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 distributions of the test statistic are derived under both nonspiked and spiked covariances. The local asymptotic power function of the test is also given. The asymptotic power function and simulations show that the proposed test is particularly powerful under spiked covariance.

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 ($1 \le i \le k$), the observations $\{X_{ij}\}_{j=1}^{n_i}$ are independent and identically distributed (iid) as $\mathcal{N}_p(\theta_i, \mathbf{\Sigma})$, the p-dimensional normal distribution with mean vector θ_i and common variance matrix $\mathbf{\Sigma}$. We would like to test the hypotheses

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_k \quad \text{v.s.} \quad H_1 : \theta_i \neq \theta_j \text{ for some } i \neq j. \tag{1.1}$$

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 $\mathbf{H} = \sum_{i=1}^{k} n_i(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^\top$ be the sum-of-squares between groups and $\mathbf{G} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{\mathbf{X}}_i)(X_{ij} - \bar{\mathbf{X}}_i)^\top$ 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 hypotheses (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 hypotheses (1.1) in high dimensional setting, i.e., p is greater than N. See, e.g., Verstynen et al. (2005) and Tsai and Chen (2009). However, when $p \geq N$, the four classical test statistics are all not defined. Researchers have done extensive work to study the testing problem (1.1) in high dimensional setting. So far, numerous tests have been proposed for the case $k = 2$. See, e.g., Bai and Saranadasa (1996), Srivastava (2007), Chen and Qin (2010), Cai et al. (2014) and Feng et al. (2015). Some tests have also been introduced for the case of general $k \geq 2$. Schott (2007) modified Hotelling-Lawley trace and proposed the test statistic

$$T_{Sc} = \frac{1}{\sqrt{N-1}}\Big(\frac{1}{k-1}\operatorname{tr}\big(\mathbf{H}\big) - \frac{1}{N-k}\operatorname{tr}\big(\mathbf{G}\big)\Big).$$

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 $\sum_{i=1}^{k} n_i \|\theta_i - \bar{\theta}\|^2$, where $\bar{\theta} = N^{-1}\sum_{i=1}^{k} n_i\theta_i$. See Srivastava and Kubokawa (2013), Hu et al. (2017), Yamada and Himeno (2015), Zhang et al. (2017), Zhou et al. (2017) and Cao et al. (2017) for some other sum-of-squares type test statistics for general $k \geq 2$. It is known that the sum-of-squares type tests are particularly powerful against dense alternatives. 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}, the test statistic T_{CX} is an extreme value type one and is very powerful against sparse alternatives.

Most existing sum-of-squares type test procedures require the condition $\operatorname{tr}(\mathbf{\Sigma}^4)/\operatorname{tr}^2(\mathbf{\Sigma}^2) \to 0$, which is equivalent to

$$\frac{\boldsymbol{\lambda}_1}{\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}} \to 0, \tag{1.2}$$

where $\boldsymbol{\lambda}_i$ is the ith largest eigenvalue of $\mathbf{\Sigma}$, $i = 1, \ldots, p$. The condition (1.2) is reasonable if $\mathbf{\Sigma}$ is nonspiked in the sense that it does not have significantly large eigenvalues. In some important situations, however, variables are heavily correlated with common factors, and the covariance matrix $\mathbf{\Sigma}$ is thus spiked in the sense that a few eigenvalues of $\mathbf{\Sigma}$ are significantly larger than the others (Fan et al., 2013; Cai et al., 2015; Wang and Fan, 2017). In such cases, the condition (1.2) can be violated, and consequently, existing sum-of-squares type tests may not have correct level. Some adjusted sum-of-squares type test procedures have been proposed to solve the problem. See, e.g., Katayama et al. (2013), Ma et al. (2015), Zhang et al. (2017) and Wang and Xu (2018a). However, the power behavior of these corrected tests may not be satisfied.

Recently, Aoshima and Yata (2018) and Wang and Xu (2018b) considered two sample mean testing problem under the spiked covariance model. These tests have better power behavior compared with sum-of-squares type tests. However, both papers imposed strong conditions on the magnitude of p. For example, under the approximate factor model in Fan et al. (2013), the test in Aoshima and Yata (2018) requires $p/n \to 0$, while the test in Wang and Xu (2018b) requires that $p/n^2 \to 0$ and the small eigenvalues of $\mathbf{\Sigma}$ are all equal.

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. However, this phenomenon is not theoretically proved.

In this paper, we propose a generalized likelihood ratio test statistic for hypotheses (1.1) called least favorable direction (LFD) test statistic, which is a generalization of the test in Zhao and Xu (2016). We give the asymptotic distributions of the test statistic under both nonspiked and spiked covariances. An adaptive LFD test procedure is constructed by consistently detecting unknown covariance structure and estimating unknown parameters. The asymptotic local power function of the LFD test is also given. Our theoretical results show that the LFD test is particularly powerful under the spiked covariance. This explains the simulation results of Zhao and Xu (2016). Compared with the work of Zhao and Xu (2016), our main contribution is that we give a thoroughly theoretical analysis of the LFD test. To prove of our main results, we carefully study the high-order asymptotic behavior of the eigenvalues and eigenspaces of the sample covariance matrix. These results are also of independent interests. We further compare the proposed test procedure with existing tests by simulations. It is shown that the LFD test has similar behavior to existing sum-of-squares tests under the nonspiked covariance, while significantly outperforms competing tests under the spiked covariance.

The rest of the paper is organized as follows. In Section 2, we propose the LFD test statistic and derive its explicit forms. The asymptotic distributions of the LFD test statistic under both nonspiked and spiked covariances are given in Section 3. Based on these theoretical results, an adaptive LFD test procedure is proposed. Section 4 complements our study with numerical simulations. In Section 5, 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 $\mathbf{G} = \mathbf{X}(\mathbf{I}_N - \mathbf{J}\mathbf{J}^\top)\mathbf{X}^\top$ 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, $i = 1, \ldots, k$. Let $n = N - k$ be the degrees of freedom of $\mathbf{G}$. Construct an $N \times n$ 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 $\tilde{\mathbf{J}}^\top \tilde{\mathbf{J}} = \mathbf{I}_n$ and $\tilde{\mathbf{J}}\tilde{\mathbf{J}}^\top = \mathbf{I}_N - \mathbf{J}\mathbf{J}^\top$. Define $\mathbf{Y} = \mathbf{X}\tilde{\mathbf{J}}$. Then $\mathbf{G}$ can be written as

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

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

$$\mathbf{H} = \mathbf{X}(\mathbf{J}\mathbf{J}^\top - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^\top)\mathbf{X}^\top = \mathbf{X}\mathbf{J}(\mathbf{I}_k - \frac{1}{N}\mathbf{J}^\top\mathbf{1}_N\mathbf{1}_N^\top\mathbf{J})\mathbf{J}^\top\mathbf{X}^\top.$$

By some matrix algebra, we have $\mathbf{I}_k - N^{-1}\mathbf{J}^\top\mathbf{1}_N\mathbf{1}_N^\top\mathbf{J} = \mathbf{C}\mathbf{C}^\top$ 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

$$\mathbf{H} = \mathbf{XJCC}^\top \mathbf{J}^\top \mathbf{X}^\top.$$

Define $\mathbf{\Theta} = (\sqrt{n_1}\theta_1, \ldots, \sqrt{n_k}\theta_k)$. Then the null hypothesis H_0 is equivalent to $\mathbf{\Theta C} = \mathbf{O}_{p\times(k-1)}$, where $\mathbf{O}_{p\times(k-1)}$ is a $p \times (k-1)$ matrix with 0 entries. Thus, the hypotheses (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 $\{\mathbf{X}_a = a^\top \mathbf{X} :\ a \in \mathbb{R}^p, a^\top a = 1\}$. This induces a decomposition of the null hypothesis and the alternative hypothesis:

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

where $H_{0a} : a^\top \mathbf{\Theta C} = \mathbf{O}_{1\times(k-1)}$ and $H_{1a} : a^\top \mathbf{\Theta C} \neq \mathbf{O}_{1\times(k-1)}$. 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 $a^\top a = 1$, the component LRT statistic

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

can be used to test H_{0a} v.s. H_{1a}. Using union intersection principle, Roy proposed the test statistic $\max_{a^\top a=1} L_1(a)/L_0(a) = (1 + \lambda_1(\mathbf{H}\mathbf{G}^{-1}))^{N/2}$, 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

$$a^* = \operatorname*{arg\,max}_{a^\top a=1} \frac{L_1(a)}{L_0(a)} \tag{2.3}$$

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 $n \geq p$, hence can not be used in high dimensional setting. In what follows, we assume $p > n$. In this case, the set

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

is not empty since $\mathbf{G}$ is singular. Consequently, the right hand side of (2.3) is not well defined since the ratio involves infinity. Hence we need a new definition for LFD in high dimensional setting. Define

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

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,

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

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

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

The null hypothesis is rejected 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 the LFD test statistic. Let $\mathbf{Y} = \mathbf{U}_\mathbf{Y}\mathbf{D}_\mathbf{Y}\mathbf{V}_\mathbf{Y}^\top$ be the singular value decomposition of $\mathbf{Y}$, where $\mathbf{U}_\mathbf{Y}$ and $\mathbf{V}_\mathbf{Y}$ are $p \times \min(n,p)$ and $n \times \min(n,p)$ column orthogonal matrices, respectively, and $\mathbf{D}_\mathbf{Y}$ is a $\min(n,p) \times \min(n,p)$ diagonal matrix whose diagonal elements are the non-increasingly ordered singular values of $\mathbf{Y}$. If $p > n$, let $\mathbf{P}_\mathbf{Y} = \mathbf{U}_\mathbf{Y}\mathbf{U}_\mathbf{Y}^\top$ be the projection matrix onto the column space of $\mathbf{Y}$. Then Lemma 1 in Appendix implies that for $p > n$,

$$T(\mathbf{X}) = \lambda_1\big(\mathbf{C}^\top\mathbf{J}^\top\mathbf{X}^\top(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{X}\mathbf{J}\mathbf{C}\big). \tag{2.4}$$

While (2.4) is convenient for theoretical analysis, it is not convenient for computation. When $p > N$, another simple form of $T(\mathbf{X})$ can be used for computation. If $p > N$, then $\mathbf{X}^\top\mathbf{X}$ is invertible. By the relationship

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

and matrix inverse formula, we have that

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

Thus,

$$T(\mathbf{X}) = \lambda_1\Big(\mathbf{C}^\top\big(\mathbf{J}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{J}\big)^{-1}\mathbf{C}\Big). \tag{2.5}$$

Compared with (2.4), the expression (2.5) doesn't involve $\mathbf{P}_\mathbf{Y}$ and is more convenient for computation.

In the case of $k = 2$, it can be seen that the least favorable direction is proportional to $(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)$ and the 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).

3. Theoretical analysis

We now turn to the analysis of the asymptotic distributions of the LFD test statistic. We shall give theoretical results under both nonspiked and spiked covariances. Based on these results, an adaptive test with asymptotically correct level can be constructed. Also, these results allow us to derive the local asymptotic power function of LFD test.

3.1 Nonspiked covariance

In this subsection, we establish the asymptotic distribution of $T(\mathbf{X})$ under the nonspiked covariance. Let $\mathbf{W}_{k-1}$ be a $(k-1) \times (k-1)$ symmetric random matrix whose entries above the main diagonal are iid $\mathcal{N}(0,1)$ random variables and the entries on the diagonal are iid $\mathcal{N}(0,2)$ random variables. The following theorem establishes the asymptotic distribution of the LFD test statistic.

Theorem 1. *Suppose as $n, p \to \infty$, the condition* (1.2) *holds. Furthermore, suppose that $n\boldsymbol{\lambda}_1/\operatorname{tr}(\boldsymbol{\Sigma}) \to 0$ and $\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_p = O(n^{-1}\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)})$. Then under the local alternative hypothesis $\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\mathbf{C}\| = O(\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)})$,*

$$\frac{T(\mathbf{X}) - (\operatorname{tr}(\boldsymbol{\Sigma}) - n\operatorname{tr}(\boldsymbol{\Sigma}^2)/\operatorname{tr}(\boldsymbol{\Sigma}))}{\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)}} \sim \lambda_1\left(\mathbf{W}_{k-1} + \frac{\mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\mathbf{C}}{\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)}}\right) + o_P(1),$$

where $\sim$ means having the same distribution.

Remark 1. The condition $n\boldsymbol{\lambda}_1/\operatorname{tr}(\boldsymbol{\Sigma}) \to 0$ implies $p/n \to \infty$. Hence $T(\mathbf{X})$ is well defined for large n. The condition $\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_p = O(n^{-1}\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)})$ requires that the range of the eigenvalues of $\boldsymbol{\Sigma}$ is not too large.

To centralize $T(\mathbf{X})$ under the conditions of Theorem 1, the parameters $\operatorname{tr}(\boldsymbol{\Sigma})$ and $\operatorname{tr}(\boldsymbol{\Sigma}^2)$ should be estimated. Let $\hat{\boldsymbol{\Sigma}} = n^{-1}\mathbf{G} = n^{-1}\mathbf{Y}\mathbf{Y}^\top$ be the sample covariance matrix. We use the following simple estimators,

$$\widehat{\operatorname{tr}(\boldsymbol{\Sigma})} = \operatorname{tr}(\hat{\boldsymbol{\Sigma}}), \quad \widehat{\operatorname{tr}(\boldsymbol{\Sigma}^2)} = \operatorname{tr}(\hat{\boldsymbol{\Sigma}}^2) - n^{-1}\operatorname{tr}^2(\hat{\boldsymbol{\Sigma}}).$$

Define

$$Q_1 = \frac{T(\mathbf{X}) - \left(\widehat{\operatorname{tr}(\boldsymbol{\Sigma})} - n\widehat{\operatorname{tr}(\boldsymbol{\Sigma}^2)}/\widehat{\operatorname{tr}(\boldsymbol{\Sigma})}\right)}{\sqrt{\widehat{\operatorname{tr}(\boldsymbol{\Sigma}^2)}}}.$$

Let $F_1(x)$ be the cumulative distribution function of $\lambda_1(\mathbf{W}_{k-1})$. Then we reject the null hypothesis if $Q_1 > F_1^{-1}(1-\alpha)$. The following corollary gives the asymptotic local power function of the proposed test under the nonspiked covariance.

Corollary 1. *Under the conditions of Theorem 1,*

$$\Pr\left(Q_1 > F_1^{-1}(1-\alpha)\right) = \Pr\left(\lambda_1\left(\mathbf{W}_{k-1} + \frac{\mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\mathbf{C}}{\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)}}\right) > F_1^{-1}(1-\alpha)\right) + o(1).$$

Corollary 1 shows that under the nonspiked covariance, the LFD test has similar power behavior to existing sum-of-squares type tests. In fact, if $k = 2$, the asymptotic local power function given by Corollary 1 is equal to the asymptotic local power function of the tests in Bai and Saranadasa (1996) and Chen and Qin (2010).

3.2 Spiked covariance

Now we derive the asymptotic results under the spiked covariance, which are much more involved than the nonspiked case. Let $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$ denote the eigenvalue decomposition of $\boldsymbol{\Sigma}$, where $\boldsymbol{\Lambda} = \operatorname{diag}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_p)$ and $\mathbf{U}$ is an orthogonal matrix. Suppose that $\boldsymbol{\Sigma}$ has r spiked eigenvalues, where $1 \leq r \leq p$ can also vary as $n, p \to \infty$. We shall first assume the spiked number r is known. Adaptation to unknown r will be considered latter. Denote

$\mathbf{\Lambda}_1 = \operatorname{diag}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r)$ and $\mathbf{\Lambda}_2 = \operatorname{diag}(\boldsymbol{\lambda}_{r+1}, \ldots, \boldsymbol{\lambda}_p)$. Correspondingly, we denote $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2)$ where $\mathbf{U}_1$ and $\mathbf{U}_2$ are the first r columns and the last $p - r$ columns of $\mathbf{U}$. Then $\mathbf{\Sigma} = \mathbf{U}_1\mathbf{\Lambda}_1\mathbf{U}_1^\top + \mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top$.

First we shall derive the asymptotic properties of the eigenvalues and eigenspaces of the sample covariance matrix $\hat{\mathbf{\Sigma}}$ since they play a key role in our latter analysis. The following proposition gives the asymptotic behavior of $\lambda_1(\hat{\mathbf{\Sigma}}), \ldots, \lambda_r(\hat{\mathbf{\Sigma}})$ and $\sum_{i=r+1}^n \lambda_i(\hat{\mathbf{\Sigma}})$.

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

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

and

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

Remark 2. Recently, the asymptotic behavior of the spiked eigenvalues of the sample covariance matrix is actively studied. See, e.g., Yata and Aoshima (2013); Shen et al. (2016); Wang and Fan (2017); Cai et al. (2017). An important improvement of Proposition 1 over existing results is that Proposition 1 does not impose any condition for the structure of $\mathbf{\Sigma}$ while still gives the correct convergence rate.

Based on Proposition 1, we propose the following estimators of $\operatorname{tr}(\mathbf{\Lambda}_2)$

and $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r$,

$$\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)} = \Big(1 - \frac{r}{n}\Big)^{-1} \sum_{i=r+1}^{n} \lambda_i(\hat{\boldsymbol{\Sigma}}), \quad \hat{\boldsymbol{\lambda}}_i = \lambda_i(\hat{\boldsymbol{\Sigma}}) - n^{-1}\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}, \quad i = 1, \ldots, r.$$

Moreover, our latter analysis requires an estimator of $\operatorname{tr}(\boldsymbol{\Lambda}_2^2)$. We propose the following estimator of $\operatorname{tr}(\boldsymbol{\Lambda}_2^2)$,

$$\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)} = \sum_{i=r+1}^{n} \Big(\lambda_i(\hat{\boldsymbol{\Sigma}}) - n^{-1}\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\Big)^2.$$

The following proposition gives the convergence rate of these estimators.

