on the numbering of the groups, that is, whether we analyze $\underset{\sim}{\Sigma}_1^{-1}\underset{\sim}{\Sigma}_2$ or $\underset{\sim}{\Sigma}_2^{-1}\underset{\sim}{\Sigma}_1$. Note that the eigenvectors of these two matrices are identical (up to scaling constants), while the eigenvalues are the inverses.

4. Estimation of covariance matrices under H_0. If q variables are found to be sufficient for the eigenvectors $\underset{\sim}{\beta}_i$ with $i\in v$, then Theorem 2 tells us that these $\underset{\sim}{\beta}_i$ can be estimated from the reduced set of q variables. However, we might also wish to estimate $\underset{\sim}{\Sigma}_1$ and $\underset{\sim}{\Sigma}_2$ subject to the constraint that some eigenvectors do not depend on some variables. Although a maximum likelihood solution of this problem has been found for p = 2 in the Wishart case (Flury 1983a), a general

solution seems rather difficult to obtain.

5. Alternative approach. In view of Theorem 3, an alternative approach to the testing problem might be based on the joint distribution of $\ell_1^{(p)}$ and $\ell_1^{(q)}$. However, the noncentral distribution of the latent roots of $S_{\sim 1}^{-1} S_{\sim 2}$ is fairly complicated (Pillai and Sugiyama 1969), and this approach seems intractable at this time.

6. Application. In the bank note example, let us test whether the largest eigenvector depends on the variables LEFT and RIGHT. The largest eigenvalue ℓ_1 seems far enough from the second root to justify the analysis of $x_{\sim 1}$ alone. The test statistic (4.24) is $R_1(6,2) = 5.62)$ with 2 degrees of freedom. Assuming that the multivariate normal model holds approximately, we can accept H_0 at an α - level of 5 percent. The linear combination with maximum ratio of variances, estimated from the four remaining variables, is 1.34 LENGTH - 2.05 BOTTOM - 1.35 TOP - 1.28 DIAGONAL.

For those who cannot fish without leering at test statistics, we give here a list of all $R_6(6,1)$ - values to test the redundancy of single variables in the linear combination $Y_{min}^{(6)}$ used in section 3: LENGTH 1.41, LEFT 3.32, RIGHT .50, BOTTOM 41.73, TOP 38.89, DIAGONAL 12.10. These chi squares, based on one degree of freedom each, are obviously in good accordance with the descriptive partial statistics.

5. THE ONE SAMPLE CASE

As already mentioned in the introduction, it might also be interesting in some cases to analyze the extreme characteristic vectors of $\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{\Sigma}$, where $\underset{\sim}{\Sigma}_0$ is a fixed hypothetical p.d.s. covariance matrix, and $\underset{\sim}{\Sigma}$ is the (unknown) covariance matrix of a population from which a sample is taken.

If we are interested in zero coefficients, we can use the fact that there exists a unique upper triangular matrix $\underset{\sim}{C}$ such that $\underset{\sim}{C}'\underset{\sim}{\Sigma}_0\underset{\sim}{C} = \underset{\sim}{I}_p$ (Choleski factorization, see e.g. Schwarz, Rutishauser and Stiefel 1973, p. 27). It is easy to show that $\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{\Sigma}$ and $\underset{\sim}{\psi} = \underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C}$ have the same eigenvalues and that every eigenvector $\underset{\sim}{\beta}_i$ of $\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{\Sigma}$ corresponds to an eigenvector $\underset{\sim}{\gamma}_i = \underset{\sim}{C}^{-1}\underset{\sim}{\beta}_i$ of $\underset{\sim}{\psi}$. Let $\underset{\sim}{\beta}_i = (\underset{\sim}{\beta}_{i1}', \underset{\sim}{\beta}_{i2}')'$ and $\underset{\sim}{\gamma}_i = (\underset{\sim}{\gamma}_{i1}', \underset{\sim}{\gamma}_{i2}')'$ be partitioned in q and p-q coefficients. Since $\underset{\sim}{C}$ and $\underset{\sim}{C}^{-1}$ are upper triangular, it follows that $\underset{\sim}{\gamma}_{i2} = \underset{\sim}{0}$ if and only if $\underset{\sim}{\beta}_{i2} = \underset{\sim}{0}$. For the purpose of finding zero coefficients in the eigenvectors of $\underset{\sim}{\Sigma}_0^{-1}\underset{\sim}{\Sigma}$ we can therefore look for zero coefficients in the eigenvectors of $\underset{\sim}{\psi}$ (assuming of course that the variables are properly ordered such as to put the "zero candidates" in the last positions). The results of this section will therefore be given in terms of a p-dimensional random vector $\underset{\sim}{Y}$ and its p.d.s. covariance matrix $\underset{\sim}{\psi}$. It will tacitly be assumed in the sequel that all eigenvectors of p.d.s. matrices are normalized in the usual way.

