The Annals of Statistics
1983, Vol. 11, No. 4, 1243–1250

A CLASS OF ASYMPTOTIC TESTS FOR PRINCIPAL COMPONENT VECTORS

BY DAVID E. TYLER

Old Dominion University

In this paper, the hypothesis that a set of vectors lie in the subspace spanned by a prescribed subset of the principal component vectors for a normal population is considered. A class of invariant asymptotic tests based on the sample covariance matrix is derived. Tests in this class are shown to be consistent and their local power functions are given. The arguments used in deriving the class of tests are not heavily dependent on the assumption of normality nor on the use of the sample covariance matrix. The results are shown to generalize when the procedures are based on any affine-invariant M-estimate of scatter and when the population is elliptical.

1. Introduction and summary. Let $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_n$ be a sample of n independent p-dimensional Normal $(\boldsymbol{\mu}, \Sigma)$ random vectors with Σ nonsingular, and let S_n be the sample covariance matrix. The population and sample principal component roots are the eigenvalues of Σ and S_n respectively, say $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$ and $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p \geq 0$. The population and sample principal component vectors are the corresponding eigenvectors of Σ and S_n respectively. By studying the asymptotic distribution of the sample vectors, Anderson (1963) has shown under the assumption $\lambda_{i-1} \neq \lambda_i \neq \lambda_{i+1}$ that

$$n(\hat{\lambda}_i \mathbf{a}' S_n^{-1} \mathbf{a} + \hat{\lambda}_i^{-1} \mathbf{a}' S_n \mathbf{a} - 2) \rightarrow \chi^2_{p-1} \tag{1.1}$$

in distribution, whenever $\Sigma \mathbf{a} = \lambda_i \mathbf{a}$ and $\mathbf{a}'\mathbf{a} = 1$. This classic result is often used to test if a hypothesized vector is a principal component vector associated with the ith largest root.

In this paper, rather than considering the principal component vectors individually, asymptotic inferences for the subspace generated by a set of principal component vectors are studied. Let i and m be fixed positive integers with $i + m - 1 \leq p$, and assume throughout the paper that $\lambda_{i-1} < \lambda_i$ (if $i > 1$) and $\lambda_{i+m-1} < \lambda_{i+m}$ (if $i + m \leq p$). Interest is to be focused on the eigenspace spanned by the eigenvectors of Σ associated with the roots $\lambda_i, \cdots, \lambda_{i+m-1}$, or equivalently on the $p \times p$ orthogonal projection matrix P_0 which projects onto this eigenspace. In particular, for a $p \times r$ matrix A with $\text{rank}(A) = r \leq m$, the following null hypothesis is considered,

$$H_0\colon P_0 A = A. \tag{1.2}$$

This null hypothesis states that the columns of A lie in the eigenspace spanned by the eigenvectors of Σ associated with $\lambda_i, \cdots, \lambda_{i+m-1}$.

The problem of testing H_0 is treated by the author in an earlier paper, Tyler (1981). There, an asymptotic chi-squared test is given. In Section 3 of the present paper, an $r \times r$ matrix-valued statistic is derived and its asymptotic distribution is shown to be Wishart. A class of invariant asymptotic tests for H_0 based on the roots of this matrix-valued statistic is then proposed. Tests in this class are shown to be consistent, and their local power functions are given. The arguments used in deriving the results in Section 3 are not heavily dependent on the assumption that the population is multivariate normal nor on the use of S_n as an estimate of Σ. The results readily generalize to any affine-invariant

Received May 1980; revised June 1983.

AMS 1970 *subject classifications*. Primary, 62H15; secondary, 62H20, 62H25, 62E20.

Key words and phrases. Elliptical distributions, invariance, noncentral Wishart, robustness, spectral decomposition.

M-estimate of scatter when sampling from a population with an elliptical distribution (Section 4).

For a $p \times s$ matrix B with rank$(B) = s > m$, one may wish to test whether the eigenspace spanned by the eigenvectors of Σ associated with $\lambda_i, \cdots, \lambda_{i+m-1}$ lies in the space spanned by the columns of B. This testing problem is similar to the problem of testing H_0 and so is not discussed further, see Remark 3 in Tyler (1981).

