High-dimensional two-sample test under spiked covariance

Rui Wang[a], Xingzhong Xu[a,b,*]

[a] *School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081,China*
[b] *Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing 100081,China*

Abstract

This paper considers testing the means of two p-variate normal samples in high dimensional setting. The covariance matrices are assumed to be spiked, which often arises in practice. We propose a new test procedure through projection on the orthogonal complement of principal space. The asymptotic normality of the new test statistic is proved and the power function of the test is given. Theoretical and simulation results show that the new test outperforms existing methods substantially when the covariance matrices are spiked. Even when the covariance matrices are not spiked, the new test is acceptable.

Keywords: high dimension, mean test, orthogonal complement of principal space, spiked covariance

1. Introduction

Suppose that $X_{k1}, \dots, X_{kn_k}$ are independent identically distributed (i.i.d.) as $N_p(\mu_k, \Sigma_k)$, where μ_k and Σ_k are unknown, $k = 1, 2$. We consider the hypothesis testing problem:

$$H_0 : \mu_1 = \mu_2 \quad \text{vs.} \quad H_1 : \mu_1 \neq \mu_2. \tag{1}$$

*Corresponding author
Email address: `xuxz@bit.edu.cn` (Xingzhong Xu)

In this paper, the dimension p varies as n_1 and n_2 increase, i.e., high dimensional setting is adopted. Testing hypotheses (1) is important in many applications, including biology, finance and economics. Quite often, these data have strong correlations between variables. When strong correlations exist, covariance matrices are often spiked in the sense that a few eigenvalues are distinctively larger than the others. The paper is devoted to testing hypotheses (1) in high dimensional setting with spiked covariance.

If $\Sigma_1 = \Sigma_2 = \Sigma$ is unknown, a classical test for hypotheses (1) is Hotelling's T^2 test. Hotelling's test statistic is $(\bar{X}_1 - \bar{X}_2)^T S^{-1}(\bar{X}_1 - \bar{X}_2)$, where S is the pooled sample covariance matrix. However, Hotelling's test is not defined when $p \geq n_1 + n_2 - 1$. Moreover, Bai and Saranadasa (1996) showed that even if $p < n_1 + n_2 - 1$, Hotelling's test suffers from low power when p is comparable to n. Perhaps, the main reason for low power of Hotelling's test is due to that S is a poor estimator of Σ when p is large compared with $n_1 + n_2$. See Chen and Qin (2010) and the references therein. In high dimensional setting, many test statistics in the literatures are based on an estimator of $(\mu_1 - \mu_2)^T A(\mu_1 - \mu_2)$ for a given positive definite matrix A. For example, Bai and Saranadasa (1996) proposed a test based on

$$T_{BS} = \|\bar{X}_1 - \bar{X}_2\|^2 - (\frac{1}{n_1} + \frac{1}{n_2})\mathrm{tr}S,$$

which is an unbiased estimator of $\|\mu_1 - \mu_2\|^2$. Chen and Qin (2010) modified T_{BS} by removing terms $\sum_{i=1}^{n_k} X_{ki}^T X_{ki}$, $k = 1, 2$ and proposed a test based on

$$\begin{aligned} T_{CQ} &= \frac{\sum_{i\neq j}^{n_1} X_{1i}^T X_{1j}}{n_1(n_1-1)} + \frac{\sum_{i\neq j}^{n_2} X_{2i}^T X_{2j}}{n_2(n_2-1)} - 2\frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} X_{1i}^T X_{2j}}{n_1 n_2} \\ &= \|\bar{X}_1 - \bar{X}_2\|^2 - \frac{1}{n_1}\mathrm{tr}S_1 - \frac{1}{n_2}\mathrm{tr}S_2, \end{aligned}$$

where S_1 and S_2 are sample covariance matrices. Statistic T_{CQ} is also an unbiased estimator of $\|\mu_1 - \mu_2\|^2$. Choosing $A = [\mathrm{diag}(\Sigma)]^{-1}$, Srivastava and Du (2008) proposed a test based on

$$T_S = (\bar{X}_1 - \bar{X}_2)^T [\mathrm{diag}(S)]^{-1}(\bar{X}_1 - \bar{X}_2),$$

where $\mathrm{diag}(A)$ is a diagonal matrix with the same diagonal elements as A's.

As Ma et al. (2015) pointed out, however, these test procedures may not be valid if strong correlations exist, i.e., Σ is far away from diagonal matrix. For example, the assumption

$$\operatorname{tr}(\Sigma^4) = o[\operatorname{tr}^2\{(\Sigma)^2\}] \tag{2}$$

adopted by Chen and Qin (2010) can be violated when $\Sigma = (1-c)I_p + c\mathbf{1}_p\mathbf{1}_p^T$ where $-1/(p-1) < c < 1$, I_p is the p dimensional identity matrix and $\mathbf{1}_p$ is the p dimensional vector with elements 1.

Strong correlations between variables do exist in practice. In gene expression analysis, genes are correlated due to genetic regulatory networks (see Thulin (2014)). Chen et al. (2011) pointed out that in terms of pathway analysis in proteomic studies, test level can not be guaranteed if correlations are incorrectly assumed to be absent. As Ma et al. (2015) argued, there're strong correlations between different stock returns since they are all affected by the market index.

Incorrectly assuming the absence of correlation between variables will result in level inflation and low power for a test procedure. A class of test procedures is proposed through random projection (see Lopes et al. (2011), Thulin (2014) and Srivastava et al. (2015)). The idea is to project data on some random lower-dimensional subspaces. It has been shown that these procedures perform well under strong correlations.

In many situations, the correlations are determined by a small number of factors. Then Σ is spiked (see Cai et al. (2013)). The random projection methods imply that test procedures are improved when data are projected on certain subspaces. We will see that the ideal subspace is the orthogonal complement of the principal space. Fortunately, the principal space can be estimated consistently even in high dimensional setting by the theory of principal component analysis (PCA). With the assumption of spiked covariance model, we propose a new test procedure through projection on the (estimated) ideal subspace. The asymptotic distribution of the test statistic is derived and hence asymptotic power is given. We will see that the test is more powerful than T_{CQ}. Moreover, even there's no strong correlation showing up, we prove that the new test performs equally well as T_{CQ} does. The idea is also generalized to the unequal variance setting and similar results still hold.

The rest of the paper is organized as follows. In Section 2, the model and some assumptions are given. In Section 3, we propose a test procedure under $\Sigma_1 = \Sigma_2$. Section 4 exploits properties of the test. In Section 5, we generalize our test procedure to the situation of $\Sigma_1 \neq \Sigma_2$. In Section 6, simulations are carried out and a real data example is given. Section 7 contains some discussion. All the technical details are in appendix.

2. Model and assumptions

Let $\{X_{k1}, \ldots, X_{kn_k}\}$, $k = 1, 2$ be two independent random samples from p dimensional normal distribution with means μ_1 and μ_2 respectively.

Assumption 1. *Assume $p \to \infty$ as $n_1, n_2 \to \infty$. Furthermore, assume two samples are balanced, that is,*

$$\frac{n_1}{n_2} \to \xi \in (0, +\infty).$$

To characterize correlations between p variables, we consider spiked covariance structure which is adopted by PCA study. See Cai et al. (2013) and the references given there.

