JOURNAL OF MULTIVARIATE ANALYSIS 52, 245–258 (1995)

Likelihood Ratio Tests for Principal Components

LUTZ DÜMBGEN*

Universität Heidelberg, Heidelberg, Germany

A particular class of tests for the principal components of a scatter matrix Σ is proposed. In the simplest case one wants to test whether a given vector is an eigenvector of Σ corresponding to its largest eigenvalue. The test statistics are likelihood ratio statistics for the classical Wishart model, and critical values are obtained parametrically as well as nonparametrically without making any assumptions on the eigenvalues of Σ. Still, the tests have asymptotic properties similar to those of classical procedures and are asymptotically admissible and optimal in some sense.

1. INTRODUCTION

Many multivariate procedures are based on eigenvalues and eigenvectors of random matrices. A major mathematical problem is that these functions are not everywhere differentiable, so that usual delta methods can fail. In the present paper it is shown how one can deal with this problem for a particular classical testing problem in principal component analysis.

Let $\Sigma \in \mathbf{X}_+$ be an unknown scatter matrix, where $\mathbf{X}$ is the set of all symmetric matrices in $\mathbf{R}^{d \times d}$, and $\mathbf{X}_+$ is the subset of positive definite $X \in \mathbf{X}$. Further, let $S \in \mathbf{X}_+$ be an estimator for Σ. Before stating general conditions on Σ, S let us us assume for the moment that

$$\mathscr{L}(nS) \text{ is a Wishart distribution } \mathscr{W}(\Sigma, n) \tag{1}$$

for some fixed $n \geqslant d$. Now consider a spectral representation of Σ,

$$\Sigma = \sum_{i=1}^{d} \lambda_i(\Sigma)\, \tau_i \tau_i',$$

Received March 19, 1993; revised June 21, 1994.

AMS 1980 subject classifications: 62G20, 62H15, 62H25.

Keywords and phrases: admissibility, bootstrap, confidence set, convex cone, eigenvalue, eigenvector, local alternatives, nonparametric, optimality, principal components, Wishart distribution.

* Work supported by Sonderforschungsbereich 343, Universität Bielefeld.

0047-259X/95 $6.00

where $\lambda_1(\Sigma) \geq \lambda_2(\Sigma) \geq \cdots \geq \lambda_d(\Sigma)$ are its eigenvalues, and $\tau_1, \tau_2, ..., \tau_d$ are the corresponding orthonormal eigenvectors. (A' denotes the transpose of a matrix A.) Sometimes theoretical considerations suggest certain candidates for the unknown principal vectors τ_i or one wants to replace the observed eigenvectors of S with vectors having a simpler form and which are easier to interpret. Then it is important to test whether these candidates are plausible. For instance, given orthonormal vectors $e_1, e_2, ..., e_k \in \mathbf{R}^d$, one might want to test whether

$$\operatorname{span}(\tau_1, \tau_2, ..., \tau_k) = \operatorname{span}(e_1, e_2, ..., e_k) \tag{2}$$

(where span($\cdot$) denotes the linear span) or whether

$$\tau_i = \pm e_i, \qquad \forall i = 1, 2, ..., k. \tag{3}$$

The special case $k = 1$ has been treated by Anderson [1]. Tyler [13, 14] modified and extended Anderson's method to test (2) for arbitrary k (as well as more general hypotheses). These tests are asymptotically valid if $n^{1/2}(\lambda_k/\lambda_{k+1} - 1)(\Sigma)$ tends to infinity. Schott [9] demonstrated by Monte Carlo simulations that these tests can be very inaccurate. Using higher order Taylor expansions for eigenvalues and eigenvectors, he derived a Bartlett adjustment, which improves the validity substantially, but still the actual level can be noticably higher than the nominal one.

The present paper is organized as follows: In Section 2 the special hypotheses (2, 3) are generalized, and the likelihood ratio (LR) test statistic under the Wishart model (1) is derived. This test statistic is used to construct a valid test, where the critical value is obtained by considering the degenerate case $\Sigma = I$ only; I stands for the identity matrix in $\mathbf{R}^{d \times d}$. This test is biased in the sense that its level is not constant for the hypothesis. The bias is the price to pay for not making any assumptions on the eigenvalues of Σ.

In Section 3 we consider the asymptotic behavior of our test if both $\Sigma = \Sigma_n$ and $S = S_n$ depend on an additional index $n \geq d$ such that $\Sigma_n^{-1/2} S_n \Sigma_n^{-1/2} = I + O_p(n^{-1/2})$. (Generally $X^{1/2}$ denotes the symmetric, positive definite square root of $X \in \mathbf{X}_+$, and $X^{-1/2} = (X^{1/2})^{-1}$.) Within this asymptotic framework one can replace parametric assumptions such as (1) by a more realistic asymptotic condition and obtain an asymptotically valid test. Letting Σ vary with n allows critical quantities such as $n^{1/2}(\lambda_k/\lambda_{k+1} - 1)(\Sigma_n)$ to be arbitrarily large but bounded as n tends to infinity. On the other hand, in many applications the ratio $(\lambda_1/\lambda_d)(S)$ is rather large, and therefore it is desirable to allow $(\lambda_1/\lambda_d)(\Sigma_n)$ to diverge. It turns out that our test has the same consistency properties as classical procedures, and in a local asymptotic framework, where $\Sigma_n = I + O(n^{-1/2})$, it is asymptotically admissible and optimal in some sense. Section 4 provides analytical tools that are needed in Section 3.

The last section contains some concluding remarks and references to related work.

2. Definition and Basic Properties of the Test

Let us first introduce some notation: For any real, symmetric matrix A let $\lambda(A)$ be the vector of its ordered eigenvalues $\lambda_1(A) \geqslant \lambda_2(A) \geqslant \lambda_3(A) \geqslant \cdots$. The set of all orthonormal matrices in $\mathbf{R}^{d \times d}$ is denoted by $\mathbf{T}$. For $X \in \mathbf{X}$ let $\mathbf{T}(X)$ be the set of all $T \in \mathbf{T}$ such that $T'XT = \text{diag}(\lambda(X))$, where $\text{diag}(v)$ stands for $(\mathbf{1}\{i=j\}\, v_i)_{1 \leqslant i,j \leqslant d}$.

