Statistica Sinica

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

Rui Wang, Xingzhong Xu

Beijing Institute of Technology

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

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

1. Introduction

Suppose there are k ($k \geq 2$) independent samples of p-dimensional data. Within the ith sample ($1 \leq i \leq k$), the observations $\{X_{ij}\}_{j=1}^{n_i}$ are independent and identically distributed (iid) as $\mathcal{N}_p(\theta_i, \mathbf{\Sigma})$, 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, for example, 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, most tests are designed for two-sample case, i.e., $k = 2$. See, for example, Bai and Saranadasa (1996), Srivastava (2007), Chen and Qin (2010), Cai et al. (2014) and Feng et al. (2015). Recently, some tests have also been introduced for the case of genenral k. Schott (2007) modified Hotelling-Lawley trace and proposed the test statistic

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

Statistic T_{SC} is a representative of the so-called sum-of-squares type statistics as it is based on an estimation of squared Euclidean norm $\sum_{i=1}^{k} n_i\|\theta_i - \bar{\theta}\|^2$, where $\bar{\theta} = N^{-1}\sum_{i=1}^{k} n_i\theta_i$. See Srivastava and Kubokawa (2013), Yamada and Himeno (2015) and Zhou et al. (2017) for some other sum-of-squares type test statistics. In another work, Cai and Xia (2014) proposed a test statistic

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

where $\Omega = (\omega)_{ij} = \mathbf{\Sigma}^{-1}$ is the precision matrix. When Ω is unknown, it is substituted by an estimator. Unlike T_{SC}, T_{CX} is an extreme value type statistic.

The likelihood ratio test (LRT) method has been very successful in leading to satisfactory procedures in many specific problems. However, the LRT statistic for hypotheses (1.1), i.e. Wilks' Lambda statistic, is not defined for $p > N - k$. In high dimensional setting, both sum-of-squares type statistics and extreme value type statistics are not based on likelihood function. This motivates us to construct a likelihood-based test in high dimensional setting. In a recent work, Zhao and Xu (2016) proposed a generalized likelihood ratio test in the context of one-sample mean vector test. They used a least favorable argument to construct a generalized likelihood ratio test statistic. Their simulation results showed that their test has good power performance, especially when the variables are correlated.

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

The rest of the paper is organized as follows. In Section 2, we propose LFD test and give the asymptotic distributions of LFD test. Section 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

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

Define $\mathbf{\Theta} = (\sqrt{n_1}\theta_1, \ldots, \sqrt{n_k}\theta_k)$ and the null hypothesis H_0 is equivalent to $\mathbf{\Theta}\mathbf{C} = \mathbf{O}_{p\times(k-1)}$, where $\mathbf{O}_{p\times(k-1)}$ is a $p \times (k-1)$ matrix with all elements equal to 0. Thus, the hypotheses (1.1) are equivalent to

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

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 $H_{0a} : a^\top \mathbf{\Theta}\mathbf{C} = \mathbf{O}_{1\times(k-1)}$ and $H_{1a} : a^\top \mathbf{\Theta}\mathbf{C} \neq \mathbf{O}_{1\times(k-1)}$. 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 $\max_{a^\top a=1} L_1(a)/L_0(a) = \lambda_1^{n/2}(\mathbf{H}\mathbf{G}^{-1})$, where $\lambda_i(\cdot)$ means the ith largest eigenvalue. This statistic is an increasing function of Roy's maximum root.

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

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

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

Unfortunately, Roy's maximum root can only be defined when $n \geq p$, hence can not be used in the 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 (2.2) is not well defined since the ratio involves infinity. Hence we need a new definition for LFD in the high dimensional setting. Define

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

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,

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

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

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

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 LFD test statistic. Let $\mathbf{Y} = \mathbf{U_Y} \mathbf{D_Y} \mathbf{V_Y}^\top$ be the singular value decomposition of $\mathbf{Y}$, where $\mathbf{U_Y}$ and $\mathbf{V_Y}$ are $p \times \min(n, p)$ and $n \times \min(n, p)$ column orthogonal matrices, respectively, and $\mathbf{D_Y}$ 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

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

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

$$T(\mathbf{X}) = \lambda_1\Big(\mathbf{C}^{\top}\big(\mathbf{J}^{\top}(\mathbf{X}^{\top}\mathbf{X})^{-1}\mathbf{J}\big)^{-1}\mathbf{C}\Big). \tag{2.4}$$

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

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

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

3. Theoretical analysis

We now turn to the analysis of the asymptotic distributions of LFD test statistic. Based on these results, a 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

Most existing high dimensional tests do not allow spiked eigenvalues. In this section, 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 $\mathcal{N}(0,1)$ and the entries on the diagonal are iid $\mathcal{N}(0,2)$.

The following theorem establishes the asymptotic distribution of LFD test statistic.

Theorem 1. *Suppose that $n\boldsymbol{\lambda}_1/\operatorname{tr}(\boldsymbol{\Sigma}) \to 0$, $\boldsymbol{\lambda}_1/\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)} \to 0$ and $\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_p = O(n^{-1}\operatorname{tr}^{1/2}(\boldsymbol{\Sigma}^2))$. 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.

To formulate a test procedure with asymptotically correct level, 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. Consider the following simple estimators,

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

Let

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

Let $F_1(x)$ be the cumulative distribution function of $\lambda_1(\mathbf{W}_{k-1})$. Then we reject the null hypothesis if $Q > F^{-1}(1-\alpha)$. The following corollary shows that this test procedure has asymptotically correct leve and gives the asymptotic local power function.

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 tests. In fact, if $k=2$, the asymptotic local power function given by Corollary 1 equals the asymptotic local power function of the tests in Bai and Saranadasa (1996) and Chen and Qin (2010).

3.2 Spiked covariance

We are especially interested in the case where variables are correlated. For some real world problems, variables are heavily correlated with common factors, then the covariance matrix $\mathbf{\Sigma}$ is spiked in the sense that a few eigenvalues of $\mathbf{\Sigma}$ are significantly larger than the others (Fan et al., 2008; Cai et al., 2013; Shen et al., 2013; Ma et al., 2015). Our results allow $\mathbf{\Sigma}$ to have r spiked eigenvalues, where $1 \leq r \leq p$ can also diverge as $n, p \to \infty$.

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 analysis. Let $\boldsymbol{\lambda}_1 \geq \cdots \geq \boldsymbol{\lambda}_p$ denote the non-increasingly ordered eigenvalues of $\mathbf{\Sigma}$. Let $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ denote the eigenvalue decomposition of $\mathbf{\Sigma}$, where $\mathbf{\Lambda} = \operatorname{diag}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_p)$. 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 $\mathbf{U}$. Denote $\mathbf{\Lambda}_1 = \operatorname{diag}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r)$ and $\mathbf{\Lambda}_2 = \operatorname{diag}(\boldsymbol{\lambda}_{r+1}, \ldots, \boldsymbol{\lambda}_p)$. Then $\mathbf{\Sigma} = \mathbf{U}_1\mathbf{\Lambda}_1\mathbf{U}_1^\top + \mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top$. We can 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. Let $\mathbf{Z} = (\mathbf{Z}_1^\top, \mathbf{Z}_2^\top)^\top$, where $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are the first r rows and last $p - r$ rows of $\mathbf{Z}$. Then $\mathbf{Y} = \mathbf{U}_1\mathbf{\Lambda}_1^{1/2}\mathbf{Z}_1 + \mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2$.