Proposition 2. *Suppose that $r = o(n)$. Then uniformly for $i = 1, \ldots, r$,*

$$\hat{\boldsymbol{\lambda}}_i = \boldsymbol{\lambda}_i + O_P\left(\boldsymbol{\lambda}_i \sqrt{\frac{r}{n}} + \sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{n}} + \boldsymbol{\lambda}_{r+1}\right);$$

and

$$\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)} = \operatorname{tr}(\boldsymbol{\Lambda}_2) + O_P\left(r\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{n}} + r\boldsymbol{\lambda}_{r+1}\right),$$
$$\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)} = \operatorname{tr}(\boldsymbol{\Lambda}_2^2) + O_P\left(\frac{r\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{n} + r\boldsymbol{\lambda}_{r+1}^2\right).$$

Remark 3. Our estimators of $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r$ and $\operatorname{tr}(\boldsymbol{\Lambda}_2)$ are similar to some existing estimators, e.g., the noise-reduction estimators in Yata and Aoshima (2012) and the estimators in Wang and Fan (2017). However, their theoretical results require that r is fixed, p is not large and $\boldsymbol{\Sigma}$ satisfies certain spiked covariance models.

Remark 4. The estimation of $\operatorname{tr}(\mathbf{\Lambda}_2^2)$ is relatively unexplored. Recently, Aoshima and Yata (2018) proposed an estimator of $\operatorname{tr}(\mathbf{\Lambda}_2^2)$ by using the cross-data-matrix methodology. They also proved the consistency of their estimator. Their method relies, however, on an arbitrary split of the data into two samples of equal size.

Next we consider the asymptotic behavior of the eigenspaces of $\hat{\mathbf{\Sigma}}$. Let $\mathbf{U}_{\mathbf{Y},1}$ denote the first r columns of $\mathbf{U}_{\mathbf{Y}}$. Then the columns of $\mathbf{U}_{\mathbf{Y},1}$ are the principal eigenvectors of $\hat{\mathbf{\Sigma}}$, and $\mathbf{P}_{\mathbf{Y},1} = \mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^\top$ is the projection matrix onto the rank r principal subspace of $\hat{\mathbf{\Sigma}}$. The properties of $\mathbf{P}_{\mathbf{Y},1}$ and individual principal eigenvectors have been extensively studied. See Cai et al. (2015), Shen et al. (2016), Wang and Fan (2017) and the references therein. The existing results include the consistency of the principal subspace and the high-order asymptotic behavior of the individual principal eigenvectors. However, these results are not enough for our latter analysis. The following proposition gives the high-order asymptotic behavior of $\mathbf{P}_{\mathbf{Y},1}$. To the best of our knowledge, such result has never been appeared in the literature before.

Write $\mathbf{Y} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}$, where $\mathbf{Z}$ is a $p \times n$ random matrix with iid $\mathcal{N}(0,1)$ entries. Then $\mathbf{Y} = \mathbf{U}_1\mathbf{\Lambda}_1^{1/2}\mathbf{Z}_1 + \mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2$, where $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are the first r rows and last $p-r$ rows of $\mathbf{Z}$.

Proposition 3. *Suppose that* $r = o(n)$, $\operatorname{tr}(\boldsymbol{\Lambda}_2)/(n\boldsymbol{\lambda}_r) \to 0$ *and* $r\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2) \to 0$. *Then*

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

where $\|\cdot\|$ *is the spectral norm,* $\mathbf{P}^{\dagger}_{\mathbf{Y},1} = \mathbf{U}_1\mathbf{U}_1^\top + \mathbf{U}_1\mathbf{Q}^\top\mathbf{U}_2^\top + \mathbf{U}_2\mathbf{Q}\mathbf{U}_1^\top$ *and* $\mathbf{Q} = \boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{Z}_1^\top(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}\boldsymbol{\Lambda}_1^{-1/2}$.

Remark 5. The condition $\operatorname{tr}(\boldsymbol{\Lambda}_2)/(n\boldsymbol{\lambda}_r) \to 0$ is commonly adopted in the study of the principal subspaces. In fact, when this condition is violated, the principal subspace will lose its relation to the rank r eigenspace of $\boldsymbol{\Sigma}$. See, e.g., Nadler (2008).

Remark 6. Recently, some high-order Davis-Kahan theorems are established, e.g., Lemma 2 in Koltchinskii and Lounici (2016) and Lemma 2 in Fan et al. (2017). These general results explicitly characterizes the linear term and high-order error on rank r eigenspace due to matrix perturbation. By applying these results to $\hat{\boldsymbol{\Sigma}}$ and $\boldsymbol{\Sigma}$, one can obtain similar results to Proposition 3. Compared with Proposition 3, however, the results so obtained are slightly weaker and requires stronger conditions.

If $p > n$, let $\mathbf{U}_{\mathbf{Y},2}$ be the $r+1$ to nth columns of $\mathbf{U}_{\mathbf{Y}}$. Then $\mathbf{P}_{\mathbf{Y},2} = \mathbf{U}_{\mathbf{Y},2}\mathbf{U}_{\mathbf{Y},2}^\top$ is the projection matrix onto the eigenspace spanned by the $r+1$ to nth eigenvectors of $\hat{\boldsymbol{\Sigma}}$. Our latter analysis also requires the asymptotic properties of $\mathbf{P}_{\mathbf{Y},2}$, which has not been considered in the literature. Let $\mathbf{V}_{\mathbf{Z}_1} = \mathbf{Z}_1^\top(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1/2}$. Then $\mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^\top = \mathbf{Z}_1^\top(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}\mathbf{Z}_1$ 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 $\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top = \mathbf{I}_n - \mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^\top$. The following proposition gives the asymptotic behavior of $\mathbf{P}_{\mathbf{Y},2}$.

Proposition 4. *Suppose that* $r = o(n)$, $\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1/(n\boldsymbol{\lambda}_r^2) \to 0$ *and* $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2) \to 0$. *Then*

$$\left\|\mathbf{P}_{\mathbf{Y},2} - \mathbf{P}_{\mathbf{Y},2}^\dagger\right\| = O_P\left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}} + \sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\mathbf{\Lambda}_2)}}\right),$$

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

Remark 7. The condition $\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1/(n\boldsymbol{\lambda}_r^2) \to 0$ is stronger than the condition $\operatorname{tr}(\mathbf{\Lambda}_2)/(n\boldsymbol{\lambda}_r) \to 0$ in Proposition 3. These two conditions are equivalent if $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_r$ are of the same order.

Now we are ready to derive the asymptotic properties of $T(\mathbf{X})$ under the spiked covariance. Let $\mathbf{W}_{k-1}^*$ be a $(k-1)\times(k-1)$ symmetric random matrix distributed as $\operatorname{Wishart}(r, \mathbf{I}_{k-1})$ and is independent of $\mathbf{W}_{k-1}$, where $\operatorname{Wishart}(m, \mathbf{\Psi})$ is the Wishart distribution with parameter $\mathbf{\Psi}$ and m degrees of freedom. The following theorem gives the asymptotic distribution of $T(\mathbf{X})$ under the null and the local alternative hypothesis.

Theorem 2. *Suppose that* $r = o(\sqrt{n})$, $r \operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1/(n\boldsymbol{\lambda}_r^2) \to 0$, $rn\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2) \to 0$, $r\boldsymbol{\lambda}_{r+1}/\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)} \to 0$ *and* $\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p = O(n^{-1}\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)})$. *Then*

(i) under the null hypothesis $\mathbf{\Theta}\mathbf{C} = \mathbf{O}_{p\times(k-1)}$,

$$\begin{aligned}&\frac{T(\mathbf{X}) - ((1+r/n)\operatorname{tr}(\mathbf{\Lambda}_2) - n\operatorname{tr}(\mathbf{\Lambda}_2^2)/\operatorname{tr}(\mathbf{\Lambda}_2))}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}}\\ &\sim\lambda_1\left(\frac{n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}}(\mathbf{W}_{k-1}^* - r\mathbf{I}_{k-1}) + \frac{\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)}}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}}\mathbf{W}_{k-1}\right) + o_P(1);\end{aligned}$$

(ii) if $r \to \infty$ *or* $\operatorname{tr}(\mathbf{\Lambda}_2)/(n\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)}) \to 0$, *then under the local alternative hypothesis* $\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\| = O(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)})$,

$$\begin{aligned}&\frac{T(\mathbf{X}) - ((1+r/n)\operatorname{tr}(\mathbf{\Lambda}_2) - n\operatorname{tr}(\mathbf{\Lambda}_2^2)/\operatorname{tr}(\mathbf{\Lambda}_2))}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}}\\ &\sim\lambda_1\left(\frac{n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}}(\mathbf{W}_{k-1}^* - r\mathbf{I}_{k-1}) + \frac{\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)}}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}}\mathbf{W}_{k-1}\right.\\ &\quad\left.+\frac{\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C}}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}}\right) + o_P(1).\end{aligned}$$

Remark 8. Suppose the approximate factor model in Fan et al. (2013) holds. That is, r is fixed, $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r$ diverge at rate $O(p)$ and $\boldsymbol{\lambda}_{r+1}, \ldots, \boldsymbol{\lambda}_p$ are bounded. Then the conditions of Theorem 2 become $p/n \to \infty$ and $\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p = O(\sqrt{p}/n)$. Hence Theorem 2 holds for ultra-high dimensional data. In contrast, recently proposed tests under the spiked covariance model can only be used for lower dimensional data. In fact, under the approximate factor model in Fan et al. (2013), Aoshima and Yata (2018) requires $p/n \to$

0, while Wang and Xu (2018b) requires $p/n^2 \to 0$ and $\boldsymbol{\lambda}_{r+1} = \cdots = \boldsymbol{\lambda}_p$. We note that if $k = 2$ and $p/n^2 \to 0$, then the coefficient of $\mathbf{W}_{k-1}^* - r\mathbf{I}_{k-1}$ is negligible, and consequently, $T(\mathbf{X})$ is asymptotically normal distributed. Thus, Theorem 2 gives the high-order behavior of $T(\mathbf{X})$.

Now we formulate a test procedure with asymptotically correct level. Define the standardized statistic as

$$Q_2 = \frac{T(\mathbf{X}) - \left((1 + r/n)\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)} - n\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}/\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\right)}{\sqrt{rn^{-2}(\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)})^2 + \widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}}.$$

Let $F_2(x; \operatorname{tr}(\boldsymbol{\Lambda}_2), \operatorname{tr}(\boldsymbol{\Lambda}_2^2))$ be the cumulative distribution function of

$$\lambda_1 \left(\frac{n^{-1} \operatorname{tr}(\boldsymbol{\Lambda}_2)}{\sqrt{rn^{-2} \operatorname{tr}^2(\boldsymbol{\Lambda}_2) + \operatorname{tr}(\boldsymbol{\Lambda}_2^2)}} (\mathbf{W}_{k-1}^* - r\mathbf{I}_{k-1}) + \frac{\sqrt{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}{\sqrt{rn^{-2} \operatorname{tr}^2(\boldsymbol{\Lambda}_2) + \operatorname{tr}(\boldsymbol{\Lambda}_2^2)}} \mathbf{W}_{k-1} \right).$$

Then we reject the null hypothesis if

$$Q_2 > F_2^{-1}\left(1 - \alpha; \widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}, \widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}\right).$$

The following corollary shows that this test procedure has asymptotically correct level, and also gives the asymptotic local power function.

Corollary 2. *Suppose the conditions of Theorem 2 hold. Then*

(i) under the null hypothesis $\boldsymbol{\Theta}\mathbf{C} = \mathbf{O}_{p\times(k-1)}$,

$$\Pr\left(Q_2 > F_2^{-1}\left(1 - \alpha; \widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}, \widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}\right)\right) = \alpha + o(1);$$

(ii) if $r \to \infty$ *or* $\operatorname{tr}(\mathbf{\Lambda}_2)/(n\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)}) \to 0$*, then under the local alternative hypothesis* $\|\mathbf{C}^\top \mathbf{\Theta}^\top \mathbf{\Theta}\mathbf{C}\| = O(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)})$,

$$
\begin{aligned}
&\Pr\left(Q_2 > F_2^{-1}\left(1-\alpha; \widehat{\operatorname{tr}(\mathbf{\Lambda}_2)}, \widehat{\operatorname{tr}(\mathbf{\Lambda}_2^2)}\right)\right)\\
=&\Pr\left(\lambda_1\left(\frac{n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}}(\mathbf{W}_{k-1}^* - r\mathbf{I}_{k-1}) + \frac{\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)}}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}}\mathbf{W}_{k-1}\right.\right.\\
&\left.\left. + \ \frac{\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C}}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}}\right) > F_2^{-1}\left(1-\alpha;\operatorname{tr}(\mathbf{\Lambda}_2),\operatorname{tr}(\mathbf{\Lambda}_2^2)\right)\right) + o(1).
\end{aligned}
$$

To gain some insight into the asymptotic behavior of $T(\mathbf{X})$, we consider $k = 2$ and compare the LFD test with the tests in Bai and Saranadasa (1996) and Chen and Qin (2010). Corollary 2 implies that if

$$
\liminf_{n\to\infty} \frac{\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C}}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}} > 0,
$$

then the LFD test has nontrivial power asymptotically. In contrast, if

$$
\limsup_{n\to\infty} \frac{\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}}{\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}} = 0,
$$

then the tests in Bai and Saranadasa (1996) and Chen and Qin (2010) has trivial power asymptotically. To compare $\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C}$ and $\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}$, we temporarily place a prior on $\mathbf{\Theta}$. Suppose that $\sqrt{n_i}\theta_i$ has prior distribution $\mathcal{N}_p(\mathbf{0}_p, \psi\mathbf{I}_p)$, $i = 1, 2$. Then $\psi^{-1}\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}$ is distributed as χ^2 distribution with p degrees of freedom. On the other hand, $\psi^{-1}\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C}$ is distributed as χ^2 distribution with $p - r$ degrees of freedom. Then we have

$$\frac{\mathbf{C}^\top \mathbf{\Theta}^\top \mathbf{U}_2 \mathbf{U}_2^\top \mathbf{\Theta} \mathbf{C}}{\mathbf{C}^\top \mathbf{\Theta}^\top \mathbf{\Theta} \mathbf{C}} \xrightarrow{P} 1.$$

So in average, the signal contained in $\mathbf{C}^\top \mathbf{\Theta}^\top \mathbf{U}_2 \mathbf{U}_2^\top \mathbf{\Theta} \mathbf{C}$ is roughly the same as that in $\mathbf{C}^\top \mathbf{\Theta}^\top \mathbf{\Theta} \mathbf{C}$. Now we compare the asymptotic variance. It is not hard to see that

$$\frac{rn^{-2} \operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Sigma}^2)} \to 0.$$

That is, the asymptotic variance of $T(\mathbf{X})$ is much smaller than the tests in Bai and Saranadasa (1996) and Chen and Qin (2010). To appreciate this phenomenon, we note that in the expression (2.4), $(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{XJC}|\mathbf{P}_\mathbf{Y} \sim \mathcal{N}_p(\mathbf{0}_p, (\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}))$. But $\mathbf{I}_p - \mathbf{P}_\mathbf{Y}$ tends to be orthogonal to $\mathbf{U}_1\mathbf{U}_1^\top$ which is the projection matrix onto the eigenspace corresponding to the leading eigenvalues of $\mathbf{\Sigma}$. Hence the projection by $\mathbf{I}_p - \mathbf{P}_\mathbf{Y}$ helps reduce the variance of $\mathbf{XJC}$.

Thus, if $\mathbf{\Theta}$ satisfies

$$\liminf_{n\to\infty} \frac{\mathbf{C}^\top \mathbf{\Theta}^\top \mathbf{\Theta} \mathbf{C}}{\sqrt{rn^{-2} \operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)}} > 0, \quad \limsup_{n\to\infty} \frac{\mathbf{C}^\top \mathbf{\Theta}^\top \mathbf{\Theta} \mathbf{C}}{\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}} = 0,$$

the LFD test has nontrivial power while the tests in Bai and Saranadasa (1996) and Chen and Qin (2010) has trivial power. Hence the LFD test tends to be more powerful than the tests in Bai and Saranadasa (1996) and Chen and Qin (2010).

In practice, one may not know weather the covariance matrix is spiked. Even if it is known that the covariance matrix is spiked, the spike number r may be unknown. So we would like to propose an adaptive test procedure. Note that Theorem 1 requires $n\boldsymbol{\lambda}_1/\operatorname{tr}(\boldsymbol{\Sigma}) \to 0$ while Theorem 2 requires $\operatorname{tr}(\boldsymbol{\Lambda}_2)/n\boldsymbol{\lambda}_r \to 0$ and $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2) \to 0$. This motivates us to consider the following adaptive test procedure. Let $\tau > 1$ be a hyperparameter. If

$$\frac{n\lambda_1(\hat{\boldsymbol{\Sigma}})}{\operatorname{tr}(\hat{\boldsymbol{\Sigma}})} < \tau,$$

then we reject the null hypothesis if $Q_1 > F^{-1}(1-\alpha)$. Otherwise, we reject the null hypothesis if $Q_2 > F_2^{-1}(1-\alpha; \widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2,)}, \widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)})$ where the unknown r is substituted by the estimator

$$\hat{r} = \min\left\{1 \le i < n : \frac{n\lambda_{i+1}(\hat{\boldsymbol{\Sigma}})}{\sum_{j=i+1}^n \lambda_j(\hat{\boldsymbol{\Sigma}})} < \tau\right\}.$$

We have the following proposition.