Let $Q = (Q_1, Q_2, ..., Q_{\bar{a}})$ be a partition of $\{1, 2, ..., d\}$ into $\bar{a} > 1$ sets such that $\max(Q_a) \leqslant \min(Q_{a+1})$ for $1 \leqslant a < \bar{a}$. We write $i \sim_Q j$ if i, j belong to the same atom of Q, and $i \nsim_Q j$ otherwise. The same letter Q is used to define the projection

$$Q(M) := (\mathbf{1}\{i \sim_Q j\}\, M_{ij})_{1 \leqslant i,j \leqslant d}$$

on $\mathbf{R}^{d \times d}$. We want to test the hypothesis that Σ belongs to the set

$$\mathbf{K} := \{X \in \mathbf{X}: \mathbf{T}(X) \cap Q(\mathbf{R}^{d \times d}) \neq \varnothing\} \subset Q(\mathbf{X}).$$

The two hypotheses (2, 3) mentioned in the Introduction correspond to the partitions

$$(\{1, 2, ..., k\}, \{k+1, k+2, ..., d\})$$
$$\text{and} \quad (\{1\}, \{2\}, ..., \{k\}, \{k+1, k+2, ..., d\}),$$

respectively, where we assume without loss of generality that $e_1, e_2, ..., e_k$ are the first k standard basis vectors of $\mathbf{R}^d$.

In the classical model (1) the LR test of a simple hypothesis $\Sigma = H$ is based on the test statistic

$$t(S, H) := \text{trace}(H^{-1}S) - d - \log(\det(H^{-1}S)) \geqslant 0;$$

see Anderson [2]. Thereore the LR test of the hypothesis $\Sigma \in \mathbf{K}$ uses

$$t_{\mathbf{K}}(S) := \inf_{H \in \mathbf{K} \cap \mathbf{X}_+} t(S, H).$$

Let $\tilde{S}$ be the unobserved random matrix

$$\tilde{S} := \Sigma^{-1/2} S \Sigma^{-1/2},$$

so that $S = \tilde{S}$ if $\Sigma = I$. Under (1) the matrix $n\tilde{S}$ has a standard Wishart distribution with n degrees of freedom, and in fact the only assumption needed here is that

$$\mathscr{L}(\tilde{S}) \text{ is known.} \tag{4}$$

Now the hypothesis $\Sigma \in \mathbf{K}$ is rejected at level $\alpha \in]0, 1[$ if $t_{\mathbf{K}}(S) > c(\alpha)$, where $c(\alpha)$ is the $(1-\alpha)$-quantile of $\mathscr{L}(t_{\mathbf{K}}(\tilde{S}))$. The following theorem implies validity and other properties of this test.

THEOREM 1. *The test statistic $t_{\mathbf{K}}(S)$ can be written as*

$$t_{\mathbf{K}}(S) = t(S, Q(S)) + t_{\mathbf{K}}(Q(S)).$$

Suppose that $\Sigma \in \mathbf{K}$. Then

$$t_{\mathbf{K}}(\tilde{S}) \geqslant t_{\mathbf{K}}(S) \geqslant t(\tilde{S}, Q(\tilde{S})),$$

where the latter inequality is an equality if

$$(\lambda_i/\lambda_{i+1})(\Sigma) \geqslant (\lambda_1/\lambda_d)(\tilde{S}) \qquad \textit{whenever} \quad i \not\sim_Q i+1.$$

If in addition $\mathscr{L}(T'\tilde{S}T) = \mathscr{L}(\tilde{S})$ for arbitrary $T \in \mathbf{T}$, then the distribution of $t_{\mathbf{K}}(S)$ is a decreasing function of the ratios $(\lambda_i/\lambda_{i+1})(\Sigma)$, $1 \leqslant i < d$.

As for the decomposition of $t_{\mathbf{K}}(S)$, it can be shown that $t(S, Q(S))$ is equivalent to the LR test statistic for the hypothesis $\Sigma \in Q(\mathbf{X})$, while $t_{\mathbf{K}}(Q(S))$ is equivalent to the LR statistic for testing $\Sigma \in \mathbf{K}$ versus $\Sigma \in Q(\mathbf{X})$. The two inequalities for $t_{\mathbf{K}}(S)$ imply that

$$\alpha \geqslant \mathbb{P}\{t_{\mathbf{K}}(S) > c(\alpha)\} \geqslant \mathbb{P}\{t(\tilde{S}, Q(\tilde{S})) > c(\alpha)\} \qquad \text{if} \quad \Sigma \in \mathbf{K}.$$

The lower bound is attained asymptotically if

$$(\lambda_i/\lambda_{i+1})(\Sigma) \to \infty \text{ whenever } i \not\sim_Q i+1.$$

Therefore the proposed test is valid, but its power is not constant on $\mathbf{K} \cap \mathbf{X}_+$, in general.

The degenerate case $\Sigma = I$ is admittedly rather extreme. With the help of the last part of Theorem 1 one could utilize moderate prior assumptions on $\lambda(\Sigma)$. Precisely, suppose that

$$(\lambda_i/\lambda_{i+1})(\Sigma) \geqslant v_i/v_{i+1} \qquad \text{for} \quad 1 \leqslant i \leqslant d$$

for some vextor $v \in \lambda(\mathbf{X}_+)$, and let $c(\alpha, v)$ be the $(1-\alpha)$-quantile of the distribution of $t_{\mathbf{K}}(\operatorname{diag}(v)^{1/2} \tilde{S} \operatorname{diag}(v)^{1/2})$. Then

$$\mathbb{P}\{t_{\mathbf{K}}(S) > c(\alpha, v)\} \leqslant \alpha,$$

provided that $\Sigma \in \mathbf{K}$ and $\mathscr{L}(\tilde{S})$ is invariant as stated in Theorem 1.