Assumption 2. *Suppose X_{ki}, $i = 1, 2, \ldots, n_k$ and $k = 1, 2$ are generated by following model*

$$X_{ki} = \mu_k + V_k D_k U_{ki} + Z_{ki},$$

where U_{ki}'s are i.i.d. random vectors distributed as r_k dimensional standard normal distribution with r_k fixed, $D_k = diag(\lambda_{k1}^{\frac{1}{2}}, \ldots, \lambda_{kr_k}^{\frac{1}{2}})$ with $\lambda_{k1} \geq \cdots \geq \lambda_{kr_k} > 0$, V_k is a $p \times r_k$ orthonormal matrix, Z_{ki}'s are i.i.d. random vectors distributed as $N_p(0, \sigma_k^2 I_p)$ independent of U_{ki}'s and $\sigma_k^2 > 0$, $k = 1, 2$.

Then $X_{ki} \sim N(\mu_k, \Sigma_k)$, where $\Sigma_k = \mathrm{Var}(X_{ki}) = V_k \Lambda_k V_k^T + \sigma_k^2 I_p$, $\Lambda_k = D_k^2 = diag(\lambda_{k1}, \ldots, \lambda_{kr_k})$. From Assumption 2, $V_k V_k^T$ is the orthogonal projection matrix on the column space of V_k. Let $\tilde{V}_k$ be a $p \times (p - r_k)$ full column rank orthonormal matrix orthogonal to columns of V_k. Note that $\tilde{V}_k$ may not be unique. But the projection matrix $\tilde{V}_k \tilde{V}_k^T$ is unique because $\tilde{V}_k \tilde{V}_k^T = I - V_k V_k^T$.

Assumption 3. *Assume that there is some constant κ such that*

$$\kappa\lambda \geq \lambda_{k1} \geq \cdots \geq \lambda_{kr_k} \geq \lambda > 0,$$

where $\lambda = cp^\beta$ for some $c > 0$ and $\beta \geq \frac{1}{2}$.

The restriction $\beta \geq 1/2$ is assumed in Assumption 3. If $\beta < 1/2$, condition (2) is meet and Chen and Qin (2010)'s method is valid. Hence $\beta = 1/2$ is the boundary of the scope between T_{CQ} and our new test.

Finally, let $\tau = (n_1 + n_2)/(n_1 n_2)$, S be the pooled sample covariance.

$$S = \frac{1}{n_1 + n_2 - 2} \sum_{k=1}^{2} \sum_{i=1}^{n_k} (X_{ki} - \bar{X}_k)(X_{ki} - \bar{X}_k)^T = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}, \tag{3}$$

where

$$S_k = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (X_{ki} - \bar{X}_k)(X_{ki} - \bar{X}_k)^T, \tag{4}$$

is the sample covariance of the sample k, $k = 1, 2$.

3. Methodology

In this section, we consider testing the hypotheses (1) with equal covariance matrices.

Assumption 4. *Assume $V_1 = V_2$, $D_1 = D_2$, $\Lambda_1 = \Lambda_2$, $\sigma_1 = \sigma_2$ and $r_1 = r_2$.*

To simplify notations, the subscript k of Σ_k, V_k, D_k, Λ_k, σ_k and r_k are dropped.

3.1. Motivation

In high dimensional setting, many test procedures for hypotheses (1) is based on a statistic $T(X)$ which estimates $(\mu_1 - \mu_2)^T A(\mu_1 - \mu_2)$. Usually, $T(X)$ satisfies $\mathrm{E}T = 0$ under null hypothesis and $\mathrm{E}T > 0$ under alternative. To determine the critical value, the asymptotic distribution of T need to be derived, say

$$\frac{T - \mathrm{E}T}{\sqrt{\mathrm{Var}(T)}} \xrightarrow{\mathcal{L}} N(0,1).$$

Since $\mathrm{Var}(T)$ may depend on parameters, a ratio consistent estimator $\widehat{\mathrm{Var}(T)}$ of $\mathrm{Var}(T)$ is necessary. Then the rejection region of a level α test can be defined as $T(X) \geq \widehat{\mathrm{Var}(T)}^{\frac{1}{2}} z_{1-\alpha}$ where $z_{1-\alpha}$ is the $1-\alpha$ quantile of $N(0,1)$. The asymptotic power of the test is

$$\Phi(\frac{\mathrm{E}T}{\sqrt{\mathrm{Var}(T)}} - z_{1-\alpha}).$$

Thus, a general idea to boost the power of test is to reduce the variance $\mathrm{Var}(T)$ while the mean $\mathrm{E}(T)$ varies relatively little.

Now we revisit T_{BS} and T_{CQ} which are both based on the estimation of $\|\mu_1 - \mu_2\|^2$. Denote the spectral decomposition of Σ by $\Sigma = \sum_{i=1}^p \lambda_i p_i p_i^T$ with $\lambda_1 \geq \cdots \geq \lambda_p$, where p_i, $i = 1, \ldots, p$, are orthonormal p dimensional vectors. The main body of both T_{BS} and T_{CQ} is

$$\frac{n_1 n_2}{n_1 + n_2} \sum_{i=1}^p (\bar{X}_1 - \bar{X}_2)^T p_i p_i^T (\bar{X}_1 - \bar{X}_2), \tag{5}$$

which is a sum of p independent terms. Since $\sqrt{n_1 n_2/(n_1+n_2)}(\bar{X}_1 - \bar{X}_2)$ is distributed as $N(0,\Sigma)$, the variance of $n_1 n_2/(n_1+n_2)(\bar{X}_1 - \bar{X}_2)^T p_i p_i^T (\bar{X}_1 - \bar{X}_2)$ is $2\lambda_i^2$ which decreases in i. By our previous argument, if a few leading terms with significantly large variance are removed, the modified test will be more powerful.

The argument is also supported by the likelihood ratio test. If Σ is known, the LRT is based on

$$(\bar{X}_1 - \bar{X}_2)^T \Sigma^{-1} (\bar{X}_1 - \bar{X}_2) = \frac{n_1 n_2}{n_1 + n_2} \sum_{i=1}^p \lambda_i^{-1} (\bar{X}_1 - \bar{X}_2)^T p_i p_i^T (\bar{X}_1 - \bar{X}_2). \tag{6}$$

The difference between (5) and (6) is the weights λ_i^{-1}. Unfortunately, λ_i's are hard to precisely estimate in high dimensional setting. See Bai and Silverstein (2010) for detail. Nevertheless, it's possible to identify which λ_i's are large. LRT implies the corresponding terms should have small weights, which coincides with our previous idea.

Under Assumption 2, $\Sigma = V\Lambda V^T + \sigma^2 I_p$. The eigenvalues of Σ are $\lambda_1 + \sigma^2, \ldots, \lambda_r + \sigma^2, \sigma^2, \ldots, \sigma^2$. The eigenvectors corresponding to the first r eigenvalues are columns of V. Follow our previous argument, if the principal subspace VV^T is known, we project X_{ki} on the orthogonal complement space $\tilde{V}\tilde{V}^T$ and invoke the statistic of Chen and Qin (2010). We define the following statistic

$$T_1 = \|\tilde{V}^T(\bar{X}_1 - \bar{X}_2)\|^2 - \frac{1}{n_1}\mathrm{tr}(\tilde{V}^T S_1 \tilde{V}) - \frac{1}{n_2}\mathrm{tr}(\tilde{V}^T S_2 \tilde{V}).$$

the asymptotic normality of T_1 can be obtained by Chen and Qin (2010)'s Theorem 1.

