Statistica Sinica

A GENERALIZED LIKELIHOOD RATIO TEST FOR MULTIVARIATE ANALYSIS OF VARIANCE IN HIGH DIMENSION

Author(s)

Affiliation(s)

Abstract: This paper considers in high dimensional setting a canonical testing problem, namely testing the equality of multiple mean vectors of normal distribution. Motivated by Roy's union-intersection principle, we propose a generalized likelihood ratio test. To investigate the asymptotic properties of the proposed test, we adopt a spiked covariance model which can characterize the strong correlation between variables. The limiting null distribution of the test statistic is derived and the local asymptotic power function is given. The asymptotic power function implies that the proposed test has particular high power when there are strong correlation between variables. We also carry out simulations to verify our theoretical results.

Key words and phrases:

1. Introduction Suppose there are k $(k \geq 2)$ 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 distributed (i.i.d.) as $N_p(\xi_i, \mathbf{\Sigma})$, the p dimensional normal distribution with mean vector ξ_i and common variance matrix $\mathbf{\Sigma}$. We would like to test the hypotheses

$$H_0 : \xi_1 = \xi_2 = \cdots = \xi_k \quad \text{v.s.} \quad H_1 : \xi_i \neq \xi_j \text{ for some } i \neq j. \tag{1.1}$$

This testing problem is known as one-way multivariate analysis of variance (MANOVA) and has been well studied when p is small compared to n, where $n = \sum_{i=1}^{k} n_i$ is the total sample size.

Let $\mathbf{H} = \sum_{i=1}^{k} n_i(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^T$ be the sum-of-squares between groups and $\mathbf{G} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{\mathbf{X}}_i)(X_{ij} - \bar{\mathbf{X}}_i)^T$ be the sum-of-squares within groups, where $\bar{\mathbf{X}}_i = n_i^{-1}\sum_{j=1}^{n_i} X_{ij}$ is the sample mean of group i and $\bar{\mathbf{X}} = n^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}$ is the pooled sample mean. There are four classical test statistics for hypothesis (1.1), which are all based on the eigenvalues of $\mathbf{HG}^{-1}$.

Wilks' Lambda:	$\lvert\mathbf{G} + \mathbf{H}\rvert / \lvert\mathbf{G}\rvert$
Pillai trace:	$\mathrm{tr}[\mathbf{H}(\mathbf{G} + \mathbf{H})^{-1}]$
Hotelling-Lawley trace:	$\mathrm{tr}[\mathbf{HG}^{-1}]$
Roy's maximum root:	$\lambda_{\max}(\mathbf{HG}^{-1})$

In some modern scientific applications, people would like to test hypothesis (1.1) in high dimensional setting, i.e., p is greater than n. See, for example, Tsai and Chen (2009); Verstynen et al. (2005). However, when $p \geq n$, the four classical test statistics can not be defined. Researchers have done extensive work to study the testing problem (1.1) in high dimensional setting. So far, most tests are designed for two sample case, i.e., $k = 2$. See, for example, Bai and Saranadasa (1996); Chen and Qin (2010); Srivastava (2009); Tony et al. (2013); Feng et al. (2016). For multiple sample case, Schott (2007) modified Hotelling-Lawley trace and proposed the test statistic

$$T_{SC} = \frac{1}{\sqrt{n-1}}\Big(\frac{1}{k-1}\,\mathrm{tr}\left(\mathbf{H}\right) - \frac{1}{n-k}\,\mathrm{tr}\left(\mathbf{G}\right)\Big).$$

In another work, Cai and Xia (2014) proposed a test statistic

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

Where $\Omega = (\omega)_{ij} = \mathbf{\Sigma}^{-1}$ 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 $\sum_{i=1}^k n_i \|\xi_i - \bar{\xi}\|^2$, where $\bar{\xi} = n^{-1}\sum_{i=1}^k n_i \xi_i$. 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. While the likelihood ratio test (LRT), i.e., Wilks' Lambda, is not defined for $p > n - k$, it remains a problem how to construct likelihood-based tests in high dimensional setting. In a recent work, Zhao and Xu (2016) proposed a generalized likelihood ratio test in the context of one-sample mean vector test. Inspired by Roy's union-intersection principle (Roy, 1953), they wrote the null hypothesis as the intersection of a class of component hypotheses. For each component hypotheses, the likelihood ratio test is constructed. They use a least favorable argument to construct test statistic based on component tests. Their simulation results showed that their test has good power performance, especially when the variables are correlated.

Following Zhao and Xu (2016)'s methodology, we propose a generalized likelihood ratio test for hypothesis (1.1). To understand the power behavior of the new test, especially when variables are strongly correlated, the asymptotic distribution of the new statistic needs to be derived. An important correlation pattern is that the variation of variables are mainly driven by a small number of common factors. In this case, the covariance matrix has a few significantly large eigenvalues (Fan et al., 2008; Cai et al., 2013; Shen et al., 2013; Ma et al., 2015). We assume there are r significantly large eigenvalues and the other eigenvalues are bounded. We derive the asymptotic null distribution of the test statistic and give the asymptotic local power function. Our theoretical results implies that the asymptotic power of the new test is not affected by large eigenvalues, while most existing tests are negatively affected by large eigenvalues. Hence the new test is particularly powerful when there are strong correlations between variables. We also conduct a simulation study to examine the numerical performance of the test.

The rest of the paper is organized as follows. In Section 2, we propose a new test. Section 3 concerns the theoretical properties of the proposed test. In Section 4, the proposed test is compared with some existing tests. Section 5 complements our study with some numerical simulations. In Section 6, we give a short discussion. Finally, the proofs are gathered in the Appendix.

2. Methodology

2.1 Discussion about existing methods

To facilitate the discussion, we introduce some notations. Let

$$\mathbf{X} = (X_{11}, X_{12}, \ldots, X_{1n_1}, X_{21}, X_{22}, \ldots, X_{2n_2}, \ldots, X_{k1}, X_{k2}, \ldots, X_{kn_k})$$

be the pooled sample matrix. The sum-of-squares within groups $\mathbf{G}$ can be written as $\mathbf{G} = \mathbf{X}(\mathbf{I}_n - \mathbf{J}\mathbf{J}^T)\mathbf{X}^T$ where

$$\mathbf{J} = \begin{pmatrix} \frac{1}{\sqrt{n_1}}\mathbf{1}_{n_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sqrt{n_2}}\mathbf{1}_{n_2} & \mathbf{0} \\ \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sqrt{n_k}}\mathbf{1}_{n_k} \end{pmatrix},$$

and $\mathbf{1}_{n_i}$ is an n_i-dimensional vector with all elements equal to 1. Let

$$\tilde{\mathbf{J}} = \begin{pmatrix} \tilde{\mathbf{J}}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{J}}_2 & \mathbf{0} \\ \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \tilde{\mathbf{J}}_k \end{pmatrix},$$

where $\tilde{\mathbf{J}}_i$ is a $n_i \times n_{i-1}$ matrix satisfying

$$\tilde{\mathbf{J}}_i = \begin{pmatrix} \frac{\sqrt{2}}{2} & & & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & & \\ & -\frac{\sqrt{2}}{2} & \ddots & \\ & & \ddots & \frac{\sqrt{2}}{2} \\ 0 & & & -\frac{\sqrt{2}}{2} \end{pmatrix}.$$

Then $\tilde{\mathbf{J}}$ is an $n \times (n-k)$ orthogonal matrix satisfying $\tilde{\mathbf{J}}\tilde{\mathbf{J}}^T = \mathbf{I}_n - \mathbf{J}\mathbf{J}^T$. Let $\mathbf{Y} = \mathbf{X}\tilde{\mathbf{J}}$. Then $\mathbf{G}$ has representation

$$\mathbf{G} = \mathbf{Y}\mathbf{Y}^T.$$

On the other hand, the sum-of-squares between groups $\mathbf{H}$ satisfies

$$\mathbf{H} = \mathbf{X}(\mathbf{J}\mathbf{J}^T - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)\mathbf{X}^T = \mathbf{X}\mathbf{J}(\mathbf{I}_k - \frac{1}{n}\mathbf{J}^T\mathbf{1}_n\mathbf{1}_n^T\mathbf{J})\mathbf{J}^T\mathbf{X}^T.$$

We can write $\mathbf{I}_k - \frac{1}{n}\mathbf{J}^T\mathbf{1}_n\mathbf{1}_n^T\mathbf{J} = \mathbf{C}\mathbf{C}^T$ where $\mathbf{C}$ be a $k \times (k-1)$ matrix satisfying

$$\mathbf{C} = \begin{pmatrix} \sqrt{\frac{n_2}{n_1+n_2}} & & & 0 \\ -\sqrt{\frac{n_1}{n_1+n_2}} & \sqrt{\frac{n_3}{n_2+n_3}} & & \\ & -\sqrt{\frac{n_2}{n_2+n_3}} & \ddots & \\ & & \ddots & \sqrt{\frac{n_k}{n_{k-1}+n_k}} \\ 0 & & & -\sqrt{\frac{n_{k-1}}{n_{k-1}+n_k}} \end{pmatrix}.$$

Then $\mathbf{H}$ has representation

$$\mathbf{H} = \mathbf{X}\mathbf{J}\mathbf{C}\mathbf{C}^T\mathbf{J}^T\mathbf{X}^T.$$

Define $\Xi = (\sqrt{n_1}\xi_1, \ldots, \sqrt{n_k}\xi_k)$ and the null hypothesis H_0 is equivalent to $\Xi\mathbf{C} = \mathbf{O}_{p\times(k-1)}$, where $\mathbf{O}_{p\times(k-1)}$ is a $p \times (k-1)$ matrix with all elements equal to 0. Thus the problem becomes testing hypotheses

$$H_0 : \Xi\mathbf{C} = \mathbf{O}_{p\times(k-1)} \quad \text{v.s.} \quad H_1 : \Xi\mathbf{C} \neq \mathbf{O}_{p\times(k-1)}$$

based on data matrix $\mathbf{X}$ when p is larger.

In low dimensional setting, the testing problem 1.1 is well studied. The difficulty occurs when $p \geq n$, where the four classical test statistics can not be defined. As the high dimensional problem is hard to deal with, a simple idea is to reduce the problem to a class of univariate problems. Following this idea, a general strategy to propose a test statistic can be summarized as three steps.

1. Construct a class of projected univariate data $\{\mathbf{X}_\gamma : \gamma \in \Gamma\}$ which contains all the information of data $\mathbf{X}$. This induces a decomposition of the null hypothesis and the alternative hypothesis:

$$H_0 = \bigcap_{\gamma \in \Gamma} H_{0\gamma} \quad \text{v.s.} \quad H_1 = \bigcup_{\gamma \in \Gamma} H_{1\gamma}.$$

2. Construct a test statistic T_γ for $H_{0\gamma}$ against $H_{1\gamma}$ such that $H_{0\gamma}$ is rejected if T_γ is large.