The test statistic $t_{\mathbf{K}}(S)$ can be computed explicitly as follows. For $X \in \mathbf{X}$ let

$$\lambda_Q(X) := (\lambda((X_{ij})_{i,j \in Q(1)})', \lambda((X_{ij})_{i,j \in Q(2)})', ..., \lambda((X_{ij})_{i,j \in Q(\bar{a})})')',$$
$$\mathbf{T}_Q(X) := \{T \in \mathbf{T} : Q(T) = T \text{ and } T'Q(X)\, T = \operatorname{diag}(\lambda_Q(X))\}.$$

Note that $\lambda_Q(\cdot) \equiv \lambda_Q(Q(\cdot)) \not\equiv \lambda(Q(\cdot))$ and $\mathbf{T}_Q(\cdot) = \mathbf{T}_Q(Q(\cdot)) \not\equiv \mathbf{T}(Q(\cdot))$. For any matrix A define $\|A\| := \sqrt{\operatorname{trace}(A'A)}$.

THEOREM 2. *Let T be any point in $\mathbf{T}_Q(S)$, and define*

$$\mu := \arg \min_{v \in \lambda(\mathbf{X})} \|v - \lambda_Q(S)\|^2.$$

Then

$$\|S - T\operatorname{diag}(\mu)\, T'\| = \operatorname{dist}(S, \mathbf{K}) := \inf_{X \in \mathbf{K}} \|S - X\|,$$

$$t(S, T\operatorname{diag}(\mu)\, T') = t_{\mathbf{K}}(S) = \sum_{i=1}^{d} \log\left(\frac{\mu_i}{\lambda_i(S)}\right).$$

For the exact computation of μ see Section 1.2 of Robertson *et al.* [8]. Theorem 2 shows that the statistics $t_{\mathbf{K}}(S)$ and $\operatorname{dist}(S, \mathbf{K})$ are closely related (though not equivalent), which is discussed in more detail in Section 4.

Proof of Theorem 1. In what follows we frequently use the fact that $Q(LM) = Q(L)\, M$ and $Q(ML) = MQ(L)$ for arbitrary $L \in \mathbf{R}^{d \times d}$ and $M \in Q(\mathbf{R}^{d \times d})$. In particular,

$$\operatorname{trace}(H^{-1}S) = \operatorname{trace}(Q(H^{-1}S)) = \operatorname{trace}(H^{-1}Q(S)) \qquad \forall H \in Q(\mathbf{X}_+). \quad (5)$$

Therefore

$$\begin{aligned} t_{\mathbf{K}}(S) &= \inf_{H \in \mathbf{K} \cap \mathbf{X}_+} (\operatorname{trace}(H^{-1}Q(S)) - d + \log(\det(H))) - \log(\det(S)) \\ &= t_{\mathbf{K}}(Q(S)) - \log(\det(Q(S)^{-1}\, S)) \\ &= t_{\mathbf{K}}(Q(S)) + t(S, Q(S)), \end{aligned}$$

because $\operatorname{trace}(Q(S)^{-1}\, S) = d$.

For $1 \leqslant a \leqslant \bar{a}$ let $\mathbf{U}_a$ denote the set of all unit vectors $u \in \mathbf{R}^d$ such that $u_i = 0$ whenever $i \notin Q_a$. Then one can write

$$\mathbf{K} = \{X \in Q(\mathbf{X}) : u'Xu \geqslant v'Xv \text{ for } u \in \mathbf{U}_a, v \in \mathbf{U}_{a+1} \text{ and } 1 \leqslant a < \bar{a}\}.$$

This representation shows that $\mathbf{K}$ is a closed, convex cone in $\mathbf{X}$. Another useful implication is that

$$\Sigma^{1/2} H \Sigma^{1/2} \in \mathbf{K} \qquad \forall H \in \mathbf{K} \cap \mathbf{X}_+, \tag{6}$$

where Σ is assumed to be in $\mathbf{K}$. $\Sigma^{1/2} H \Sigma^{1/2}$ is easily seen to be in $Q(\mathbf{X}_+)$. Moreover, if $h(u) := \|\Sigma^{1/2} u\|^{-1} \Sigma^{1/2} u$, then $h(\mathbf{U}_a) = \mathbf{U}_a$ for all a. Hence

$$u'\Sigma^{1/2} H \Sigma^{1/2} u = u'\Sigma u h(u)' \, Hh(u) \geqslant v'\Sigma v h(v)' \, Hh(v) = v'\Sigma^{1/2} H \Sigma^{1/2} v$$

for $u \in \mathbf{U}_a$, $v \in \mathbf{U}_{a+1}$, and $1 \leqslant a \leqslant \bar{a}$. But $t(\cdot, \cdot)$ is invariant in that

$$t(X, Y) = t(M'XM, M'YM) \qquad \text{for nonsingular } M \in \mathbf{R}^{d \times d} \text{ and } X, Y \in \mathbf{X}_+. \tag{7}$$

Consequently, (6) and (7) together imply that

$$t_{\mathbf{K}}(\tilde{S}) = \inf_{H \in \mathbf{K} \cap \mathbf{X}_+} t(S, \Sigma^{1/2} H \Sigma^{1/2}) \geqslant t_{\mathbf{K}}(S).$$

Since $t_{\mathbf{K}}(\cdot)$ is nonnegative, the first part of Theorem 1 implies that $t_{\mathbf{K}}(S) \geqslant t(S, Q(S))$, and (7) yields

$$t(S, Q(S)) = t(\tilde{S}, \Sigma^{-1/2} Q(S) \Sigma^{-1/2}) = t(\tilde{S}, Q(\Sigma^{-1/2} S \Sigma^{-1/2})) = t(\tilde{S}, Q(\tilde{S})).$$

Note that

$$\lambda_d(\tilde{S}) \leqslant v'Sv/v'\Sigma v \leqslant \lambda_1(\tilde{S}) \qquad \forall v \in \mathbf{R}^d \setminus \{0\}. \tag{8}$$

