Statistica Sinica

LEAST FAVORABLE DIRECTION TEST FOR MULTIVARIATE ANALYSIS OF VARIANCE IN HIGH DIMENSION

Rui Wang, Xingzhong Xu

Beijing Institute of Technology

Abstract: This paper considers the problem of multivariate analysis of variance for normal samples. When the sample dimension is larger than the sample size, the classical likelihood ratio test is not defined since the likelihood function is unbounded. Based on the unboundedness of the likelihood function, we propose a new test called least favorable direction test. The asymptotic null distribution of the test statistic is derived and the local asymptotic power function of the test is also given. The asymptotic power function and simulations show that the proposed test has particular high power when variables are strongly correlated.

Key words and phrases: High dimensional data, least favorable direction test, multivariate analysis of variance, principal component analysis, spiked covariance.

1. Introduction

Suppose there are k ($k \geq 2$) independent samples of p-dimensional data. Within the ith sample ($1 \leq i \leq k$), the observations $\{X_{ij}\}_{j=1}^{n_i}$ are independent and identically distributed (iid) as $\mathcal{N}_p(\xi_i, \mathbf{\Sigma})$, 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 with 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:	$\lambda_1(\mathbf{HG}^{-1})$

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, Verstynen et al. (2005) and Tsai and Chen (2009). However, when $p \geq n$, the four classical test statistics are not 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), Srivastava (2007), Chen and Qin (2010), Cai et al. (2014) and Feng et al. (2015). Recently, some tests have also been introduced for the case of genenral k. Schott (2007) modified Hotelling-Lawley trace and proposed the test statistic

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

Statistic T_{SC} is a representative of the so-called sum-of-squares type statistics as it is based on an estimation of squared Euclidean norm $\sum_{i=1}^{k} n_i\|\xi_i - \bar{\xi}\|^2$, where $\bar{\xi} = n^{-1}\sum_{i=1}^{k} n_i\xi_i$. See Srivastava and Kubokawa (2013), Yamada and Himeno (2015) and Zhou et al. (2017) for some other sum-of-squares type test statistics. 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, it is substituted by an estimator. Unlike T_{SC}, T_{CX} is an extreme value type statistic.

The likelihood ratio test (LRT) method has been very successful in leading to satisfactory procedures in many specific problems. However, the LRT statistic for hypotheses (1.1), i.e. Wilks' Lambda statistic, is not defined for $p > n-k$. In high dimensional setting, both sum-of-squares type statistics and extreme value type statistics are not based on likelihood function. This motivates us to construct a likelihood-based test 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. They used a least favorable argument to construct a generalized likelihood ratio test statistic. Their simulation results showed that their test has good power performance, especially when the variables are correlated.

In this paper, we propose a generalized likelihood ratio test statistic for hypotheses (1.1) called least favorable direction (LFD) test statistic. The asymptotic distributions of the test statistic are derived. These asymptotic distributions are valid when the eigenvalues of covariance matrix are bounded or the covariance matrix has r significantly large eigenvalues. The latter covariance structure, known as spiked covariance model, can characterize the strong correlations between variables. See, for example, Fan et al. (2008), Cai et al. (2013), Shen et al. (2013) and Ma et al. (2015). The asymptotic null distribution of the proposed test statistic involves some unknown parameters. We substitute the unknown parameters by their consistent estimators and formulate a test with asymptotically correct level. The asymptotic local power function of LFD test is also given. It will be seen that the asymptotic local power function of LFD test doesn't rely on the large eigenvalues of the covariance matrix. For most existing tests, however, the asymptotic power decreases as the large eigenvalues of the covariance matrix increase. Thus, LFD test is particularly powerful when variables are strongly correlated. Further simulations show the good performance of LFD test.

The rest of the paper is organized as follows. In Section 2, we propose LFD test and give the asymptotic distributions of LFD test. Section 3 complements our study with numerical simulations. In Section 4, we give a short discussion. Finally, the proofs are gathered in the Appendix.

2. Least favorable direction test

We introduce some notations. Define the $p \times n$ pooled sample matrix $\mathbf{X}$ as

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

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

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

is an $n \times k$ matrix and $\mathbf{1}_{n_i}$ is an n_i-dimensional vector with all elements equal to 1, $i = 1, \ldots, k$. Construct an $n \times (n-k)$ matrix $\tilde{\mathbf{J}}$ as

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

where $\tilde{\mathbf{J}}_i$ is an $n_i \times (n_i - 1)$ matrix defined as

$$\tilde{\mathbf{J}}_i = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \cdots & \frac{1}{\sqrt{(n_i-2)(n_i-1)}} & \frac{1}{\sqrt{(n_i-1)n_i}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \cdots & \frac{1}{\sqrt{(n_i-2)(n_i-1)}} & \frac{1}{\sqrt{(n_i-1)n_i}} \\ 0 & -\frac{2}{\sqrt{6}} & \cdots & \vdots & \vdots \\ \vdots & \vdots & \cdots & -\frac{n_i-2}{\sqrt{(n_i-2)(n_i-1)}} & \frac{1}{\sqrt{(n_i-1)n_i}} \\ 0 & 0 & \cdots & 0 & -\frac{n_i-1}{\sqrt{(n_i-1)n_i}} \end{pmatrix}.$$

The matrix $\tilde{\mathbf{J}}$ is a column orthogonal matrix satisfying $\tilde{\mathbf{J}}^T\tilde{\mathbf{J}} = \mathbf{I}_{n-k}$ and $\tilde{\mathbf{J}}\tilde{\mathbf{J}}^T = \mathbf{I}_n - \mathbf{J}\mathbf{J}^T$. Let $\mathbf{Y} = \mathbf{X}\tilde{\mathbf{J}}$ be a $p \times (n-k)$ random matrix. Then $\mathbf{G}$

can be written as

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

The sum-of-squares between groups $\mathbf{H}$ can be written as

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

By some matrix algebra, we have $\mathbf{I}_k - \frac{1}{n}\mathbf{J}^T\mathbf{1}_n\mathbf{1}_n^T\mathbf{J} = \mathbf{C}\mathbf{C}^T$ where $\mathbf{C}$ is a $k \times (k-1)$ matrix defined as $\mathbf{C} = \mathbf{C}_1\mathbf{C}_2$, and

$$\mathbf{C}_1 = \begin{pmatrix} \sqrt{n_1} & \sqrt{n_1} & \cdots & \sqrt{n_1} & \sqrt{n_1} \\ -\frac{n_1}{\sqrt{n_2}} & \sqrt{n_2} & \cdots & \sqrt{n_2} & \sqrt{n_2} \\ 0 & -\frac{n_1+n_2}{\sqrt{n_3}} & \cdots & \vdots & \vdots \\ \vdots & \vdots & \cdots & -\frac{\sum_{i=1}^{k-2} n_i}{\sqrt{n_{k-1}}} & \sqrt{n_{k-1}} \\ 0 & 0 & \cdots & 0 & -\frac{\sum_{i=1}^{k-1} n_i}{\sqrt{n_k}} \end{pmatrix},$$

$$\mathbf{C}_2 = \begin{pmatrix} \frac{n_1(n_1+n_2)}{n_2} & 0 & \cdots & 0 \\ 0 & \frac{(\sum_{i=1}^{2} n_i)(\sum_{i=1}^{3} n_i)}{n_3} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \frac{(\sum_{i=1}^{k-1} n_i)(\sum_{i=1}^{k} n_i)}{n_k} \end{pmatrix}^{-\frac{1}{2}}.$$

Then $\mathbf{H}$ can be written as

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

Define $\mathbf{\Theta} = (\sqrt{n_1}\xi_1, \ldots, \sqrt{n_k}\xi_k)$ and the null hypothesis H_0 is equivalent to $\mathbf{\Theta}\mathbf{C} = \mathbf{O}_{p\times(k-1)}$, where $\mathbf{O}_{p\times(k-1)}$ is a $p \times (k-1)$ matrix with all elements equal to 0. Thus, the hypotheses in (1.1) are equivalent to

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

In low dimensional setting, the testing problem (1.1) is well studied. A classical test statistic is Roy's maximum root which is constructed by ROY (1953) using his well-known union intersection principle. The key idea is to decompose data $\mathbf{X}$ into a set of univariate data $\{\mathbf{X}_a = a^T\mathbf{X} : a \in \mathbb{R}^p, a^T a = 1\}$. This induces a decomposition of the null hypothesis and the alternative hypothesis:

$$H_0 = \bigcap_{a\in\mathbb{R}^p, a^T a=1} H_{0a} \quad \text{v.s.} \quad H_1 = \bigcup_{a\in\mathbb{R}^p, a^T a=1} H_{1a},$$

where $H_{0a} : a^T\mathbf{\Theta C} = \mathbf{O}_{1\times(k-1)}$ and $H_{1a} : a^T\mathbf{\Theta C} \neq \mathbf{O}_{1\times(k-1)}$. Let $L_0(a)$ and $L_1(a)$ be the maximum likelihood of $\mathbf{X}_a$ under H_{0a} and H_{1a}, respectively. For each a satisfying $a^T a = 1$, the component LRT statistic

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

can be used to test H_{0a} v.s. H_{1a}. Using union intersection principle, Roy proposed the test statistic $\max_{a^T a=1} L_1(a)/L_0(a) = \lambda_1^{n/2}(\mathbf{H}\mathbf{G}^{-1})$, where $\lambda_i(\cdot)$ means the ith largest eigenvalue. This statistic is an increasing function of Roy's maximum root.

From a likelihood point of view, log likelihood ratio is an estimator of the Kullback-Leibler divergence between the true distribution and the null distribution. Hence the component LRT statistic $L_1(a)/L_0(a)$ characterizes the discrepancy between the true distribution and the null distribution along the direction a. This motivates us to consider the direction

$$a^* = \arg\max_{a^T a=1} \frac{L_1(a)}{L_0(a)} \tag{2.2}$$

which can hopefully achieve the largest discrepancy between the true distribution and the null distribution. Thus, H_{0a^*} is the component null hypothesis most likely to be not true. We shall call a^* the least favorable direction. Roy's maximum root is in fact the component LRT statistic along the least favorable direction.

Unfortunately, Roy's maximum root can only be defined when $n-k \geq p$, hence can not be used in the high dimensional setting. In what follows, we assume $p > n-k$. In this case, the set

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

is not empty since $\mathbf{G}$ is singular. Consequently, the right hand side of (2.2) is not well defined since the ratio involves infinity. Hence we need a new definition for LFD in the high dimensional setting. Define

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

It can be seen that $\mathcal{B} \subset \mathcal{A}$. Moreover, by the independence of $\mathbf{G}$ and $\mathbf{H}$, with probability 1, we have $\mathcal{A} \cap \mathcal{B}^c \neq \emptyset$. Then for any direction a, there are three possible scenarios: $L_1(a) < +\infty$ and $L_0(a) < +\infty$; $L_1(a) = +\infty$ and $L_0(a) < +\infty$; $L_1(a) = +\infty$ and $L_0(a) = +\infty$. To maximizes the discrepancy between $L_1(a)$ and $L_0(a)$, one may consider the direction a such that $L_1(a) = +\infty$ and $L_0(a) < +\infty$. This suggests that the least favorable direction a^*, which hopefully maximizes the discrepancy between $L_1(a)$ and $L_0(a)$, should be defined as $a^* = \arg\min_{a \in \mathcal{A} \cap \mathcal{B}^c} L_0(a)$. Equivalently,

$$a^* = \underset{a \in \mathcal{A} \cap \mathcal{B}^c}{\arg\min}\, L_0(a) = \underset{a^T a = 1, a^T G a = 0}{\arg\max}\, a^T \mathbf{H} a.$$

Based on a^* and likelihood $L_0(a)$, we propose a new test statistic

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

We reject the null hypothesis when $T(\mathbf{X})$ is large enough. We shall call $T(\mathbf{X})$ the LFD test statistic. Since the least favorable direction a^* is obtained from the component likelihood function, the statistic $T(\mathbf{X})$ is also a generalized likelihood ratio test statistic.