Let us first apply the results of section 2 to the one matrix case. By putting $\underset{\sim}{S} = \underset{\sim}{I}_p$ in theorem 1 we get the familiar spectral decomposition theorem for a p.d.s. matrix $\underset{\sim}{T}$. Theorem 2 can be simplified due to $\underset{\sim}{S}_{11} = \underset{\sim}{I}_q$ and $\underset{\sim}{S}_{12} = \underset{\sim}{0}$. However, theorem 2 has two alternative versions (one of them to be given in rectangular brackets) which are somehow more "statistical". Note that by "does not depend on $\underset{\sim}{Y}_2$" we mean that the coefficients associated with the random variables in $\underset{\sim}{Y}_2$ are zero.

Theorem 2*: Let the p-dimensional normal random vector $\underset{\sim}{Y} = (\underset{\sim}{Y}_1', \underset{\sim}{Y}_2')$ be partitioned in q and p-q components, and let $\underset{\sim}{\psi} = \begin{pmatrix} \underset{\sim}{\psi}_{11} & \underset{\sim}{\psi}_{12} \\ \underset{\sim}{\psi}_{21} & \underset{\sim}{\psi}_{22} \end{pmatrix}$ denote its p.d.s. covariance matrix. Let $\underset{\sim}{Y}_{1.2}$ denote a random vector having the conditional distribution of $\underset{\sim}{Y}_1$ given $\underset{\sim}{Y}_2$.

(i) Let $U = \underset{\sim}{b}'\underset{\sim}{Y}$ denote a principal component of $\underset{\sim}{Y}$. If U does not depend on $\underset{\sim}{Y}_2$, then U is a principal component of $\underset{\sim}{Y}_1$, and the multiple correlation between U and $\underset{\sim}{Y}_2$ is zero [and, simultaneously, U is a principal component of $\underset{\sim}{Y}_{1.2}$].

(ii) If U* is a principal component of $\underset{\sim}{Y}_1$ and the multiple correlation between U* and $\underset{\sim}{Y}_2$ is zero [and U* is simultaneously a principal component of $\underset{\sim}{Y}_{1.2}$], then U* is a principal component of $\underset{\sim}{Y}$.

The "translations" of theorems 2 (iii) and 3 into statistical terms are analogous and need not be given here. All proofs are straightforward. In the proof of 2* (ii) note that $\underset{\sim}{\psi}_{12}\underset{\sim}{\psi}_{22}^{-1}\underset{\sim}{\psi}_{21}\ \underset{\sim}{b}^* = \underset{\sim}{0}$ implies $\underset{\sim}{\psi}_{21}\underset{\sim}{b}^* = \underset{\sim}{0}$ since $\underset{\sim}{\psi}_{22}$ is p.d.s. Theorem 2* is potentially useful for estimation purposes, but we are not going to pursue this line in the present paper. The one-matrix analog of theorem 4 follows by putting $S_{22} = 1$.

The method of section 3 needs only little (obvious) modification to apply to the one sample case. As in the two sample case, the largest and smallest roots' criterion (or some other test for H_0: $\underset{\sim}{\Sigma} = \underset{\sim}{\Sigma}_0$) should be applied to the data prior to any analysis of the extreme eigenvectors. The null distribution of the extreme roots of Wishart matrices has been tabulated by Harumara and Thompson (1968) and by Pillai and Chang (1970).

