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 principal, we propose a generalized likelihood ratio test. The critical value is determined by permutation method. We introduce an algorthm for permuting procedure, whose complexity does not depend on data dimension. The limiting distribution of the test statistic is derived in two different setting: non-spiked covariance and spiked covariance. Theoretical results and simulation studies show that the test is particularly powerful under spiked covariance.

*Key words and phrases:*

**1. Introduction** Suppose there are $k$ ($k \geq 2$) groups of $p$ dimensional data. Within the $i$th 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 $\xi_i$ 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 $\mathbf{H}\mathbf{G}^{-1}$.

| | |
|---|---|
| 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: | $\mathrm{tr}[\mathbf{H}\mathbf{G}^{-1}]$ |
| Roy's maximum root: | $\lambda_{\max}(\mathbf{H}\mathbf{G}^{-1})$ |

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). 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

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

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

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

Where $\Omega = (\omega)_{ij} = \mathbf{\Sigma}^{-1}$ is the precision matrix. When $\Omega$ 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 if $p > n-k$, 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 dependent.

Following Zhao and Xu (2016)'s methodology, we proposed a generalized likelihood ratio test for hypothesis (1.1). To understand the power behavior of the new test, we derive the asymptotic distribution of the new statistic under two different settings. In first setting, we assume the eigenvalues of $\mathbf{\Sigma}$ are bounded. It's a common assumption in high dimensional statistics. In fact, most existing tests for hypothesis (1.1) imposed conditions which prevent from large leading eigenvalues of $\mathbf{\Sigma}$. However, when the correlations between variables are determined by a small number of factors, $\mathbf{\Sigma}$ is spiked in the sense that a few leading eigenvalues are much larger than the others. See, for example Cai et al. (2013) and Shen et al. (2013). We then derive the asymptotic distribution of the test statistic under spiked covariance. From the theoretical results we give, it can be seen that the new test is particularly powerful under spiked covariance. 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

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. Define

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

where $\mathbf{1}_{n_i}$ is an $n_i$-dimensional vector with all elements equal to 1. Then the matrices $\mathbf{I}_n - \mathbf{J}\mathbf{J}^T$, $\mathbf{J}\mathbf{J}^T - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$ and $\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$ are three $n \times n$ projection matrices which are pairwise orthogonal with rank $n-k$, $k-1$ and 1 respectively. 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

$$\mathbf{G} = \mathbf{X}(\mathbf{I}_n - \mathbf{J}\mathbf{J}^T)\mathbf{X}^T = \mathbf{Y}\mathbf{Y}^T.$$

Note that $\mathbf{I}_k - \frac{1}{n}\mathbf{J}^T\mathbf{1}_n\mathbf{1}_n^T\mathbf{J}$ is a $k \times k$ projection matrix with rank $k-1$.

Let $\mathbf{C}$ 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 we have $\mathbf{C}\mathbf{C}^T = \mathbf{I}_k - \frac{1}{n}\mathbf{J}^T\mathbf{1}_n\mathbf{1}_n^T\mathbf{J}$ and thus $\mathbf{H}$ has representation

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

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.

### 2.1 Roy's maximum root

Roy's maximum root test statistic is derived in Roy (1953) as an application of his union intersection principle. Roy's union intersection principle can be decomposed into 3 main steps:

1. Decompose the hypothesis $H_0$ and $H_1$ into component hypotheses

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

   where $\Gamma$ is an index set.

2. For each $\gamma$, construct a component test for $H_{0\gamma}$ against $H_{1\gamma}$.

3. Accept $H_0$ if all component tests accept the null hypotheses. Or equivalently, reject $H_0$ if any component test reject the null hypothesis.

Roy's union intersection principle is particularly useful when $H_0$ and $H_1$ themselves are complicated but can be decomposed into a class of simple hypotheses.

The decomposition in step 1 of union intersection principle is often induced by a data transformation. The data matrix $\mathbf{X}$ is not easy to deal with since it is multivariate. Note that there is a one-to-one mapping between the data $\mathbf{X}$ and the set $\{\mathbf{X}_a : a \in \mathbb{R}^p, a^T a = 1\}$, where $\mathbf{X}_a = a^T\mathbf{X}$ is the univariate data obtained by projecting $\mathbf{X}$ on direction $a$. This naturally induces the decomposition

$$H_0 = \bigcap_{a\in\mathbb{R}^p, a^T a=1} H_{0a} \quad \text{and} \quad K = \bigcup_{a\in\mathbb{R}^p, a^T a=1} H_{1a},$$

where

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

Based on $\mathbf{X}_a$, the likelihood ratio test statistic for $H_{0a}$ against $H_{1a}$ is

$$\mathrm{LR}_a = \left(1 + \frac{a^T\mathbf{H}a}{a^T\mathbf{G}a}\right)^{n/2}.$$

By Roy's union intersection principle, $H_0$ is rejected when $\max_{a^T a=1} \mathrm{LR}_a$ is large. If $p \leq n-k$, $\mathbf{G}$ is invertible and $\max_{a^T a=1} \mathrm{LR}_a = (1+\lambda_{\max}(\mathbf{H}\mathbf{G}^{-1}))^{n/2}$, which is an increasing function of Roy's maximum root test statistic.