3. Summarize the component test statistics $\{T_\gamma : \gamma \in \Gamma\}$ into a global test statistic.

It turns out that many tests in the literature can be derived by the above strategy. While the LRT may be the best choice of univariate problems in step 2, there are more choices in step 1 and step 3. In step 3, Roy's union intersection principle suggests to use $\max_{\gamma \in \Gamma} T_\gamma$ as global test statistic (Roy, 1953), while another choice is to integrate T_γ according some measure $\mu(\gamma)$ and use $\int_\gamma T_\gamma \, \mu(d\gamma)$ as global test statistic. For step 1, we consider two different constructions of data projection.

i Consider the class of univariate data $\{\mathbf{X}_i = e_i^T\mathbf{X} : i = 1, \ldots, p\}$, where e_i is the ith standard basis in $\mathbb{R}^p$. Hence $H_0 = \bigcap_{i=1}^p H_{0i}$ and $H_1 = \bigcup_{i=1}^p H_{1i}$, where

$$H_{0i} : e_i^T \Xi \mathbf{C} = \mathbf{O}_{1\times(k-1)} \quad \text{and} \quad H_{1a} : e_i^T \Xi \mathbf{C} \neq \mathbf{O}_{1\times(k-1)}.$$

ii Consider the class of univariate data $\{\mathbf{X}_a = a^T\mathbf{X} : i = 1, a \in \mathbb{R}^p, a^T a = 1\}$. Hence $H_0 = \bigcap_{a\in\mathbb{R}^p, a^T a=1} H_{0a}$ and $H_1 = \bigcup_{a\in\mathbb{R}^p, a^T a=1} H_{1a}$, where

$$H_{0a} : a^T \Xi \mathbf{C} = \mathbf{O}_{1\times(k-1)} \quad \text{and} \quad H_{1a} : a^T \Xi \mathbf{C} \neq \mathbf{O}_{1\times(k-1)}.$$

First, we consider the construction i in step 1. Suppose component test statistics

$$T_i = (k-1)^{-1} e_i^T \mathbf{H} e_i - (n-k)^{-1} e_i^T \mathbf{G} e_i \quad i = 1, \ldots, p$$

are used in step 2, and in step 3 we integrate T_i according to the uniform measure on $1, \ldots, p$. Then the resulting statistic is $p^{-1}\sum_{i=1}^p T_i$ which is equivalent to T_{SC}. If instead the likelihood ratio test statistic $e_i^T\mathbf{H}e_i/e_i^T\mathbf{G}e_i$ is used in step 2, one obtains a scalar invariant test statistic which is a direct generalization of Srivastava (2009). On the other hand, by using data $\Omega^{-1}\mathbf{X}$ and component test statistics

$$T_i^* = \sum_{1\le j<l\le k} \frac{n_j n_l}{n_j + n_l} \frac{(\Omega(\bar{\mathbf{X}}_j - \bar{\mathbf{X}}_l))_i^2}{\omega_{ii}},$$

we have $T_{CX} = \max_{1\le i\le p} T_i^*$. Here the component test statistic T_i^* is similar to likelihood ratio tests.

While many existing tests can be derived by the construction i in step 1, this construction has limitation in that it relies on the choice of an orthogonal basis of $\mathbb{R}^p$. In fact, test statistics resulting from this construction mostly requires certain prior information about the covariance matrix. For example, Schott (2007) requires that $\operatorname{tr}(\mathbf{\Sigma}^{2j})/p \to \tau_j \in (0,\infty)$, $j = 1, 2$, and Cai and Xia (2014) requires a consistent estimator of Ω.

Next, we consider using construction ii in step 1, which does not rely on the basis of $\mathbb{R}^p$. Suppose the likelihood ratio test statistic $T_a = a^T\mathbf{H}a/a^T\mathbf{G}a$ is used in step 2. If we use the integrating strategy in step 3 and choose μ equal to the uniform distribution on the sphere, then the test statistic becomes

$$\int_{a^Ta=1} \frac{a^T\mathbf{H}a}{a^T\mathbf{G}a}\,\mu(da).$$

Although it is hard to give an explicit form of the integration, it can be approximated by random projection. More specifically, one can randomly generate unit vectors $a_1, \ldots, a_M$ and the statistics can be approximated by $M^{-1}\sum_{i=1}^M a_i^T\mathbf{H}a_i/a_i^T\mathbf{G}a_i$. This statistic is well defined in high dimensional setting. A similar method is proposed by Lopes et al. (2015) for $k = 2$ from a different point of view. Their analysis and simulations show that such random projection method has relative good performance especially when variables are correlated. On the other hand, if $n-k \geq p$, Roy's union intersection principle can be used in step 3, the resulting statistic is the well known Roy's maximum root:

$$\max_{a^T a=1} T_a = \lambda_{\max}(\mathbf{H}\mathbf{G}^{-1}).$$

In fact, this statistic is first derived in Roy (1953) as an example of his union intersection principle.

2.2 A new test statistic

Roy's maximum root is constructed from the component statistics $\{T_a : a^T a = 1\}$ and doesn't require prior knowledge of covariance matrix. However, it can only be defined when $n-k \geq p$. In fact, if $p > n-k$, G is not invertible and T_a is not defined for some a. We will follow Zhao and Xu (2016)'s idea and propose a new test statistic for $p > n-k$.

Let $L_0(a)$ and $L_1(a)$ be the maximum likelihood of $\mathbf{X}_a$ under H_{01} and H_{01}, respectively. The log likelihood ratio

$$\log \frac{L_1(a)}{L_0(a)} = \Big(\frac{a^T(\mathbf{G}+\mathbf{H})a}{a^T\mathbf{G}a}\Big)^{n/2}$$

is an increase function of T_a. From a likelihood point view, log likelihood ratio is an estimator of the Kullback-Leibler divergence between the true distribution and the null distribution. Thus, the component LRT statistic characterize the discrepancy between H_{0a} and H_{1a}. Thus, by maximizing $\log L_1(a) - \log L_0(a)$, or equivalently maximizing T_a, one obtains component hypothesis H_{0a^*}, where $a^* = \arg\max_{a^Ta=1} T_a$. We shall call H_{0a^*} the least favorable hypothesis since it is the component null hypothesis most like to be not true.

While it is hard to generalize Roy's maximum root to high dimensional setting, the least favorable hypothesis a^* can be formally generalized to high dimensional setting. In what follows, we shall assume $p > n - k$. In this case, Roy's maximum root is not defined since

$$\mathcal{A} \overset{def}{=} \{a : L_1(a) = +\infty,\, a^Ta = 1\} = \{a : a^T\mathbf{G}a = 0,\, a^Ta = 1\}$$

is not empty. Note that

$$\{a : L_0(a) < +\infty,\, a^Ta = 1\} = \{a : a^T(\mathbf{G} + \mathbf{H})a \neq 0,\, a^Ta = 1\}.$$

By the independence of $\mathbf{G}$ and $\mathbf{H}$, with probability 1, we have $\mathcal{A} \cap \{a : L_0(a) < +\infty,\, a^Ta = 1\} \neq \emptyset$. This suggests that the a^* which makes the discrepancy between $L_1(a^*)$ and $L_0(a^*)$ the most should be the one which is in the set $\mathcal{A}$ and minimize $L_0(a)$. Thus, we define $a^* = \arg\min_{a\in\mathcal{A}} L_0(a)$ and take H_{a^*} as the least favorable hypothesis. Equivalently,

$$a^* = \arg\min_{a\in\mathcal{A}} L_0(a) = \arg\max_{a^Ta=1, a^TGa=0} a^T\mathbf{H}a.$$

This motivates us to propose the new test statistic as

$$T(\mathbf{X}) = a^{*T}\mathbf{H}a^* = \max_{a^Ta=1, a^T\mathbf{G}a=0} a^T\mathbf{H}a.$$

We reject the null hypothesis when T is large enough.

Now we derive the explicit forms of the test statistic. Let $\mathbf{Y} = \mathbf{U_Y D_Y V_Y^{\mathit{T}}}$ be the singular value decomposition of $\mathbf{Y}$, where $\mathbf{U_Y}$ and $\mathbf{V_Y}$ are $p\times(n-k)$ and $(n-k)\times(n-k)$ both column orthogonal matrices, $\mathbf{D}_Y$ is an $(n-k)\times(n-k)$ diagonal matrix. Let $\mathbf{P_Y} = \mathbf{U_Y U_Y^{\mathit{T}}}$ be the projection matrix on the column space of $\mathbf{Y}$. Then Proposition 4 implies that

$$T(\mathbf{X}) = \lambda_{\max}\big(\mathbf{C}^T\mathbf{J}^T\mathbf{X}^T(\mathbf{I}_p - \mathbf{P_Y})\mathbf{XJC}\big). \tag{2.2}$$

While (2.2) is convenient for theoretical analysis, it involves a projection matrix $\mathbf{P_Y}$ which is not easy to compute. Next we derive another simple form of $T(\mathbf{X})$. By the relationship

$$\begin{pmatrix} \mathbf{J}^T\mathbf{X}^T\mathbf{XJ} & \mathbf{J}^T\mathbf{X}^T\mathbf{X}\tilde{\mathbf{J}} \\ \tilde{\mathbf{J}}^T\mathbf{X}^T\mathbf{XJ} & \tilde{\mathbf{J}}^T\mathbf{X}^T\mathbf{X}\tilde{\mathbf{J}} \end{pmatrix}^{-1} = \left(\begin{pmatrix} \mathbf{J}^T \\ \tilde{\mathbf{J}}^T \end{pmatrix} \mathbf{X}^T\mathbf{X} \begin{pmatrix} \mathbf{J} & \tilde{\mathbf{J}} \end{pmatrix} \right)^{-1} = \begin{pmatrix} \mathbf{J}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{J} & \mathbf{J}^T(\mathbf{X}^T\mathbf{X})^{-1}\tilde{\mathbf{J}} \\ \tilde{\mathbf{J}}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{J} & \tilde{\mathbf{J}}^T(\mathbf{X}^T\mathbf{X})^{-1}\tilde{\mathbf{J}} \end{pmatrix}$$

and matrix inverse formula, we have that

$$\big(\mathbf{J}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{J}\big)^{-1} = \mathbf{J}^T\mathbf{X}^T\mathbf{XJ} - \mathbf{J}^T\mathbf{X}^T\mathbf{X}\tilde{\mathbf{J}}(\tilde{\mathbf{J}}^T\mathbf{X}^T\mathbf{X}\tilde{\mathbf{J}})^{-1}\tilde{\mathbf{J}}^T\mathbf{X}^T\mathbf{XJ} = \mathbf{J}^T\mathbf{X}^T(\mathbf{I}_p - \mathbf{P_Y})\mathbf{XJ}.$$

Thus,

$$T(\mathbf{X}) = \lambda_{\max}\Big(\mathbf{C}^T\big(\mathbf{J}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{J}\big)^{-1}\mathbf{C}\Big). \tag{2.3}$$

Compared with (2.2), (2.3) doesn't involve $\mathbf{P_Y}$. Hence (2.3) is convenient for computation.

3. Main results

To understand the statistical properties of $T(\mathbf{X})$, we derive the asymptotic distribution of $T(\mathbf{X})$. We are specially interested in the case when variables are correlated. For some real world problems, variables are heavily correlated with common factors, then the covariance matrix $\mathbf{\Sigma}$ is spiked in the sense that a few eigenvalues of $\mathbf{\Sigma}$ are significantly larger than the others (Fan et al., 2008; Cai et al., 2013; Shen et al., 2013; Ma et al., 2015). To characterize this correlation pattern, we make the following assumption for the eigenvalues of $\mathbf{\Sigma}$.

Assumption 1. *Let $\boldsymbol{\lambda}_1 \geq \boldsymbol{\lambda}_2 \geq \cdots \geq \boldsymbol{\lambda}_p$ be the eigenvalues of $\mathbf{\Sigma}$. Suppose the first r $(r \geq 0)$ eigenvalues are significantly larger than the others. We assume $r = o(n)$. We assume $c_1 \geq \boldsymbol{\lambda}_{r+1} \geq \ldots \geq \boldsymbol{\lambda}_p \geq c_2$ for some absolute constants c_1 and c_2. If $r \neq 0$, we assume*

$$\frac{\boldsymbol{\lambda}_r n}{p} \to \infty, \quad \frac{\boldsymbol{\lambda}_1^2 p r^2}{\boldsymbol{\lambda}_r^2 n^2} \to 0 \quad \frac{\log \boldsymbol{\lambda}_r}{n} \to 0.$$

Remark 1. The spiked covariance model is commonly assumed in the study of PCA theory. Most existing work assumed r is fixed. Here we allow r to vary as a smaller order of n. The condition $\boldsymbol{\lambda}_r n/p \to \infty$ requires $\boldsymbol{\lambda}_r$ to be much larger than p/n. This is satisfied, for example, for the factor model adopted by Ma et al. (2015). The most harsh condition is $\boldsymbol{\lambda}_1^2 pr^2/(\boldsymbol{\lambda}_r^2 n^2) \to 0$. If $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_r$ are of same order and r is fixed, this condition is equivalent to $p/n^2 \to 0$. We require this condition since the PCA consistency results are not valid when p is too large. See, for example, (Cai et al., 2013). If $r > 0$, this condition is unavoidable and the asymptotic behavior of $T(\mathbf{X})$ is different if this condition is violated.

