A generalized likelihood ratio test for multivariate analysis of variance in high dimension

Elsevier[1]

Radarweg 29, Amsterdam

Elsevier Inc[a,b], *Global Customer Service*[b,*]

[a]*1600 John F Kennedy Boulevard, Philadelphia*
[b]*360 Park Avenue South, New York*

Abstract

This paper considers in the high dimensional setting a canonical testing problem, namely testing the equality of multiple mean vectors of normal distribution. Motivated by Roy's union-intersection principal, we propose a generalized likelihood ratio test. The critical value is determined by permutation method. We introduce an algorthm for permuting procedure, whose complexity does not depend on data dimension. The limiting distribution of the test statistic is derived in two different setting: non-spiked covariance and spiked covariance. Theoretical results and simulation studies show that the test is particularly powerful under spiked covariance.

Keywords: `elsarticle.cls`, LaTeX, Elsevier, template
2010 MSC: 00-01, 99-00

1. Introduction

Suppose there are k groups of p dimensional data. Within the ith group ($1 \leq i \leq k$), we have observations $\{X_{ij}\}_{j=1}^{n_i}$ which are independent and identically

---

☆Fully documented templates are available in the elsarticle package on CTAN.
*Corresponding author
Email address: `support@elsevier.com` (Global Customer Service)
URL: `www.elsevier.com` (Elsevier Inc)
[1]Since 1880.

distributed (i.i.d.) as $N_p(\mu_i, \Sigma)$, the p dimensional normal distribution with mean vector μ_i and variance matrix Σ. We would like to test

$$H : \mu_1 = \mu_2 = \cdots = \mu_k \quad \text{v.s.} \quad K : \mu_i \neq \mu_j \text{ for some } i \neq j. \tag{1}$$

The problem is known as one-way multivariate analysis of variance (MANOVA) and is well studied in classical statistic literatures. There are four classical tests for hypothesis (1): Wilks' Lambda (which is also the LRT), Hotelling-Lawley trace, Pillai Trace and Roy's maximum root.

In some modern scientific applications, people would like to test hypothesis (1) in high dimensional setting, i.e., p is greater than $n = \sum_{i=1}^k n_i$. See, for example, [1]. However, when $p > n - k$, the LRT for hypothesis (1) is not well defined. Researchers have done extensive work to study the testing problem (1) in high dimensional setting. Sor far, most tests in the literature are designed for two sample case, i.e. $k = 2$. See, for example, [2], [3], [4], [5] and [6]. For the multiple sample case, [7] modified the Dempster's trace test and proposed the test statistic

$$T_{SC} = \frac{1}{\sqrt{n-1}}\Big(\frac{1}{k-1}\operatorname{tr}\Big(\sum_{i=1}^k n_i \bar{X}_i \bar{X}_i^T - n\bar{X}\bar{X}^T\Big) - \frac{1}{n-k}\operatorname{tr}\Big(\sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}X_{ij}^T - \sum_{i=1}^k n_i \bar{X}_i \bar{X}_i^T\Big)\Big),$$

where $\bar{\mathbf{X}}_i = n_i^{-1} \sum_{j=1}^{n_i} X_{ij}$ and $\bar{\mathbf{X}} = n^{-1} \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}$. In another work, [8] proposed a test statistic

$$T_{CX} = \max_{1 \le i \le p} \sum_{1 \le j < l \le k} \frac{n_j n_l}{n_j + n_l} \frac{(\Omega(\bar{X}_j - \bar{X}_l))_i^2}{\omega_{ii}},$$

Where $\Omega = (\omega)_{ij} = \Sigma^{-1}$ is the precision matrix. When Ω is unknown, they substitute it by an estimator $\hat{\Omega}$. Stitistics T_{SC} and T_{CX} are the representatives of two popular methodologies for high dimensional tests. T_{SC} is a so-called sum-of-squares type statistic as it is based on an estimation of squared Euclidean norm $\sum_{i=1}^k n_i \|\mu_i - \bar{\mu}\|^2$, where $\bar{\mu} = n^{-1} \sum_{i=1}^k n_i \mu_i$. T_{CX} is an extreme value type statistic.

Note that both sum-of-squares type statistic and extreme value type statistic are not based on likelihood function. It remains a problem how to construct likelihood-based tests in high dimensional setting. In a recent work, [9] proposed a generalized likelihood ratio test in the context of one-sample mean vector test. Inspired by Roy's union-intersection tests ([10]), they write the null hypothesis as the intersection of a class of component hypotheses. For each component hypotheses, the likelihood ratio test is constructed. Using a least favorable argument, they construct a test statistics based on these tests. Their simulation results showed that their test has particular good power performance when the variables are dependent.

Following [9]'s methodology, we proposed a generalized likelihood ratio test for hypothesis (1). Most existing tests for hypothesis (1) imposed conditions which prevent from large leading eigenvalues of Σ. However, when the correlations between variables are determined by a small number of factors, Σ is spiked in the sense that a few leading eigenvalues are much larger than the others. See, for example [11] and [12]. We derive the asymptotic distribution of the test statistic under both spiked and non-spiked covariance. Our theoretical results imply that the new test is particularly powerful under spiked covariance. We conduct a simulation study to examine the numerical performance of the test.

The rest of the paper is organized as follows.

Higher criticism CX are special case of UIT.

2. Methodology

2.1. Roy's maximum root

Roy's maximum root test statistic is derived in [10] as an example of Roy's union intersection principle. The idea of Roy's union intersection principle is to reduce testing problem to a class of pseudo-univariate problems. For $a \in \mathbb{R}^p$ and $a^T a = 1$, define the hypothesis H_a and K_a as

$$H_a : a^T\mu_1 = a^T\mu_2 = \cdots = a^T\mu_k \quad \text{v.s.} \quad K_a : a^T\mu_i \neq a^T\mu_j \text{ for some } i \neq j.$$

Then the hypothesis (1) can be written as

$$H = \bigcap_{a^T a=1} H_a \quad \text{v.s.} \quad K = \bigcup_{a^T a=1} K_a.$$

The union-intersection principle tells that H is rejected if and only if any one of H_a is rejected. Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{in_i})$ be the ith sample, $i = 1, \ldots, k$. Let $\mathbf{Z} = (\mathbf{X}_1, \ldots, \mathbf{X}_k)$ be the pool sample. Note that the likelihood function based on $a^T\mathbf{Z}$ is

$$f_a(a^T\mathbf{Z}; \mu_1, \ldots, \mu_k, \Sigma) = (2\pi)^{-n/2}|a^T\Sigma a|^{-n/2} \exp\Big(-\frac{1}{2a^T\Sigma a}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(a^T X_{ij} - a^T\mu_i)^2\Big).$$

The likelihood ratio test statistic for H_a v.s. K_a (which is also uniformly most powerful unbiased test) is

$$\mathrm{LRT}_a = \frac{\sup\limits_{\mu_1,\ldots,\mu_k,\Sigma} f_a(a^T\mathbf{Z}; \mu_1, \ldots, \mu_k, \Sigma)}{\sup\limits_{\mu,\Sigma} f_a(a^T\mathbf{Z}; \mu, \ldots, \mu, \Sigma)} = \Big(1 + \frac{a^T F a}{a^T G a}\Big)^{n/2},$$

where $G = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{\mathbf{X}}_i)(X_{ij} - \bar{\mathbf{X}}_i)^T$ and $F = \sum_{i=1}^{k} n_i(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^T$. Follow the type II method of [10], the union intersection test statistic is $\max_{a^T a=1} \mathrm{LRT}_a = (1 + \lambda_{\max}(FG^{-1}))^{n/2}$ which is an increase function of $\lambda_{\max}(FG^{-1}))$, Roy's maximum root test statistic. Note that here we need the assumption $p \leq n - k$, or else G is not invertible.