Proposition 1. *Under Assumptions 1-4, if local alternative holds, that is,* $\frac{n_1+n_2}{p}\|\mu_1 - \mu_2\|^2 \to 0$, *then we have*

$$\frac{T_1 - \|\tilde{V}(\mu_1 - \mu_2)\|^2}{\sigma^2\sqrt{2\tau^2 p}} \xrightarrow{\mathcal{L}} N(0,1).$$

Remark 1. The asymptotic variance of T_1 is of order $\tau^2 p$ while the asymptotic variance of T_{CQ} is of order $\tau^2 p^{2\beta}$ by Chen and Qin (2010)'s Theorem 1. The asymptotic variance is reduced significantly if $\beta > 1/2$ and p is sufficiently large.

3.2. New Test

Actually, T_1 is not a statistic since VV^T and $\tilde{V}\tilde{V}^T$ are unknown. Fortunately, V and $\tilde{V}$ can be estimated by $\hat{V}$ and $\hat{\tilde{V}}$, where $\hat{V}$ and $\hat{\tilde{V}}$ are the first r and last $p-r$ eigenvectors of S, respectively. These simple estimators actually reach the optimal convergence rate. See Cai et al. (2013).

Define

$$T_2 = \|\hat{\tilde{V}}^T(\bar{X}_1 - \bar{X}_2)\|^2 - \frac{1}{n_1}tr(\hat{\tilde{V}}^T S_1 \hat{\tilde{V}}) - \frac{1}{n_2}tr(\hat{\tilde{V}}^T S_2 \hat{\tilde{V}}).$$

We propose our new test statistic as

$$Q = \frac{T_2}{\hat{\sigma}^2\sqrt{2p\tau^2}}, \tag{7}$$

where $\hat{\sigma}^2$ is a ratio consistent estimator of σ^2. In next section, it will be proved that the asymptotic distribution of Q is $N(0,1)$. We reject the null hypothesis if Q is larger than the upper α quantile of $N(0,1)$.

Remark 2. Compared with random projection method, our projection is determined by the structure of S. Hence we don't project multiple times as random projection method did, which leads to reproducibility.

Remark 3. The statistic T_2 is invariant under shift transformation, that is, T_2 is invariant when adding a vector to X_{1i} and X_{2j} simultaneously: $X_{1i} \mapsto X_{1i} + \mu$ and $X_{2j} \mapsto X_{2j} + \mu$, $i = 1, \ldots, n_1$, $j = 1, \ldots, n_2$.

Remark 4. If r is an unknown positive number, a consistent estimator of r is

$$\hat{r} = \mathrm{argmax}_{l \le R} \frac{\lambda_l(S)}{\lambda_{l+1}(S)}, \tag{8}$$

where R is the maximum value of r to be tested. See Ahn and Horenstein (2013) for detail. Therefore, without loss of generality, we will assume that r is known.

Theoretical results will show that the asymptotic variance of T_2 is significantly smaller than T_{CQ}. On the other hand, the new test statistic estimates $\|\tilde{V}(\mu_1 - \mu_2)\|^2$. Hence the superiority of the new test will be established if

$$\frac{\|\tilde{V}(\mu_1 - \mu_2)\|}{\|\mu_1 - \mu_2\|} \approx 1. \tag{9}$$

Unfortunately, (9) is not always the case since there always exists some $\tilde{V}$ and $\mu_1 - \mu_2$ such that $\|\tilde{V}(\mu_1 - \mu_2)\| = 0$. However, (9) is reasonable since $\tilde{V}\tilde{V}^T$ is nearly an identity matrix in the sense that $\|I_p - \tilde{V}\tilde{V}^T\|_F^2 / \|I_p\|_F^2 = r/p \to 0$. In bayes framework, if we assume that the elements of μ_k are independently generated from certain probability distribution, it can be established that

$$\frac{\|\tilde{V}(\mu_1 - \mu_2)\|}{\|\mu_1 - \mu_2\|} \xrightarrow{P} 1.$$

Such assumption of μ_k will be used in Theorem 1.

Next we concern the problem of estimating σ^2. We give two different estimators and later we will prove their consistency using different techniques. Note that σ^2 can be written as

$$\sigma^2 = \sum_{i=r+1}^{p} \lambda_i(\Sigma) = \frac{1}{p-r} \mathrm{tr} \tilde{V}^T \Sigma \tilde{V}. \tag{10}$$

The first estimator we propose is to directly estimate (10) by

$$\hat{\sigma}^2_{(1)} = \frac{1}{p-r} \mathrm{tr} \hat{\tilde{V}}^T S \hat{\tilde{V}}.$$

The estimator $\hat{\sigma}^2_{(1)}$ uses the smallest $p-r$ eigenvalues of S. However some of the smallest $p-r$ eigenvalues may be also contaminated by Λ. Hence besides $\hat{\sigma}^2_{(1)}$, we give another estimator of σ^2:

$$\hat{\sigma}^2_{(2)} = \frac{1}{p-4r} \sum_{i=2r+1}^{p-2r} \lambda_i(S).$$

If estimator $\hat{\sigma}^2$ of σ^2 is ratio consistent, the asymptotic distribution of (7) will not change if we replace σ^2 by $\hat{\sigma}^2$ due to Slutsky's theorem.

4. Theoretical results

In this section, we derive some properties of the new test statistic. Our main results require the following relationship of n_1, n_2 and p.

Assumption 5. *We assume*

$$\frac{\sqrt{p}}{n_1 + n_2} \to 0.$$

First we establish $\hat{\sigma}^2_{(1)}$ and $\hat{\sigma}^2_{(2)}$ are consistent estimators of σ^2.

Proposition 2. *Under Assumptions 1-4, $\hat{\sigma}^2_{(2)}$ is consistent. If we further assume 5, then $\hat{\sigma}^2_{(1)}$ is consistent.*

Remark 5. Compared with $\hat{\sigma}^2_{(1)}$, $\hat{\sigma}^2_{(2)}$ uses less information. Proposition 2, however, shows that $\hat{\sigma}^2_{(2)}$ is consistent without requirement of the relationship between n and p, while $\hat{\sigma}^2_{(1)}$ requires Assumption 5.

Next we derive the asymptotic normality of the new test statistic. Consider the case when the eigenvalues of Σ is bounded, i.e., there is no clear correlation between variables. In many practical problems, the alternative is 'dense', i.e., under H_1 the signals in $\mu_1 - \mu_2$ spread out over a large number of co-ordinates. See Cai et al. (2014). We characterize 'dense' alternative from bayes framework and assume that elements of μ_k are independently generated from normal distribution. Next theorem shows that the power of our new test is asymptotically the same as Chen and Qin (2010)'s test in this case.

Theorem 1. *Assume $X_{ki} \sim N(\mu_k, \Sigma)$, $i = 1, \ldots, n_k$, $k = 1, 2$. Suppose that assumption 1 holds, $0 < c \leq \lambda_p(\Sigma) \leq \lambda_1(\Sigma) \leq C < \infty$ where c and C are constant, each element of μ_k is independently generated by $N(0, (n_k\sqrt{p})^{-1}\psi)$ for $k = 1, 2$, ψ is a constant and $\hat{r}$ is estimated by (8). If $p \to \infty$ as $n \to \infty$ and $p = o((n_1 + n_2)^2)$, then we have*

$$\frac{T_2 - \|\mu_1 - \mu_2\|^2}{\sqrt{2\tau^2 \mathrm{tr}\Sigma^2}} \xrightarrow{\mathcal{L}} N(0, 1).$$

Next we establish the asymptotic normality of the new test statistic under spiked covariance model. Our first step is to prove the asymptotic normality under null hypothesis.