In particular, for $u \in \mathbf{U}_a$, $v \in \mathbf{U}_{a+1}$, and $1 \leqslant a < \bar{a}$,

$$u'Q(S)\, u = u'Su \geqslant \lambda_d(\tilde{S})\, u'\Sigma u \geqslant \lambda_d(\tilde{S})\, \lambda_{\max(Q_a)}(\Sigma)$$

and

$$v'Q(S)\, v \leqslant \lambda_1(\tilde{S})\, \lambda_{\min(Q_{a+1})}(\Sigma).$$

Consequently, $Q(S) \in \mathbf{K}$ if $(\lambda_i/\lambda_{i+1})(\Sigma) \geqslant (\lambda_1/\lambda_d)(\tilde{S})$ whenever $i \not\to_Q i+1$, which entails $t_{\mathbf{K}}(Q(S)) = 0$.

One easily verifies that $T'(\mathbf{K} \cap \mathbf{X}_+)\, T = \mathbf{K} \cap \mathbf{X}_+$ for all $T \in \mathbf{T} \cap Q(\mathbf{R}^{d \times d})$. Thus it follows from (7) and the invariance assumption on $\mathscr{L}(\tilde{S})$ that the (hypothetical) distribution of $t_{\mathbf{K}}(S)$ depends only on $\lambda(\Sigma)$. Now let

$\Sigma_1 = \mathrm{diag}(\lambda(\Sigma_1))$ and $\Sigma_2 = \mathrm{diag}(\lambda(\Sigma_2))$ be two matrices in $\mathbf{K} \cap \mathbf{X}_+$ such that

$$(\lambda_i/\lambda_{i+1})(\Sigma_1) \geqslant (\lambda_i/\lambda_{i+1})(\Sigma_2) \qquad \text{for} \quad 1 \leqslant i < d.$$

Then $\Gamma := \Sigma_2^{-1}\Sigma_1 \in \mathbf{K} \cap \mathbf{X}_+$, and

$$t_{\mathbf{K}}(\Sigma_1^{1/2}\tilde{S}\Sigma_1^{1/2}) = t_{\mathbf{K}}(\Gamma^{1/2}(\Sigma_2^{1/2}\tilde{S}\Sigma_2^{1/2})\,\Gamma^{1/2}) \leqslant t_{\mathbf{K}}(\Sigma_2^{1/2}\tilde{S}\Sigma_2^{1/2}). \quad \blacksquare$$

Proof of Theorem 2. Two well-known inequalities, due to Hoffman and Wielandt [5] and von Neumann [7], respectively, are

$$\|\lambda(D) - \lambda(E)\|^2 \leqslant \|D - E\|^2 \qquad \forall D, E \in \mathbf{X}; \tag{9}$$

$$\mathrm{trace}(E^{-1}D) \geqslant \mathrm{trace}(\mathrm{diag}(\lambda(E))^{-1}\,\mathrm{diag}(\lambda(D))) \qquad \forall D, E \in \mathbf{X}_+. \tag{10}$$

One can apply (9) to the $\bar{a}$ main subblocks of $Q(S)$ and $X \in \mathbf{K}$ separately to show that

$$\begin{aligned} \|S - X\|^2 &= \|S - Q(S)\|^2 + \|Q(S) - X\|^2 \\ &\geqslant \|S - Q(S)\|^2 + \|Q(S) - T\,\mathrm{diag}(\mu)\,T'\|^2 \\ &= \|S - T\,\mathrm{diag}(\mu)\,T'\|^2. \end{aligned}$$

Hence $\|S - T\,\mathrm{diag}(\mu)\,T'\| = \mathrm{dist}(S, \mathbf{K})$. Analogously one can apply (10) to the $\bar{a}$ main subblocks of $Q(S)$ and $H \in \mathbf{K} \cap \mathbf{X}_+$ separately for showing that

$$\begin{aligned} \mathrm{trace}(H^{-1}Q(S)) &\geqslant \mathrm{trace}((T\,\mathrm{diag}(\lambda(H))\,T')^{-1}\,Q(S)) \\ &= \mathrm{trace}(\mathrm{diag}(\lambda(H))^{-1}\,\mathrm{diag}(\lambda_Q(S))). \end{aligned}$$

Therefore

$$\begin{aligned} t_{\mathbf{K}}(S) &= \inf_{H \in \mathbf{K} \cap \mathbf{X}_+} (\mathrm{trace}(H^{-1}Q(S)) - d + \log(\det(H))) - \log(\det(S)) \\ &= \inf_{v \in \lambda(\mathbf{X}_+)} \sum_{i=1}^{d} (v_i^{-1}\lambda_{Qi}(S) + \log(v_i) - \log(\lambda_i(S)) - d. \end{aligned}$$

It follows from Theorem 1.5.1 in Robertson *et al.* [8] with $\Phi(r) := -\log(r)$ that the latter infimum is attained in μ, whence $t_{\mathbf{K}}(S) = t(S, T\,\mathrm{diag}(\mu)\,T')$. Finally, since $r\mu \in \lambda(\mathbf{X}_+)$ for all $r > 0$,

$$0 = \frac{d}{dr}\bigg|_{r=1} \sum_{i=1}^{d} ((r\mu_i)^{-1}\lambda_{Qi}(S) + \log(r\mu_i)) = d - \sum_{i=1}^{d} \mu_i^{-1}\lambda_{Qi}(S).$$

Thus $t_{\mathbf{K}}(S) = \sum_{i=1}^{d} \log(\mu_i/\lambda_i(S))$. $\blacksquare$

3. Asymptotic Properties of the Test

Unless stated otherwise, all results in this section follow from analytical properties of the functions $t(\cdot,\cdot)$ and $t_{\mathbf{K}}(\cdot)$, which are derived in Section 4. It is assumed that $\Sigma=\Sigma_n$ and $S=S_n$ for $n\geqslant d$ such that

$$W_n:=n^{1/2}(\tilde{S}_n-I)\to W \qquad \text{in distribution} \tag{11}$$

(as $n\to\infty$), where $W\in\mathbf{X}$ is a random matrix with a centered, nonsingular Gaussian distribution. Then S_n is a consistent estimator for Σ_n in the sense that

$$\Sigma_n^{-1/2}S_n^{1/2}=I+O_p(n^{-1/2}). \tag{12}$$

It is well-known that in the classical model (1) assumption (11) is satisfied with a random matrix $W\in\mathbf{X}$ having independent components $W_{ij}=W_{ji}\sim\mathcal{N}(0,1+\mathbf{1}\{i=j\})$ for $1\leqslant i\leqslant j\leqslant d$.

