Ann. Inst. Statist. Math.
Vol. 45, No. 1, 147–158 (1993)

AN APPROXIMATE TEST FOR COMMON PRINCIPAL COMPONENT SUBSPACES IN TWO GROUPS

TERUO FUJIOKA

Department of Mathematics, Faculty of Science, Hiroshima University, Hiroshima 730, Japan

(Received March 27, 1990; revised April 27, 1992)

Abstract. Principal component analysis has made an important contribution to data reduction. In two sample problems, one great interest is whether we can reduce the number of variables to a smaller number in similar fashions for both samples. More precisely, we consider the hypothesis H_m that the subspaces spanned by the latent vectors of the population covariance matrices corresponding to the first principal components are the same in two groups. In this paper, we propose a simple and easily interpreted test procedure for H_m.

Key words and phrases: Data reduction, latent root, latent vector, principal component, conditional Haar distribution.

1. Introduction

Let $x^{(g)}$ $(g = 1, 2)$ be vector variables with p components. In practical applications, the same p variables are being measured on objects in different but related two groups. The data reduction in principal component analysis is done by using the first m principal components whose coefficient vectors are orthonormal with relatively large variabilities in each groups. We treat the hypothesis H_m that the subspaces spanned by the m latent vectors of the population covariance matrices corresponding to the first m principal components are the same in two groups. Krzanowski (1979, 1982) discussed the similarity measure of two subspaces, its geometric interpretation and the simulation studies instead of considering the asymptotic distribution. In this paper, the test procedure proposed by Krzanowski is modified so that it is available for multiple roots cases. Schott (1988) has also given a procedure for testing H_m for general cases. His test criterion, however, depends on the rotational freedom in the hypothesis H_m, which is not uniquely determined for multiple roots cases. In testing the hypothesis H_m, the rotational freedom in H_m is an obstacle to constructing similar tests, even asymptotically. In order to get over these difficulties, one attempt is the approximation using the maximization on that sorts of freedom described afterwards.

Flury (1984, 1986, 1987) has developed the common principal components (CPC) method and recently partial CPC method. However, his common subspace does not correspond to the first m principal components in each groups. Chen and Robinson (1989) treated test statistics for a common factor space in factor analysis. Their hypothesis is generally different from H_m when restricting to PCA, but their test statistic is the same as ours only if we consider the testing probrem of H_m when the first m latent roots are all distinct and the remainders are all the same in each groups.

A test statistic which underlies our test procedure is described in Section 2, and its asymptotic distribution for general case is derived in Section 3. We complete our test procedure by using an approximation of the critical points of our statistic in Section 4. The approximate critical points of our statistic and actual ones obtained by simulation study are tabulated in Table 1 and the influence of the maximization in approximating the critical points is examined in Table 2. An illustrative example is given in Section 5.

2. The test statistics

Assume that $x_1^{(g)}, \dots, x_{N_g}^{(g)}$ $(g = 1, 2)$ are random samples of independent random vectors $x^{(g)}$ normally distributed with covariance matrices $\Sigma^{(g)}$. Write $\Sigma^{(g)}$ in spectral decomposition as $\Sigma^{(g)} = \Gamma^{(g)}\Delta^{(g)}\Gamma^{(g)\prime}$ where $\Delta^{(g)} = \mathrm{diag}(\delta_1^{(g)}, \dots, \delta_p^{(g)})$ with $\delta_1^{(g)} \geq \cdots \geq \delta_p^{(g)}$, $\Gamma^{(g)} = [\Gamma_1^{(g)} : \Gamma_2^{(g)}] = [\gamma_1^{(g)}, \dots, \gamma_m^{(g)} : \gamma_{m+1}^{(g)}, \dots, \gamma_p^{(g)}]$, and $I_p = \Gamma^{(g)}\Gamma^{(g)\prime}$. Then the hypothesis H_m means that $\Re(\Gamma_1^{(1)}) = \Re(\Gamma_1^{(2)})$. Here I_p is the $p \times p$ identity matrix, and $\Re(\Gamma_1^{(g)})$ denotes the subspace spanned by the column vectors of $\Gamma_1^{(g)}$. Using the fact that $\Re(\Gamma_1^{(g)})^{\perp} = \Re(\Gamma_2^{(g)})$, it is easy to show that H_m is equivalent to $\Re(\Gamma_2^{(1)}) \perp \Re(\Gamma_1^{(2)})$ or $\Gamma_2^{(1)\prime}\Gamma_1^{(2)} = O$. Consequently, we obtain the prominent property:

$$\mathrm{tr}\,\Gamma_2^{(1)\prime}\Gamma_1^{(2)}\Gamma_1^{(2)\prime}\Gamma_2^{(1)} \begin{cases} = 0 & \text{if } H_m \text{ is true,} \\ > 0 & \text{otherwise.} \end{cases}$$