For motivation, consider the following hypothetical example. A battery of six tests, $(y_1, y_2, \cdots, y_6)'$, are given to a group of subjects. The first three tests are considered measures of mathematical ability and the last three measures of verbal ability. Suppose that an analysis of the principal component roots indicates that most of the variability in the tests can be accounted for by the first two principal component variables. One may then wish to test whether the space spanned by the first two principal component vectors corresponds to the space spanned by the vectors $(1, 1, 1, 0, 0, 0)'$ and $(0, 0, 0, 1, 1, 1)'$. Alternatively, one may wish to test whether the space spanned by the first two principal component vectors measures verbal ability only. This can be tested by considering the hypothesis that the set of vectors $\{(1, 0, 0, 0, 0, 0)', (0, 1, 0, 0, 0, 0)', (0, 0, 1, 0, 0, 0)'\}$ lie in the space spanned by the last four principal component vectors.

In the above example, hypotheses on the individual vectors may also be of interest provided the corresponding roots λ_1 and λ_2 are "well spaced", see Remark 1 in Tyler (1981). Other motivations for comparing principal component spaces rather than solely comparing principal component vectors are discussed for the two populations case by Krzanowski (1979), wherein an illuminating example can be found.

In practice, principal components analysis is usually used as an exploratory procedure. As such, one may not have any a priori hypotheses concerning the principal component vectors, but rather use principal components analysis to discover interesting linear combinations of the original variables. The sample vectors, however, are often cumbersome and difficult to interpret. Thus, it is common practice to note that the sample vectors look "close" to some more parsimonious or scientifically meaningful set of linear combinations, or possibly to the principal component vectors of a known population. This concept of closeness can be made more rigorous by considering the p-value of an appropriate test of hypothesis for the vectors.

2. Sample principal component spaces. The asymptotic distribution of S_n as $n \to \infty$ when sampling from a normal population can be expressed in the following manner:

$$n^{1/2}\{\text{vec}(S_n - \Sigma)\} \to \text{Normal}\{\mathbf{0}, (I + K_{p,p})(\Sigma \otimes \Sigma)\} \tag{2.1}$$

in distribution. If B is a $b \times t$ matrix, then vec(B) is the bt-dimensional vector formed by stacking the columns of B, while if C is $c \times u$, then $B \otimes C$ is the $bc \times tu$ Kronecker product of B and C. The commutation matrix or permuted identity matrix is the $ab \times ab$ matrix $K_{a,b} = \sum_{i=1}^{a} \sum_{j=1}^{b} E_{ij} \otimes E'_{ij}$ where E_{ij} is an $a \times b$ matrix with a one in the (i, j) position and zeroes elsewhere. Algebraic properties involving the "vec" transformation, the Kronecker product and the commutation matrix have been extensively investigated by Magnus and Neudecker (1979). Three important properties are

$$\text{vec}(ABC) = (C' \otimes A)\text{vec}(B), \tag{2.2}$$

$$K_{a,b}\text{vec}(A_{a\times b}) = \text{vec}(A'), \tag{2.3}$$

$$K_{a,b}(A_{a\times c} \otimes B_{b\times d}) = (B \otimes A)K_{c,d}. \tag{2.4}$$

The spectral representation of Σ is $\Sigma = \sum_{\lambda \in \mathscr{L}} \lambda P_\lambda$, where $\mathscr{L}$ represents the set of distinct roots of Σ and P_λ is the unique orthogonal projection onto the space spanned by the eigenvectors of Σ associated with the eigenvalue λ. The matrix P_0 defined in Section 1 is thus $P_0 = \sum_{\lambda \in w} P_\lambda$ where w represents the set of distinct values from $\{\lambda_i, \cdots, \lambda_{i+m-1}\}$.

Corresponding, for $n > p$, let $\hat{w} = \{\hat{\lambda}_i, \cdots, \hat{\lambda}_{i+m-1}\}$ and $\hat{P}_0 = \sum_{\lambda\in\hat{w}}\hat{P}_\lambda$, where $\hat{P}_\lambda$ is obtained from $S_n = \sum_{\lambda\in\hat{\jmath}}\lambda\hat{P}_\lambda$. The elements of $\hat{\jmath} = \{\hat{\lambda}_1, \cdots, \hat{\lambda}_p\}$ are distinct and nonzero with probability one if $n > p$. Also, whether H_0 is true or false, if $n > p$, then

$$\text{rank}(\hat{P}_0A) = r \quad \text{almost surely.} \tag{2.5}$$

Statement (2.5) can be verified by noting that if $\text{rank}(\hat{P}_0A) < r$, then $\text{rank}(S_nA) < r$ and hence $\text{rank}(S_n) < p$. The last inequality can occur only on a set with probability zero.