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

$$T(\mathbf{X}) = \lambda_1\big(\mathbf{C}^T\mathbf{J}^T\mathbf{X}^T(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{X}\mathbf{J}\mathbf{C}\big). \tag{2.3}$$

Next, we derive another simple form of $T(\mathbf{X})$. By the relationship

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

and matrix inverse formula, we have that

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

Thus,

$$T(\mathbf{X}) = \lambda_{\max}\Big(\mathbf{C}^T\big(\mathbf{J}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{J}\big)^{-1}\mathbf{C}\Big). \tag{2.4}$$

Compared with (2.3), (2.4) doesn't involve $\mathbf{P}_\mathbf{Y}$. Hence (2.4) is convenient for computation. In the case of $k = 2$, the least favorable direction is propotional to $(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)$ and LFD test statistic has expression

$$T(\mathbf{X}) = \frac{n_1 n_2}{n_1 + n_2}\|(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)\|^2.$$

In this case, the least favorable direction coincides with the maximal data piling direction proposed by Ahn and Marron (2010).

Now we derive the asymptotic distribution of LFD test statistic. We are especially interested in the case where variables are correlated. For some real world problems, variables are heavily correlated with common factors, then the covariance matrix $\boldsymbol{\Sigma}$ is spiked in the sense that a few eigenvalues of $\boldsymbol{\Sigma}$ 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 $\boldsymbol{\Sigma}$.

Assumption 1. *Let $\boldsymbol{\lambda}_1 \geq \boldsymbol{\lambda}_2 \geq \cdots \geq \boldsymbol{\lambda}_p$ be the eigenvalues of $\boldsymbol{\Sigma}$. Suppose there are $r \geq 0$ eigenvalues significantly larger than the others. We assume that*

- *$r = o(n)$.*
- *$c_1 \geq \boldsymbol{\lambda}_{r+1} \geq \ldots \geq \boldsymbol{\lambda}_p \geq c_2$ for some positive 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. If $r = 0$, Assumption 1 becomes $c_1 \geq \boldsymbol{\lambda}_1 \geq \ldots \geq \boldsymbol{\lambda}_p \geq c_2$. This is equivalent to Condition (C1) in Cai and Xia (2014). In fact, most existing high dimensional tests require conditions like this to avoid large eigenvalues. If $r > 0$, the covariance matrix is spiked. The spiked covariance model is commonly assumed in the study of PCA theory. Many existing works in PCA study assume that r is fixed. Here we allow r to vary as a smaller order of n.

Remark 2. If $r > 0$, the harshest 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 the 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). 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 iid $\mathcal{N}(0,1)$ and the entries on the diagonal are iid $\mathcal{N}(0,2)$. 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 LFD test statistic.

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

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

we have

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

where $\sim$ *means having the same distribution.*

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

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

The asymptotic variance of $T(\mathbf{X})$ is $2\operatorname{tr}(\mathbf{\Lambda}_2^2)$. If $r = 0$, it equals to $2\operatorname{tr}(\mathbf{\Sigma}^2)$ which is also the asymptotic variance of Bai and Saranadasa (1996) and Chen and Qin (2010)'s statistics. In comparison, if $r > 0$, $2\operatorname{tr}(\mathbf{\Lambda}_2^2)$ tends to be 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 matrix $\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}}$ appeared in expression (2.3) can remove large variance terms of $\mathbf{XJC}$.

To formulate a test procedure with asymptotically correct level, the unknown parameters r, $\operatorname{tr}(\mathbf{\Lambda}_2)$ and $\operatorname{tr}(\mathbf{\Lambda}_2^2)$ should be estimated. 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 a hyper parameter slowly tending 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 3. For the factor model adopted by Ma et al. (2015), λ_r is of order p. Hence we can take $\gamma_n = \sqrt{n}$.

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 a $p \times (n-k-2)$ matrix obtained by deleting the ith and jth columns from $\mathbf{Y}$. Let $\mathbf{Y}_{(i,j)} = \mathbf{U}_{\mathbf{Y};(i,j)}\mathbf{D}_{\mathbf{Y};(i,j)}\mathbf{V}_{\mathbf{Y};(i,j)}^T$ denote the singular value decomposition of $\mathbf{Y}_{(i,j)}$. Here $\mathbf{U}_{\mathbf{Y};(i,j)}$ is a $p \times (n-k-2)$ column orthogonal matrix. 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 ratio consistent.

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 LFD test procedure with asymptotic correct level α. Let

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

Let $F(x)$ be the cumulative distribution function of $\lambda_1(\mathbf{W}_{k-1})$. LFD test 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. Moreover, by Theorem 1, the asymptotic local power function of

LFD test procedure is

$$\Pr\Big(\lambda_1\big(\mathbf{W}_{k-1} + \tfrac{1}{\sqrt{\mathrm{tr}(\mathbf{\Lambda}_2^2)}}\mathbf{C}^T\mathbf{\Theta}^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Theta}\mathbf{C}\big) \geq F_{\mathbf{W}}^{-1}(1-\alpha)\Big).$$

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

$$\Pr\Big(\lambda_1\big(\mathbf{W}_1 + \tfrac{1}{\sqrt{\mathrm{tr}(\mathbf{\Sigma}^2)}}\mathbf{C}^T\mathbf{\Theta}^T\mathbf{\Theta}\mathbf{C}\big) \geq F_{\mathbf{W}}^{-1}(1-\alpha)\Big).$$

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

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

There's a random term $\mathbf{P}_{\mathbf{Y}}$ in the expression. To overcome this, suppose $\sqrt{n_i}\xi_i$ has prior distribution $\mathcal{N}_p(0, \psi\mathbf{I}_p)$, $i = 1, 2$. In this case, $\psi^{-1}\mathbf{C}^T\mathbf{\Theta}^T\mathbf{\Theta}\mathbf{C}$ is distributed as χ^2 distribution with p degrees of freedom. On the other hand, $\mathbf{C}^T\mathbf{\Theta}^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Theta}\mathbf{C}$ is distributed as χ^2 distribution with $p - n + k$ degrees of freedom. In this case, we have

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

Thus, when

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

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

3. Numerical study

In this section, we compare the numerical performance of LFD test with some existing tests, including the MANOVA tests of Schott (2007) and Cai and Xia (2014) and the two sample tests of Srivastava (2007), Chen and Qin (2010), Cai et al. (2014) and Feng et al. (2015). Note that the critical values of these existing tests are not valid under spiked covariance model. Hence we use permutation method to determine the critical values throughout our simulations. The test procedures resulting from permutation method have exact levels 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 LFD test statistic, the permutation method has a simple implementation. By expression (2.4), a permuted statistic can be written as

$$T(\mathbf{X}\Gamma) = \lambda_{\max}\Big(\mathbf{C}^T\big(\mathbf{J}^T\Gamma^T(\mathbf{X}^T\mathbf{X})^{-1}\Gamma\mathbf{J}\big)^{-1}\mathbf{C}\Big), \tag{3.5}$$

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 LFD test statistic can be summarized as:

1. Calculate $T(\mathbf{X})$ according to (2.4), keep 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 (3.5).

3. Calculate the p-value by $\tilde{p} = (M+1)^{-1}\big[1+\sum_{i=1}^{M} I\{T(\mathbf{X}\Gamma_i) \geq T(\mathbf{X})\}\big]$. 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 a negligible effect on total computational complexity.

Now we evaluate the empirical power performance of LFD test and competing tests. Define signal-to-noise ratio (SNR) as

$$\mathrm{SNR} = \frac{\|\boldsymbol{\Theta}\mathbf{C}\|_F^2}{\sqrt{\boldsymbol{\Lambda}_2^2}}.$$

We use SNR to characterize the signal strength.

In the first simulation study, we take $k = 3$. For comparison, we also carry out simulations for the tests of Schott (2007) and Cai and Xia (2014). We denote these two tests by SC and CX, respectively. We take $r = 2$ and $\boldsymbol{\Sigma} = \mathrm{diag}(1.5p, p, 1, 1, \ldots, 1)$. We consider two different structures 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 specific values. 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 specific values. The empirical power is computed based on 1000 simulations. The simulation results are summarized in Tables 1-4. It can be seen from the results that the proposed test outperforms the other two tests for both non-sparse and sparse alternatives. This verifies our theoretical results that LFD test performs well under spiked covariance.

In our second simulation study, we would like to investigate the effect of correlations between variables. We take $k = 2$ so that we can compare our test with some existing two sample tests. For comparison, we carry out simulations for the test of Srivastava (2007), Chen and Qin (2010), Cai et al. (2014) and Feng et al. (2015). We denote these tests by SR, CQ, CLX and FZWZ, respectively. Let the diagonal elements of Σ be 1 and the off-diagonal elements of Σ be ρ with $0 \leq \rho < 1$. The parameter ρ characterizes the correlations between variables. We set $\xi_1 = \kappa(\mathbf{1}_{p/2}^T, -\mathbf{1}_{p/2}^T)^T$ and $\xi_2 = \mathbf{0}_p$, where κ is selected such that SNR equals to 5. Figure 1 plots the empirical power versus ρ, where empirical power is computed based on 1000 simulations. We can see that the empirical power of LFD test holds nearly constant as ρ varies while the empirical powers of Chen and Qin (2010) and Feng et al. (2015)'s tests decrease as ρ increases. When ρ is small, LFD test has reasonable performance. When ρ is larger than 0.1, LFD test outperforms all other tests.

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

SNR	$p = 50$			$p = 75$			$p = 100$		
	CX	SC	LFD	CX	SC	LFD	CX	SC	LFD
0	0.047	0.044	0.052	0.051	0.050	0.051	0.059	0.047	0.047
1	0.074	0.056	0.074	0.089	0.050	0.089	0.062	0.062	0.093
2	0.120	0.045	0.133	0.090	0.040	0.119	0.071	0.049	0.127
3	0.107	0.046	0.197	0.118	0.057	0.242	0.102	0.057	0.220
4	0.160	0.062	0.271	0.131	0.057	0.328	0.146	0.053	0.339
5	0.207	0.064	0.386	0.149	0.052	0.458	0.146	0.067	0.484
6	0.199	0.061	0.485	0.192	0.047	0.583	0.160	0.057	0.588
7	0.234	0.071	0.577	0.221	0.074	0.685	0.185	0.057	0.707
8	0.266	0.072	0.648	0.263	0.078	0.775	0.201	0.062	0.829
9	0.319	0.081	0.718	0.245	0.068	0.838	0.230	0.064	0.896
10	0.304	0.075	0.784	0.297	0.089	0.904	0.288	0.062	0.913

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

SNR	$p = 100$			$p = 150$			$p = 200$		
	CX	SC	LFD	CX	SC	LFD	CX	SC	LFD
0	0.045	0.041	0.054	0.052	0.046	0.043	0.048	0.043	0.049
1	0.074	0.061	0.099	0.054	0.056	0.082	0.057	0.061	0.107
2	0.092	0.066	0.128	0.086	0.050	0.146	0.079	0.065	0.174
3	0.097	0.070	0.207	0.094	0.058	0.258	0.087	0.053	0.307
4	0.117	0.050	0.249	0.116	0.053	0.375	0.127	0.061	0.412
5	0.147	0.057	0.334	0.139	0.058	0.535	0.122	0.034	0.570
6	0.204	0.057	0.444	0.169	0.070	0.666	0.139	0.055	0.738
7	0.215	0.065	0.523	0.190	0.054	0.774	0.165	0.061	0.847
8	0.247	0.074	0.618	0.200	0.064	0.851	0.181	0.055	0.915
9	0.274	0.073	0.650	0.229	0.059	0.915	0.212	0.052	0.943
10	0.291	0.069	0.729	0.245	0.064	0.930	0.225	0.051	0.977

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

SNR	$p = 50$			$p = 75$			$p = 100$		
	CX	SC	LFD	CX	SC	LFD	CX	SC	LFD
0	0.038	0.043	0.037	0.046	0.058	0.059	0.049	0.044	0.047
1	0.064	0.054	0.076	0.067	0.061	0.088	0.066	0.053	0.084
2	0.101	0.052	0.097	0.085	0.048	0.114	0.111	0.058	0.114
3	0.144	0.060	0.169	0.132	0.050	0.188	0.112	0.049	0.166
4	0.181	0.060	0.220	0.161	0.052	0.239	0.157	0.063	0.249
5	0.236	0.063	0.295	0.194	0.061	0.313	0.216	0.057	0.311
6	0.285	0.070	0.333	0.253	0.065	0.419	0.243	0.060	0.398
7	0.344	0.081	0.425	0.299	0.061	0.506	0.291	0.066	0.543
8	0.401	0.082	0.513	0.363	0.077	0.620	0.299	0.065	0.611
9	0.455	0.079	0.600	0.407	0.067	0.667	0.392	0.060	0.709
10	0.522	0.076	0.641	0.467	0.086	0.784	0.417	0.071	0.766