Now we constitute the test criterion based on this property. It is natural to adopt the sample covariance matrices $S^{(g)}$ as the estimators of $\Sigma^{(g)}$. Then $n_g S^{(g)}$ is distributed as a Wishart distribution $W_p(n_g, \Sigma^{(g)})$, where $n_g = N_g - 1$. Decompose $S^{(g)}$ in the same manner as $\Sigma^{(g)}$: $S^{(g)} = C^{(g)}D^{(g)}C^{(g)\prime}$ where $D^{(g)} = \mathrm{diag}(d_1^{(g)}, \dots, d_p^{(g)})$ with $d_1^{(g)} \geq \cdots \geq d_p^{(g)}$, $C^{(g)} = [C_1^{(g)} : C_2^{(g)}] = [c_1^{(g)}, \dots, c_m^{(g)} : c_{m+1}^{(g)}, \dots, c_p^{(g)}]$ and $I_p = C^{(g)}C^{(g)\prime}$. When $\delta_m^{(g)} > \delta_{m+1}^{(g)}$ $(g = 1, 2)$, $C_1^{(2)}C_1^{(2)\prime}$ and $C_2^{(1)}C_2^{(1)\prime}$ are consistent estimators of $\Gamma_1^{(2)}\Gamma_1^{(2)\prime}$ and $\Gamma_2^{(1)}\Gamma_2^{(1)\prime}$, respectively, due to the general results given in Chapter 2 of Kato (1966). This implies that if H_m is true,

$$T_m = \mathrm{tr}\,C_2^{(1)\prime}C_1^{(2)}C_1^{(2)\prime}C_2^{(1)} \tag{2.1}$$

should not be too large. Therefore, we propose T_m as a test statistic for H_m.

It is easily shown that

$$\begin{aligned} \mathrm{tr}\,\Gamma_2^{(1)\prime}\Gamma_1^{(2)}\Gamma_1^{(2)\prime}\Gamma_2^{(1)} &= \mathrm{tr}\,\Gamma_1^{(1)\prime}\Gamma_2^{(2)}\Gamma_2^{(2)\prime}\Gamma_1^{(1)} \\ &= m - \mathrm{tr}\,\Gamma_1^{(1)\prime}\Gamma_1^{(2)}\Gamma_1^{(2)\prime}\Gamma_1^{(1)} \\ &= p - m - \mathrm{tr}\,\Gamma_2^{(1)\prime}\Gamma_2^{(2)}\Gamma_2^{(2)\prime}\Gamma_2^{(1)}. \end{aligned} \tag{2.2}$$

The quantity $\operatorname{tr}\Gamma_1^{(1)\prime}\Gamma_1^{(2)}\Gamma_1^{(2)\prime}\Gamma_1^{(1)}$ has been proposed as a measure of total similarity between $\Re(\Gamma_1^{(1)})$ and $\Re(\Gamma_1^{(2)})$ in Krzanowski (1979). We can also give a geometric interpretation that the quantity $\operatorname{tr}\Gamma_2^{(1)\prime}\Gamma_1^{(2)}\Gamma_1^{(2)\prime}\Gamma_2^{(1)}$ is equal to the sum of squares of the cosines of the angles between each of the column vectors in $\Gamma_2^{(1)}$ and each one in $\Gamma_1^{(2)}$. Note that the identities replaced $\Gamma_i^{(g)}$ by $C_i^{(g)}$ in (2.2) are all available as the definition of T_m and convenient to diagnose the asymptotic null distribution of T_m given in the next section.

3. The asymptotic null distribution of T_m

Under H_m, we have the relation

$$\Gamma^{(2)} = \Gamma^{(1)} \begin{bmatrix} Q & O \\ O & R \end{bmatrix},$$

called the rotational freedom in H_m, where $Q = \Gamma_1^{(1)\prime}\Gamma_1^{(2)}$ and $R = \Gamma_2^{(1)\prime}\Gamma_2^{(2)}$ are $m \times m$, $(p-m) \times (p-m)$ orthogonal matrices, respectively. Let $n_g = \kappa_g n$ where κ_g are fixed positive numbers satisfying $\kappa_1 + \kappa_2 = 1$ and let $\bar{n} = n_1 n_2/(n_1 + n_2)$. We investigate the limiting distribution of $\bar{n}T_m$ as n tends to infinity with the help of the results on the asymptotic behavior of $C^{(g)}$ summarized in Anderson (1963).

Assume that $\delta_m^{(g)} > \delta_{m+1}^{(g)}$ $(g = 1, 2)$ so that H_m may make sense. We first study the limiting distribution of $\bar{n}T_m$ for simple roots case. Put $U^{(g)} = \Gamma^{(g)\prime}C^{(g)}$, $U^{(g)} = (u_{ij}^{(g)})$, $1 \le i, j \le p$. Then $u_{ii}^{(g)}$ and $u_{ij}^{(g)}$ for $i > j$ are independent in the limiting distribution. The $n_g^{1/2}u_{ij}^{(g)}$ for $i > j$ are asymptotically independently distributed as normal with mean 0, variance $\delta_i^{(g)}\delta_j^{(g)}/(\delta_i^{(g)} - \delta_j^{(g)})^2$, and $u_{ii}^{(g)}$ and $u_{ij}^{(g)} - u_{ji}^{(g)}$ for $i \ne j$ converge stochastically to 1 and 0, respectively.