To establish the asymptotic distribution of $T(\mathbf{X})$ under Assumption 1, we need following notations. Let $\mathbf{W}_{k-1}$ 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 $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ denote the eigenvalue decomposition of $\boldsymbol{\Sigma}$, where $\boldsymbol{\Lambda} = \operatorname{diag}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_p)$. We denote $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2)$ where $\mathbf{U}_1$ is $p \times r$ and $\mathbf{U}_2$ is $p \times (p-r)$. Denote $\boldsymbol{\Lambda}_1 = \operatorname{diag}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r)$ and $\boldsymbol{\Lambda}_2 = \operatorname{diag}(\boldsymbol{\lambda}_{r+1}, \ldots, \boldsymbol{\lambda}_p)$. Then $\boldsymbol{\Sigma} = \mathbf{U}_1\boldsymbol{\Lambda}_1\mathbf{U}_1^T + \mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^T$.

The following theorem establishes the asymptotic distribution of $T(\mathbf{X})$.

Theorem 1. *Under Assumption 1, suppose $p/n \to \infty$ and*

$$\operatorname{tr}\left(\boldsymbol{\Lambda}_2 - \frac{1}{p-r}\operatorname{tr}(\boldsymbol{\Lambda}_2)\mathbf{I}_{p-r}\right)^2 = o\left(\frac{p}{n}\right).$$

Then under local alternative hypothesis

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

we have

$$\frac{T(\mathbf{X}) - \frac{p-r-n+k}{p-r}\operatorname{tr}(\mathbf{\Lambda}_2)}{\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)}} \sim \lambda_{\max}\Big(\mathbf{W}_{k-1} + \frac{1}{\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)}}\mathbf{C}^T\Xi^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\Xi\mathbf{C}\Big) + o_P(1),$$

where $\sim$ means they have the same distribution.

To gain some insight into the asymptotic behavior of $T(\mathbf{X})$, suppose null hypothesis holds and $k = 2$, then Theorem 1 implies that

$$\frac{T(\mathbf{X}) - \frac{p-r-n+k}{p-r}\operatorname{tr}(\mathbf{\Lambda}_2)}{\sqrt{2\operatorname{tr}(\mathbf{\Lambda}_2^2)}} \xrightarrow{\mathcal{L}} N(0,1).$$

The asymptotic variance of $T(\mathbf{X})$ is $2\operatorname{tr}(\mathbf{\Lambda}_2^2)$. If $r = 0$, this equals to $2\operatorname{tr}(\mathbf{\Sigma}^2)$ which is the asymptotic variance of Bai and Saranadasa (1996) and Chen and Qin (2010)'s statistic. While for $r > 0$, $2\operatorname{tr}(\mathbf{\Lambda}_2^2)$ is smaller than $2\operatorname{tr}(\mathbf{\Sigma}^2)$. In fact, if $\liminf_{n\to\infty}\boldsymbol{\lambda}_1/p \in (0, +\infty]$, we have

$$\liminf_{n\to\infty}\frac{2\operatorname{tr}(\mathbf{\Sigma}^2)}{2\operatorname{tr}(\mathbf{\Lambda}_2^2)} \in (1, +\infty].$$

The reason for this is because the projection $\mathbf{I}_p - \mathbf{P}_\mathbf{Y}$ appeared in expression (2.2) can remove large variance terms of $\mathbf{XJC}$.

To formulate a test procedure with asymptotic correct level, $\operatorname{tr}(\mathbf{\Lambda}_2)$ and $\operatorname{tr}(\mathbf{\Lambda}_2^2)$ should be estimated. Since they relies on the unknown parameter r, we use the following statistic to estimate r:

$$\hat{r} = \begin{cases} \arg\max_{1\le i\le n-k-1} \frac{\lambda_i(\mathbf{Y}^T\mathbf{Y})}{\lambda_{i+1}(\mathbf{Y}^T\mathbf{Y})} \ge \gamma_n & \text{if } \max_{1\le i\le n-k-1} \frac{\lambda_i(\mathbf{Y}^T\mathbf{Y})}{\lambda_{i+1}(\mathbf{Y}^T\mathbf{Y})} \ge \gamma_n \\ 0 & \text{otherwise} \end{cases}$$

where γ_n is slowly tends to $+\infty$ as $n \to \infty$. The following proposition establishes the consistency of $\hat{r}$.

Proposition 1. *Suppose $p/n \to \infty$, $r = o(n)$, $\boldsymbol{\lambda}_r n/p \to \infty$ and $c_1 \ge \boldsymbol{\lambda}_{r+1} \ge \ldots \ge \boldsymbol{\lambda}_p \ge c_2$. If $\gamma_n \to \infty$ and $\gamma_n = o(n\boldsymbol{\lambda}_r/p)$, then* $\Pr(\hat{r} = r) \to 1$.

Remark 2. For the factor model adopted by Ma et al. (2015), γ is of order p. Then we can set $\gamma_n = \sqrt{n}$.

We use the following statistic to estimate $\operatorname{tr}(\boldsymbol{\Lambda}_2)$:

$$\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)} = \frac{1}{n-k} \sum_{i=\hat{r}+1}^{n-k} \lambda_i(\mathbf{Y}^T\mathbf{Y}).$$

Proposition 2. *Under the assumptions of Theorem 1, suppose $\gamma_n \to \infty$ and $\gamma_n = o(n\boldsymbol{\lambda}_r/p)$, then*

$$\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)} = \operatorname{tr}(\boldsymbol{\Lambda}_2) + o_P(\sqrt{p}).$$

To estimate $\operatorname{tr}(\boldsymbol{\Lambda}_2^2)$, we use the idea of leave-two-out. Let $\mathbf{Y}_{(i,j)}$ be the matrix obtained from deleting the ith and jth columns from $\mathbf{Y}$. Let $\mathbf{Y}_{(i,j)} = \mathbf{U}_{\mathbf{Y};(i,j)}\mathbf{D}_{\mathbf{Y};(i,j)}\mathbf{V}_{\mathbf{Y};(i,j)}^T$ be the singular value decomposition of $\mathbf{Y}_{(i,j)}$.

Here $\mathbf{U}_{\mathbf{Y};(i,j)}$ is $p\times(n-k-2)$. Let $\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}$ be a $p\times(p-n+k+2)$ orthogonal matrix satisfying $\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T = \mathbf{I}_p - \mathbf{U}_{\mathbf{Y};(i,j)}\mathbf{U}_{\mathbf{Y};(i,j)}^T$.

Let w_{ij} be the (i,j)th element of $\mathbf{Y}^T\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T\mathbf{Y}$. Define

$$\widehat{\operatorname{tr}(\mathbf{\Lambda}_2^2)} = \frac{2}{(n-k)(n-k-1)}\sum_{1\le i<j\le n-k} w_{ij}^2.$$

We use $\widehat{\operatorname{tr}(\mathbf{\Lambda}_2^2)}$ to estimate $\operatorname{tr}(\mathbf{\Lambda}_2^2)$. The following proposition shows that $\widehat{\operatorname{tr}(\mathbf{\Lambda}_2^2)}$ is a ratio consistent estimator.

Proposition 3. *Under the assumptions of Theorem 1, we have*

$$\frac{\widehat{\operatorname{tr}(\mathbf{\Lambda}_2^2)}}{\operatorname{tr}(\mathbf{\Lambda}_2^2)} \xrightarrow{P} 1.$$

Now we can construct a test procedure with asymptotic correct level α. Let

$$Q = \frac{T(\mathbf{X}) - \frac{p-r-n+k}{p-r}\widehat{\operatorname{tr}(\mathbf{\Lambda}_2)}}{\sqrt{\widehat{\operatorname{tr}(\mathbf{\Lambda}_2^2)}}}.$$

Let $F(x)$ be the cumulative distribution function of $\lambda_{\max}(\mathbf{W}_{k-1})$. We reject the null hypothesis if $Q > F^{-1}(1-\alpha)$.

Theorem 1, Proposition 2 and Proposition 3 implies that the resulting test procedure has asymptotic correct level under the assumptions of Theorem 1. And by Theorem 1, the asymptotic local power function of the proposed test procedure is

$$\Pr\Big(\lambda_{\max}\big(\mathbf{W}_{k-1} + \tfrac{1}{\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)}}\mathbf{C}^T\Xi^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\Xi\mathbf{C}\big) \ge F_{\mathbf{W}}^{-1}(1-\alpha)\Big).$$

If $k = 2$, the asymptotic local power function of Bai and Saranadasa (1996) and Chen and Qin (2010)'s method can be written as

$$\Pr\Big(\lambda_{\max}\big(\mathbf{W}_1 + \tfrac{1}{\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)}}\mathbf{C}^T\Xi^T\Xi\mathbf{C}\big) \geq F_{\mathbf{W}}^{-1}(1-\alpha)\Big).$$

Hence the asymptotic relative efficiency between our method and Bai and Saranadasa (1996) and Chen and Qin (2010)'s method is

$$\sqrt{\frac{\mathrm{tr}(\boldsymbol{\Sigma}^2)}{\mathrm{tr}(\boldsymbol{\Lambda}_2^2)}}\frac{\mathbf{C}^T\Xi^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\Xi\mathbf{C}}{\mathbf{C}^T\Xi^T\Xi\mathbf{C}}.$$

There's a random term $\mathbf{P}_{\mathbf{Y}}$ in the expression. To gain some insight into this, suppose, for example, $\sqrt{n_i}\xi_i$ is from prior distribution $N_p(0, \psi\mathbf{I}_p)$, $i = 1, \ldots, k$. Then $\psi^{-1}\mathbf{C}^T\Xi^T\Xi\mathbf{C}$ is distributed as $\mathrm{Wishart}_{k-1}(p, \mathbf{I}_{k-1})$, the $k-1$ dimensional Wishart distribution with degree of freedom p and parameter $\mathbf{I}_{k-1}$. On the other hand, $\mathbf{C}^T\Xi^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\Xi\mathbf{C}$ is distributed as $\mathrm{Wishart}_{k-1}(p-n+k, \mathbf{I}_{k-1})$. In this case, we have

$$\frac{\mathbf{C}^T\Xi^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\Xi\mathbf{C}}{\mathbf{C}^T\Xi^T\Xi\mathbf{C}} \xrightarrow{P} 1.$$

Thus, when

$$\liminf_{n\to\infty}\frac{\mathrm{tr}(\boldsymbol{\Sigma}^2)}{\mathrm{tr}(\boldsymbol{\Lambda}_2^2)} \in (1, +\infty],$$

the new test tends to be more powerful than Chen and Qin (2010)'s test.

4. Numerical study

4.1 Permutation method

Most existing test can not be used under spiked covariance model, since spiked covariance model violated their assumption. To compare the new test with other tests, we use permutation method to determine the critical value in our simulation.