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

SNR	$p = 100$			$p = 150$			$p = 200$		
	CX	SC	LFD	CX	SC	LFD	CX	SC	LFD
0	0.068	0.051	0.051	0.046	0.053	0.043	0.065	0.049	0.052
1	0.074	0.049	0.062	0.062	0.046	0.109	0.084	0.048	0.103
2	0.100	0.060	0.123	0.064	0.055	0.149	0.093	0.055	0.155
3	0.105	0.048	0.157	0.104	0.054	0.228	0.114	0.065	0.270
4	0.152	0.064	0.246	0.133	0.056	0.320	0.129	0.054	0.303
5	0.194	0.054	0.280	0.190	0.036	0.419	0.151	0.048	0.434
6	0.232	0.059	0.311	0.210	0.057	0.500	0.203	0.051	0.553
7	0.298	0.061	0.405	0.246	0.054	0.586	0.220	0.057	0.661
8	0.367	0.061	0.477	0.314	0.051	0.707	0.261	0.077	0.765
9	0.405	0.064	0.499	0.351	0.057	0.783	0.275	0.064	0.823
10	0.455	0.067	0.587	0.405	0.061	0.828	0.367	0.059	0.900

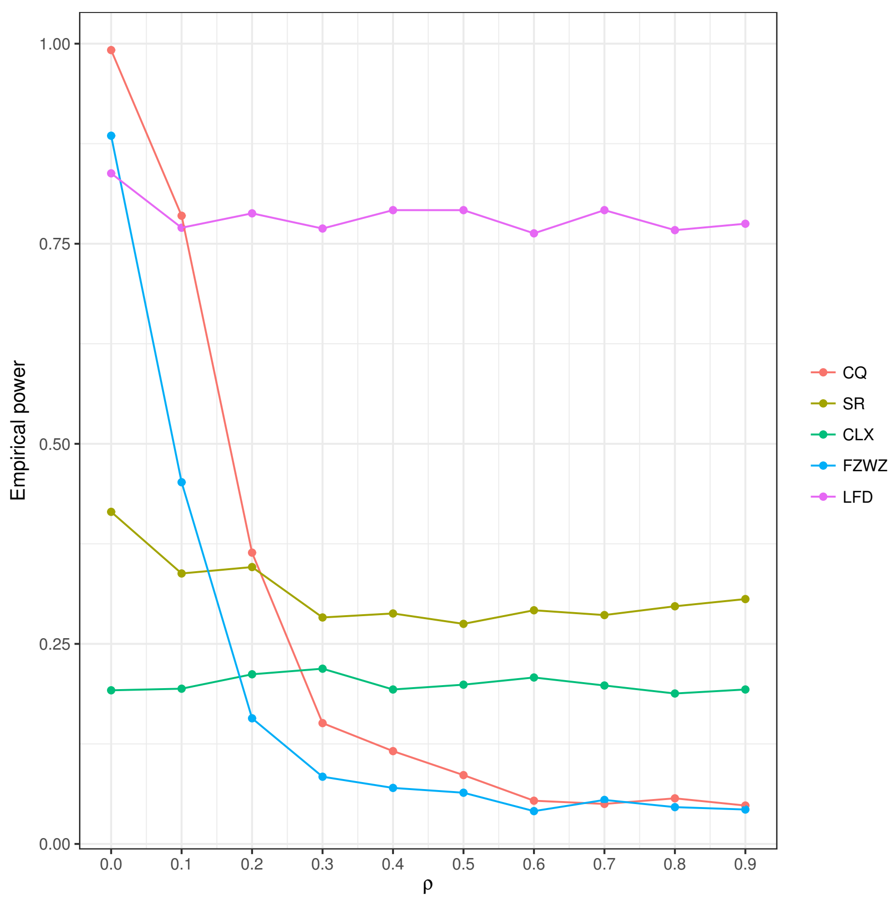

Figure 1: The empirical powers of tests. $\alpha = 0.05$, $k = 2$, $n_1 = n_2 = 20$, $p = 150$.

4. Concluding remarks

In this paper, using the idea of least favorable direction, we proposed LFD test for MANOVA in high dimensional setting. We derived the asymptotic distribution of LFD test statistic. We also gave the asymptotic local power function. Our theoretic work and simulation studies show that when the covariance matrix is spiked, LFD test tends to be more powerful than existing tests.

Our proof relies on the normality of the observations. It is interesting to investigate whether the theorems are still valid without normal assumption. Moreover, we assumed that p doesn't grow too fast. Without prior knowledge of Σ, this condition is unavoidable since when p is large, it's impossible to consistently estimate the principal subspace. See, for example, Cai et al. (2013). On the other hand, if we know some prior knowledge of Σ, for example, Σ is sparse, it's possible to construct a better test. We leave it for future research.

Appendix A Technical details

Proposition 4. *Suppose* $\mathbf{A}$ *is a* $p \times r$ *matrix with rank* r *and* $\mathbf{B}$ *is a* $p \times p$ *non-zero positive semi-definite matrix. Denote by* $\mathbf{A} = \mathbf{U}_{\mathbf{A}}\mathbf{D}_{\mathbf{A}}\mathbf{V}_{\mathbf{A}}^T$ *the singular value decomposition of* $\mathbf{A}$*, where* $\mathbf{U}_{\mathbf{A}}$ *and* $\mathbf{V}_{\mathbf{A}}$ *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 matrix on the column space of* $\mathbf{A}$*. Then*

$$\max_{a^T a=1, a^T \mathbf{A}\mathbf{A}^T a=0} a^T \mathbf{B} a = \lambda_{\max}\big(\mathbf{B}(\mathbf{I}_p - \mathbf{P_A})\big). \tag{A.6}$$

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

$$\max_{a^T a=1, a^T \mathbf{A}\mathbf{A}^T a=0} a^T \mathbf{B} a = \max_{a^T a=1, \mathbf{P_A} a=0} a^T (\mathbf{I}_p - \mathbf{P_A})\mathbf{B}(\mathbf{I}_p - \mathbf{P_A})a, \tag{A.7}$$

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 (A.7) and (A.7) 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 $\lambda_{\max}\big((\mathbf{I} - \mathbf{P_A})\mathbf{B}(\mathbf{I} - \mathbf{P_A})\big) = \lambda_{\max}\big(\mathbf{B}(\mathbf{I} - \mathbf{P_A})\big)$.

$\square$

Lemma 1 (Weyl's inequality)**.** *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two symmetric* $n \times n$ *matrices. If* $r + s - 1 \le i \le j + k - n$*, we have*

$$\lambda_j(\mathbf{A}) + \lambda_k(\mathbf{B}) \le \lambda_i(\mathbf{A} + \mathbf{B}) \le \lambda_r(\mathbf{A}) + \lambda_s(\mathbf{B}).$$

See, for example, Horn and Johnson (2012) Theorem 4.3.1.

Lemma 2. *Let $\{Z_i\}_{i=1}^n$ be iid p-dimensional random vectors with common distribution $\mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p)$. Then for any n-dimensional vector $\omega = (\omega_1, \ldots, \omega_n)^T$, we have*

$$\left\| \sum_{i=1}^n \omega_i (Z_i Z_i^T - \mathbf{I}_p) \right\| = O_P(|\omega|_2 \sqrt{p} + |\omega|_\infty p).$$

Proof. This proof is adapted from the proof of Theorem 5.39 in Vershynin (2010). By Lemma 5.2 and Lemma 5.4 of Vershynin (2010), there exists a set $\mathcal{C} \subset \{x \in \mathbb{R}^p : |x|_2 = 1\}$ satisfying $\mathrm{Card}(\mathcal{C}) \le 9^p$ such that for any $p \times p$ symmetric matrix $\mathbf{A}$,

$$\|A\| \le 2 \max_{x \in \mathcal{C}} x^T \mathbf{A} x. \tag{A.8}$$

For $t > 1$, we have

$$\begin{aligned}
&\Pr\left(\left\| \sum_{i=1}^n \omega_i (Z_i Z_i^T - \mathbf{I}_p) \right\| > t(|\omega|_2 \sqrt{p} + |\omega|_\infty p) \right) \\
\le &\Pr\left(2 \sup_{x \in \mathcal{C}} \left| \sum_{i=1}^n \omega_i (x^T Z_i Z_i^T x - 1) \right| > t(|\omega|_2 \sqrt{p} + |\omega|_\infty p) \right) \\
\le &\sum_{x \in \mathcal{C}} \Pr\left(\left| \sum_{i=1}^n \omega_i (x^T Z_i Z_i^T x - 1) \right| > |\omega|_2 \sqrt{\frac{pt}{4}} + 2|\omega|_\infty \frac{pt}{4} \right) \\
\le &2 \cdot 9^p \exp\left(-\frac{pt}{4} \right) = 2\exp\left((2\log 3 - t/4)p \right) \le 2\exp\left(2\log 3 - t/4 \right),
\end{aligned}$$

where the first inequality follows from (A.8), the second inequality follows from union bound and the third inequality follows Lemma 1 of Laurent and

Massart (2000). The upper bound $2\exp(2\log 3 - t/4)$ can be arbitrarily small as long as t is large enough. This completes the proof. $\square$

Let $\mathbf{U}_{\mathbf{Y},1}$ and $\mathbf{U}_{\mathbf{Y},2}$ denote the first r columns and the last $n-r$ columns of $\mathbf{U}_{\mathbf{Y}}$, respectively. We can write $\mathbf{Y} = \mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{Z}$, where $\mathbf{Z}$ be a $p \times n$ random matrix with iid $\mathcal{N}(0,1)$ entries. Let $\mathbf{Z} = (\mathbf{Z}_1^T, \mathbf{Z}_2^T)^T$, where $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are the first r rows and last $p-r$ rows of $\mathbf{Z}$. Let $\mathbf{V}_{\mathbf{Z}_1} = \mathbf{Z}_1^T(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1/2}$. Then $\mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^T = \mathbf{Z}_1^T(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1}\mathbf{Z}_1$ is the projection matrix onto the row space of $\mathbf{Z}_1$. Let $\tilde{\mathbf{V}}_{\mathbf{Z}_1}$ be a $n \times (n-r)$ column orthogonal matrix which satisfies $\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T = \mathbf{I}_n - \mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^T$. Let $\mathbf{Q} = \boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1/2}\boldsymbol{\Lambda}_1^{-1/2} = \boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{Z}_1^T(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1}\boldsymbol{\Lambda}_1^{-1/2}$. Let $\mathbf{U}_{\mathbf{Y},2}$ denote the $(r+1)$ to nth columns of $\mathbf{U}_{\mathbf{Y}}$. Let $\hat{\boldsymbol{\Sigma}} = n^{-1}\mathbf{Y}\mathbf{Y}^T$.

Proposition 5. *Suppose that $r \le n$. Then uniformly for $i = 1, \ldots, r$,*

$$\lambda_i(\hat{\boldsymbol{\Sigma}}) = \boldsymbol{\lambda}_i + n^{-1}\sum_{i=r+1}^{p}\boldsymbol{\lambda}_i + O_P\left(\boldsymbol{\lambda}_i\sqrt{\frac{r}{n}} + \sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{n}} + \boldsymbol{\lambda}_{r+1}\right). \tag{A.9}$$

And

$$\sum_{i=r+1}^{p}\lambda_i(\hat{\boldsymbol{\Sigma}}) = \operatorname{tr}(\boldsymbol{\Lambda}_2) - \frac{r}{n}\sum_{i=r+1}^{p}\boldsymbol{\lambda}_i + O_P\left(r\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{n}} + r\boldsymbol{\lambda}_{r+1}\right). \tag{A.10}$$

Proof. We only need to deal with the matrix $n^{-1}\mathbf{Z}^T\boldsymbol{\Lambda}\mathbf{Z}$ since it shares the same non-zero eigenvalues as $\hat{\boldsymbol{\Sigma}}$. Write

$$
\begin{aligned}
n^{-1}\mathbf{Z}^T\boldsymbol{\Lambda}\mathbf{Z} =& n^{-1}\mathbf{Z}_1^T\boldsymbol{\Lambda}_1\mathbf{Z}_1 + n^{-1}\mathbf{Z}_2^T\boldsymbol{\Lambda}_2\mathbf{Z}_2 \\
=& n^{-1}\mathbf{Z}_1^T\boldsymbol{\Lambda}_1\mathbf{Z}_1 + n^{-1}(\sum_{i=r+1}^{p}\boldsymbol{\lambda}_i)\mathbf{I}_n + n^{-1}\left(\mathbf{Z}_2^T\boldsymbol{\Lambda}_2\mathbf{Z}_2 - (\sum_{i=r+1}^{p}\boldsymbol{\lambda}_i)\mathbf{I}_n\right).
\end{aligned}
$$

Then Weyl's inequality implies that for $i = 1, \ldots, r$,

$$
\left|\lambda_i\left(n^{-1}\mathbf{Z}^T\boldsymbol{\Lambda}\mathbf{Z}\right) - \lambda_i(n^{-1}\mathbf{Z}_1^T\boldsymbol{\Lambda}_1\mathbf{Z}_1) - n^{-1}\sum_{i=r+1}^{p}\boldsymbol{\lambda}_i\right| \le n^{-1}\left\|\mathbf{Z}_2^T\boldsymbol{\Lambda}_2\mathbf{Z}_2 - (\sum_{i=r+1}^{p}\boldsymbol{\lambda}_i)\mathbf{I}_n\right\|. \tag{A.11}
$$

Using Weyl's inequality, we can derive the following lower bound for $\lambda_i(n^{-1}\mathbf{Z}_1^T\boldsymbol{\Lambda}_1\mathbf{Z}_1)$, $i = 1, \ldots, r$.