Partitioning the matrices $U^{(g)}$ into submatrices with m, $p-m$ rows and columns,

$$U^{(g)} = \begin{bmatrix} U_{11}^{(g)} & U_{12}^{(g)} \\ U_{21}^{(g)} & U_{22}^{(g)} \end{bmatrix},$$

we get

$$\bar{n}^{1/2}C_2^{(1)\prime}C_1^{(2)} = n_1^{1/2}U_{12}^{(1)\prime} \cdot \kappa_2^{1/2}QU_{11}^{(2)} + \kappa_1^{1/2}U_{22}^{(1)\prime}R \cdot n_2^{1/2}U_{21}^{(2)}. \tag{3.1}$$

The elements of $n_1^{1/2}U_{12}^{(1)\prime}$ and $n_2^{1/2}U_{21}^{(2)}$ in (3.1) are asymptotically normally distributed; $U_{22}^{(1)\prime}$ and $U_{11}^{(2)}$ converge stochastically to I_m and I_{p-m}, respectively.

In order to describe the distribution of matrices, we use the following notations: if A is $a \times t$ matrix, then $\operatorname{vec}(A)$ is the at dimensional vector formed by stacking the columns of A, while if B is $b \times u$ matrix, then $A \otimes B$ is the $ab \times tu$ Kronecker product of B and C. We use the following three important properties: $\operatorname{vec}(ABC) = (C' \otimes A)\operatorname{vec}(B)$, $(A \otimes B)(C \otimes D) = AC \otimes BD$ and $(A \otimes B)' = A' \otimes B'$. Then (3.1) can be rewritten as

$$\begin{aligned} \operatorname{vec}(\bar{n}^{1/2}C_2^{(1)\prime}C_1^{(2)}) = {} & \kappa_2^{1/2}(U_{11}^{(2)\prime}Q' \otimes I_{p-m})\operatorname{vec}(n_1^{1/2}U_{12}^{(1)\prime}) \\ & + \kappa_1^{1/2}(I_m \otimes U_{22}^{(1)\prime}R)\operatorname{vec}(n_2^{1/2}U_{21}^{(2)}). \end{aligned} \tag{3.2}$$

Therefore, the limiting distribution of $\mathrm{vec}(\bar{n}^{1/2}C_2^{(1)\prime}C_1^{(2)})$ as $n \to \infty$ is normal with mean 0, and covariance matrix,

$$\Psi = \kappa_2(Q' \otimes I_{p-m})\Lambda^{(1)}(Q \otimes I_{p-m}) + \kappa_1(I_m \otimes R)\Lambda^{(2)}(I_m \otimes R'), \tag{3.3}$$

with

$$\Lambda^{(g)} = \mathrm{diag}(\Lambda_1^{(g)}, \ldots, \Lambda_m^{(g)}), \qquad g = 1, 2, \tag{3.4}$$

$$\Lambda_i^{(g)} = \mathrm{diag}(\delta_i^{(g)}\delta_{m+1}^{(g)}/(\delta_i^{(g)} - \delta_{m+1}^{(g)})^2, \ldots, \delta_i^{(g)}\delta_p^{(g)}/(\delta_i^{(g)} - \delta_p^{(g)})^2), \quad g = 1, 2, \quad i = 1, \ldots, m. \tag{3.5}$$

On the other hand, since $\bar{n}T_m = (\mathrm{vec}(\bar{n}^{1/2}C_2^{(1)\prime}C_1^{(2)}))'(\mathrm{vec}(\bar{n}^{1/2}C_2^{(1)\prime}C_1^{(2)}))$, we have the following theorem.

THEOREM 3.1. *For simple roots case, the limiting distribution of $\bar{n}T_m$ as $n \to \infty$ is the same as the distribution of the sum of squares of normal variates with mean* 0 *and covariance matrix* Ψ.

Next, we consider the limiting distribution of $\bar{n}T_m$ as $n \to \infty$ under H_m for multiple roots case. Due to Anderson (1963), the $n_1^{1/2}U_{12}^{(1)\prime}$, $U_{11}^{(2)}$, $n_2^{1/2}U_{21}^{(2)}$ and $U_{22}^{(1)}$ in (3.2) are independent in the limiting distribution. The $U_{11}^{(2)}$ and $U_{22}^{(1)}$ have the limiting distributions on orthogonal matrices according to the multiplicities of the roots; the $n_1^{1/2}U_{12}^{(1)\prime}$ and $n_2^{1/2}U_{21}^{(2)}$ have the same distributions as given for simple roots case. The $QU_{11}^{(2)}$ and $U_{22}^{(1)\prime}R$ converge to V_1 and V_2 in distribution, respectively, where V_1 is distributed on the set of $m \times m$ orthogonal matrices, and V_2 is distributed on the set of $(p-m) \times (p-m)$ orthogonal matrices. The conditional limiting distribution of $\mathrm{vec}(\bar{n}^{1/2}C_2^{(1)\prime}C_1^{(2)})$ as $n \to \infty$ giving V_1 and V_2 is normal with mean 0 and covariance matrix $\tilde{\Psi}$, where $\tilde{\Psi}$ is defined by replacing Q and R by V_1 and V_2, respectively, in Ψ. Thus, we have the following theorem.