2.2. The new test statistic

We are interested in the case when $p > n-k$. In this setting, $\max_{a^T a=1} \mathrm{LRT}_a = +\infty$ and Roy's maximum root test is not defined. In another viewpoint, union intersection principal finds an direction a along which the evidence against null hypothesis is maximized. Such an a is data dependent. In the classical setting, the evidence of direction a is LRT_a. In the current context, there are a class of a such that LRT_a achieve the infinity, the largest evidence in classical sense. We need to further choose a single a from $\{a \,|\, \mathrm{LRT}_a = +\infty \text{ and } a^T a = 1\}$. From the expression of LRT_a, we would like to make the largest discrepancy between $a^T F a$ and $a^T G a$. Note that if $\mathrm{LRT}_\alpha = +\infty$, then $a^T G a = 0$. Hence it's natural to choose a as

$$a^* = \mathop{\arg\max}_{a^T a=1, a^T G a=0} a^T F a.$$

Since $a^{*T} G a^* = 0$, we propose the following test statistic for H:

$$T = a^{*T} F a^* = \max_{a^T a=1, a^T G a=0} a^T F a.$$

When T is large enough, we reject H. The above strategy is first proposed by [9] in the context of testing one sample mean vector.

Next we derive the explicit forms of the test statistic. Let $J = \operatorname{diag}(n_1^{-1/2}\mathbf{1}_{n_1}, \ldots, n_k^{-1/2}\mathbf{1}_{n_k})$. Then the matrices $I_n - JJ^T$, $JJ^T - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$ and $\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$ are three $n \times n$ projection matrices which are pairwise orthogonal with rank $n-k$, $k-1$ and 1. Let $\tilde{J}$ be a $n \times (n-k)$ matrix satisfying $\tilde{J}\tilde{J}^T = I - JJ^T$. Note that $I_k - \frac{1}{n}J^T\mathbf{1}_n\mathbf{1}_n^T J$ is a $k \times k$ projection matrix with rank $k-1$. Let C be a $k \times (k-1)$ matrix satisfying $CC^T = I_k - \frac{1}{n}J^T\mathbf{1}_n\mathbf{1}_n^T J$. Then

$$G = Z(I_n - JJ^T)Z^T = Z\tilde{J}\tilde{J}^T Z^T.$$

and

$$F = \sum_{i=1}^k n_i(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^T = Z(JJ^T - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)Z^T = ZJ(I_k - \frac{1}{n}J^T\mathbf{1}_n\mathbf{1}_n^T J)J^T Z^T = ZJCC^T J^T Z^T.$$

Let $Z\tilde{J} = U_{Z\tilde{J}} D_{Z\tilde{J}} V_{Z\tilde{J}}^T$ be the singular value decomposition of $Z\tilde{J}$, where $U_{Z\tilde{J}}$ and $V_{Z\tilde{J}}$ are $p \times (n-k)$ and $(n-k) \times (n-k)$ column orthogonal matrices respectively, $D_{Z\tilde{J}}$ is $(n-k) \times (n-k)$ diagonal matrix. Let $H_{Z\tilde{J}} = U_{Z\tilde{J}} U_{Z\tilde{J}}^T$ be the projection on the column space of A. Then by Proposition 1,

$$T(Z) = \lambda_{\max}\big(ZJCC^T J^T Z^T (I_p - H_{Z\tilde{J}})\big) = \lambda_{\max}\big(C^T J^T Z^T (I_p - H_{Z\tilde{J}}) ZJC\big). \tag{2}$$

Next we introduce another form of T. By the relationship

$$\begin{pmatrix} J^T Z^T Z J & J^T Z^T Z\tilde{J} \\ \tilde{J}^T Z^T Z J & \tilde{J}^T Z^T Z \tilde{J} \end{pmatrix}^{-1} = \Big(\begin{pmatrix} J^T \\ \tilde{J}^T \end{pmatrix} Z^T Z \begin{pmatrix} J & \tilde{J} \end{pmatrix} \Big)^{-1} = \begin{pmatrix} J^T (Z^T Z)^{-1} J & J^T (Z^T Z)^{-1} \tilde{J} \\ \tilde{J}^T (Z^T Z)^{-1} J & \tilde{J}^T (Z^T Z)^{-1} \tilde{J} \end{pmatrix}$$

and matrix inverse formula, we have that

$$\big(J^T (Z^T Z)^{-1} J\big)^{-1} = J^T Z^T Z J - J^T Z^T Z\tilde{J}(\tilde{J}^T Z^T Z \tilde{J})^{-1} \tilde{J}^T Z^T Z J = J^T Z^T (I_p - H_{Z\tilde{J}}) Z J.$$

Thus,

$$T(Z) = \lambda_{\max}\Big(C^T\big(J^T(Z^TZ)^{-1}J\big)^{-1}C\Big). \tag{3}$$

While the form (2) is used for theoretical analysis, the form (3) is well suited for computation, as we shall see.

2.3. Permutation method

Permutation method is a powerful tool to determine the critical value of a test statistic. The test procedure resulting from permutation method is exact as long as the null distribution of observations are exchangeable. See, for example, [13]. The major down-side to permutation method is that it can be computationally intensive. Fortunately, for our statistic, there is a fast implementation of the permutation method. Using expression (3), a permuted statistic can be written as

$$T(Z\Gamma) = \lambda_{\max}\Big(C^T\big(J^T\Gamma^T\big(Z^TZ\big)^{-1}\Gamma J\big)^{-1}C\Big), \tag{4}$$

where Γ is an $n \times n$ permutation matrix. Note that $\big(Z^TZ\big)^{-1}$, the most time-consuming component, can be calculated aforehand. The permutation procedure for our statistic can be summarized as:

(I) Calculate $T(Z)$ according to (3), hold intermediate result $\big(Z^TZ\big)^{-1}$.

(II) For a large M, independently generate M random permutation matrix $\Gamma_1, \dots, \Gamma_M$ and calculate $T(Z\Gamma_1), \dots, T(Z\Gamma_M)$ according to (4).

(III) Calculate the p-value by $\tilde{p} = (M+1)^{-1}\big[1 + \sum_{i=1}^M I\{T(Z\Gamma_i) \geq T(Z)\}\big]$. Reject the null hypothesis if $\tilde{p} \leq \alpha$.

Here M is the permutation times. It can be shown that for any integer $M > 0$, the resulting test controls the Type I error. More precisely, we have $\Pr(\tilde{p} \leq u) \leq u$ for all $0 \leq u \leq 1$. Moreover, as M tends to ∞, $\lim_{M\to\infty} \Pr(\tilde{p} \leq u) = u$. See, for example, [14], Chapter 15.

It can be seen that the time complexities of step (I) and step (II) are $O(n^2p + n^3)$ and $O(n^2M)$, respectively. In large sample or high dimensional setting, $M/(p+n)$ is small. In this case, the permutation procedure has negligible effect on total time complexity.

3. Theory

Let

$$\Xi \stackrel{def}{=} (\sqrt{n_1}\mu_1, \ldots, \sqrt{n_k}\mu_k).$$

Then $\mathrm{E}\, Z = \Xi J^T$.

$H : \Xi C = O_{p\times(k-1)}$

The uniformly minimum variance unbiased estimator of Ξ is $ZJ = (\sqrt{n_1}\bar{\mathbf{X}}_1, \ldots, \sqrt{n_k}\bar{\mathbf{X}}_k)$.

Suppose M is a $(k-1)\times p$ matrix.