### 2.2 A new test

Despite the wide use of Roy's maximum root, it is not defined for $p > n-k$. In fact, if $p > n-k$, $G$ is not invertible and $\max_{a^T a=1} \mathrm{LR}_a = +\infty$.

The derivation of Roy's maximum root implies that it is based on the likelihood ratio of projected data $\mathbf{X}_a$. From a likelihood point view, log likelihood ratio is an estimator of the KL divergence between the alternative distribution and the null distribution. Thus, by maximizing $\mathrm{LR}_a$, one obtains the direction $a^* = \arg\max_{a^T a=1} \mathrm{LR}_a$ which hopefully distinct the null distribution and the alternative distribution of $\mathbf{X}_a$.

While it is hard to generalize Roy's maximum root to high dimensional setting, $a^*$ can be formally generalized to high dimensional setting. Note that with probability 1, we have $\{a : \mathrm{LR}_a = +\infty\} = \{a : a^T \mathbf{G} a = 0\}$.

When $p > n-k$, we have the following formal argument

$$\begin{aligned}
a^* &= \underset{a^T a=1}{\arg\max}\, \mathrm{LR}_a \\
&= \underset{a^T a=1, \mathrm{LR}_a=+\infty}{\arg\max} \Big(1+\frac{a^T\mathbf{H}a}{a^T\mathbf{G}a}\Big)^{n/2} \\
&= \underset{a^T a=1, a^T Ga=0}{\arg\max} \Big(1+\frac{a^T\mathbf{H}a}{0}\Big)^{n/2} \\
&= \underset{a^T a=1, a^T Ga=0}{\arg\max}\ a^T\mathbf{H}a.
\end{aligned}$$

This motivates us to propose the test statistic

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

We reject the null hypothesis when $T$ is large enough.

Next we derive the explicit forms of the test statistic. Let $\mathbf{Y} = \mathbf{U_Y}\mathbf{D_Y}\mathbf{V_Y}^T$ 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 $\mathbf{P_Y} = \mathbf{U_Y}\mathbf{U_Y}^T$ be the projection matrix on the column space of $\mathbf{Y}$. By Proposition 3, we have

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

Next we introduce another form of $T$. 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

$$\left(\mathbf{J}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{J}\right)^{-1} = \mathbf{J}^T\mathbf{X}^T\mathbf{X}\mathbf{J} - \mathbf{J}^T\mathbf{X}^T\mathbf{X}\tilde{\mathbf{J}}(\tilde{\mathbf{J}}^T\mathbf{X}^T\mathbf{X}\tilde{\mathbf{J}})^{-1}\tilde{\mathbf{J}}^T\mathbf{X}^T\mathbf{X}\mathbf{J} = \mathbf{J}^T\mathbf{X}^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{X}\mathbf{J}.$$

Thus,

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

We will use (2.2) for theoretical analysis, (2.3) for computation.

## 2.3 Permutation method

Permutation method is a powerful tool to determine the critical value of a test statistic. The test procedure resulting from permutation method is exact as long as the null distribution of observations are exchangeable (Romano, 1990). The major down-side to permutation method is that it can be computationally intensive. Fortunately, the permutation method can be computationally fast. By expression (2.3), a permuted statistic can be written as

$$T(\mathbf{X}\Gamma) = \lambda_{\max}\Big(\mathbf{C}^T\left(\mathbf{J}^T\Gamma^T(\mathbf{X}^T\mathbf{X})^{-1}\Gamma\mathbf{J}\right)^{-1}\mathbf{C}\Big), \tag{2.4}$$

where $\Gamma$ 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 (2.4).

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.

## 3. Theoretical results

In this section, we investigate the asymptotic behavior of our test statistic when $p$ is much larger than $n$. In high dimensional setting, it is a common phenomenon that the asymptotic distribution of a statistic relies on the covariance structure (Ma et al., 2015). We shall derive the asymptotic distribution of our statistic under two different covariance structures: non-spiked covariance and spiked covariance.

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. $N(0,2)$. The random matrix $\mathbf{W}_{k-1}$ will appear in the asymptotic distribution of $T(\mathbf{X})$. Let $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ be the eigenvalue decomposition of

$\boldsymbol{\Sigma}$, where $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_p)$.

The following theorem establishes the asymptotic distribution of $T(\mathbf{X})$ under non-spiked covariance.

**Theorem 1.** *Suppose $p/n \to \infty$, $c_1 \geq \boldsymbol{\lambda}_1 \geq \cdots \geq \boldsymbol{\lambda}_p \geq c_2$ and*

$$\mathrm{tr}\left(\boldsymbol{\Sigma} - \frac{1}{p}(\mathrm{tr}\,\boldsymbol{\Sigma})\mathbf{I}_p\right)^2 = o\left(\frac{p}{n}\right).$$

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

$$\frac{T(\mathbf{X}) - \frac{p-n+k}{p}\,\mathrm{tr}(\boldsymbol{\Sigma})}{\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)}} \sim \lambda_{\max}\left(\mathbf{W}_{k-1} + \frac{1}{\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)}}\mathbf{C}^T\Xi^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\Xi\mathbf{C}\right) + o_P(1).$$

For some real world problems, variables are heavily correlated with common factors, then a few eigenvalues of $\boldsymbol{\Sigma}$ are significantly larger than the others Ma et al. (2015). To characterize this correlation, we make the following assumption for the eigenvalues of $\boldsymbol{\Sigma}$.