Proposition 5. *Let $\tau > 1$ be a constant.*

(i) Under the conditions of Theorem 1,

$$\Pr\left(\frac{n\lambda_1(\hat{\boldsymbol{\Sigma}})}{\operatorname{tr}(\hat{\boldsymbol{\Sigma}})} < \tau\right) \to 1;$$

(ii) Under the conditions of Theorem 2,

$$\Pr\left(\frac{n\lambda_1(\hat{\boldsymbol{\Sigma}})}{\operatorname{tr}(\hat{\boldsymbol{\Sigma}})} < \tau\right) \to 0, \quad \Pr(\hat{r} = r) \to 1.$$

Proposition 5 implies that the spiked covariance structure can be consistently detected. So the proposed adaptive LFD test procedure can indeed adapt to the unknown covariance structure.

4. Numerical study

In this section, we compare the numerical performance of the adaptive LFD test procedure with some existing tests, including the MANOVA tests in Schott (2007), Cai and Xia (2014), Hu et al. (2017) and Zhang et al. (2017). These competing tests are denoted by Sc, CX, HBWW and ZGZ, respectively. In the simulations, we take $k = 3$. For the adaptive LFD test, we take $\tau = 5$. For CX, we use their oracle procedure.

First, we simulate the empirical level and power under various models for $\mathbf{\Sigma}$ and $\mathbf{\Theta}$. To characterize the signal strength, we define signal-to-noise ratio (SNR) as

$$\mathrm{SNR} = \frac{\mathbf{C}^\top \mathbf{\Theta}^\top \mathbf{\Theta} \mathbf{C}}{\sqrt{\mathrm{tr}(\mathbf{\Sigma}^2)}}.$$

We consider the following four models for $\mathbf{\Sigma}$.

- Model 1: $\mathbf{\Sigma} = \mathbf{I}_p$.
- Model 2: $\mathbf{\Sigma} = (\sigma_{ij})$ where $\sigma_{ij} = 0.6^{|i-j|}$.
- Model 3: $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ where $\mathbf{U}$ is a $p \times p$ orthogonal matrix generated

from Haar distribution and $\mathbf{\Lambda} = \mathrm{diag}(3p, 2p, p, 1, \ldots, 1)$.

- Model 4: $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top + \mathbf{A}\mathbf{A}^\top$ where $\mathbf{U}$ is a $p \times p$ orthogonal matrix generated from Haar distribution, $\mathbf{\Lambda} = \mathrm{diag}(p, p, 1, \ldots, 1)$ and $\mathbf{A}$ is a $p \times p$ matrix whose elements are independently generated from Bernoulli distribution with success probability 0.01.

Under the null hypothesis, we shall always take $\theta_1 = \cdots = \theta_k = \mathbf{0}_p$. We consider two different structures of alternative hypotheses: the non-sparse alternative and the sparse alternative. In the non-sparse case, we take $\theta_1 = \kappa\mathbf{1}_p$, $\theta_2 = -\kappa\mathbf{1}_p$ and $\theta_3 = \mathbf{0}_p$, where κ is selected to make SNR equal to specific values. In the sparse case, we take $\theta_1 = \kappa(\mathbf{1}_{p/5}^\top, \mathbf{0}_{4p/5}^\top)^\top$, $\theta_2 = \kappa(\mathbf{0}_{p/5}^\top, \mathbf{1}_{p/5}^\top, \mathbf{0}_{3p/5}^\top)^\top$ and $\theta_3 = \mathbf{0}_p$. Again, κ is selected to make SNR equal to specific values. The empirical power is computed based on 5000 simulations. The simulation results under Model 1 and Model 2 are summarized in Tables 1 and 2. The covariance matrices are nonspiked under these two models. It can be seen that the LFD test is less powerful than the sum-of-squares type tests under the nonspiked covariances. Nevertheless, the power performance of the LFD test improves as n and p increase. The simulation results under Model 3 and Model 4 are summarized in Tables 3 and 4. The covariance matrices are spiked under these two models. It can be seen that the LFD test outperforms the competing tests significantly in most cases. This verifies our theoretical results that the LFD test is particularly powerful under the spiked covariance.

In our second simulation study, we would like to investigate the effect of correlations between variables. We consider the compound symmetry structure, that is, the diagonal elements of $\mathbf{\Sigma}$ are 1 and the off-diagonal elements are ρ with $0 \leq \rho < 1$. The parameter ρ characterizes the correlations between variables. We take $\theta_1 = \kappa(\mathbf{1}_{p/5}^\top, \mathbf{0}_{4p/5}^\top)^\top$, $\theta_2 = \kappa(\mathbf{0}_{p/5}^\top, \mathbf{1}_{p/5}^\top, \mathbf{0}_{3p/5}^\top)^\top$ and $\theta_3 = \mathbf{0}_p$, where κ is selected such that $\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}/(\sum_{i=2}^p \boldsymbol{\lambda}_i^2)^{1/2} = 5$. Figure 1 plots the empirical powers of various tests versus ρ, which is computed based on 5000 simulations. We can see that the empirical power of the LFD test holds nearly constant as ρ varies while the empirical powers of competing sum-of-squares type tests decrease as ρ increases. When ρ is non-zero, the LFD test outperforms competing tests significantly.

5. Concluding remarks

In this paper, using the idea of least favorable direction, we proposed the LFD test for MANOVA in high dimensional setting. We derived the asymptotic distribution of the LFD test statistic under both nonspiked and spiked covariances. The asymptotic local power functions are also given. From our theoretic results and simulation studies, it is seen that the LFD test has

Table 1: Empirical sizes and powers of tests. $\alpha = 0.05$, $n_1 = n_2 = n_3 = 20$, $p = 300$.

SNR	Model 1					Model 2				
	Sc	CX	HBWW	ZGZ	LFD	Sc	CX	HBWW	ZGZ	LFD
Non-sparse case										
0	0.0544	0.0362	0.0554	0.0542	0.0446	0.0604	0.0316	0.0598	0.0578	0.0554
0.2	0.0672	0.0350	0.0664	0.0670	0.0500	0.0710	0.0340	0.0704	0.0690	0.0640
0.4	0.0828	0.0380	0.0820	0.0826	0.0610	0.0874	0.0364	0.0872	0.0858	0.0700
0.8	0.1078	0.0420	0.1096	0.1078	0.0776	0.1254	0.0324	0.1250	0.1232	0.0930
1.6	0.2070	0.0490	0.2064	0.2070	0.1594	0.2112	0.0316	0.2112	0.2074	0.1420
3.2	0.4728	0.0628	0.4752	0.4720	0.4280	0.4668	0.0354	0.4666	0.4640	0.3098
Sparse case										
0	0.0516	0.0308	0.0520	0.0516	0.0412	0.0570	0.0324	0.0566	0.0560	0.0544
0.2	0.0662	0.0360	0.0652	0.0662	0.0478	0.0708	0.0340	0.0708	0.0684	0.0632
0.4	0.0782	0.0418	0.0790	0.0780	0.0554	0.0972	0.0336	0.0968	0.0952	0.0714
0.8	0.1042	0.0374	0.1032	0.1040	0.0726	0.1252	0.0322	0.1252	0.1234	0.0856
1.6	0.2074	0.0460	0.2088	0.2072	0.1318	0.2148	0.0314	0.2150	0.2116	0.1378
3.2	0.4670	0.0802	0.4650	0.4668	0.3214	0.4594	0.0318	0.4598	0.4558	0.2718

Table 2: Empirical sizes and powers of tests. $\alpha = 0.05$, $n_1 = n_2 = n_3 = 30$, $p = 800$.

SNR	Model 1					Model 2				
	Sc	CX	HBWW	ZGZ	LFD	Sc	CX	HBWW	ZGZ	LFD
Non-sparse case										
0	0.0522	0.0244	0.0530	0.0522	0.0460	0.0536	0.0268	0.0534	0.0512	0.0498
0.2	0.0624	0.0300	0.0626	0.0624	0.0546	0.0722	0.0250	0.0712	0.0704	0.0616
0.4	0.0792	0.0300	0.0782	0.0792	0.0686	0.0790	0.0286	0.0798	0.0774	0.0726
0.8	0.1142	0.0308	0.1152	0.1142	0.0962	0.1172	0.0278	0.1170	0.1154	0.0992
1.6	0.1990	0.0356	0.1984	0.1990	0.1882	0.2080	0.0228	0.2074	0.2058	0.1608
3.2	0.4710	0.0476	0.4710	0.4710	0.4954	0.4772	0.0328	0.4796	0.4754	0.3978
Sparse case										
0	0.0502	0.0278	0.0498	0.0502	0.0448	0.0570	0.0302	0.0570	0.0556	0.0484
0.2	0.0652	0.0254	0.0662	0.0652	0.0612	0.0630	0.0236	0.0626	0.0626	0.0624
0.4	0.0712	0.0252	0.0714	0.0712	0.0636	0.0856	0.0232	0.0862	0.0854	0.0704
0.8	0.1096	0.0286	0.1104	0.1096	0.0848	0.1114	0.0250	0.1122	0.1104	0.0900
1.6	0.1992	0.0368	0.1992	0.1990	0.1550	0.2068	0.0294	0.2054	0.2038	0.1452
3.2	0.4822	0.0422	0.4828	0.4822	0.3836	0.4722	0.0246	0.4718	0.4696	0.3230

Table 3: Empirical sizes and powers of tests. $\alpha = 0.05$, $n_1 = n_2 = n_3 = 20$, $p = 300$.

SNR	Model 3					Model 4				
	Sc	CX	HBWW	ZGZ	LFD	Sc	CX	HBWW	ZGZ	LFD
Non-sparse case										
0	0.0724	0.0374	0.0728	0.0634	0.0430	0.0776	0.0352	0.0772	0.0706	0.0530
0.2	0.0906	0.2228	0.0902	0.0764	0.9990	0.0824	0.0426	0.0824	0.0760	0.0860
0.4	0.0920	0.5272	0.0938	0.0772	1.0000	0.0982	0.0596	0.0990	0.0910	0.1726
0.8	0.1214	0.9612	0.1202	0.1002	1.0000	0.1166	0.0932	0.1168	0.1084	0.4016
1.6	0.1900	1.0000	0.1884	0.1562	1.0000	0.1942	0.2238	0.1930	0.1784	0.8486
3.2	0.4050	1.0000	0.4078	0.3462	1.0000	0.4164	0.5358	0.4178	0.3826	1.0000
Sparse case										
0	0.0680	0.0344	0.0688	0.0550	0.0466	0.0694	0.0360	0.0696	0.0636	0.0560
0.2	0.0864	0.3582	0.0860	0.0700	0.9894	0.0876	0.0604	0.0882	0.0804	0.0912
0.4	0.0930	0.8438	0.0932	0.0748	1.0000	0.0908	0.1280	0.0918	0.0840	0.1718
0.8	0.1134	0.9998	0.1138	0.0910	1.0000	0.1200	0.3744	0.1204	0.1086	0.4450
1.6	0.1828	1.0000	0.1818	0.1518	1.0000	0.1794	0.7412	0.1792	0.1628	0.9428
3.2	0.4028	1.0000	0.4032	0.3426	1.0000	0.4104	0.9998	0.4110	0.3806	1.0000

Table 4: Empirical sizes and powers of tests. $\alpha = 0.05$, $n_1 = n_2 = n_3 = 30$, $p = 800$.

SNR	Model 3					Model 4				
	Sc	CX	HBWW	ZGZ	LFD	Sc	CX	HBWW	ZGZ	LFD
Non-sparse case										
0	0.0748	0.0250	0.0750	0.0566	0.0426	0.0756	0.0266	0.0750	0.0698	0.0528
0.2	0.0872	0.1984	0.0874	0.0688	1.0000	0.0906	0.0310	0.0900	0.0824	0.0814
0.4	0.0954	0.5260	0.0948	0.0768	1.0000	0.0884	0.0366	0.0886	0.0828	0.1118
0.8	0.1096	0.9672	0.1106	0.0886	1.0000	0.1184	0.0390	0.1192	0.1098	0.2058
1.6	0.1904	1.0000	0.1908	0.1560	1.0000	0.1844	0.0676	0.1838	0.1742	0.4970
3.2	0.3982	1.0000	0.3986	0.3276	1.0000	0.4348	0.1470	0.4346	0.4178	0.9188
Sparse case										
0	0.0764	0.0230	0.0770	0.0564	0.0478	0.0718	0.0264	0.0716	0.0688	0.0548
0.	0.0794	0.3896	0.0796	0.0604	1.0000	0.0816	0.0548	0.0802	0.0758	0.0802
0.4	0.0936	0.9104	0.0934	0.0762	1.0000	0.0978	0.1120	0.0976	0.0918	0.1424
0.8	0.1246	1.0000	0.1250	0.0990	1.0000	0.1072	0.3346	0.1080	0.1014	0.2820
1.6	0.1812	1.0000	0.1806	0.1450	1.0000	0.1788	0.7834	0.1796	0.1702	0.7376
3.2	0.4058	1.0000	0.4060	0.3324	1.0000	0.4212	0.9998	0.4208	0.4028	0.9992

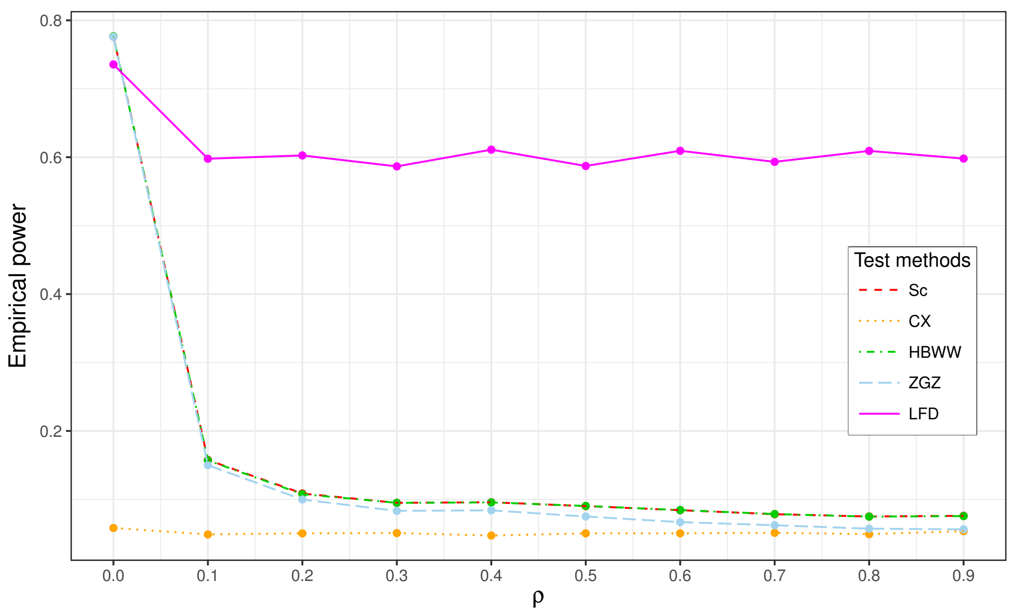

Figure 1: Empirical powers of tests. $\alpha = 0.05$, $n_1 = n_2 = n_3 = 35$, $p = 1000$.

similar power behavior to existing tests when the covariance matrix is non-spiked, while tends to be much more powerful than existing tests when the covariance matrix is spiked.

For the case where the covariance structure is unknown, we proposed an adaptive LFD test procedure by consistently detecting unknown covariance structure and estimating the unknown r. However, this procedure relies on a hyperparameter τ. How to choose an optimal τ is an interesting problem. In addition, our proof relies on the normality of the observations. It is unclear whether the same results hold without normal assumption. We leave these topics for future research.

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grant Nos. 11471035, 11471030.