Theorem 2. *Under Assumptions 1-5 and $\sqrt{p}/(n_1 + n_2) \to 0$, if the null hypothesis holds, then*

$$\frac{T_2}{\sigma^2\sqrt{2p\tau^2}} \xrightarrow{\mathcal{L}} N(0, 1).$$

Remark 6. The critical value of the new test can be determined by Theorem 2, that is, reject when $T_2/(\sigma^2\sqrt{2p\tau^2}) > z_{1-\alpha}$.

Then we generalize the Theorem 2 to the case of local alternative.

Theorem 3. *Under Assumptions 1-5, if the local alternative holds, that is,*

$$\frac{(n_1 + n_2)}{\sqrt{p}}\|\mu_1 - \mu_2\|^2 = O(1),$$

then

$$\frac{T_2 - \|\tilde{V}^T(\mu_1 - \mu_2)\|^2}{\sigma^2\sqrt{2p\tau^2}} \xrightarrow{\mathcal{L}} N(0, 1).$$

By Proposition 2 and Theorem 3, the power of our new test can be obtained immediately.

Theorem 4. *Under Assumptions 1-5, if we reject the null hypothesis when Q is larger than $1-\alpha$ quantile of $N(0,1)$, then the power of our test is*

$$\Phi\Big(-\Phi^{-1}(1-\alpha)+\frac{\|\tilde{V}(\mu_1-\mu_2)\|^2}{\sigma^2\sqrt{2\tau^2 p}}\Big).$$

Remark 7. The power of T_{CQ} is of the form

$$\Phi\Big(-\Phi^{-1}(1-\alpha)+\frac{\|\mu_1-\mu_2\|^2}{\sqrt{2\tau^2\mathrm{tr}\Sigma^2}}\Big).$$

If $\|\tilde{V}\mu\|/\|\mu\|\to 1$, the relative efficiency of our test with respect to Chen's test is

$$\sqrt{\frac{\mathrm{tr}\Sigma^2}{(p-r)\sigma^4}}\sim p^{\beta-1/2}.$$

5. Unequal Variance

In this section, we concern the situation with unequal covariance matrices. Assume $\{X_{11},\ldots,X_{1n_1}\}$ and $\{X_{21},\ldots,X_{2n_2}\}$ are both generated from the model in Assumption 2. Denote by $\hat{V}_k$ the first r_k eigenvectors of S_k for $k=1,2$. With a little abuse of notation, let VV^T be the projection on the sum of column spaces of V_1 and V_2, that is,

$$VV^T=(V_1,V_2)\big((V_1,V_2)^T(V_1,V_2)\big)^+(V_1,V_2)^T.$$

where A^+ is the Moore-Penrose inverse of a matrix A. Similarly, let $\hat{V}\hat{V}^T$ be the projection matrix on the sum of column spaces of $\hat{V}_1$ and $\hat{V}_2$. We define $\tilde{V}\tilde{V}^T=I_p-VV^T$ and $\hat{\tilde{V}}\hat{\tilde{V}}^T=I_p-\hat{V}\hat{V}^T$.

The previous statistic can not be directly used since the principal subspace is different for X_{1i} and X_{2j}. The idea here is to remove all large variance terms from T_{CQ} by projecting data on the space $\tilde{V}\tilde{V}^T$. Thus, we propose a new test statistic as

$$T_3=\|\hat{\tilde{V}}^T(\bar{X}_1-\bar{X}_2)\|^2-\frac{1}{n_1}\mathrm{tr}(\hat{\tilde{V}}^TS_1\hat{\tilde{V}})-\frac{1}{n_2}\mathrm{tr}(\hat{\tilde{V}}^TS_2\hat{\tilde{V}}).$$

The theoretical results are parallel to those in equal variance setting.

Theorem 5. *Under Assumptions 1-3 and 5, if the null hypothesis holds, then*

$$\frac{T_3}{\sqrt{\sigma_n^2}} \xrightarrow{\mathcal{L}} N(0,1),$$

where $\sigma_n^2 = \frac{2(p-r_1-r_2)}{n_1(n_1-1)}\sigma_1^4 + \frac{2(p-r_1-r_2)}{n_2(n_2-1)}\sigma_2^4 + \frac{4(p-r_1-r_2)}{n_1 n_2}\sigma_1^2\sigma_2^2$.

Remark 8. Even if $\hat{\tilde{V}}_k\hat{\tilde{V}}_k^T$ is an consistent estimator of $\tilde{V}_k\tilde{V}_k^T$ for $k = 1, 2$, $\hat{\tilde{V}}\hat{\tilde{V}}^T$ may not be an consistent estimator of $\tilde{V}\tilde{V}^T$. Nevertheless, the asymptotic normality still holds.

Theorem 6. *Under Assumptions 1-3 and 5, if the local alternative holds, that is,*

$$\frac{n_1+n_2}{\sqrt{p}}\|\mu_1-\mu_2\|^2 = O(1),$$

then

$$\frac{T_3 - \|\tilde{V}^T(\mu_1-\mu_2)\|^2}{\sqrt{\sigma_n^2}} \xrightarrow{\mathcal{L}} N(0,1).$$

σ_n^2 can be estimated by ratio consistent estimator of σ_k^2 for $k = 1, 2$. Thus, if n_k and p are large and $\sqrt{p}/(n_1+n_2)$ is small, we reject when $T_3/\sqrt{\hat{\sigma}_n^2} > z_{1-\alpha}$.

6. Numerical studies

6.1. Simulation results

Our simulation study focus on equal variance case. We generate X_{ki} by the model in Assumption 2, where each element of U_{ki} and Z_{ki} are generated from $N(0,1)$. V is a random orthonormal matrix. We generate λ_i as p^β plus a random error from $U(0,1)$.

First we simulate the level of the new test. The nominal level $\alpha = 0.05$ and we set $r = 2$. Samples are repeatedly generated 1000 times to calculate empirical level. For comparison, we also give corresponding 'oracle' level which is calculated by 'statistic' $T_1/(\sigma^2\sqrt{2p\tau^2})$ whose asymptotic normality can be guaranteed by Theorem 1 in Chen and Qin (2010). The results are listed in Table 1. From the results, we can find that for small n and p, even oracle level

Table 1: Test level simulation

		β=0.5		β=1		β=2	
n	p	NEW	ORACLE	NEW	ORACLE	NEW	ORACLE
300	200	0.075	0.062	0.079	0.062	0.074	0.070
300	400	0.074	0.065	0.061	0.044	0.046	0.040
300	600	0.058	0.041	0.070	0.052	0.071	0.055
300	800	0.066	0.047	0.071	0.052	0.062	0.048
600	200	0.061	0.055	0.052	0.051	0.058	0.056
600	400	0.051	0.048	0.051	0.042	0.059	0.051
600	600	0.061	0.058	0.056	0.054	0.051	0.047
600	800	0.053	0.046	0.060	0.050	0.056	0.048

is not satisfied. Level of the new test is a little inflated compared with oracle level and it performs better when n is larger.

Then we simulate the empirical power of our test and Chen and Qin (2010)'s test. The simulation results of Ma et al. (2015) have showed that the level of the Chen and Qin (2010)'s test can't be guaranteed when covariance is spiked. To be fair, we use permutation method to compute critical value. The validity of permutation method can be found in Lehmann and Romano (2005)'s Example 15.2.2. We plot the empirical power versus $\|\mu_1 - \mu_2\|$ when other parameters hold constant. The results are illustrated in figure 1. From the results, we can find that when Σ is spiked, the new test outperforms T_{CQ} substantially; when Σ is not spiked, the new test and T_{CQ} are comparable.