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 $i = 1, \ldots, r$,*

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

and

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

Remark 2. Recently, the asympotic 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 $\boldsymbol{\Sigma}$ while still gives the correct convergence rate.

Based on Proposition 1, we propose the following estimators of $\operatorname{tr}(\boldsymbol{\Lambda}_2)$ and $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r$,

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

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

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

The following proposition gives the convergence rate of these estimators.

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

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

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 $\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_r$ 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 $\operatorname{tr}(\boldsymbol{\Lambda}_2^2)$ is relatively unexplored. Recently, Aoshima and Yata (2018) proposed an estimator of $\operatorname{tr}(\boldsymbol{\Lambda}_2^2)$ 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. In comparison, our estimator does not suffer from this problem. Moreover, we give the consistency rate of our estimator.

Next we consider the asymptotic behavior of the eigenspaces of $\hat{\boldsymbol{\Sigma}}$. 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.

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

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

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

Remark 5. The condition $\operatorname{tr}(\mathbf{\Lambda}_2)/(n\boldsymbol{\lambda}_r) \to 0$ 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 $\mathbf{\Sigma}$. 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 $\hat{\mathbf{\Sigma}}$ and $\mathbf{\Sigma}$, 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 n 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 n eigenvectors of $\hat{\mathbf{\Sigma}}$. 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 properties 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 $\operatorname{tr}(\boldsymbol{\Lambda}_2)/(n\boldsymbol{\lambda}_r) \to 0$ 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 asymptotc 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 Wishart$(r, \mathbf{I}_{k-1})$ and is independent of $\mathbf{W}_{k-1}$, where Wishart$(m, \boldsymbol{\Psi})$ is the Wishart distribution with parameter $\boldsymbol{\Psi}$ 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 $r = o(\sqrt{n})$, $r\operatorname{tr}(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1/(n\boldsymbol{\lambda}_r^2) \to 0$, $rn\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2) \to 0$, $r\boldsymbol{\lambda}_{r+1}/\sqrt{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)} \to 0$ and $\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p = O(n^{-1}\operatorname{tr}^{1/2}(\boldsymbol{\Lambda}_2^2))$. Then*

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

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

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

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

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

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

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

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

Then we reject the null hypothesis if

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

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

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

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

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

(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* $\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\| = O(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)})$,

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

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

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

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

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

then the tests in Bai and Saranadasa (1996) and Chen and Qin (2010). It is not hard to see that

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

To compare $\mathbf{C}^\top \boldsymbol{\Theta}^\top \mathbf{U}_2 \mathbf{U}_2^\top \boldsymbol{\Theta} \mathbf{C}$ and $\mathbf{C}^\top \boldsymbol{\Theta}^\top \boldsymbol{\Theta} \mathbf{C}$, we temporarily play a prior on $\boldsymbol{\Theta}$. Suppose that $\sqrt{n_i}\theta_i$ has prior distribution $\mathcal{N}_p(0, \psi \mathbf{I}_p)$, $i=1,2$. Then $\psi^{-1}\mathbf{C}^\top \boldsymbol{\Theta}^\top \boldsymbol{\Theta} \mathbf{C}$ is distributed as χ^2 distribution with p degrees of freedom. On the other hand, $\psi^{-1}\mathbf{C}^\top \boldsymbol{\Theta}^\top \mathbf{U}_2 \mathbf{U}_2^\top \boldsymbol{\Theta} \mathbf{C}$ is distributed as χ^2 distribution with $p-r$ degrees of freedom. Then we have

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

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

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

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

In practice, one may not know if the covariance matrix is spiked. Even if it is known that the covariance matrix is spiked, the spike number r may be unknown. Now we consider the detection of spiked covariance and the estimation of r. 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 $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2) \to 0$. We use the following statistic to estimate r:

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

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

Proposition 5. *Suppose $p/n \to \infty$, $r = o(n)$, $\boldsymbol{\lambda}_r n/p \to \infty$ and $c_1 \geq \boldsymbol{\lambda}_{r+1} \geq \ldots \geq \boldsymbol{\lambda}_p \geq c_2$. If $\gamma_n \to \infty$ and $\gamma_n = o(n\boldsymbol{\lambda}_r/p)$, then* $\Pr(\hat{r} = r) \to 1$.

Remark 8. For the factor model adopted by Ma et al. (2015), λ_r is of order p. Hence we can take $\gamma_n = \sqrt{n}$.

4. Numerical study

In this section, we compare the numerical performance of LFD test with some existing tests, including the MANOVA tests of Schott (2007) and Cai and Xia (2014) and the two sample tests of Srivastava (2007), Chen and Qin (2010), Cai et al. (2014) and Feng et al. (2015). Note that the critical values of these existing tests are not valid under spiked covariance model. Hence we use permutation method to determine the critical values throughout our simulations. The test procedures resulting from permutation method have exact levels as long as the null distribution of observations are exchangeable (ROMANO, 1990). The major down-side to permutation method is that it can be computationally intensive. Fortunately, for LFD test statistic, the permutation method has a simple implementation. By expression (2.4), a permuted statistic can be written as

$$T(\mathbf{X}\Gamma) = \lambda_{\max}\Big(\mathbf{C}^\top\big(\mathbf{J}^\top\Gamma^\top(\mathbf{X}^\top\mathbf{X})^{-1}\Gamma\mathbf{J}\big)^{-1}\mathbf{C}\Big), \tag{4.5}$$

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

1. Calculate $T(\mathbf{X})$ according to (2.4), keep intermediate result $(\mathbf{X}^\top\mathbf{X})^{-1}$.

2. For a large M, independently generate M random permutation matrix

$\Gamma_1, \ldots, \Gamma_M$ and calculate $T(\mathbf{X}\Gamma_1), \ldots, T(\mathbf{X}\Gamma_M)$ according to (4.5).

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

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

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

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

We use SNR to characterize the signal strength.

In the first simulation study, we take $k = 3$. For comparison, we also carry out simulations for the tests of Schott (2007) and Cai and Xia (2014). We denote these two tests by SC and CX, respectively. We take $r = 2$ and $\mathbf{\Sigma} = \mathrm{diag}(1.5p, p, 1, 1, \ldots, 1)$. We consider two different structures of alternative hypotheses: the non-sparse alternative and the sparse alternative. In the non-sparse case, we set $\theta_1 = \kappa 1_p$, $\theta_2 = -\kappa 1_p$ and $\theta_3 = \mathbf{0}_p$, where κ is selected to make the SNR equal to specific values. In the sparse case, we set $\theta_1 = \kappa(1_{p/5}^\top, \mathbf{0}_{4p/5}^\top)^\top$, $\theta_2 = \kappa(\mathbf{0}_{p/5}^\top, 1_{p/5}^\top, \mathbf{0}_{3p/5}^\top)^\top$ and $\theta_3 = \mathbf{0}_p$. Again, κ

is selected to make the SNR equal to specific values. The empirical power is computed based on 1000 simulations. The simulation results are summarized in Tables 1-4. It can be seen from the results that the proposed test outperforms the other two tests for both non-sparse and sparse alternatives. This verifies our theoretical results that LFD test performs well under spiked covariance.