Appendix A Technical lemmas

Lemma 1. *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* $\mathbf{A} = \mathbf{U_A D_A V_A^\top}$ *the singular value decomposition of* $\mathbf{A}$*, where* $\mathbf{U_A}$ *and* $\mathbf{V_A}$ *are* $p \times r$ *and* $r \times r$ *column orthogonal matrices, respectively, and* $\mathbf{D_A}$ *is a* $r \times r$ *diagonal matrix. Let*

$\mathbf{P}_{\mathbf{A}} = \mathbf{U}_{\mathbf{A}}\mathbf{U}_{\mathbf{A}}^\top$ *be the projection matrix onto the column space of* $\mathbf{A}$*. Then*

$$\max_{a^\top a=1, a^\top \mathbf{A}\mathbf{A}^\top a=0} a^\top \mathbf{B} a = \lambda_1\big(\mathbf{B}(\mathbf{I}_p - \mathbf{P}_{\mathbf{A}})\big).$$

Proof. It can be seen that $a^\top \mathbf{A}\mathbf{A}^\top a = 0$ if and only if $a = (\mathbf{I}_p - \mathbf{P}_{\mathbf{A}})a$. Then

$$\max_{a^\top a=1, a^\top \mathbf{A}\mathbf{A}^\top a=0} a^\top \mathbf{B} a = \max_{a^\top a=1, \mathbf{P}_{\mathbf{A}} a=0} a^\top (\mathbf{I}_p - \mathbf{P}_{\mathbf{A}})\mathbf{B}(\mathbf{I}_p - \mathbf{P}_{\mathbf{A}})a, \tag{A.6}$$

which is obviously no greater than $\lambda_1\big((\mathbf{I} - \mathbf{P}_{\mathbf{A}})\mathbf{B}(\mathbf{I} - \mathbf{P}_{\mathbf{A}})\big)$. To prove that they are equal, without loss of generality, we can assume $\lambda_1\big((\mathbf{I} - \mathbf{P}_{\mathbf{A}})\mathbf{B}(\mathbf{I} - \mathbf{P}_{\mathbf{A}})\big) > 0$. Let α_1 be one eigenvector corresponding to the largest eigenvalue of $(\mathbf{I} - \mathbf{P}_{\mathbf{A}})\mathbf{B}(\mathbf{I} - \mathbf{P}_{\mathbf{A}})$. Since $(\mathbf{I} - \mathbf{P}_{\mathbf{A}})\mathbf{B}(\mathbf{I} - \mathbf{P}_{\mathbf{A}})\mathbf{P}_{\mathbf{A}} = (\mathbf{I} - \mathbf{P}_{\mathbf{A}})\mathbf{B}(\mathbf{P}_{\mathbf{A}} - \mathbf{P}_{\mathbf{A}}) = \mathbf{O}_{p\times p}$ and $\mathbf{P}_{\mathbf{A}}$ is symmetric, the rows of $\mathbf{P}_{\mathbf{A}}$ are eigenvetors of $(\mathbf{I} - \mathbf{P}_{\mathbf{A}})\mathbf{B}(\mathbf{I} - \mathbf{P}_{\mathbf{A}})$ corresponding to eigenvalue 0. It follows that $\mathbf{P}_{\mathbf{A}}\alpha_1 = 0$. Therefore, α_1 satisfies the constraint of (A.6) and thus (A.6) is no less than $\lambda_1\big((\mathbf{I} - \mathbf{P}_{\mathbf{A}})\mathbf{B}(\mathbf{I} - \mathbf{P}_{\mathbf{A}})\big)$. The conclusion now follows by noting that $\lambda_1\big((\mathbf{I} - \mathbf{P}_{\mathbf{A}})\mathbf{B}(\mathbf{I} - \mathbf{P}_{\mathbf{A}})\big) = \lambda_1\big(\mathbf{B}(\mathbf{I} - \mathbf{P}_{\mathbf{A}})\big)$.

□

Lemma 2. *Let* $\xi_{n,i}$*,* $i = 1, \ldots, n$*,* $n = 1, 2, \ldots$*, be iid* s*-dimensional random vectors with mean zero, covariance matrix* $\mathbf{M}$ *and finite fourth moment. For* $n = 1, 2, \ldots$*, let* $\{a_{n,i}\}_{i=1}^n$ *be real random variables which are independent of*

$\{\xi_{n,i}\}_{i=1}^n$ *and satisfy*

$$\frac{\max_{1\le i\le n} a_{n,i}^2}{\sum_{i=1}^n a_{n,i}^2} \xrightarrow{P} 0. \tag{A.7}$$

Then

$$(\sum_{i=1}^n a_{n,i}^2)^{-1/2} \sum_{i=1}^n a_{n,i}\xi_{n,i} \xrightarrow{\mathcal{L}} \mathcal{N}_s(\mathbf{0}_s, \mathbf{M}).$$

Proof. First we observe that if $\{a_{n,i}\}_{i=1}^n$ are fixed numbers satisfying (A.7), then Lyapunov central limit theorem and continuity theorem imply that for any $t \in \mathbb{R}^s$,

$$\mathrm{E}\left[\exp\left((\sum_{i=1}^n a_{n,i}^2)^{-1/2} \sum_{i=1}^n a_{n,i} i t^\top \xi_{n,i}\right)\right] \to \exp\left(-\frac{1}{2} t^\top \mathbf{M} t\right).$$

We only need to prove that for every subsequence of $\{n\}$, there is a further subsequence along which the conclusion holds. Let $\{m(n)\}$ be a subsequence of $\{n\}$. We can find a further subsequence of $\{m(n)\}$ along which (A.7) holds almost surely. Then along this subsequence, our previous argument implies that for any $t \in \mathbb{R}^s$,

$$\mathrm{E}\left[\exp\left((\sum_{i=1}^n a_{n,i}^2)^{-1/2} \sum_{i=1}^n a_{n,i} i t^\top \xi_{n,i}\right) \middle| a_{n,1}, \ldots, a_{n,n}\right] \to \exp\left(-\frac{1}{2} t^\top \mathbf{M} t\right)$$

almost surely. Then by dominated convergence theorem, we have

$$\mathrm{E}\left[\exp\left((\sum_{i=1}^n a_{n,i}^2)^{-1/2} \sum_{i=1}^n a_{n,i} i t^\top \xi_{n,i}\right)\right] \to \exp\left(-\frac{1}{2} t^\top \mathbf{M} t\right)$$

along this further subsequence. This implies the conclusion holds along this further subsequence, which completes the proof.

$\square$

Lemma 3 (Weyl's inequality)**.** *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two symmetric* $n \times n$ *matrices. If* $r + s - 1 \leq i \leq j + k - n$*, we have*

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

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

Lemma 4 (von Neumann's trace theorem)**.** *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two* $m \times n$ *matrices. Let* $\sigma_1(\mathbf{A}) \geq \ldots \geq \sigma_q(\mathbf{A})$ *and* $\sigma_1(\mathbf{B}) \geq \cdots \geq \sigma_q(\mathbf{B})$ *denote the non-increasingly ordered singular values of* $\mathbf{A}$ *and* $\mathbf{B}$*, respectively. Then*

$$\mathrm{tr}(\mathbf{A}\mathbf{B}^\top) \leq \sum_{i=1}^{\min(m,n)} \sigma_i(\mathbf{A})\sigma_i(\mathbf{B}).$$

See, for example, Horn and Johnson (2012) Theorem 7.4.1.1.

Lemma 5. *Let* $\{Z_i\}_{i=1}^n$ *be iid* m*-dimensional random vectors with common distribution* $\mathcal{N}_m(\mathbf{0}_m, \mathbf{I}_m)$*. Then for any* n*-dimensional vector* $\omega = (\omega_1, \ldots, \omega_n)^\top$*, we have*

$$\left\| \sum_{i=1}^n \omega_i (Z_i Z_i^\top - \mathbf{I}_m) \right\| = O_P(|\omega|_2 \sqrt{m} + |\omega|_\infty m),$$

where $|\omega|_2 = \sqrt{\sum_{i=1}^n \omega_i^2}$ *and* $|\omega|_\infty = \max_{1 \leq i \leq n} |\omega_i|$*.*

Remark 9. Our proof implies that the conclusion is still valid if ω is random and is independent of $\{Z_i\}_{i=1}^n$.

Proof. Our 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 $\mathcal{C} \subset \{x \in \mathbb{R}^m : |x|_2 = 1\}$ satisfying $\text{Card}(\mathcal{C}) \leq 9^m$ such that for any $m \times m$ symmetric matrix $\mathbf{A}$,

$$\|A\| \leq 2 \max_{x \in \mathcal{C}} \left| x^\top \mathbf{A} x \right|. \tag{A.8}$$

Then for $t > 4$,

$$\begin{aligned}
&\Pr\left(\left\| \sum_{i=1}^n \omega_i (Z_i Z_i^\top - \mathbf{I}_m) \right\| > t(|\omega|_2 \sqrt{m} + |\omega|_\infty m) \right) \\
\leq &\Pr\left(2 \max_{x \in \mathcal{C}} \left| \sum_{i=1}^n \omega_i (x^\top Z_i Z_i^\top x - 1) \right| > t(|\omega|_2 \sqrt{m} + |\omega|_\infty m) \right) \\
\leq &\sum_{x \in \mathcal{C}} \Pr\left(\left| \sum_{i=1}^n \omega_i (x^\top Z_i Z_i^\top x - 1) \right| > 2|\omega|_2 \sqrt{\frac{mt}{4}} + 2|\omega|_\infty \frac{mt}{4} \right) \\
\leq &2 \cdot 9^m \exp\left(-\frac{mt}{4} \right) = 2 \exp\left((2 \log 3 - t/4) m \right),
\end{aligned}$$

where the first inequality follows from (A.8), the second inequality follows from the union bound and the third inequality follows Lemma 1 of Laurent and Massart (2000). The upper bound $2 \exp\left((2 \log 3 - t/4) m \right)$ can be arbitrarily small as long as t is large enough. This completes the proof. $\square$

Appendix B Proofs of Propositions 1-4

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

$$
\begin{aligned}
n^{-1}\mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z} =& n^{-1}\mathbf{Z}_1^\top\mathbf{\Lambda}_1\mathbf{Z}_1 + n^{-1}\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2 \\
=& n^{-1}\mathbf{Z}_1^\top\mathbf{\Lambda}_1\mathbf{Z}_1 + n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_n + n^{-1}\left(\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2 - \operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_n\right).
\end{aligned}
$$

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

$$
\left|\lambda_i\left(n^{-1}\mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z}\right) - \lambda_i(n^{-1}\mathbf{Z}_1^\top\mathbf{\Lambda}_1\mathbf{Z}_1) - n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\right| \leq n^{-1}\left\|\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2 - \operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_n\right\|. \tag{B.9}
$$

Using Weyl's inequality, we can derive the following lower bound for $\lambda_i(\mathbf{Z}_1^\top\mathbf{\Lambda}_1\mathbf{Z}_1)$, $i = 1, \ldots, r$.

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

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

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

The above lower bound and upper bound imply

$$
\begin{aligned}
\big|\lambda_i(n^{-1}\mathbf{Z}_1^\top \boldsymbol{\Lambda}_1 \mathbf{Z}_1) - \boldsymbol{\lambda}_i\big| \le&\boldsymbol{\lambda}_i \max\big(|\lambda_1(n^{-1}\mathbf{Z}_1\mathbf{Z}_1^\top) - 1|, |\lambda_r(n^{-1}\mathbf{Z}_1\mathbf{Z}_1^\top) - 1|\big)\\
=&\boldsymbol{\lambda}_i\|n^{-1}\mathbf{Z}_1\mathbf{Z}_1^\top - \mathbf{I}_r\|.
\end{aligned}
\tag{B.10}
$$

Combining the bounds (B.9) and (B.10) gives that for $i = 1, \ldots, r$,

$$
\begin{aligned}
&\big|\lambda_i\left(n^{-1}\mathbf{Z}^\top \boldsymbol{\Lambda}\mathbf{Z}\right) - \boldsymbol{\lambda}_i - n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\big|\\
\le&n^{-1}\left\|\mathbf{Z}_2^\top \boldsymbol{\Lambda}_2 \mathbf{Z}_2 - \operatorname{tr}(\boldsymbol{\Lambda}_2)\mathbf{I}_n\right\| + \boldsymbol{\lambda}_i\|n^{-1}\mathbf{Z}_1\mathbf{Z}_1^\top - \mathbf{I}_r\|.
\end{aligned}
$$

From Lemma 5, we have

$$
\|n^{-1}\mathbf{Z}_1\mathbf{Z}_1^\top - \mathbf{I}_r\| = O_P\left(\sqrt{\frac{r}{n}}\right), \tag{B.11}
$$

$$
n^{-1}\left\|\mathbf{Z}_2^\top \boldsymbol{\Lambda}_2 \mathbf{Z}_2 - \operatorname{tr}(\boldsymbol{\Lambda}_2)\mathbf{I}_n\right\| = O_P\left(\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{n}} + \boldsymbol{\lambda}_{r+1}\right). \tag{B.12}
$$

This proves the first statement.

Next we prove the second statement. Note that

$$
\begin{aligned}
\sum_{i=r+1}^{n} \lambda_i(\hat{\mathbf{\Sigma}}) = &\sum_{i=r+1}^{n} \lambda_i(n^{-1}\ \mathbf{Z}^\top \mathbf{\Lambda} \mathbf{Z}) \\
&= \operatorname{tr}(n^{-1}\mathbf{Z}^\top \mathbf{\Lambda} \mathbf{Z}) - \sum_{i=1}^{r} \lambda_i(n^{-1}\ \mathbf{Z}^\top \mathbf{\Lambda} \mathbf{Z}) \\
&= \operatorname{tr}(n^{-1}\mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2) - \frac{r}{n} \operatorname{tr}(\mathbf{\Lambda}_2) \\
&\quad - \left(\sum_{i=1}^{r} \lambda_i(n^{-1}\ \mathbf{Z}^\top \mathbf{\Lambda} \mathbf{Z}) - \operatorname{tr}(n^{-1}\mathbf{Z}_1^\top \mathbf{\Lambda}_1 \mathbf{Z}_1) - \frac{r}{n} \operatorname{tr}(\mathbf{\Lambda}_2) \right).
\end{aligned}
$$

It follows from inequalities (B.9) and (B.12) that

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

Thus,

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

It is straightforward to show that

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

Hence

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

This completes the proof of the second statement. $\square$

Proof of Proposition 2. The first two statements are direct consequences of Proposition 1 and the condition $r = o(n)$. Next we prove the third statement. We have $\widehat{\mathrm{tr}(\mathbf{\Lambda}_2^2)} = n^{-2}\sum_{i=r+1}^{n}\lambda_i^2(\mathbf{Y}^\top\mathbf{Y} - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n)$. Note that Weyl's inequality implies that for $i = r+1, \ldots, n$,

$$\lambda_i(\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n) \leq \lambda_i(\mathbf{Y}^\top\mathbf{Y} - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n) \leq \lambda_{i-r}(\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n).$$

Define

$$\begin{aligned}
\mathcal{C}_1 &= \left\{i : 1 \leq i \leq n,\, \lambda_i\left(\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n\right) > 0\right\}, \\
\mathcal{C}_2 &= \left\{i : r+1 \leq i \leq n,\, \lambda_{i-r}\left(\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n\right) \leq 0\right\}.
\end{aligned}$$

It can be seen that $\mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$ and $\mathrm{Card}(\mathcal{C}_1 \cup \mathcal{C}_2) \geq n - r$. For $i \geq r+1$ and $i \in \mathcal{C}_1$,

$$\lambda_i^2(\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n) \leq \lambda_i^2(\mathbf{Y}^\top\mathbf{Y} - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n) \leq \lambda_{i-r}^2(\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n);$$

for $i \in \mathcal{C}_2$,

$$\lambda_{i-r}^2(\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n) \leq \lambda_i^2(\mathbf{Y}^\top\mathbf{Y} - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n) \leq \lambda_i^2(\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n);$$

for $i \geq r+1$ and $i \notin \mathcal{C}_1 \cup \mathcal{C}_2$,

$$\lambda_i^2(\mathbf{Y}^\top\mathbf{Y} - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n) \leq \max\left(\lambda_{i-r}^2(\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n), \lambda_i^2(\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2 - \widehat{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n)\right).$$

Therefore,

$$
\begin{aligned}
&\left|\sum_{i=r+1}^{n} \lambda_i^2 \left(\mathbf{Y}^\top \mathbf{Y} - \widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n\right) - \mathrm{tr}(\mathbf{Z}_2^\top \boldsymbol{\Lambda}_2 \mathbf{Z}_2 - \widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n)^2\right| \\
\le & \left|\sum_{i>r,\, i\in\mathcal{C}_1} \lambda_i^2 \left(\mathbf{Y}^\top \mathbf{Y} - \widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n\right) - \sum_{i\in\mathcal{C}_1} \lambda_i^2 \left(\mathbf{Z}_2^\top \boldsymbol{\Lambda}_2 \mathbf{Z}_2 - \widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n\right)\right| \\
& + \left|\sum_{i>r,\, i\in\mathcal{C}_2} \lambda_i^2 \left(\mathbf{Y}^\top \mathbf{Y} - \widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n\right) - \sum_{i\notin\mathcal{C}_1} \lambda_i^2 \left(\mathbf{Z}_2^\top \boldsymbol{\Lambda}_2 \mathbf{Z}_2 - \widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n\right)\right| \\
& + \left|\sum_{i>r,\, i\notin\mathcal{C}_1\cup\mathcal{C}_2} \lambda_i^2 \left(\mathbf{Y}^\top \mathbf{Y} - \widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n\right)\right| \\
\le & 3r\|\mathbf{Z}_2^\top \boldsymbol{\Lambda}_2 \mathbf{Z}_2 - \widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n\|^2 \\
\le & 3r\left(\|\mathbf{Z}_2^\top \boldsymbol{\Lambda}_2 \mathbf{Z}_2 - \mathrm{tr}(\boldsymbol{\Lambda}_2)\mathbf{I}_n\| + \left|\mathrm{tr}(\boldsymbol{\Lambda}_2) - \widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2)}\right|\right)^2 \\
= & O_P\left(rn\,\mathrm{tr}(\boldsymbol{\Lambda}_2^2) + rn^2 \boldsymbol{\lambda}_{r+1}^2\right).
\end{aligned}
\tag{B.13}
$$

where the last equality follows from (B.12) and the second statement of the proposition.

