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. By a strategy similar to Roy's union-intersection test, we propose a generalized likelihood ratio test. The critical value is determined by permutation method, which produces an exact test. The limiting distribution of the test statistic is analysed under non-spiked 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 we have independent observations $X_{ij} \in \mathbb{R}^p$ ($j = 1, \dots, n_k$; $i = 1, \dots, K$) with distribution $N_p(\mu_i, \Sigma)$, where μ_i, $i = 1, \dots, K$, and $\Sigma > 0$ are unknown. We would like to test

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

☆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.

The problem is known as one-way multivariate analysis of variance (MANOVA). A commonly used test for hypothesis (1) is the so-called Wilks' Lambda which is also the LRT.

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 a 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 $\bar{X}_i = n_i^{-1}\sum_{j=1}^{n_i} X_{ij}$ and $\bar{X} = n^{-1}\sum_{i=1}^{K}\sum_{j=1}^{n_i} X_{ij}$. 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.

Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{in_i})$ for $i = 1, \ldots, K$. 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. They write the null hypothesis as the intersection of simpler hypotheses for which the likelihood ratio test statistics exist. Inspired by Roy's union-intersection tests (UIT), they proposed a procedure to combine the obtained test statistics to a single statistic. 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 [10] and [11]. 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. GLRT

Let $\mathbf{Z} = (X_1, \ldots, X_k)$.

$$f(Z; \mu_1, \ldots, \mu_k, \Sigma) = \prod_{i=1}^{k} \Big[(2\pi)^{-n_i p/2} |\Sigma|^{-n_i/2} \exp(-\frac{1}{2} \mathrm{tr} \Sigma^{-1} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)(x_{ij} - \mu_i)^T) \Big].$$

Assume $n = \sum_{i=1}^{p} n_i < p$. Let $a \in \mathbb{R}^p$ be a vector satisfying $a^T a = 1$. Then

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

$$\max_{\mu_1, \ldots, \mu_k, \Sigma} f_a(a^T Z, \mu_1, \ldots, \mu_k, \Sigma) = (2\pi)^{-n/2} \big(\sum_{i=1}^{k} \sum_{j=1}^{n_i} (a^T x_{ij} - a^T \bar{\mathbf{X}}_i)^2 \big)^{-n/2} e^{-n/2} \tag{2}$$

Let $S_i = \sum_{j=1}^{n_i} (x_{ij} - \bar{\mathbf{X}}_i)(x_{ij} - \bar{\mathbf{X}}_i)^T$ and $S = \sum_{i=1}^{k} S_i$.

Under H, we have

$$\max_{\mu, \Sigma} f_a(a^T Z, \mu, \ldots, \mu, \Sigma) = (2\pi)^{-n/2} \big(\sum_{i=1}^{k} \sum_{j=1}^{n_i} (a^T x_{ij} - a^T \bar{\mathbf{X}})^2 \big)^{-n/2} e^{-n/2} \tag{3}$$

The generalized likelihood ratio test statistic is defined as

$$T(Z) = \max_{a^T a=1, a^T S a=0} a^T \sum_{i=1}^{k} n_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^T a \tag{4}$$

Let $J = \operatorname{diag}(n_1^{-1/2}\mathbf{1}_{n_1}, \ldots, n_k^{-1/2}\mathbf{1}_{n_k})$. Then $S = Z(I_n - JJ^T)Z^T$ and

$$\sum_{i=1}^{k} n_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^T = Z(JJ^T - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)Z^T. \tag{5}$$

The matrix $I_n - JJ^T$, $JJ^T - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$ and $\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$ are all projection matrix and pairwise orthogonal with rank $n-k$, $k-1$ and 1.

Let $\tilde{J}$ be a $n \times (n-k)$ matrix satisfied $\tilde{J}\tilde{J}^T = I - JJ^T$. Then $S = Z\tilde{J}\tilde{J}^T Z^T$ and Note that

$$Z(JJ^T - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)Z^T = ZJ(I_k - \frac{1}{n}J^T\mathbf{1}_n\mathbf{1}_n^T J)J^T Z^T.$$

Note that $I_k - \frac{1}{n}J^T\mathbf{1}_n\mathbf{1}_n^T J$ is a projection matrix with rank $k-1$. Let C be a $k \times (k-1)$ matrix satisfied $CC^T = I_k - \frac{1}{n}J^T\mathbf{1}_n\mathbf{1}_n^T J$.