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 (Romano, 1990). The major down-side to permutation method is that it can be computationally intensive. Fortunately, for the test statistic proposed in this paper, the permutation method can be computationally fast. By expression (2.3), a permuted statistic can be written as

$$T(\mathbf{X}\Gamma) = \lambda_{\max}\Big(\mathbf{C}^T\big(\mathbf{J}^T\Gamma^T(\mathbf{X}^T\mathbf{X})^{-1}\Gamma\mathbf{J}\big)^{-1}\mathbf{C}\Big), \tag{4.4}$$

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

1. Calculate $T(\mathbf{X})$ according to (2.3), hold intermediate result $(\mathbf{X}^T\mathbf{X})^{-1}$.

2. For a large M, independently generate M random permutation matrix $\Gamma_1, \ldots, \Gamma_M$ and calculate $T(\mathbf{X}\Gamma_1), \ldots, T(\mathbf{X}\Gamma_M)$ according to (4.4).

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

Here M is the permutation times. It can be seen that step 1 and step 2 cost $O(n^2 p + n^3)$ and $O(n^2 M)$ operations respectively. In large sample or high dimensional setting, step 2 has negligible effect on total computational complexity.

4.2 Simulation results

In this section, we evaluate the numerical performance of the new test. For comparison, we also carry out simulations for the test of Cai and Xia (2014) and the test of Schott (2007). These tests are denoted respectively by NEW, CX and SC. Since the critical value of CX and SC may not be valid under spiked covariance model, we use permutation method to determine the critical value for all three test. The empirical power is computed based on 1000 simulations.

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 $\mathbf{\Sigma}$. We consider two different structure of $\mathbf{\Sigma}$.

- Covariance structure I: $\mathbf{\Sigma} = \text{diag}(p, 1, \ldots, 1)$.

- Covariance structure II: $\mathbf{\Sigma} = \text{diag}(\rho_1, \ldots, \rho_p)$, where $\rho_1 \geq \cdots \geq \rho_p$ are order statistics of p i.i.d. random variables which have uniform distribution between 0 and 1.

Define signal-to-noise ratio (SNR) as

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

We use SNR to characterize the signal strength. We consider two structure of alternative hypotheses: the non-sparse alternative and the sparse alternative. In the non-sparse case, we set $\xi_1 = \kappa 1_p$, $\xi_2 = -\kappa 1_p$ and $\xi_3 = \mathbf{0}_p$, where κ is selected to make the SNR equal to the given value. In the sparse case, we set $\xi_1 = \kappa(1_{p/5}^T, \mathbf{0}_{4p/5}^T)^T$, $\xi_2 = \kappa(\mathbf{0}_{p/5}^T, 1_{p/5}^T, \mathbf{0}_{3p/5}^T)^T$ and $\xi_3 = \mathbf{0}_p$. Again, κ is selected to make the SNR equal to the given value.

The simulation results are summarized in Tables 1-6. It can be seen from the results that under spiked covariance, the proposed test outperforms the other two tests for both non-sparse and sparse alternatives. Under non-spiked covariance, the power of the new test is a little lower than that of SC. As p increase, the power of the new test approaches to that of SC.

Table 1: Empirical powers of tests under covariance structure I and non-sparse alternative. $\alpha = 0.05$, $k = 3$, $n_1 = n_2 = n_3 = 10$.

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 covariance structure I and non-sparse alternative. $\alpha = 0.05$, $k = 3$, $n_1 = n_2 = n_3 = 25$.

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 covariance structure I and sparse alternative. $\alpha = 0.05$, $k = 3$, $n_1 = n_2 = n_3 = 10$.

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 covariance structure I and sparse alternative. $\alpha = 0.05$, $k = 3$, $n_1 = n_2 = n_3 = 25$.

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 covariance structure II and non-sparse alternative. $\alpha = 0.05$, $k = 3$, $n_1 = n_2 = n_3 = 25$.

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 covariance structure II and sparse alternative. $\alpha = 0.05$, $k = 3$, $n_1 = n_2 = n_3 = 25$.

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

5. Concluding remarks

In this paper, motivated by Roy's union intersection principle, we proposed a generalized likelihood ratio statistic for MANOVA in high dimensional setting. We derive the asymptotic distribution of the new test statistic. We also gives the asymptotic local power function. Our theoretic work and simulation study shows that when the covariance matrix is spiked, the proposed test tends to be more powerful than existing tests.

Appendix

Proposition 4. *Suppose* $\mathbf{A}$ *is a* $p \times r$ *matrix with rank* r *and* $\mathbf{B}$ *is a* $p \times p$ *non-zero semi-definite matrix. Denote by* $\mathbf{A} = \mathbf{U_A}\mathbf{D_A}\mathbf{V_A}^T$ *the singular value decomposition of* $\mathbf{A}$*, where* $\mathbf{U_A}$ *and* $\mathbf{V_A}$ *are* $p \times r$ *and* $r \times r$ *column orthogonal matrix,* $\mathbf{D_A}$ *is a* $r \times r$ *diagonal matrix. Let* $\mathbf{P_A} = \mathbf{U_A}\mathbf{U_A}^T$ *be the projection on the column space of* $\mathbf{A}$*. Then*

$$\max_{a^T a=1, a^T\mathbf{A}\mathbf{A}^T a=0} a^T\mathbf{B}a = \lambda_{\max}\big(\mathbf{B}(\mathbf{I}_p - \mathbf{P_A})\big). \tag{5.5}$$

Proof. Note that $a^T\mathbf{A}\mathbf{A}^T a = 0$ is equivalent to $\mathbf{P_A}a = 0$ which in turn is equivalent to $a = (\mathbf{I}_p - \mathbf{P_A})a$. Then

$$\max_{a^T a=1, a^T\mathbf{A}\mathbf{A}^T a=0} a^T\mathbf{B}a = \max_{a^T a=1, \mathbf{P_A}a=0} a^T(\mathbf{I}_p - \mathbf{P_A})\mathbf{B}(\mathbf{I}_p - \mathbf{P_A})a, \tag{5.6}$$

which is obviously no greater than $\lambda_{\max}\big((\mathbf{I}-\mathbf{P_A})\mathbf{B}(\mathbf{I}-\mathbf{P_A})\big)$. To prove that they are equal, without loss of generality, we can assume $\lambda_{\max}\big((\mathbf{I}-\mathbf{P_A})\mathbf{B}(\mathbf{I}-\mathbf{P_A})\big) > 0$. Let α_1 be one eigenvector corresponding to the largest eigenvalue of $(\mathbf{I}-\mathbf{P_A})\mathbf{B}(\mathbf{I}-\mathbf{P_A})$. Since $(\mathbf{I}-\mathbf{P_A})\mathbf{B}(\mathbf{I}-\mathbf{P_A})\mathbf{P_A} = (\mathbf{I}-\mathbf{P_A})\mathbf{B}(\mathbf{P_A}-\mathbf{P_A}) = \mathbf{O}_{p\times p}$ and $\mathbf{P_A}$ is symmetric, the rows of $\mathbf{P_A}$ are eigenvetors of $(\mathbf{I}-\mathbf{P_A})\mathbf{B}(\mathbf{I}-\mathbf{P_A})$ corresponding to eigenvalue 0. It follows that $\mathbf{P_A}\alpha_1 = 0$. Therefore, α_1 satisfies the constraint of (5.6) and (5.6) is no less than $\lambda_{\max}\big((\mathbf{I}-\mathbf{P_A})\mathbf{B}(\mathbf{I}-\mathbf{P_A})\big)$. The conclusion now follows by noting that $\lambda_{\max}\big((\mathbf{I}-\mathbf{P_A})\mathbf{B}(\mathbf{I}-\mathbf{P_A})\big) = \lambda_{\max}\big(\mathbf{B}(\mathbf{I}-\mathbf{P_A})\big)$.

□

Lemma 1 (Davidson and Szarek (2001) Theorem II.7)**.** *Let* $\mathbf{A}$ *be* $m \times n$ *with iid* $N(0,1)$ *entries. If* $m > n$*, then for any* $t > 0$*,*

$$\Pr(\sqrt{\lambda_1(\mathbf{A}\mathbf{A}^T)} > \sqrt{m} + \sqrt{n} + t) \le \exp(-t^2/2),$$
$$\Pr(\sqrt{\lambda_n(\mathbf{A}\mathbf{A}^T)} < \sqrt{m} - \sqrt{n} - t) \le \exp(-t^2/2).$$

Proves of the main results It can be seen that $\mathbf{XJC}$ is independent of $\mathbf{Y}$. Since $\mathrm{E}\,\mathbf{Y} = \mathbf{O}_{p\times(n-k)}$, we can write $\mathbf{Y} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{G}_1$, where $\mathbf{G}_1$ is a $p\times(n-k)$ matrix with i.i.d. $N(0,1)$ entries. We write $\mathbf{XJC} = \xi_f + \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{G}_2$, where $\mathbf{G}_2$ is a $p\times(k-1)$ matrix with i.i.d. $N(0,1)$ entries.

Then

$$\mathbf{C}^T\mathbf{J}^T\mathbf{X}^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{XJC} = \mathbf{G}_2^T\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{G}_2 + \xi_f^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\xi_f + \xi_f^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{G}_2 + \mathbf{G}_2^T\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\xi_f. \tag{5.7}$$

The first term of (5.7) can be represented as

$$\mathbf{G}_2^T\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{G}_2 = \sum_{i=1}^{p} \lambda_i(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2})\xi_i\xi_i^T, \tag{5.8}$$

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

Let $\mathbf{G}_1 = (\mathbf{G}_{1A}^T, \mathbf{G}_{1B}^T)^T$, where $\mathbf{G}_{1A}$ is the first r rows of $\mathbf{G}_1$ and $\mathbf{G}_{1B}$ is the last $p - r$ rows of $\mathbf{G}_1$. The following lemma gives the asymptotic property of $\lambda_i(\mathbf{Y}^T\mathbf{Y})$, $i = 1, \ldots, r$.

Lemma 2. *Suppose $p/n \to \infty$, $r = o(n)$, $\boldsymbol{\lambda}_r n/p \to \infty$ and $c_1 \geq \boldsymbol{\lambda}_{r+1} \geq \ldots \geq \boldsymbol{\lambda}_p \geq c_2$. Then*

$$\sup_{1\leq i\leq r} \left| \frac{\lambda_i(\mathbf{Y}^T\mathbf{Y})}{n\boldsymbol{\lambda}_i} - 1 \right| \to 0, \tag{5.9}$$

$$\limsup_{n\to+\infty} \frac{\lambda_{r+1}(\mathbf{Y}^T\mathbf{Y})}{p} \leq c_1, \tag{5.10}$$

$$\liminf_{n\to+\infty} \frac{\lambda_{n-k}(\mathbf{Y}^T\mathbf{Y})}{p} \geq c_2, \tag{5.11}$$

almost surely.