THEOREM 3.2. *For multiple roots case, the limiting distribution of $\bar{n}T_m$ as $n \to \infty$ is the same as that of the random variable $X'X$ with*

$$X = \kappa_2^{1/2}(V_1 \otimes I_{p-m})Z_1 + \kappa_1^{1/2}(I_m \otimes V_2)Z_2,$$

where V_1 is distributed on the set of $m \times m$ orthogonal matrices, V_2 is distributed on the set of $(p-m) \times (p-m)$ orthogonal matrices, Z_1 is $N_{m(p-m)}(0, \Lambda^{(1)})$ and Z_2 is $N_{m(p-m)}(0, \Lambda^{(2)})$, and V_1, V_2, Z_1 and Z_2 are all independent.

In observing the limiting null distribution of $\bar{n}T_m$, it is convenient to give another expression for it. From the identities (2.2) with replacing $\Gamma_i^{(g)}$ by $C_i^{(g)}$, we also have $\bar{n}T_m = (\mathrm{vec}(\bar{n}^{1/2}C_2^{(2)\prime}C_1^{(1)}))'(\mathrm{vec}(\bar{n}^{1/2}C_2^{(2)\prime}C_1^{(1)}))$. As in the argument for deriving Theorem 3.2, we have the following theorem.

THEOREM 3.3. *For multiple roots case, the limiting distribution of $\bar{n}T_m$ as $n \to \infty$ is the same as that of the random variable $Y'Y$ with*

$$Y = \kappa_2^{1/2}(I_m \otimes W_2)Z_1 + \kappa_1^{1/2}(W_1 \otimes I_{p-m})Z_2,$$

where W_1 is distributed on the set of $m \times m$ orthogonal matrices, W_2 is distributed on the set of $(p-m) \times (p-m)$ orthogonal matrices, Z_1 is $N_{m(p-m)}(0, \Lambda^{(1)})$ and Z_2 is $N_{m(p-m)}(0, \Lambda^{(2)})$, and W_1, W_2, Z_1 and Z_2 are all independent.

In Theorem 3.2, if the conditional covariance matrix of X giving V_i's, is free from V_i's, then X is normally distributed. In Theorem 3.3, the similar observation shows the normality of Y. Thus, we can clarify that the limiting null distribution of $\bar{n}T_m$ is weighted sum of chi-squared for the cases where the condition "$\delta_1^{(1)} = \cdots = \delta_m^{(1)}$, $\delta_{m+1}^{(2)} = \cdots = \delta_p^{(2)}$", or the conditon "$\delta_1^{(2)} = \cdots = \delta_m^{(2)}$, $\delta_{m+1}^{(1)} = \cdots = \delta_p^{(1)}$" holds. In particular, if the above two conditions hold simultaneously, then $\bar{n}T_m$ multiplied by some constant is asymptotically distributed as chi-squared distribution with $m(p-m)$ degrees of freedom. Notice that the joint use of the two expressions $X'X$ and $Y'Y$ facilitates the above clarification. As can be seen in Theorems 3.2 and 3.3, the limiting null distribution of $\bar{n}T_m$ as $n \to \infty$ is rather complicated because it involves the conditional Haar distribution.

We should have explained the asymptotic null distribution for multiple roots case more precisely to constitute the test criterion. However, only if it is known that V_1, V_2, W_1, and W_2 are orthogonal matrices, the devices proposed in the next section completes the test criterion.

4. A test procedure

The asymptotic null distribution of $\bar{n}T_m$ as $n \to \infty$ is cumbersome in general. It depends on the unknown parameters Q, R and $\delta_i^{(g)}$, $g = 1, 2$, $i = 1, \ldots, p$, and involves the conditional Haar distribution according to the multiplicities of the population roots. Although one thing to do for approximating the distribution of $\bar{n}T_m$ is to substitute estimates of unknown parameters, it is difficult to estimate the multiplicities of the roots, and the substitution cannot work well. Some counterplans for them should be made up minding that the resultant critical region will be conservative. We use $a\chi_f^2$ as an approximation to the null distribution of $\bar{n}T_m$, where a and f are constants and χ_f^2 denotes a chi-squared variable with f degrees of freedom. Then, we determine the critical point so that the test is conservative with respect to the rotational freedom in H_m and the multiplicities of the roots. The unknown parameters $\delta_i^{(g)}$'s are replaced by their estimates. First, the constants a and f are determined so that the mean and variance of $a\chi_f^2$ are the same as the asymptotic mean, say e, and variance, say v, of $\bar{n}T_m$, respectively. This implies that $a = v/2e$, and $f = 2e^2/v$. For simple roots case we have

$$e = \operatorname{tr}\Psi = \kappa_1 \operatorname{tr}\Lambda^{(1)} + \kappa_2 \operatorname{tr}\Lambda^{(2)}, \tag{4.1}$$

$$\begin{aligned} v = 2\operatorname{tr}\Psi^2 &= 2\kappa_1^2 \operatorname{tr}\Lambda^{(1)2} + 2\kappa_2^2 \operatorname{tr}\Lambda^{(2)2} \\ &\quad + 4\kappa_1\kappa_2 \operatorname{tr}(Q \otimes R)'\Lambda^{(1)}(Q \otimes R)\Lambda^{(2)}, \qquad \text{say} \quad v(Q, R). \end{aligned} \tag{4.2}$$