$H_M : \mathrm{tr}(M\Xi C) = 0,\ K_M : \mathrm{tr}(M\Xi C) \geq 0.$

$$\mathrm{tr}\left(MZJC\right) = \mathrm{tr}\left(CMZJ\right)$$

Note that $CM\sqrt{n_i}\bar{\mathbf{X}}_i \sim N_{k-1}(\sqrt{n_i}CM\mu_i, CM\Sigma M^T C^T)$. Hence under H_M, we have that

$$\mathrm{tr}\left(CMZJ\right) \sim N(\mathrm{tr}(CM\Xi), \mathrm{tr}(CM\Sigma M^T C^T)) \sim N(\mathrm{tr}(M\Xi C), \mathrm{tr}(M\Sigma M^T)).$$

Hence define

$$T_M = \frac{\mathrm{tr}\left(MZJC\right)}{\sqrt{\mathrm{tr}(MGM^T)}}.$$

By Cauchy inequality $\max_B \mathrm{tr}(AB^T)/\,\mathrm{tr}^{1/2}(BB^T) = \mathrm{tr}^{1/2}(AA^T)$, we have

$$\begin{aligned}
\max_M T_M &= \max_M \frac{\mathrm{tr}\left(MG^{1/2}G^{-1/2}ZJC\right)}{\sqrt{\mathrm{tr}(MG^{1/2}(MG^{1/2})^T)}} \\
&= \mathrm{tr}^{1/2}((ZJC)^T G^{-1} ZJC) \\
&= \mathrm{tr}^{1/2}(ZJC(ZJC)^T G^{-1}) \\
&= \mathrm{tr}^{1/2}(FG^{-1}).
\end{aligned}$$

In this section, we investigate the asymptotic behavior of our test statistic when p is much larger than n. More precisely, we shall assume $p/n \to \infty$. In high dimensional setting, it is a common phenomenon that the asymptotic distribution of statistic relies on the covariance structure. See, for example, [15] and Rui Wang's paper. We shall investigate the asymptotics of our statistic under two different covariance structures: non-spiked covariance and spiked covariance.

Let W_{k-1} be a $(k-1) \times (k-1)$ symmetric random matrix whose entries above the main diagonal are i.i.d. $N(0,1)$ and the entries on the diagonal are i.i.d. $N(0,2)$.

Let $\Sigma = U\Lambda U^T$ be the eigenvalue decomposition of Σ, where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$.

Theorem 1. *Suppose $p/n \to \infty$, $c \le \lambda_p(\Sigma) \le \cdots \le \lambda_1(\Sigma) \le C$ and*

$$\operatorname{tr}\Big(\Sigma - \frac{1}{p}(\operatorname{tr}\Sigma)I_p\Big)^2 = o\big(\frac{p}{n}\big).$$

Under local alternative $p^{-1}\|\Xi C\|_F^2 \to 0$, we have

$$\frac{T(Z) - \frac{p-n+k}{p}\operatorname{tr}(\Sigma)}{\sqrt{2\operatorname{tr}(\Sigma^2)}} \sim \lambda_{\max}\Big(W_{k-1} + \big(2\operatorname{tr}(\Sigma^2)\big)^{-1/2} C^T \Xi^T (I_p - H_{Z\tilde{J}})\Xi C\Big) + o_P(1).$$

The spiked covariance model assumes that a few eigenvalues of Σ are significantly larger than the others. This model is a standard model in many problems and takes factor model as a special case. See, for example,.

Assumption 1. *Let r be a fixed integer. Suppose $\lambda_r n/p \to \infty$ and $C \ge \lambda_{r+1} \ge \ldots \ge \lambda_p \ge c$, where c and C are absolute constant.*

Let $U = (U_1, U_2)$ where U_1 is $p \times r$ and U_2 is $p \times (p-r)$. Let $\Lambda_1 = \operatorname{diag}(\lambda_1, \ldots, \lambda_r)$ and $\Lambda_2 = \operatorname{diag}(\lambda_{r+1}, \ldots, \lambda_p)$. Then $\Sigma = U_1\Lambda_1 U_1^T + U_2 \Lambda_2 U_2^T$.

Theorem 2. *Under Assumption (1), suppose $p/n \to \infty$, $\frac{\lambda_1^2 p}{\lambda_r^2 n^2} \to 0$ and*

$$\operatorname{tr}\Big(\Lambda_2 - \frac{1}{p-r}(\operatorname{tr}\Lambda_2)I_{p-r}\Big)^2 = o\big(\frac{p}{n}\big).$$

Then under local alternative

$$\frac{1}{\sqrt{p}}\|\Xi C\|_F^2 = O(1),$$

we have

$$\frac{T(Z) - \frac{p-r-n+k}{p-r}\operatorname{tr}(\Lambda_2)}{\sqrt{2\operatorname{tr}(\Lambda_2^2)}} \sim \lambda_{\max}\Big(W_{k-1} + \big(2\operatorname{tr}(\Lambda_2^2)\big)^{-1/2} C^T \Xi^T (I_p - H_{Z\tilde{J}})\Xi C\Big) + o_P(1).$$

4. Simulation Results

In this section, we evaluate the numerical performance of the new test. For comparison, we also carried out simulation for the test of Tony Cai and Yin Xia and the test of Schott. These tests are denoted respectively by NEW, CX and SC.

In the simulations, we set $k = 3$. Note that the new test is invariant under orthogonal transformation. Without loss of generality, we only consider diagonal Σ. We set $\Sigma = \operatorname{diag}(p, 1, \ldots, 1)$. Define signal-to-noise ratio (SNR) as

$$\mathrm{SNR} = \frac{\|\mu_f\|_F^2}{\sqrt{\sum_{i=2}^p \lambda_i(\Sigma)^2}}.$$

We use SNR to characterize the signal strength. We consider two alternative hypotheses: the non-sparse alternative and the sparse alternative. In the non-sparse case, we set $\mu_1 = \kappa 1_p$, $\mu_2 = -\kappa 1_p$ and $\mu_3 = 0_p$, where κ is selected to make the SNR equal to the given value. In the sparse case, we set $\mu_1 = \kappa(1_{p/5}^T, 0_{4p/5}^T)^T$, $\mu_2 = \kappa(0_{p/5}^T, 1_{p/5}^T, 0_{3p/5}^T)^T$ and $\mu_3 = 0_p$. Again, κ is selected to make the SNR equal to the given value.

5. Appendix

Proposition 1. *Suppose A is a $p \times r$ matrix with rank r and B is a $p \times p$ non-zero semi-definite matrix. Denote by $A = U_A D_A V_A^T$ the singular value decomposition of A, where U_A and V_A are $p \times r$ and $r \times r$ column orthogonal matrix, D_A is a $r \times r$ diagonal matrix. Let $H_A = U_A U_A^T$ be the projection on the column space of A. Then*

$$\max_{a^T a = 1, a^T A A^T a = 0} a^T B a = \lambda_{\max}\big(B(I_p - H_A)\big). \tag{5}$$

Proof. Note that $a^T A A^T a = 0$ is equivalent to $H_A a = 0$ which in turn is equivalent to $a = (I_p - H_A)a$. Then

$$\max_{a^T a = 1, a^T A A^T a = 0} a^T B a = \max_{a^T a = 1, H_A a = 0} a^T (I_p - H_A) B (I_p - H_A) a, \tag{6}$$

Table 1: Empirical powers of tests under non-sparse alternative with $\alpha = 0.05$, $k = 3$, $n_1 = n_2 = n_3 = 10$. Based on 1000 replications.