6.2. Real data analysis

In this section, we study the same practical problem as Ma et al. (2015) did. That is testing whether Monday stock returns are equal to those of other trading days on average. Define an observation be the log return of stocks in a day. Hence p is the total number of stocks. Let sample 1 and sample 2 be the observations on Monday and the other trading days, respectively. Then we

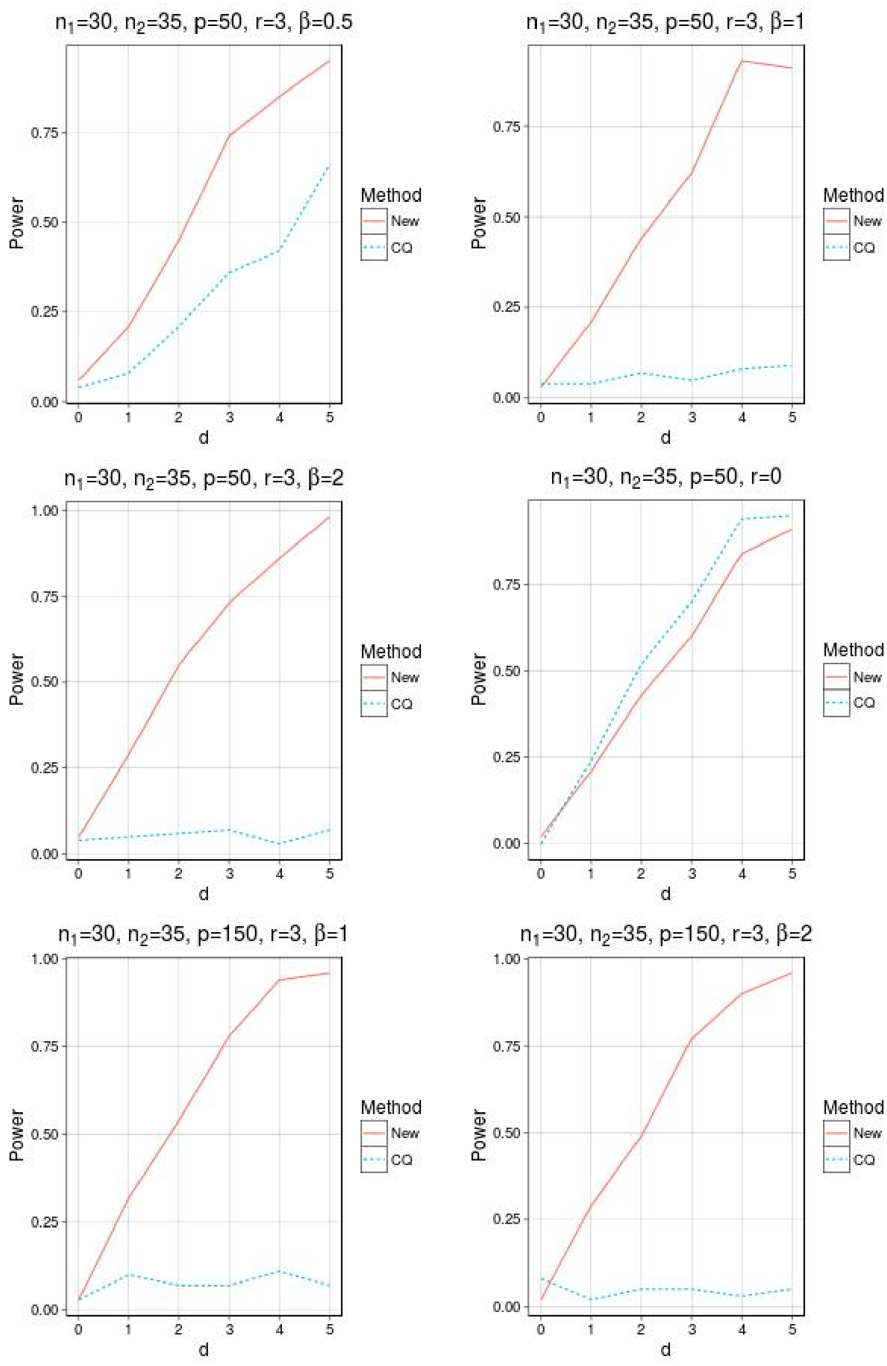

Figure 1: Empirical power simulation. α is set to be 0.05. d is proportional to $\|\mu_1 - \mu_2\|^2$. For each simulation, we do 50 permutations to determine critical value. We generate 100 independent samples to compute empirical power.

would like to test $H_0 : \mu_1 = \mu_2$ v.s. $H_1 : \mu_1 \neq \mu_2$. We collected the data of $p = 710$ stocks of China from 01/04/2013 to 12/31/2014. There are total $n_1 = 95$ Monday and $n_2 = 388$ other trading days.

We assume $\Sigma_1 = \Sigma_2$. The first eigenvaule of S is 0.14, which is significantly larger than the others. In fact, the second eigenvalue is 0.02. Hence there's clearly a spiked eigenvalue. We set $r = 1$ and perform our new test. The p value is 0.149, which is obtained by 1000 permutations. Hence, the null hypothesis can not be rejected for $\alpha = 0.05$. We draw the same conclusion as Ma et al. (2015).

7. Conclusion remark

This paper is concerned with the problem of testing the equality of means in the setting of high dimension and spiked covariance. We removes big variance terms from T_{CQ} and it's power is boosted substantially. The asymptotic normality of the new statistic is proved and the asymptotic power is given.

In another paper, Zhao and Xu (2016) proved their test statistic can be written in the form of projection. Their simulation results showed that their test performs well under strong correlations. Our work partially explains why their test performs well although the projections are slightly different.

Spiked covariance is an important correlation pattern and has been widely studied in terms of PCA. In PCA, authors focus on the principal subspace. However, our work shows that in some circumstance, the complement of principal subspace is more useful.

Our theoretical results rely on the assumption $\sqrt{p}/(n_1 + n_2) \to 0$. In the situation of small sample or very large p, the critical value of the new test can be determined by permutation method. Our simulation shows that the new test still performs well. It remains a theoretical interest to derive the power function in these situations.

Appendix

We denote by $\|\cdot\|$ and $\|\cdot\|_F$ the operator and Frobenius norm of matrix, separately.

Lemma 1 (Weyl's inequality)**.** *Let H and P be two symmetric matrices and $M = H + P$. If $j + k - n \geq i \geq r + s - 1$, we have*

$$\lambda_j(H) + \lambda_k(P) \leq \lambda_i(M) \leq \lambda_r(H) + \lambda_s(P).$$

Corollary 7. *Let H and P be two symmetric matrices and $M = H + P$. If $\mathrm{rank}(P) < k$, then*

$$\lambda_k(M) \leq \lambda_1(H).$$

Lemma 2. *Let H and P be two $p \times p$ positive semi-definite matrices and $M = H + P$. Suppose $\mathrm{Rank}(H) = d_1$, $\mathrm{Rank}(P) = d_2$, where d_1 and d_2 may depend on n. Assume $\mathrm{Rank}(M) = d_1 + d_2$, $\lambda_1(M) \to 1$ and $\lambda_{d_1+d_2}(M) \to 1$, then we have*

$$\lambda_1(H) \to 1, \quad \lambda_{d_1}(H) \to 1, \tag{11}$$

and

$$\lambda_1(P) \to 1, \quad \lambda_{d_2}(P) \to 1, \tag{12}$$

Remark 9. This generalizes the Cochran's theorem (See Anderson).