In our second simulation study, we would like to investigate the effect of correlations between variables. We take $k = 2$ so that we can compare our test with some existing two sample tests. For comparison, we carry out simulations for the test of Srivastava (2007), Chen and Qin (2010), Cai et al. (2014) and Feng et al. (2015). We denote these tests by SR, CQ, CLX and FZWZ, respectively. Let the diagonal elements of Σ be 1 and the off-diagonal elements of Σ be ρ with $0 \leq \rho < 1$. The parameter ρ characterizes the correlations between variables. We set $\theta_1 = \kappa(\mathbf{1}_{p/2}^\top, -\mathbf{1}_{p/2}^\top)^\top$ and $\theta_2 = \mathbf{0}_p$, where κ is selected such that SNR equals to 5. Figure 1 plots the empirical power versus ρ, where empirical power is computed based on 1000 simulations. We can see that the empirical power of LFD test holds nearly constant as ρ varies while the empirical powers of Chen and Qin (2010) and Feng et al. (2015)'s tests decrease as ρ increases. When ρ is small, LFD test has reasonable performance. When ρ is larger than 0.1, LFD test outperforms all other tests.

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

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

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

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

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

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

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

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

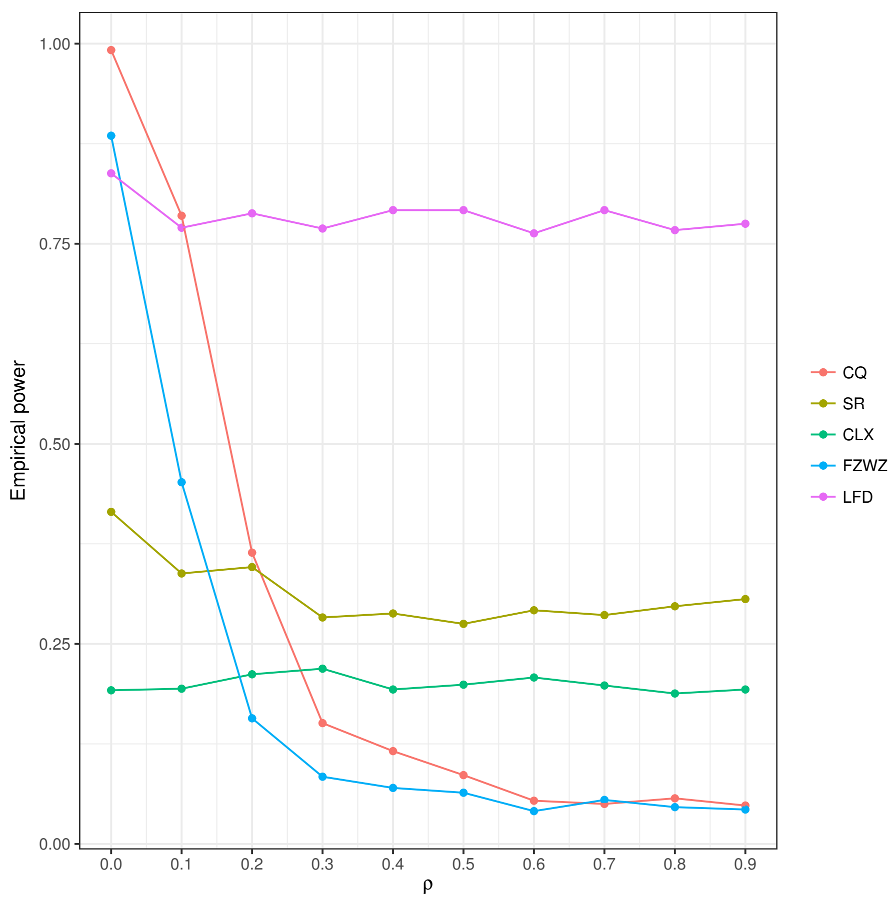

Figure 1: The empirical powers of tests. $\alpha = 0.05$, $k = 2$, $n_1 = n_2 = 20$, $p = 150$.

5. Concluding remarks

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

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

Appendix A Technical details

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 $\mathbf{A} = \mathbf{U_A}\mathbf{D_A}\mathbf{V_A}^\top$ 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 matrix, $\mathbf{D}_\mathbf{A}$ 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 on the column space of $\mathbf{A}$. Then*

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

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

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

which is obviously no greater than $\lambda_{\max}\big((\mathbf{I} - \mathbf{P}_\mathbf{A})\mathbf{B}(\mathbf{I} - \mathbf{P}_\mathbf{A})\big)$. To prove that they are equal, without loss of generality, we can assume $\lambda_{\max}\big((\mathbf{I} - \mathbf{P}_\mathbf{A})\mathbf{B}(\mathbf{I} - \mathbf{P}_\mathbf{A})\big) > 0$. 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.7) and (A.7) is no less than $\lambda_{\max}\big((\mathbf{I} - \mathbf{P}_\mathbf{A})\mathbf{B}(\mathbf{I} - \mathbf{P}_\mathbf{A})\big)$. The conclusion now follows by noting that $\lambda_{\max}\big((\mathbf{I} - \mathbf{P}_\mathbf{A})\mathbf{B}(\mathbf{I} - \mathbf{P}_\mathbf{A})\big) = \lambda_{\max}\big(\mathbf{B}(\mathbf{I} - \mathbf{P}_\mathbf{A})\big)$.

□

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

$$\frac{\max_{1\le i\le n} a_{n,i}^2}{\sum_{i=1}^n a_{n,i}^2} \xrightarrow{P} 0. \tag{A.8}$$

Then

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

Proof. First we observe that if $\{a_{n,i}\}_{i=1}^n$ are fixed numbers satisfying (A.8), then Lyapunov central limit theoremi 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.8) 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,

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

holds 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* $r + s - 1 \le i \le j + k - n$*, we have*

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

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

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

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

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* $|\omega|_\infty = \max_{1 \le i \le n} |\omega_i|$*.*

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 $\mathbf{A}$,

$$\|A\| \leq 2 \max_{x \in \mathcal{C}} x^\top \mathbf{A} x. \tag{A.9}$$

Then for $t > 4$,

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

where the first inequality follows from (A.9), 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. □

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 $\hat{\boldsymbol{\Sigma}}$. Write

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

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

$$
\left|\lambda_i\left(n^{-1}\mathbf{Z}^\top\boldsymbol{\Lambda}\mathbf{Z}\right) - \lambda_i(n^{-1}\mathbf{Z}_1^\top\boldsymbol{\Lambda}_1\mathbf{Z}_1) - n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\right| \leq n^{-1}\left\|\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2 - \operatorname{tr}(\boldsymbol{\Lambda}_2)\mathbf{I}_n\right\|. \tag{A.10}
$$

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

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

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

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

The above lower bound and upper bound imply

$$\begin{aligned}\left|\lambda_i(n^{-1}\mathbf{Z}_1^\top\mathbf{\Lambda}_1\mathbf{Z}_1) - \boldsymbol{\lambda}_i\right| \leq & \boldsymbol{\lambda}_i \max\left(|\lambda_1(n^{-1}\mathbf{Z}_1\mathbf{Z}_1^\top) - 1|, |\lambda_r(n^{-1}\mathbf{Z}_1\mathbf{Z}_1^\top) - 1|\right) \\ = & \boldsymbol{\lambda}_i\|n^{-1}\mathbf{Z}_1\mathbf{Z}_1^\top - \mathbf{I}_r\|.\end{aligned} \tag{A.11}$$