SNR	$p = 50$			$p = 75$			$p = 100$		
	CX	SC	NEW	CX	SC	NEW	CX	SC	NEW
0	0.035	0.048	0.052	0.057	0.052	0.057	0.053	0.048	0.045
1	0.060	0.049	0.096	0.081	0.050	0.092	0.063	0.062	0.085
2	0.100	0.058	0.140	0.073	0.045	0.169	0.086	0.055	0.171
3	0.145	0.066	0.234	0.119	0.070	0.266	0.117	0.056	0.307
4	0.126	0.064	0.317	0.121	0.059	0.380	0.122	0.061	0.402
5	0.179	0.072	0.392	0.178	0.068	0.541	0.141	0.071	0.579
6	0.198	0.070	0.513	0.189	0.071	0.639	0.143	0.066	0.717
7	0.249	0.085	0.629	0.227	0.084	0.777	0.206	0.073	0.822
8	0.268	0.092	0.685	0.252	0.084	0.822	0.217	0.078	0.894
9	0.324	0.100	0.786	0.256	0.090	0.911	0.246	0.074	0.949
10	0.342	0.115	0.828	0.303	0.097	0.937	0.270	0.075	0.973

Table 2: Empirical powers of tests under non-sparse alternative with $\alpha = 0.05$, $k = 3$, $n_1 = n_2 = n_3 = 25$. Based on 1000 replications.

SNR	$p = 100$			$p = 150$			$p = 200$		
	CX	SC	NEW	CX	SC	NEW	CX	SC	NEW
0	0.050	0.043	0.050	0.056	0.066	0.048	0.062	0.045	0.054
1	0.069	0.048	0.063	0.046	0.052	0.091	0.068	0.048	0.095
2	0.097	0.046	0.131	0.086	0.053	0.164	0.068	0.057	0.173
3	0.113	0.061	0.200	0.117	0.057	0.270	0.101	0.045	0.313
4	0.135	0.053	0.247	0.130	0.054	0.402	0.118	0.066	0.485
5	0.158	0.065	0.357	0.134	0.066	0.526	0.134	0.073	0.616
6	0.198	0.081	0.433	0.161	0.052	0.668	0.138	0.067	0.765
7	0.217	0.068	0.514	0.191	0.067	0.759	0.174	0.068	0.862
8	0.229	0.063	0.582	0.223	0.075	0.853	0.187	0.060	0.927
9	0.264	0.094	0.680	0.218	0.080	0.918	0.227	0.067	0.966
10	0.298	0.091	0.758	0.245	0.076	0.934	0.228	0.052	0.982

Table 3: Empirical powers of tests under sparse alternative with $\alpha = 0.05$, $k = 3$, $n_1 = n_2 = n_3 = 10$. Based on 1000 replications.

SNR	$p = 50$			$p = 75$			$p = 100$		
	CX	SC	NEW	CX	SC	NEW	CX	SC	NEW
0	0.063	0.056	0.052	0.048	0.049	0.048	0.057	0.047	0.042
1	0.087	0.058	0.071	0.069	0.044	0.096	0.076	0.051	0.080
2	0.091	0.066	0.116	0.113	0.037	0.133	0.080	0.058	0.139
3	0.155	0.065	0.177	0.131	0.062	0.228	0.113	0.058	0.218
4	0.184	0.065	0.246	0.174	0.076	0.308	0.144	0.061	0.310
5	0.225	0.081	0.337	0.214	0.075	0.386	0.176	0.083	0.417
6	0.270	0.088	0.425	0.266	0.085	0.507	0.228	0.071	0.508
7	0.364	0.080	0.501	0.307	0.078	0.571	0.302	0.087	0.629
8	0.405	0.105	0.549	0.381	0.080	0.698	0.362	0.089	0.721
9	0.470	0.121	0.634	0.408	0.078	0.774	0.391	0.070	0.797
10	0.547	0.128	0.702	0.484	0.109	0.819	0.415	0.088	0.877

Table 4: Empirical powers of tests under sparse alternative with $\alpha = 0.05$, $k = 3$, $n_1 = n_2 = n_3 = 25$. Based on 1000 replications.

SNR	$p = 100$			$p = 150$			$p = 200$		
	CX	SC	NEW	CX	SC	NEW	CX	SC	NEW
0	0.048	0.045	0.046	0.053	0.046	0.043	0.051	0.034	0.046
1	0.079	0.055	0.082	0.066	0.063	0.079	0.063	0.059	0.100
2	0.097	0.054	0.119	0.088	0.055	0.138	0.085	0.055	0.160
3	0.133	0.069	0.167	0.113	0.066	0.223	0.114	0.054	0.235
4	0.149	0.062	0.212	0.126	0.084	0.298	0.132	0.057	0.344
5	0.204	0.060	0.281	0.169	0.066	0.427	0.154	0.057	0.469
6	0.252	0.060	0.352	0.227	0.070	0.548	0.195	0.072	0.641
7	0.310	0.072	0.429	0.252	0.059	0.614	0.220	0.061	0.711
8	0.372	0.088	0.529	0.314	0.085	0.719	0.297	0.060	0.800
9	0.427	0.083	0.547	0.362	0.085	0.794	0.300	0.057	0.881
10	0.449	0.093	0.619	0.396	0.072	0.853	0.340	0.076	0.911

Table 5: Empirical powers of tests under non-sparse alternative with $\alpha = 0.05$, $k = 3$, $n_1 = n_2 = n_3 = 25$. The diagonal elements of Σ are generated from sort(Unif(1,100)). Based on 1000 replications.

SNR	$p = 100$			$p = 150$			$p = 200$		
	CX	SC	NEW	CX	SC	NEW	CX	SC	NEW
0	0.063	0.054	0.058	0.052	0.040	0.042	0.045	0.049	0.070
1	0.141	0.120	0.115	0.126	0.120	0.112	0.103	0.110	0.102
2	0.181	0.209	0.169	0.330	0.260	0.210	0.200	0.227	0.201
3	0.692	0.367	0.244	0.759	0.385	0.341	0.468	0.413	0.394
4	0.753	0.539	0.420	0.744	0.573	0.515	0.516	0.554	0.561
5	0.828	0.690	0.509	0.871	0.697	0.693	0.556	0.724	0.727
6	0.809	0.812	0.622	0.822	0.824	0.766	0.959	0.838	0.859
7	1.000	0.882	0.780	0.979	0.916	0.903	0.990	0.923	0.947
8	0.993	0.955	0.789	1.000	0.965	0.954	0.999	0.972	0.971
9	1.000	0.979	0.911	0.999	0.981	0.979	0.964	0.986	0.987
10	1.000	0.991	0.877	0.989	0.996	0.988	0.996	0.996	0.997

Table 6: Empirical powers of tests under sparse alternative with $\alpha = 0.05$, $k = 3$, $n_1 = n_2 = n_3 = 25$. The diagonal elements of Σ are generated from sort(Unif(1,100)). Based on 1000 replications.

SNR	$p = 100$			$p = 150$			$p = 200$		
	CX	SC	NEW	CX	SC	NEW	CX	SC	NEW
0	0.052	0.055	0.047	0.055	0.057	0.053	0.044	0.055	0.057
1	0.068	0.124	0.065	0.070	0.130	0.085	0.049	0.116	0.087
2	0.085	0.233	0.112	0.076	0.239	0.149	0.067	0.241	0.161
3	0.110	0.388	0.161	0.090	0.408	0.215	0.097	0.417	0.227
4	0.120	0.530	0.184	0.112	0.552	0.282	0.103	0.556	0.309
5	0.167	0.708	0.238	0.142	0.699	0.387	0.140	0.687	0.394
6	0.196	0.807	0.261	0.168	0.820	0.472	0.162	0.823	0.547
7	0.217	0.875	0.318	0.177	0.892	0.505	0.173	0.896	0.646
8	0.234	0.935	0.378	0.220	0.951	0.625	0.195	0.948	0.749
9	0.312	0.965	0.407	0.222	0.970	0.672	0.224	0.979	0.809
10	0.334	0.976	0.505	0.292	0.987	0.773	0.254	0.989	0.881

which is obviously no greater than $\lambda_{\max}\big((I-H_A)B(I-H_A)\big)$. To prove that they are equal, without loss of generality, we can assume $\lambda_{\max}\big((I-H_A)B(I-H_A)\big) > 0$. Let α_1 be one eigenvector corresponding to the largest eigenvalue of $(I-H_A)B(I-H_A)$. Since $(I-H_A)B(I-H_A)H_A = (I-H_A)B(H_A-H_A) = O_{p\times p}$ and H_A is symmetric, the rows of H_A are eigenvetors of $(I-H_A)B(I-H_A)$ corresponding to eigenvalue 0. It follows that $H_A\alpha_1 = 0$. Therefore, α_1 satisfies the constraint of (6) and (6) is no less than $\lambda_{\max}\big((I-H_A)B(I-H_A)\big)$. The conclusion now follows by noting that $\lambda_{\max}\big((I-H_A)B(I-H_A)\big) = \lambda_{\max}\big(B(I-H_A)\big)$.