Let $\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_{i+1}, \cdots, \hat{\mathbf{x}}_{i+m-1}$ be a set of orthonormal eigenvectors of S_n with $S_n\hat{\mathbf{x}}_j = \hat{\lambda}_j\hat{\mathbf{x}}_j$ for $j = i, \cdots, i + m - 1$. The asymptotic distribution of these sample principal component vectors depend upon the multiplicities of the elements in w, see Anderson (1963). For a multiple root, the sample vectors are not asymptotically normal. To develop tests for H_0 by using the asymptotic distribution of the sample vectors, assumptions concerning the multiplicities of the elements in w are needed. It would then be necessary to show that any resulting test does not depend on these additional assumptions. These complications can be avoided by using the sample projection $\hat{P}_0$ in formulating tests for H_0. Some basic results concerning the asymptotic distribution of $\hat{P}_0$ are given below.

LEMMA 2.1.
(i) *If* $P_0A = A$, *then* $\text{vec}\{n^{1/2}(I - \hat{P}_0)A\} \to \text{Normal}\{\mathbf{0}, \Sigma_0(A)\}$ *in distribution, where* $\Sigma_0(A) = \sum_{\mu\notin w}U(\mu, A) \otimes P_\mu$ and $U(\mu, A) = A'[\sum_{\lambda\in w}\{\lambda\mu/(\lambda - \mu)^2\}P_\lambda]A$.
(ii) *Let* $\hat{\Sigma}_0(A) = \sum_{\mu\notin\hat{w}}U_n(\mu, A) \otimes \hat{P}_\mu$ *where* $U_n(\mu, A) = A'[\sum_{\lambda\in\hat{w}}\{\lambda\mu/(\lambda - \mu)^2\}\hat{P}_\lambda]A$, *then* $\hat{\Sigma}_0(A) \to \Sigma_0(P_0A)$ *in probability.*
(iii) $\text{rank}\{\Sigma_0(A)\} = (p - m)\text{rank}(P_0A)$ *and* $\text{rank}\{\hat{\Sigma}_0(A)\} = (p - m)r$ *almost surely.*

PROOF. Part (i) is a special case of Theorem 4.1 in Tyler (1981). The form of $\Sigma_0(A)$ given here follows from (2.1), (2.4) and the facts $P'_\lambda = P_\lambda$, $P_\lambda P_\mu = 0$ for $\mu \neq \lambda$. Similarly, part (ii) is a special case of Equation (4.3) and Theorem 4.2 in Tyler (1981). Part (iii) is a consequence of Theorem 5.1. ii and Corollary 5.2.ii in Tyler (1981) and statement (2.5) above. □

REMARK 2.1. For computational purposes, the expression $U_n(\mu, A) = A'X_n\Delta_n(\mu)X'_nA$ can be used, where $\Delta_n(\mu)$ is an $m \times m$ diagonal matrix with entries $\mu\hat{\lambda}_j/(\hat{\lambda}_j - \mu)^2$ $j = i, \cdots, i + m - 1$ and $X_n = [\hat{\mathbf{x}}_i\ \hat{\mathbf{x}}_{i+1} \cdots \hat{\mathbf{x}}_{i+m-1}]$ with $\hat{\mathbf{x}}_i$ being defined prior to Lemma 2.1.

3. A class of invariant tests. The results of Lemma 2.1 can be used to develop an asymptotic chi-squared statistic for testing H_0, as in Tyler (1981). However, a more general asymptotic Wishart statistic can be constructed. This is done in Theorem 3.1. Before presenting the theorem and its proof, some additional notation and preliminary results are needed.

If $\text{vec}(X_{\nu\times d}) \sim \text{Normal}\{\text{vec}(C_{\nu\times d}), \Gamma_{d\times d} \otimes I_\nu\}$ with Γ nonsingular, then $X'X \sim \text{Wishart}_d(\nu, \Gamma, \frac{1}{2}\Gamma^{-1}C'C)$, a noncentral Wishart distribution of order d on ν degrees of freedom with covariance matrix argument Γ and noncentrality matrix argument $\frac{1}{2}\Gamma^{-1}CC'$. If $\Gamma = I$, then the distribution of the noncentral Wishart depends on the noncentrality matrix only through its roots. If $C = 0$, then $X'X \sim \text{Wishart}_d(\nu, \Gamma)$, a central Wishart distribution.