Within this asymptotic framework we weaken the parametric assumption (4) of Section 2 and require only that there is an estimator $\hat{D}_n$ for the distribution of $\tilde{S}_n$ such that

$$\mathscr{L}_{\hat{S}_n\sim\hat{D}_n}(n^{1/2}(\hat{S}_n-I))\to\mathscr{L}(W) \qquad \text{weakly in probability.} \tag{13}$$

Now the (unknown) quantile $c(\alpha)=c_n(\alpha)$ is estimated by the $(1-\alpha)$-quantile $\hat{c}_n(\alpha)$ of $\mathscr{L}_{\hat{S}_n\sim\hat{D}_n}(t_{\mathbf{K}}(\hat{S}_n))$.

Proposition 1. *Both* $\mathscr{L}(nt_{\mathbf{K}}(\tilde{S}_n))$ *and* $\mathscr{L}_{\hat{S}_n\sim\hat{D}_n}(nt_{\mathbf{K}}(\hat{S}_n))$ *converge weakly to the continuous distribution* $\mathscr{L}(\mathrm{dist}(W,\mathbf{K})^2/2)$ *in probability.*

Theorem 1 and Proposition 1 together imply that the test $\mathbf{1}\{t_{\mathbf{K}}(S_n)>\hat{c}_n(\alpha)\}$ is asymptotically valid; i.e.,

$$\limsup_{n\to\infty}\mathbb{P}\{t_{\mathbf{K}}(S_n)>\hat{c}_n(\alpha)\}\leqslant\alpha \qquad \text{if} \quad \Sigma_n\in\mathbf{K}\quad\forall n\geqslant d.$$

It is shown in Section 4 that the quantity $t_{\mathbf{K}}(\Sigma_n)$ is a natural measure of "distance" from Σ_n to the hypothesis $\mathbf{X}_+\cap\mathbf{K}$. In terms of eigenvalues and eigenvectors one can say that $nt_{\mathbf{K}}(\Sigma_n)$ is bounded if, and only if,

$$\max_{U\in\mathbf{T}(\Sigma_n),\,V\in\mathbf{T}_Q(\Sigma_n)}\sum_{i,j=1}^{d}\frac{(\lambda_i-\lambda_j)^2}{\lambda_i\lambda_j}(\Sigma_n)(U'V)_{ij}^2=O(n^{-1}). \tag{14}$$

The test $\mathbf{1}\{t_{\mathbf{K}}(S_n)>\hat{c}_n(\alpha)\}$ is consistent in that it detects large values of $nt_{\mathbf{K}}(\Sigma_n)$ with high probability.

PROPOSITION 2.

$$\mathbb{P}\{t_{\mathbf{K}}(S_n) > \hat{c}_n(\alpha)\} \geqslant \mathbb{P}\{\operatorname{dist}(W, \mathbf{K}) < \sqrt{2nt_{\mathbf{K}}(\Sigma_n)} - \gamma(\alpha)\} + o(1),$$

where $\gamma(\alpha)$ *is the* $(1-\alpha)$*-quantile of* $\mathscr{L}(\operatorname{dist}(W, \mathbf{K}))$.

More precise information about the asymptotic behavior of our test can be obtained in a local asymptotic framework.

PROPOSITION 3. *Suppose that*

$$\Sigma_n = I + n^{-1/2}\Delta_n \qquad \text{such that} \quad \Delta_n \to \Delta \in \mathbf{X}. \tag{15}$$

Then

$$nt_{\mathbf{K}}(\Sigma_n) \to \operatorname{dist}(\Delta, \mathbf{K})^2/2,$$
$$nt_{\mathbf{K}}(S_n) \to \operatorname{dist}(\Delta + W, \mathbf{K})^2/2 \qquad \text{in distribution},$$
$$\mathbb{P}\{t_{\mathbf{K}}(S_n) > \hat{c}_n(\alpha)\} \to \mathbb{P}\{\operatorname{dist}(\Delta + W, \mathbf{K}) > \gamma(\alpha)\}.$$

In other words, (15) leads to a simple shift model, where one observes $Z = \Delta + W$ with unknown $\Delta \in \mathbf{X}$ and unobserved Gaussian error $W \in \mathbf{X}$, whose distribution is known. Then the hypothesis $\Delta \in \mathbf{K}$ is rejected if $\operatorname{dist}(Z, \mathbf{K}) > \gamma(\alpha)$. It follows from a general theorem of Stein [12] that this test is admissible among all tests of $\Delta \in \mathbf{K}$ versus $\Delta \in \mathbf{X}\setminus\mathbf{K}$ at level α. In addition, suppose that $\operatorname{Var}(\operatorname{trace}(XB))/\|X\|^2$ is constant in $X \in \mathbf{X}\setminus\{0\}$ (which is true under (1) with constant 2). Then a variation of Stein's [12] arguments shows that

$$\frac{1 - \inf_{\operatorname{dist}(\Delta, \mathbf{K}) \geqslant \delta} \mathbb{E}(\phi(\Delta + W))}{1 - \inf_{\operatorname{dist}(\Delta, \mathbf{K}) \geqslant \delta} \mathbb{P}\{\operatorname{dist}(\Delta + W, \mathbf{K}) > \gamma(\alpha)\}} \to \infty \qquad \text{as} \quad \delta \to \infty$$

for any test $\phi: \mathbf{X} \to [0, 1]$ such that $\mathbb{E}(\phi(W)) \leqslant \alpha$ and $\phi(W) \neq \mathbf{1}\{\operatorname{dist}(W, \mathbf{K}) > \gamma(\alpha)\}$ with positive probability; see Dümbgen [4]. Thus the test $\mathbf{1}\{t_{\mathbf{K}}(S_n) > \hat{c}_n(\alpha)\}$ is locally asymptotically admissible and minimax in a vague sense.