Combining the bounds (A.10) and (A.11) gives that for $i = 1, \ldots, r$,

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

From Lemma 5, we have

$$\|n^{-1}\mathbf{Z}_1\mathbf{Z}_1^\top - \mathbf{I}_r\| = O_P\left(\sqrt{\frac{r}{n}}\right), \tag{A.12}$$

$$n^{-1}\left\|\mathbf{Z}_2^\top\mathbf{\Lambda}_2\mathbf{Z}_2 - \operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_n\right\| = O_P\left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{n}} + \boldsymbol{\lambda}_{r+1}\right). \tag{A.13}$$

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 (A.10) and (A.13) that

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

Thus,

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

It is straightforward to show that

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

Hence

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

This completes the proof of the second statement. □

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 $\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}=n^{-2}\sum_{i=r+1}^{n}\lambda_i^2(\mathbf{Y}^\top\mathbf{Y}-\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n)$. Note that Weyl's inequality implies that for $i=r+1,\ldots,n$,

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

Define

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

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

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

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

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

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

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

Therefore,

$$
\begin{aligned}
&\left|\sum_{i=r+1}^{n}\lambda_i^2\left(\mathbf{Y}^\top\mathbf{Y}-\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n\right)-\operatorname{tr}(\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2-\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n)^2\right|\\
\leq&\left|\sum_{i>r,\,i\in\mathcal{C}_1}\lambda_i^2\left(\mathbf{Y}^\top\mathbf{Y}-\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n\right)-\sum_{i\in\mathcal{C}_1}\lambda_i^2\left(\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2-\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n\right)\right|\\
&+\left|\sum_{i>r,\,i\in\mathcal{C}_2}\lambda_i^2\left(\mathbf{Y}^\top\mathbf{Y}-\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n\right)-\sum_{i\notin\mathcal{C}_1}\lambda_i^2\left(\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2-\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n\right)\right|\\
&+\left|\sum_{i>r,\,i\notin\mathcal{C}_1\cup\mathcal{C}_2}\lambda_i^2\left(\mathbf{Y}^\top\mathbf{Y}-\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n\right)\right|\\
\leq&3r\|\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2-\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n\|^2\\
\leq&3r\left(\|\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2-\operatorname{tr}(\boldsymbol{\Lambda}_2)\mathbf{I}_n\|+\left|\operatorname{tr}(\boldsymbol{\Lambda}_2)-\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\right|\right)^2\\
=&O_P\left(rn\operatorname{tr}(\boldsymbol{\Lambda}_2^2)+rn^2\boldsymbol{\lambda}_{r+1}^2\right).
\end{aligned}
\tag{A.14}
$$

where the last equality follows from (A.13) and the first statement of the proposition.

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

$$
\operatorname{tr}(\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2-\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\mathbf{I}_n)^2=\sum_{i=1}^{n}(Z_{2,i}^\top\boldsymbol{\Lambda}_2Z_{2,i}-\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)})^2+2\sum_{1\leq i<j\leq n}(Z_{2,i}^\top\boldsymbol{\Lambda}_2Z_{2,j})^2.
$$

For the first term, we have

$$
\sum_{i=1}^{n}(Z_{2,i}^\top\boldsymbol{\Lambda}_2Z_{2,i}-\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)})^2\leq 2\sum_{i=1}^{n}(Z_{2,i}^\top\boldsymbol{\Lambda}_2Z_{2,i}-\operatorname{tr}(\boldsymbol{\Lambda}_2))^2+2n(\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}-\operatorname{tr}(\boldsymbol{\Lambda}_2))^2.
$$

Then it follows from the first 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

$$\sum_{i=1}^n (Z_{2,i}^\top \mathbf{\Lambda}_2 Z_{2,i} - \widehat{\operatorname{tr}(\mathbf{\Lambda}_2)})^2 = O_P\left((n+r^2)\operatorname{tr}(\mathbf{\Lambda}_2^2) + r^2 n \boldsymbol{\lambda}_{r+1}^2\right). \qquad \text{(A.15)}$$

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 (A.15) 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 (A.14), 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.

$\square$

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

$$\|\mathbf{P}_{\mathbf{Y},1} - \mathbf{P}^*_{\mathbf{Y},1}\| \le \frac{2\|\hat{\boldsymbol{\Sigma}} - \hat{\boldsymbol{\Sigma}}_0\|}{\lambda_r(\hat{\boldsymbol{\Sigma}}_0)}. \tag{A.16}$$

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

$$\begin{aligned}
\|\hat{\boldsymbol{\Sigma}} - \hat{\boldsymbol{\Sigma}}_0\| =& n^{-1}\left\|\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^\top - \mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^\top\right\| \\
=& n^{-1}\left\|\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2(\mathbf{I}_n - \mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^\top)\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\right\| \\
\leq& n^{-1}\left\|\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2\right\| \\
\leq& n^{-1}\left\|\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2 - \operatorname{tr}(\boldsymbol{\Lambda}_2)\mathbf{I}_n\right\| + n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2) \\
=& O_P\left(\sqrt{\frac{\operatorname{tr}\left(\boldsymbol{\Lambda}_2^2\right)}{n}} + \boldsymbol{\lambda}_{r+1} + n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\right) \\
=& O_P\left(\boldsymbol{\lambda}_{r+1} + n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\right),
\end{aligned} \tag{A.17}$$

where the second last equality follows from (A.13) and the last equality follows from

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

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

$$\begin{aligned}\left\|n^{-1}\mathbf{V}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2 \mathbf{Z}_2 \mathbf{V}_{\mathbf{Z}_1} - n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\mathbf{I}_r\right\| =& O_P\left(n^{-1}\sqrt{r\operatorname{tr}(\mathbf{\Lambda}_2^2)} + rn^{-1}\boldsymbol{\lambda}_{r+1}\right)\\ =& O_P\left(n^{-1}\sqrt{r\boldsymbol{\lambda}_{r+1}\operatorname{tr}(\mathbf{\Lambda}_2)} + rn^{-1}\boldsymbol{\lambda}_{r+1}\right)\\ =& o_P\left(n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)\right),\end{aligned} \tag{A.18}$$

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, (A.11) and (A.12) 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

$$\lambda_r(\hat{\mathbf{\Sigma}}_0) = (1 + o_P(1))(\boldsymbol{\lambda}_r + n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)). \tag{A.19}$$

Then the conclusion follows from (A.16), (A.17) and (A.19). □

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

$$\begin{aligned}
\|\mathbf{Q}^{\top}\mathbf{Q}\| =&\left\|\mathbf{\Lambda}_1^{-1/2}(\mathbf{Z}_1\mathbf{Z}_1^{\top})^{-1/2}\mathbf{V}_{\mathbf{Z}_1}^{\top}\mathbf{Z}_2^{\top}\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}(\mathbf{Z}_1\mathbf{Z}_1^{\top})^{-1/2}\mathbf{\Lambda}_1^{-1/2}\right\| \\
\le&\boldsymbol{\lambda}_r^{-1}\left\|(\mathbf{Z}_1\mathbf{Z}_1^{\top})^{-1}\right\|\left\|\mathbf{V}_{\mathbf{Z}_1}^{\top}\mathbf{Z}_2^{\top}\mathbf{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right\| \\
=&O_P\left(\frac{\mathrm{tr}(\mathbf{\Lambda}_2)}{n\boldsymbol{\lambda}_r}\right),
\end{aligned} \tag{A.20}$$

where the second last equality follows from the fact $\|(\mathbf{Z}_1\mathbf{Z}_1^{\top})^{-1}\| = \lambda_r(\mathbf{Z}_1\mathbf{Z}_1^{\top})^{-1}$, (A.12), (A.18) and Weyl's inequality. Therefore, we have

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

This completes the proof.