An asymptotic chi square statistic for testing $H_v(p,q)$ in the one sample case can be derived from the results of section 4 as follows: write the statistic (4.24) as

$$R_v(p,q) = n_2 \sum_{j \in v} \underset{\sim}{b}_{j2}' \Big(\sum_{i \in w} \frac{n_2 \ell_j^2 / n_1 + \ell_i \ell_j}{(\ell_i - \ell_j)^2} \underset{\sim}{b}_{i2} \underset{\sim}{b}_{i2}' \Big)^{-1} \underset{\sim}{b}_{j2} \qquad (5.1)$$

and let n_1 go to infinity, holding n_2 constant. Then, writing n instead of n_2, we get the one-sample statistic

$$R_v^*(p,q) = n \sum_{j \in w} \underset{\sim}{b}_{j2}' \Big(\sum_{i \in w} \frac{\ell_i \ell_j}{(\ell_i - \ell_j)^2} \underset{\sim}{b}_{i2} \underset{\sim}{b}_{i2}' \Big)^{-1} \underset{\sim}{b}_{j2}. \qquad (5.2)$$

Here, n is the number of degrees of freedom of a Wishart matrix $\underset{\sim}{S}$, and $\underset{\sim}{b}_h = (\underset{\sim}{b}_{h1}', \underset{\sim}{b}_{h2}')'$ are the eigenvectors of $\underset{\sim}{\Sigma}_0^{-1} \underset{\sim}{S}$, normalized such that $\underset{\sim}{b}_h' \underset{\sim}{\Sigma}_0 \underset{\sim}{b}_h = 1$, ℓ_h are the associated eigenvalues, and $\underset{\sim}{\Sigma}_0$ is the hypothetical covariance matrix. If we are working with the transformed matrix $\underset{\sim}{T} = \underset{\sim}{C}' \underset{\sim}{S} \underset{\sim}{C}$ instead (where $\underset{\sim}{C}' \underset{\sim}{\Sigma}_0 \underset{\sim}{C} = \underset{\sim}{I}_p$, C upper triangular), then it is easy to show that the same statistic (5.2) can be used, where $\underset{\sim}{b}_h$ denotes now the h-th eigenvector of $\underset{\sim}{T}$, normalized such that $\underset{\sim}{b}_h' \underset{\sim}{b}_h = 1$. (Remember, however, that $\underset{\sim}{C}$ depends on the order of the variables - the ones to be tested for redundancy must be moved to the last p-q positions before switching from $\underset{\sim}{S}$ to $\underset{\sim}{T}$).

The asymptotic ($n \to \infty$) null distribution of $R_v^*(p,q)$ is again chi square with m(p-q) degrees of freedom, where m is the number of elements in v. This result can also be established from Tyler's application of his asymptotic theory to principal component analysis (Tyler 1981, formula 7.4). Still another way of proving this result is based on the asymptotic distribution of

the eigenvectors of a Wishart matrix $\underset{\sim}{S}$ (Anderson 1963, p. 130), using Wald's method (Wald 1943, Moore 1977) and the fact that the eigenvalues of $\underset{\sim}{S}$ are consistent estimates of the population eigenvalues.

For recent developments in the one sample case see Tyler (1983).

6. CONCLUSIONS

The comparison of covariance matrices has been a stepchild of applied multivariate analysis. Although most textbooks on multivariate methods give considerable attention to the analysis of multivariate structure in the one sample case (Principal component analysis, factor analysis), the comparison of the multivariate variability of several groups has most often been treated on the crude level of equality versus inequality. I hope to have shown in this paper that comparing two covariance matrices by analyzing the linear combinations with extreme ratios of variances is a very interesting method itself, giving much more insight into differences between multivariate scatters than just an overall test of equality.

An interesting related approach (in the one sample problem) has recently been proposed by Krzanowski (1984). While the descriptive method in section 3 of this paper is based on the idea that changes in a coefficient of a linear combination are reflected by changes in the associated ratio of variances, Krzanowski's approach is just reverse: the key idea is to compute the maximum changes in the coefficients of a principal component that are associated with a given small change in the variance of the component. This method could probably be generalized to the two sample case.

7. REFERENCES

ANDERSON, T. W. (1963), "Asymptotic theory for principal component analysis," Annals of Mathematical Statistics, 34, 122-148.

——— (1970), "Estimation of covariance matrices which are linear combinations or whose inverse are linear combinations of given matrices," in Essays in Probability and Statistics, ed. R. C. Bose et al, Chapel Hill: University of North Carolina Press, 1-24.

BASILEVSKY, A. (1983), Applied Matrix Algebra in the Statistical Sciences, New York: North Holland.

BIBBY, J.M. (1980), "Some effects of rounding optimal estimates," Sankhya B, 42, 165-178.

BREIMAN, L., and FREEDMAN, D. (1983), "How many variables should be entered in a regression equation?," Journal of the American Statistical Association, 78, 131-136.