Now we deal with $\mathrm{tr}(\mathbf{Z}_2^\top \boldsymbol{\Lambda}_2 \mathbf{Z}_2 - \widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n)^2$. Let $Z_{2,i}$ be the ith column of $\mathbf{Z}_2$, $i = 1, \ldots, n$. Then

$$
\mathrm{tr}(\mathbf{Z}_2^\top \boldsymbol{\Lambda}_2 \mathbf{Z}_2 - \widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n)^2 = \sum_{i=1}^{n} (Z_{2,i}^\top \boldsymbol{\Lambda}_2 Z_{2,i} - \widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2)})^2 + 2\sum_{1\le i<j\le n} (Z_{2,i}^\top \boldsymbol{\Lambda}_2 Z_{2,j})^2.
$$

For the first term, we have

$$
\sum_{i=1}^{n} (Z_{2,i}^\top \boldsymbol{\Lambda}_2 Z_{2,i} - \widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2)})^2 \le 2\sum_{i=1}^{n} (Z_{2,i}^\top \boldsymbol{\Lambda}_2 Z_{2,i} - \mathrm{tr}(\boldsymbol{\Lambda}_2))^2 + 2n(\widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2)} - \mathrm{tr}(\boldsymbol{\Lambda}_2))^2.
$$

Then it follows from the second statement of the proposition and the fact $\mathrm{E}\sum_{i=1}^{n}(Z_{2,i}^{\top}\mathbf{\Lambda}_2 Z_{2,i}-\operatorname{tr}(\mathbf{\Lambda}_2))^2 = 2n\operatorname{tr}(\mathbf{\Lambda}_2^2)$ that

$$\sum_{i=1}^{n}(Z_{2,i}^{\top}\mathbf{\Lambda}_2 Z_{2,i}-\widehat{\operatorname{tr}(\mathbf{\Lambda}_2)})^2 = O_P\left((n+r^2)\operatorname{tr}(\mathbf{\Lambda}_2^2)+r^2 n\boldsymbol{\lambda}_{r+1}^2\right). \tag{B.14}$$

For the second term, it is straightforward to show that $\mathrm{E}\,2\sum_{1\le i<j\le n}(Z_{2,i}^{\top}\mathbf{\Lambda}_2 Z_{2,j})^2 = n(n-1)\operatorname{tr}(\mathbf{\Lambda}_2^2)$. Furthermore, Chen et al. (2010), Proposition A.2 implies that

$$\begin{aligned}\operatorname{Var}\left(2\sum_{1\le i<j\le n}(Z_{2,i}^{\top}\mathbf{\Lambda}_2 Z_{2,j})^2\right) =& O\left(n^2\operatorname{tr}^2(\mathbf{\Lambda}_2^2)+n^3\operatorname{tr}(\mathbf{\Lambda}_2^4)\right)\\ =& O\left(n^2\operatorname{tr}^2(\mathbf{\Lambda}_2^2)+n\operatorname{tr}(\mathbf{\Lambda}_2^2)n^2\boldsymbol{\lambda}_{r+1}^2\right)\\ =& O\left(n^2\operatorname{tr}^2(\mathbf{\Lambda}_2^2)+n^4\boldsymbol{\lambda}_{r+1}^4\right).\end{aligned}$$

Thus,

$$2\sum_{1\le i<j\le n}(Z_{2,i}^{\top}\mathbf{\Lambda}_2 Z_{2,j})^2 = n^2\operatorname{tr}(\mathbf{\Lambda}_2^2)+O_P\left(n\operatorname{tr}(\mathbf{\Lambda}_2^2)+n^2\boldsymbol{\lambda}_{r+1}^2\right).$$

Combining the last display and (B.14) yields

$$\operatorname{tr}(\mathbf{Z}_2^{\top}\mathbf{\Lambda}_2\mathbf{Z}_2-\widehat{\operatorname{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n)^2 = n^2\operatorname{tr}(\mathbf{\Lambda}_2^2)+O_P\left((n+r^2)\operatorname{tr}(\mathbf{\Lambda}_2^2)+(n+r^2)n\boldsymbol{\lambda}_{r+1}^2\right).$$

Combine the last display and (B.13), we have

$$\sum_{i=r+1}^{n}\lambda_i^2\left(\mathbf{Y}^{\top}\mathbf{Y}-\widehat{\operatorname{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_n\right) = O_P\left(rn\operatorname{tr}(\mathbf{\Lambda}_2^2)+rn^2\boldsymbol{\lambda}_{r+1}^2\right).$$

This completes the proof.

□

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

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

where

$$\mathbf{P}^*_{\mathbf{Y},1} = \mathbf{U}\begin{pmatrix}\mathbf{I}_r \\ \mathbf{Q}\end{pmatrix}\left(\mathbf{I}_r + \mathbf{Q}^\top\mathbf{Q}\right)^{-1}\begin{pmatrix}\mathbf{I}_r & \mathbf{Q}^\top\end{pmatrix}\mathbf{U}^\top.$$

Proof. The following intermediate matrix

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

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

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

Consequently, $\hat{\boldsymbol{\Sigma}}_0$ is a positive semi-definite matrix with rank r, and $\mathbf{P}^*_{\mathbf{Y},1}$ is the projection matrix onto the rank r principal subspace of $\hat{\boldsymbol{\Sigma}}_0$.

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

$$\|\mathbf{P}_{\mathbf{Y},1} - \mathbf{P}^*_{\mathbf{Y},1}\| \le \frac{2\|\hat{\boldsymbol{\Sigma}} - \hat{\boldsymbol{\Sigma}}_0\|}{\lambda_r(\hat{\boldsymbol{\Sigma}}_0)}. \tag{B.15}$$

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

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

where the second last equality follows from (B.12) and the last equality follows from

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

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

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

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 5 implies that

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

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

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

On the other hand, (B.10) and (B.11) imply that

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

Hence we have

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

Then the conclusion follows from (B.15), (B.16) and (B.18). $\square$

Proof of Proposition 3. Note that

$$\left\|\mathbf{P}_{\mathbf{Y},1}-\mathbf{P}^{\dagger}_{\mathbf{Y},1}\right\| \le \left\|\mathbf{P}_{\mathbf{Y},1}-\mathbf{P}^{*}_{\mathbf{Y},1}\right\| + \left\|\mathbf{P}^{*}_{\mathbf{Y},1}-\mathbf{P}^{\dagger}_{\mathbf{Y},1}\right\|.$$

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

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

So we only need to deal with $\|\mathbf{P}^{*}_{\mathbf{Y},1}-\mathbf{P}^{\dagger}_{\mathbf{Y},1}\|$. We have

$$\begin{aligned}
&\left\|\mathbf{P}^{*}_{\mathbf{Y},1}-\mathbf{P}^{\dagger}_{\mathbf{Y},1}\right\| \\
\le&\left\|\mathbf{P}^{*}_{\mathbf{Y},1}-\mathbf{U}\begin{pmatrix}\mathbf{I}_r\\ \mathbf{Q}\end{pmatrix}\begin{pmatrix}\mathbf{I}_r & \mathbf{Q}^\top\end{pmatrix}\mathbf{U}^\top\right\| + \left\|\mathbf{U}\begin{pmatrix}\mathbf{I}_r\\ \mathbf{Q}\end{pmatrix}\begin{pmatrix}\mathbf{I}_r & \mathbf{Q}^\top\end{pmatrix}\mathbf{U}^\top-\mathbf{P}^{\dagger}_{\mathbf{Y},1}\right\| \\
=&\left\|\begin{pmatrix}\mathbf{I}_r\\ \mathbf{Q}\end{pmatrix}\left(\left(\mathbf{I}_r+\mathbf{Q}^\top\mathbf{Q}\right)^{-1}-\mathbf{I}_r\right)\begin{pmatrix}\mathbf{I}_r & \mathbf{Q}^\top\end{pmatrix}\right\| + \left\|\mathbf{U}_2\mathbf{Q}\mathbf{Q}^\top\mathbf{U}_2^\top\right\| \\
=&\left\|\left(\left(\mathbf{I}_r+\mathbf{Q}^\top\mathbf{Q}\right)^{-1}-\mathbf{I}_r\right)\left(\mathbf{I}_r+\mathbf{Q}^\top\mathbf{Q}\right)\right\| + \left\|\mathbf{U}_2\mathbf{Q}\mathbf{Q}^\top\mathbf{U}_2^\top\right\| \\
=&2\left\|\mathbf{Q}^\top\mathbf{Q}\right\|.
\end{aligned}$$

Note that

$$\begin{aligned}
\|\mathbf{Q}^\top\mathbf{Q}\| =& \left\|\mathbf{\Lambda}_1^{-1/2}(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1/2}\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1/2}\mathbf{\Lambda}_1^{-1/2}\right\| \\
\le&\boldsymbol{\lambda}_r^{-1}\left\|(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}\right\|\left\|\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right\| \\
=&O_P\left(\frac{\mathrm{tr}(\mathbf{\Lambda}_2)}{n\boldsymbol{\lambda}_r}\right),
\end{aligned} \tag{B.19}$$

where the second last equality follows from the fact $\|(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}\| = \lambda_r(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}$, (B.11), (B.17) and Weyl's inequality. 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*

$$\left\|\mathbf{P}_{\mathbf{Y},2} - \mathbf{P}^*_{\mathbf{Y},2}\right\| = O_P\left(\min\left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}}, 1\right)\right).$$

where $\mathbf{P}^*_{\mathbf{Y},2} = \mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\left(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\right)^{-1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top$.

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 $\mathbf{P}_{\mathbf{Y},2}$ and $\mathbf{P}^*_{\mathbf{Y},2}$ are both projection matrices, we have $\left\|\mathbf{P}_{\mathbf{Y},2} - \mathbf{P}^*_{\mathbf{Y},2}\right\| \leq 2$. 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 leading $n-r$ eigenvectors of $(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})\hat{\mathbf{\Sigma}}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})$. Under the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2) \to 0$, Proposition 3 implies that

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

It can be seen that

$$\begin{aligned}&\left\|(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1})\hat{\boldsymbol{\Sigma}}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1})-(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})\hat{\boldsymbol{\Sigma}}(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})\right\|\\ \le&\left\|(\mathbf{P}^{\dagger}_{\mathbf{Y},1}-\mathbf{P}_{\mathbf{Y},1})\hat{\boldsymbol{\Sigma}}(\mathbf{P}^{\dagger}_{\mathbf{Y},1}-\mathbf{P}_{\mathbf{Y},1})\right\|+2\left\|(\mathbf{P}^{\dagger}_{\mathbf{Y},1}-\mathbf{P}_{\mathbf{Y},1})\hat{\boldsymbol{\Sigma}}(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})\right\|.\end{aligned}$$

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

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

Then

$$\left\|(\mathbf{P}^{\dagger}_{\mathbf{Y},1}-\mathbf{P}_{\mathbf{Y},1})\hat{\boldsymbol{\Sigma}}(\mathbf{P}^{\dagger}_{\mathbf{Y},1}-\mathbf{P}_{\mathbf{Y},1})\right\|\le\|\hat{\boldsymbol{\Sigma}}\|\left\|\mathbf{P}^{\dagger}_{\mathbf{Y},1}-\mathbf{P}_{\mathbf{Y},1}\right\|^2=O_P\left(\frac{\operatorname{tr}^2(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1}{n^2\boldsymbol{\lambda}_r^2}\right). \tag{B.20}$$

On the other hand, we have

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

It is straightforward to show that

$$\mathbf{Z}^\top\boldsymbol{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})=\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^\top-\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2\mathbf{Z}_1^\top(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}\boldsymbol{\Lambda}_1^{-1/2}\mathbf{U}_1^\top. \tag{B.21}$$

Then

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

It follows from (B.12) and the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2) \to 0$ that

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

Consequently,

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

Thus,

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

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

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

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

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

Then

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

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

Combine the above bounds, we obtain

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

The matrix $n^{-1}\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^\top$ shares the same non-zero eigenvalues as $n^{-1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}$. 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 5 and the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2) \to 0$ that

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

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

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

It can be seen that the matrix $\mathbf{P}_{\mathbf{Y},2}^*$ is the projection matrix onto the rank $n-r$ principal subspace of $n^{-1}\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^\top$. Therefore, Cai et al. (2015), Proposition 1 implies that

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

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

Proof of Proposition 4. By some algebra, it can be seen that

$$
\begin{aligned}
\left\|\mathbf{P}^*_{\mathbf{Y},2}-\mathbf{P}^\dagger_{\mathbf{Y},2}\right\|=&(\operatorname{tr}(\boldsymbol{\Lambda}_2))^{-1}\left\|\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}-\operatorname{tr}(\boldsymbol{\Lambda}_2)\mathbf{I}_{n-r}\right\|\\
=&O_P\left(\frac{\sqrt{n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}+\frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\right)\\
=&O_P\left(\sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}}\right),
\end{aligned}
$$

where the second last equality follows from (B.25) and the last equality follows from the fact $\sqrt{n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}/\operatorname{tr}(\boldsymbol{\Lambda}_2)\leq\sqrt{n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2)}$ and the condition $\sqrt{n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2)}\to 0$. Then the conclusion follows from the last display and Proposition 7.

□

Appendix C Proofs of Theorems 1 and 2

It can be seen that $\mathbf{XJC}$ is independent of $\mathbf{Y}$. We write $\mathbf{XJC} = \mathbf{\Theta C} + \mathbf{U\Lambda}^{1/2}\mathbf{Z}^{\dagger}$, where $\mathbf{Z}^{\dagger}$ is a $p \times (k-1)$ matrix with iid $\mathcal{N}(0,1)$ entries and is independent of $\mathbf{Z}$. Then

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

It can be seen that the first term of (C.27) can be written as

$$\mathbf{Z}^{\dagger\top}\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p - \mathbf{P_Y})\mathbf{U\Lambda}^{1/2}\mathbf{Z}^{\dagger} = \sum_{i=1}^{p}\lambda_i((\mathbf{I}_p - \mathbf{P_Y})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P_Y}))\eta_i\eta_i^\top,$$

where $\eta_1, \dots, \eta_p$ are independent $\mathcal{N}_{k-1}(\mathbf{0}_{k-1}, \mathbf{I}_{k-1})$ random vectors and are independent of $\mathbf{P_Y}$.

Lemma 6. *Suppose that* $n\boldsymbol{\lambda}_1/\operatorname{tr}(\mathbf{\Sigma}) \to 0$. *Then*

$$\operatorname{tr}((\mathbf{I}_p - \mathbf{P_Y})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P_Y})) = \operatorname{tr}(\mathbf{\Sigma}) - \frac{n\operatorname{tr}(\mathbf{\Sigma}^2)}{\operatorname{tr}(\mathbf{\Sigma})} + O_P\left(n(\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_p)\sqrt{\frac{n\boldsymbol{\lambda}_1}{\operatorname{tr}(\mathbf{\Sigma})}}\right)$$

And

$$\operatorname{tr}((\mathbf{I}_p - \mathbf{P_Y})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P_Y}))^2 = \operatorname{tr}(\mathbf{\Sigma}^2) - \frac{n\operatorname{tr}^2(\mathbf{\Sigma}^2)}{\operatorname{tr}^2(\mathbf{\Sigma})} + O_P(n\boldsymbol{\lambda}_1(\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_p)).$$

Proof. First we approximate $\mathbf{P_Y}$ by a simple expression. We have

$$\begin{aligned}\left\|\mathbf{P_Y} - (\operatorname{tr}(\mathbf{\Sigma}))^{-1}\mathbf{YY}^\top\right\| =& \left\|\mathbf{Y}(\mathbf{Y}^\top\mathbf{Y})^{-1}\mathbf{Y}^\top - (\operatorname{tr}(\mathbf{\Sigma}))^{-1}\mathbf{YY}^\top\right\|\\ =& (\operatorname{tr}(\mathbf{\Sigma}))^{-1}\left\|\mathbf{Y}^\top\mathbf{Y} - \operatorname{tr}(\mathbf{\Sigma})\mathbf{I}_n\right\|.\end{aligned}$$

Then from Lemma 5, we have

$$\begin{aligned}
\left\|\mathbf{P}_{\mathbf{Y}}-(\operatorname{tr}(\boldsymbol{\Sigma}))^{-1}\mathbf{Y}\mathbf{Y}^{\top}\right\| =&(\operatorname{tr}(\boldsymbol{\Sigma}))^{-1}\left\|\mathbf{Z}^{\top}\boldsymbol{\Sigma}\mathbf{Z}-\operatorname{tr}(\boldsymbol{\Sigma})\mathbf{I}_n\right\| \\
=&O_P\left(\frac{\sqrt{n\operatorname{tr}(\boldsymbol{\Sigma}^2)}}{\operatorname{tr}(\boldsymbol{\Sigma})}+\frac{n\boldsymbol{\lambda}_1}{\operatorname{tr}(\boldsymbol{\Sigma})}\right) \\
=&O_P\left(\frac{\sqrt{n\boldsymbol{\lambda}_1\operatorname{tr}(\boldsymbol{\Sigma})}}{\operatorname{tr}(\boldsymbol{\Sigma})}+\frac{n\boldsymbol{\lambda}_1}{\operatorname{tr}(\boldsymbol{\Sigma})}\right) \\
=&O_P\left(\sqrt{\frac{n\boldsymbol{\lambda}_1}{\operatorname{tr}(\boldsymbol{\Sigma})}}\right).
\end{aligned} \tag{C.28}$$

Now we deal with $\operatorname{tr}\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)$. It can be seen that

$$\begin{aligned}
\operatorname{tr}\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right) =& \operatorname{tr}(\boldsymbol{\Sigma})-\operatorname{tr}(\boldsymbol{\Sigma}\mathbf{P}_{\mathbf{Y}}) \\
=& \operatorname{tr}(\boldsymbol{\Sigma})-\operatorname{tr}\left(\left(\boldsymbol{\Sigma}-\frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{I}_p\right)\mathbf{P}_{\mathbf{Y}}\right)-\frac{n\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}.
\end{aligned} \tag{C.29}$$