□

Proof of the main results. It can be seen that ZJC is independent of $Z\tilde{J}$. Since $\mathrm{E}(Z\tilde{J}) = O_{p\times(n-k)}$, we can write $Z\tilde{J} = U\Lambda^{1/2}G_1$, where G_1 is a $p\times(n-k)$ matrix with i.i.d. $N(0,1)$ entries. We write $ZJC = \mu_f + U\Lambda^{1/2}G_2$, where G_2 is a $p\times(k-1)$ matrix with i.i.d. $N(0,1)$ entries.

Then

$$\begin{aligned}C^TJ^TZ^T(I_p-H_{Z\tilde{J}})ZJC =& G_2^T\Lambda^{1/2}U^T(I_P-H_{Z\tilde{J}})U\Lambda^{1/2}G_2 + \mu_f^T(I_p-H_{Z\tilde{J}})\mu_f +\\ &\mu_f^T(I_p-H_{Z\tilde{J}})U\Lambda^{1/2}G_2 + G_2^T\Lambda^{1/2}U^T(I_P-H_{Z\tilde{J}})\mu_f.\end{aligned} \tag{7}$$

The first term of (7) can be represented as

$$G_2^T\Lambda^{1/2}U^T(I_p-H_{Z\tilde{J}})U\Lambda^{1/2}G_2 = \sum_{i=1}^{p}\lambda_i(\Lambda^{1/2}U^T(I_p-H_{Z\tilde{J}})U\Lambda^{1/2})\xi_i\xi_i^T, \tag{8}$$

where $\xi_i \overset{i.i.d.}{\sim} N(0, I_{k-1})$.

Proof of Theorem 1. First we deal with the first term of (7). Note that for $i = 1,\dots,p$, we have

$$\lambda_i(\Lambda^{1/2}U^T(I_p-H_{Z\tilde{J}})U\Lambda^{1/2}) \le \lambda_i(\Lambda). \tag{9}$$

Note that $H_{Z\tilde{J}}$ has rank $n-k$. For $i = 1,\dots,p-n+k$, by Weyl's inequality, we have

$$\lambda_i(\Lambda^{1/2}U^T(I_p-H_{Z\tilde{J}})U\Lambda^{1/2}) \ge \lambda_{i+n-k}(\Lambda). \tag{10}$$

Then we have

$$\frac{\lambda_1^2(\Lambda^{1/2}U^T(I_p - H_{Z\tilde{Z}})U\Lambda^{1/2})}{\sum_{i=1}^p \lambda_i^2(\Lambda^{1/2}U^T(I_p - H_{Z\tilde{Z}})U\Lambda^{1/2})} \leq \frac{C}{c(p-n+k)} \to 0.$$

Apply Lyapunov central limit theorem conditioning on $Z\tilde{J}$, we have

$$\Big(2\sum_{i=1}^p \lambda_i^2(\Lambda^{1/2}U^T(I_p - H_{Z\tilde{Z}})U\Lambda^{1/2})\Big)^{-1/2}$$
$$\Big(G_2^T\Lambda^{1/2}U^T(I_p - H_{Z\tilde{J}})U\Lambda^{1/2}G_2 - \sum_{i=1}^p \lambda_i(\Lambda^{1/2}U^T(I_p - H_{Z\tilde{Z}})U\Lambda^{1/2})I_{k-1}\Big) \xrightarrow{\mathcal{L}} W_{k-1}.$$

Also by (9) and (10), we have

$$\sum_{i=n-k+1}^p \lambda_i^2 \leq \operatorname{tr}\big[(\Lambda^{1/2}U^T(I_p - H_{Z\tilde{J}})U\Lambda^{1/2})^2\big] \leq \operatorname{tr}(\Lambda^2).$$

Hence we have

$$\operatorname{tr}\big[(\Lambda^{1/2}U^T(I_p - H_{Z\tilde{J}})U\Lambda^{1/2})^2\big] = \operatorname{tr}(\Lambda^2) + O_P(n) = \big(1 + O_P(\frac{n}{p})\big)\operatorname{tr}(\Lambda^2).$$

Note that

$$\operatorname{tr}(\Lambda^{1/2}U^T(I_p - H_{Z\tilde{J}})U\Lambda^{1/2}) = \operatorname{tr}(\Lambda) - \operatorname{tr}(H_{Z\tilde{J}}U\Lambda U^T).$$

and

$$\big|\operatorname{tr}(H_{Z\tilde{J}}U\Lambda U^T) - \frac{n-k}{p}\operatorname{tr}(\Lambda)\big| = \big|\operatorname{tr}\Big(H_{Z\tilde{J}}U\big(\Lambda - \frac{1}{p}(\operatorname{tr}\Lambda)I_p\big)U^T\Big)\big|$$
$$\leq \sqrt{\operatorname{tr}\big(H_{Z\tilde{J}}^2\big)}\sqrt{\operatorname{tr}\Big(\Lambda - \frac{1}{p}\big(\operatorname{tr}\Lambda\big)I_p\Big)^2} = \sqrt{(n-k)\operatorname{tr}\Big(\Lambda - \frac{1}{p}\big(\operatorname{tr}\Lambda\big)I_p\Big)^2} = o(\sqrt{p}).$$

Hence

$$\operatorname{tr}(\Lambda^{1/2}U^T(I_p - H_{Z\tilde{J}})U\Lambda^{1/2}) = \frac{p-n+k}{p}\operatorname{tr}(\Lambda) + o(\sqrt{p}).$$

It follows that

$$\Big(2\sum_{i=1}^p \lambda_i^2(\Lambda^{1/2}U^T(I_p - H_{Z\tilde{Z}})U\Lambda^{1/2})\Big)^{-1/2}$$
$$\Big(G_2^T\Lambda^{1/2}U^T(I_p - H_{Z\tilde{J}})U\Lambda^{1/2}G_2 - \sum_{i=1}^p \lambda_i(\Lambda^{1/2}U^T(I_p - H_{Z\tilde{Z}})U\Lambda^{1/2})I_{k-1}\Big)$$
$$=\Big(2(1 + O_P(\frac{n}{p}))\operatorname{tr}(\Lambda^2)\Big)^{-1/2}\Big(G_2^T\Lambda^{1/2}U^T(I_p - H_{Z\tilde{J}})U\Lambda^{1/2}G_2 - (\frac{p-n+k}{p}\operatorname{tr}(\Lambda) + O_P(\sqrt{p}))I_{k-1}\Big)$$

By Slutsky's theorem, we have that

$$\Big(2\operatorname{tr}(\Lambda^2)\Big)^{-1/2}\Big(G_2^T\Lambda^{1/2}U^T(I_p-H_{Z\tilde{J}})U\Lambda^{1/2}G_2-\frac{p-n+k}{p}\operatorname{tr}(\Lambda)I_{k-1}\Big)\xrightarrow{\mathcal{L}}W_{k-1}$$