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

Similarly, we can derive the following upper bound for $\lambda_i(\mathbf{Z}_1^T\boldsymbol{\Lambda}_1\mathbf{Z}_1)$, $i = 1, \ldots, r$.

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

The above lower bound and upper bound imply

$$\begin{aligned}\left|\lambda_i(n^{-1}\mathbf{Z}_1^T\mathbf{\Lambda}_1\mathbf{Z}_1) - \boldsymbol{\lambda}_i\right| \leq& \boldsymbol{\lambda}_i \max\left(|\lambda_1(n^{-1}\mathbf{Z}_1\mathbf{Z}_1^T) - 1|, |\lambda_r(n^{-1}\mathbf{Z}_1\mathbf{Z}_1^T) - 1|\right) \\ =& \boldsymbol{\lambda}_i\|n^{-1}\mathbf{Z}_1\mathbf{Z}_1^T - \mathbf{I}_r\|.\end{aligned} \tag{A.12}$$

Combining the bounds (A.11) and (A.12) gives that for $i = 1, \ldots, r$,

$$\begin{aligned}&\left|\lambda_i\left(n^{-1}\mathbf{Z}^T\mathbf{\Lambda}\mathbf{Z}\right) - \boldsymbol{\lambda}_i - n^{-1}\sum_{i=r+1}^{p}\boldsymbol{\lambda}_i\right| \\ \leq& n^{-1}\left\|\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2 - (\sum_{i=r+1}^{p}\boldsymbol{\lambda}_i)\mathbf{I}_n\right\| + \boldsymbol{\lambda}_i\|n^{-1}\mathbf{Z}_1\mathbf{Z}_1^T - \mathbf{I}_r\|.\end{aligned}$$

By Lemma 2, we have

$$\|n^{-1}\mathbf{Z}_1\mathbf{Z}_1^T - \mathbf{I}_r\| = O_P\left(\sqrt{\frac{r}{n}}\right), \tag{A.13}$$

$$n^{-1}\left\|\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2 - (\sum_{i=r+1}^{p}\boldsymbol{\lambda}_i)\mathbf{I}_n\right\| = O_P\left(\sqrt{\frac{\mathrm{tr}(\mathbf{\Lambda}_2^2)}{n}} + \boldsymbol{\lambda}_{r+1}\right). \tag{A.14}$$

Then (A.9) follows.

To prove (A.10), we note that

$$\begin{aligned}\sum_{i=r+1}^{p}\lambda_i(\hat{\mathbf{\Sigma}}) =& \sum_{i=r+1}^{n}\lambda_i(n^{-1}\ \mathbf{Z}^T\mathbf{\Lambda}\mathbf{Z}) \\ =& \mathrm{tr}(n^{-1}\mathbf{Z}^T\mathbf{\Lambda}\mathbf{Z}) - \sum_{i=1}^{r}\lambda_i(n^{-1}\ \mathbf{Z}^T\mathbf{\Lambda}\mathbf{Z}) \\ =& \mathrm{tr}(n^{-1}\mathbf{Z}_1^T\mathbf{\Lambda}_1\mathbf{Z}_1) + \mathrm{tr}(n^{-1}\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2) - \sum_{i=1}^{r}\lambda_i(n^{-1}\ \mathbf{Z}^T\mathbf{\Lambda}\mathbf{Z}).\end{aligned}$$

It follows from inequalities (A.11) and (A.14) that

$$\begin{aligned}&\left|\sum_{i=1}^{r}\lambda_i(n^{-1}\ \mathbf{Z}^T\mathbf{\Lambda}\mathbf{Z})-\operatorname{tr}(n^{-1}\mathbf{Z}_1^T\mathbf{\Lambda}_1\mathbf{Z}_1)-\frac{r}{n}\sum_{i=r+1}^{p}\boldsymbol{\lambda}_i\right|\\ \leq&\frac{r}{n}\left\|\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2-(\sum_{i=r+1}^{p}\boldsymbol{\lambda}_i)\mathbf{I}_n\right\|=O_P\left(r\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{n}}+r\boldsymbol{\lambda}_{r+1}\right).\end{aligned}$$

Thus,

$$\sum_{i=r+1}^{p}\lambda_i(\hat{\mathbf{\Sigma}})=\operatorname{tr}(n^{-1}\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2)-\frac{r}{n}\sum_{i=r+1}^{p}\boldsymbol{\lambda}_i+O_P\left(r\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{n}}+r\boldsymbol{\lambda}_{r+1}\right).$$

It is straightforward to show that

$$\mathrm{E}\operatorname{tr}(n^{-1}\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2)=\operatorname{tr}(\mathbf{\Lambda}_2),\quad \operatorname{Var}\left(\operatorname{tr}(n^{-1}\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2)\right)=\frac{2}{n}\operatorname{tr}(\mathbf{\Lambda}_2^2).$$

Hence

$$\begin{aligned}&\sum_{i=r+1}^{p}\lambda_i(\hat{\mathbf{\Sigma}})\\ =&\operatorname{tr}(\mathbf{\Lambda}_2)+O_P\left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{n}}\right)-\frac{r}{n}\sum_{i=r+1}^{p}\boldsymbol{\lambda}_i+O_P\left(r\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{n}}+r\boldsymbol{\lambda}_{r+1}\right)\\ =&\operatorname{tr}(\mathbf{\Lambda}_2)-\frac{r}{n}\sum_{i=r+1}^{p}\boldsymbol{\lambda}_i+O_P\left(r\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{n}}+r\boldsymbol{\lambda}_{r+1}\right).\end{aligned}$$

This completes the proof of (A.10).

□

Proposition 6. *Suppose that $r=o(n)$ and $r\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2)\to 0$. Then*

$$\|\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T-\mathbf{P}_{\mathbf{Y},1}^*\|=O_P\left(\frac{\boldsymbol{\lambda}_{r+1}+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)}{\boldsymbol{\lambda}_r+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)}\right),$$

where

$$\mathbf{P}^*_{\mathbf{Y},1} = \mathbf{U} \begin{pmatrix} \mathbf{I}_r \\ \mathbf{Q} \end{pmatrix} \left(\mathbf{I}_r + n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{\Lambda}_1^{-1}\right)^{-1} \begin{pmatrix} \mathbf{I}_r & \mathbf{Q}^T \end{pmatrix} \mathbf{U}^T.$$

Proof. The following intermediate matrix

$$\begin{aligned}\hat{\mathbf{\Sigma}}_0 =& n^{-1}\mathbf{U}_1\mathbf{\Lambda}_1^{1/2}\mathbf{Z}_1\mathbf{Z}_1^T\mathbf{\Lambda}_1^{1/2}\mathbf{U}_1^T + n^{-1}\mathbf{U}_1\mathbf{\Lambda}_1^{1/2}\mathbf{Z}_1\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T + n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{Z}_1^T\mathbf{\Lambda}_1^{1/2}\mathbf{U}_1^T \\ &+ n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T\end{aligned}$$

plays a key role in the proof. It can be seen that

$$\hat{\mathbf{\Sigma}}_0 = n^{-1}\mathbf{U} \begin{pmatrix} \mathbf{I}_r \\ \mathbf{Q} \end{pmatrix} \mathbf{\Lambda}_1^{1/2}\mathbf{Z}_1\mathbf{Z}_1^T\mathbf{\Lambda}_1^{1/2} \begin{pmatrix} \mathbf{I}_r & \mathbf{Q}^T \end{pmatrix} \mathbf{U}^T.$$

Consequently, $\hat{\mathbf{\Sigma}}_0$ is a positive semi-definite matrix with rank r, and the matrix

$$\mathbf{P}_0 = \mathbf{U} \begin{pmatrix} \mathbf{I}_r \\ \mathbf{Q} \end{pmatrix} \left(\mathbf{I}_r + \mathbf{Q}^T\mathbf{Q}\right)^{-1} \begin{pmatrix} \mathbf{I}_r & \mathbf{Q}^T \end{pmatrix} \mathbf{U}^T$$

is the projection matrix onto the rank r principal subspace of $\hat{\mathbf{\Sigma}}_0$.

From Cai et al. (2015), Proposition 1, we have

$$\|\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T - \mathbf{P}_0\| \leq \frac{2\|\hat{\mathbf{\Sigma}} - \hat{\mathbf{\Sigma}}_0\|}{\lambda_r(\hat{\mathbf{\Sigma}}_0)}. \tag{A.15}$$

We have the following upper bound for $\|\hat{\mathbf{\Sigma}} - \hat{\mathbf{\Sigma}}_0\|$.

$$
\begin{aligned}
\|\hat{\mathbf{\Sigma}} - \hat{\mathbf{\Sigma}}_0\| =& n^{-1} \left\| \mathbf{U}_2 \mathbf{\Lambda}_2^{1/2} \mathbf{Z}_2 \mathbf{Z}_2^T \mathbf{\Lambda}_2^{1/2} \mathbf{U}_2^T - \mathbf{U}_2 \mathbf{\Lambda}_2^{1/2} \mathbf{Z}_2 \mathbf{V}_{\mathbf{Z}_1} \mathbf{V}_{\mathbf{Z}_1}^T \mathbf{Z}_2^T \mathbf{\Lambda}_2^{1/2} \mathbf{U}_2^T \right\| \\
=& n^{-1} \left\| \mathbf{\Lambda}_2^{1/2} \mathbf{Z}_2 (\mathbf{I}_n - \mathbf{V}_{\mathbf{Z}_1} \mathbf{V}_{\mathbf{Z}_1}^T) \mathbf{Z}_2^T \mathbf{\Lambda}_2^{1/2} \right\| \\
\leq& n^{-1} \left\| \mathbf{Z}_2^T \mathbf{\Lambda}_2 \mathbf{Z}_2 \right\| \\
\leq& n^{-1} \left\| \mathbf{Z}_2^T \mathbf{\Lambda}_2 \mathbf{Z}_2 - \operatorname{tr}(\mathbf{\Lambda}_2) \mathbf{I}_n \right\| + n^{-1} \operatorname{tr}(\mathbf{\Lambda}_2) \\
=& O_P \left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{n}} + \boldsymbol{\lambda}_{r+1} + n^{-1} \operatorname{tr}(\mathbf{\Lambda}_2) \right) \\
=& O_P \left(\boldsymbol{\lambda}_{r+1} + n^{-1} \operatorname{tr}(\mathbf{\Lambda}_2) \right),
\end{aligned}
\tag{A.16}
$$

where the second last equality follows from (A.14) and the last equality follows from

$$
\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{n}} \leq \sqrt{\frac{\boldsymbol{\lambda}_{r+1} \operatorname{tr}(\mathbf{\Lambda}_2)}{n}} \leq \frac{1}{2} \left(\boldsymbol{\lambda}_{r+1} + n^{-1} \operatorname{tr}(\mathbf{\Lambda}_2) \right).
$$