**Assumption 1.** *Let $r$ be a fixed integer. 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 $c_2$. For large eigenvalues $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r$, we assume*

$$\frac{\lambda_r(\boldsymbol{\Sigma})n}{p} \to \infty, \quad \frac{\lambda_1(\boldsymbol{\Sigma})^2 p}{\lambda_r(\boldsymbol{\Sigma})^2 n^2} \to 0.$$

To state the asymptotic distribution of $T(\mathbf{X})$ under Assumption 1. We need following notations. Let $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2)$ where $\mathbf{U}_1$ is $p \times r$ and $\mathbf{U}_2$ is

$p \times (p-r)$. Let $\mathbf{\Lambda}_1 = \operatorname{diag}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r)$ and $\mathbf{\Lambda}_2 = \operatorname{diag}(\boldsymbol{\lambda}_{r+1}, \ldots, \boldsymbol{\lambda}_p)$. Then $\mathbf{\Sigma} = \mathbf{U}_1\mathbf{\Lambda}_1\mathbf{U}_1^T + \mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^T$.

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

**Theorem 2.** *Under Assumption* (1), *suppose $p/n \to \infty$ and*

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

*Then under local alternative*

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

*we have*

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

### 3.1 Variance estimation

Under non-spiked covariance, an unbiased estimator of $\operatorname{tr}(\mathbf{\Sigma}^2)$ is

$$\widehat{\operatorname{tr}(\mathbf{\Sigma}^2)} = \frac{2}{(n-k)(n-k-1)}\sum_{1\le i<j\le n-k}(Y_i^TY_j)^2.$$

**Proposition 1.** *As $n \to \infty$,*

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

*Proof.* Since $\mathrm{E}(Y_i^T Y_j)^2 = \mathrm{tr}(\boldsymbol{\Sigma}^2)$, we have

$$\mathrm{E}\left(\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}\right) = \mathrm{tr}(\boldsymbol{\Sigma}^2).$$

We only need to show that

$$\mathrm{Var}\left(\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}\right) = o(\mathrm{tr}^2(\boldsymbol{\Sigma}^2)).$$

Note that

$$\begin{aligned}
&\mathrm{E}\left(\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}\right)^2 = \frac{4}{(n-k)^2(n-k-1)^2}\Big(\sum_{1\le i<j\le n-k}(Y_i^T Y_j)^2\Big)^2 \\
=&\frac{4}{(n-k)^2(n-k-1)^2}\,\mathrm{E}\Big(\sum_{i<j}(Y_i^T Y_j)^4 + \sum_{i<j,k<l:\{i,j\}\{k,l\}=\emptyset}(Y_i^T Y_j)^2(Y_k^T Y_l)^2 \\
&+ 2\sum_{i<j<k}\big((Y_i^T Y_j)^2(Y_i Y_k)^2 + (Y_i^T Y_j)^2(Y_j Y_k)^2 + (Y_i^T Y_k)^2(Y_j Y_k)^2\big)\Big) \\
=&\frac{4}{(n-k)^2(n-k-1)^2}\Big(\frac{(n-k)(n-k-1)}{2}\big(6\,\mathrm{tr}(\boldsymbol{\Sigma}^4) + 3\,\mathrm{tr}^2(\boldsymbol{\Sigma}^2)\big) \\
&+ \frac{(n-k)(n-k-1)(n-k-2)(n-k-3)}{4}\,\mathrm{tr}^2(\boldsymbol{\Sigma}^2) \\
&+ (n-k)(n-k-1)(n-k-2)\big(2\,\mathrm{tr}(\boldsymbol{\Sigma}^4) + \mathrm{tr}^2(\boldsymbol{\Sigma}^2)\big)\Big) \\
=&\,\mathrm{tr}^2(\boldsymbol{\Sigma}^2)(1+o(1)).
\end{aligned}$$

Then

$$\mathrm{Var}\left(\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}\right) = \mathrm{E}\left(\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}\right)^2 - \left(\mathrm{E}\left(\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}\right)\right)^2 = o(\mathrm{tr}^2(\boldsymbol{\Sigma}^2)).$$

This completes the proof. $\square$

We use a leave-two-out estimator to estimate $\mathrm{tr}(\mathbf{\Lambda}_2^2)$. Let $w_{ij}$ be the $(i,j)$th element of $\mathbf{Y}^T\mathbf{U}_{\mathbf{Y},2;(i,j)}\mathbf{U}_{\mathbf{Y},2;(i,j)}^T\mathbf{Y}$. We use

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

to estimate $\mathrm{tr}(\mathbf{\Lambda}_2^2)$.