Note that

$$\begin{aligned}
&\mathrm{E}\left[\|C^T\Xi^T(I_p-H_{Z\tilde{J}})U\Lambda^{1/2}G_2\|_F^2\right]\\
=&(k-1)\,\mathrm{E}\left[\operatorname{tr}\left(C^T\Xi^T(I_p-H_{Z\tilde{J}})U\Lambda U^T(I_p-H_{Z\tilde{J}})\Xi C\right)\right]\\
\leq&(k-1)\,\mathrm{E}\left[\lambda_1\big((I_p-H_{Z\tilde{J}})U\Lambda U^T(I_p-H_{Z\tilde{J}})\big)\right]\|\Xi C\|_F^2\\
\leq&(k-1)\lambda_1(\Lambda)\|\Xi C\|_F^2\leq(k-1)C\|\Xi C\|_F^2=o(p),
\end{aligned}$$

we have

$$\Big(2\operatorname{tr}(\Sigma^2)\Big)^{-1/2}\Big(C^TJ^TZ^T(I_p-H_{Z\tilde{J}})ZJC-\frac{p-n+k}{p}\operatorname{tr}(\Sigma)I_{k-1}-C^T\Xi^T(I_p-H_{Z\tilde{J}})\Xi C\Big)\xrightarrow{\mathcal{L}}W_{k-1}.$$

Equivalently, we have

$$\begin{aligned}
&\Big(2\operatorname{tr}(\Sigma^2)\Big)^{-1/2}\Big(C^TJ^TZ^T(I_p-H_{Z\tilde{J}})ZJC-\frac{p-n+k}{p}\operatorname{tr}(\Sigma)I_{k-1}\Big)\\
\sim&\Big(2\operatorname{tr}(\Sigma^2)\Big)^{-1/2}C^T\Xi^T(I_p-H_{Z\tilde{J}})\Xi C+W_{k-1}+o_P(1).
\end{aligned}$$

Then the conclusion follows by taking the maximum eigenvalue. $\square$

The following lemma gives the asympototics of $\lambda_i(\tilde{J}^TZ^TZ\tilde{J})$, $i=1,\ldots,r$.

Lemma 1. *Under the Assumptions of Theorem 2, we have* $\lambda_i(\tilde{J}^TZ^TZ\tilde{J})=\lambda_in(1+o_P(1))$, $i=1,\ldots,r$.

Proof. Note that $\tilde{J}^TZ^TZ\tilde{J}=G_1^T\Lambda G_1=V_{Z\tilde{J}}D_{Z\tilde{J}}^2V_{Z\tilde{J}}^T$, and $G_1^T\Lambda G_1=G_{1[1:r,]}^T\Lambda_1G_{1[1:r,]}+G_{1[(r+1):p,]}^T\Lambda_2G_{1[(r+1):p,]}$. We have

$$V_{Z\tilde{J}}D_{Z\tilde{J}}^2V_{Z\tilde{J}}^T=G_{1[1:r,]}^T\Lambda_1G_{1[1:r,]}+G_{1[(r+1):p,]}^T\Lambda_2G_{1[(r+1):p,]}.$$

For $i=1,\ldots,r$,

$$\begin{aligned}
&\lambda_i(G_{1[1:r,]}^T\Lambda_1G_{1[1:r,]})\geq\lambda_i(G_{1[1:r,]}^T\operatorname{diag}(\lambda_iI_i,O_{(r-i)\times(r-i)})G_{1[1:r,]})\\
=&\lambda_i\lambda_i(G_{1[1:i,]}G_{1[1:i,]}^T)=\lambda_in(1+o_P(1)),
\end{aligned}\tag{11}$$

where the last equality holds since $n^{-1}G_{1[1:i,]}G_{1[1:i,]}^T \xrightarrow{P} I_i$ by law of large numbers. On the other hand, for $i = 1, \ldots, r$,

$$
\begin{aligned}
&\lambda_i(G_{1[1:r,]}^T \Lambda_1 G_{1[1:r,]})\\
=&\lambda_i\Big(G_{1[1:r,]}^T\big(\operatorname{diag}(\lambda_1, \ldots, \lambda_{i-1}, O_{(r-i+1)\times(r-i+1)}) + \operatorname{diag}(O_{(i-1)\times(i-1)}, \lambda_i, \ldots, \lambda_r)\big)G_{1[1:r,]}\Big)\\
\le&\lambda_1(G_{1[1:r,]}^T \operatorname{diag}(O_{(i-1)\times(i-1)}, \lambda_i, \ldots, \lambda_r)G_{1[1:r,]}) \le \lambda_1(G_{1[1:r,]}^T \operatorname{diag}(O_{(i-1)\times(i-1)}, \lambda_i I_{r-i+1})G_{1[1:r,]})\\
=&\lambda_i\lambda_1(G_{1[i:r,]}G_{1[i:r,]}^T) = \lambda_i n(1 + o_P(1))
\end{aligned}
\tag{12}
$$

where the first inequality holds by Weyl's inequality. It follows from (11) and (12) that $\lambda_i(G_{1[1:r,]}^T \Lambda_1 G_{1[1:r,]}) = \lambda_i n(1 + o_P(1))$ for $i = 1, \ldots, r$.

Note that $\lambda_{\max}(G_{1[(r+1):p,]}^T \Lambda_2 G_{1[(r+1):p,]}) \le C\lambda_{\max}(G_{1[(r+1):p,]}^T G_{1[(r+1):p,]}) = O_P(p)$ by Bai-Yin's law. By assumption $\lambda_r n/p \to \infty$, we can deduce that $D_{Z\tilde{J}[i,i]}^2 = \lambda_i(G_1^T \Lambda G_1) = \lambda_i n(1 + o_P(1))$, $i = 1, \ldots, r$.

□

The next lemma gives the asymptotics of $U_{Z\tilde{J}[,1:r]}$.

Lemma 2. *Under the Assumptions of Theorem 2, we have*

$$\lambda_{\max}(I_r - U_1^T U_{Z\tilde{J}[,1:r]} U_{Z\tilde{J}[,1:r]}^T U_1) = O_P(\frac{p}{\lambda_r n}).$$

Proof. Note that $U\Lambda^{1/2}G_1G_1^T\Lambda^{1/2}U^T = U_{Z\tilde{J}}D_{Z\tilde{J}}^2 U_{Z\tilde{J}}^T$, we have $G_1G_1^T = \Lambda^{-1/2}U^T U_{Z\tilde{J}}D_{Z\tilde{J}}^2 U_{Z\tilde{J}}^T U\Lambda^{-1/2}$. Thus,

$$
\begin{aligned}
&G_{1[(r+1):p,]}G_{1[(r+1):p,]}^T = \Lambda_2^{-1/2}U_{[,(r+1):p]}^T U_{Z\tilde{J}}D_{Z\tilde{J}}^2 U_{Z\tilde{J}}^T U_{[,(r+1):p]}\Lambda_2^{-1/2}\\
\ge&\Lambda_2^{-1/2}U_{[,(r+1):p]}^T U_{Z\tilde{J}[,1:r]}D_{Z\tilde{J}[1:r,1:r]}^2 U_{Z\tilde{J}[,1:r]}^T U_{[,(r+1):p]}\Lambda_2^{-1/2}\\
\ge&D_{Z\tilde{J}[r,r]}^2\Lambda_2^{-1/2}U_{[,(r+1):p]}^T U_{Z\tilde{J}[,1:r]}U_{Z\tilde{J}[,1:r]}^T U_{[,(r+1):p]}\Lambda_2^{-1/2}.
\end{aligned}
$$

It follows that

$$\lambda_{\max}(U_{[,(r+1):p]}^T U_{Z\tilde{J}[,1:r]}U_{Z\tilde{J}[,1:r]}^T U_{[,(r+1):p]}) \le \frac{C}{D_{Z\tilde{J}[r,r]}^2}\lambda_1(G_{1[(r+1):p,]}G_{1[(r+1):p,]}^T) = O_P(\frac{p}{\lambda_r n}),$$

where the last equality follows by Lemma 1 and Weyl's inequality.