For a symmetric nonnegative definite matrix M, let $M^{1/2}$ represent the unique symmetric square-root of M and let M^+ represent the Moore-Penrose generalized inverse. These operations have the property

$$(Q'MQ)^{1/2} = Q'M^{1/2}Q \quad \text{and} \quad (Q'MQ)^+ = Q'M^+Q \tag{3.1}$$

for any orthogonal Q. Furthermore, if M_n represents a sequence of symmetric nonnegative definite random matrices which converge in probability to the nonrandom matrix M, then

$$M_n^{1/2} \to M^{1/2} \quad \text{in probability.} \tag{3.2}$$

If, in addition, $\text{rank}(M_n) \le \text{rank}(M)$ almost surely, then

$$M_n^+ \to M^+ \quad \text{in probability.} \tag{3.3}$$

Statement (3.2) follows from the continuity of the square-root operation, and statement (3.3) is a special case of Lemma 2.1 in Tyler (1981).

Finally, it should be noted that the matrix $U_n(\mu, A)$, defined in Lemma 2.1, is almost surely nonsingular whether H_0 is true or false. This statement easily follows from (2.5), Remark 2.1, and the identity $\hat{P}_0 = X_n X'_n$.

THEOREM 3.1. *Let* $W_n(A) = n\Sigma_{\mu\notin\hat{w}} \{U_n(\mu, A)\}^{-1/2} A'\hat{P}_\mu A \{U_n(\mu, A)\}^{-1/2}$.
(i) *If* $P_0 A = A$, *then* $W_n(A) \to \text{Wishart}_r(p - m, I)$ *in distribution.*
(ii) *If* $P_0 A \neq A$, *then for any fixed* x, $\text{Prob}[\text{trace}\{W_n(A)\} > x] \to 1$.

PROOF. (i) Let $t_n(A)$ be a $p \times r$ matrix defined such that

$$\text{vec}\{t_n(A)\} = \{\hat{\Sigma}_0(A)^{1/2}\}^+ \text{vec}\{n^{1/2}(I - \hat{P}_0)A\}. \tag{3.4}$$

It then follows from Lemma 2.1 and statements (3.2) and (3.3) that $\text{vec}\{t_n(A)\} \to \text{Normal}\{\mathbf{0}, I \otimes (I - P_0)\}$ in distribution. The asymptotic covariance of $\text{vec}\{t_n(A)\}$ is obtained as a special case of Theorem 5.1.i in Tyler (1981). By using properties (2.2) – (2.4) it can be shown that $t_n(A) = n^{1/2} \Sigma_{\mu\notin\hat{w}} \hat{P}_\mu A \{U_n(\mu, A)\}^{-1/2}$ since $\{\hat{\Sigma}_0(A)^{1/2}\} = \Sigma_{\mu\notin\hat{w}} U_n(\mu, A)^{-1/2} \otimes \hat{P}_\mu$, and $\hat{P}_\mu(I - \hat{P}_0) = \hat{P}_\mu$ for $\mu \notin \hat{w}$. Also, since $\hat{P}_\mu \hat{P}_\lambda = 0$ for $\mu \neq \lambda$ and $\hat{P}_\mu = \hat{P}'_\mu$, it follows that

$$W_n(A) = \{t_n(A)\}'\{t_n(A)\}. \tag{3.5}$$

Part (i) then follows by noting that $I - P_0$ is idempotent with rank $p - m$.

(ii) This is a special case of Theorem 5.3.iii in Tyler (1981). □

A sequence of local alternatives to H_0 can be constructed in the following manner. Let Σ_n be a sequence of nonrandom symmetric positive definite matrices of order p with the roots of Σ_n independent of n, and where as $n \to \infty$, $\Sigma_n \to \Sigma$ with Σ satisfying H_0. By the continuity of eigenvalues, the roots of Σ are the same as the roots of Σ_n. Let $P_{0,n}$ be the orthogonal projection matrix which projects onto the space spanned by the eigenvectors of Σ_n associated with the roots in w. Furthermore, let Σ_n be defined so that $n^{1/2}(I - P_{0,n}) A \to D$. Note that $P_0 D = 0$ since $P_{0,n} \to P_0$. Define the sequence of hypotheses

$$H_{1,n}\colon \mathbf{Y}_j \sim \text{Normal}\ (\mu, \Sigma_n), \quad 1 \le j \le n \tag{3.6}$$

with $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_n$ mutually independent.