A Nonparametric Example for Conditions (11) *and* (13). Let S_n be the sample covariance matrix of independent, identically distributed random vectors $\mathbf{y}_{0n}, \mathbf{y}_{n1}, ..., \mathbf{y}_{nn} \in \mathbf{R}^d$, i.e.,

$$S_n := n^{-1} \sum_{i=0}^{n} (\mathbf{y}_{in} - \bar{\mathbf{y}}_n)(\mathbf{y}_{in} - \bar{\mathbf{y}}_n)', \qquad \mathbf{y}_n := (n+1)^{-1} \sum_{i=0}^{n} \mathbf{y}_{in}.$$

It is assumed that the distribution of $\mathbf{y}_{0n}$ is completely unknown but has finite fourth moments and nonsingular covariance matrix Σ_n. Here $\tilde{S}_n$ is just the sample covariance matrix of the standardized vectors

$$\tilde{\mathbf{y}}_{in} := \Sigma_n^{-1/2}(\mathbf{y}_{in} - \mathbb{E}(\mathbf{y}_{0n})), \qquad 0 \leqslant i \leqslant n.$$

This motivates the following bootstrap estimator for $\mathscr{L}(\tilde{S}_n)$: Let $\hat{S}_n$ be the sample covariance matrix of the random vectors

$$\hat{\mathbf{y}}_{in} := S_n^{-1/2}(\mathbf{y}_{I(i),n} - \bar{\mathbf{y}}_n), \qquad 0 \leqslant i \leqslant n,$$

where $I(0), I(1), ..., I(n)$, $\mathbf{y}_{0n}, \mathbf{y}_{1n}, ..., \mathbf{y}_{nn}$ are independent, and each $I(i)$ is uniformly distributed on $\{0, 1, ..., n\}$. Then define

$$\hat{D}_n := \mathscr{L}(\hat{S}_n \mid \mathbf{y}_{0n}, \mathbf{y}_{1n}, ..., \mathbf{y}_{nn}).$$

Under mild regularity conditions on $\mathscr{L}(\tilde{\mathbf{y}}_{0n})$, assumptions (11) and (13) are satisfied regardless of the sequence $(\Sigma_n)_{n \geqslant d}$. Suppose that for some random vector $\tilde{\mathbf{y}} \in \mathbf{R}^d$,

$$\mathscr{L}(\tilde{\mathbf{y}}_{0n}) \to \mathscr{L}(\tilde{\mathbf{y}}) \quad \text{weakly}, \qquad \mathbb{E}(\|\tilde{\mathbf{y}}_{0n}\|^4) \to \mathbb{E}(\|\tilde{\mathbf{y}}\|^4) < \infty. \tag{16}$$

Then W_n converges in distribution to a centered Gaussian random matrix $W \in \mathbf{X}$ with covariances

$$\mathrm{Cov}(W_{ij}, W_{kl}) = \mathrm{Cov}(\tilde{\mathbf{y}}_i \tilde{\mathbf{y}}_j, \tilde{\mathbf{y}}_k \tilde{\mathbf{y}}_l);$$

see Beran and Srivastava [3]. In particular, $S_n \in \mathbf{X}_+$ with probability tending to one. Moreover, the distribution of W is nonsingular, unless $\tilde{\mathbf{y}}' X \tilde{\mathbf{y}}$ is constant almost surely for some fixed $X \in \mathbf{X} \setminus \{0\}$. The consistency of $\hat{D}_n$ can be verified as follows: In view of (16) it suffices to show that

$$\begin{aligned} &\mathscr{L}(\hat{\mathbf{y}}_{0n} \mid \mathbf{y}_{0n}, \mathbf{y}_{1n}, ..., \mathbf{y}_{nn}) \to \mathscr{L}(\tilde{\mathbf{y}}) && \text{weakly in probability,} \\ &\mathbb{E}(\|\hat{\mathbf{y}}_{0n}\|^4 \mid \mathbf{y}_{0n}, \mathbf{y}_{1n}, ..., \mathbf{y}_{nn}) \to \mathbb{E}(\|\tilde{\mathbf{y}}\|^4) && \text{in probability.} \end{aligned} \tag{17}$$

Since $S_n^{-1/2}\Sigma_n^{1/2} = I + O_p(n^{-1/2})$ and $\Sigma_n^{-1/2}(\bar{\mathbf{y}}_n - \mathbb{E}(\mathbf{y}_{0n})) = O_p(n^{-1/2})$, one may replace $\hat{\mathbf{y}}_{0n}$ in (17) with the resampling variable $\tilde{\mathbf{y}}_{I(0),n}$. But then (17) follows via standard arguments.

4. Analytical Tools

THEOREM 3. *For $X, Y \in \mathbf{X}_+$ let*

$$B = B(X, Y) := Y^{-1/2} X Y^{-1/2} - I.$$

Then

$$\|Y^{-1/2}X^{1/2}-I\|^2,\ \|X^{-1/2}Y^{1/2}-I\|^2 \leqslant \frac{\|B\|^2}{1+\lambda_d(B)}; \tag{18}$$

$$2t(X,Y)\in\left[\frac{\|B\|^2}{1+\max\{\lambda_1(B),0\}},\frac{\|B\|^2}{1+\min\{\lambda_d(B),0\}}\right]. \tag{19}$$

Since W_n in Section 3 equals $n^{1/2}B(S_n,\Sigma_n)$, the consistency of S_n, Eq. (12), follows directly from (18). An interesting fact in connection with (18) is that $B(X,Y)\to 0$ does not imply $Y^{-1}X\to I$. One can easily construct a counterexample for $d=2$.