$\square$

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

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

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

$$\left\|(\mathbf{P}^{\dagger}_{\mathbf{Y},1}-\mathbf{P}_{\mathbf{Y},1})\hat{\boldsymbol{\Sigma}}(\mathbf{P}^{\dagger}_{\mathbf{Y},1}-\mathbf{P}_{\mathbf{Y},1})\right\|\le\|\hat{\boldsymbol{\Sigma}}\|\left\|\mathbf{P}^{\dagger}_{\mathbf{Y},1}-\mathbf{P}_{\mathbf{Y},1}\right\|^2=O_P\left(\frac{\operatorname{tr}^2(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1}{n^2\boldsymbol{\lambda}_r^2}\right).\tag{A.21}$$

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

$$\mathbf{Z}^\top\boldsymbol{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p-\mathbf{P}^{\dagger}_{\mathbf{Y},1})=\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^\top-\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2\mathbf{Z}_1^\top(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}\boldsymbol{\Lambda}_1^{-1/2}\mathbf{U}_1^\top.\tag{A.22}$$

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 (A.13) and the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2) \to 0$ that

$$\|\mathbf{Z}_2^\top \boldsymbol{\Lambda}_2 \mathbf{Z}_2\| = (1 + o_P(1))\operatorname{tr}(\boldsymbol{\Lambda}_2). \tag{A.23}$$

Consequently,

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

Thus,

$$\left\|(\mathbf{P}^\dagger_{\mathbf{Y},1} - \mathbf{P}_{\mathbf{Y},1})\hat{\boldsymbol{\Sigma}}(\mathbf{I}_p - \mathbf{P}^\dagger_{\mathbf{Y},1})\right\| = O_P\left(\frac{\operatorname{tr}^{3/2}(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1^{1/2}}{n^{3/2}\boldsymbol{\lambda}_r}\right). \tag{A.24}$$

Combine (A.21) and (A.24), we obtain

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

Now we deal with $(\mathbf{I}_p - \mathbf{P}^\dagger_{\mathbf{Y},1})\hat{\boldsymbol{\Sigma}}(\mathbf{I}_p - \mathbf{P}^\dagger_{\mathbf{Y},1})$. In view of (A.22), we have

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

Then

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

where the last equality follows from (A.20) and (A.23).

Combine the above bounds, we obtain

$$\begin{aligned}&\left\|(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1})\hat{\boldsymbol{\Sigma}}(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y},1})-n^{-1}\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^\top\right\|\\=&O_P\left(\frac{\operatorname{tr}^2(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1}{n^2\boldsymbol{\lambda}_r^2}+\frac{\operatorname{tr}^{3/2}(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1^{1/2}}{n^{3/2}\boldsymbol{\lambda}_r}\right).\end{aligned}\tag{A.25}$$

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

$$\begin{aligned}\left\|n^{-1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}-n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\mathbf{I}_{n-r}\right\|=&O_P\left(n^{-1/2}\sqrt{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}+\boldsymbol{\lambda}_{r+1}\right)\\=&O_P\left(n^{-1/2}\sqrt{\boldsymbol{\lambda}_{r+1}\operatorname{tr}(\boldsymbol{\Lambda}_2)}+\boldsymbol{\lambda}_{r+1}\right)\\=&o_P\left(n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\right).\end{aligned}\tag{A.26}$$

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

$$\lambda_{n-r}\left(n^{-1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\right)=(1+o_P(1))n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_1).\tag{A.27}$$

As a result, the matrix $n^{-1}\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^\top$ is of rank $n-r$. 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

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

where the second last equality follows from (A.25) and (A.27). This completes the proof. $\square$

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

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

where the second last equality follows from (A.26) and the last equality follows from the fact $\sqrt{n\operatorname{tr}(\mathbf{\Lambda}_2^2)}/\operatorname{tr}(\mathbf{\Lambda}_2) \leq \sqrt{n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2)}$ and the condition $\sqrt{n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2)} \to 0$. Then the conclusion follows from the last display and Proposition 7.

$\square$

Appendix B Proofs of the main results

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

$$\begin{aligned}\mathbf{C}^{\top}\mathbf{J}^{\top}\mathbf{X}^{\top}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{XJC} =& \mathbf{Z}^{\dagger\top}\mathbf{\Lambda}^{1/2}\mathbf{U}^{\top}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U\Lambda}^{1/2}\mathbf{Z}^{\dagger} + \mathbf{C}^{\top}\mathbf{\Theta}^{\top}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Theta C} \\ &+ \mathbf{C}^{\top}\mathbf{\Theta}^{\top}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U\Lambda}^{1/2}\mathbf{Z}^{\dagger} + \mathbf{Z}^{\dagger\top}\mathbf{\Lambda}^{1/2}\mathbf{U}^{\top}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{\Theta C}.\end{aligned} \tag{B.28}$$

It can be seen that the first term of (B.28) can be written as

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

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

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

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

And

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

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

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

Then from Lemma 5, we have

$$\begin{aligned}\left\|\mathbf{P_Y}-(\operatorname{tr}(\boldsymbol{\Sigma}))^{-1}\mathbf{Y}\mathbf{Y}^\top\right\| &=(\operatorname{tr}(\boldsymbol{\Sigma}))^{-1}\left\|\mathbf{Z}^\top\boldsymbol{\Sigma}\mathbf{Z}-\operatorname{tr}(\boldsymbol{\Sigma})\mathbf{I}_n\right\| \\ &=O_P\left(\frac{\sqrt{n\operatorname{tr}(\boldsymbol{\Sigma}^2)}}{\operatorname{tr}(\boldsymbol{\Sigma})}+\frac{n\boldsymbol{\lambda}_1}{\operatorname{tr}(\boldsymbol{\Sigma})}\right) \\ &=O_P\left(\frac{\sqrt{n\boldsymbol{\lambda}_1\operatorname{tr}(\boldsymbol{\Sigma})}}{\operatorname{tr}(\boldsymbol{\Sigma})}+\frac{n\boldsymbol{\lambda}_1}{\operatorname{tr}(\boldsymbol{\Sigma})}\right) \\ &=O_P\left(\sqrt{\frac{n\boldsymbol{\lambda}_1}{\operatorname{tr}(\boldsymbol{\Sigma})}}\right).\end{aligned} \tag{B.29}$$

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

$$\begin{aligned}\operatorname{tr}((\mathbf{I}_p-\mathbf{P_Y})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P_Y})) &= \operatorname{tr}(\boldsymbol{\Sigma})-\operatorname{tr}(\boldsymbol{\Sigma}\mathbf{P_Y}) \\ &= \operatorname{tr}(\boldsymbol{\Sigma})-\operatorname{tr}\left(\left(\boldsymbol{\Sigma}-\frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{I}_p\right)\mathbf{P_Y}\right)-\frac{n\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}.\end{aligned} \tag{B.30}$$

For the second term, we have

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

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 (B.29) 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

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

and

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

Thus,

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

Then the first statement follows from the last display and (B.30).