Proof. Note that

$$\mathbf{Y}^T\mathbf{Y} = \mathbf{G}_1^T\mathbf{\Lambda}\mathbf{G}_1 = \mathbf{G}_{1A}^T\mathbf{\Lambda}_1\mathbf{G}_{1A} + \mathbf{G}_{1B}^T\mathbf{\Lambda}_2\mathbf{G}_{1B}.$$

For $1 \le i \le r$, we have

$$\lambda_i(\mathbf{G}_{1A}^T \boldsymbol{\Lambda}_1 \mathbf{G}_{1A}) \le \lambda_i(\mathbf{Y}^T\mathbf{Y}) \le \lambda_i(\mathbf{G}_{1A}^T \boldsymbol{\Lambda}_1 \mathbf{G}_{1A}) + c_1 \lambda_1(\mathbf{G}_{1B}^T \mathbf{G}_{1B}). \quad (5.12)$$

Using Weyl's inequality, we can derive a lower bound for $\lambda_i(\mathbf{G}_{1A}^T \boldsymbol{\Lambda}_1 \mathbf{G}_{1A})$, $i = 1, \ldots, r$.

$$\begin{aligned}
&\lambda_i(\mathbf{G}_{1A}^T \boldsymbol{\Lambda}_1 \mathbf{G}_{1A}) \ge \lambda_i(\mathbf{G}_{1A}^T \operatorname{diag}(\boldsymbol{\lambda}_i \mathbf{I}_i, \mathbf{O}_{(r-i)\times(r-i)})\mathbf{G}_{1A}) \\
=&\lambda_i\Big(\boldsymbol{\lambda}_i \mathbf{G}_{1A}^T \mathbf{G}_{1A} - \boldsymbol{\lambda}_i \mathbf{G}_{1A}^T \operatorname{diag}(\mathbf{O}_{i\times i}, \mathbf{I}_{r-i})\mathbf{G}_{1A}\Big) \\
\ge&\lambda_r\Big(\boldsymbol{\lambda}_i \mathbf{G}_{1A}^T \mathbf{G}_{1A}\Big) + \lambda_{p+i-r}\Big(- \boldsymbol{\lambda}_i \mathbf{G}_{1A}^T \operatorname{diag}(\mathbf{O}_{i\times i}, \mathbf{I}_{r-i})\mathbf{G}_{1A}\Big) \\
=&\boldsymbol{\lambda}_i \lambda_r(\mathbf{G}_{1A}\mathbf{G}_{1A}^T).
\end{aligned} \quad (5.13)$$

Similarly, we can obtain the upper bound.

$$\begin{aligned}
&\lambda_i(\mathbf{G}_{1A}^T \boldsymbol{\Lambda}_1 \mathbf{G}_{1A}) \\
=&\lambda_i\Big(\mathbf{G}_{1A}^T\big(\operatorname{diag}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{i-1}, \mathbf{O}_{(r-i+1)\times(r-i+1)}) + \operatorname{diag}(\mathbf{O}_{(i-1)\times(i-1)}, \boldsymbol{\lambda}_i, \ldots, \boldsymbol{\lambda}_r)\big)\mathbf{G}_{1A}\Big) \\
\le&\lambda_1(\mathbf{G}_{1A}^T \operatorname{diag}(\mathbf{O}_{(i-1)\times(i-1)}, \boldsymbol{\lambda}_i \mathbf{I}_{r-i+1})\mathbf{G}_{1A}) \le \boldsymbol{\lambda}_i \lambda_1(\mathbf{G}_{1A}\mathbf{G}_{1A}^T).
\end{aligned} \quad (5.14)$$

The inequality (5.12), (5.13) and (5.14) implies that

$$\sup_{1\le i\le r}\Big|\frac{\lambda_i(\mathbf{Y}^T\mathbf{Y})}{n\boldsymbol{\lambda}_i} - 1\Big| \le \max\Big(\Big|\frac{\lambda_1(\mathbf{G}_{1A}\mathbf{G}_{1A}^T)}{n} - 1\Big|, \Big|\frac{\lambda_r(\mathbf{G}_{1A}\mathbf{G}_{1A}^T)}{n} - 1\Big|\Big) + \frac{c_1}{n\boldsymbol{\lambda}_r}\lambda_1(\mathbf{G}_{1B}^T\mathbf{G}_{1B}).$$

We only need to prove the right hand side converges to 0 almost surely.

By Lemma 1, for every $t > 0$, we have

$$\Pr\left(\sqrt{1-\frac{k}{n}} - \sqrt{\frac{r}{n}} - \frac{t}{\sqrt{n}} \leq \sqrt{\frac{\lambda_r(\mathbf{G}_{1A}\mathbf{G}_{1A}^T)}{n}} \leq \sqrt{\frac{\lambda_1(\mathbf{G}_{1A}\mathbf{G}_{1A}^T)}{n}} \leq \sqrt{1-\frac{k}{n}} + \sqrt{\frac{r}{n}} + \frac{t}{\sqrt{n}}\right)$$
$$\geq 1 - 2\exp(-\frac{t^2}{2}). \tag{5.15}$$

Let $t = n^{1/4}$. Since $r = o(n)$, we have

$$\sqrt{1-\frac{k}{n}} - \sqrt{\frac{r}{n}} - \frac{t}{\sqrt{n}} \to 1 \quad \text{and} \quad \sqrt{1-\frac{k}{n}} + \sqrt{\frac{r}{n}} + \frac{t}{\sqrt{n}} \to 1.$$

This, together with Borel-Cantelli lemma, yields

$$\frac{\lambda_r(\mathbf{G}_{1A}\mathbf{G}_{1A}^T)}{n} \to 1 \quad \frac{\lambda_1(\mathbf{G}_{1A}\mathbf{G}_{1A}^T)}{n} \to 1,$$

almost surely. As for $\lambda_1(\mathbf{G}_{1B}^T\mathbf{G}_{1B})$, by Lemma 1, we have

$$\Pr\left(\frac{c_1}{n\boldsymbol{\lambda}_r}\lambda_1(\mathbf{G}_{1B}\mathbf{G}_{1B}^T) \leq \frac{c_1}{n\boldsymbol{\lambda}_r}(\sqrt{n-k} + \sqrt{p-r} + t)^2\right) \geq 1 - \exp(-\frac{t^2}{2}). \tag{5.16}$$

Let $t = n^{1/2}$, since we have assumed $\boldsymbol{\lambda}_r n/p \to \infty$, we have

$$\frac{c_1}{n\boldsymbol{\lambda}_r}\lambda_1(\mathbf{G}_{1B}\mathbf{G}_{1B}^T) \to 0$$

almost surely. Then (5.9) follows.

Inequality (5.10) and (5.11) follows from the fact

$$\lambda_{r+1}(\mathbf{Y}^T\mathbf{Y}) \leq \lambda_1(\mathbf{G}_{1B}^T\boldsymbol{\Lambda}_2\mathbf{G}_{1B}) \leq c_1\lambda_1(\mathbf{G}_{1B}^T\mathbf{G}_{1B}),$$

$$\lambda_{n-k}(\mathbf{Y}^T\mathbf{Y}) \geq \lambda_{n-k}(\mathbf{G}_{1B}^T\boldsymbol{\Lambda}_2\mathbf{G}_{1B}) \geq c_2\lambda_{n-k}(\mathbf{G}_{1B}^T\mathbf{G}_{1B}),$$

and Lemma 1. $\square$

Let $\mathbf{U_Y} = (\mathbf{U}_{\mathbf{Y},1}, \mathbf{U}_{\mathbf{Y},2})$, where $\mathbf{U}_{\mathbf{Y},1}$ and $\mathbf{U}_{\mathbf{Y},2}$ are the first r and last $p-r$ columns of $\mathbf{U_Y}$ respectively.

Lemma 3. *Under the assumptions of Lemma 2, we have*

$$\lambda_{\max}(\mathbf{I}_r - \mathbf{U}_1^T \mathbf{U}_{\mathbf{Y},1} \mathbf{U}_{\mathbf{Y},1}^T \mathbf{U}_1) = O_P(\frac{p}{\boldsymbol{\lambda}_r n}).$$

If in addition, we assume

$$\frac{\log \boldsymbol{\lambda}_r}{n} \to 0, \tag{5.17}$$

then

$$\mathrm{E}\,\lambda_{\max}(\mathbf{I}_r - \mathbf{U}_1^T \mathbf{U}_{\mathbf{Y},1} \mathbf{U}_{\mathbf{Y},1}^T \mathbf{U}_1) = O(\frac{p}{\boldsymbol{\lambda}_r n}).$$

Proof. From $\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{G}_1\mathbf{G}_1^T\boldsymbol{\Lambda}^{1/2}\mathbf{U}^T = \mathbf{U_Y}\mathbf{D}_\mathbf{Y}^2\mathbf{U}_\mathbf{Y}^T$, we have

$$\begin{pmatrix} \boldsymbol{\Lambda}_1^{\frac{1}{2}}\mathbf{G}_{1A}\mathbf{G}_{1A}^T\boldsymbol{\Lambda}_1^{\frac{1}{2}} & \boldsymbol{\Lambda}_1^{\frac{1}{2}}\mathbf{G}_{1A}\mathbf{G}_{1B}^T\boldsymbol{\Lambda}_2^{\frac{1}{2}} \\ \boldsymbol{\Lambda}_2^{\frac{1}{2}}\mathbf{G}_{1B}\mathbf{G}_{1A}^T\boldsymbol{\Lambda}_1^{\frac{1}{2}} & \boldsymbol{\Lambda}_2^{\frac{1}{2}}\mathbf{G}_{1B}\mathbf{G}_{1B}^T\boldsymbol{\Lambda}_2^{\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} \mathbf{U}_1^T\mathbf{U_Y}\mathbf{D}_\mathbf{Y}^2\mathbf{U}_\mathbf{Y}^T\mathbf{U}_1 & \mathbf{U}_1^T\mathbf{U_Y}\mathbf{D}_\mathbf{Y}^2\mathbf{U}_\mathbf{Y}^T\mathbf{U}_2 \\ \mathbf{U}_2^T\mathbf{U_Y}\mathbf{D}_\mathbf{Y}^2\mathbf{U}_\mathbf{Y}^T\mathbf{U}_1 & \mathbf{U}_2^T\mathbf{U_Y}\mathbf{D}_\mathbf{Y}^2\mathbf{U}_\mathbf{Y}^T\mathbf{U}_2 \end{pmatrix}$$

It follows that

$$\boldsymbol{\Lambda}_2^{\frac{1}{2}}\mathbf{G}_{1B}\mathbf{G}_{1B}^T\boldsymbol{\Lambda}_2^{\frac{1}{2}} \geq \lambda_r(\mathbf{Y}^T\mathbf{Y})\mathbf{U}_2^T\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T\mathbf{U}_2.$$

Hence

$$\lambda_1(\mathbf{U}_2^T\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T\mathbf{U}_2) \leq \frac{c_1}{\lambda_r(\mathbf{Y}^T\mathbf{Y})}\lambda_1(\mathbf{G}_{1B}\mathbf{G}_{1B}^T). \tag{5.18}$$

By Lemma 1, for every $t > 0$, we have

$$\Pr\Big(\frac{1}{p}(\sqrt{p-r}-\sqrt{n-k}-t)^2 \leq \frac{1}{p}\lambda_1(\mathbf{G}_{1B}\mathbf{G}_{1B}^T) \leq \frac{1}{p}(\sqrt{p-r}+\sqrt{n-k}+t)^2\Big) \geq 1-2\exp(-\frac{t^2}{2}). \tag{5.19}$$

Let $t = n^{1/2}$, then Borel-Cantelli lemma implies that

$$\frac{1}{p}\lambda_1(\mathbf{G}_{1B}\mathbf{G}_{1B}^T) \to 1 \tag{5.20}$$

almost surely. Then (5.20), (5.18) and Lemma 2 implies that

$$\lambda_1(\mathbf{U}_2^T\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T\mathbf{U}_2) = O_P(\frac{p}{\boldsymbol{\lambda}_r n}).$$