THEOREM 3.2. *Under the sequence of hypotheses* $H_{1,n}$, $W_n(A) \to \text{Wishart}_r\{p - m, I, \Omega(A, D)\}$ *in distribution, where* $\Omega(A, D) = \tfrac{1}{2}\Sigma_{\mu\notin w} \{U(\mu, A)\}^{-1/2} D' P_\mu D \{U(\mu, A)\}^{-1/2}$ *and with* $U(\mu, A)$ *defined in Lemma* 2.1.

PROOF. Since the roots of Σ_n are the same as the roots of Σ, there exists a sequence of orthogonal matrices Q_n such that $\Sigma_n = Q_n \Sigma Q'_n$ with $Q_n \to I$. The transformation $S_n \to Q_n S_n Q'_n$ induces the transformations $\hat{P}_\mu \to Q_n \hat{P}_\mu Q'_n$ and $\hat{P}_0 \to Q_n \hat{P}_0 Q'_n$ and so it follows from (3.1) that the distribution of $W_n(A)$ under $H_{1,n}$ is the same as the distribution of $W_n(Q'_n A)$ under H_0. Thus, it suffices to find the asymptotic distribution of the latter. By expressing $\hat{\Sigma}_0(Q'_n A) = (A' Q_n \otimes I)\{\hat{\Sigma}_0(I_p)\}(Q'_n A \otimes I)$ it follows from Lemma 2.1.i that $\hat{\Sigma}_0(Q'_n A) \to (A' \otimes I)\{\Sigma_0\ (P_0)\}(A \otimes I) = \Sigma_0(A)$ in probability. Furthermore,

$\text{vec}\{n^{1/2}(I - \hat{P}_0)Q'_nA\} \to \text{Normal}\{\text{vec}(D), \Sigma_0(A)\}$ in distribution. To verify, expand $n^{1/2}(I - \hat{P}_0)Q'_nA = n^{1/2}(I - \hat{P}_0)A + D_n$ where $D_n = n^{1/2}(I - \hat{P}_0)(Q'_n - I)A$. Note that $n^{1/2}(I - P_{0,n})A = n^{1/2}Q_n(I - P_0)Q'_nA = n^{1/2}Q_n(I - P_0)(Q'_n - I)A$ and so $D_n \to D$ in probability. The aforementioned statement then follows from Lemma 2.1.i. The remainder of the proof of the theorem is analogous to the proof of Theorem 3.1.i. □

The problem of testing H_0 is invariant under the transformations $\mathbf{Y}_j \to Q\mathbf{Y}_j + \mathbf{b}$, $1 \le j \le n$, $A \to QAB$, where Q is an orthogonal matrix, $\mathbf{b}$ is a $p \times 1$ vector, and B is a nonsingular matrix. After reduction by sufficiency, it is easily shown that any invariant test is defined by a test statistic $T(S_n, A)$ having the property

$$T(S_n, A) = T(QS_nQ', QAB), \tag{3.7}$$

for any orthogonal matrix Q and any nonsingular matrix B. The asymptotic Chi squared statistic proposed in Tyler (1981), which in terms of the notation of the present paper is $\text{trace}\{W_n(A)\}$, possesses this property. More generally,

THEOREM 3.3. *Let $A_0 = A(A'A)^{-1/2}$ and let $w_1(S_n, A) \ge \cdots \ge w_r(S_n, A)$ represent the ordered eigenvalues of $W_n(A_0)$. Any function of $w_1(S_n, A), \cdots, w_r(S_n, A)$ satisfies property* (3.7)

PROOF. The transformation $A \to QAB$ induces the transformation $A_0 \to QA_0R$, where $R = (A'A)^{1/2}B(B'A'AB)^{-1/2}$ is an orthogonal matrix. The transformation $S_n \to QS_nQ'$ induces the transformations $\hat{P}_\mu \to Q\hat{P}_\mu Q'$ and $U_n(\mu, A) \to U_n(\mu, Q'A)$. By using property (3.1), these results imply that the transformation $(S_n, A) \to (QS_nQ', QAB)$ induces the transformation $W_n(A_0) \to QW_n(A_0)Q'$. Part (i) then follows since the roots of $W_n(A_0)$ and $QW_n(A_0)Q'$ are the same for any orthogonal matrix Q. □

REMARK 3.1. (i) Theorem 3.3 generalizes the Chi squared statistic in Tyler (1981) since $\text{trace}\{W_n(A)\} = \text{trace}\{W_n(A_0)\} = \sum_{j=1}^r w_j(S_n, A)$. (ii) The roots of $W_n(A)$ itself do not possess the invariance property (3.7). Introducing the matrix A_0 can be viewed as expressing the null hypothesis H_0 in a more canonical form, namely H_0: $P_0A_0 = A_0$ with A_0 having orthonormal columns. This form is unique up to post-multiplication of A_0 by an orthogonal matrix. (iii) It can be shown that the only functions of $t_n(A_0)$, the studentized difference between $\hat{P}_0A_0$ and A_0 defined by (3.4), which satisfy (3.7) are functions of $w_1(S_n, A), \cdots, w_r(S_n, A)$.