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

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

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

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

and the third term satisfies

$$\left|\operatorname{tr}((\boldsymbol{\Sigma}\mathbf{P}_{\mathbf{Y}})^2) - \frac{n\operatorname{tr}^2(\boldsymbol{\Sigma}^2)}{\operatorname{tr}^2(\boldsymbol{\Sigma})}\right| = \left|\operatorname{tr}\left(\left(\boldsymbol{\Sigma} + \frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{I}_p\right)\mathbf{P}_{\mathbf{Y}}\left(\boldsymbol{\Sigma} - \frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\mathbf{I}_p\right)\mathbf{P}_{\mathbf{Y}}\right)\right|$$

$$\leq 2n\boldsymbol{\lambda}_1(\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_p).$$

This completes the proof of the second statement. □

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

$$\operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) = \operatorname{tr}(\boldsymbol{\Sigma}) - \frac{n\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})} + o_P(\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)}), \tag{B.31}$$

$$\operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)^2 = (1 + o_P(1))\operatorname{tr}(\boldsymbol{\Sigma}^2). \tag{B.32}$$

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 (B.32) imply that the first term of (B.28) 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 (B.31), (B.32) and Slutsky's theorem that

$$\frac{\mathbf{Z}^{\dagger\top}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{Z}^\dagger - \left(\operatorname{tr}(\boldsymbol{\Sigma}) - n\operatorname{tr}(\boldsymbol{\Sigma}^2)/\operatorname{tr}(\boldsymbol{\Sigma})\right)\mathbf{I}_{k-1}}{\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)}} \xrightarrow{\mathcal{L}} \mathbf{W}_{k-1}. \tag{B.33}$$

Next we consider the second term of (B.28). Note that

$$\begin{aligned}\mathbf{C}^\top\boldsymbol{\Theta}^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Theta}\mathbf{C} =& \mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\mathbf{C} - \mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{Y}(\mathbf{Y}^\top\mathbf{Y})^{-1}\mathbf{Y}^\top\boldsymbol{\Theta}\mathbf{C}\\ =& \mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\mathbf{C} - \mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{Z}(\mathbf{Z}^\top\boldsymbol{\Lambda}\mathbf{Z})^{-1}\mathbf{Z}^\top\boldsymbol{\Lambda}^{1/2}\mathbf{U}^\top\boldsymbol{\Theta}\mathbf{C}.\end{aligned}$$

We have

$$
\begin{aligned}
&\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}(\mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{\Lambda}^{1/2}\mathbf{U}^\top\mathbf{\Theta}\mathbf{C}-\operatorname{tr}(\mathbf{\Sigma})^{-1}\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}\mathbf{Z}^\top\mathbf{\Lambda}^{1/2}\mathbf{U}^\top\mathbf{\Theta}\mathbf{C}\right\|\\
\leq&\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}\mathbf{Z}^\top\mathbf{\Lambda}^{1/2}\mathbf{U}^\top\mathbf{\Theta}\mathbf{C}\right\|\left\|(\mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z})^{-1}-\operatorname{tr}(\mathbf{\Sigma})^{-1}\mathbf{I}_n\right\|\\
\leq&\left\|\operatorname{tr}(\mathbf{\Sigma})^{-1}\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}\mathbf{Z}^\top\mathbf{\Lambda}^{1/2}\mathbf{U}^\top\mathbf{\Theta}\mathbf{C}\right\|\left\|(\mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z})^{-1}\right\|\left\|\mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z}-\operatorname{tr}(\mathbf{\Sigma})\mathbf{I}_n\right\|.
\end{aligned}
$$

From Lemma 5, we have $\left\|\mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z}-\operatorname{tr}(\mathbf{\Sigma})\mathbf{I}_n\right\| = O_P(\sqrt{n\operatorname{tr}(\mathbf{\Sigma}^2)}+n\boldsymbol{\lambda}_1) = o_P(\operatorname{tr}(\mathbf{\Sigma}))$. Then $\left\|(\mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z})^{-1}\right\| = \lambda_n^{-1}(\mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z}) = (1+o_P(1))\operatorname{tr}(\mathbf{\Sigma})$. Therefore,

$$
\begin{aligned}
&\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}(\mathbf{Z}^\top\mathbf{\Lambda}\mathbf{Z})^{-1}\mathbf{Z}^\top\mathbf{\Lambda}^{1/2}\mathbf{U}^\top\mathbf{\Theta}\mathbf{C}-\operatorname{tr}(\mathbf{\Sigma})^{-1}\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}\mathbf{Z}^\top\mathbf{\Lambda}^{1/2}\mathbf{U}^\top\mathbf{\Theta}\mathbf{C}\right\|\\
=&o_P\left(\left\|\operatorname{tr}(\mathbf{\Sigma})^{-1}\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}\mathbf{Z}^\top\mathbf{\Lambda}^{1/2}\mathbf{U}^\top\mathbf{\Theta}\mathbf{C}\right\|\right).
\end{aligned}
$$

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

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

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

$$
\begin{aligned}
\left\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{\Theta}\mathbf{C}-\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right\| &\le (1+o_P(1))\frac{n}{\operatorname{tr}(\mathbf{\Sigma})}\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Sigma}\mathbf{\Theta}\mathbf{C}\right\| \\
&\le (1+o_P(1))\frac{n\boldsymbol{\lambda}_1}{\operatorname{tr}(\mathbf{\Sigma})}\left\|\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}\right\| \\
&= o_P\left(\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}\right).
\end{aligned}
\tag{B.34}
$$

Now we deal with the cross term of (B.28). Note that

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

Therefore,

$$
\begin{aligned}
\|\mathbf{C}^\top\mathbf{\Theta}^\top(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}^\dagger\| &= O_P\left(\sqrt{\boldsymbol{\lambda}_1\operatorname{tr}(\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C})}\right) \\
&= o_P\left(\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}\right),
\end{aligned}
\tag{B.35}
$$

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

It follows from (B.34), (B.35) and Weyl's inequality that

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

But (B.33) implies that

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

This completes the proof.

$\square$

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

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

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

$$
\begin{aligned}
\operatorname{Var}\left(2\sum_{1\leq i<j\leq n}(Z_i^\top\boldsymbol{\Lambda}Z_j)^2\right)=&O\left(n^2\operatorname{tr}^2(\boldsymbol{\Sigma}^2)+n^3\operatorname{tr}(\boldsymbol{\Sigma}^4)\right)\\
=&O\left(n^2\operatorname{tr}^2(\boldsymbol{\Sigma}^2)+n\operatorname{tr}(\boldsymbol{\Sigma}_2^2)n^2\boldsymbol{\lambda}_1^2\right)\\
=&O\left(n^2\operatorname{tr}^2(\boldsymbol{\Sigma}^2)+n^4\boldsymbol{\lambda}_1^4\right).
\end{aligned}
$$

Hence $\widehat{\operatorname{tr}(\boldsymbol{\Sigma}^2)} = \operatorname{tr}(\boldsymbol{\Sigma}^2) + O_P(n^{-1}\operatorname{tr}(\boldsymbol{\Sigma}^2) + \boldsymbol{\lambda}_1^2)$.