In Proposition 1, letting $A = Z\tilde{J}$ and $B = ZJCC^TJ^TZ^T$ yields

$$\begin{aligned} T(Z) &= \lambda_{max}\big((I_p - Z\tilde{J}(\tilde{J}^T Z^T Z\tilde{J})^{-1}\tilde{J}^T Z^T)ZJCC^T J^T Z^T (I_p - Z\tilde{J}(\tilde{J}^T Z^T Z\tilde{J})^{-1}\tilde{J}^T Z^T)\big) \\ &= \lambda_{max}\big(C^T J^T Z^T (I_p - Z\tilde{J}(\tilde{J}^T Z^T Z\tilde{J})^{-1}\tilde{J}^T Z^T)ZJC\big). \end{aligned}$$

Note that

$$\begin{aligned} &\Big(\begin{pmatrix} J^T \\ \tilde{J}^T \end{pmatrix} Z^T Z \begin{pmatrix} J & \tilde{J} \end{pmatrix} \Big)^{-1} \\ =& \begin{pmatrix} J^T Z^T Z J & J^T Z^T Z \tilde{J} \\ \tilde{J}^T Z^T Z J & \tilde{J}^T Z^T Z \tilde{J} \end{pmatrix}^{-1} = \begin{pmatrix} J^T (Z^T Z)^{-1} J & J^T (Z^T Z)^{-1} \tilde{J} \\ \tilde{J}^T (Z^T Z)^{-1} J & \tilde{J}^T (Z^T Z)^{-1} \tilde{J} \end{pmatrix}. \end{aligned} \tag{6}$$

It follows that

$$\begin{aligned} &\big(J^T (Z^T Z)^{-1} J\big)^{-1} \\ =& J^T Z^T Z J - J^T Z^T Z\tilde{J}(\tilde{J}^T Z^T Z\tilde{J})^{-1}\tilde{J}^T Z^T Z J \\ =& J^T Z^T (I_p - Z\tilde{J}(\tilde{J}^T Z^T Z\tilde{J})^{-1}\tilde{J}^T Z^T) Z J \end{aligned} \tag{7}$$

It follows that

$$T(Z) = \lambda_{\max}\Big(C^T \big(J^T (Z^T Z)^{-1} J\big)^{-1} C\Big) \tag{8}$$

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. Let $H_A = A(A^T A)^{-1} A^T$. Then*

$$\max_{a^T a=1, a^T A A^T a=0} a^T B a = \lambda_{\max}\big((I_p - H_A)B(I_p - H_A)\big). \tag{9}$$

Proof. Note that $a^T A A^T a = 0$ is equivalent to $A^T a = 0$ and is in turn equivalent to $H_A a = 0$. In this circumstance, $a = (I_p - H_A)a$. Then

$$\begin{aligned} \max_{a^T a=1, a^T A A^T a=0} a^T B a &= \max_{a^T a=1, H_A a=0} a^T B a \\ &= \max_{a^T a=1, H_A a=0} a^T (I_p - H_A) B (I_p - H_A) a. \end{aligned} \tag{10}$$

It's obvious that (10) $\leq \lambda_{\max}\big((I - H_A)B(I - H_A)\big)$. On the other hand, let α_1 be one eigenvector corresponding to the largest eigenvalue of $(I - H_A)B(I - H_A)$. Note that the row of H_A are all eigenvetors of $(I - H_A)B(I - H_A)$ corresponding to eigenvalue 0. It follows that $H_A \alpha_1 = 0$. Now that α_1 satisfies the constraint of (10), (10) is maximized when $a = \alpha_1$.

$\square$

3. Schott's method

$$E = ZZ^T - \sum_{i=1}^{k} n_i \bar{X}_i \bar{X}_i^T.$$

$$H = \sum_{i=1}^{k} n_i \bar{X}_i \bar{X}_i^T - n \bar{X} \bar{X}^T.$$

$$\operatorname{tr} E = \operatorname{tr} Z^T Z - \operatorname{tr} J^T Z^T Z J.$$

$$\operatorname{tr} H = \operatorname{tr} J^T Z^T Z J - \frac{1}{n} 1_n^T Z^T Z 1_n$$

$$T_{SC} = \frac{1}{\sqrt{n-1}} \left(\frac{1}{k-1} \operatorname{tr} H - \frac{1}{n-k} \operatorname{tr} E\right)$$

4. Theory

Let $\Sigma = U\Lambda U^T$ be the eigenvalue decomposition of Σ, where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$. 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 $\Sigma = U_1\Lambda_1 U_1^T + U_2\Lambda_2 U_2^T$.