For general cases we have

$$(4.3)\quad e = E[\operatorname{tr}\Psi] = \kappa_2 \operatorname{tr}\Lambda^{(1)} + \kappa_1 \operatorname{tr}\Lambda^{(2)},$$

$$\begin{aligned}(4.4)\quad v &= E[2\operatorname{tr}\Psi^2] = E[v(V_1, V_2)] \\ &= 2\kappa_2^2 \operatorname{tr}\Lambda^{(1)2} + 2\kappa_1^2 \operatorname{tr}\Lambda^{(2)2} + 4\kappa_1\kappa_2 E[\operatorname{tr}(V_1 \otimes V_2)'\Lambda^{(1)}(V_1 \otimes V_2)\Lambda^{(2)}],\end{aligned}$$

where the expectation is taken with respect to V_1 and V_2. The multiplicities of the roots are generally unknown and cannot be easily estimated, it seems complicated to compute v explicitly even though the multiplicities are known.

Now we propose a test procedure based on $\bar{n}T_m$ by determining the critical point r so that $P(a\chi_f^2 > r \mid H_m)$ is maximized with respect to the orthogonal matrices Q, R for simple roots case. Since, as is described later, the maximum of $P(a\chi_f^2 > r \mid H_m)$ is obtained when $v = v(Q, R)$ is maximized, the inequality

$$(4.5)\qquad \min_{Q,R} v(Q, R) \le E[v(V_1, V_2)] \le \max_{Q,R} v(Q, R)$$

implies that the test procedure proposed above will do well for general cases. Note that (4.5) is valid for all possible multiplicities of the roots.

From (4.1), (4.2), the asymptotic mean of $\bar{n}T_m$, e, is independent of Q, R, and the asymptotic variance of $\bar{n}T_m$, $v(Q, R)$, depends on Q, R. The following facts show that the maximum of $P(a\chi_f^2 > r \mid H_m)$ is attained when $v(Q, R)$ is maximized. The first fact is that the null distribution of $\bar{n}T_m$ can be approximated as $e\chi_f^2/f$ where $f = 2e^2/v(Q, R)$. The second one is that the percent point of bZ is equal to b times the percent point of Z for each positive number b and a random variable Z. The last one is that the upper 100α percent point of χ_f^2/f is monotonically decreasing in f for properly small α, which is checked by a table for the percent point of chi-squared distribution but is not analytically proved yet.

THEOREM 4.1. *Let $v(Q, R)$, $\Lambda^{(g)}$ $(g = 1, 2)$ be given in* (4.2), (3.4), (3.5). *Then we get*

$$(4.6)\qquad \max_{Q,R} v(Q, R) = 2\operatorname{tr}(\kappa_2\Lambda^{(1)} + \kappa_1\Lambda^{(2)})^2,$$

and the maximum is attained when $Q = I_m$, $R = I_{p-m}$.

PROOF. It suffices to maximize

$$(4.7)\qquad \operatorname{tr}(Q \otimes R)'\Lambda^{(1)}(Q \otimes R)\Lambda^{(2)},$$

subject to orthogonal matrices Q, R. For this purpose, the basical tool is the following results due to Anderson (1963): let A and B be symmetric matrices with latent roots $a_1 \ge \cdots \ge a_p$ and $b_1 \ge \cdots \ge b_p$, respectively, and let H be a $p \times p$ orthogonal matrix. Then $\max_H \operatorname{tr} H'AHB = \sum a_j b_j$.

Let $Q = (q_{ij})$, $1 \le i, j \le m$, (4.7) can be written as

$$(4.8)\qquad \sum\sum q_{ij}^2 \operatorname{tr} R'\Lambda_i^{(1)} R\Lambda_j^{(2)}.$$

Since the $\Lambda_i^{(1)}$ and $\Lambda_j^{(2)}$ are diagonal matrices with decreasingly ordered diagonal elements, all summands in (4.8) simultaneously attain each maximum values when $R = I_{p-m}$. Putting

$$\Pi_k^{(g)} = \operatorname{diag}(\delta_1^{(g)}\delta_{m+k}^{(g)}/(\delta_1^{(g)} - \delta_{m+k}^{(g)})^2, \ldots, \delta_m^{(g)}\delta_{m+k}^{(g)}/(\delta_m^{(g)} - \delta_{m+k}^{(g)})^2), \qquad g = 1, 2, \quad k = 1, \ldots, p-m,$$

the maximum value of (4.7) subject to orthogonal matrices R can be expressed as

$$\sum \operatorname{tr} Q'\Pi_k^{(1)} Q \Pi_k^{(2)}. \tag{4.9}$$

Since the $\Pi_k^{(1)}$ and $\Pi_k^{(2)}$ are diagonal matrices with increasingly ordered diagonal elements, all summands in (4.9) simultaneously attain each maximum values when $Q = I_m$. This completes the proof.