The first conclusion then follows by the following simple relationship

$$\begin{aligned}
&\lambda_{\max}(\mathbf{U}_2^T\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T\mathbf{U}_2) = \lambda_{\max}(\mathbf{U}_{\mathbf{Y},1}^T\mathbf{U}_2\mathbf{U}_2^T\mathbf{U}_{\mathbf{Y},1})\\
=&\lambda_{\max}(\mathbf{U}_{\mathbf{Y},1}^T(\mathbf{I}_p - \mathbf{U}_1\mathbf{U}_1^T)\mathbf{U}_{\mathbf{Y},1}) = \lambda_{\max}(\mathbf{I}_r - \mathbf{U}_{\mathbf{Y},1}^T\mathbf{U}_1\mathbf{U}_1^T\mathbf{U}_{\mathbf{Y},1})\\
=&1 - \lambda_{\min}(\mathbf{U}_{\mathbf{Y},1}^T\mathbf{U}_1\mathbf{U}_1^T\mathbf{U}_{\mathbf{Y},1}) = 1 - \lambda_{\min}(\mathbf{U}_1^T\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T\mathbf{U}_1)\\
=&\lambda_{\max}(\mathbf{I}_r - \mathbf{U}_1^T\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T\mathbf{U}_1).
\end{aligned}$$

Next we prove the second conclusion of the lemma. In (5.15), (5.16) and (5.19), we take $t = \sqrt{2\log(\boldsymbol{\lambda}_r n/p)}$. Then these inequalities, (5.18) and condition (5.17) implies that

$$\lambda_1(\mathbf{U}_2^T\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T\mathbf{U}_2) \le \frac{c^* p}{\boldsymbol{\lambda}_r n}$$

with probability at least $1 - 3p/\boldsymbol{\lambda}_r n$ for some constant c^*. Since $\lambda_1(\mathbf{U}_2^T\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T\mathbf{U}_2) \le 1$, we have

$$\mathrm{E}\,\lambda_1(\mathbf{U}_2^T\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T\mathbf{U}_2) \le \frac{c^* p}{\boldsymbol{\lambda}_r n}(1 - \frac{3p}{\boldsymbol{\lambda}_r n}) + \frac{3p}{\boldsymbol{\lambda}_r n} = O(\frac{p}{\boldsymbol{\lambda}_r n}).$$

This completes the proof. $\square$

Lemma 4. *Under the assumptions of Lemma 2, we have the following upper and lower bound for* $\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}$.

$$\lambda_i(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}) \geq \boldsymbol{\lambda}_{i+n-k}, \quad i = 1, \ldots, p-n+k, \tag{5.21}$$

$$\lambda_i\big(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\big) = O_P(\frac{\boldsymbol{\lambda}_1 p}{\boldsymbol{\lambda}_r n}) + c_1, \quad i = 1, \ldots, r, \tag{5.22}$$

$$\lambda_i(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}) \leq \boldsymbol{\lambda}_i, \quad i = r+1, \ldots, p. \tag{5.23}$$

Proof. The inequality (5.23) follows from the fact $\mathbf{I}_p - \mathbf{P}_\mathbf{Y} \leq \mathbf{I}_p$. The inequality (5.21) follows from the fact that $\text{Rank}(\mathbf{P}_\mathbf{Y}) \leq n-k$ and Weyl's inequality. As for inequality (5.22), note that the positive eigenvalues of $\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}$ equal to the positive eigenvalues of $(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})$. We write $(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})$ as the sum of two terms

$$\begin{aligned}
&(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}) \\
=&(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}_1\mathbf{\Lambda}_1\mathbf{U}_1^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}) + (\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}) \stackrel{def}{=} \mathbf{R}_1 + \mathbf{R}_2.
\end{aligned}$$

Lemma 3 can be applied to control the largest eigenvalue of $\mathbf{R}_1$:

$$\begin{aligned}
&\lambda_1\big(\mathbf{R}_1\big) = \lambda_1\big(\mathbf{\Lambda}_1^{1/2}\mathbf{U}_1^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}_1\mathbf{\Lambda}_1^{1/2}\big) \leq \lambda_1\big(\mathbf{\Lambda}_1^{1/2}\mathbf{U}_1^T(\mathbf{I}_p - \mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T)\mathbf{U}_1\mathbf{\Lambda}_1^{1/2}\big) \\
\leq&\boldsymbol{\lambda}_1\lambda_1\big(\mathbf{U}_1^T(\mathbf{I}_p - \mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T)\mathbf{U}_1\big) = \boldsymbol{\lambda}_1\lambda_1\big(\mathbf{I}_r - \mathbf{U}_1^T\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T\mathbf{U}_1\big) = O_P(\frac{\boldsymbol{\lambda}_1 p}{\boldsymbol{\lambda}_r n}).
\end{aligned}$$

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

$$\lambda_i(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}) \leq \lambda_1(\mathbf{R}_1) + \lambda_1(\mathbf{R}_2) = O_P(\frac{\boldsymbol{\lambda}_1 p}{\boldsymbol{\lambda}_r n}) + c_1.$$

$\square$

Lemma 5. *Under the assumptions of Theorem 1, we have*

$$\operatorname{tr}\left(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\right)=\frac{p-r-n+k}{p-r}\operatorname{tr}(\mathbf{\Lambda}_2)+o_P(\sqrt{p}), \tag{5.24}$$

and

$$\operatorname{tr}\left(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\right)^2=(1+o_P(1))\operatorname{tr}(\mathbf{\Lambda}_2^2). \tag{5.25}$$

Proof. By Lemma 4, we have

$$\sum_{i=n-k+1}^{p}\boldsymbol{\lambda}_i^2\leq\operatorname{tr}\left(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\right)^2\leq r(O_P(\frac{\boldsymbol{\lambda}_1 p}{\boldsymbol{\lambda}_r n})+c_1)^2+\sum_{i=r+1}^{p}\boldsymbol{\lambda}_i^2.$$

Hence

$$\begin{aligned}&\left|\operatorname{tr}\left(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\right)^2-\sum_{i=r+1}^{p}\boldsymbol{\lambda}_i^2\right|\\&\leq\max\left(\sum_{i=r+1}^{n-k}\boldsymbol{\lambda}_i^2,r(O_P(\frac{\boldsymbol{\lambda}_1 p}{\boldsymbol{\lambda}_r n})+c_1)^2\right)\\&\leq r(O_P(\frac{\boldsymbol{\lambda}_1 p}{\boldsymbol{\lambda}_r n})+c_1)^2+O(n)=o_P(p).\end{aligned}$$

Then (5.25) holds.

Now we prove (5.24). Note that

$$\operatorname{tr}\left(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\right)=\operatorname{tr}\left(\mathbf{\Lambda}_1^{1/2}\mathbf{U}_1^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}_1\mathbf{\Lambda}_1^{1/2}\right)+\operatorname{tr}\left(\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\right).$$

By Lemma 4, we have

$$\operatorname{tr}\left(\mathbf{\Lambda}_1^{1/2}\mathbf{U}_1^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}_1\mathbf{\Lambda}_1^{1/2}\right)=O_P(\frac{\boldsymbol{\lambda}_1 pr}{\boldsymbol{\lambda}_r n})=o_P(\sqrt{p}).$$

The second term can be written as $\operatorname{tr}\left(\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\right) = \operatorname{tr}(\mathbf{\Lambda}_2)-\operatorname{tr}(\mathbf{P}_\mathbf{Y}\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^T)$. For $\operatorname{tr}(\mathbf{P}_\mathbf{Y}\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^T)$, we have

$$\left|\operatorname{tr}(\mathbf{P}_\mathbf{Y}\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^T)-\frac{n-k}{p-r}\operatorname{tr}(\mathbf{\Lambda}_2)\right| = \left|\operatorname{tr}\left(\mathbf{P}_\mathbf{Y}\mathbf{U}\big(\mathbf{\Lambda}_2-\frac{1}{p-r}(\operatorname{tr}\mathbf{\Lambda}_2)\mathbf{I}_{p-r}\big)\mathbf{U}^T\right)\right|$$
$$\leq\sqrt{\operatorname{tr}\left(\mathbf{P}_\mathbf{Y}^2\right)}\sqrt{\operatorname{tr}\left(\mathbf{\Lambda}_2-\frac{1}{p-r}\big(\operatorname{tr}\mathbf{\Lambda}_2\big)\mathbf{I}_{p-r}\right)^2} = \sqrt{(n-k)\operatorname{tr}\left(\mathbf{\Lambda}_2-\frac{1}{p-r}\big(\operatorname{tr}\mathbf{\Lambda}_2\big)\mathbf{I}_{p-r}\right)^2} = o(\sqrt{p}).$$

Hence

$$\operatorname{tr}\left(\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\right) = \frac{p-r-n+k}{p-r}\operatorname{tr}(\mathbf{\Lambda}_2)+o(\sqrt{p}).$$

Then (5.24) holds. $\square$

Proof of Theorem 1. We deal with the three terms of (5.7) separately. Lemma (4) implies that the first term satisfies the Lyapunov condition

$$\frac{\lambda_1\Big(\big(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\big)^2\Big)}{\operatorname{tr}\Big(\big(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\big)^2\Big)} = \frac{\big(O_P(\frac{\boldsymbol{\lambda}_1 p}{\boldsymbol{\lambda}_r n})+c_1\big)^2}{(1+o_P(1))\operatorname{tr}(\mathbf{\Lambda}_2)} \xrightarrow{P} 0.$$

Apply Lyapunov central limit theorem conditioning on $\mathbf{P}_\mathbf{Y}$, we have

$$\Big(\operatorname{tr}\Big(\big(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\big)^2\Big)\Big)^{-1/2}$$
$$\big(\mathbf{G}_2^T\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{G}_2-\operatorname{tr}\big(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\big)\mathbf{I}_{k-1}\big) \xrightarrow{\mathcal{L}} \mathbf{W}_{k-1}.$$

This, combined with Lemma 5 and Slutsky's theorem, yields

$$\frac{1}{\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)}}\big(\mathbf{G}_2^T\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{G}_2-\tfrac{p-r-n+k}{p-r}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_{k-1}\big) \xrightarrow{\mathcal{L}} \mathbf{W}_{k-1}.$$

Next we show that the cross term of (5.7) is negligible. Note that

$$
\begin{aligned}
&\mathrm{E}[\|\mathbf{C}^T\Xi^T(\mathbf{I}_p-\mathbf{P}_{Z\tilde{J}})\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{G}_2\|_F^2|\mathbf{Y}] \\
=&(k-1)\,\mathrm{tr}(\mathbf{C}^T\Xi^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\Xi\mathbf{C}) \\
\le&(k-1)\lambda_1\big((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\big)\|\Xi\mathbf{C}\|_F^2 \\
\le&(k-1)\lambda_1\big(\boldsymbol{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\boldsymbol{\Lambda}^{1/2}\big)\|\Xi\mathbf{C}\|_F^2 \\
=&(k-1)O_P(\frac{\lambda_1 p}{\lambda_r n}+c_1)\|\Xi\mathbf{C}\|_F^2 \\
=&(k-1)O_P\big(\frac{\lambda_1\sqrt{p}}{\lambda_r n}+\frac{c_1}{\sqrt{p}}\big)\sqrt{p}\|\Xi\mathbf{C}\|_F^2=o_P(p),
\end{aligned}
$$

where the last equality holds since we have assumed $\frac{1}{\sqrt{p}}\|\Xi\mathbf{C}\|_F^2 = O(1)$. Hence $\|\mathbf{C}^T\Xi^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{G}_2\|_F^2 = o_P(p)$. Now,

$$
\frac{1}{\sqrt{\mathrm{tr}(\boldsymbol{\Lambda}_2^2)}}\big(\mathbf{C}^T\mathbf{Y}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{Y}\mathbf{C}-\tfrac{p-r-n+k}{p-r}\,\mathrm{tr}(\boldsymbol{\Lambda}_2)\mathbf{I}_{k-1}-\mathbf{C}^T\Xi^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\Xi\mathbf{C}\big)\xrightarrow{\mathcal{L}}\mathbf{W}_{k-1}.
$$