Let $W \sim \text{Wishart}_r(p - m, I, \Delta)$ where Δ is an $r \times r$ diagonal matrix with diagonal elements $\delta_1 \ge \cdots \ge \delta_r \ge 0$, and $\mathbf{w} = \{w_1, \cdots, w_r\}$ where $w_1 \ge \cdots \ge w_r \ge 0$ are the roots of W. Let h be a continuous function on R^r and let h_α be defined so that $\text{Prob}\{h(\mathbf{w}) > h_\alpha\} = \alpha$ whenever $\Delta = 0$. Furthermore, let h be defined so that the following conditions hold:

CONDITION 3.1. $h(\mathbf{v}) \to \sup\{h(\mathbf{v})\}$ as $\mathbf{v}'\mathbf{v} \to \infty$.

CONDITION 3.2. The function $\phi(\Delta) = \text{Prob}\{h(\mathbf{w}) > h_\alpha\}$ is monotonically increasing in each δ_j, $1 \le j \le r$.

Functions satisfying Condition 3.2 have been studied in relationship to MANOVA problems with known covariance matrices. In general, it is not easy to show that a given function h satisfies Condition 3.2. Perlman and Olkin (1980) give sufficient conditions on h for Condition 3.2 to hold. Examples of functions which satisfy both Condition 3.1 and 3.2 are $h(\mathbf{v}) = v_1$, $h(\mathbf{v}) = \pi_{i=1}^r(1 + v_i)$ and $h(\mathbf{v}) = \sum_{i=1}^r v_i$.

A consistent invariant asymptotic α-level test for H_0 with monotonically increasing local power function can then be obtained by defining as the rejection region

$$C(S_n, A) = \{S_n: h[w_1(S_n, A), \cdots, w_r(S_n, A)] > h_\alpha\}. \tag{3.8}$$

More specifically, as a consequence of the continuity property of eigenvalues, the following corollary to Theorems 3.1, 3.2 and 3.3 is obtained.

COROLLARY 3.1.
(i) $C(S_n, A) = C(QS_nQ', QAB)$ *for any orthogonal matrix* Q *and any nonsingular matrix* B.
(ii) *If* $P_0A = A$, *then* $\text{Prob}\{C(S_n, A)\} \to \alpha$.
(iii) *If* $P_0A \neq A$, *then* $\text{Prob}\{C(S_n, A)\} \to 1$.
(iv) *Under the sequence* $H_{1,n}$, $\text{Prob}\{C(S_n, A)\} \to \phi\{\Delta(A, D)\}$ *where* $\Delta(A, D)$ *is a diagonal matrix whose diagonal entries are the ordered eigenvalues of* $\Omega(A_0, D_0)$ *with* $D_0 = D(A'A)^{-1/2}$, *and where the function* ϕ *is defined in Condition* 5.2.

The matrix D and the diagonal matrix $\Delta(A, D)$ have the following relationships. Since $P_0D = 0$, D can be expressed as $\sum_{j\in J} d_j \mathbf{x}_j \mathbf{q}_j'$ where $\{\mathbf{x}_j, j \in J\}$ is a fixed set of orthonormal eigenvectors of Σ which span the range of $(I - P_0)$, and $\{\mathbf{q}_j, j \in J\}$ is a set of vectors in R^r. For any fixed set of vector $\{\mathbf{q}_j, j \in J\}$, the diagonal elements of $\Delta(A, D)$ are increasing functions in each d_j^2, with at least one of the diagonal elements of $\Delta(A, D)$ strictly increasing. This shows that Condition 5.2 is desirable. Furthermore, Condition 5.2 insures that asymptotically the rejection region $C(S_n, A)$ is locally unbiased.