Now we deal with $\lambda_r(\hat{\mathbf{\Sigma}}_0)$. We have

$$
\begin{aligned}
\lambda_r(\hat{\mathbf{\Sigma}}_0) =& \lambda_r \left(n^{-1} \mathbf{Z}_1^T \mathbf{\Lambda}_1^{1/2} (\mathbf{I}_r + \mathbf{Q}^T \mathbf{Q}) \mathbf{\Lambda}_1^{1/2} \mathbf{Z}_1 \right) \\
=& \lambda_r \left(n^{-1} \mathbf{\Lambda}_1^{1/2} (\mathbf{I}_r + \mathbf{Q}^T \mathbf{Q}) \mathbf{\Lambda}_1^{1/2} \mathbf{Z}_1 \mathbf{Z}_1^T \right) \\
=& \lambda_r \left(n^{-1} (\mathbf{Z}_1 \mathbf{Z}_1^T)^{1/2} \mathbf{\Lambda}_1^{1/2} (\mathbf{I}_r + \mathbf{Q}^T \mathbf{Q}) \mathbf{\Lambda}_1^{1/2} (\mathbf{Z}_1 \mathbf{Z}_1^T)^{1/2} \right) \\
=& \lambda_r \left(n^{-1} (\mathbf{Z}_1 \mathbf{Z}_1^T)^{1/2} \mathbf{\Lambda}_1 (\mathbf{Z}_1 \mathbf{Z}_1^T)^{1/2} + n^{-1} \mathbf{V}_{\mathbf{Z}_1}^T \mathbf{Z}_2^T \mathbf{\Lambda}_2 \mathbf{Z}_2 \mathbf{V}_{\mathbf{Z}_1} \right).
\end{aligned}
$$

It can be seen that $\mathbf{Z}_2 \mathbf{V}_{\mathbf{Z}_1}$ is a $(p-r) \times r$ random matrix with iid $\mathcal{N}(0,1)$

entries. Then Lemma 2 implies that

$$\begin{aligned}\left\|n^{-1}\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}-n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_r\right\| =&O_P\left(n^{-1}\sqrt{r\operatorname{tr}\left(\mathbf{\Lambda}_2^2\right)}+rn^{-1}\boldsymbol{\lambda}_{r+1}\right)\\ =&O_P\left(n^{-1}\sqrt{r\boldsymbol{\lambda}_{r+1}\operatorname{tr}\left(\mathbf{\Lambda}_2\right)}+rn^{-1}\boldsymbol{\lambda}_{r+1}\right)\\ =&o_P\left(n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\right),\end{aligned} \tag{A.17}$$

where the last equality follows from the condition $r\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2)\to 0$. Then it follows from Weyl's inequality that

$$\begin{aligned}&\left|\lambda_r(\hat{\mathbf{\Sigma}}_0)-\lambda_r\left(n^{-1}(\mathbf{Z}_1\mathbf{Z}_1^T)^{1/2}\mathbf{\Lambda}_1(\mathbf{Z}_1\mathbf{Z}_1^T)^{1/2}+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_r\right)\right|\\ \le&\left\|n^{-1}\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}-n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_r\right\|\\ =&o_P\left(n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\right).\end{aligned}$$

On the other hand, (A.12) and (A.13) imply that

$$\begin{aligned}&\lambda_r\left(n^{-1}(\mathbf{Z}_1\mathbf{Z}_1^T)^{1/2}\mathbf{\Lambda}_1(\mathbf{Z}_1\mathbf{Z}_1^T)^{1/2}+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_r\right)\\ =&\lambda_r\left(n^{-1}\mathbf{Z}_1^T\mathbf{\Lambda}_1\mathbf{Z}_1\right)+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\\ =&\boldsymbol{\lambda}_r+o_P(\boldsymbol{\lambda}_r)+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2).\end{aligned}$$

Hence we have

$$\lambda_r(\hat{\mathbf{\Sigma}}_0)=(1+o_P(1))(\boldsymbol{\lambda}_r+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)). \tag{A.18}$$

Combining (A.15), (A.16) and (A.18) yields

$$\|\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T-\mathbf{P}_0\|=O_P\left(\frac{\boldsymbol{\lambda}_{r+1}+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)}{\boldsymbol{\lambda}_r+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)}\right).$$

Now we deal with $\mathbf{P}_0$. By some algebra, we have

$$
\begin{aligned}
&\left\|\mathbf{P}_0-\mathbf{P}_{\mathbf{Y},1}^*\right\| \\
=&\left\|\left(\mathbf{I}_r+\mathbf{Q}^T\mathbf{Q}\right)^{1/2}\left(\left(\mathbf{I}_r+\mathbf{Q}^T\mathbf{Q}\right)^{-1}-\left(\mathbf{I}_r+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{\Lambda}_1^{-1}\right)^{-1}\right)\left(\mathbf{I}_r+\mathbf{Q}^T\mathbf{Q}\right)^{1/2}\right\| \\
=&\left\|\mathbf{I}_r-\left(\mathbf{I}_r+\mathbf{Q}^T\mathbf{Q}\right)^{1/2}\left(\mathbf{I}_r+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{\Lambda}_1^{-1}\right)^{-1}\left(\mathbf{I}_r+\mathbf{Q}^T\mathbf{Q}\right)^{1/2}\right\| \\
=&\left\|\mathbf{I}_r-\left(\mathbf{I}_r+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{\Lambda}_1^{-1}\right)^{-1/2}\left(\mathbf{I}_r+\mathbf{Q}^T\mathbf{Q}\right)\left(\mathbf{I}_r+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{\Lambda}_1^{-1}\right)^{-1/2}\right\| \\
=&\left\|\left(\mathbf{I}_r+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{\Lambda}_1^{-1}\right)^{-1/2}\left(\mathbf{Q}^T\mathbf{Q}-n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{\Lambda}_1^{-1}\right)\left(\mathbf{I}_r+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{\Lambda}_1^{-1}\right)^{-1/2}\right\| \\
=&\left\|\left(\mathbf{\Lambda}_1+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\right)^{-1/2}\left(\mathbf{\Lambda}_1^{1/2}\mathbf{Q}^T\mathbf{Q}\mathbf{\Lambda}_1^{1/2}-n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_r\right)\left(\mathbf{\Lambda}_1+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\right)^{-1/2}\right\|.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\left\|\mathbf{P}_0-\mathbf{P}_{\mathbf{Y},1}^*\right\| \\
\leq&\left(\boldsymbol{\lambda}_r+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\right)^{-1}\left\|(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1}\mathbf{Z}_1\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{Z}_1^T(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1}-n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_r\right\| \\
\leq&\left(\boldsymbol{\lambda}_r+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\right)^{-1}\left\|(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1}\mathbf{Z}_1\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{Z}_1^T(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1}-\operatorname{tr}(\mathbf{\Lambda}_2)\left(\mathbf{Z}_1\mathbf{Z}_1^T\right)^{-1}\right\| \\
&+\left(\boldsymbol{\lambda}_r+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\right)^{-1}\left\|\operatorname{tr}(\mathbf{\Lambda}_2)\left(\mathbf{Z}_1\mathbf{Z}_1^T\right)^{-1}-n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_r\right\| \\
\leq&\left(\boldsymbol{\lambda}_r+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\right)^{-1}\left\|(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1}\right\|\left\|\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}-\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_r\right\| \\
&+\left(\boldsymbol{\lambda}_r+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\right)^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\left\|(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1}\right\|\left\|n^{-1}\mathbf{Z}_1\mathbf{Z}_1^T-\mathbf{I}_r\right\| \\
=&o_P\left(\frac{n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)}{\boldsymbol{\lambda}_r+n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)}\right),
\end{aligned}
$$

where the last equality follows from the fact $\|(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1}\| = \lambda_r(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1}$, (A.12), (A.13) and (A.17). This completes the proof. □

Corollary 1. *Suppse the conditions of Proposition 6 hold. If we further* *assume that* $\operatorname{tr}(\boldsymbol{\Lambda}_2)/(n\boldsymbol{\lambda}_r) \to 0$*, then*

$$\|\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T - \mathbf{P}_{\mathbf{Y},1}^{\dagger}\| = O_P\left(\frac{\boldsymbol{\lambda}_{r+1}}{\boldsymbol{\lambda}_r} + \frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)}{n\boldsymbol{\lambda}_r}\right),$$

where $\mathbf{P}_{\mathbf{Y},1}^{\dagger} = \mathbf{U}_1\mathbf{U}_1^T + \mathbf{U}_1\mathbf{Q}^T\mathbf{U}_2^T + \mathbf{U}_2\mathbf{Q}\mathbf{U}_1^T$.

Proof. Note that

$$\begin{aligned}
&\|\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T - \mathbf{P}_{\mathbf{Y},1}^{\dagger}\| \\
\leq &\|\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T - \mathbf{P}_{\mathbf{Y},1}^{*}\| + \left\|\mathbf{P}_{\mathbf{Y},1}^{*} - \mathbf{P}_{\mathbf{Y},1}^{\dagger}\right\|.
\end{aligned}$$

Under the condition $\operatorname{tr}(\boldsymbol{\Lambda}_2)/(n\boldsymbol{\lambda}_r) \to 0$, Proposition 6 implies that

$$\|\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T - \mathbf{P}_{\mathbf{Y},1}^{*}\| = O_P\left(\frac{\boldsymbol{\lambda}_{r+1}}{\boldsymbol{\lambda}_r} + \frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)}{n\boldsymbol{\lambda}_r}\right).$$

So we only need to deal with the second term. We have

$$\begin{aligned}
&\left\|\mathbf{P}_{\mathbf{Y},1}^{*} - \mathbf{P}_{\mathbf{Y},1}^{\dagger}\right\| \\
\leq&\left\|\mathbf{P}_{\mathbf{Y},1}^{*} - \mathbf{U}\begin{pmatrix}\mathbf{I}_r \\ \mathbf{Q}\end{pmatrix}\begin{pmatrix}\mathbf{I}_r & \mathbf{Q}^T\end{pmatrix}\mathbf{U}^T\right\| + \left\|\mathbf{U}\begin{pmatrix}\mathbf{I}_r \\ \mathbf{Q}\end{pmatrix}\begin{pmatrix}\mathbf{I}_r & \mathbf{Q}^T\end{pmatrix}\mathbf{U}^T - \mathbf{P}_{\mathbf{Y},1}^{\dagger}\right\| \\
=&\left\|\begin{pmatrix}\mathbf{I}_r \\ \mathbf{Q}\end{pmatrix}\left(\left(\mathbf{I}_r + n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\boldsymbol{\Lambda}_1^{-1}\right)^{-1} - \mathbf{I}_r\right)\begin{pmatrix}\mathbf{I}_r & \mathbf{Q}^T\end{pmatrix}\right\| + \|\mathbf{U}_2\mathbf{Q}\mathbf{Q}^T\mathbf{U}_2^T\| \\
\leq&(1 + \|\mathbf{Q}^T\mathbf{Q}\|)\left\|\left(\mathbf{I}_r + n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\boldsymbol{\Lambda}_1^{-1}\right)^{-1} - \mathbf{I}_r\right\| + \|\mathbf{Q}^T\mathbf{Q}\| \\
\leq&(1 + \|\mathbf{Q}^T\mathbf{Q}\|)\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)/(n\boldsymbol{\lambda}_r)}{1 + \operatorname{tr}(\boldsymbol{\Lambda}_2)/(n\boldsymbol{\lambda}_r)} + \left\|\mathbf{Q}^T\mathbf{Q}\right\|.
\end{aligned}$$

Note that

$$\|\mathbf{Q}^T\mathbf{Q}\| \leq \boldsymbol{\lambda}_r^{-1}\|(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1}\|\|\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\boldsymbol{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\| = O_P\left(\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)}{n\boldsymbol{\lambda}_r}\right), \quad \text{(A.19)}$$

---

where the second last equality follows from the fact $\|(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1}\| = \lambda_r(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1}$, (A.12), (A.13) and (A.17). Therefore, we have

$$\left\|\mathbf{P}^*_{\mathbf{Y},1} - \mathbf{P}^\dagger_{\mathbf{Y},1}\right\| = O_P\left(\frac{\operatorname{tr}(\mathbf{\Lambda}_2)}{n\boldsymbol{\lambda}_r}\right).$$

This completes the proof.