Let $Z\tilde{J} = U_{Z\tilde{J}} D_{Z\tilde{J}} V_{Z\tilde{J}}^T$ be the singular value decomposition of $Z\tilde{J}$. Let $H_{Z\tilde{J}} = U_{Z\tilde{J}} U_{Z\tilde{J}}^T$. Then $T(Z) = \lambda_{\max}(C^T J^T Z^T (I_p - H_{Z\tilde{J}}) ZJC)$. Note that

$$\mathrm{E}(ZJC) = (\sqrt{n_1}\mu_1, \ldots, \sqrt{n_k}\mu_k)C \stackrel{def}{=} \mu_f.$$

Assumption 1. *Assume $C \geq \lambda_{r+1} \geq \ldots \geq \lambda_p \geq c$, where c and C are absolute constant.*

Theorem 1. *Suppose Assumption (1) holds. Suppose*

$$p/n \to \infty, \quad \text{and} \quad \frac{\lambda_1^2 p}{\lambda_r^2 n^2} \to 0. \tag{11}$$

Suppose

$$\frac{\lambda_r n}{p} \to \infty. \tag{12}$$

Suppose

$$\frac{1}{\sqrt{p}} \|\mu_f\|_F^2 = O(1). \tag{13}$$

Then

$$(\operatorname{tr}\Lambda_2^2)^{-1/2}\big(C^T J^T Z^T (I_p - H_{Z\tilde{J}}) ZJC - (\operatorname{tr}\Lambda_2) I_{k-1} - \mu_f^T (I_p - H_{Z\tilde{Z}})\mu_f\big) \xrightarrow{\mathcal{L}} W_{k-1}, \tag{14}$$

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. $N(0,2)$.

5. 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 $\Sigma = \mathrm{diag}(p, 1, \ldots, 1)$. 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.

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	$p = 50$			$p = 75$			$p = 100$		
	SC	CX	NEW	SC	CX	NEW	SC	CX	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	$p = 100$			$p = 150$			$p = 200$		
	SC	CX	NEW	SC	CX	NEW	SC	CX	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	$p = 50$			$p = 75$			$p = 100$		
	SC	CX	NEW	SC	CX	NEW	SC	CX	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	$p = 100$			$p = 150$			$p = 200$		
	SC	CX	NEW	SC	CX	NEW	SC	CX	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

6. Appendix

Proof of Theorem 1. 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

$$\begin{aligned} C^TJ^TZ^T(I_p - H_{Z\tilde{J}})ZJC =& G_2^T\Lambda^{1/2}U^T(I_P - H_{Z\tilde{J}})U\Lambda_{1/2}G_2 + \mu_f^T(I_p - H_{Z\tilde{J}})\mu_f + \\ & \mu_f^T(I_p - H_{Z\tilde{J}})U\Lambda^{1/2}G_2 + G_2^T\Lambda^{1/2}U^T(I_P - H_{Z\tilde{J}})\mu_f. \end{aligned} \tag{15}$$

To deal the first term, we note that

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

where $\xi_i \overset{i.i.d.}{\sim} N(0, I_{k-1})$. The key to its asymptotic behavior is the positive eigenvalues of $\Lambda^{1/2}U^T(I_p - H_{Z\tilde{J}})U\Lambda^{1/2}$, which in turn 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

$$\begin{aligned} &(I_p - H_{Z\tilde{J}})U\Lambda U^T(I_p - H_{Z\tilde{J}}) \\ =&(I_p - H_{Z\tilde{J}})U_1\Lambda_1U_1^T(I_p - H_{Z\tilde{J}}) + (I_p - H_{Z\tilde{J}})U_2\Lambda_2U_2^T(I_p - H_{Z\tilde{J}}) \overset{def}{=} R_1 + R_2. \end{aligned}$$

Note that

$$\begin{aligned} &\lambda_{\max}\big(R_1\big) = \lambda_{\max}\big(\Lambda_1^{1/2}U_1^T(I_p - H_{Z\tilde{J}})U_1\Lambda_1^{1/2}\big) \le \lambda_{\max}\big(\Lambda_1^{1/2}U_1^T(I_p - U_{Z\tilde{J}[,1:r]}U_{Z\tilde{J}[,1:r]}^T)U_1\Lambda_1^{1/2}\big) \\ \le&\lambda_1\lambda_{\max}\big(U_1^T(I_p - U_{Z\tilde{J}[,1:r]}U_{Z\tilde{J}[,1:r]}^T)U_1\big) = \lambda_1\lambda_{\max}\big(I_r - U_1^TU_{Z\tilde{J}[,1:r]}U_{Z\tilde{J}[,1:r]}^TU_1\big). \end{aligned}$$