Thus, we have $\widehat{\operatorname{tr}(\boldsymbol{\Sigma}^2)} = (1+o(1))\operatorname{tr}(\boldsymbol{\Sigma}^2)$ and

$$
\begin{aligned}
&\widehat{\operatorname{tr}(\boldsymbol{\Sigma})} - n\widehat{\operatorname{tr}(\boldsymbol{\Sigma}^2)}/\widehat{\operatorname{tr}(\boldsymbol{\Sigma})} \\
&= \operatorname{tr}(\boldsymbol{\Sigma}) + O_P(\sqrt{n^{-1}\operatorname{tr}(\boldsymbol{\Sigma}^2)}) - \frac{n\operatorname{tr}(\boldsymbol{\Sigma}^2)(1+O_P(n^{-1}+\boldsymbol{\lambda}_1^2/\operatorname{tr}(\boldsymbol{\Sigma}^2)))}{\operatorname{tr}(\boldsymbol{\Sigma})(1+O_P(\sqrt{n^{-1}\operatorname{tr}(\boldsymbol{\Sigma}^2)/\operatorname{tr}^2(\boldsymbol{\Sigma})}))} \\
&= \operatorname{tr}(\boldsymbol{\Sigma}) + O_P(\sqrt{n^{-1}\operatorname{tr}(\boldsymbol{\Sigma}^2)}) - \frac{n\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})}\left(1+O_P\left(\frac{1}{n}+\frac{\boldsymbol{\lambda}_1^2}{\operatorname{tr}(\boldsymbol{\Sigma}^2)}+\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Sigma}^2)}{n\operatorname{tr}^2(\boldsymbol{\Sigma})}}\right)\right) \\
&= \operatorname{tr}(\boldsymbol{\Sigma}) - \frac{n\operatorname{tr}(\boldsymbol{\Sigma}^2)}{\operatorname{tr}(\boldsymbol{\Sigma})} + o_P(\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)}).
\end{aligned}
$$

Therefore,

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

Then the conclusion follows from Theorem 1. □

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

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

Proof. Note that

$$
(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}}) = (\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},2})(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},2}). \tag{B.36}
$$

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

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

From the decomposition

$$(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}) =(\mathbf{I}_p - \mathbf{P}^\dagger_{\mathbf{Y},1})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}^\dagger_{\mathbf{Y},1}) + (\mathbf{P}^\dagger_{\mathbf{Y},1} - \mathbf{P}_{\mathbf{Y},1})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}^\dagger_{\mathbf{Y},1})$$
$$+ (\mathbf{I}_p - \mathbf{P}^\dagger_{\mathbf{Y},1})\boldsymbol{\Sigma}(\mathbf{P}^\dagger_{\mathbf{Y},1} - \mathbf{P}_{\mathbf{Y},1}) + (\mathbf{P}^\dagger_{\mathbf{Y},1} - \mathbf{P}_{\mathbf{Y},1})\boldsymbol{\Sigma}(\mathbf{P}^\dagger_{\mathbf{Y},1} - \mathbf{P}_{\mathbf{Y},1}),$$

we have

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

Note that

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

where the last equality follows from (A.18), (A.20) and the condition $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2) \to 0$. Thus,

$$\left\|(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y},1}) - (\mathbf{I}_p - \mathbf{P}^\dagger_{\mathbf{Y},1})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}^\dagger_{\mathbf{Y},1})\right\| = O_P\left(\frac{\operatorname{tr}^{3/2}(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1^{1/2}}{n^{3/2}\boldsymbol{\lambda}_r}\right). \tag{B.37}$$

From the decomposition

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

we have

$$
\begin{aligned}
\left\|(\mathbf{I}_p - \mathbf{P}^{\dagger}_{\mathbf{Y},1})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}^{\dagger}_{\mathbf{Y},1}) - \mathbf{U}_2\mathbf{Q}\boldsymbol{\Lambda}_1\mathbf{Q}^\top\mathbf{U}_2^\top\right\| \leq &\boldsymbol{\lambda}_{r+1}(1 + 2\|\mathbf{Q}^\top\mathbf{Q}\|^{1/2} + \|\mathbf{Q}^\top\mathbf{Q}\|) \\
=&O_P\left(\boldsymbol{\lambda}_{r+1}\right),
\end{aligned}
\tag{B.38}
$$

where the last equality follows from (A.20). Note that $\mathbf{U}_2\mathbf{Q}\boldsymbol{\Lambda}_1\mathbf{Q}^\top\mathbf{U}_2^\top = \mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^\top$. We have

$$
\begin{aligned}
&\left\|\mathbf{U}_2\mathbf{Q}\boldsymbol{\Lambda}_1\mathbf{Q}^\top\mathbf{U}_2^\top - n^{-1}\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^\top\right\| \\
\leq&\left\|\mathbf{U}_2\boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2^{1/2}\mathbf{U}_2^\top\right\|\left\|(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1} - n^{-1}\mathbf{I}_r\right\| \\
\leq&\left\|\mathbf{V}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\boldsymbol{\Lambda}_2\mathbf{Z}_2\mathbf{V}_{\mathbf{Z}_1}\right\|\left\|(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}\right\|\left\|n^{-1}\mathbf{Z}_1\mathbf{Z}_1^\top - \mathbf{I}_r\right\| \\
=&O_P\left(\frac{r^{1/2}\operatorname{tr}(\boldsymbol{\Lambda}_2)}{n^{3/2}}\right),
\end{aligned}
\tag{B.39}
$$

where the last equality follows from (A.12) and (A.18). From (B.36), (B.37), (B.38) and (B.39), we obtain that

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

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

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

Note that

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

where the last equality follows from (A.18) and Proposition 4. Then it follows from Weyl's inequality that uniformly for $i = 1, \ldots, r$,

$$\begin{aligned}&\lambda_i \left((\mathbf{I}_p - \mathbf{P}_\mathbf{Y})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_\mathbf{Y}) \right) \\ =& n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2) - (n\operatorname{tr}(\boldsymbol{\Lambda}_2))^{-1} \lambda_{r+1-i} \left(\mathbf{V}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \boldsymbol{\Lambda}_2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1} \tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \boldsymbol{\Lambda}_2 \mathbf{Z}_2 \mathbf{V}_{\mathbf{Z}_1} \right) \\ &+ O_P\left(\left(\sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}} + \sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\boldsymbol{\Lambda}_2)}} + \sqrt{\frac{r}{n}} \right) \frac{\operatorname{tr}(\boldsymbol{\Lambda}_2)}{n} \right).\end{aligned} \tag{B.40}$$

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

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

By some algebra, we have

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

It is straightforward to show that

$$\operatorname{E}\operatorname{tr}(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2^2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}) = (n-r)\operatorname{tr}(\mathbf{\Lambda}_2^2), \quad \operatorname{Var}\left(\operatorname{tr}(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2^2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1})\right) = 2(n-r)\operatorname{tr}(\mathbf{\Lambda}_2^4).$$

Hence $\operatorname{tr}(\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2^2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}) = (n-r)\operatorname{tr}(\mathbf{\Lambda}_2^2) + O_P(\sqrt{n}\operatorname{tr}(\mathbf{\Lambda}_2^2))$. On the other hand, Lemma 5 implies that $\|\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top \mathbf{Z}_2^\top \mathbf{\Lambda}_2^2 \mathbf{Z}_2 \tilde{\mathbf{V}}_{\mathbf{Z}_1}\| = O_P(\operatorname{tr}(\mathbf{\Lambda}_2^2) + n\boldsymbol{\lambda}_{r+1}^2)$. Combine these bounds, we have

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

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

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

Then (B.40) and the last display implies that uniformly for $i = 1, \ldots, r$,

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

This completes the proof.