REMARK 3.2. It can be shown that $W_n(A_0)$ is asymptotically equivalent to $W_n(A_1)$, where $A_1 = A(A'\hat{P}_0A)^{-1/2}$, in the sense that $W_n(A_1) - W_n(A_0) \to 0$ in probability under both H_0 and under the sequence $H_{1,n}$. Furthermore, $\text{tr}\{W_n(A_0)\} = \text{tr}\{W_n(A_1)\}$, and the roots of $W_n(A_1)$ satisfy property (3.7). Thus, Corollary 3.1 holds if the roots of $W_n(A_0)$ are replaced by the roots of $W_n(A_1)$ in defining $C(S_n, A)$.

For the special case $r = m$, the roots of $W_n(A_1)$ are the same as the roots of $W_n(A_2)$ where $A_2 = A(X_n'A)^{-1}$ and X_n is defined as in Remark 2.1. This follows since the roots of $W_n(A_1)$ are unchanged if A is postmultiplied by the nonsingular matrix $(X_n'A)^{-1}$ and since $\hat{P}_0 = X_nX_n'$. If A_2 and $W_n(A_2)$ are represented by $A_2 = [\mathbf{a}_i\ \mathbf{a}_{i+1} \cdots \mathbf{a}_{i+m-1}]$ and $W_n(A_2) = \{\hat{w}_{jk}\}$ for $j, k = i, \cdots, i + m - 1$ respectively, then

$$\hat{w}_{jk} = n[(\hat{\lambda}_j\hat{\lambda}_k)^{1/2}\mathbf{a}_j'S_n^{-1}\mathbf{a}_k + (\hat{\lambda}_j\hat{\lambda}_k)^{-1/2}\mathbf{a}_j'S_n\mathbf{a}_k - \{(\hat{\lambda}_j/\hat{\lambda}_k)^{1/2} + (\hat{\lambda}_k/\hat{\lambda}_j)^{1/2}\}\mathbf{a}_j'\mathbf{a}_k]. \tag{3.9}$$

Note the similarity between (3.9) and (1.1). For computational purposes, it is simpler to calculate $W_n(A_2)$ by using (3.9) than it is to calculate $W_n(A_0)$.

For $r < m$, the author has not been able to derive in general a simple computational form either for $W_n(A_0)$ or for any matrix whose roots are asymptotically equivalent to the roots of $W_n(A_0)$. However, if one makes the additional assumption that $\lambda_i = \lambda_{i+1} = \cdots = \lambda_{i+m-1}$, then it can be shown under both H_0 and under the sequence $H_{1,n}$ that $W_n(A_0)$ is asymptotically equivalent to

$$n\{\hat{\lambda}A_0'S_n^{-1}A_0 + \hat{\lambda}^{-1}A_0'S_nA_0 - 2I\}, \tag{3.10}$$

where $\hat{\lambda} = m^{-1}\sum_{j=i}^{i+m-1}\hat{\lambda}_j$.

REMARK 3.3. For comparing two subspaces of order m, Krzanowski (1979) suggests using the angular values $\{\text{Cos}^{-1}(\gamma_j^{1/2}), 1 \le j \le m\}$ where $\gamma_1 \ge \cdots \ge \gamma_m$ are the roots of $M_1'M_2M_2'M_1$ with M_1 and M_2 being $p \times m$ matrices whose columns are orthonormal and span the respective spaces. Applying this concept to the problem of testing H_0 for the case $r = m$, it suggests comparing the sample and hypothesized principal component spaces via the statistics $\{\text{Cos}^{-1}(\hat{\gamma}_j^{1/2}), 1 \le j \le m\}$ where $\hat{\gamma}_1 \ge \cdots \ge \hat{\gamma}_m$ are the roots of $A_0'\hat{P}_0A_0$. These statistics possess the invariance property (3.7). The asymptotic distribution, however, is somewhat intractable. Under H_0, $n(A_0'\hat{P}_0A_0 - I) = \{n^{1/2}(I - \hat{P}_0)A_0\}'\{n^{1/2}(I - \hat{P}_0)A_0\}$ which converges in distribution to $N_0'N_0$ where $\text{vec}(N_0) \sim \text{Normal}\{\mathbf{0}, \Sigma_0(A_0)\}$. Thus, the limiting null distribution of $\{n\text{Cos}^{-1}(\hat{\gamma}_j^{1/2}), 1 \le j \le m\}$ corresponds to the joint distribution of the Arcosines of the square-roots of the ordered eigenvalues of $N_0'N_0$.