To investigate the behavior of $U_{Z\tilde{J}}$, we need to investigate the behavior of $D_{Z\tilde{J}}$ first. Note that $G_1^T\Lambda G_1 = \tilde{J}^TZ^TZ\tilde{J} = V_{Z\tilde{J}}D_{Z\tilde{J}}^2V_{Z\tilde{J}}^T$, 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 $i = 1,\dots,r$,

$$\begin{aligned} &\lambda_i(G_{1[1:r,]}^T\Lambda_1G_{1[1:r,]}) \ge \lambda_i(G_{1[1:r,]}^T\operatorname{diag}(\lambda_iI_i, O_{(r-i)\times(r-i)})G_{1[1:r,]}) \\ =&\lambda_i\lambda_i(G_{1[1:i,]}G_{1[1:i,]}^T) = \lambda_in(1 + o_P(1)), \end{aligned} \tag{16}$$

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 $i = 1, \ldots, r$,

$$
\begin{aligned}
&\lambda_i(G_{1[1:r,]}^T \Lambda_1 G_{1[1:r,]}) \\
=&\lambda_i\Big(G_{1[1:r,]}^T\big(\operatorname{diag}(\lambda_1, \ldots, \lambda_{i-1}, O_{(r-i+1)\times(r-i+1)}) + \operatorname{diag}(O_{(i-1)\times(i-1)}, \lambda_i, \ldots, \lambda_r)\big)G_{1[1:r,]}\Big) \\
\leq&\lambda_1(G_{1[1:r,]}^T \operatorname{diag}(O_{(i-1)\times(i-1)}, \lambda_i, \ldots, \lambda_r)G_{1[1:r,]}) \leq \lambda_1(G_{1[1:r,]}^T \operatorname{diag}(O_{(i-1)\times(i-1)}, \lambda_i I_{r-i+1})G_{1[1:r,]}) \\
=&\lambda_i\lambda_1(G_{1[i:r,]}G_{1[i:r,]}^T) = \lambda_i n(1 + o_P(1))
\end{aligned}
\tag{17}
$$

where the first inequality holds by Weyl's inequality. It follows from (16) and (17) 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 $\lambda_{\max}(G_{1[(r+1):p,]}^T \Lambda_2 G_{1[(r+1):p,]}) \leq C\lambda_{\max}(G_{1[(r+1):p,]}^T G_{1[(r+1):p,]}) = O_P(p)$ by Bai-Yin's law. By assumption $\lambda_r n/p \to \infty$, we can deduce that $D_{Z\tilde{J}[i,i]}^2 = \lambda_i(G_1^T \Lambda G_1) = \lambda_i n(1 + o_P(1))$, $i = 1, \ldots, r$.

Now we are ready to investigate the behavior of $U_{Z\tilde{J}}$. Since $U\Lambda^{1/2}G_1G_1^T\Lambda^{1/2}U^T = U_{Z\tilde{J}}D_{Z\tilde{J}}^2 U_{Z\tilde{J}}^T$, we have $G_1G_1^T = \Lambda^{-1/2}U^T U_{Z\tilde{J}}D_{Z\tilde{J}}^2 U_{Z\tilde{J}}^T U\Lambda^{-1/2}$, which further indicates

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

Thus,

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

Note that we have the simple relationship

$$
\begin{aligned}
&\lambda_{\max}(U_{[,(r+1):p]}^T U_{Z\tilde{J}[,1:r]}U_{Z\tilde{J}[,1:r]}^T U_{[,(r+1):p]}) = \lambda_{\max}(U_{Z\tilde{J}[,1:r]}^T U_{[,(r+1):p]}U_{[,(r+1):p]}^T U_{Z\tilde{J}[,1:r]}) \\
=&\lambda_{\max}(U_{Z\tilde{J}[,1:r]}^T(I_p - U_1U_1^T)U_{Z\tilde{J}[,1:r]}) = \lambda_{\max}(I_r - U_{Z\tilde{J}[,1:r]}^T U_1U_1^T U_{Z\tilde{J}[,1:r]}) \\
=&1 - \lambda_{\min}(U_{Z\tilde{J}[,1:r]}^T U_1U_1^T U_{Z\tilde{J}[,1:r]}) = 1 - \lambda_{\min}(U_1^T U_{Z\tilde{J}[,1:r]}U_{Z\tilde{J}[,1:r]}^T U_1) \\
=&\lambda_{\max}(I_r - U_1^T U_{Z\tilde{J}[,1:r]}U_{Z\tilde{J}[,1:r]}^T U_1).
\end{aligned}
$$

Therefore $\lambda_{\max}(I_r - U_1^T U_{Z\tilde{J}[,1:r]}U_{Z\tilde{J}[,1:r]}^T U_1) = O_P(\frac{p}{\lambda_r n})$, and we can conclude $\lambda_{\max}(R_1) = O_P(\frac{\lambda_1 p}{\lambda_r n})$.