Equivalently,

$$
\begin{aligned}
&\frac{1}{\sqrt{\mathrm{tr}(\boldsymbol{\Lambda}_2^2)}}\Big(\mathbf{C}^T\mathbf{Y}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{Y}\mathbf{C}-\frac{p-r-n+k}{p-r}\,\mathrm{tr}(\boldsymbol{\Lambda}_2)\mathbf{I}_{k-1}\Big) \\
\sim&\frac{1}{\sqrt{\mathrm{tr}(\boldsymbol{\Lambda}_2^2)}}\mathbf{C}^T\Xi^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\Xi\mathbf{C}+\mathbf{W}_{k-1}+o_P(1).
\end{aligned}
$$

The conclusion follows by taking the maximum eigenvalue. □

Proof of Proposition 1. First we consider the case of $r>0$. By the construction of $\hat{r}$,

$$
\{\hat{r}=r\}\supseteq\{\frac{\lambda_r(\mathbf{Y}^T\mathbf{Y})}{\lambda_{r+1}(\mathbf{Y}^T\mathbf{Y})}\ge\gamma_n\}\cap\{\frac{\lambda_{r+1}(\mathbf{Y}^T\mathbf{Y})}{\lambda_{n-k}(\mathbf{Y}^T\mathbf{Y})}\le\gamma_n\}.
$$

Suppose $0 < \epsilon < 1$ is a fixed number. By assumption, there exists an n_0^*, for $n \geq n_0^*$, $\gamma_n \leq (1-\epsilon)n\boldsymbol{\lambda}_r/(c_1 p)$ and $\gamma_n \geq (1+\epsilon)c_1/c_2$. Thus

$$\{\hat{r} = r\} \supseteq \{\frac{\lambda_r(\mathbf{Y}^T\mathbf{Y})}{\lambda_{r+1}(\mathbf{Y}^T\mathbf{Y})} \geq (1-\epsilon)\frac{n\boldsymbol{\lambda}_r}{c_1 p}\} \cap \{\frac{\lambda_{r+1}(\mathbf{Y}^T\mathbf{Y})}{\lambda_{n-k}(\mathbf{Y}^T\mathbf{Y})} \leq (1+\epsilon)\frac{c_1}{c_2}\}.$$

Lemma 2 implies that almost surely, there exists an n_0, for $n \geq n_0$, we have

$$\frac{\lambda_r(\mathbf{Y}^T\mathbf{Y})}{\lambda_{r+1}(\mathbf{Y}^T\mathbf{Y})} \geq (1-\epsilon)\frac{n\boldsymbol{\lambda}_r}{c_1 p}, \quad \frac{\lambda_{r+1}(\mathbf{Y}^T\mathbf{Y})}{\lambda_{n-k}(\mathbf{Y}^T\mathbf{Y})} \leq (1+\epsilon)\frac{c_1}{c_2}.$$

This yields $\Pr(\hat{r} = r) \to 1$ for $r > 0$. The case of $r = 0$ can be similarly proved by noting that

$$\{\hat{r} = r\} \supseteq \{\frac{\lambda_1(\mathbf{Y}^T\mathbf{Y})}{\lambda_{n-k}(\mathbf{Y}^T\mathbf{Y})} \leq \gamma_n\}.$$

□

Proof of Proposition 2. Since $\hat{r}$ is a consistent estimator of r, we only need to prove

$$\frac{1}{n-k}\sum_{i=r+1}^{n-k} \lambda_i(\mathbf{Y}^T\mathbf{Y}) = \mathrm{tr}(\boldsymbol{\Lambda}_2) + O_P(\sqrt{p}).$$

Note that

$$\mathbf{Y}^T\mathbf{Y} = \mathbf{G}_1^T\boldsymbol{\Lambda}\mathbf{G}_1 = \mathbf{G}_{1A}^T\boldsymbol{\Lambda}_1\mathbf{G}_{1A} + \mathbf{G}_{1B}^T\boldsymbol{\Lambda}_2\mathbf{G}_{1B}.$$

By Weyl's inequality, for $i = r+1, \ldots, n-k$, we have

$$\lambda_i(\mathbf{G}_{1B}^T\boldsymbol{\Lambda}_2\mathbf{G}_{1B}) \leq \lambda_i(\mathbf{Y}^T\mathbf{Y}) \leq \lambda_{i-r}(\mathbf{G}_{1B}^T\boldsymbol{\Lambda}_2\mathbf{G}_{1B}).$$

It follows that

$$\sum_{i=r+1}^{n-k} \lambda_i(\mathbf{G}_{1B}^T \mathbf{\Lambda}_2 \mathbf{G}_{1B}) \leq \sum_{i=r+1}^{n-k} \lambda_i(\mathbf{Y}^T \mathbf{Y}) \leq \sum_{i=1}^{n-k-r} \lambda_i(\mathbf{G}_{1B}^T \mathbf{\Lambda}_2 \mathbf{G}_{1B}).$$

Hence

$$\Big|\sum_{i=r+1}^{n-k} \lambda_i(\mathbf{Y}^T \mathbf{Y}) - \operatorname{tr}(\mathbf{G}_{1B}^T \mathbf{\Lambda}_2 \mathbf{G}_{1B})\Big| \leq r\lambda_1(\mathbf{G}_{1B}^T \mathbf{\Lambda}_2 \mathbf{G}_{1B}) = O_P(rp).$$

But central limit theorem implies that

$$\operatorname{tr}(\mathbf{G}_{1B}^T \mathbf{\Lambda}_2 \mathbf{G}_{1B}) - (n-k)\operatorname{tr}(\mathbf{\Lambda}_2) = O_P(\sqrt{np}).$$

Thus

$$\frac{1}{n-k}\sum_{i=r+1}^{n-k} \lambda_i(\mathbf{Y}^T \mathbf{Y}) = \operatorname{tr}(\mathbf{\Lambda}_2) + \frac{1}{n-k}\Big|\sum_{i=r+1}^{n-k} \lambda_i(\mathbf{Y}^T \mathbf{Y}) - (n-k)\operatorname{tr}(\mathbf{\Lambda}_2)\Big|$$

$$= \operatorname{tr}(\mathbf{\Lambda}_2) + O_P(\frac{rp}{n}) = \operatorname{tr}(\mathbf{\Lambda}_2) + O_P(\sqrt{p}),$$

where the last equality follows form Assumption 1. $\square$

Proof of Proposition 3. Let $\mathbf{U}_{\mathbf{Y},1;(i,j)}$ be the first r columns of $\mathbf{U}_{\mathbf{Y};(i,j)}$. Let $\mathbf{U}_{\mathbf{Y},2;(i,j)}$ be a $p \times (p-r)$ orthogonal matrix satisfying $\mathbf{U}_{\mathbf{Y},2;(i,j)}\mathbf{U}_{\mathbf{Y},2;(i,j)}^T = \mathbf{I}_p - \mathbf{U}_{\mathbf{Y},1;(i,j)}\mathbf{U}_{\mathbf{Y},1;(i,j)}^T$. Then by Lemma 3, we have

$$\lambda_1(\mathbf{U}_1^T \mathbf{U}_{\mathbf{Y},2;(i,j)}\mathbf{U}_{\mathbf{Y},2;(i,j)}^T \mathbf{U}_1) = \lambda_1(\mathbf{I}_r - \mathbf{U}_1^T \mathbf{U}_{\mathbf{Y},1;(i,j)}\mathbf{U}_{\mathbf{Y},1;(i,j)}^T \mathbf{U}_1) = O_P(\frac{p}{\boldsymbol{\lambda}_r n}).$$

First, we prove that w_{ij}^2 is an approximation of $Y_i^T \mathbf{U}_2 \mathbf{U}_2^T Y_j$. For $1 \leq$

$i < j \le n-k$, define $\epsilon_{ij} = w_{ij} - Y_i^T \mathbf{U}_2 \mathbf{U}_2^T Y_j$, then we have

$$\begin{aligned}
\epsilon_{ij} =& Y_i^T (\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)} \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T - \mathbf{U}_2 \mathbf{U}_2^T) Y_j \\
=& Y_i^T (\mathbf{U}_1 \mathbf{U}_1^T + \mathbf{U}_2 \mathbf{U}_2^T)(\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)} \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T - \mathbf{U}_2 \mathbf{U}_2^T)(\mathbf{U}_1 \mathbf{U}_1^T + \mathbf{U}_2 \mathbf{U}_2^T) Y_j \\
=& Y_i^T \mathbf{U}_2 \mathbf{U}_2^T (\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)} \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T - \mathbf{U}_2 \mathbf{U}_2^T) \mathbf{U}_2 \mathbf{U}_2^T Y_j \\
&+ Y_i^T \mathbf{U}_1 \mathbf{U}_1^T \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)} \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T \mathbf{U}_2 \mathbf{U}_2^T Y_j \\
&+ Y_i^T \mathbf{U}_2 \mathbf{U}_2^T \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)} \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T \mathbf{U}_1 \mathbf{U}_1^T Y_j \\
&+ Y_i^T \mathbf{U}_1 \mathbf{U}_1^T \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)} \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T \mathbf{U}_1 \mathbf{U}_1^T Y_j \\
=& Y_i^T \mathbf{U}_2 \mathbf{U}_2^T (\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)} \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T - \mathbf{U}_{\mathbf{Y},2;(i,j)} \mathbf{U}_{\mathbf{Y},2;(i,j)}^T) \mathbf{U}_2 \mathbf{U}_2^T Y_j \\
&+ Y_i^T \mathbf{U}_2 \mathbf{U}_2^T (\mathbf{U}_{\mathbf{Y},2;(i,j)} \mathbf{U}_{\mathbf{Y},2;(i,j)}^T - \mathbf{U}_2 \mathbf{U}_2^T) \mathbf{U}_2 \mathbf{U}_2^T Y_j \\
&+ Y_i^T \mathbf{U}_1 \mathbf{U}_1^T \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)} \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T \mathbf{U}_2 \mathbf{U}_2^T Y_j \\
&+ Y_i^T \mathbf{U}_2 \mathbf{U}_2^T \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)} \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T \mathbf{U}_1 \mathbf{U}_1^T Y_j \\
&+ Y_i^T \mathbf{U}_1 \mathbf{U}_1^T \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)} \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T \mathbf{U}_1 \mathbf{U}_1^T Y_j \\
\stackrel{def}{=}& \epsilon_{ij}^{(1)} + \epsilon_{ij}^{(2)} + \epsilon_{ij}^{(3)} + \epsilon_{ij}^{(4)} + \epsilon_{ij}^{(5)}.
\end{aligned} \tag{5.26}$$

We deal with the five terms separately. First we deal with $\epsilon_{ij}^{(1)}$. It can be seen that Y_i, Y_j and $\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}$ are mutually independent and $\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)} \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T - \mathbf{U}_{\mathbf{Y},2;(i,j)} \mathbf{U}_{\mathbf{Y},2;(i,j)}^T$ is a projection matrix whose rank is not larger than $n-$

$k-2-r$. Then

$$
\begin{aligned}
&\mathrm{E}(\epsilon_{ij}^{(1)})^2\\
=&\mathrm{E}\,\mathrm{tr}(\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T(\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T-\mathbf{U}_{\mathbf{Y},2;(i,j)}\mathbf{U}_{\mathbf{Y},2;(i,j)}^T)\mathbf{U}_2\mathbf{\Lambda}_2^{1/2})^2\\
\le& c_1^2(n-k-2-r)=o(p).
\end{aligned}
$$