**Proposition 2.** *Under Assumption* (1) *and $p/n\to\infty$, we have*

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

*Proof.* For $1\le i<j\le n-k$, we have

$$\begin{aligned}
w_{ij} - Y_i^T\mathbf{U}_2\mathbf{U}_2^TY_j =& Y_i^T(\mathbf{U}_{\mathbf{Y},2;(i,j)}\mathbf{U}_{\mathbf{Y},2;(i,j)}^T - \mathbf{U}_2\mathbf{U}_2^T)Y_j\\
=& Y_i^T\mathbf{U}_2\mathbf{U}_2^T(\mathbf{U}_{\mathbf{Y},2;(i,j)}\mathbf{U}_{\mathbf{Y},2;(i,j)}^T - \mathbf{U}_2\mathbf{U}_2^T)\mathbf{U}_2\mathbf{U}_2^TY_j\\
&+ Y_i^T\mathbf{U}_1\mathbf{U}_1^T\mathbf{U}_{\mathbf{Y},2;(i,j)}\mathbf{U}_{\mathbf{Y},2;(i,j)}^T\mathbf{U}_2\mathbf{U}_2^TY_j\\
&+ Y_i^T\mathbf{U}_2\mathbf{U}_2^T\mathbf{U}_{\mathbf{Y},2;(i,j)}\mathbf{U}_{\mathbf{Y},2;(i,j)}^T\mathbf{U}_1\mathbf{U}_1^TY_j\\
&+ Y_i^T\mathbf{U}_1\mathbf{U}_1^T\mathbf{U}_{\mathbf{Y},2;(i,j)}\mathbf{U}_{\mathbf{Y},2;(i,j)}^T\mathbf{U}_1\mathbf{U}_1^TY_j
\end{aligned}$$

For the first term, we have

$$\begin{aligned}
&Y_i^T\mathbf{U}_2\mathbf{U}_2^T(\mathbf{U}_{\mathbf{Y},2;(i,j)}\mathbf{U}_{\mathbf{Y},2;(i,j)}^T - \mathbf{U}_2\mathbf{U}_2^T)\mathbf{U}_2\mathbf{U}_2^TY_j\\
=&O_P(1)\sqrt{p}\lambda_1(\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T(\mathbf{U}_{\mathbf{Y},2;(i,j)}\mathbf{U}_{\mathbf{Y},2;(i,j)}^T - \mathbf{U}_2\mathbf{U}_2^T)\mathbf{U}_2\mathbf{\Lambda}_2^{1/2})\\
=&O_P(1)\sqrt{p}\lambda_1(\mathbf{U}_2^T\mathbf{U}_{\mathbf{Y},2;(i,j)}\mathbf{U}_{\mathbf{Y},2;(i,j)}^T\mathbf{U}_2 - \mathbf{I}_{p-r}) = O_P(\sqrt{p}\frac{p}{\lambda_r n}).
\end{aligned}$$

For the second term, we have

$$
\begin{aligned}
& Y_i^T \mathbf{U}_1 \mathbf{U}_1^T \mathbf{U}_{\mathbf{Y},2;(i,j)} \mathbf{U}_{\mathbf{Y},2;(i,j)}^T \mathbf{U}_2 \mathbf{U}_2^T Y_j \\
=& O_P(1)\sqrt{r} \|\mathbf{\Lambda}_1^{1/2} \mathbf{U}_1^T \mathbf{U}_{\mathbf{Y},2;(i,j)} \mathbf{U}_{\mathbf{Y},2;(i,j)}^T \mathbf{U}_2 \mathbf{\Lambda}_2^{1/2}\| \\
=& O_P(1)\sqrt{r\lambda_1} \|\mathbf{U}_1^T \mathbf{U}_{\mathbf{Y},2;(i,j)}\| \\
=& O_P(1)\sqrt{r\lambda_1}\sqrt{\frac{p}{\lambda_r n}} \\
=& O_P(\sqrt{\frac{pr\lambda_1}{\lambda_r n}}).
\end{aligned}
$$

For the third term, we have

$$
\begin{aligned}
& Y_i^T \mathbf{U}_1 \mathbf{U}_1^T \mathbf{U}_{\mathbf{Y},2;(i,j)} \mathbf{U}_{\mathbf{Y},2;(i,j)}^T \mathbf{U}_1 \mathbf{U}_1^T Y_j \\
=& O_P(1)\sqrt{r} \|\mathbf{\Lambda}_1^{1/2} \mathbf{U}_1^T \mathbf{U}_{\mathbf{Y},2;(i,j)} \mathbf{U}_{\mathbf{Y},2;(i,j)}^T \mathbf{U}_1 \mathbf{\Lambda}_1^{1/2}\| \\
=& O_P(1)\sqrt{r}\lambda_1 \|\mathbf{U}_1^T \mathbf{U}_{\mathbf{Y},2;(i,j)} \mathbf{U}_{\mathbf{Y},2;(i,j)}^T \mathbf{U}_1\| \\
=& O_P(\sqrt{r}\lambda_1 \frac{p}{\lambda_r n}).
\end{aligned}
$$

Thus,

$$
w_{ij} = Y_i^T \mathbf{U}_2 \mathbf{U}_2^T Y_j + O_P(\frac{p\sqrt{p}}{\lambda_r n} + \sqrt{\frac{pr\lambda_1}{\lambda_r n}} + \frac{\sqrt{r}\lambda_1 p}{\lambda_r n}) = Y_i^T \mathbf{U}_2 \mathbf{U}_2^T Y_j + o_P(\sqrt{p})
$$

It follows that $w_{ij} = (Y_i^T \mathbf{U}_2 \mathbf{U}_2^T Y_j)^2 + o_P(p)$. Thus

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

Then the conclusion follows from Proposition 1. □

## 4. Comparison with existing tests

In this section, we revist some existing high dimensional tests in the point of view of union intersection principle. This will help to compare the proposed test and other tests.