Proof. By Weyl's inequality, $\lambda_1(H) \leq \lambda_1(M) \to 1$ and $\lambda_{d_1}(H) \geq \lambda_{d_1+d_2}(M) \to 1$. Hence (11) holds. And (12) follows in the same way. □

Lemma 3 (Convergence rate of principal space estimation)**.** *Under the model in Assumption 4 with condition (1), assume $\Sigma_1 = \Sigma_2$. Suppose $p \to \infty$ as $n \to \infty$ and $\frac{\sqrt{p}}{n_1+n_2} \to 0$. We have*

$$E\|\hat{V}\hat{V}^T - VV^T\|_F^2 = O(\frac{p}{p^\beta(n_1+n_2)}).$$

Proof. By theorem 5 of Cai et al. (2013), sample principal subspace $\hat{V}\hat{V}^T$ is a minimax rate estimator of VV^T, namely, it reaches the minimax convergence rate

$$\inf_{\hat{V}} \sup_{\Sigma} E\|\hat{V}\hat{V}^T - VV^T\|_F^2 \asymp r \wedge (p-r) \wedge \frac{r(p-r)}{(n_1+n_2-2)h(\lambda)} \tag{13}$$

as long as the right hand side tends to 0. Here $h(\lambda) = \frac{\lambda^2}{\lambda+1}$, $a_n \asymp b_n$ represents $a_n \geq cb_n$ and $a_n \leq Cb_n$ for some positive c, C for every n. In model of Assumption 2, r is fixed, $\lambda = cp^\beta$. It's obvious that the right hand side of (13) is of order $\frac{p^{1-\beta}}{n_1+n_2}$. We note that it is assumed $\beta \geq \frac{1}{2}$ in model of Assumption 2, together with $\frac{\sqrt{p}}{n_1+n_2} \to 0$ we have $\frac{p^{1-\beta}}{n_1+n_2} \to 0$. Hence $\hat{V}\hat{V}^T$ reaches the optimal rate.

□

Lemma 4. *Suppose that W_n is a $p \times p$ matrix distributed as Wishart$_p(n, I_p)$, $p/n \to c \in [0, +\infty]$. Then as $n \to \infty$,*

$$\lambda_1(W_n) = O_P(\max(n, p)).$$

Proof. If p is bounded as $n \to \infty$, then p is fixed when n is sufficiently large and the lemma can be established by weak law of large numbers. Otherwise, if $p \to \infty$ as $n \to \infty$, the lemma is a direct corollary of Theorem 1.1 of Johnstone (2001) and Theorem 2 of El Karoui (2003). □

Lemma 5. *Suppose X_n is a k_n dimensional standard normal random vector and A_n is a $k_n \times k_n$ symmetric matrix. Then a necessary and sufficient condition for*

$$\frac{X_n^T A_n X_n - \mathrm{E} X_n^T A_n X_n}{[\mathrm{Var}(X_n^T A_n X_n)]^{1/2}} \xrightarrow{\mathcal{L}} N(0,1) \tag{14}$$

is that

$$\frac{\lambda_{\max}(A_n^2)}{\mathrm{tr}(A_n^2)} \to 0. \tag{15}$$

Remark 10. This lemma is the Example 5.1 of Jiming Jiang (1996). Here we give a proof by ch.f.

Proof. Let $\lambda_1(A_n) \geq \cdots \geq \lambda_{k_n}(A_n)$ be the eigenvalues of A_n, then

$$\frac{X_n^T A_n X_n - \mathrm{E} X_n^T A_n X_n}{[\mathrm{Var}(X_n^T A_n X_n)]^{1/2}} = \sum_{i=1}^{k_n} \frac{\lambda_i(A_n)}{\left[2\mathrm{tr}(A_n^2)\right]^{1/2}} (Z_{ni}^2 - 1), \tag{16}$$

where Z_{ni}'s $(i = 1, \ldots, k_n)$ are independent standard normal random variables.

If 15 holds, then

$$\begin{aligned}
&\sum_{i=1}^{k_n} \mathrm{E}\Big[\frac{\lambda_i^2(A_n)}{2\mathrm{tr}(A_n^2)}(Z_{ni}^2-1)^2\Big\{\frac{\lambda_i^2(A_n)}{2\mathrm{tr}(A_n^2)}(Z_{ni}^2-1)^2 \ge \epsilon\Big\}\Big] \\
\le&\sum_{i=1}^{k_n} \frac{\lambda_i^2(A_n)}{2\mathrm{tr}(A_n^2)}\mathrm{E}\Big[(Z_{n1}^2-1)^2\Big\{\frac{\lambda_{\max}(A_n^2)}{2\mathrm{tr}(A_n^2)}(Z_{n1}^2-1)^2 \ge \epsilon\Big\}\Big] \\
=&\frac{1}{2}\mathrm{E}\Big[(Z_{n1}^2-1)^2\Big\{\frac{\lambda_{\max}(A_n^2)}{2\mathrm{tr}(A_n^2)}(Z_{n1}^2-1)^2 \ge \epsilon\Big\}\Big] \to 0.
\end{aligned}$$

Hence 14 follows by Lindeberg theorem.

Conversely, if 14 holds, we will prove that there is a subsequence of $\{n\}$ along which 15 holds. Then 15 will hold by a standard contradiction argument.

Denote $c_{ni} = \lambda_i(A_n)/\big[2\mathrm{tr}(A_n^2)\big]^{1/2}$ $(i = 1, \ldots, k_n)$, we have $c_{ni} \in [-\sqrt{2}/2, \sqrt{2}/2]$. Since 14 holds, the characteristic function of $\sum_{i=1}^{k_n} c_{ni}(Z_{ni}^2-1)$ converges to $\exp(-t^2/2)$ for every t. For $t \in (-1, 1)$, we have

$$\begin{aligned}
&\log \mathrm{E}\exp\big(it\sum_{j=1}^{k_n} c_{nj}(Z_{nj}^2-1)\big) \\
=&-i(\sum_{j=1}^{k_n} c_{nj})t - \frac{1}{2}\sum_{j=1}^{k_n}\log(1-i2c_{nj}t) \\
=&-i(\sum_{j=1}^{k_n} c_{nj})t + \frac{1}{2}\sum_{j=1}^{k_n}\sum_{l=1}^{+\infty}\frac{1}{l}(i2c_{nj}t)^l \\
=&-i(\sum_{j=1}^{k_n} c_{nj})t + \frac{1}{2}\sum_{l=1}^{+\infty}\Big[\sum_{j=1}^{k_n}(c_{nj})^l\Big]\frac{1}{l}(i2t)^l \\
=&-\frac{1}{2}t^2 + \frac{1}{2}\sum_{l=3}^{+\infty}\Big[\sum_{j=1}^{k_n}(c_{nj})^l\Big]\frac{1}{l}(i2t)^l.
\end{aligned}$$