4. Robustness. In this section assume that $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_n$ is a random sample from a p-dimensional population with density function of the form $f_g(\mathbf{y}; \mu, \Lambda) = |\Lambda|^{-1/2}g\{(\mathbf{y} - \mu)'\Lambda^{-1}(\mathbf{y} - \mu)\}$ for some symmetric positive definite matrix Λ and some nonnegative function g, where g is not dependent on μ or Λ. Properties of distributions with elliptically contoured density functions have been studied by Kelker (1970). In particular, if the second moments of the distribution exist, then the population covariance matrix is proportion to Λ.

Maronna (1976) defines affine-invariant M-estimates of location and scatter to be solutions to a system of equations of the form $n^{-1}\Sigma_i u_1(d_i)(\mathbf{Y}_i - \hat{\mu}) = \mathbf{0}$, and $n^{-1}\Sigma_i u_2(d_i^2)(\mathbf{Y}_i - \mu)(\mathbf{Y}_i - \hat{\mu})' = \hat{V}$, where $d_i^2 = (\mathbf{Y}_i - \hat{\mu})'\hat{V}^{-1}(\mathbf{Y}_i - \hat{\mu})$. The functions u_1 and u_2 satisfy a set of general assumptions given in Section 2 of Maronna's paper. The solutions $(\hat{\mu}, \hat{V})$ are estimates for the parameters (μ, V) where $V = \sigma^{-1}\Lambda$. The parameter σ is the solution of an integral equation and depends on the function g. Maronna (1976) shows that $n^{1/2}(\hat{V} - V) \to Z$ in distribution, where Z is multivariate normal with mean zero. The covariance matrix of $\text{vec}(Z)$ is given in Tyler (1982) as

$$(4.1) \qquad \sigma_1(I + K_{p,p})(V \otimes V) + \sigma_2\{\text{vec}(V)\}\{\text{vec}(V)\}',$$

where σ_1 and σ_2 are parameters depending on g. For more details, see Tyler (1982).

Let the notation established in Section 1 be used when (Σ, S_n) is replaced by $(V, \hat{V})$. Since V is proportional to Λ, the hypothesis H_0: $P_0A = A$ is the same whether the eigenprojection P_0 refers to the matrix V or the matrix Λ. Analogous to Lemma 2.1.i, it can be shown that if $P_0A = A$, then

$$(4.2) \qquad \text{vec}\{n^{1/2}(I - \hat{P}_0)A\} \to \text{Normal}\{\mathbf{0}, \sigma_1\Sigma_0(A)\}$$

in distribution, where $\Sigma_0(A)$ is defined as in Lemma 2.1 with Σ replaced by V. This result then leads to the following generalization of Corollary 3.1.

THEOREM 4.1. *Corollary 3.1 holds whenever S_n, Σ, $\{w_j(S_n, A), 1 \le j \le r\}$ and $\Delta(A_0, D_0)$ are replaced by $\hat{V}$, V, $\{\hat{\sigma}_1^{-1}w_j(\hat{V}, A), 1 \le j \le r\}$ and $\sigma_1^{-1}\Delta(A_0, D_0)$ respectively, where $\hat{\sigma}_1$ represents a consistent estimate of σ_1.*

Acknowledgements. The author is grateful for the numerous comments by the referees which have improved the quality of the manuscript, especially with regards to the invariance arguments and the definition of local alternatives. I also wish to belatedly thank the associate editor who handled a previous paper (Tyler, 1981) for suggesting that my results could be extended from normal populations to elliptical populations in general.

REFERENCES

ANDERSON, T. W. (1963). Asymptotic theory for principal components analysis. *Ann. Math. Statist.* **34** 122–148.

KELKER, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā Ser. A* **32** 419–430.

KRZANOWSKI, T. E. (1979). Between-groups comparison of principal components. *J. Amer. Statist. Assoc.* **74** 703–707.

MAGNUS, J. R. and NEUDECKER, H. (1979). The commutation matrix: some properties and applications. *Ann. Statist.* **7** 381–394.

MARONNA, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4** 51–67.

PERLMAN, M. D. and OLKIN, I. (1980). Unbiasedness of invariant tests for MANOVA and other multivariate problems. *Ann. Statist.* **8** 1326–1341.

TYLER, D. E. (1981). Asymptotic inference for eigenvectors. *Ann. Statist.* **9** 725–736.

TYLER, D. E. (1982). Radial estimates and the test for sphericity. *Biometrika* **69** 429–436.

DEPARTMENT OF STATISTICS
HILL CENTER
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903