Theorem 3 implies that $\sqrt{t(\cdot,\cdot)}$ behaves almost like a metric on $\mathbf{X}_+$. Precisely, it follows from (18) that

$$\begin{aligned}
B(Y,X) &= X^{-1/2}(Y-X)X^{-1/2}\\
&= -X^{-1/2}Y^{1/2}B(X,Y)Y^{1/2}X^{-1/2}\\
&= -B(X,Y)+O(\|B(X,Y)\|^2),\\
B(X,Z) &= Z^{-1/2}Y^{1/2}B(X,Y)Y^{1/2}Z^{-1/2}+B(Y,Z)\\
&= B(X,Y)+B(Y,Z)+O(\|B(X,Y)\|^2+\|B(Y,Z)\|^2)
\end{aligned}$$

as $B(X,Y)$, $B(Y,Z)\to 0$. Combined with (19) this yields

$$\begin{aligned}
&\sqrt{t(Y,X)}=\sqrt{t(X,Y)}+O(t(X,Y)),\\
&\sqrt{t(X,Z)}\leqslant\sqrt{t(X,Y)}+\sqrt{t(Y,Z)}+O(t(X,Y)+t(Y,Z))
\end{aligned}$$

as $t(X,Y)+t(Y,Z)\to 0$.

THEOREM 4. *For* $Y\in\mathbf{X}_+$ *let*

$$\tilde{t}_{\mathbf{K}}(Y):=\max_{U\in\mathbf{T}(Y),\,V\in\mathbf{T}_Q(Y)}\sum_{i,j=1}^{d}\frac{(\lambda_i-\lambda_j)^2}{\lambda_i\lambda_j}(Y)(V'U)_{ij}^2/2.$$

Then

$$\sqrt{\frac{\tilde{t}_{\mathbf{K}}(Y)}{t_{\mathbf{K}}(Y)}}\to[1,2] \qquad \textit{as} \quad \min\{t_{\mathbf{K}}(Y),\tilde{t}_{\mathbf{K}}(Y)\}\to 0. \tag{20}$$

For $X\in\mathbf{X}_+$ *and* $B=B(X,Y)$,

$$\sqrt{2t_{\mathbf{K}}(I+B)}=\operatorname{dist}(B,\mathbf{K})+O(\|B\|^2) \tag{21}$$

and

$$\sqrt{2t_{\mathbf{K}}(X)} \leqslant \sqrt{2t_{\mathbf{K}}(Y)} + \text{dist}(B, \mathbf{K}) + O(t_{\mathbf{K}}(Y) + \|B\|^2) \qquad (22)$$

as $t_{\mathbf{K}}(Y) + \|B\|^2 \to 0$.

The first part of Theorem 4 yields the equivalence of (14) and the boundedness of $nt_{\mathbf{K}}(\Sigma_n)$. Equation (21) implies Propositions 1 and 3. The continuity of the distribution of dist$(W, \mathbf{K})$ follows from the nonsingularity of $\mathscr{L}(W)$ together with the fact that each set $\{X \in \mathbf{X}: \text{dist}(X, \mathbf{K}) = r\}$ is contained in the boundary of a convex set in $\mathbf{X}$. As for Proposition 2, one may assume without loss of generality that $\sqrt{2nt_{\mathbf{K}}(\Sigma_n)} \to \delta \in [0, \infty]$. Since $B(S_n, \Sigma_n) = n^{-1/2} W_n = -B(\Sigma_n, S_n) + O_p(n^{-1})$, one can apply the inequality (22) to $(X, Y) = (S_n, \Sigma_n)$ and $(X, Y) = (\Sigma_n, S_n)$ to show that

$$\sqrt{2nt_{\mathbf{K}}(S_n)} \begin{cases} \to_p \infty & \text{if } \delta = \infty, \\ \geqslant \delta - \text{dist}(-W_n, \mathbf{K}) + o_p(1) & \text{if } \delta < \infty. \end{cases}$$

Since $\hat{c}_n(\alpha) \to_p \gamma(\alpha)$ and $-W_n \to W$ in distribution, this yields Proposition 2.

Proof of Theorem 3. Let $L := \text{diag}(\lambda(X))$, $M := \text{diag}(\lambda(Y))$, and $T \in \mathbf{T}(X)$, $U \in \mathbf{T}(Y)$. Then

$$\begin{aligned}
\|Y^{-1/2}X^{1/2} - I\|^2 &= \|U'(Y^{-1/2}X^{1/2} - I)T\|^2 \\
&= \|M^{-1/2}U'TL^{1/2} - U'T\|^2 \\
&= \sum_{i,j=1}^{d} (\sqrt{\lambda_j(X)/\lambda_i(Y)} - 1)^2 (U'T)_{ij}^2 \\
&\leqslant \sum_{i,j=1}^{d} (\sqrt{\lambda_j(X)/\lambda_i(Y)} - \sqrt{\lambda_i(Y)/\lambda_j(X)})^2 (U'T)_{ij}^2 \\
&= \|Y^{-1/2}X^{1/2} - Y^{1/2}X^{-1/2}\|^2 \\
&= \text{trace}(Y^{-1/2}XY^{-1/2} + (Y^{-1/2}XY^{-1/2})^{-1} - 2I) \\
&= \sum_{i=1}^{d} \lambda_i(B)^2/(1 + \lambda_i(B)) \\
&\in [\|B\|^2/(1 + \lambda_1(B)), \|B\|^2/(1 + \lambda_d(B))].
\end{aligned}$$

Since $\|Y^{-1/2}X^{1/2} - Y^{1/2}X^{-1/2}\|$ is symmetric in X and Y, one can interchange them and obtain (18).