We now deal with R_1+R_2. For $i=1,\ldots,r$,

$$\lambda_i(R_1+R_2) \le \lambda_1(R_1+R_2) \le \lambda_1(R_1)+\lambda_1(R_2) \le O_P(\frac{\lambda_1 p}{\lambda_r n})+C.$$

For $i=r+1,\ldots,p-r$,

$$\lambda_i(R_1+R_2) \le \lambda_{i-r}(R_2) = \lambda_{i-r}\big(\Lambda_2^{1/2}U_2^T(I_p-H_{Z\tilde{J}})U_2\Lambda_2^{1/2}\big) \le \lambda_{i-r}(\Lambda_2)=\lambda_i.$$

On the other hand, for $i=1,\ldots,p-r-n+k$,

$$\begin{aligned}&\lambda_i(R_1+R_2) \ge \lambda_i(R_2) = \lambda_i\big(\Lambda_2^{1/2}U_2^T(I_p-H_{Z\tilde{J}})U_2\Lambda_2^{1/2}\big)\\ =&\lambda_i\big(\Lambda_2-\Lambda_2^{1/2}U_2^T H_{Z\tilde{J}}U_2\Lambda_2^{1/2}\big) \ge \lambda_{i+n-k}.\end{aligned}$$

The last equality holds since $U_2^T H_{Z\tilde{J}} U_2$ is at most of rank $n-k$.

As a consequence of these bounds, we have

$$\sum_{i=1}^{p-r-n+k} \lambda_{i+n-k}^2 \le \operatorname{tr}[(R_1+R_2)^2] \le r(O_P(\frac{\lambda_1 p}{\lambda_r n})+C)^2+\sum_{i=r+1}^{p-r}\lambda_i^2,$$

or

$$|\operatorname{tr}[(R_1+R_2)^2]-\sum_{i=r+1}^{p}\lambda_i^2| \le \sum_{i=r+1}^{n-k}\lambda_i^2+\sum_{i=p-r+1}^{p}\lambda_i^2+r(O_P(\frac{\lambda_1 p}{\lambda_r n})+C)^2.$$

Similarly,

$$|\operatorname{tr}[(R_1+R_2)]-\sum_{i=r+1}^{p}\lambda_i| \le \sum_{i=r+1}^{n-k}\lambda_i+\sum_{i=p-r+1}^{p}\lambda_i+r(O_P(\frac{\lambda_1 p}{\lambda_r n})+C).$$

These, conbined with the assumptions, yield

$$\operatorname{tr}[(R_1+R_2)^2] = (1+o_P(1))\sum_{i=r+1}^{p}\lambda_i^2,$$

and

$$\operatorname{tr}[(R_1+R_2)] = \sum_{i=r+1}^{p}\lambda_i+O(n)+O_P(\frac{\lambda_1 p}{\lambda_r n}).$$

Now we have the Lyapunov condition

$$\frac{\lambda_1[(R_1+R_2)^2]}{\operatorname{tr}[(R_1+R_2)^2]} = \frac{\big(O_P(\frac{\lambda_1 p}{\lambda_r n})+C\big)^2}{(1+o_P(1))\sum_{i=r+1}^{p}\lambda_i^2} \xrightarrow{P} 0.$$

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

$$\big(\operatorname{tr}[(R_1+R_2)^2]\big)^{-1/2}\big(G_2^T\Lambda^{1/2}U^T(I_p-H_{Z\tilde{J}})U\Lambda_{1/2}G_2-\operatorname{tr}(R_1+R_2)I_{k-1}\big)\xrightarrow{\mathcal{L}}W_{k-1}$$

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

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

As for the cross term of (15), we have

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

The last equality holds when we assume $\frac{1}{\sqrt{p}}\|\mu_f\|_F^2=O(1)$. Hence $\|\mu_f^T(I_p-H_{Z\tilde{J}})U\Lambda^{1/2}G_2\|_F^2=o_P(p)$. This completes the proof of the theorem.

□

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

[11] 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.