To see the effects of Q and R, it is also important to mind the minimum of $P(a\chi_f^2 > r \mid H_m)$ subject to orthogonal matrices Q, R. As in the proof of Theorem 4.1 above, we can obtain

$$\min_{Q,R} v(Q, R) = 2\operatorname{tr}(\kappa_2 \Lambda^{(1)} + \kappa_1 \Lambda^{(2)*})^2, \tag{4.10}$$

with

$$\Lambda^{(2)*} = \operatorname{diag}(\Lambda_m^{(2)*}, \ldots, \Lambda_1^{(2)*}),$$
$$\Lambda_i^{(2)*} = \operatorname{diag}(\delta_i^{(2)}\delta_p^{(2)}/(\delta_i^{(2)} - \delta_p^{(2)})^2, \ldots, \delta_i^{(2)}\delta_{m+1}^{(2)}/(\delta_i^{(2)} - \delta_{m+1}^{(2)})^2), \qquad i = 1, \ldots, m.$$

The minimum value is attained when

$$Q = \begin{bmatrix} O & & & 1 \\ & & \cdot & \\ & \cdot & & \\ 1 & & & O \end{bmatrix}, \qquad R = \begin{bmatrix} O & & & 1 \\ & & \cdot & \\ & \cdot & & \\ 1 & & & O \end{bmatrix}. \tag{4.11}$$

Although the minimum value is not necessarily attained due to the restriction to $\Gamma^{(g)}$, we can regard it as a lower bound at least. The difference between (4.6) and (4.10) tends to be large when the hypothesis H_m itself is awkward and indistinct, that is, $\delta_m^{(g)}$ and $\delta_{m+1}^{(g)}$ are close to each other. On the contrary, it tends to be small when $\delta_1^{(g)}, \ldots, \delta_m^{(g)}$ are close to each other and $\delta_{m+1}^{(g)}, \ldots, \delta_p^{(g)}$ are close to each other. Further note that (4.10) coincides with (4.6) when $\delta_1^{(g)} = \cdots = \delta_m^{(g)}$ for either of two groups and $\delta_{m+1}^{(g)} = \cdots = \delta_p^{(g)}$ for either of two groups. In Table 1, we compare the critical points of our test with actual ones obtained through a simulation study for the cases $Q = I_m$, $R = I_{p-m}$, where the rotational freedom does not affect, in order to examine the adequacy of the asymptotic chi-squared approximation.

Table 1. The simulated critical points of $\bar{n}T_m$ and approximate ones used actually.

(a) $\kappa_1 = \kappa_2 = 0.5$

$\delta_1^{(1)}$ $\delta_1^{(2)}$	$\delta_2^{(1)}$ $\delta_2^{(2)}$	$\delta_3^{(1)}$ $\delta_3^{(2)}$	α	(n_1, n_2) (40, 40)	(100, 100)	(200, 200)	(400, 400)	r
6	6	6	0.10	5.99	5.58	5.46	5.40	5.35
6	6	6	0.05	6.81	6.28	6.11	6.06	6.00
			0.01	8.59	7.78	7.49	7.43	7.34
8	6	4	0.10	7.22	6.73	6.58	6.53	6.48
8	6	4	0.05	8.33	7.66	7.46	7.41	7.32
			0.01	10.94	9.71	9.37	9.31	9.08
10	6	2	0.10	18.87	21.05	20.52	20.44	20.34
10	6	2	0.05	20.25	25.29	24.32	24.17	23.81
			0.01	21.68	35.94	32.86	32.18	31.12
6	6	6	0.10	6.59	6.13	5.98	5.93	5.88
8	6	4	0.05	7.55	6.93	6.74	6.68	6.60
			0.01	9.76	8.65	8.34	8.27	8.11
6	6	6	0.10	14.83	13.83	13.14	12.85	12.62
10	6	2	0.05	18.04	16.88	15.53	15.10	14.55
			0.01	21.09	25.74	21.31	20.01	18.65

(b) $\kappa_1 = 0.8$, $\kappa_2 = 0.2$

$\delta_1^{(1)}$ $\delta_1^{(2)}$	$\delta_2^{(1)}$ $\delta_2^{(2)}$	$\delta_3^{(1)}$ $\delta_3^{(2)}$	α	(n_1, n_2) (40, 160)	(100, 400)	(200, 800)	(400, 1600)	r
6	6	6	0.10	6.07	5.61	5.48	5.42	5.35
6	6	6	0.05	6.97	6.32	6.16	6.10	6.00
			0.01	9.01	7.83	7.58	7.45	7.34
8	6	4	0.10	7.36	6.77	6.62	6.55	6.48
8	6	4	0.05	8.59	7.77	7.55	7.44	7.32
			0.01	11.64	9.92	9.53	9.38	9.08
10	6	2	0.10	22.86	21.84	21.04	20.67	20.34
10	6	2	0.05	27.89	26.85	25.20	24.51	23.81
			0.01	32.38	41.37	34.81	33.24	31.12
6	6	6	0.10	7.16	6.54	6.37	6.32	6.23
8	6	4	0.05	8.37	7.48	7.26	7.14	7.02
			0.01	11.37	9.51	9.12	8.93	8.67
6	6	6	0.10	21.67	19.37	18.21	17.71	17.24
10	6	2	0.05	27.28	24.20	21.96	21.08	20.10
			0.01	32.42	39.04	30.99	28.51	26.13

Table 2. The influence of the maximization in our test procedure.