For the second term, we have

$$\begin{aligned}
&\left|\operatorname{tr}\left(\left(\boldsymbol{\Sigma}-\frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{I}_p\right)\mathbf{P}_{\mathbf{Y}}\right)-(\operatorname{tr}(\boldsymbol{\Sigma}))^{-1}\operatorname{tr}\left(\left(\boldsymbol{\Sigma}-\frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{I}_p\right)\mathbf{Y}\mathbf{Y}^{\top}\right)\right| \\
=&\left|\operatorname{tr}\left(\left(\boldsymbol{\Sigma}-\frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{I}_p\right)\left(\mathbf{P}_{\mathbf{Y}}-(\operatorname{tr}(\boldsymbol{\Sigma}))^{-1}\mathbf{Y}\mathbf{Y}^{\top}\right)\right)\right| \\
\leq&2n\left\|\boldsymbol{\Sigma}-\frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{I}_p\right\|\left\|\mathbf{P}_{\mathbf{Y}}-(\operatorname{tr}(\boldsymbol{\Sigma}))^{-1}\mathbf{Y}\mathbf{Y}^{\top}\right\| \\
=&O_P\left(n(\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_p)\sqrt{\frac{n\boldsymbol{\lambda}_1}{\operatorname{tr}(\boldsymbol{\Sigma})}}\right),
\end{aligned}$$

where the last inequality follows from von Neumann's trace theorem and the fact $\operatorname{Rank}\left(\mathbf{P}_{\mathbf{Y}}-(\operatorname{tr}(\boldsymbol{\Sigma}))^{-1}\mathbf{Y}\mathbf{Y}^{\top}\right)\leq 2n$, and the last equality follows from (C.28) and the fact $\operatorname{tr}(\boldsymbol{\Sigma}^2)/\operatorname{tr}(\boldsymbol{\Sigma})\in[\boldsymbol{\lambda}_p,\boldsymbol{\lambda}_1]$. On the other hand, it is straightforward to show that

$$\mathrm{E}\left((\mathrm{tr}(\mathbf{\Sigma}))^{-1}\,\mathrm{tr}\left(\left(\mathbf{\Sigma}-\frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{\mathrm{tr}(\mathbf{\Sigma})}\mathbf{I}_p\right)\mathbf{Y}\mathbf{Y}^\top\right)\right)=0,$$

and

$$\begin{aligned}
&\mathrm{Var}\left((\mathrm{tr}(\mathbf{\Sigma}))^{-1}\,\mathrm{tr}\left(\left(\mathbf{\Sigma}-\frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{\mathrm{tr}(\mathbf{\Sigma})}\mathbf{I}_p\right)\mathbf{Y}\mathbf{Y}^\top\right)\right)\\
=&\frac{2n}{\mathrm{tr}^2(\mathbf{\Sigma})}\,\mathrm{tr}\left(\mathbf{\Sigma}^2-\frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{\mathrm{tr}(\mathbf{\Sigma})}\mathbf{\Sigma}\right)^2\\
=&\frac{2n}{\mathrm{tr}^2(\mathbf{\Sigma})}\sum_{i=1}^{p}\boldsymbol{\lambda}_i^2\left(\boldsymbol{\lambda}_i-\frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{\mathrm{tr}(\mathbf{\Sigma})}\right)^2\\
\leq&\frac{2n\boldsymbol{\lambda}_1(\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_p)^2}{\mathrm{tr}(\mathbf{\Sigma})}.
\end{aligned}$$

Thus,

$$\mathrm{tr}\left(\left(\mathbf{\Sigma}-\frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{\mathrm{tr}(\mathbf{\Sigma})}\mathbf{I}_p\right)\mathbf{P}_{\mathbf{Y}}\right)=O_P\left(n(\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_p)\sqrt{\frac{n\boldsymbol{\lambda}_1}{\mathrm{tr}(\mathbf{\Sigma})}}\right).$$

Then the first statement follows from the last display and (C.29).

Next we deal with $\mathrm{tr}\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)^2$. We have

$$\mathrm{tr}\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)^2=\mathrm{tr}(\mathbf{\Sigma}^2)-2\,\mathrm{tr}(\mathbf{\Sigma}^2\mathbf{P}_{\mathbf{Y}})+\mathrm{tr}((\mathbf{\Sigma}\mathbf{P}_{\mathbf{Y}})^2).$$

From von Neumann's trace theorem, the second term satisfies

$$\left|\mathrm{tr}(\mathbf{\Sigma}^2\mathbf{P}_{\mathbf{Y}})-\frac{n\,\mathrm{tr}^2(\mathbf{\Sigma}^2)}{\mathrm{tr}^2(\mathbf{\Sigma})}\right|=\left|\mathrm{tr}\left(\left(\mathbf{\Sigma}^2-\frac{\mathrm{tr}^2(\mathbf{\Sigma}^2)}{\mathrm{tr}^2(\mathbf{\Sigma})}\mathbf{I}_p\right)\mathbf{P}_{\mathbf{Y}}\right)\right|\leq n\boldsymbol{\lambda}_1(\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_p),$$

and the third term satisfies

$$\begin{aligned}
\left|\mathrm{tr}((\mathbf{\Sigma}\mathbf{P}_{\mathbf{Y}})^2)-\frac{n\,\mathrm{tr}^2(\mathbf{\Sigma}^2)}{\mathrm{tr}^2(\mathbf{\Sigma})}\right|=&\left|\mathrm{tr}\left(\left(\mathbf{\Sigma}+\frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{\mathrm{tr}(\mathbf{\Sigma})}\mathbf{I}_p\right)\mathbf{P}_{\mathbf{Y}}\left(\mathbf{\Sigma}-\frac{\mathrm{tr}(\mathbf{\Sigma}^2)}{\mathrm{tr}(\mathbf{\Sigma})}\mathbf{I}_p\right)\mathbf{P}_{\mathbf{Y}}\right)\right|\\
\leq&2n\boldsymbol{\lambda}_1(\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_p).
\end{aligned}$$

This completes the proof of the second statement. $\square$

Proof of Theorem 1. In the current context, Lemma 6 implies that

$$\operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) = \operatorname{tr}(\boldsymbol{\Sigma}) - \frac{n \operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})} + o_P(\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)}), \tag{C.30}$$

$$\operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)^2 = (1 + o_P(1)) \operatorname{tr}(\boldsymbol{\Sigma}^2). \tag{C.31}$$

The fact $\lambda_1\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) \leq \boldsymbol{\lambda}_1$ and (C.31) imply that the first term of (C.27) satisfies the Lyapunov condition

$$\frac{\lambda_1\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)}{\sqrt{\operatorname{tr}\left(((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}}))^2\right)}} \leq \frac{\boldsymbol{\lambda}_1}{\sqrt{(1 + o_P(1)) \operatorname{tr}^2(\boldsymbol{\Sigma})}} \xrightarrow{P} 0.$$

From Lemma 2, we have

$$\frac{\mathbf{Z}^{\dagger\top}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{Z}^\dagger - \operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)\mathbf{I}_{k-1}}{\sqrt{\operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)^2}} \xrightarrow{\mathcal{L}} \mathbf{W}_{k-1}.$$

Then it follows from (C.30), (C.31) and Slutsky's theorem that

$$\frac{\mathbf{Z}^{\dagger\top}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{Z}^\dagger - \left(\operatorname{tr}(\boldsymbol{\Sigma}) - n \operatorname{tr}(\boldsymbol{\Sigma}^2)/\operatorname{tr}(\boldsymbol{\Sigma})\right)\mathbf{I}_{k-1}}{\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)}} \xrightarrow{\mathcal{L}} \mathbf{W}_{k-1}. \tag{C.32}$$

Next we consider the second term of (C.27). Note that

$$\left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Theta}\mathbf{C} - \mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\mathbf{C}\right\| = \left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{Y}(\mathbf{Y}^\top\mathbf{Y})^{-1}\mathbf{Y}^\top\boldsymbol{\Theta}\mathbf{C}\right\|.$$

We have

$$\begin{aligned}
&\left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{Y}(\mathbf{Y}^\top\mathbf{Y})^{-1}\mathbf{Y}^\top\boldsymbol{\Theta}\mathbf{C} - \operatorname{tr}(\boldsymbol{\Sigma})^{-1}\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{Y}\mathbf{Y}^\top\boldsymbol{\Theta}\mathbf{C}\right\| \\
\leq &\left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{Y}\mathbf{Y}^\top\boldsymbol{\Theta}\mathbf{C}\right\| \left\|(\mathbf{Y}^\top\mathbf{Y})^{-1} - \operatorname{tr}(\boldsymbol{\Sigma})^{-1}\mathbf{I}_n\right\| \\
\leq &\left\|\operatorname{tr}(\boldsymbol{\Sigma})^{-1}\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{Y}\mathbf{Y}^\top\boldsymbol{\Theta}\mathbf{C}\right\| \left\|(\mathbf{Y}^\top\mathbf{Y})^{-1}\right\| \left\|\mathbf{Y}^\top\mathbf{Y} - \operatorname{tr}(\boldsymbol{\Sigma})\mathbf{I}_n\right\|.
\end{aligned}$$

From Lemma 5, we have

$$\begin{aligned}\left\|\mathbf{Y}^\top\mathbf{Y}-\operatorname{tr}(\mathbf{\Sigma})\mathbf{I}_n\right\| = &\left\|\mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z}-\operatorname{tr}(\mathbf{\Sigma})\mathbf{I}_n\right\| \\ =&O_P(\sqrt{n\operatorname{tr}(\mathbf{\Sigma}^2)}+n\boldsymbol{\lambda}_1) \\ =&o_P(\operatorname{tr}(\mathbf{\Sigma})).\end{aligned}$$

Then $\left\|(\mathbf{Y}^\top\mathbf{Y})^{-1}\right\| = \lambda_n^{-1}(\mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z}) = (1+o_P(1))\operatorname{tr}(\mathbf{\Sigma})$. Therefore,

$$\begin{aligned}&\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{Y}(\mathbf{Y}^\top\mathbf{Y})^{-1}\mathbf{Y}^\top\mathbf{\Theta}\mathbf{C}-\operatorname{tr}(\mathbf{\Sigma})^{-1}\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{Y}\mathbf{Y}^\top\mathbf{\Theta}\mathbf{C}\right\| \\ =&o_P\left(\left\|\operatorname{tr}(\mathbf{\Sigma})^{-1}\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{Y}\mathbf{Y}^\top\mathbf{\Theta}\mathbf{C}\right\|\right).\end{aligned}$$

Note that the columns of $\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{Y} = \mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}$ are iid $\mathcal{N}_{k-1}(\mathbf{0}_{k-1},\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Sigma}\mathbf{\Theta}\mathbf{C})$ random vectors. Hence we can write $\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{Y} = (\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Sigma}\mathbf{\Theta}\mathbf{C})^{1/2}\mathbf{Z}^*$, where $\mathbf{Z}^*$ is a $(k-1)\times n$ random matrix with iid $\mathcal{N}(0,1)$ entries. Then

$$\begin{aligned}&\left\|\operatorname{tr}(\mathbf{\Sigma})^{-1}\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{Y}\mathbf{Y}^\top\mathbf{\Theta}\mathbf{C}-\frac{n}{\operatorname{tr}(\mathbf{\Sigma})}\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Sigma}\mathbf{\Theta}\mathbf{C}\right\| \\ \le&\frac{n}{\operatorname{tr}(\mathbf{\Sigma})}\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Sigma}\mathbf{\Theta}\mathbf{C}\right\|\left\|n^{-1}\mathbf{Z}^*\mathbf{Z}^{*\top}-\mathbf{I}_{k-1}\right\| \\ =&o_P\left(\frac{n}{\operatorname{tr}(\mathbf{\Sigma})}\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Sigma}\mathbf{\Theta}\mathbf{C}\right\|\right),\end{aligned}$$

where the last equality follows from the law of large numbers. Combine the above arguments, we have

$$\begin{aligned}\left\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Theta}\mathbf{C}-\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right\| =& (1+o_P(1))\frac{n}{\operatorname{tr}(\mathbf{\Sigma})}\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Sigma}\mathbf{\Theta}\mathbf{C}\right\| \\ \le& (1+o_P(1))\frac{n\boldsymbol{\lambda}_1}{\operatorname{tr}(\mathbf{\Sigma})}\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right\| \\ =&o_P\left(\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}\right).\end{aligned} \tag{C.33}$$

Now we deal with the cross term of (C.27). Note that

$$\begin{aligned}\mathrm{E}[\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger\|_F^2|\mathbf{Y}] =&(k-1)\operatorname{tr}\left(\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{\Theta}\mathbf{C}\right)\\ \leq&(k-1)\boldsymbol{\lambda}_1\operatorname{tr}\left(\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right).\end{aligned}$$

Therefore,

$$\begin{aligned}\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger\| =&O_P\left(\sqrt{\boldsymbol{\lambda}_1\operatorname{tr}\left(\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right)}\right)\\ =&o_P\left(\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}\right),\end{aligned}\tag{C.34}$$

where the last equality follows from the conditions $\boldsymbol{\lambda}_1/\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}\to 0$ and $\operatorname{tr}\left(\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right)\leq(k-1)\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right\|=O(\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)})$.

It follows from (C.33), (C.34) and Weyl's inequality that

$$\begin{aligned}&\left|T(\mathbf{X})-\left(\lambda_1\left(\mathbf{Z}^{\dagger\top}\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger+\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right)\right)\right|\\ \leq&\left\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{\Theta}\mathbf{C}-\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}+\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger+\mathbf{Z}^{\dagger\top}\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{\Theta}\mathbf{C}\right\|\\ \leq&\left\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{\Theta}\mathbf{C}-\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right\|+2\left\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger\right\|\\ =&o_P\left(\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}\right).\end{aligned}$$

But (C.32) implies that

$$\begin{aligned}&\frac{\lambda_1\left(\mathbf{Z}^{\dagger\top}\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger+\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right)-(\operatorname{tr}(\mathbf{\Sigma})-n\operatorname{tr}(\mathbf{\Sigma}^2)/\operatorname{tr}(\mathbf{\Sigma}))}{\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}}\\ =&\lambda_1\left(\frac{\mathbf{Z}^{\dagger\top}\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger-(\operatorname{tr}(\mathbf{\Sigma})-n\operatorname{tr}(\mathbf{\Sigma}^2)/\operatorname{tr}(\mathbf{\Sigma}))\,\mathbf{I}_{k-1}}{\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}}+\frac{\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}}{\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}}\right)\\ \sim&\lambda_1\left(\mathbf{W}_{k-1}+\frac{\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}}{\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}}\right)+o_P(1).\end{aligned}$$

This completes the proof.

□

Proof of Corollary 1. It is straightforward to show that $\operatorname{E} \widehat{\operatorname{tr}(\mathbf{\Sigma})} = \operatorname{tr}(\mathbf{\Sigma})$ and $\operatorname{Var}\left(\widehat{\operatorname{tr}(\mathbf{\Sigma})}\right) = 2n^{-1}\operatorname{tr}(\mathbf{\Sigma}^2)$. Then $\widehat{\operatorname{tr}(\mathbf{\Sigma})} = \operatorname{tr}(\mathbf{\Sigma}) + O_P(\sqrt{n^{-1}\operatorname{tr}(\mathbf{\Sigma}^2)})$. Let $Z_1, \ldots, Z_n$ be the columns of $\mathbf{Z}$. Then we have

$$\begin{aligned}\widehat{\operatorname{tr}(\mathbf{\Sigma}^2)} =& n^{-2}\operatorname{tr}(\mathbf{Z}^\top \mathbf{\Lambda} \mathbf{Z} - n^{-1}\operatorname{tr}(\mathbf{Z}^\top \mathbf{\Lambda}\mathbf{Z})\mathbf{I}_n)^2 \\ =& n^{-2}\sum_{i=1}^n (Z_i^\top \mathbf{\Lambda} Z_i - n^{-1}\sum_{i=1}^n Z_i^\top \mathbf{\Lambda} Z_i)^2 + 2n^{-2}\sum_{1\leq i<j\leq n}(Z_i^\top \mathbf{\Lambda} Z_i)^2.\end{aligned}$$

It can be seen that $n^{-2}\sum_{i=1}^n (Z_i^\top \mathbf{\Lambda} Z_i - n^{-1}\sum_{i=1}^n Z_i^\top \mathbf{\Lambda} Z_i)^2 = O_P(n^{-1}\operatorname{tr}(\mathbf{\Sigma}^2))$. On the other hand, we have $\operatorname{E} 2\sum_{1\leq i<j\leq n}(Z_i^\top \mathbf{\Lambda} Z_i)^2 = n(n-1)\operatorname{tr}(\mathbf{\Sigma}^2)$. Furthermore, Chen et al. (2010), Proposition A.2 implies that

$$\operatorname{Var}\left(2\sum_{1\leq i<j\leq n}(Z_i^\top \mathbf{\Lambda} Z_j)^2\right) = O\left(n^2\operatorname{tr}^2(\mathbf{\Sigma}^2) + n^3\operatorname{tr}(\mathbf{\Sigma}^4)\right) = O\left(n^3\operatorname{tr}^2(\mathbf{\Sigma}^2)\right).$$

Hence $\widehat{\operatorname{tr}(\mathbf{\Sigma}^2)} = (1 + O_P(n^{-1/2}))\operatorname{tr}(\mathbf{\Sigma}^2)$.