Elementary calculations show that

$$r^2/(1 + \max\{r, 0\}) \leqslant 2(r - \log(1 + r)) \leqslant r^2/(1 + \min\{r, 0\})$$

for all $r > -1$, and thus (19) follows from the identity

$$t(X, Y) = \sum_{i=1}^{d} (\lambda_i(B) - \log(1 + \lambda_i(B))). \quad \blacksquare$$

Proof of Theorem 4. For $M := \mathrm{diag}(\lambda(Y))$ and suitable $V \in \mathbf{T}_Q(Y)$ one can write

$$\begin{aligned}\sqrt{\tilde{t}_{\mathbf{K}}(Y)} &= 2^{-1/2} \|(VMV')^{-1/2} Y^{1/2} - (VMV')^{1/2} Y^{-1/2}\| \\ &= \sqrt{t(Y, VMV')}\,(1 + o(1)) \\ &\geqslant \sqrt{t_{\mathbf{K}}(Y)}\,(1 + o(1))\end{aligned}$$

as $\min\{t(Y, VMV'), \tilde{t}_{\mathbf{K}}(Y)\} \to 0$; see the proof of Theorem 3. On the other hand, $t_{\mathbf{K}}(Y) = t(Y, Z)$ for some $Z \in \mathbf{K} \cap \mathbf{X}_+$ such that $V \in \mathbf{T}(Z)$, by Theorem 2. Then inequality (10) implies that

$$t(Y, Z) \geqslant t(VMV', Z).$$

Consequently,

$$\begin{aligned}\sqrt{t(Y, VMV')} &\leqslant \sqrt{t(Y, Z)} + \sqrt{t(VMV', Z)} + O(t(Y, Z) + t(VMV', Z)) \\ &\leqslant 2\sqrt{t_{\mathbf{K}}(Y)} + O(t_{\mathbf{K}}(Y))\end{aligned}$$

as $t_{\mathbf{K}}(Y) \to 0$, and (20) follows.

It follows from Theorem 2 that $t_{\mathbf{K}}(I + B) = t(I + B, I + C)$, where $C \in \mathbf{K}$ such that $\|B - C\| = \mathrm{dist}(B, \mathbf{K})$. Moreover, $\|C\| \leqslant \|B\| + \mathrm{dist}(B, \mathbf{K}) \leqslant 2\|B\|$, so that

$$B(I + B, I + C) = B - C + O(\|B\|^2)$$

as $B \to 0$, and (21) follows from (19).

If $\rho := \|B\|^2 + t_{\mathbf{K}}(Y) \to 0$, then

$$\begin{aligned}\sqrt{2t_{\mathbf{K}}(X)} &= \sqrt{2t_{\mathbf{K}}(Z^{1/2}(I + B(X, Z))\,Z^{1/2})} \\ &\leqslant \sqrt{2t_{\mathbf{K}}(I + B(X, Z))} \\ &= \sqrt{2t_{\mathbf{K}}(I + B + B(Y, Z) + O(\rho))} \\ &= \mathrm{dist}(B + B(Y, Z), \mathbf{K}) + O(\rho) \\ &\leqslant \mathrm{dist}(B, \mathbf{K}) + \|B(Y, Z)\| + O(\rho) \\ &= \mathrm{dist}(B, \mathbf{K}) + \sqrt{2t_{\mathbf{K}}(Y)} + O(\rho),\end{aligned}$$

where the first inequality follows from Theorem 1. $\blacksquare$

5. Concluding Remarks

The test hypothesis discussed here can be viewed as a special case of the more general hypothesis $\Sigma \in I + \tilde{\mathbf{K}}$, where $\tilde{\mathbf{K}}$ is any closed, convex cone in $\mathbf{X}$. The hypothesis H_1 of Kuriki [6] is of this type, where $\tilde{\mathbf{K}}$ is the cone of nonnegative definite matrices in $\mathbf{X}$. Similarly to Sections 2 and 3 of the present paper, one could extend Kuriki's [6] tests and some of his results to situations where (4) or only (11, 13) hold.

An interesting open question is whether the present approach of looking for a least favorable parameter Σ can be applied to other testing problems. It is conjectured that the general hypothesis in Tyler [13, 14] or Schott [10, 11] can be treated similarly.

Acknowledgments

I thank the referees for constructive criticism and useful references.

References

[1] Anderson, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Statist.* **34** 122–148.

[2] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York.

[3] Beran, R., and Srivastava, M. S. (1985). Bootstrap test and confidence regions for functions of a covariance matrix. *Ann. Statist.* **13** 95–115.

[4] Dümbgen, L. (1993). Minimax tests for convex cones, *Ann. Inst. Statist. Math.*, to appear.

[5] Hoffman, A. J., and Wielandt, H. W. (1953). The variation of the spectrum of a normal matrix. *Duke Math. J.* **20** 37–39.

[6] Kuriki, S. (1993). Likelihood ratio tests for covariance structure in random effects models. *J. Multivariate Anal.* **46** 175–197.

[7] von Neumann, J. (1937). Some matrix-inequalities and metrization of matrix-space. *Tomsk Univ. Rev.* **1** 283–300. Also in (1962) *John von Neumann Collected Works* (A. H. Taub, Ed.), Vol. 4, pp. 205–219, Pergamon, Oxford.

[8] Robertson, T., Wright, F. T., and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.

[9] Schott, J. R. (1989). An adjustment for a test concerning a principal component subspace. *Statist. Probab. Lett.* **7** 425–430.

[10] Schott, J. R. (1991). Testing for the redundancy of variables in principal component analysis. *Statist. Probab. Lett.* **11** 495–501.

[11] Schott, J. R. (1991). Some test for common principal component subspaces in several groups. *Biometrika* **78** 771–777.

[12] Stein, C. (1956). The admissibility of Hotelling's T^2-test. *Ann. Math. Statist.* **27** 616–623.

[13] Tyler, D. E. (1981). Asymptotic inference for eigenvectors. *Ann. Statist.* **9** 725–736.

[14] Tyler, D. E. (1983). A class of asymptotic tests for principle component vectors. *Ann. Statist.* **11** 1243–1250.