□

Proposition 7. *Suppose that $r = o(n)$ and $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2) \to 0$. Then*

$$\left\|\mathbf{U}_{\mathbf{Y},2}\mathbf{U}_{\mathbf{Y},2}^T - \mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\left(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\right)^{-1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T\right\| = O_P\left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}}\right).$$

Proof. We only need to prove that for any subsequence of $\{n\}$, there is a further subsequence along which the conclusion holds. Thus, without loss of generality, we assume $\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1/(n\boldsymbol{\lambda}_r^2) \to c \in [0, +\infty]$. Since $\mathbf{U}_{\mathbf{Y},2}\mathbf{U}_{\mathbf{Y},2}^T$ and $\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\left(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\right)^{-1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T$ are both projection matrix, we have

$$\left\|\mathbf{U}_{\mathbf{Y},2}\mathbf{U}_{\mathbf{Y},2}^T - \mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\left(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\right)^{-1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T\right\| \leq 2.$$

Therefore, the conclusion holds if $c > 0$. In the rest of the proof, we assume $c = 0$, that is $\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1/(n\boldsymbol{\lambda}_r^2) \to 0$.

Note that $\mathbf{U}_{\mathbf{Y},2}$ is in fact the first $n-r$ eigenvectors of $(\mathbf{I}_p - \mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T)\hat{\mathbf{\Sigma}}(\mathbf{I}_p - \mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T)$. Under the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2) \to 0$, Corollary 1 implies that

$$\|\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T - \mathbf{P}^\dagger_{\mathbf{Y},1}\| = O_P\left(\frac{\operatorname{tr}(\mathbf{\Lambda}_2)}{n\boldsymbol{\lambda}_r}\right).$$

It can be seen that

$$
\begin{aligned}
&(\mathbf{I}_p - \mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T)\hat{\mathbf{\Sigma}}(\mathbf{I}_p - \mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T) - (\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^\dagger)\hat{\mathbf{\Sigma}}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^\dagger)\\
=&(\mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T)\hat{\mathbf{\Sigma}}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^\dagger) + (\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^\dagger)\hat{\mathbf{\Sigma}}(\mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T)\\
&+ (\mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T)\hat{\mathbf{\Sigma}}(\mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T).
\end{aligned}
$$

Under the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2) \to 0$, Proposition 5 implies that

$$
\|\hat{\mathbf{\Sigma}}\| = \boldsymbol{\lambda}_1\left(1 + \frac{\operatorname{tr}(\mathbf{\Lambda}_2)}{n\boldsymbol{\lambda}_1} + O_P\left(\sqrt{\frac{r}{n}} + \sqrt{\frac{\boldsymbol{\lambda}_{r+1}}{\boldsymbol{\lambda}_r}\frac{\operatorname{tr}(\mathbf{\Lambda}_2)}{n\boldsymbol{\lambda}_1}} + \frac{\boldsymbol{\lambda}_{r+1}}{\boldsymbol{\lambda}_r}\right)\right) = \boldsymbol{\lambda}_1(1+o_P(1)).
$$

Then

$$
\begin{aligned}
&\left\|(\mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T)\hat{\mathbf{\Sigma}}(\mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T)\right\|\\
\le&\|\hat{\mathbf{\Sigma}}\|\left\|\mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T\right\|^2 \qquad \text{(A.20)}\\
=&O_P\left(\frac{\operatorname{tr}^2(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n^2\boldsymbol{\lambda}_r^2}\right).
\end{aligned}
$$

On the other hand, we have

$$
\begin{aligned}
&\left|(\mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T)\hat{\mathbf{\Sigma}}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^\dagger)\right|\\
\le&\left\|\mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T\right\|\|n^{-1}\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}\|\left\|\mathbf{Z}^T\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^\dagger)\right\|\\
=&n^{-1/2}\left\|\mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T\right\|\left\|\hat{\mathbf{\Sigma}}\right\|^{1/2}\left\|\mathbf{Z}^T\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^\dagger)\right\|\\
=&O_P\left(\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1^{1/2}}{n^{3/2}\boldsymbol{\lambda}_r}\right)\left\|\mathbf{Z}^T\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^\dagger)\right\|.
\end{aligned}
$$

It is straightforward to show that

$$
\mathbf{Z}^T\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^\dagger) = \tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T - \mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{Z}_1^T(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1}\mathbf{\Lambda}_1^{-1/2}\mathbf{U}_1^T. \qquad \text{(A.21)}
$$

Then

$$\left\|\mathbf{Z}^T\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\right\| \leq \left\|\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\right\|^{1/2}+\boldsymbol{\lambda}_r^{-1/2}\left\|\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\right\|\left\|(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1}\right\|^{1/2}.$$

It follows from (A.14) and the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2)\to 0$ that

$$\|\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\| = (1+o_P(1))\operatorname{tr}(\mathbf{\Lambda}_2). \tag{A.22}$$

Consequently,

$$\left\|\mathbf{Z}^T\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\right\| = O_P\left(\operatorname{tr}^{1/2}(\mathbf{\Lambda}_2)\right)+O_P\left(\frac{\operatorname{tr}(\mathbf{\Lambda}_2)}{\sqrt{n\boldsymbol{\lambda}_r}}\right) = O_P\left(\operatorname{tr}^{1/2}(\mathbf{\Lambda}_2)\right).$$

Thus,

$$\left|(\mathbf{P}_{\mathbf{Y},1}^{\dagger}-\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T)\hat{\mathbf{\Sigma}}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\right| = O_P\left(\frac{\operatorname{tr}^{3/2}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1^{1/2}}{n^{3/2}\boldsymbol{\lambda}_r}\right). \tag{A.23}$$

Combine (A.20) and (A.23), we obtain

$$\begin{aligned}&\left\|(\mathbf{I}_p-\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T)\hat{\mathbf{\Sigma}}(\mathbf{I}_p-\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T)-(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\hat{\mathbf{\Sigma}}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\right\|\\ =&O_P\left(\frac{\operatorname{tr}^2(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n^2\boldsymbol{\lambda}_r^2}+\frac{\operatorname{tr}^{3/2}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1^{1/2}}{n^{3/2}\boldsymbol{\lambda}_r}\right).\end{aligned}$$

Now we deal with $(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\hat{\mathbf{\Sigma}}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})$. In view of (A.21), we have

$$\begin{aligned}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\hat{\mathbf{\Sigma}}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger}) =&n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T\\ &-n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{Q}\mathbf{U}_1^T\\ &-n^{-1}\mathbf{U}_1\mathbf{Q}^T\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T\\ &+n^{-1}\mathbf{U}_1\mathbf{Q}^T\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{Q}\mathbf{U}_1^T.\end{aligned}$$

Then

$$
\begin{aligned}
&\left\|(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})\hat{\boldsymbol{\Sigma}}(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})-n^{-1}\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2(\mathbf{I}_n-\mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^T)\mathbf{Z}_2^T\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^T\right\|\\
\leq& n^{-1}\left\|\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2(\mathbf{I}_n-\mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^T)\mathbf{Z}_2^T\boldsymbol{\Lambda}_2^{1/2}\mathbf{Q}\right\|+n^{-1}\left\|\mathbf{Q}^T\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{Z}_2^T\boldsymbol{\Lambda}_2^{1/2}\mathbf{Q}\right\|\\
\leq& n^{-1}\|\mathbf{Z}_2^T\boldsymbol{\Lambda}_2\mathbf{Z}_2\|\|\mathbf{Q}^T\mathbf{Q}\|^{1/2}+n^{-1}\|\mathbf{Z}_2^T\boldsymbol{\Lambda}_2\mathbf{Z}_2\|\|\mathbf{Q}^T\mathbf{Q}\|\\
=&O_P\left(\frac{\operatorname{tr}^{3/2}(\boldsymbol{\Lambda}_2)}{n^{3/2}\boldsymbol{\lambda}_r^{1/2}}\right),
\end{aligned}
$$

where the last equality follows from (A.19) and (A.22).

Combine the above bounds, we obtain

$$
\begin{aligned}
&\left\|(\mathbf{I}_p-\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T)\hat{\boldsymbol{\Sigma}}(\mathbf{I}_p-\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T)-n^{-1}\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^T\right\|\\
=&O_P\left(\frac{\operatorname{tr}^2(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1}{n^2\boldsymbol{\lambda}_r^2}+\frac{\operatorname{tr}^{3/2}(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1^{1/2}}{n^{3/2}\boldsymbol{\lambda}_r}\right).
\end{aligned}
\tag{A.24}
$$

The matrix $n^{-1}\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^T$ shares the same non-zero eigenvalues as $n^{-1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\boldsymbol{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}$. Note that $\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}$ is a $p\times(n-r)$ random matrix with iid $\mathcal{N}(0,1)$ entries. Then it follows from Lemma 2 and the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2)\to 0$ that

$$
\begin{aligned}
\left\|n^{-1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\boldsymbol{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}-n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\mathbf{I}_{n-r}\right\|=&O_P\left(n^{-1/2}\sqrt{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}+\boldsymbol{\lambda}_{r+1}\right)\\
\leq&O_P\left(n^{-1/2}\sqrt{\boldsymbol{\lambda}_{r+1}\operatorname{tr}(\boldsymbol{\Lambda}_2)}+\boldsymbol{\lambda}_{r+1}\right)\\
=&o_P\left(n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\right).
\end{aligned}
\tag{A.25}
$$

This bound, combined with Weyl's inequality, leads to

$$\lambda_{n-r}\left(n^{-1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\right)=(1+o_P(1))n^{-1}\operatorname{tr}(\mathbf{\Lambda}_1). \tag{A.26}$$

As a result, the matrix $n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T$ is of rank $n-r$. It can be seen that the matrix $\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\left(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\right)^{-1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T$ is the projection matrix onto the rank $n-r$ principal subspace of $n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T$. Therefore, Cai et al. (2015), proposition 1 implies that

$$\begin{aligned}
&\left\|\mathbf{U}_{\mathbf{Y},2}\mathbf{U}_{\mathbf{Y},2}^T-\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\left(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\right)^{-1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T\right\|\\
\leq&\frac{2\left\|(\mathbf{I}_p-\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T)\hat{\mathbf{\Sigma}}(\mathbf{I}_p-\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T)-n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T\right\|}{\lambda_{n-r}\left(n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T\right)}\\
=&O_P\left(\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}+\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}}\right)\\
=&O_P\left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}}\right).
\end{aligned}$$

where the second last equality follows from (A.24) and (A.26). This completes the proof.

□

Corollary 2. *Under the conditions of Proposition 7, we have*

$$\left\|\mathbf{P}_{\mathbf{Y},2}-\mathbf{P}_{\mathbf{Y},2}^\dagger\right\|=o_P(1),$$

where $\mathbf{P}_{\mathbf{Y},2}^\dagger=(\operatorname{tr}(\mathbf{\Lambda}_2))^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T$.

Proof. By some algebra, it can be seen that

$$\begin{aligned}
&\left\|\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\left(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\right)^{-1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T - \mathbf{P}_{\mathbf{Y},2}^{\dagger}\right\| \\
=&(\operatorname{tr}(\mathbf{\Lambda}_2))^{-1}\left\|\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1} - \operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_{n-r}\right\| \\
=&o_P(1),
\end{aligned}$$

where the last equality follows from (A.25). Then the conclusion follows from the last display and Proposition 7.

□

Appendix B Proofs 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 $\mathbf{Y} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}$, where $\mathbf{Z}$ is a $p \times (n-k)$ matrix with iid $\mathcal{N}(0,1)$ entries. We write $\mathbf{XJC} = \mathbf{\Theta}\mathbf{C} + \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^{\dagger}$, where $\mathbf{Z}^{\dagger}$ is a $p \times (k-1)$ matrix with iid $\mathcal{N}(0,1)$ entries.

Then

$$\begin{aligned}
\mathbf{C}^T\mathbf{J}^T\mathbf{X}^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{XJC} =&\mathbf{Z}^{\dagger T}\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^{\dagger} + \mathbf{C}^T\mathbf{\Theta}^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Theta}\mathbf{C} + \\
&\mathbf{C}^T\mathbf{\Theta}^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^{\dagger} + \mathbf{Z}^{\dagger T}\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Theta}\mathbf{C}.
\end{aligned} \tag{B.27}$$

The first term of (B.27) can be written as

$$\mathbf{Z}^{\dagger T}\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^{\dagger} = \sum_{i=1}^{p}\lambda_i((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}}))\eta_i\eta_i^T, \tag{B.28}$$

where $\eta_i \overset{iid}{\sim} \mathcal{N}(0, \mathbf{I}_{k-1})$, $i = 1, \ldots, p$.