Thus, we have

$$\begin{aligned}&\widehat{\operatorname{tr}(\mathbf{\Sigma})} - n\widehat{\operatorname{tr}(\mathbf{\Sigma}^2)}/\widehat{\operatorname{tr}(\mathbf{\Sigma})} \\ =& \operatorname{tr}(\mathbf{\Sigma}) + O_P(\sqrt{n^{-1}\operatorname{tr}(\mathbf{\Sigma}^2)}) - \frac{n\operatorname{tr}(\mathbf{\Sigma}^2)(1+O_P(n^{-1/2}))}{\operatorname{tr}(\mathbf{\Sigma})(1+O_P(\sqrt{n^{-1}\operatorname{tr}(\mathbf{\Sigma}^2)/\operatorname{tr}^2(\mathbf{\Sigma})}))} \\ =& \operatorname{tr}(\mathbf{\Sigma}) + O_P(\sqrt{n^{-1}\operatorname{tr}(\mathbf{\Sigma}^2)}) - \frac{n\operatorname{tr}(\mathbf{\Sigma}^2)}{\operatorname{tr}(\mathbf{\Sigma})}\left(1 + O_P\left(\frac{1}{\sqrt{n}} + \sqrt{\frac{\operatorname{tr}(\mathbf{\Sigma}^2)}{n\operatorname{tr}^2(\mathbf{\Sigma})}}\right)\right) \\ =& \operatorname{tr}(\mathbf{\Sigma}) - \frac{n\operatorname{tr}(\mathbf{\Sigma}^2)}{\operatorname{tr}(\mathbf{\Sigma})} + o_P(\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}).\end{aligned}$$

Therefore,

$$Q_1 = \frac{T(\mathbf{X}) - (\operatorname{tr}(\mathbf{\Sigma}) - n\operatorname{tr}(\mathbf{\Sigma}^2)/\operatorname{tr}(\mathbf{\Sigma}))}{\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}} + o_P(1).$$

Then the conclusion follows from Theorem 1. $\square$

Lemma 7. *Suppose that $r = o(n)$, $\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1/(n\boldsymbol{\lambda}_r^2) \to 0$, $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2) \to 0$. Then uniformly for $i = 1, \ldots, r$,*

$$\lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) = n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\left(1 + O_P\left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}} + \sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\mathbf{\Lambda}_2)}} + \sqrt{\frac{r}{n}}\right)\right).$$

Proof. Note that

$$(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}}) = (\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},2})(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},2}). \tag{C.35}$$

We first deal with $(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})$. Under the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2) \to 0$, Proposition 3 implies that

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

From the decomposition

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

we have

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

Note that

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

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

$$
\left\|(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1})-(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\right\| = O_P\left(\frac{\operatorname{tr}^{3/2}(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1^{1/2}}{n^{3/2}\boldsymbol{\lambda}_r}\right). \tag{C.36}
$$

From the decomposition

$$
\begin{aligned}
&(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger}) \\
=&\mathbf{U}_2\mathbf{Q}\boldsymbol{\Lambda}_1\mathbf{Q}^\top\mathbf{U}_2^\top+\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^\top-\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{Q}\mathbf{U}_1^\top-\mathbf{U}_1\mathbf{Q}^\top\boldsymbol{\Lambda}_2\mathbf{U}_2^\top+\mathbf{U}_1\mathbf{Q}^\top\boldsymbol{\Lambda}_2\mathbf{Q}\mathbf{U}_1^\top,
\end{aligned}
$$

we have

$$
\begin{aligned}
\left\|(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1}^{\dagger})-\mathbf{U}_2\mathbf{Q}\boldsymbol{\Lambda}_1\mathbf{Q}^\top\mathbf{U}_2^\top\right\| \le& \boldsymbol{\lambda}_{r+1}(1+2\|\mathbf{Q}^\top\mathbf{Q}\|^{1/2}+\|\mathbf{Q}^\top\mathbf{Q}\|) \\
=& O_P\left(\boldsymbol{\lambda}_{r+1}\right),
\end{aligned} \tag{C.37}
$$

where the last equality follows from (B.19). Note that $\mathbf{U}_2\mathbf{Q}\boldsymbol{\Lambda}_1\mathbf{Q}^\top\mathbf{U}_2^\top =$

$\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top$. We have

$$\begin{aligned}
&\left\|\mathbf{U}_2\mathbf{Q}\mathbf{\Lambda}_1\mathbf{Q}^\top\mathbf{U}_2^\top - n^{-1}\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top\right\| \\
\leq &\left\|\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top\right\| \left\|(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1} - n^{-1}\mathbf{I}_r\right\| \\
\leq &\left\|\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right\| \left\|(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}\right\| \left\|n^{-1}\mathbf{Z}_1\mathbf{Z}_1^\top - \mathbf{I}_r\right\| \\
= &O_P\left(\frac{r^{1/2}\,\mathrm{tr}(\mathbf{\Lambda}_2)}{n^{3/2}}\right),
\end{aligned} \tag{C.38}$$

where the last equality follows from (B.11) and (B.17). From (C.35), (C.36), (C.37) and (C.38), we obtain that

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

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

$$\begin{aligned}
\lambda_i\left((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\right) = &n^{-1}\lambda_i\left(\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top(\mathbf{I} - \mathbf{P}_{\mathbf{Y},2})\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right) \\
&+ O_P\left(\left(\sqrt{\frac{\mathrm{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}} + \frac{n\boldsymbol{\lambda}_{r+1}}{\mathrm{tr}(\mathbf{\Lambda}_2)} + \sqrt{\frac{r}{n}}\right)\frac{\mathrm{tr}(\mathbf{\Lambda}_2)}{n}\right).
\end{aligned}$$

Note that

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

where the last equality follows from (B.17) and Proposition 4. Then it follows from Weyl's inequality that uniformly for $i=1,\ldots,r$,

$$
\begin{aligned}
&\lambda_i\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right) \\
=&n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)-(n\operatorname{tr}(\boldsymbol{\Lambda}_2))^{-1}\,\lambda_{r+1-i}\left(\mathbf{V}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right) \\
&+O_P\left(\left(\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}}+\sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}}+\sqrt{\frac{r}{n}}\right)\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)}{n}\right).
\end{aligned}
\tag{C.39}
$$

Now we deal with the matrix $\mathbf{V}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}$. 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. Then Lemma 5 implies that

$$
\begin{aligned}
&\Big\|\mathbf{V}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}-\operatorname{tr}(\boldsymbol{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2)\mathbf{I}_r\Big\| \\
=&O_P\left(\sqrt{r\operatorname{tr}(\boldsymbol{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2)^2}+r\Big\|\boldsymbol{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\Big\|\right).
\end{aligned}
$$

By some algebra, we have

$$
\begin{aligned}
&\left\|\mathbf{V}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \boldsymbol{\Lambda}_2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1} \tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \boldsymbol{\Lambda}_2 \mathbf{Z}_2 \mathbf{V}_{\mathbf{Z}_1} - \operatorname{tr}(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \boldsymbol{\Lambda}_2^2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}) \mathbf{I}_r\right\| \\
=&O_P\left(\sqrt{r\left\|\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \boldsymbol{\Lambda}_2^2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}\right\| \operatorname{tr}(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \boldsymbol{\Lambda}_2^2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1})} + r\left\|\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \boldsymbol{\Lambda}_2^2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}\right\|\right).
\end{aligned}
$$

Since $\operatorname{E}\operatorname{tr}(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \boldsymbol{\Lambda}_2^2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}) = (n-r)\operatorname{tr}(\boldsymbol{\Lambda}_2^2)$, we have

$$
\operatorname{tr}(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \boldsymbol{\Lambda}_2^2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}) = O_P\left(n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)\right) = O_P\left(n\boldsymbol{\lambda}_{r+1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\right).
$$

On the other hand, Lemma 5 implies that

$$
\|\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \boldsymbol{\Lambda}_2^2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}\| = O_P\left(\operatorname{tr}(\boldsymbol{\Lambda}_2^2) + n\boldsymbol{\lambda}_{r+1}^2\right) = O_P\left(\boldsymbol{\lambda}_{r+1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\right).
$$

Combine these bounds, we have

$$
\left\|\mathbf{V}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \boldsymbol{\Lambda}_2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1} \tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \boldsymbol{\Lambda}_2 \mathbf{Z}_2 \mathbf{V}_{\mathbf{Z}_1} - n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)\mathbf{I}_r\right\| = O_P\left(\sqrt{rn}\boldsymbol{\lambda}_{r+1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\right).
$$

The last display, combined with Weyl's inequality, implies that uniformly for $i = 1, \ldots, r$,

$$
(n\operatorname{tr}(\boldsymbol{\Lambda}_2))^{-1}\lambda_i\left(\mathbf{V}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \boldsymbol{\Lambda}_2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1} \tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \boldsymbol{\Lambda}_2 \mathbf{Z}_2 \mathbf{V}_{\mathbf{Z}_1}\right) = O_P(\boldsymbol{\lambda}_{r+1}).
$$

Then (C.39) and the last display implies that uniformly for $i = 1, \ldots, r$,

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

This completes the proof.

□

Lemma 8. *Suppose that* $r = o(n)$, $\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1/(n\boldsymbol{\lambda}_r^2) \to 0$, $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2) \to 0$. *Then*

$$\begin{aligned}&\sum_{i=r+1}^{p} \lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)\\ =&\operatorname{tr}(\mathbf{\Lambda}_2) - \frac{n\operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Lambda}_2)} + O_P\left(n(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}} + \sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\mathbf{\Lambda}_2)}}\right) + r\boldsymbol{\lambda}_{r+1}\right).\end{aligned}$$

Proof. Write $\mathbf{\Sigma} = \mathbf{U}_1\mathbf{\Lambda}_1\mathbf{U}_1^\top + \mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top$. Note that $\mathbf{U}_1\mathbf{\Lambda}_1\mathbf{U}_1^\top$ is of rank r. Then Weyl's inequality implies that for $i = r+1, \ldots, p$,

$$\lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) \geq \lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right), \tag{C.40}$$

$$\lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) \leq \lambda_{i-r}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right). \tag{C.41}$$

Hence we have

$$\begin{aligned}&\left|\sum_{i=r+1}^{p} \lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) - \operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)\right|\\ \leq& r\lambda_1\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)\\ \leq& r\boldsymbol{\lambda}_{r+1}.\end{aligned} \tag{C.42}$$

Write

$$\begin{aligned}&\operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)\\ =&\operatorname{tr}\left(\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\right)\\ =&\operatorname{tr}(\mathbf{\Lambda}_2) - \operatorname{tr}\left(\left(\mathbf{\Lambda}_2 - \frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^\top\mathbf{P}_{\mathbf{Y}}\mathbf{U}_2\right) - \frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Lambda}_2)}\operatorname{tr}\left(\mathbf{U}_2^\top\mathbf{P}_{\mathbf{Y}}\mathbf{U}_2\right).\end{aligned} \tag{C.43}$$

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

$$\left|\operatorname{tr}\left(\mathbf{U}_2^\top \mathbf{P}_\mathbf{Y} \mathbf{U}_2\right) - n\right| \leq r. \tag{C.44}$$

Next we deal with the second term. We have

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

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

$$\begin{aligned}
&\left|\operatorname{tr}\left(\left(\boldsymbol{\Lambda}_2 - \frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^\top\left(\mathbf{P}_\mathbf{Y} - \mathbf{P}^\dagger_{\mathbf{Y},1} - \mathbf{P}^\dagger_{\mathbf{Y},2}\right)\mathbf{U}_2\right)\right| \\
\leq& 2n(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\left\|\mathbf{P}_\mathbf{Y} - \mathbf{P}^\dagger_{\mathbf{Y},1} - \mathbf{P}^\dagger_{\mathbf{Y},2}\right\| \\
\leq& 2n(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\left(\left\|\mathbf{P}_{\mathbf{Y},1} - \mathbf{P}^\dagger_{\mathbf{Y},1}\right\| + \left\|\mathbf{P}_{\mathbf{Y},2} - \mathbf{P}^\dagger_{\mathbf{Y},2}\right\|\right) \\
=& O_P\left(n(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\left(\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}} + \sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}}\right)\right),
\end{aligned} \tag{C.45}$$

where the last equality follows from Proposition 3 and Proposition 4. Note that

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

It is straightforward to show that

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

and

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

Thus,

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

The last display, combined with (C.45), leads to

$$\mathrm{tr}\left(\left(\mathbf{\Lambda}_2 - \frac{\mathrm{tr}(\mathbf{\Lambda}_2^2)}{\mathrm{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^\top\mathbf{P}_{\mathbf{Y}}\mathbf{U}_2\right) = O_P\left(n(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\left(\sqrt{\frac{\mathrm{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}} + \sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\mathrm{tr}(\mathbf{\Lambda}_2)}}\right)\right).$$

It then follows from (C.43), (C.44) and the last display that

$$\begin{aligned}&\mathrm{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) \\ =&\,\mathrm{tr}(\mathbf{\Lambda}_2) - \frac{n\,\mathrm{tr}(\mathbf{\Lambda}_2^2)}{\mathrm{tr}(\mathbf{\Lambda}_2)} + O_P\left(n(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p)\left(\sqrt{\frac{\mathrm{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}} + \sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\mathrm{tr}(\mathbf{\Lambda}_2)}}\right) + r\boldsymbol{\lambda}_{r+1}\right).\end{aligned}$$

Then the conclusion follows from (C.42) and the last display. □

Lemma 9. *Suppose $p > n$, we have*

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

Proof. From (C.40) and (C.41), we have

$$
\begin{aligned}
&\left|\sum_{i=r+1}^{p} \lambda_i^2\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)-\operatorname{tr}\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)^2\right| \\
\leq & r\lambda_1^2\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right) \\
\leq & r\boldsymbol{\lambda}_{r+1}^2.
\end{aligned}
\tag{C.46}
$$

It is straightforward to show that

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

For the second term, we have

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

where the last equality follows from von Neumann's trace theorem. The last display, combined with (C.44), implies that

$$
\operatorname{tr}(\mathbf{\Lambda}_2^2\mathbf{U}_2^\top\mathbf{P}_{\mathbf{Y}}\mathbf{U}_2)=\frac{n\operatorname{tr}^2(\mathbf{\Lambda}_2^2)}{\operatorname{tr}^2(\mathbf{\Lambda}_2)}+O_P\left(n\boldsymbol{\lambda}_{r+1}(\boldsymbol{\lambda}_{r+1}-\boldsymbol{\lambda}_p)+r\boldsymbol{\lambda}_{r+1}^2\right).
$$

For the third term, von Neumann's trace theorem implies that

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

Note that

$$\begin{aligned}\operatorname{tr}(\mathbf{U}_2^\top \mathbf{P}_\mathbf{Y}\mathbf{U}_2)^2 =& \operatorname{tr}\left(\mathbf{P}_\mathbf{Y} - \mathbf{P}_\mathbf{Y}\mathbf{U}_1\mathbf{U}_1^\top\right)^2\\ =& n - 2\operatorname{tr}(\mathbf{P}_\mathbf{Y}\mathbf{U}_1\mathbf{U}_1^\top) + \operatorname{tr}(\mathbf{P}_\mathbf{Y}\mathbf{U}_1\mathbf{U}_1^\top)^2\\ =& n + O_P(r).\end{aligned}$$

Therefore, the third term satisfies

$$\operatorname{tr}(\mathbf{\Lambda}_2\mathbf{U}_2^\top \mathbf{P}_\mathbf{Y}\mathbf{U}_2)^2 = \frac{n\operatorname{tr}^2(\mathbf{\Lambda}_2^2)}{\operatorname{tr}^2(\mathbf{\Lambda}_2)} + O_P\left(n\boldsymbol{\lambda}_{r+1}(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p) + r\boldsymbol{\lambda}_{r+1}^2\right).$$

Thus,

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

Then the conclusion follows from the last display and (C.46). □

Proof of Theorem 2. We have

$$\begin{aligned}&\mathbf{Z}^{\dagger\top}\mathbf{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger\\ =&\sum_{i=1}^r \lambda_i((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}))\eta_i\eta_i^\top + \sum_{i=r+1}^p \lambda_i((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}))\eta_i\eta_i^\top.\end{aligned}$$

From Lemma 7, the first term satisfies

$$\sum_{i=1}^r \lambda_i((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}))\eta_i\eta_i^\top = (1 + o_P(r^{-1/2}))n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\sum_{i=1}^r \eta_i\eta_i^\top.$$

Then

$$\frac{\sum_{i=1}^r \lambda_i((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\mathbf{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}))\eta_i\eta_i^\top - rn^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_{k-1}}{\sqrt{r}n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)} = \frac{\sum_{i=1}^r \eta_i\eta_i^\top - r\mathbf{I}_{k-1}}{\sqrt{r}} + o_P(1). \tag{C.47}$$

Next we deal with the term $\sum_{i=r+1}^{p}\lambda_i((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}}))\eta_i\eta_i^\top$. In the current context, Lemma 8 and Lemma 9 imply that

$$\sum_{i=r+1}^{p}\lambda_i\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)=\operatorname{tr}(\mathbf{\Lambda}_2)-\frac{n\operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Lambda}_2)}+o_P\left(\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)}\right), \tag{C.48}$$

$$\sum_{i=r+1}^{p}\lambda_i^2\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)=(1+o_P(1))\operatorname{tr}(\mathbf{\Lambda}_2^2). \tag{C.49}$$

By Weyl's inequality, we have

$$\begin{aligned}
&\lambda_{r+1}((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}}))\\
=&\lambda_{r+1}\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}_1\mathbf{\Lambda}_1\mathbf{U}_1^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})+(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)\\
\le&\lambda_1\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)\\
\le&\boldsymbol{\lambda}_{r+1}.
\end{aligned}$$

The last display and (C.48) imply that

$$\frac{\lambda_{r+1}^2\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)}{\sum_{i=r+1}^{p}\lambda_i^2\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)}\le\frac{\boldsymbol{\lambda}_{r+1}^2}{(1+o_P(1))\operatorname{tr}(\mathbf{\Lambda}_2^2)}\xrightarrow{P}0.$$

Then Lemma 2 implies that

$$\frac{\sum_{i=r+1}^{p}\lambda_i((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}}))\eta_i\eta_i^\top-\sum_{i=r+1}^{p}\lambda_i((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}}))\mathbf{I}_{k-1}}{\sqrt{\sum_{i=r+1}^{p}\lambda_i^2((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}}))}}\xrightarrow{\mathcal{L}}\mathbf{W}_{k-1}.$$

The last display, combined with (C.48) and (C.49), leads to

$$\frac{\sum_{i=r+1}^{p}\lambda_i((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}}))\eta_i\eta_i^\top-\left(\operatorname{tr}(\mathbf{\Lambda}_2)-n\operatorname{tr}(\mathbf{\Lambda}_2^2)/\operatorname{tr}(\mathbf{\Lambda}_2)\right)\mathbf{I}_{k-1}}{\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)}}\xrightarrow{\mathcal{L}}\mathbf{W}_{k-1}. \tag{C.50}$$

Note that $\sum_{i=1}^{r}\eta_i\eta_i^\top$ is independent of $\sum_{i=r+1}^{p}\lambda_i((\mathbf{I}_p-\mathbf{P_Y})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P_Y}))\eta_i\eta_i^\top$. Then(C.47) and (C.50) implies that

$$\frac{\mathbf{Z}^{\dagger\top}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p-\mathbf{P_Y})\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{Z}^\dagger-((1+r/n)\operatorname{tr}(\boldsymbol{\Lambda}_2)-n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)/\operatorname{tr}(\boldsymbol{\Lambda}_2))\,\mathbf{I}_{k-1}}{\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2)+\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}$$
$$\sim\frac{n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)}{\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2)+\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}(\mathbf{W}_{k-1}^*-r\mathbf{I}_{k-1})+\frac{\sqrt{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}{\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2)+\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}\mathbf{W}_{k-1}+o_P(1). \tag{C.51}$$

This completes the proof of the first statement.