For high dimensional testing problem, the step 1 of Roy's union intersection principle is often induced by a data projection and component tests are univariate problem. For univariate testing problem, likelihood ratio test statistic is often the best choice. As we have obtained a set of component test statistic, we need to summarize them to obtain a global test statistic. Union intersection principle suggest using the maximum of component test statistics. But it is not the only choice. In summary, a generalized union intersection principle consists the following 3 steps.

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

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

2. Construct a test statistic $T_\gamma$ for $H_{0\gamma}$ against $H_{1\gamma}$.

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

For step 1, we consider two different constructions of data projection.

i Consider the set $\{\mathbf{X}_i = e_i^T \mathbf{X} : i = 1, \ldots, p\}$, where $e_i$ is the $i$th standard basis.

ii Consider the set $\{\mathbf{X}_a = a^T \mathbf{X} : i = 1, a \in \mathbb{R}^p, a^T a = 1\}$.

For step 3, we consider two different strategy of summarization.

I Integrating $T_\gamma$ according some measure $\mu(\gamma)$ and use $\int_\gamma T_\gamma \, \mu(d\gamma)$ as global test statistic.

II Use $\max_{\gamma \in \Gamma} T_\gamma$ as global test statistic.

First, we consider using construction i in step 1. If component statistics

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

are used in step 2 and strategy I with $\mu$ being the uniform measure on $1, \ldots, p$ is used in step 3, one obtains $T_{SC}$. If 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). By using data $\Omega^{-1}\mathbf{X}$, the test statistic $T_{CX}$ can be obtained with strategy II. The component test statistic corresponding $T_{CX}$ is similar to likelihood ratio tests. Maybe $T_{CX}$ can be improved by replace their component tests by likelihood ratio tests.

We can see that the test statistics resulting from the construction i mostly requires that certain prior information about the covariance structure of data is known. 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 $\Omega$. This may due to that the construction i chooses a orthogonal basis of $\mathbb{R}^p$.

Next, we consider using construction ii in step 1. Suppose the likelihood ratio test $T_a = a^T \mathbf{H} a / a^T \mathbf{G} a$ is used in step 2. In step 3, if we choose strategy I with $\mu$ equals to the uniform distribution on the sphere, then the test statistic becomes

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

Although it is hard to give the 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 $M^{-1} \sum_{i=1}^{M} a^T \mathbf{H} a / a^T \mathbf{G} a$. A similar random projection method is proposed by Lopes et al. (2015) for $k = 2$. There analysis and simulations show that such random projection method has relative good performance especially when variables are correlated.

Our new test statistic comes from construction ii in step 1, the likelihood ratio test statistics in step 2 and strategy II in step 3. Theorems 1 and 2 al-low us to analyze the properties of the proposed test. Suppose $\sqrt{n_i}\mu_i$ is from prior distribution $N_p(0, \psi\mathbf{I}_p)$, $i = 1, \dots, k$. Then $\psi^{-1}\mathbf{C}^T\Xi^T\Xi\mathbf{C}$ is distributed as $\text{Wishart}_{k-1}(p, \mathbf{I}_{k-1})$ (Wishart distribution with freedom $p$ and parameter $\mathbf{I}_{k-1}$) and $\psi^{-1}\mathbf{C}^T\Xi^T\mathbf{P}_\mathbf{Y}\Xi\mathbf{C}$ is distributed as $\text{Wishart}_{k-1}(n-k, \mathbf{I}_{k-1})$. In this case, we have

$$\psi^{-1}\mathbf{C}^T\Xi^T(\mathbf{I}_P - \mathbf{P}_\mathbf{Y})\Xi\mathbf{C} = (1 + o_P(1))\psi^{-1}\mathbf{C}^T\Xi^T\Xi\mathbf{C}.$$

If the conditions of Theorem 1 hold and $k = 2$, the asymptotic power of the proposed test is the same as that of Bai and Saranadasa (1996)and Chen and Qin (2010)'s method. Since the method of Schott (2007) is a direct generalization of Bai and Saranadasa (1996)'s method, it can be shown the asymptotic power of the proposed test is the same as that of Schott (2007) for general $k$. Next, suppose the covariance matrix is spiked and the conditions of Theorem 2 hold. Theorem 2 implies that the proposed test does not depend on large eigenvalues $\lambda_1, \dots, \lambda_r$ while other existing test procedures are negatively affected by large eigenvalues $\lambda_1, \dots, \lambda_r$. Thus, the new test has particular good power behavior when $\lambda_1, \dots, \lambda_r$ are large. This property is not surprising since our statistic is from construction ii. As a result, our statistic has a wider applicable range compared to the tests from construction i.

## 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 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 $\kappa$ 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, $\kappa$ 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.