The conclusion follows by the following simple relationship

$$
\begin{aligned}
&\lambda_{\max}(U_{[,(r+1):p]}^T U_{Z\tilde{J}[,1:r]} U_{Z\tilde{J}[,1:r]}^T U_{[,(r+1):p]}) = \lambda_{\max}(U_{Z\tilde{J}[,1:r]}^T U_{[,(r+1):p]} U_{[,(r+1):p]}^T U_{Z\tilde{J}[,1:r]}) \\
=&\lambda_{\max}(U_{Z\tilde{J}[,1:r]}^T (I_p - U_1 U_1^T) U_{Z\tilde{J}[,1:r]}) = \lambda_{\max}(I_r - U_{Z\tilde{J}[,1:r]}^T U_1 U_1^T U_{Z\tilde{J}[,1:r]}) \\
=&1 - \lambda_{\min}(U_{Z\tilde{J}[,1:r]}^T U_1 U_1^T U_{Z\tilde{J}[,1:r]}) = 1 - \lambda_{\min}(U_1^T U_{Z\tilde{J}[,1:r]} U_{Z\tilde{J}[,1:r]}^T U_1) \\
=&\lambda_{\max}(I_r - U_1^T U_{Z\tilde{J}[,1:r]} U_{Z\tilde{J}[,1:r]}^T U_1).
\end{aligned}
$$

□

Proof of Theorem 2. As in the proof of Theorem 1, for $i = r+1, \ldots, p$, we have that

$$\lambda_i(\Lambda^{1/2} U^T (I_p - H_{Z\tilde{J}}) U \Lambda^{1/2}) \le \lambda_i(\Lambda). \tag{13}$$

And for $i = 1, \ldots, p - n + k$, we have

$$\lambda_i(\Lambda^{1/2} U^T (I_p - H_{Z\tilde{J}}) U \Lambda^{1/2}) \ge \lambda_{i+n-k}(\Lambda). \tag{14}$$

Next, we need to give an upper bound for $\lambda_i(\Lambda^{1/2} U^T (I_p - H_{Z\tilde{J}}) U \Lambda^{1/2})$, $i = 1, \ldots, r$. Note that the positive eigenvalues of $\Lambda^{1/2} U^T (I_p - H_{Z\tilde{J}}) U \Lambda^{1/2}$ are equal to the eigenvalues of $(I_p - H_{Z\tilde{J}}) U \Lambda U^T (I_p - H_{Z\tilde{J}})$. Write $(I_p - H_{Z\tilde{J}}) U \Lambda U^T (I_p - H_{Z\tilde{J}})$ as the sum of two terms

$$
\begin{aligned}
&(I_p - H_{Z\tilde{J}}) U \Lambda U^T (I_p - H_{Z\tilde{J}}) \\
=&(I_p - H_{Z\tilde{J}}) U_1 \Lambda_1 U_1^T (I_p - H_{Z\tilde{J}}) + (I_p - H_{Z\tilde{J}}) U_2 \Lambda_2 U_2^T (I_p - H_{Z\tilde{J}}) \stackrel{def}{=} R_1 + R_2.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\lambda_{\max}\big(R_1\big) = \lambda_{\max}\big(\Lambda_1^{1/2} U_1^T (I_p - H_{Z\tilde{J}}) U_1 \Lambda_1^{1/2}\big) \le \lambda_{\max}\big(\Lambda_1^{1/2} U_1^T (I_p - U_{Z\tilde{J}[,1:r]} U_{Z\tilde{J}[,1:r]}^T) U_1 \Lambda_1^{1/2}\big) \\
\le&\lambda_1 \lambda_{\max}\big(U_1^T (I_p - U_{Z\tilde{J}[,1:r]} U_{Z\tilde{J}[,1:r]}^T) U_1\big) = \lambda_1 \lambda_{\max}\big(I_r - U_1^T U_{Z\tilde{J}[,1:r]} U_{Z\tilde{J}[,1:r]}^T U_1\big) = O_P(\frac{\lambda_1 p}{\lambda_r n}).
\end{aligned}
$$

The last equality follows by Lemma 2.

Thus, for $i = 1, \ldots, r$, we have

$$\lambda_i\big((I_p - H_{Z\tilde{J}}) U \Lambda U^T (I_p - H_{Z\tilde{J}})\big) = \lambda_i(R_1 + R_2) \le \lambda_1(R_1 + R_2) \le \lambda_1(R_1) + \lambda_1(R_2) = O_P(\frac{\lambda_1 p}{\lambda_r n}) + C.$$

As a consequence of these bounds, we have

$$\sum_{i=n-k+1}^{p} \lambda_i^2 \le \operatorname{tr}\big((I_p - H_{Z\tilde{J}}) U \Lambda U^T (I_p - H_{Z\tilde{J}})\big)^2 \le r(O_P(\frac{\lambda_1 p}{\lambda_r n}) + C)^2 + \sum_{i=r+1}^{p} \lambda_i^2,$$

or

$$\big|\operatorname{tr}\left((I_p-H_{Z\tilde{J}})U\Lambda U^T(I_p-H_{Z\tilde{J}})\right)^2-\sum_{i=r+1}^{p}\lambda_i^2\big|\leq r(O_P(\frac{\lambda_1 p}{\lambda_r n})+C)^2+O(n). \quad (15)$$

Note that

$$\operatorname{tr}(R_2)=\operatorname{tr}(\Lambda_2)-\operatorname{tr}(H_{Z\tilde{J}}U_2\Lambda_2U_2^T).$$

and

$$\big|\operatorname{tr}(H_{Z\tilde{J}}U_2\Lambda_2U_2^T)-\frac{n-k}{p-r}\operatorname{tr}(\Lambda_2)\big|=\big|\operatorname{tr}\Big(H_{Z\tilde{J}}U\big(\Lambda_2-\frac{1}{p-r}(\operatorname{tr}\Lambda_2)I_{p-r}\big)U^T\Big)\big|$$

$$\leq\sqrt{\operatorname{tr}\left(H_{Z\tilde{J}}^2\right)}\sqrt{\operatorname{tr}\Big(\Lambda_2-\frac{1}{p-r}\big(\operatorname{tr}\Lambda_2\big)I_{p-r}\Big)^2}=\sqrt{(n-k)\operatorname{tr}\Big(\Lambda_2-\frac{1}{p-r}\big(\operatorname{tr}\Lambda_2\big)I_{p-r}\Big)^2}=o(\sqrt{p}).$$

Hence

$$\operatorname{tr}(R_2)=\frac{p-r-n+k}{p-r}\operatorname{tr}(\Lambda_2)+o(\sqrt{p}).$$

Then

$$\big|\operatorname{tr}[(R_1+R_2)]-\frac{p-r-n+k}{p-r}\operatorname{tr}(\Lambda_2)\big|\leq rO_P\big(\frac{\lambda_1 p}{\lambda_r n}\big)+o(\sqrt{p}). \qquad (16)$$