$\delta_1^{(1)}$ / $\delta_1^{(2)}$	$\delta_2^{(1)}$ / $\delta_2^{(2)}$	$\delta_3^{(1)}$ / $\delta_3^{(2)}$		(a) $\kappa_1 = \kappa_2 = 0.5$			(b) $\kappa_1 = 0.8$, $\kappa_2 = 0.2$		
			α	0.10	0.05	0.01	0.10	0.05	0.01
6	6	6	r	5.35	6.00	7.34	5.35	6.00	7.34
6	6	6	r^*	5.35	6.00	7.34	5.35	6.00	7.34
			P^*	0.100	0.050	0.010	0.100	0.050	0.010
8	6	4	r	6.48	7.32	9.08	6.48	7.32	9.08
8	6	4	r^*	6.32	7.08	8.67	6.38	7.17	8.82
			P^*	0.086	0.040	0.006	0.091	0.043	0.008
10	6	2	r	20.34	23.82	31.12	20.34	23.82	31.12
10	6	2	r^*	18.25	20.67	25.76	19.08	21.92	27.92
			P^*	0.055	0.019	0.002	0.074	0.031	0.004
6	6	6	r	5.88	6.60	8.11	6.23	7.02	8.67
8	6	4	r^*	5.88	6.60	8.11	6.23	7.02	8.67
			P^*	0.100	0.050	0.010	0.100	0.050	0.010
6	6	6	r	12.62	14.55	18.65	17.24	20.10	26.13
10	6	2	r^*	12.62	14.55	18.65	17.24	20.10	26.13
			P^*	0.100	0.050	0.010	0.100	0.050	0.010
4	2	2	r	34.33	38.90	48.49	34.33	38.90	48.49
4	2	2	r^*	33.38	37.52	46.12	33.73	38.03	45.99
			P^*	0.086	0.039	0.006	0.091	0.043	0.008
6	2	2	r	33.21	37.79	47.43	33.21	37.79	47.43
6	2	2	r^*	31.97	35.97	44.31	32.43	36.65	45.46
			P^*	0.081	0.036	0.005	0.088	0.041	0.007
4	4	2	r	23.32	26.78	32.56	23.32	26.78	32.56
4	4	2	r^*	21.98	24.80	30.69	22.49	25.55	31.97
			P^*	0.072	0.030	0.004	0.083	0.037	0.006
6	6	2	r	20.98	24.44	31.75	20.98	24.44	31.75
6	6	2	r^*	19.17	21.73	27.10	19.88	22.79	28.92
			P^*	0.062	0.023	0.002	0.077	0.033	0.005
4	2	2	r	28.42	32.24	40.26	25.24	28.78	36.25
4	4	2	r^*	27.27	30.55	37.37	24.45	27.61	34.23
			P^*	0.079	0.034	0.005	0.084	0.038	0.006

Table 2 contains the critical points of our test, ones obtained using (4.10) and $P(a\chi_f^2 > r)$ where r is the critical point of our test and a and f are calculated using (4.10), say P^* in Table 2, to check the influence of the maximization in our test procedure. We set $(p,m) = (8,3)$ and $\delta_i^{(g)} = 1$ ($g = 1,2$, $i = 4,\ldots,8$). Here the significance levels are 0.10, 0.05 and 0.01 and the sample sizes are chosen satisfying (a) $\kappa_1 = \kappa_2 = 0.5$, and (b) $\kappa_1 = 0.8$, $\kappa_2 = 0.2$.

Finally, we summarize the use of our test procedure to actual data. The dimensionality m may be decided according to the latent roots for aimed purpose rather than specified in advance. The test statistic $\bar{n}T_m$ is calculated using n_g and $C^{(g)}$. The critical points at specified significance levels are estimated using n_g and $D^{(g)}$ as follows; the approximate mean $\hat{e}$ and variance $\hat{v}$ are estimated by replacing $\Delta^{(g)}$ by $D^{(g)}$, κ_g by $n_g/(n_1+n_2)$, Q by I_m, and R by I_{p-m}, respectively: the $a\chi_f^2$ approximation are performed using $\hat{e}$ and $\hat{v}$. A supplementary task is to replace Q and R by (4.11) in order to check that the maximization in our test procedure does not deeply effect the result. Table 2 shows that the influences are negligible as far as $\delta_{m+1}^{(g)}$ is comparatively away from $\delta_m^{(g)}$, which is the case that testing H_m is meaningful.