$\square$

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

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

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

$$
\lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) \geq \lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right), \tag{B.41}
$$

$$
\lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) \leq \lambda_{i-r}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right). \tag{B.42}
$$

Hence we have

$$
\begin{aligned}
&\left|\sum_{i=r+1}^{p} \lambda_i\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) - \operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)\right| \\
\leq & r\lambda_1\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) \\
\leq & r\boldsymbol{\lambda}_{r+1}.
\end{aligned} \tag{B.43}
$$

Write

$$
\begin{aligned}
&\operatorname{tr}\left((\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\right)\\
=&\operatorname{tr}\left(\mathbf{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p-\mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\right)\\
=&\operatorname{tr}(\mathbf{\Lambda}_2)-\operatorname{tr}\left(\left(\mathbf{\Lambda}_2-\frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^\top\mathbf{P}_{\mathbf{Y}}\mathbf{U}_2\right)-\frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Lambda}_2)}\operatorname{tr}\left(\mathbf{U}_2^\top\mathbf{P}_{\mathbf{Y}}\mathbf{U}_2\right).
\end{aligned}
\tag{B.44}
$$

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

$$
\left|\operatorname{tr}\left(\mathbf{U}_2^\top\mathbf{P}_{\mathbf{Y}}\mathbf{U}_2\right)-n\right|\le r. \tag{B.45}
$$

Next we deal with the second term. We have

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

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

$$
\begin{aligned}
&\left|\operatorname{tr}\left(\left(\mathbf{\Lambda}_2-\frac{\operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Lambda}_2)}\mathbf{I}_{p-r}\right)\mathbf{U}_2^\top\left(\mathbf{P}_{\mathbf{Y}}-\mathbf{P}_{\mathbf{Y},1}^\dagger-\mathbf{P}_{\mathbf{Y},2}^\dagger\right)\mathbf{U}_2\right)\right|\\
\le&2n(\boldsymbol{\lambda}_{r+1}-\boldsymbol{\lambda}_p)\left\|\mathbf{P}_{\mathbf{Y}}-\mathbf{P}_{\mathbf{Y},1}^\dagger-\mathbf{P}_{\mathbf{Y},2}^\dagger\right\|\\
\le&2n(\boldsymbol{\lambda}_{r+1}-\boldsymbol{\lambda}_p)\left(\left\|\mathbf{P}_{\mathbf{Y},1}-\mathbf{P}_{\mathbf{Y},1}^\dagger\right\|+\left\|\mathbf{P}_{\mathbf{Y},2}-\mathbf{P}_{\mathbf{Y},2}^\dagger\right\|\right)\\
=&O_P\left(n(\boldsymbol{\lambda}_{r+1}-\boldsymbol{\lambda}_p)\left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}}+\sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\mathbf{\Lambda}_2)}}\right)\right),
\end{aligned}
\tag{B.46}
$$

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

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

It is straightforward to show that

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

and

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

Thus,

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

The last display, combined with (B.46), leads to

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

It then follows from (B.44), (B.45) and the last display that

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

Then the conclusion follows from (B.43) and the last display. □

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

$$
\sum_{i=r+1}^{p}\lambda_i^2\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) = \operatorname{tr}(\boldsymbol{\Lambda}_2^2) - \frac{n\operatorname{tr}^2(\boldsymbol{\Lambda}_2^2)}{\operatorname{tr}^2(\boldsymbol{\Lambda}_2)} + O_P\left(n\boldsymbol{\lambda}_{r+1}(\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p) + r\boldsymbol{\lambda}_{r+1}^2\right). \tag{B.47}
$$

Proof. From (B.41) and (B.42), we have

$$
\begin{aligned}
&\left|\sum_{i=r+1}^{p}\lambda_i^2\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) - \operatorname{tr}\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right)^2\right| \\
\le& r\lambda_1^2\left((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}_2\boldsymbol{\Lambda}_2\mathbf{U}_2^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\right) \\
\le& r\boldsymbol{\lambda}_{r+1}^2.
\end{aligned} \tag{B.48}
$$

It is straightforward to show that

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

For the second term, we have

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

where the second last equality follows from von Neumann's trace theorem.

The last display, combined with (B.45), implies that

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

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

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

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

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

Thus,

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

Then the conclusion follows from the last display and (B.48). $\square$

Proof of Theorem 2. We have

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

From Lemma 7, we have

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

Then

$$
\frac{\sum_{i=1}^{r}\lambda_i((\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_\mathbf{Y}))\eta_i\eta_i^\top-rn^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)\mathbf{I}_{k-1}}{\sqrt{r}n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)}=\frac{\sum_{i=1}^{r}\eta_i\eta_i^\top-r\mathbf{I}_{k-1}}{\sqrt{r}}+o_P(1). \tag{B.49}
$$

Next we deal with $\sum_{i=r+1}^{p}\lambda_i((\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\boldsymbol{\Sigma}(\mathbf{I}_p-\mathbf{P}_\mathbf{Y}))\eta_i\eta_i^\top$. By Weyl's inequality,

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

The last display and Lemma 9 imply that

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

From Lemma 2, we have

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

Then it follows from Lemma 8 and Lemma 9 and Slutsky's theorem that

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

Since $\sum_{i=1}^{r} \eta_i\eta_i^\top$ is independent of $\sum_{i=r+1}^{p} \lambda_i((\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Sigma}(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}}))\eta_i\eta_i^\top$, (B.49) and (B.50) implies that

$$\begin{aligned}
&\frac{\mathbf{Z}^{\dagger\top}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathbf{Z}^\dagger - ((1 + r/n)\operatorname{tr}(\boldsymbol{\Lambda}_2) - n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)/\operatorname{tr}(\boldsymbol{\Lambda}_2))\,\mathbf{I}_{k-1}}{\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2) + \operatorname{tr}(\boldsymbol{\Lambda}_2^2)}} \\
=&\frac{n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)}{\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2) + \operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}(\mathbf{W}_{k-1}^* - r\mathbf{I}_{k-1}) + \frac{\sqrt{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}{\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2) + \operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}\mathbf{W}_{k-1} + o_P(1).
\end{aligned} \tag{B.51}$$

This completes the proof of the first statement.

Now we prove the second statement. For the second term of (B.28), we have $\mathbf{C}^\top\boldsymbol{\Theta}^\top(\mathbf{I}_p - \mathbf{P}_{\mathbf{Y}})\boldsymbol{\Theta}\mathbf{C} = \mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\mathbf{C} - \mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{P}_{\mathbf{Y}}\boldsymbol{\Theta}\mathbf{C}$. Note that Proposition 3 implies that

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

It follows from the last display and Proposition 4 that

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

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

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

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

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

Hence $\left\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{P}_{\mathbf{Y},2}^\dagger\boldsymbol{\Theta}\mathbf{C}\right\| = o_P\left(\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2)+\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}\right)$. Combine the above arguments, we have

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

Next we consider the cross term of (B.28). Note that

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

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

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

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

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

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

□

Proof of Corollary 2. From Proposition 2, we have

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

and

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

Therefore,

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

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

Proof of Proposition 5. First we consider the case of $r > 0$. By the construction of $\hat{r}$,

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

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

Thus,

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

Lemma **??** implies that almost surely, there exists an n_0 such that $n \geq n_0$ implies

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

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

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

□

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

References

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

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

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

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

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

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

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.

Koltchinskii, V. and K. Lounici (2016, 11). Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Ann. Inst. H. Poincar Probab. Statist. 52*(4), 1976–2013.

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

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

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

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

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

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

Shen, D., H. Shen, and J. S. Marron (2013). Consistency of sparse PCA in High Dimension, Low Sample Size contexts. *Journal of Multivariate Analysis 115*, 317–333.

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, W. and J. Fan (2017, 06). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann. Statist. 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.

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