DAVIS, A.W. (1982), "On the effects of moderate multivariate non-normality on Roy's largest root test," Journal of the American Statistical Association, 77, 896-900.

DRAPER, N., and SMITH, H. (1981), Applied Regression Analysis, 2nd Edition, New York: Wiley.

EFRON, B. (1982), The Jackknife, the Bootstrap and Other Resampling Plans, Philadelphia: Society for Industrial and Applied Mathematics.

FLURY, B.N. (1983a), "Maximum likelihood estimation of some patterned 2 x 2 covariance matrices, "Technical Report #83-9, Dept. of Statistics, Purdue Univeristy.

——— (1983 b), "Some relations between the comparison of covariance matrices and principal component analysis," Computational Statistics and Data Analysis, 1, 97-109.

——— (1983 c), "A generalization of principal component analysis to k groups," Technical Report #83-14, Dept. of Statistics, Purdue University.

FLURY, B. and RIEDWYL, H. (1983), Angewandte Multivariate Statistik, Stuttgart & New York: Gustav Fischer.

FREEDMAN, D.A. (1983), "A note on screening regression equations," The American Statistician, 37, 152-155.

HANUMARA, R.C., and THOMPSON, W.A. (1968), "Percentage points of the extreme roots of a Wishart matrix," Biometrika, 55, 505-512.

HECK, D.L. (1960), "Charts of some upper percentage points of the distribution of the largest characteristic root," Annals of Mathematical Statistics, 31, 625-642.

KRES, H. (1975), Statistische Tafeln zur Multivariaten Analysis, Berlin: Springer.

KRZANOWSKI, W.J. (1984), "Sensitivity of principal components," Journal of the Royal Statistical Society, to be published.

MAGNUS, J.R., and NEUDECKER, H. (1979), "The commutation matrix: some properties and applications," Annals of Statistics, 7, 381-394.

MARDIA, K.V., KENT, J.T. and BIBBY, J.M. (1979), Multivariate Analysis, London: Academic Press.

MOORE, D.S. (1977), "Generalized inverses, Wald's method, and the construction of chi-squared tests of fit," Journal of the American Statistical Association, 72, 131-137.

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

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

NERRING, E.D. (1970), Linear Algebra and Matrix Theory, 2nd Ed., New York: Wiley.

NEUDECKER, H. (1968), "The Kronecker matrix product and some of its applications," Statistica Neerlandica, 22, 69-82.

PILLAI, K.C.S. (1967), "Upper percentage points of the largest root of a matrix in multivariate analysis," Biometrika, 54, 189-194.

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.

PILLAI, K.C.S., and CHANG, T.C. (1970), "An approximation to the C.D.F. of the largest root of a covariance matrix," Annals of the Institute of Statistical Mathematics, Supplement b, 115-124.

PILLAI, K.C.S., and CHU, S.S. (1979), "Power comparison of two-sided tests of equality of two covariance matrices based on six criteria," Annals of the Institute of Statistical Mathematics, 31, Part A, 185-205.

PILLAI, K.C.S., and SUDJANA (1975), "Exact robustness studies of tests of two multivariate hypotheses based on four criteria and their distribution problems under violations," Annals of Statistics, 3, 617-636.

PILLAI, K.C.S., and SUGIYAMA, T. (1969), "Non-central distributions of the largest latent roots of three matrices in multivariate analysis," Annals of the Institute of Statistical Mathematics, 21, 321-327.

RAO, C.R. (1973), Linear Statistical Inference and its Application, 2nd Ed., New York: Wiley.

——— (1983), "Likelihood ratio tests for relationships between two covariance matrices," in Studies in Econometrics, Time Series, and Multivariate Statistics, ed. S. Karlin, T. Amemiya and L.A. Goodman, New York: Academic Press.

RAO, C.R., and MITRA, S.K. (1971), Generalized Inverse of Matrices and its Applications, New York: Wiley.

RENCHER, A.C., and PUN, F.C. (1980), "Inflation of R^2 in best subset regression," Technometrics, 22, 49-53.

ROY, S.N. (1957), Some Aspects of Multivariate Analysis, New York: Wiley.

SCHWARZ, H.R., RUTISHAUSER, H., and STIEFEL, E. (1973), Numerical Analysis of Symmetric Matrices, Englewood Cliffs, N.J.: Prentice-Hall.

SEARLE, S.R. (1982), Matrix Algebra Useful for Statistics, New York: Wiley.