5. An example

The testing problem for H_m is considered by Schott (1988). We illustrate how to use our test procedure in practical situations by applying it to the example in Schott (1988). The data are originally treated in a study by Airoldi and Hoffmann (1984). The data rise from the skull measurements on two groups of the vole species *Microtus californicus*: (1) males; (2) females. The sample sizes $N_i = n_i + 1$ are 82 and 60, respectively. The four variables are condylo-incisive length, alveolar length of upper molar toothrow, zygomatic width, and interiorbital width. Flury (1987) analyses the data under partial common principal components model. Schott (1988) tests the hypothesis H_2 on the data and H_2 is not rejected at any reasonable significance levels. In our study, testing hypothsis H_2 are considered because of the same observation on the latent roots like Schott (1988). The sample covariance matrices, with their spectral decompositions, are given in Table 3. To test H_2, we calculate $\bar{n}T_m = 0.17$, $\hat{e} = 0.88$, and $\hat{v} = 0.51$. Consequently, for the $a\chi_f^2$ approximation, $a = 0.29$ and $f = 3.03$, so that H_2 is not rejected at any reasonable significance levels, which is the same conclusion as in Schott (1988).

6. Discussion

In our study, the test procedure is proposed based on T_m so that it is conservative with respect to Q, R, V_i and W_i with estimating δ_i^g. On the other hand, we can constitute an asymptotic similar test under the assumption that all latent roots are distinct and positive. When H_m is true, the limiting distribution of $\mathrm{vec}(n^{1/2}C_2^{(1)\prime}C_1^{(2)})$ as $n \to \infty$ is normal with mean 0, and covariance matrix Ψ which is positive definite. Therefore, replacing Ψ by the estimator $\hat{\Psi}$, we can consider the test statistic

$$\bar{n}T_m^* = (\mathrm{vec}(n^{1/2}C_2^{(1)\prime}C_1^{(2)}))'\hat{\Psi}^{-1}(\mathrm{vec}(n^{1/2}C_2^{(1)\prime}C_1^{(2)})), \tag{6.1}$$

which is asymptotically distributed as chi-squared with $m(p-m)$ degrees of freedom under H_m. Similarly, from Theorem 3.2, when X is normally distributed, the modification like (6.1) using an estimate of the covariance matrix leads to a test

Table 3. Data on the vole species *Microtus californicus*.

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

Male ($n_1 = 81$) Female ($n_2 = 59$)

$$S^{(1)} = \begin{bmatrix} 13.55 & 10.00 & 10.65 & -2.05 \\ 10.00 & 18.19 & 8.90 & 3.58 \\ 10.65 & 8.90 & 15.32 & 3.00 \\ -2.05 & 3.58 & 3.00 & 27.10 \end{bmatrix} \quad S^{(2)} = \begin{bmatrix} 18.74 & 14.78 & 15.62 & 2.51 \\ 14.78 & 19.00 & 15.51 & 7.02 \\ 15.62 & 15.51 & 19.45 & 5.91 \\ 2.51 & 7.02 & 5.91 & 23.31 \end{bmatrix}$$

(b) Latent roots

$D^{(1)} = \mathrm{diag}(36.32, 27.01, 8.05, 2.78)$

$D^{(2)} = \mathrm{diag}(57.44, 21.15, 3.75, 3.17)$

(c) Latent vectors

$$C^{(1)} = \begin{bmatrix} .789 & .187 & .314 & .494 \\ -.239 & -.755 & .101 & .602 \\ -.541 & .625 & .123 & .549 \\ .168 & .064 & -.936 & .303 \end{bmatrix} \quad C^{(2)} = \begin{bmatrix} .737 & .286 & .309 & .529 \\ -.109 & -.820 & .062 & .559 \\ -.653 & .483 & .140 & .566 \\ .138 & .116 & -.938 & .295 \end{bmatrix}$$

statistic distributed as chi-squared under certain assumptions. For multiple roots case, however, the limiting distribution of $\mathrm{vec}(n^{1/2}C_2^{(1)\prime}C_1^{(2)})$ is no longer normal, and so, the same device as (5.1) is not successful. The same observaton can be applied to Y in Theorem 3.3. Further, because of the rotational freedom in H_m, it might be better to use the test by $\bar{n}T_m$ whether the multiplicities of roots are known or not.

Acknowledgements

The author is very grateful to the referees for their valuable comments. He also wishes to thank Professor Y. Fujikoshi for some helpful discussions.

References

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

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

Chen, K. H. and Robinson, J. (1989). Comparison of factor spaces of two related populations, *J. Multivariate Anal.*, **28**, 190–203.

Flury, B. N. (1984). Common principal components in k groups, *J. Amer. Statist. Assoc.*, **79**, 892–898.

Flury, B. N. (1986). Asymptotic theory for common principal component analysis, *Ann. Statist.*, **14**, 418–430.

Flury, B. N. (1987). Two generalizations of the common principal component model, *Biometrika*, **74**, 59–69.

Kato, T. (1966). *Perturbation Theory for Linear Operators*, Springer, Berlin.

Krzanowski, W. J. (1979). Between-group comparison of principal components, *J. Amer. Statist. Assoc.*, **74**, 703–707 (Corregenda: *ibid.* (1981), **76**, 1022).

Krzanowski, W. J. (1982). Between group comparison of principal components—some sampling results, *J. Statist. Comput. Simulation*, **15**, 141–154.

Schott, J. R. (1988). Common principal component subspaces in two groups, *Biometrika*, **75**, 229–236.