Next we deal with $\epsilon_{ij}^{(2)}$. we have

$$
\begin{aligned}
&\mathrm{E}(\epsilon_{ij}^{(2)})^2\\
=&\mathrm{E}\,\mathrm{tr}(\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T(\mathbf{U}_{\mathbf{Y},2;(i,j)}\mathbf{U}_{\mathbf{Y},2;(i,j)}^T-\mathbf{U}_2\mathbf{U}_2^T)\mathbf{U}_2\mathbf{\Lambda}_2^{1/2})^2\\
\le& c_1^2\,\mathrm{E}\,\mathrm{tr}(\mathbf{U}_2^T(\mathbf{U}_{\mathbf{Y},2;(i,j)}\mathbf{U}_{\mathbf{Y},2;(i,j)}^T-\mathbf{U}_2\mathbf{U}_2^T)\mathbf{U}_2)^2\\
=& c_1^2\,\mathrm{E}\,\mathrm{tr}(\mathbf{I}_{p-r}-\mathbf{U}_2^T\mathbf{U}_{\mathbf{Y},2;(i,j)}\mathbf{U}_{\mathbf{Y},2;(i,j)}^T\mathbf{U}_2)^2\\
=& c_1^2\,\mathrm{E}\,\mathrm{tr}(\mathbf{I}_{r}-\mathbf{U}_1^T\mathbf{U}_{\mathbf{Y},1;(i,j)}\mathbf{U}_{\mathbf{Y},1;(i,j)}^T\mathbf{U}_1)^2\\
\le& c_1^2 r\,\mathrm{E}\,\lambda_1^2(\mathbf{I}_{r}-\mathbf{U}_1^T\mathbf{U}_{\mathbf{Y},1;(i,j)}\mathbf{U}_{\mathbf{Y},1;(i,j)}^T\mathbf{U}_1)\\
=& O(r(\frac{p}{\boldsymbol{\lambda}_r n})^2)=o(p).
\end{aligned}
$$

Note that $\epsilon_{i,j}^{(3)}$ and $\epsilon_{i,j}^{(4)}$ have the same distribution, we have

$$
\begin{aligned}
&\mathrm{E}(\epsilon_{ij}^{(3)})^2 = \mathrm{E}(\epsilon_{ij}^{(4)})^2 \\
=&\mathrm{E}\,\mathrm{tr}(\mathbf{\Lambda}_1^{1/2}\mathbf{U}_1^T\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^T\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T\mathbf{U}_1\mathbf{\Lambda}_1^{1/2}) \\
\leq&c_1\boldsymbol{\lambda}_1\,\mathrm{E}\,\mathrm{tr}(\mathbf{U}_1^T\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T\mathbf{U}_2\mathbf{U}_2^T\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T\mathbf{U}_1) \\
\leq&c_1\boldsymbol{\lambda}_1\,\mathrm{E}\,\mathrm{tr}(\mathbf{U}_1^T\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T\mathbf{U}_1) \\
\leq&c_1\boldsymbol{\lambda}_1\,\mathrm{E}\,\mathrm{tr}(\mathbf{U}_1^T\mathbf{U}_{\mathbf{Y},2;(i,j)}\mathbf{U}_{\mathbf{Y},2;(i,j)}^T\mathbf{U}_1) \\
\leq&c_1\boldsymbol{\lambda}_1 r\frac{p}{\boldsymbol{\lambda}_r n} = o(p).
\end{aligned}
$$

As for $\epsilon_{i,j}^{(5)}$, we have

$$
\begin{aligned}
&\mathrm{E}(\epsilon_{ij}^{(5)})^2 \\
=&\mathrm{E}\,\mathrm{tr}(\mathbf{\Lambda}_1^{1/2}\mathbf{U}_1^T\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T\mathbf{U}_1\mathbf{\Lambda}_1^{1/2})^2 \\
\leq&\boldsymbol{\lambda}_1^2\,\mathrm{E}\,\mathrm{tr}(\mathbf{U}_1^T\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T\mathbf{U}_1)^2 \\
\leq&\boldsymbol{\lambda}_1^2\,\mathrm{E}\,\mathrm{tr}(\mathbf{U}_1^T\mathbf{U}_{\mathbf{Y},2;(i,j)}\mathbf{U}_{\mathbf{Y},2;(i,j)}^T\mathbf{U}_1)^2 \\
\leq&\boldsymbol{\lambda}_1^2 r(\frac{p}{\boldsymbol{\lambda}_r n})^2 = o(p).
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\widehat{\mathrm{tr}(\mathbf{\Lambda}_2^2)} \\
=&\frac{2}{(n-k)(n-k-1)}\sum_{1\leq i<j\leq n-k}(Y_i^T\mathbf{U}_2\mathbf{U}_2^TY_j)^2 \\
&+\frac{2}{(n-k)(n-k-1)}\sum_{1\leq i<j\leq n-k}\big((\sum_{l=1}^5\epsilon_{ij}^{(l)})(Y_i^T\mathbf{U}_2\mathbf{U}_2^TY_j)+(\sum_{l=1}^5\epsilon_{ij}^{(l)})^2\big).
\end{aligned}
$$

We have

$$
\begin{aligned}
&\mathrm{E}\Big|\frac{2}{(n-k)(n-k-1)}\sum_{1\le i<j\le n-k}\big((\sum_{l=1}^{5}\epsilon_{ij}^{(l)})(Y_i^T\mathbf{U}_2\mathbf{U}_2^TY_j)+(\sum_{l=1}^{5}\epsilon_{ij}^{(l)})^2\big)\Big| \\
\le&\,\mathrm{E}\Big|(\sum_{l=1}^{5}\epsilon_{12}^{(l)})(Y_1^T\mathbf{U}_2\mathbf{U}_2^TY_2)+(\sum_{l=1}^{5}\epsilon_{12}^{(l)})^2\Big| \\
\le&\sqrt{\mathrm{E}\,\big(\sum_{l=1}^{5}\epsilon_{12}^{(l)}\big)^2\,\mathrm{E}\,\big(Y_1^T\mathbf{U}_2\mathbf{U}_2^TY_2\big)^2}+\mathrm{E}(\sum_{l=1}^{5}\epsilon_{12}^{(l)})^2=o(p).
\end{aligned}
$$

It follows that

$$
\widehat{\mathrm{tr}(\mathbf{\Lambda}_2^2)}=\frac{2}{(n-k)(n-k-1)}\sum_{1\le i<j\le n-k}(Y_i^T\mathbf{U}_2\mathbf{U}_2^TY_j)^2+o_P(p).
$$

Now we only need to prove that

$$
\frac{2}{(n-k)(n-k-1)}\sum_{1\le i<j\le n-k}(Y_i^T\mathbf{U}_2\mathbf{U}_2^TY_j)^2
$$

is ratio consistent. Since $\mathrm{E}(Y_i^T\mathbf{U}_2\mathbf{U}_2^TY_j)^2=\mathrm{tr}(\mathbf{\Lambda}_2^2)$ for $i<j$, we have

$$
\mathrm{E}\,\frac{2}{(n-k)(n-k-1)}\sum_{1\le i<j\le n-k}(Y_i^T\mathbf{U}_2\mathbf{U}_2^TY_j)^2=\mathrm{tr}(\mathbf{\Lambda}_2^2).
$$

To prove the proposition, we only need to show that

$$
\mathrm{Var}\Big(\frac{2}{(n-k)(n-k-1)}\sum_{1\le i<j\le n-k}(Y_i^T\mathbf{U}_2\mathbf{U}_2^TY_j)^2\Big)=o(\mathrm{tr}^2(\mathbf{\Lambda}_2^2)).
$$

Note that

$$
\begin{aligned}
&\mathrm{E}\Big(\frac{2}{(n-k)(n-k-1)}\sum_{1\le i<j\le n-k}(Y_i^T\mathbf{U}_2\mathbf{U}_2^TY_j)^2\Big)^2\\
=&\frac{4}{(n-k)^2(n-k-1)^2}\Big(\sum_{1\le i<j\le n-k}(Y_i^T\mathbf{U}_2\mathbf{U}_2^TY_j)^2\Big)^2\\
=&\frac{4}{(n-k)^2(n-k-1)^2}\,\mathrm{E}\Big(\sum_{i<j}(Y_i^T\mathbf{U}_2\mathbf{U}_2^TY_j)^4\\
&+\sum_{i<j,k<l:\{i,j\}\{k,l\}=\emptyset}(Y_i^T\mathbf{U}_2\mathbf{U}_2^TY_j)^2(Y_k^T\mathbf{U}_2\mathbf{U}_2^TY_l)^2\\
&+2\sum_{i<j<k}\big((Y_i^T\mathbf{U}_2\mathbf{U}_2^TY_j)^2(Y_i\mathbf{U}_2\mathbf{U}_2^TY_k)^2\\
&+(Y_i^T\mathbf{U}_2\mathbf{U}_2^TY_j)^2(Y_j\mathbf{U}_2\mathbf{U}_2^TY_k)^2+(Y_i^T\mathbf{U}_2\mathbf{U}_2^TY_k)^2(Y_j\mathbf{U}_2\mathbf{U}_2^TY_k)^2\big)\Big)\\
=&\frac{4}{(n-k)^2(n-k-1)^2}\Big(\frac{(n-k)(n-k-1)}{2}\big(6\operatorname{tr}(\mathbf{\Lambda}_2^4)+3\operatorname{tr}^2(\mathbf{\Lambda}_2^2)\big)\\
&+\frac{(n-k)(n-k-1)(n-k-2)(n-k-3)}{4}\operatorname{tr}^2(\mathbf{\Lambda}_2^2)\\
&+(n-k)(n-k-1)(n-k-2)\big(2\operatorname{tr}(\mathbf{\Lambda}_2^4)+\operatorname{tr}^2(\mathbf{\Lambda}_2^2)\big)\Big)\\
=&\operatorname{tr}^2(\mathbf{\Lambda}_2^2)(1+o(1)).
\end{aligned}
$$

It follows that

$$
\mathrm{Var}\Big(\frac{2}{(n-k)(n-k-1)}\sum_{1\le i<j\le n-k}(Y_i^T\mathbf{U}_2\mathbf{U}_2^TY_j)^2\Big)=o(\operatorname{tr}^2(\mathbf{\Lambda}_2^2)).
$$

This completes the proof. □

Supplementary Materials

Contain the brief description of the online supplementary materials.

Acknowledgements

Write the acknowledgements here.

References

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

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

Cai, T. T. and Y. Xia (2014). High-dimensional sparse manova. *Journal of Multivariate Analysis 131*(4), 174–196.

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

Davidson, K. R. and S. J. Szarek (2001). *Handbook of the Geometry of Banach Spaces*, Volume 1. Amsterdam: North-Holland. Handbook of the Geometry of Banach Spaces.

Fan, J., Y. Fan, and J. Lv (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics 147*(1), 186–197.

Feng, L., C. Zou, and Z. Wang (2016). Multivariate-sign-based high-dimensional tests for the two-sample location problem. *Journal of the American Statistical Association*.

Lopes, M. E., L. J. Jacob, and M. J. Wainwright (2015). A more powerful two-sample test in high dimensions using random projection. *Statistics*, 1206–1214.

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

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

Roy, S. N. (1953, jun). On a heuristic method of test construction and its use in multivariate analysis. *The Annals of Mathematical Statistics 24*(2), 220–238.

Schott, J. R. (2007). Some high-dimensional tests for a one-way manova. *Journal of Multivariate Analysis 98*(9), 1825–1839.

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

Srivastava, M. S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis 100*(3), 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.

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

Tsai, C.-A. and J. J. Chen (2009). Multivariate analysis of variance test for gene set analysis. *Bioinformatics 25*(7), 897.

Verstynen, T., J. Diedrichsen, N. Albert, P. Aparicio, and R. B. Ivry (2005). Ipsilateral motor cortex activity during unimanual hand movements relates to task complexity. *Journal of Neurophysiology 93*(3), 1209–1222.

Zhao, J. and X. Xu (2016). A generalized likelihood ratio test for normal mean when p is greater than n. *Computational Statistics & Data Analysis*.

first author affiliation

E-mail: (first author email)

second author affiliation

E-mail: (second author email)