Now we prove the second statement. For the second term of (C.27), we have $\mathbf{C}^\top\boldsymbol{\Theta}^\top(\mathbf{I}_p-\mathbf{P_Y})\boldsymbol{\Theta}\mathbf{C}=\mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\mathbf{C}-\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{P_Y}\boldsymbol{\Theta}\mathbf{C}$. We need to deal with $\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{P_Y}\boldsymbol{\Theta}\mathbf{C}$. Note that Proposition 3 implies that

$$\left\|\mathbf{P}_{\mathbf{Y},1}-\mathbf{U}_1\mathbf{U}_1^\top\right\|\le\left\|\mathbf{P}_{\mathbf{Y},1}-\mathbf{P}_{\mathbf{Y},1}^\dagger\right\|+2\left\|\mathbf{Q}\right\|=o_P(1).$$

It follows from the last display and Proposition 4 that

$$\begin{aligned}
&\left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{P_Y}\boldsymbol{\Theta}\mathbf{C}-\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{U}_1\mathbf{U}_1^\top\boldsymbol{\Theta}\mathbf{C}-\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{P}_{\mathbf{Y},2}^\dagger\boldsymbol{\Theta}\mathbf{C}\right\|\\
\le&\left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{P}_{\mathbf{Y},1}\boldsymbol{\Theta}\mathbf{C}-\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{U}_1\mathbf{U}_1^\top\boldsymbol{\Theta}\mathbf{C}\right\|+\left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{P}_{\mathbf{Y},2}\boldsymbol{\Theta}\mathbf{C}-\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{P}_{\mathbf{Y},2}^\dagger\boldsymbol{\Theta}\mathbf{C}\right\|\\
\le&\left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\mathbf{C}\right\|\left(\left\|\mathbf{P}_{\mathbf{Y},1}-\mathbf{U}_1\mathbf{U}_1^\top\right\|+\left\|\mathbf{P}_{\mathbf{Y},2}-\mathbf{P}_{\mathbf{Y},2}^\dagger\right\|\right)\\
=&o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2)+\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}\right).
\end{aligned}$$

We have

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

Note that $\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}$ is a $(p-r)\times(n-r)$ matrix with iid $\mathcal{N}(0,1)$ entries. Then the columns of $\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}$ are iid $\mathcal{N}_{k-1}(\mathbf{0}_{k-1},\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C})$ random vectors. Write $\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}=(\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C})^{1/2}\mathbf{Z}^*$, where $\mathbf{Z}^*$ is a $(k-1)\times(n-r)$ random matrix with iid $\mathcal{N}(0,1)$ entries. Then

$$\begin{aligned}
&\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{P}_{\mathbf{Y},2}^\dagger\mathbf{\Theta}\mathbf{C}-\frac{n}{\operatorname{tr}(\mathbf{\Lambda}_2)}\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C}\right\|\\
\leq&\frac{n}{\operatorname{tr}(\mathbf{\Lambda}_2)}\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C}\right\|\left\|n^{-1}\mathbf{Z}^*\mathbf{Z}^{*\top}-\mathbf{I}_{k-1}\right\|\\
=&o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}\right),
\end{aligned}$$

where the last equality follows from the law of large numbers, the local alternative condition and the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2)\to 0$. But

$$\frac{n}{\operatorname{tr}(\mathbf{\Lambda}_2)}\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C}\right\|\leq\frac{n\boldsymbol{\lambda}_2}{\operatorname{tr}(\mathbf{\Lambda}_2)}\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right\|=o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}\right).$$

Hence $\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{P}_{\mathbf{Y},2}^\dagger\mathbf{\Theta}\mathbf{C}\right\|=o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}\right)$. Consequently, $\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{P}_{\mathbf{Y}}\mathbf{\Theta}\mathbf{C}-\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_1\mathbf{U}_1^\top\mathbf{\Theta}\mathbf{C}\right\|=o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}\right)$. Thus, the second term of (C.27) satisfies

$$\left\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Theta}\mathbf{C}-\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C}\right\|=o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}\right). \tag{C.52}$$

Next we consider the cross term of (C.27). Note that

$$
\begin{aligned}
&\mathrm{E}[\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger\|_F^2|\mathbf{Y}]\\
=&(k-1)\operatorname{tr}(\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Theta}\mathbf{C})\\
\le&(k-1)\lambda_1\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)\operatorname{tr}(\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C})\\
=&O_P\left(n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right\|\right),
\end{aligned}
$$

where the last equality follows from Lemma 7. Under the condition $r\to\infty$ or $\operatorname{tr}(\mathbf{\Lambda}_2)/(n\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)})\to 0$, we have $n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)=o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}\right)$. Therefore,

$$
\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger\|=o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}\right).
$$

It follows from the last display, (C.52) and Weyl's inequality that

$$
\begin{aligned}
&\left|T(\mathbf{X})-\lambda_1\left(\mathbf{Z}^{\dagger\top}\mathbf{\Lambda}^{1/2}\mathbf{U}^\dagger(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger+\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C}\right)\right|\\
=&o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}\right).
\end{aligned}
$$

Then the second statement follows from the last display and (C.51).

□

Proof of Corollary 2. From Proposition 2, we have

$$
rn^{-2}(\widehat{\operatorname{tr}(\mathbf{\Lambda}_2)})^2+\widehat{\operatorname{tr}(\mathbf{\Lambda}_2^2)}=(1+o_P(1))(rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)),
$$

and

$$
\begin{aligned}
&(1+r/n)\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)} - n\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}/\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)} \\
=&(1+r/n)\operatorname{tr}(\boldsymbol{\Lambda}_2) + O_P\left(r\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{n}} + r\boldsymbol{\lambda}_{r+1}\right) \\
&- \frac{n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)\left(1+O_P\left(r/n + r\boldsymbol{\lambda}_{r+1}^2/\operatorname{tr}(\boldsymbol{\Lambda}_2^2)\right)\right)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)\left(1+O_P\left(r\sqrt{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)/n\operatorname{tr}^2(\boldsymbol{\Lambda}_2)} + r\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2)\right)\right)} \\
=&(1+r/n)\operatorname{tr}(\boldsymbol{\Lambda}_2) + O_P\left(r\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{n}} + r\boldsymbol{\lambda}_{r+1}\right) \\
&- \frac{n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\left(1+O_P\left(\frac{r}{n} + \frac{r\boldsymbol{\lambda}_{r+1}^2}{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)} + r\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{n\operatorname{tr}^2(\boldsymbol{\Lambda}_2)}} + \frac{r\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\right)\right) \\
=&(1+r/n)\operatorname{tr}(\boldsymbol{\Lambda}_2) - \frac{n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}(\boldsymbol{\Lambda}_2)} + o_P\left(\sqrt{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}\right).
\end{aligned}
$$

Therefore,

$$
Q_2 = \frac{T(\mathbf{X}) - ((1+r/n)\operatorname{tr}(\boldsymbol{\Lambda}_2) - n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)/\operatorname{tr}(\boldsymbol{\Lambda}_2))}{\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2) + \operatorname{tr}(\boldsymbol{\Lambda}_2^2)}} + o_P(1).
$$

On the other hand, it is not hard to see that the ratio consistency of $\widehat{\operatorname{tr}(\boldsymbol{\lambda}_2)}$ and $\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}$ imply $F_2^{-1}(1-\alpha;\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)},\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}) = F_2^{-1}(1-\alpha;\operatorname{tr}(\boldsymbol{\Lambda}_2),\operatorname{tr}(\boldsymbol{\Lambda}_2^2)) + o_P(1)$. Then the conclusion follows from Theorem 2 and Slutsky's theorem.

□

Proof of Proposition 5. Under the conditions of Theorem 1, we have $n\boldsymbol{\lambda}_1/\operatorname{tr}(\boldsymbol{\Sigma}) \to 0$. From Lemma 5 and Weyl's inequality, we have

$$
\lambda_1(\hat{\boldsymbol{\Sigma}}) = n^{-1}\lambda_1(\mathbf{Z}^\top\boldsymbol{\Lambda}\mathbf{Z}) = n^{-1}\operatorname{tr}(\boldsymbol{\Sigma}) + O_P\left(\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{n}} + \boldsymbol{\lambda}_1\right) = (1+o_P(1))n^{-1}\operatorname{tr}(\boldsymbol{\Sigma}).
$$

From the proof of Corollary 1, we have $\operatorname{tr}(\hat{\mathbf{\Sigma}}) = (1+o_P(1))\operatorname{tr}(\mathbf{\Sigma})$. Therefore,

$$\frac{n\lambda_1(\hat{\mathbf{\Sigma}})}{\operatorname{tr}(\hat{\mathbf{\Sigma}})} \xrightarrow{P} 1.$$

This completes the proof of (i).

Now we prove (ii). Under the conditions of Theorem 2, Proposition 1 implies that

$$\begin{aligned}
\frac{n\lambda_1(\hat{\mathbf{\Sigma}})}{\operatorname{tr}(\hat{\mathbf{\Sigma}})} =& \frac{n\lambda_1(\hat{\mathbf{\Sigma}})}{\sum_{i=1}^r \lambda_i(\hat{\mathbf{\Sigma}}) + \sum_{i=r+1}^n \lambda_i(\hat{\mathbf{\Sigma}})} \\
=&(1+o_P(1))\frac{n\boldsymbol{\lambda}_1 + \operatorname{tr}(\mathbf{\Lambda}_2)}{\sum_{i=1}^r \boldsymbol{\lambda}_i + \operatorname{tr}(\mathbf{\Lambda}_2)} \\
\geq&(1+o_P(1))\frac{n\boldsymbol{\lambda}_1}{r\boldsymbol{\lambda}_1 + \operatorname{tr}(\mathbf{\Lambda}_2)} \xrightarrow{P} \infty.
\end{aligned}$$

It follows that

$$\Pr\left(\frac{n\lambda_1(\hat{\mathbf{\Sigma}})}{\operatorname{tr}(\hat{\mathbf{\Sigma}})} < \tau\right) \to 0.$$

Next we consider the consistency of $\hat{r}$. Note that

$$\{\hat{r} = r\} = \left\{\frac{n\lambda_{i+1}(\hat{\mathbf{\Sigma}})}{\sum_{j=i+1}^n \lambda_j(\hat{\mathbf{\Sigma}})} \geq \tau, i = 1, \ldots, r-1\right\} \cap \left\{\frac{n\lambda_{r+1}(\hat{\mathbf{\Sigma}})}{\sum_{j=r+1}^n \lambda_j(\hat{\mathbf{\Sigma}})} < \tau\right\}.$$

But Proposition 1 implies that uniformly for $i = 1, \ldots, r-1$,

$$\begin{aligned}
\frac{n\lambda_{i+1}(\hat{\mathbf{\Sigma}})}{\sum_{j=i+1}^n \lambda_j(\hat{\mathbf{\Sigma}})} \geq& \frac{n\lambda_{i+1}(\hat{\mathbf{\Sigma}})}{(r-i)\lambda_{i+1}(\hat{\mathbf{\Sigma}}) + \sum_{j=r+1}^n \lambda_j(\hat{\mathbf{\Sigma}})} \\
=&(1+o_P(1))\frac{n\boldsymbol{\lambda}_{i+1} + \operatorname{tr}(\mathbf{\Lambda}_2)}{(r-i)\boldsymbol{\lambda}_{i+1} + (1-i/n)\operatorname{tr}(\mathbf{\Lambda}_2)} \xrightarrow{P} \infty.
\end{aligned}$$

Thus, we only need to prove that

$$\Pr\left(\frac{n\lambda_{r+1}(\hat{\mathbf{\Sigma}})}{\sum_{j=r+1}^n \lambda_j(\hat{\mathbf{\Sigma}})} < \tau\right) \to 1.$$

Weyl' inequality implies that $n\lambda_{r+1}(\hat{\mathbf{\Sigma}}) = \lambda_{r+1}(\mathbf{Z}_1^\top \mathbf{\Lambda}_1 \mathbf{Z}_1 + \mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2) \leq \lambda_1(\mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2)$. Then using Lemma 5, we have $n\lambda_{r+1}(\hat{\mathbf{\Sigma}}) \leq (1+o_P(1)) \operatorname{tr}(\mathbf{\Lambda}_2)$. Thus,

$$\frac{n\lambda_{r+1}(\hat{\mathbf{\Sigma}})}{\sum_{j=r+1}^n \lambda_j(\hat{\mathbf{\Sigma}})} \leq (1 + o_P(1)).$$

This completes the proof.

$\square$

References

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

Aoshima, M. and K. Yata (2018). Two-sample tests for high-dimension, strongly spiked eigenvalue models. *Statistica Sinica 28*(1), 43–62.

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., X. Han, and G. Pan (2017). Limiting Laws for Divergent Spiked Eigenvalues and Largest Non-spiked Eigenvalue of Sample Covariance Matrices. *ArXiv e-prints*.

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

Cao, M., J. Park, and D. He (2017). A test for k sample Behrens-Fisher problem in high dimensional data. *ArXiv e-prints*.

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.

Chen, S. X., L. Zhang, and P. Zhong (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association 105*(490), 810–819.

Fan, J., D. Wang, K. Wang, and Z. Zhu (2017). Distributed Estimation of Principal Eigenspaces. *ArXiv e-prints*.

Fan, J., L. Yuan, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*(4), 603–680.

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.

Hu, J., Z. Bai, C. Wang, and W. Wang (2017). On testing the equality of high dimensional mean vectors with unequal covariance matrices. *Annals of the Institute of Statistical Mathematics 69*(2), 365–387.

Katayama, S., Y. Kano, and M. S. Srivastava (2013). Asymptotic distributions of some test criteria for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis 116*, 410–421.

Koltchinskii, V. and K. Lounici (2016). Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 52*(4), 1976–2013.

Laurent, B. and P. Massart (2000). Adaptive estimation of a quadratic functional by model selection. *The 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.

Nadler, B. (2008). Finite sample approximation results for principal component analysis: A matrix perturbation approach. *The Annals of Statistics 36*(6), 2791–2817.

Roy, S. N. (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 Multivariate Analysis 98*(9), 1825–1839.

Shen, D., H. Shen, and J. S. Marron (2016). A general framework for consistency of principal component analysis. *Journal of Machine Learning Research 17*(150), 1–34.

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.

Wang, R. and X. Xu (2018a). A feasible high dimensional randomization test for the mean vector. *Journal of Statistical Planning and Inference*.

Wang, R. and X. Xu (2018b). On two-sample mean tests under spiked covariances. *Journal of Multivariate Analysis 167*, 225 – 249.

Wang, W. and J. Fan (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *The Annals of Statistics 45*(3), 1342–1374.

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

Yata, K. and M. Aoshima (2012). Effective pca for high-dimension, low-sample-size data with noise reduction via geometric representations. *Journal of Multivariate Analysis 105*(1), 193 – 215.

Yata, K. and M. Aoshima (2013). Pca consistency for the power spiked model in high-dimensional settings. *Journal of Multivariate Analysis 122*, 334 – 354.

Zhang, J.-T., J. Guo, and B. Zhou (2017). Linear hypothesis testing in high-dimensional one-way manova. *Journal of Multivariate Analysis 155*, 200 – 216.

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