SELVIN, H.C. and STUART, A. (1966), "Data-dredging procedures in survey analysis," The American Statistician, 20, no. 3, 20-23.

SRIVASTAVA, M.S. and CARTER, E.M. (1983), An Introduction to Applied Multivariate Statistics, New York: North-Holland.

SZATROWSKI, T.H. (1976), "Estimation and testing for block compound symmetry and other patterned covariance matrices with linear and non-linear structure," Technical Report #107, Dept. of Statistics, Stanford University.

TYLER, D.E. (1981), "Asymptotic inference for eigenvectors," Annals of Statistics, 9, 725-736.

——— (1983), "A class of asymptotic tests for principal component vectors," Annals of Statistics, 11, 1243-1250.

WALD, A. (1943), "Tests of statistical hypotheses concerning several parameters when the number of observations is large, "Transactions of the American Mathematical Society, 54, 425-482.

Table 1: Covariance matrices of real and forged Swiss bank notes

a) real notes ($n_1 = 100$)

$$\underset{\sim}{S}_1 = \begin{array}{c} \begin{array}{cccccc} \text{LENGTH} & \text{LEFT} & \text{RIGHT} & \text{BOTTOM} & \text{TOP} & \text{DIAGONAL} \end{array} \\ \begin{pmatrix} 0.1502 & 0.0580 & 0.0573 & 0.0571 & 0.0145 & 0.0055 \\ 0.0580 & 0.1326 & 0.0859 & 0.0567 & 0.0491 & -0.0431 \\ 0.0573 & 0.0859 & 0.1236 & 0.0582 & 0.0306 & -0.0238 \\ 0.0571 & 0.0567 & 0.0582 & 0.4132 & -0.2635 & -0.0002 \\ 0.0145 & 0.0491 & 0.0306 & -0.2635 & 0.4212 & -0.0753 \\ 0.0055 & -0.0431 & -0.0238 & -0.0002 & -0.0753 & 0.1998 \end{pmatrix} \end{array} \begin{array}{l} \text{LENGTH} \\ \text{LEFT} \\ \text{RIGHT} \\ \text{BOTTOM} \\ \text{TOP} \\ \text{DIAGONAL} \end{array}$$

---

b) forged notes ($n_2 = 100$)

$$\underset{\sim}{S}_2 = \begin{array}{c} \begin{array}{cccccc} \text{LENGTH} & \text{LEFT} & \text{RIGHT} & \text{BOTTOM} & \text{TOP} & \text{DIAGONAL} \end{array} \\ \begin{pmatrix} 0.1240 & 0.0315 & 0.0240 & -0.1006 & 0.0194 & 0.0116 \\ 0.0315 & 0.0650 & 0.0468 & -0.0240 & -0.0119 & -0.0050 \\ 0.0240 & 0.0468 & 0.0889 & -0.0186 & 0.0001 & 0.0342 \\ -0.1006 & -0.0240 & -0.0186 & 1.2813 & -0.4902 & 0.2385 \\ 0.0194 & -0.0119 & 0.0001 & -0.4902 & 0.4045 & -0.0221 \\ 0.0116 & -0.0050 & 0.0342 & 0.2358 & -0.0221 & 0.3112 \end{pmatrix} \end{array} \begin{array}{l} \text{LENGTH} \\ \text{LEFT} \\ \text{RIGHT} \\ \text{BOTTOM} \\ \text{TOP} \\ \text{DIAGONAL} \end{array}$$

---

c) eigenvalues and eigenvectors of $\underset{\sim}{S}_1^{-1}\underset{\sim}{S}_2$

<u>eigenvalues</u>:

F_{max} = 6.2225 1.6745 1.0516 0.9003 0.5455 0.2839 = F_{min}

<u>eigenvectors</u>:

0.9751	-0.0718	-1.4129	1.9840	-1.3421	-0.3961	LENGTH
0.7054	0.0426	1.0120	1.3528	3.3632	-1.1742	LEFT
0.4192	1.4190	1.9213	-1.6155	-2.5544	-0.3740	RIGHT
-2.2562	-0.4762	-0.3505	-0.0446	-0.2471	-0.5121	BOTTOM
-1.5528	0.4905	-1.3088	-0.7537	0.0319	-0.8418	TOP
-1.0667	1.9275	0.1204	0.5800	0.6345	0.5866	DIAGONAL