## 6. 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 proved that the proposed test has similar asymptotic power with $T_{SC}$ under non-spiked covariance. On the other hand, if covariance matrix is spiked, the asymptotic power of the proposed test is not affected by the large eigenvalues. We give a discussion of existing MANOVA tests from union intersection principle point of view, this explains why the proposed test has good power behavior.

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 |

## Appendix

**Proposition 3.** *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* $\mathbf{A} = \mathbf{U_A D_A V_A}^T$ *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* $\mathbf{P_A} = \mathbf{U_A U_A}^T$ *be the projection on the column space of* $\mathbf{A}$*. Then*

$$\max_{a^T a=1, a^T \mathbf{A}\mathbf{A}^T a=0} a^T \mathbf{B} a = \lambda_{\max}\big(\mathbf{B}(\mathbf{I}_p - \mathbf{P_A})\big). \tag{6.5}$$

*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

$$\max_{a^T a=1, a^T \mathbf{A}\mathbf{A}^T a=0} a^T \mathbf{B} a = \max_{a^T a=1, \mathbf{P_A} a=0} a^T (\mathbf{I}_p - \mathbf{P_A})\mathbf{B}(\mathbf{I}_p - \mathbf{P_A})a, \tag{6.6}$$

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 $\alpha_1$ 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, $\alpha_1$ satisfies the constraint of (6.6) and (6.6) 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 $\lambda_{\max}\big((\mathbf{I} - \mathbf{P_A})\mathbf{B}(\mathbf{I} - \mathbf{P_A})\big) = \lambda_{\max}\big(\mathbf{B}(\mathbf{I} - \mathbf{P_A})\big)$.

$\square$

**Proves of the main results**

**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* $t > 0$*,*

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

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

$$\begin{aligned}\mathbf{C}^T\mathbf{J}^T\mathbf{X}^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{XJC} =&\mathbf{G}_2^T\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{G}_2 + \xi_f^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\xi_f +\\ &\xi_f^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{G}_2 + \mathbf{G}_2^T\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\xi_f.\end{aligned} \tag{6.7}$$

The first term of (6.7) can be represented as

$$\mathbf{G}_2^T\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{G}_2 = \sum_{i=1}^{p} \lambda_i(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2})\xi_i\xi_i^T, \tag{6.8}$$

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

*Proof of Theorem 1.* We deal with the three terms of (6.7) separately. The first term of (6.7) relies on the eigenvalues of $\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}$. Trivially, the eigenvalues of $\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}$ have upper bound

$$\lambda_i(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}) \leq \boldsymbol{\lambda}_i, \quad i=1,\ldots,p. \tag{6.9}$$

On the other hand, we note that $\mathbf{P}_\mathbf{Y}$ has rank $n-k$. By Weyl's inequality, we have lower bound

$$\lambda_i(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}) \geq \boldsymbol{\lambda}_{i+n-k}, \quad i=1,\ldots,p-n+k. \tag{6.10}$$

Inequality (6.9) and (6.10) imply that

$$\frac{\lambda_1^2(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2})}{\sum_{i=1}^p \lambda_i^2(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2})} \leq \frac{c_1}{c_2(p-n+k)} \to 0.$$

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

$$\Big(\operatorname{tr}\big(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\big)^2\Big)^{-1/2}$$
$$\Big(\mathbf{G}_2^T\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{G}_2-\operatorname{tr}\big(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\big)\mathbf{I}_{k-1}\Big) \xrightarrow{\mathcal{L}} \mathbf{W}_{k-1}. \tag{6.11}$$

Another implication of (6.9) and (6.10) is that

$$\sum_{i=n-k+1}^p \boldsymbol{\lambda}_i^2 \leq \operatorname{tr}\Big((\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2})^2\Big) \leq \operatorname{tr}(\mathbf{\Lambda}^2).$$

It follows that

$$\operatorname{tr}\Big((\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2})^2\Big) = \operatorname{tr}(\mathbf{\Sigma}^2)+O_P(n) = \big(1+O_P(\frac{n}{p})\big)\operatorname{tr}(\mathbf{\Sigma}^2). \tag{6.12}$$

To deal with $\operatorname{tr}(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}) = \operatorname{tr}(\mathbf{\Sigma})-\operatorname{tr}(\mathbf{P}_{\mathbf{Y}}\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T)$, we note that

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

Hence

$$\operatorname{tr}(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}) = \frac{p-n+k}{p}\operatorname{tr}(\mathbf{\Sigma})+o(\sqrt{p}). \tag{6.13}$$

It follows from (6.11), (6.12) and (6.13) that

$$\begin{aligned}
&\Big(\sum_{i=1}^{p}\lambda_i^2(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2})\Big)^{-1/2} \\
&\Big(\mathbf{G}_2^T\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{G}_2 - \sum_{i=1}^{p}\lambda_i(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2})\mathbf{I}_{k-1}\Big) \\
=&\Big((1+O_P(\frac{n}{p}))\operatorname{tr}(\mathbf{\Sigma}^2)\Big)^{-1/2}\Big(\mathbf{G}_2^T\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{G}_2 - \big(\frac{p-n+k}{p}\operatorname{tr}(\mathbf{\Sigma})+O_P(\sqrt{p})\big)\mathbf{I}_{k-1}\Big)
\end{aligned}$$