Denote $b_{nl} = \sum_{j=1}^{k_n}(c_{nj})^l$, $n = 1, 2, \cdots$ and $l = 3, 4, \cdots$. For $l \ge 3$, $\big|\sum_{j=1}^{k_n}(c_{nj})^l\big| \le \big|\sum_{j=1}^{k_n}(c_{nj})^2\big| = 1/2$. By Helly's selection theorem, there's a subsequence of $\{n\}$ along which $\lim_{n\to\infty} b_{nl} = b_l$ exists for every l. Apply dominated convergence theorem to this subsequence we have $\log \mathrm{E}\exp\big(it\sum_{j=1}^{k_n} c_{nj}(Z_{nj}^2-1)\big) \to -\frac{1}{2}t^2 + \frac{1}{2}\sum_{l=3}^{+\infty} b_l \frac{1}{l}(i2t)^l$ for $t \in (-1/2, 1/2)$. By the property of power series, we have $b_l = 0$ for $l \ge 3$. Then 15 follows by noting that $b_{n4} \ge \max_j (c_{nj})^4$. $\square$

Theorem 8. *Suppose X_{ki}'s are distributed as $N_p(0, \Sigma)$ $(k = 1, 2,\ i =$*

$1, 2, \ldots, n_k$), $n_1/n_2 \to \psi \in (0, +\infty)$. *Then a necessary and sufficient condition for*

$$\frac{T_{CQ} - \mathrm{E}T_{CQ}}{\left[\mathrm{Var}(T_{CQ})\right]^{1/2}} \xrightarrow{L} N(0,1) \tag{17}$$

is that

$$\frac{\lambda_{\max}(\Sigma)}{\left[\mathrm{tr}\Sigma^2\right]^{1/2}} \to 0. \tag{18}$$

Remark 11. The condition 18 is equivalent to condition 2.

Proof. Let $X_{ki} = \Sigma^{1/2} Z_{ki}$ with Z_{ki} distributed as $N_p(0, I_p)$. Denote $Z = [Z_{11}, \ldots, Z_{1n_1}, Z_{21}, \ldots, Z_{2n_2}]^T$. Then T_{CQ} is a quadratic form of Z:

$$T_{CQ} = Z^T \big(B_n \otimes \Sigma\big) Z, \tag{19}$$

where

$$B_n = \begin{pmatrix} \frac{1}{n_1(n_1-1)}(n_1 \alpha\alpha^T - I_{n_1}) & -\frac{1}{\sqrt{n_1 n_2}}\alpha\beta^T \\ -\frac{1}{\sqrt{n_1 n_2}}\beta\alpha^T & \frac{1}{n_2(n_2-1)}(n_2\beta\beta^T - I_{n_2}) \end{pmatrix}, \tag{20}$$

α is an n_1 dimensional vector with all elements equal to $1/\sqrt{n_1}$ and β is an n_2 dimensional vector with all elements equal to $1/\sqrt{n_2}$.

By direct calculation, it can be seen that B_n's eigenvalues are $-1/n_1(n_1-1)$, $-1/n_2(n_2-1)$, $(n_1+n_2)/n_1 n_2$ and 0 with multiplicities $n_1 - 1$, $n_2 - 1$, 1 and 1 respectively. The eigenspace corresponding to $-1/n_1(n_1-1)$ is

$$\{(\eta^T, \underbrace{0, \ldots, 0}_{n_2})^T | \eta \text{ is of } n_1 \text{ dimension and } \eta^T \alpha = 0\}.$$

The eigenspace corresponding to $-1/n_2(n_2-1)$ is

$$\{(\underbrace{0, \ldots, 0}_{n_1}, \eta^T)^T | \eta \text{ is of } n_2 \text{ dimension and } \eta^T \beta = 0\}.$$

The eigenvector corresponding to $(n_1+n_2)/n_1 n_2$ is

$$\Big(-\sqrt{\frac{n_2}{n_1+n_2}}\alpha^T, \sqrt{\frac{n_1}{n_1+n_2}}\beta^T\Big)^T.$$

The eigenvector corresponding to 0 is

$$\Big(\sqrt{\frac{n_1}{n_1+n_2}}\alpha^T, \sqrt{\frac{n_2}{n_1+n_2}}\beta^T\Big)^T.$$

It follows that

$$\mathrm{tr}(B_n \otimes \Sigma)^2 = \mathrm{tr}(B_n^2)\mathrm{tr}\Sigma^2 = (\frac{1}{n_1(n_1-1)} + \frac{1}{n_1(n_1-1)} + \frac{2}{n_1 n_2})\mathrm{tr}\Sigma^2.$$

And

$$\lambda_{\max}\Big((B_n \otimes \Sigma)^2\Big) = \lambda_{\max}(B_n^2)\lambda_{\max}(\Sigma^2) = \Big(\frac{1}{n_1} + \frac{1}{n_2}\Big)^2 \lambda_{\max}(\Sigma^2).$$

The theorem follows by previous lemma. $\square$

8. PCA Theory

We give some PCA theory here. Compared with existing results, we impose less assumptions since our main task is to obtain the properties of principal space.

Assumption 6. *Suppose $X_i = \mu + U\Lambda^{1/2}Z_i$, $i = 1, \ldots, n$, where Z_i is a random vector with p indepent entries with $\mathrm{E}Z_{ij} = 0$ and $\mathrm{Var}Z_{ij} = 1$, $i = 1, \ldots, n$, $j = 1, \ldots, p$, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ with $\lambda_1 \geq \cdots \geq \lambda_p$ and U is a p dimensional orthogonal matrix. Suppose $\lambda_{r+1} \asymp \lambda_p \asymp 1$.*

Denote by $S = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$ the sample covariance matrix, $S = \hat{U}\hat{\Lambda}\hat{U}^T$ the spectral decomposition of S where $\hat{\Lambda} = \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_p)$ and $\hat{U}$ is a orthogonal matrix.

Let u_i be the ith column of U, $1 \leq i \leq p$. And denote $U = (V, \tilde{V})$, where V is the first r columns of U and $\tilde{V}$ is the larst $p - r$ columns of U. Similarly define $\hat{u}_i$, $\hat{V}$ and $\hat{\tilde{V}}$.

The proof is similar to Dan Shen's paper.

Let $Z \stackrel{def}{=} (Z_1, \ldots, Z_n) \stackrel{def}{=} (\tilde{Z}_1^T, \ldots, \tilde{Z}_p^T)^T$ and $\tilde{Z}_{(1)} = (\tilde{Z}_1^T, \ldots, \tilde{Z}_r^T)^T$, $\tilde{Z}_{(2)} = (\tilde{Z}_{r+1}^T, \ldots, \tilde{Z}_p^T)^T$. Let $\Lambda_{(1)} = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$ and $\Lambda_{(2)} = \mathrm{diag}(\lambda_{r+1}, \ldots, \lambda_p)$. Define $\hat{\Lambda}_{(1)}$ and $\hat{\Lambda}_{(2)}$ in a similar way. Deote by J the $p \times p$ matrix with all elements equal to 1. Then

$$S = \frac{1}{n}U\Lambda^{1/2}Z(I - \frac{1}{n}J)Z^T\Lambda^{1/2}U^T.$$

The positive eigenvalues of S are the same as those of $S^* = \frac{1}{n}(I - \frac{1}{n}J)Z^T\Lambda Z(I - \frac{1}{n}J)$, the dual matrix of S.

Theorem 9. *Suppose Assumption 6 holds and $p/n \to \infty$. For $1 \le i \le r$, we have:*

If $\frac{p}{n\lambda_i} \to 0$, then $\frac{\hat{\lambda}_i}{\lambda_i} \xrightarrow{a.s.} 1$. If $\frac{p}{n\lambda_i} \to \infty$, then $\hat{\lambda}_i \asymp \frac{p}{n}$ almost surely.