Lemma 3. *Suppose the conditions of Proposition 7 hold. Then uniformly for $i = 1, \ldots, r$,*

$$\lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) = (1 + o_P(1))n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2).$$

Proof. Note that

$$(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}}) = (\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},2})(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},2}). \tag{B.29}$$

We first deal with $(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})$. Under the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2) \to 0$, Corollary 1 implies that

$$\|\mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^T - \mathbf{P}_{\mathbf{Y},1}^{\dagger}\| = O_P\left(\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)}{n\boldsymbol{\lambda}_r}\right).$$

From the decomposition

$$\begin{aligned}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}) =& (\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^{\dagger})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^{\dagger}) + (\mathbf{P}_{\mathbf{Y},1}^{\dagger} - \mathbf{P}_{\mathbf{Y},1})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^{\dagger}) \\ &+ (\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^{\dagger})\boldsymbol{\Sigma}(\mathbf{P}_{\mathbf{Y},1}^{\dagger} - \mathbf{P}_{\mathbf{Y},1}) + (\mathbf{P}_{\mathbf{Y},1}^{\dagger} - \mathbf{P}_{\mathbf{Y},1})\boldsymbol{\Sigma}(\mathbf{P}_{\mathbf{Y},1}^{\dagger} - \mathbf{P}_{\mathbf{Y},1}),\end{aligned}$$

we have

$$\begin{aligned}&\left\|(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}) - (\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^{\dagger})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^{\dagger})\right\| \\ \le& 2\left\|\mathbf{P}_{\mathbf{Y},1}^{\dagger} - \mathbf{P}_{\mathbf{Y},1}\right\|\left\|\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^{\dagger})\right\| + \boldsymbol{\lambda}_1\|\mathbf{P}_{\mathbf{Y},1}^{\dagger} - \mathbf{P}_{\mathbf{Y},1}\|^2. \\ =& O_P\left(\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)}{n\boldsymbol{\lambda}_r}\right)\left\|\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}^{\dagger})\right\| + O_P\left(\frac{\operatorname{tr}^2(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1}{n^2\boldsymbol{\lambda}_r^2}\right).\end{aligned}$$

Note that

$$\begin{aligned}\left\|\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})\right\| =& \left\|\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^T-\mathbf{U}_1\mathbf{\Lambda}_1\mathbf{Q}^T\mathbf{U}_2^T-\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{Q}\mathbf{U}_1^T\right\| \\ \leq& \boldsymbol{\lambda}_{r+1}+\left\|\mathbf{\Lambda}_1\mathbf{Q}^T\right\|+\boldsymbol{\lambda}_{r+1}\left\|\mathbf{Q}\right\| \\ =& \boldsymbol{\lambda}_{r+1}+\left\|\mathbf{\Lambda}_1^{1/2}(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1}\mathbf{Z}_1\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\right\|+\boldsymbol{\lambda}_{r+1}\|\mathbf{Q}^T\mathbf{Q}\|^{1/2} \\ \leq& \boldsymbol{\lambda}_{r+1}+\boldsymbol{\lambda}_1^{1/2}\left\|(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1/2}\right\|\left\|\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right\|^{1/2}+\boldsymbol{\lambda}_{r+1}\|\mathbf{Q}^T\mathbf{Q}\|^{1/2} \\ =& O_P\left(\sqrt{\frac{\boldsymbol{\lambda}_1\operatorname{tr}(\mathbf{\Lambda}_2)}{n}}\right),\end{aligned}$$

where the last equality follows from (A.17), (A.19) and the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2)\to 0$. Thus,

$$\left\|(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1})-(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})\right\|=O_P\left(\frac{\operatorname{tr}^{3/2}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1^{1/2}}{n^{3/2}\boldsymbol{\lambda}_r}\right). \tag{B.30}$$

It is straightforward to show that

$$(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})=\mathbf{U}_2\mathbf{Q}\mathbf{\Lambda}_1\mathbf{Q}^T\mathbf{U}_2^T+\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^T-\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{Q}\mathbf{U}_1^T-\mathbf{U}_1\mathbf{Q}^T\mathbf{\Lambda}_2\mathbf{U}_2^T+\mathbf{U}_1\mathbf{Q}^T\mathbf{\Lambda}_2\mathbf{Q}\mathbf{U}_1^T.$$

Therefore,

$$\begin{aligned}\left\|(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})-\mathbf{U}_2\mathbf{Q}\mathbf{\Lambda}_1\mathbf{Q}^T\mathbf{U}_2^T\right\| \leq& \boldsymbol{\lambda}_{r+1}(1+2\|\mathbf{Q}^T\mathbf{Q}\|^{1/2}+\|\mathbf{Q}^T\mathbf{Q}\|) \\ =& o_P\left(n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\right),\end{aligned} \tag{B.31}$$

where the last equality follows from (A.19) and the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2)\to 0$. Note that $\mathbf{U}_2\mathbf{Q}\mathbf{\Lambda}_1\mathbf{Q}^T\mathbf{U}_2^T=\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1}\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T$. We have

$$
\begin{aligned}
&\left\|\mathbf{U}_2\mathbf{Q}\boldsymbol{\Lambda}_1\mathbf{Q}^T\mathbf{U}_2^T - n^{-1}\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^T\right\| \\
\le&\left\|\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^T\right\| \left\|(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1} - n^{-1}\mathbf{I}_r\right\| \\
\le&\left\|\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\boldsymbol{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right\| \left\|(\mathbf{Z}_1\mathbf{Z}_1^T)^{-1}\right\| \left\|n^{-1}\mathbf{Z}_1\mathbf{Z}_1^T - \mathbf{I}_r\right\| \\
=&o_P\left(n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\right),
\end{aligned}
\tag{B.32}
$$

where the last equality follows from (A.13) and (A.17). From (B.29), (B.30), (B.31) and (B.32), we obtain that

$$
\begin{aligned}
&\left\|(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}) - n^{-1}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},2})\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},2})\right\| \\
=&o_P\left(n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\right).
\end{aligned}
$$

Thus, the last display, together with Weyl's inequality, implies that uniformly for $i = 1, \ldots, r$,

$$
\begin{aligned}
&\lambda_i\left((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\right) \\
=&n^{-1}\lambda_i\left(\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^T(\mathbf{I} - \mathbf{P}_{\mathbf{Y},2})\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right) + o_P\left(n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\right).
\end{aligned}
\tag{B.33}
$$

From Corollary 2 we have

$$
\begin{aligned}
&\Big\|\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^T(\mathbf{I} - \mathbf{P}_{\mathbf{Y},2})\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1} \\
&- \left(\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\boldsymbol{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1} - (\operatorname{tr}(\boldsymbol{\Lambda}_2))^{-1}\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\boldsymbol{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\boldsymbol{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right)\Big\| \\
\le&\left\|\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\boldsymbol{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right\| \left\|\mathbf{P}_{\mathbf{Y},2} - (\operatorname{tr}(\boldsymbol{\Lambda}_2))^{-1}\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\boldsymbol{\Lambda}_2^{1/2}\right\| \\
=&o_P(\operatorname{tr}(\boldsymbol{\Lambda}_2)),
\end{aligned}
\tag{B.34}
$$

where the last equality follows from (A.17), Now we deal with the term $\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}$. Note that $\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}$ and $\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}$ both have iid $\mathcal{N}(0,1)$ entries and they are mutually independent. So we can write

$$
\begin{aligned}
&\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1} \\
=&(\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1})^{1/2}\mathbf{Z}_2^{\dagger}\mathbf{Z}_2^{\dagger T}(\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1})^{1/2}.
\end{aligned}
$$

where $\mathbf{Z}_2^{\dagger} = (\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1})^{-1/2}\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}$ is a $r\times(n-r)$ random matrix with iid $\mathcal{N}(0,1)$ entries and is independent of $\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}$. Hence

$$
\begin{aligned}
\left\|\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right\| \le& \left\|\mathbf{Z}_2^{\dagger}\mathbf{Z}_2^{\dagger T}\right\| \left\|\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right\| \\
\le&\boldsymbol{\lambda}_{r+1}\left\|\mathbf{Z}_2^{\dagger}\mathbf{Z}_2^{\dagger T}\right\| \left\|\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right\| \\
=&o_P\left(\operatorname{tr}^2(\mathbf{\Lambda}_2)\right),
\end{aligned}
$$

where the last equality follows from (A.17), the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2)\to 0$ and the fact $\|\mathbf{Z}_2^{\dagger}\mathbf{Z}_2^{\dagger T}\| = O_P(n)$. It then follows from (A.17), (B.34) and the last display that

$$
\left\|\mathbf{V}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^T(\mathbf{I}-\mathbf{P}_{\mathbf{Y},2})\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1} - \operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_r\right\| = o_P\left(\operatorname{tr}(\mathbf{\Lambda}_2)\right).
$$

This, together with (B.33) and Weyl's inequality, implies that uniformly for $i=1,\ldots,r$,

$$
\lambda_i\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right) = (1+o_P(1))n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2).
$$

□

Lemma 4. *Suppose the conditions of Proposition 7 hold. Then*

$$\begin{aligned}&\sum_{i=r+1}^{p}\lambda_i\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)\\ =&\operatorname{tr}(\boldsymbol{\Lambda}_2)-\frac{n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}+o_P\left(n(\boldsymbol{\lambda}_{r+1}-\boldsymbol{\lambda}_p)\right)+O_P\left(r\boldsymbol{\lambda}_{r+1}\right).\end{aligned}$$

Proof. Write $\boldsymbol{\Sigma}=\mathbf{U}_1\boldsymbol{\Lambda}_1\mathbf{U}_1^T+\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^T$. Note that $\mathbf{U}_1\boldsymbol{\Lambda}_1\mathbf{U}_1^T$ is of rank r. Then Weyl's inequality implies that for $i=r+1,\ldots,p-n$,

$$\begin{aligned}&\lambda_i\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)\geq\lambda_i\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right),\\ &\lambda_i\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)\leq\lambda_{i-r}\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right).\end{aligned}$$

Hence we have

$$\begin{aligned}&\left\|\sum_{i=r+1}^{p}\lambda_i\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)-\operatorname{tr}\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)\right\|\\ \leq&r\lambda_1\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)\\ \leq&r\boldsymbol{\lambda}_{r+1}.\end{aligned}\tag{B.35}$$

Write

$$\begin{aligned}&\operatorname{tr}\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)\\ =&\operatorname{tr}\left(\boldsymbol{\Lambda}_2\mathbf{U}_2^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\right)\\ =&\operatorname{tr}(\boldsymbol{\Lambda}_2)-\operatorname{tr}\left(\left(\boldsymbol{\Lambda}_2-\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^T\mathbf{P}_{\mathbf{Y}}\mathbf{U}_2\right)-\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\operatorname{tr}\left(\mathbf{U}_2^T\mathbf{P}_{\mathbf{Y}}\mathbf{U}_2\right).\end{aligned}\tag{B.36}$$

For the third term, note that $\operatorname{tr}\left(\mathbf{U}_2^T\mathbf{P}_{\mathbf{Y}}\mathbf{U}_2\right) = \operatorname{tr}(\mathbf{P}_{\mathbf{Y}}) - \operatorname{tr}\left(\mathbf{P}_{\mathbf{Y}}\mathbf{U}_1\mathbf{U}_1^T\right)$. Since $\mathbf{P}_{\mathbf{Y}}$ is of rank n and $\mathbf{U}_1$ is of rank r, we have

$$|\operatorname{tr}\left(\mathbf{U}_2^T\mathbf{P}_{\mathbf{Y}}\mathbf{U}_2\right) - n| \le r. \tag{B.37}$$

Next we deal with the second term. We have

$$\begin{aligned}
&\left|\operatorname{tr}\left(\left(\boldsymbol{\Lambda}_2 - \frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^T\mathbf{P}_{\mathbf{Y}}\mathbf{U}_2\right) - \operatorname{tr}\left(\left(\boldsymbol{\Lambda}_2 - \frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^T\left(\mathbf{P}_{\mathbf{Y},1}^\dagger + \mathbf{P}_{\mathbf{Y},2}^\dagger\right)\mathbf{U}_2\right)\right| \\
=&\left|\operatorname{tr}\left(\left(\boldsymbol{\Lambda}_2 - \frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^T\left(\mathbf{P}_{\mathbf{Y}} - \mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{P}_{\mathbf{Y},2}^\dagger\right)\mathbf{U}_2\right)\right|.
\end{aligned}$$