By Slutsky's theorem, we have that

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

Note that

$$\begin{aligned}
&\operatorname{E}\left(\|\mathbf{C}^T\Xi^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{G}_2\|_F^2\right) \\
=&(k-1)\operatorname{E}\Big(\operatorname{tr}\left(\mathbf{C}^T\Xi^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\Xi\mathbf{C}\right)\Big) \\
\leq&(k-1)\operatorname{E}\Big(\lambda_1\big((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\big)\Big)\|\Xi\mathbf{C}\|_F^2 \\
\leq&(k-1)\boldsymbol{\lambda}_1\|\Xi\mathbf{C}\|_F^2 \leq (k-1)c_1\|\Xi\mathbf{C}\|_F^2 = o(p),
\end{aligned}$$

we have

$$\frac{1}{\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)}}\Big(\mathbf{C}^T\mathbf{J}^T\mathbf{X}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{X}\mathbf{J}\mathbf{C}-\frac{p-n+k}{p}\operatorname{tr}(\boldsymbol{\Sigma})\mathbf{I}_{k-1}-\mathbf{C}^T\Xi^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\Xi\mathbf{C}\Big)\xrightarrow{\mathcal{L}}\mathbf{W}_{k-1}.$$

Equivalently, we have

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

Then the conclusion follows by taking the maximum eigenvalue. □

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.** *Under the Assumptions of Theorem 2, $r=o(n)$,*

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

*almost surely.*

*Proof.* Note that

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

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

$$\lambda_i(\mathbf{G}_{1A}^T\boldsymbol{\Lambda}_1\mathbf{G}_{1A})\le\lambda_i(\mathbf{Y}^T\mathbf{Y})\le\lambda_i(\mathbf{G}_{1A}^T\boldsymbol{\Lambda}_1\mathbf{G}_{1A})+c_1\lambda_1(\mathbf{G}_{1B}^T\mathbf{G}_{1B}).\quad(6.14)$$

---

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

$$\begin{aligned}
&\lambda_i(\mathbf{G}_{1A}^T\mathbf{\Lambda}_1\mathbf{G}_{1A}) \geq \lambda_i(\mathbf{G}_{1A}^T \operatorname{diag}(\boldsymbol{\lambda}_i\mathbf{I}_i, \mathbf{O}_{(r-i)\times(r-i)})\mathbf{G}_{1A}) \\
=&\lambda_i\Big(\boldsymbol{\lambda}_i\mathbf{G}_{1A}^T\mathbf{G}_{1A} - \boldsymbol{\lambda}_i\mathbf{G}_{1A}^T \operatorname{diag}(\mathbf{O}_{i\times i}, \mathbf{I}_{r-i})\mathbf{G}_{1A}\Big) \\
\geq&\lambda_r\Big(\boldsymbol{\lambda}_i\mathbf{G}_{1A}^T\mathbf{G}_{1A}\Big) + \lambda_{p+i-r}\Big(-\boldsymbol{\lambda}_i\mathbf{G}_{1A}^T \operatorname{diag}(\mathbf{O}_{i\times i}, \mathbf{I}_{r-i})\mathbf{G}_{1A}\Big) \\
=&\boldsymbol{\lambda}_i\lambda_r(\mathbf{G}_{1A}\mathbf{G}_{1A}^T).
\end{aligned} \tag{6.15}$$

Similarly, we can obtain the upper bound.

$$\begin{aligned}
&\lambda_i(\mathbf{G}_{1A}^T\mathbf{\Lambda}_1\mathbf{G}_{1A}) \\
=&\lambda_i\Big(\mathbf{G}_{1A}^T\big(\operatorname{diag}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{i-1}, \mathbf{O}_{(r-i+1)\times(r-i+1)}) + \operatorname{diag}(\mathbf{O}_{(i-1)\times(i-1)}, \boldsymbol{\lambda}_i, \ldots, \boldsymbol{\lambda}_r)\big)\mathbf{G}_{1A}\Big) \\
\leq&\lambda_1(\mathbf{G}_{1A}^T \operatorname{diag}(\mathbf{O}_{(i-1)\times(i-1)}, \boldsymbol{\lambda}_i\mathbf{I}_{r-i+1})\mathbf{G}_{1A}) \leq \boldsymbol{\lambda}_i\lambda_1(\mathbf{G}_{1A}\mathbf{G}_{1A}^T).
\end{aligned} \tag{6.16}$$

The inequality (6.14), (6.15) and (6.16) implies that

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

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

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

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

Let $t = n^{1/4}$, then Borel-Cantelli lemma implies

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

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

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

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. This completes the proof. $\square$

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

**Lemma 3.** *Under the Assumptions of Theorem 2, we have*

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

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

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

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

$$\lambda_1(\mathbf{U}_2^T\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T\mathbf{U}_2) \leq \frac{c_1}{\lambda_r(\mathbf{Y}^T\mathbf{Y})}\lambda_1(\mathbf{G}_{1B}\mathbf{G}_{1B}^T). \tag{6.17}$$