Equation (15) and (16), conbined with the assumptions, yield

$$\operatorname{tr}\left((I_p-H_{Z\tilde{J}})U\Lambda U^T(I_p-H_{Z\tilde{J}})\right)^2=(1+o_P(1))\operatorname{tr}(\Lambda_2),$$

and

$$\operatorname{tr}\left((I_p-H_{Z\tilde{J}})U\Lambda U^T(I_p-H_{Z\tilde{J}})\right)=\frac{p-r-n+k}{p-r}\operatorname{tr}(\Lambda_2)+o_P(\sqrt{p}).$$

Now we have the Lyapunov condition

$$\frac{\lambda_1\Big(\left((I_p-H_{Z\tilde{J}})U\Lambda U^T(I_p-H_{Z\tilde{J}})\right)^2\Big)}{\operatorname{tr}\Big(\left((I_p-H_{Z\tilde{J}})U\Lambda U^T(I_p-H_{Z\tilde{J}})\right)^2\Big)}=\frac{\left(O_P(\frac{\lambda_1 p}{\lambda_r n})+C\right)^2}{(1+o_P(1))\operatorname{tr}(\Lambda_2)}\xrightarrow{P}0.$$

Apply Lyapunov central limit theorem conditioning on $H_{Z\tilde{J}}$, we have

$$\Big(2\operatorname{tr}\Big(\left((I_p-H_{Z\tilde{J}})U\Lambda U^T(I_p-H_{Z\tilde{J}})\right)^2\Big)\Big)^{-1/2}$$
$$\big(G_2^T\Lambda^{1/2}U^T(I_p-H_{Z\tilde{J}})U\Lambda^{1/2}G_2-\operatorname{tr}\left((I_p-H_{Z\tilde{J}})U\Lambda U^T(I_p-H_{Z\tilde{J}})\right)I_{k-1}\big)\xrightarrow{\mathcal{L}}W_{k-1},$$

where W_{k-1} is a $(k-1)\times(k-1)$ symmetric random matrix whose entries above the main diagonal are i.i.d. $N(0,1)$ and the entries on the diagonal are i.i.d. $N(0,2)$. By Slutsky's theorem, we have

$$\Big(2\operatorname{tr}(\Lambda_2)\Big)^{-1/2}\big(G_2^T\Lambda^{1/2}U^T(I_p-H_{Z\tilde{J}})U\Lambda^{1/2}G_2-\tfrac{p-r-n+k}{p-r}\operatorname{tr}(\Lambda_2)I_{k-1}\big)\xrightarrow{\mathcal{L}}W_{k-1}.$$

As for the cross term of (7), we have

$$\begin{aligned}
&\mathrm{E}[\|C^T\Xi^T(I_p-H_{Z\tilde{J}})U\Lambda^{1/2}G_2\|_F^2|Z\tilde{J}]\\
=&(k-1)\operatorname{tr}(C^T\Xi^T(I_p-H_{Z\tilde{J}})U\Lambda U^T(I_p-H_{Z\tilde{J}})\Xi C)\\
\le&(k-1)\lambda_1\big((I_p-H_{Z\tilde{J}})U\Lambda U^T(I_p-H_{Z\tilde{J}})\big)\|\Xi C\|_F^2\\
=&(k-1)O_P(\frac{\lambda_1 p}{\lambda_r n})\|\Xi C\|_F^2\\
=&(k-1)O_P(\frac{\lambda_1\sqrt{p}}{\lambda_r n})\sqrt{p}\|\Xi C\|_F^2=o_P(p)
\end{aligned}$$

The last equality holds when we assume $\frac{1}{\sqrt{p}}\|\Xi C\|_F^2=O(1)$. Hence $\|C^T\Xi^T(I_p-H_{Z\tilde{J}})U\Lambda^{1/2}G_2\|_F^2=o_P(p)$, and we have

$$\Big(2\operatorname{tr}(\Lambda_2)\Big)^{-1/2}\big(C^TJ^TZ^T(I_p-H_{Z\tilde{J}})ZJC-\tfrac{p-r-n+k}{p-r}\operatorname{tr}(\Lambda_2)I_{k-1}-C^T\Xi^T(I_p-H_{Z\tilde{J}})\Xi C\big)\xrightarrow{\mathcal{L}}W_{k-1}.$$

Equivalently, we have

$$\begin{aligned}
&\Big(2\operatorname{tr}(\Lambda_2^2)\Big)^{-1/2}\Big(C^TJ^TZ^T(I_p-H_{Z\tilde{J}})ZJC-\frac{p-r-n+k}{p-r}\operatorname{tr}(\Lambda_2)I_{k-1}\Big)\\
\sim&\Big(2\operatorname{tr}(\Lambda_2^2)\Big)^{-1/2}C^T\Xi^T(I_p-H_{Z\tilde{J}})\Xi C+W_{k-1}+o_P(1).
\end{aligned}$$

Then the conclusion follows by taking the maximum eigenvalue. □

References

[1] C.-A. Tsai, J. J. Chen, Multivariate analysis of variance test for gene set analysis, Bioinformatics 25 (7) (2009) 897. `arXiv:/oup/backfile/content_public/journal/bioinformatics/25/7/10.1093_bioinformatics_btp098/1/btp098.pdf`, `doi:10.1093/bioinformatics/btp098`.
URL `+http://dx.doi.org/10.1093/bioinformatics/btp098`

[2] Z. Bai, H. Saranadasa, Effect of high dimension: by an example of a two sample problem, Statistica Sinica 6 (2) (1996) 311–329.

[3] S. X. Chen, Y. L. Qin, A two-sample test for high-dimensional data with applications to gene-set testing, Annals of Statistics 38 (2) (2010) 808–835.

[4] M. S. Srivastava, A test for the mean vector with fewer observations than the dimension under non-normality, Journal of Multivariate Analysis 100 (3) (2009) 518–532As the access to this document is restricted, you may want to look for a different version under "Related research" (further below) orfor a different version of it.

[5] L. Feng, C. Zou, Z. Wang, Multivariate-sign-based high-dimensional tests for the two-sample location problem, Journal of the American Statistical Association.

[6] C. T. Tony, W. Liu, Y. Xia, P. Fryzlewicz, I. V. Keilegom, Two-sample test of high dimensional means under dependence, Journal of the Royal Statistical Society 76 (2) (2013) 349–372.

[7] J. R. Schott, Some high-dimensional tests for a one-way manova, Journal of Multivariate Analysis 98 (9) (2007) 1825–1839. `doi:10.1016/j.jmva.2006.11.007`.

[8] T. T. Cai, Y. Xia, High-dimensional sparse manova, Journal of Multivariate Analysis 131 (4) (2014) 174–196. `doi:10.1016/j.jmva.2014.07.002`.

[9] J. Zhao, X. Xu, A generalized likelihood ratio test for normal mean when p is greater than n, Computational Statistics & Data Analysis.

[10] S. N. Roy, On a heuristic method of test construction and its use in multivariate analysis, The Annals of Mathematical Statistics 24 (2) (1953) 220–238. `doi:10.1214/aoms/1177729029`.
URL https://doi.org/10.1214/aoms/1177729029

[11] T. T. Cai, Z. Ma, Y. Wu, Sparse pca: Optimal rates and adaptive estimation, Annals of Statistics 41 (6) (2013) 3074–3110.

[12] D. Shen, H. Shen, J. S. Marron, Consistency of sparse pca in high dimension, low sample size contexts, Journal of Multivariate Analysis 115 (1) (2013) 317–333.

[13] J. P. Romano, On the behavior of randomization tests without a group invariance assumption, Journal of the American Statistical Association 85 (411) (1990) 686–692.

[14] J. P. R. E. L. Lehmann, Testing Statistical Hypotheses, Springer New York, 2005. `doi:10.1007/0-387-27605-X`.

[15] Y. Ma, W. Lan, H. Wang, A high dimensional two-sample test under a low dimensional factor structure, Journal of Multivariate Analysis 140 (2015) 162–170.