Since $\operatorname{tr}(\boldsymbol{\Lambda}_2^2)/\operatorname{tr}(\boldsymbol{\Lambda}_2) \in [\boldsymbol{\lambda}_p, \boldsymbol{\lambda}_{r+1}]$, we have $\|\boldsymbol{\Lambda}_2 - (\operatorname{tr}(\boldsymbol{\Lambda}_2^2)/\operatorname{tr}(\boldsymbol{\Lambda}_2))\mathbf{I}_{p-r}\| \le \boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p$. Also note that the rank of the matrix $\mathbf{P}_{\mathbf{Y}} - \mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{P}_{\mathbf{Y},2}^\dagger$ is at most $2n$. Therefore, von Neumann's trace theorem implies that

$$\begin{aligned}
&\left|\operatorname{tr}\left(\left(\boldsymbol{\Lambda}_2 - \frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^T\left(\mathbf{P}_{\mathbf{Y}} - \mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{P}_{\mathbf{Y},2}^\dagger\right)\mathbf{U}_2\right)\right| \\
\le&2n(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\left\|\mathbf{P}_{\mathbf{Y}} - \mathbf{P}_{\mathbf{Y},1}^\dagger - \mathbf{P}_{\mathbf{Y},2}^\dagger\right\| \\
\le&2n(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\left(\left\|\mathbf{P}_{\mathbf{Y},1} - \mathbf{P}_{\mathbf{Y},1}^\dagger\right\| + \left\|\mathbf{P}_{\mathbf{Y},2} - \mathbf{P}_{\mathbf{Y},2}^\dagger\right\|\right) \\
=&o_P\left(n(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\right),
\end{aligned} \tag{B.38}$$

where the last equality follows from Corollary 1 and Corollary 2. Note that

$$\begin{aligned}
&\operatorname{tr}\left(\left(\boldsymbol{\Lambda}_2 - \frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^T\left(\mathbf{P}_{\mathbf{Y},1}^\dagger + \mathbf{P}_{\mathbf{Y},2}^\dagger\right)\mathbf{U}_2\right) \\
=&\operatorname{tr}\left(\left(\boldsymbol{\Lambda}_2 - \frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^T\mathbf{P}_{\mathbf{Y},2}^\dagger\mathbf{U}_2\right) \\
=&\frac{1}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\operatorname{tr}\left(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T\mathbf{Z}_2^T\left(\boldsymbol{\Lambda}_2^2 - \frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\boldsymbol{\Lambda}_2\right)\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\right)
\end{aligned}$$

It is straightforward to show that

$$\mathrm{E}\,\mathrm{tr}\left(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T \mathbf{Z}_2^T \left(\boldsymbol{\Lambda}_2^2 - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_2^2)}{\mathrm{tr}(\boldsymbol{\Lambda}_2)}\boldsymbol{\Lambda}_2\right) \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}\right) = 0,$$

and

$$\begin{aligned}
&\mathrm{Var}\left(\mathrm{tr}\left(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^T \mathbf{Z}_2^T \left(\boldsymbol{\Lambda}_2^2 - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_2^2)}{\mathrm{tr}(\boldsymbol{\Lambda}_2)}\boldsymbol{\Lambda}_2\right) \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}\right)\right) \\
=&2(n-r)\,\mathrm{tr}\left(\boldsymbol{\Lambda}_2^2 - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_2^2)}{\mathrm{tr}(\boldsymbol{\Lambda}_2)}\boldsymbol{\Lambda}_2\right)^2 \\
\leq&2n\,\mathrm{tr}(\Lambda_2^2)(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)^2 \\
\leq&2n\boldsymbol{\lambda}_{r+1}\,\mathrm{tr}(\boldsymbol{\Lambda}_2)(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)^2 \\
=&o_P\left(\mathrm{tr}^2(\boldsymbol{\Lambda}_2)(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)^2\right).
\end{aligned}$$

Thus,

$$\mathrm{tr}\left(\left(\boldsymbol{\Lambda}_2 - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_2^2)}{\mathrm{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^T\left(\mathbf{P}_{\mathbf{Y},1}^{\dagger} + \mathbf{P}_{\mathbf{Y},2}^{\dagger}\right)\mathbf{U}_2\right) = o_P\left(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p\right).$$

The last display, combined with (B.38), leads to

$$\mathrm{tr}\left(\left(\boldsymbol{\Lambda}_2 - \frac{\mathrm{tr}(\boldsymbol{\Lambda}_2^2)}{\mathrm{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^T\mathbf{P}_{\mathbf{Y}}\mathbf{U}_2\right) = o_P\left(n(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\right).$$

It follows from (B.36), (B.37) and the last display that

$$\begin{aligned}
&\mathrm{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^T(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) \\
=&\,\mathrm{tr}(\boldsymbol{\Lambda}_2) - \frac{n\,\mathrm{tr}(\boldsymbol{\Lambda}_2^2)}{\mathrm{tr}(\boldsymbol{\Lambda}_2)} + o_P\left(n(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\right) + O_P\left(\frac{r\,\mathrm{tr}(\boldsymbol{\Lambda}_2^2)}{\mathrm{tr}(\boldsymbol{\Lambda}_2)}\right) \\
=&\,\mathrm{tr}(\boldsymbol{\Lambda}_2) - \frac{n\,\mathrm{tr}(\boldsymbol{\Lambda}_2^2)}{\mathrm{tr}(\boldsymbol{\Lambda}_2)} + o_P\left(n(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\right) + O_P\left(r\boldsymbol{\lambda}_{r+1}\right).
\end{aligned}$$

Then the conclusion follows from (B.35) and the last display. $\square$

Lemma 5.

$$\operatorname{tr}\left(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\right)^2=(1+o_P(1))\operatorname{tr}(\mathbf{\Lambda}_2^2). \tag{B.39}$$

Proof. $\square$

Proof of Theorem 1. Lemma (3) and Lemma (5) imply that the first term of (B.27) 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

$$\Big(\operatorname{tr}\Big(\big(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\big)^2\Big)\Big)^{-1/2}$$

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

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

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

Next we show that the cross term of (B.27) is negligible. Note that

$$\begin{aligned}
&\mathrm{E}[\|\mathbf{C}^T\mathbf{\Theta}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}_2\|_F^2|\mathbf{Y}]\\
=&(k-1)\operatorname{tr}(\mathbf{C}^T\mathbf{\Theta}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Theta}\mathbf{C})\\
\le&(k-1)\lambda_1\big((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\big)\|\mathbf{\Theta}\mathbf{C}\|_F^2\\
\le&(k-1)\lambda_1\big(\mathbf{\Lambda}^{1/2}\mathbf{U}^T(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\big)\|\mathbf{\Theta}\mathbf{C}\|_F^2\\
=&(k-1)O_P\big(\frac{\boldsymbol{\lambda}_1 p}{\boldsymbol{\lambda}_r n}\big)\|\mathbf{\Theta}\mathbf{C}\|_F^2\\
=&(k-1)O_P\big(\frac{\boldsymbol{\lambda}_1\sqrt{p}}{\boldsymbol{\lambda}_r n}\big)\sqrt{p}\|\mathbf{\Theta}\mathbf{C}\|_F^2=o_P(p),
\end{aligned}$$

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

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

Equivalently,

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

The conclusion follows by taking the maximum eigenvalue. □

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

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

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

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

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

$$\{\hat{r} = 0\} \supseteq \{\frac{\lambda_1(\mathbf{Y}^T\mathbf{Y})}{\lambda_{n-k}(\mathbf{Y}^T\mathbf{Y})} \leq \gamma_n\}.$$

$\square$

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

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

Note that

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

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

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

It follows that

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

Hence

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

Where the last equality holds by Lemma **??**. But central limit theorem implies that

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

Thus,

$$\frac{1}{n-k}\sum_{i=r+1}^{n-k}\lambda_i(\mathbf{Y}^T\mathbf{Y})$$
$$=\operatorname{tr}(\mathbf{\Lambda}_2)+\frac{1}{n-k}\big(\sum_{i=r+1}^{n-k}\lambda_i(\mathbf{Y}^T\mathbf{Y})-\operatorname{tr}(\mathbf{Z}_{1B}^T\mathbf{\Lambda}_2\mathbf{Z}_{1B})\big)+\frac{1}{n-k}\big(\operatorname{tr}(\mathbf{Z}_{1B}^T\mathbf{\Lambda}_2\mathbf{Z}_{1B})-(n-k)\operatorname{tr}(\mathbf{\Lambda}_2)\big)$$
$$=\operatorname{tr}(\mathbf{\Lambda}_2)+O_P(\frac{rp}{n})+O_P(\sqrt{\frac{p}{n}})=\operatorname{tr}(\mathbf{\Lambda}_2)+O_P(\sqrt{p}),$$

where the last equality follows form Assumption 1. □

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 **??**, we have

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

and

$$\mathrm{E}\,\lambda_1^2(\mathbf{U}_1^T\mathbf{U}_{\mathbf{Y},2;(i,j)}\mathbf{U}_{\mathbf{Y},2;(i,j)}^T\mathbf{U}_1)=O(\frac{p^2}{\boldsymbol{\lambda}_r^2 n^2}).$$

First, we prove that w_{ij}^2 is an approximation of $Y_i^T\mathbf{U}_2\mathbf{U}_2^TY_j$. 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{B.40}
$$

We deal with the five terms separately. First we deal with $\epsilon_{ij}^{(1)}$. Note that

$$
\mathbf{U}_{\mathbf{Y},2;(i,j)} \mathbf{U}_{\mathbf{Y},2;(i,j)}^T - \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)} \tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}^T = \mathbf{U}_{\mathbf{Y};(i,j)} \mathbf{U}_{\mathbf{Y};(i,j)}^T - \mathbf{U}_{\mathbf{Y},1;(i,j)} \mathbf{U}_{\mathbf{Y},1;(i,j)}^T
$$

is a projection matrix whose rank is not larger than $n-k-2-r$. By definition, Y_i, Y_j and $\tilde{\mathbf{U}}_{\mathbf{Y};(i,j)}$ are mutually independent. Then

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

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

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

The terms $\epsilon_{i,j}^{(3)}$ 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

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

Hence we have bound $\mathbf{E}(\epsilon_{ij}^2) = o(p)$.

Note that

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

We have

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

It follows that

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

Now we only need to prove that

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

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

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

To prove the proposition, we only need to show that

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

Note that

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

It follows that

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

This completes the proof. □

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grant No. 11471035, 11471030.

References

Ahn, J. and J. S. Marron (2010). The maximal data piling direction for discrimination. *Biometrika 97*(1), 254–259.

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., Z. Ma, and Y. Wu (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probability Theory & Related Fields 161*(3-4), 781–815.

Cai, T. T., W. Liu, and Y. Xia (2014). Two-sample test of high dimensional means under dependence. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES B-STATISTICAL METHODOLOGY 76*(2), 349–372.

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

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

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

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, Z. Wang, and L. Zhu (2015). Two-sample behrens-fisher problem for high-dimensional data. *Statistica Sinica 25*(4), 1297–1312.

Horn, R. A. and C. R. Johnson (2012). *Matrix Analysis* (2nd ed.). New York: Cambridge University Press.

Laurent, B. and P. Massart (2000, October). Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics 28*(5), 1302–1338.

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

ROY, S. (1953). 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 Multi-*

variate 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*, 317–333.

Srivastava, M. S. (2007). Multivariate theory for analyzing high dimensional data. *Journal of the Japan Statistical Society 37*(1), 53–86.

Srivastava, M. S. and T. Kubokawa (2013). Tests for multivariate analysis of variance in high dimension under non-normality. *Journal of Multivariate Analysis 115*, 204–216.

Tsai, C.-A. and J. J. Chen (2009). Multivariate analysis of variance test for gene set analysis. *Bioinformatics 25*(7), 897–903.

Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. *Eprint Arxiv*.

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

Yamada, T. and T. Himeno (2015). Testing homogeneity of mean vectors under heteroscedasticity in high-dimension. *Journal of Multivariate Analysis 139*, 7–27.

Zhao, J. and X. Xu (2016). A generalized likelihood ratio test for normal mean when p is greater than n. *Computational Statistics & Data Analysis 99*, 91–104.

Zhou, B., J. Guo, and J.-T. Zhang (2017). High-dimensional general linear hypothesis testing under heteroscedasticity. *Journal of Statistical Planning and Inference 188*, 36–54.

School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, 100081,China

E-mail: wangruiphd@bit.edu.cn

School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, 100081,China

and Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing 100081,China

E-mail: xuxz@bit.edu.cn