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

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

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

$$\frac{1}{p}\lambda_1(\mathbf{G}_{1B}\mathbf{G}_{1B}^T) \to 1 \tag{6.18}$$

almost surely. Then (6.18), (6.17) 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 conclusion then follows by the following simple relationship

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

□

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

$$\lambda_i(\boldsymbol{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\boldsymbol{\Lambda}^{1/2}) \leq \lambda_i(\boldsymbol{\Lambda}). \tag{6.19}$$

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

$$\lambda_i(\boldsymbol{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\boldsymbol{\Lambda}^{1/2}) \geq \lambda_{i+n-k}(\boldsymbol{\Lambda}). \tag{6.20}$$

Next, we need to give an upper bound for $\lambda_i(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2})$, $i=1,\ldots,r$. 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 $(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})$. 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

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

Note that

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

The last equality follows by Lemma 3. Thus, for $i=1,\ldots,r$, we have

$$
\lambda_i\big((\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\big)\le\lambda_1(\mathbf{R}_1)+\lambda_1(\mathbf{R}_2)=O_P(\frac{\lambda_1 p}{\lambda_r n})+c_1. \tag{6.21}
$$

As a consequence of the bound (6.19), (6.20) and (6.21), we have

$$
\sum_{i=n-k+1}^{p}\lambda_i^2\le\operatorname{tr}\big((\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\big)^2\le r(O_P(\frac{\lambda_1 p}{\lambda_r n})+c_1)^2+\sum_{i=r+1}^{p}\lambda_i^2.
$$

Hence

$$
\Big|\operatorname{tr}\big((\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\big)^2-\sum_{i=r+1}^{p}\lambda_i^2\Big|\le r(O_P(\frac{\lambda_1 p}{\lambda_r n})+c_1)^2+O(n). \tag{6.22}
$$

Note that

$$
\operatorname{tr}(\mathbf{R}_2)=\operatorname{tr}(\mathbf{\Lambda}_2)-\operatorname{tr}(\mathbf{P}_\mathbf{Y}\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^T).
$$

and

$$\left|\operatorname{tr}(\mathbf{P_Y}\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^T)-\frac{n-k}{p-r}\operatorname{tr}(\boldsymbol{\Lambda}_2)\right|=\left|\operatorname{tr}\left(\mathbf{P_Y}\mathbf{U}\big(\boldsymbol{\Lambda}_2-\frac{1}{p-r}(\operatorname{tr}\boldsymbol{\Lambda}_2)\mathbf{I}_{p-r}\big)\mathbf{U}^T\right)\right|$$

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

Hence

$$\operatorname{tr}(\mathbf{R}_2)=\frac{p-r-n+k}{p-r}\operatorname{tr}(\boldsymbol{\Lambda}_2)+o(\sqrt{p}).$$

Then

$$\left|\operatorname{tr}[(\mathbf{R}_1+\mathbf{R}_2)]-\frac{p-r-n+k}{p-r}\operatorname{tr}(\boldsymbol{\Lambda}_2)\right|\leq rO_P\big(\frac{\lambda_1 p}{\lambda_r n}\big)+o(\sqrt{p}). \qquad (6.23)$$

Equation (6.22) and (6.23), combined with the assumptions, yield

$$\operatorname{tr}\left((\mathbf{I}_p-\mathbf{P_Y})\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P_Y})\right)^2=(1+o_P(1))\operatorname{tr}(\boldsymbol{\Lambda}_2),$$

and

$$\operatorname{tr}\left((\mathbf{I}_p-\mathbf{P_Y})\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P_Y})\right)=\frac{p-r-n+k}{p-r}\operatorname{tr}(\boldsymbol{\Lambda}_2)+o_P(\sqrt{p}).$$

Now we have the Lyapunov condition

$$\frac{\lambda_1\Big(\left((\mathbf{I}_p-\mathbf{P_Y})\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P_Y})\right)^2\Big)}{\operatorname{tr}\left(\left((\mathbf{I}_p-\mathbf{P_Y})\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P_Y})\right)^2\right)}=\frac{\big(O_P(\frac{\lambda_1 p}{\lambda_r n})+c_1\big)^2}{(1+o_P(1))\operatorname{tr}(\boldsymbol{\Lambda}_2)}\xrightarrow{P}0.$$

Apply Lyapunov central limit theorem conditioning on $\mathbf{P_Y}$, we have

$$\left(\operatorname{tr}\left(\left((\mathbf{I}_p-\mathbf{P_Y})\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P_Y})\right)^2\right)\right)^{-1/2}$$

$$\left(\mathbf{G}_2^T\boldsymbol{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P_Y})\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{G}_2-\operatorname{tr}\left((\mathbf{I}_p-\mathbf{P_Y})\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P_Y})\right)\mathbf{I}_{k-1}\right)\xrightarrow{\mathcal{L}}\mathbf{W}_{k-1},$$

where $\mathbf{W}_{k-1}$ 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

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

As for the cross term of (6.7), we have

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

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

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

Equivalently, we have

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

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

## 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.

Davis, C. and W. M. Kahan (1970). The rotation of eigenvectors by a perturbation. iii. *Siam Journal on Numerical Analysis 7*(1), 1–46.

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.

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)