Proof. Let

$$S^* = \frac{1}{n}(I-\frac{1}{n}J)\tilde{Z}_{(1)}^T\Lambda_{(1)}\tilde{Z}_{(1)}(I-\frac{1}{n}J)+\frac{1}{n}(I-\frac{1}{n}J)\tilde{Z}_{(2)}^T\Lambda_{(2)}\tilde{Z}_{(2)}(I-\frac{1}{n}J) \stackrel{def}{=} A+B.$$

For $1 \le i \le r$, we prove the ith sample eigenvalue is consistent. By Corollary 7, we have

$$\begin{aligned}\frac{\lambda_i(A)}{\lambda_i} &\le \frac{1}{\lambda_i}\lambda_{\max}\Big(\frac{1}{n}(I-\frac{1}{n}J)\tilde{Z}_{(1)}^T\mathrm{diag}(\underbrace{0,\dots,0}_{i-1},\underbrace{\lambda_i,\dots,\lambda_i}_{r-i+1})\tilde{Z}_{(1)}(I-\frac{1}{n}J)\Big)\\ &\le \lambda_{\max}\Big(\frac{1}{n}\tilde{Z}_{(1)}(I-\frac{1}{n}J)\tilde{Z}_{(1)}^T\Big)\end{aligned}$$

But $\frac{1}{n}\tilde{Z}_{(1)}(I-\frac{1}{n}J)\tilde{Z}_{(1)}^T \xrightarrow{a.s.} I_r$ by law of large number. It follows that the right hand side tends to 1 almost surely. On the other hand, by Weyl's inequility, we have

$$\frac{\lambda_i(A)}{\lambda_i} \ge \frac{1}{\lambda_i}\lambda_i\Big(\frac{1}{n}(I-\frac{1}{n}J)\tilde{Z}_{(1)}^T\mathrm{diag}(\underbrace{\lambda_i,\dots,\lambda_i}_{i},\underbrace{0,\dots,0}_{r-i})\tilde{Z}_{(1)}(I-\frac{1}{n}J)\Big).$$

The right hand side tends to 1 almost surely. As a result, $\lambda_i(A)/\lambda_i$ tends to 1 almost surely.

By Weyl's inequality, we have that

$$\frac{\max(\lambda_i(A),\lambda_i(B))}{\lambda_i} \le \frac{\lambda_i(S)}{\lambda_i} \le \frac{\lambda_i(A)}{\lambda_i} + \frac{\lambda_{\max}(B)}{\lambda_i}. \tag{21}$$

By Bai Yin's law, $\frac{n}{p}\lambda_i(B) \asymp 1$ almost surely for $1 \le i \le n-1$. It follows that

$$\frac{\lambda_i(B)}{\lambda_i} \asymp \frac{p}{n\lambda_i} \quad \text{and} \quad \frac{\lambda_{\max}(B)}{\lambda_i} \asymp \frac{p}{n\lambda_i}, \tag{22}$$

where $1 \le i \le n-1$. The theorem follows from (21) and (22). ∎

Theorem 10. *Suppose Assumption 6 holds and $p/n \to \infty$. If $\frac{p}{n\lambda_r} \to 0$, then almost surely we have*

$$\|\hat{V}\hat{V}^T - VV^T\|_F^2 \asymp \frac{p}{n\lambda_r}. \tag{23}$$

If $\frac{p}{n\lambda_r} \to \infty$, *then*

$$r - \frac{1}{2}\|\hat{V}\hat{V}^T - VV^T\|_F^2 = O_{a.s.}(\frac{n\lambda_1}{p}). \tag{24}$$

Proof. Since

$$S = \hat{U}\hat{\Lambda}\hat{U}^T = \frac{1}{n}U\Lambda^{1/2}Z(I - \frac{1}{n}J)Z^T\Lambda^{1/2}U^T,$$

we have

$$\Lambda^{-1/2}U^T\hat{U}\hat{\Lambda}\hat{U}^TU\Lambda^{-1/2} = \frac{1}{n}Z(I - \frac{1}{n}J)Z^T. \tag{25}$$

First, we prove (23). It follows from (25) that

$$\Lambda_{(2)}^{-1/2}\tilde{V}^T\hat{U}\hat{\Lambda}\hat{U}^T\tilde{V}\Lambda_{(2)}^{-1/2} = \frac{1}{n}\tilde{Z}_{(2)}(I - \frac{1}{n}J)\tilde{Z}_{(2)}^T. \tag{26}$$

The left hand side of (26) equals to $C+D$, where $C = \Lambda_{(2)}^{-1/2}\tilde{V}^T\hat{V}\hat{\Lambda}_{(1)}\hat{V}^T\tilde{V}\Lambda_{(2)}^{-1/2}$ and $D = \Lambda_{(2)}^{-1/2}\tilde{V}^T\hat{\tilde{V}}\hat{\Lambda}_{(2)}\hat{\tilde{V}}^T\tilde{V}\Lambda_{(2)}^{-1/2}$. Note that $\text{Rank}(C) = r$, $\text{Rank}(D) = n-1-r$ and $\text{Rank}(C+D) = n-1$. By Bai Yin's law, we have that

$$\lambda_1\big(\frac{1}{p}\tilde{Z}_{(2)}(I - \frac{1}{n}J)\tilde{Z}_{(2)}^T\big) \to 1, \quad \lambda_{n-1}\big(\frac{1}{p}\tilde{Z}_{(2)}(I - \frac{1}{n}J)\tilde{Z}_{(2)}^T\big) \to 1 \quad a.s..$$

By Lemma 2, $\lambda_1(C) \xrightarrow{a.s.} 1$ and $\lambda_r(C) \xrightarrow{a.s.} 1$. It follows that

$$\frac{n}{p}\hat{\Lambda}_{(1)}^{1/2}\hat{V}^T\tilde{V}\Lambda_{(2)}^{-1}\tilde{V}^T\hat{V}\hat{\Lambda}_{(1)}^{1/2} \xrightarrow{a.s.} I_r. \tag{27}$$

When $\frac{p}{n\lambda_r} \to 0$, $\hat{\lambda}_i$'s are ratio consistent for $1 \le i \le r$. That is, $\Lambda_{(1)}^{-1}\hat{\Lambda}_{(1)} \to I_r$ almost surely. Then it follows from (27) that

$$\frac{n}{p}\Lambda_{(1)}^{1/2}\hat{V}^T\tilde{V}\Lambda_{(2)}^{-1}\tilde{V}^T\hat{V}\Lambda_{(1)}^{1/2} \xrightarrow{a.s.} I_r. \tag{28}$$

Notice that

$$\frac{n}{p}\text{tr}\big(\Lambda_{(1)}^{1/2}\hat{V}^T\tilde{V}\Lambda_{(2)}^{-1}\tilde{V}^T\hat{V}\Lambda_{(1)}^{1/2}\big) \ge \frac{n}{p}\lambda_r\text{tr}\big(\hat{V}^T\tilde{V}\Lambda_{(2)}^{-1}\tilde{V}^T\hat{V}\big) \ge \frac{n}{p}e_r^T\hat{\Lambda}_{(1)}^{1/2}\hat{V}^T\tilde{V}\Lambda_{(2)}^{-1}\tilde{V}^T\tilde{V}\hat{\Lambda}_{(1)}^{1/2}e_1$$

where $e_r = (\underbrace{0,\ldots,0}_{r-1},1)$. It follows that the medium term is bounded above and below asymptotically. Notice that

$$\frac{n}{p}\lambda_r\text{tr}\big(\hat{V}^T\tilde{V}\Lambda_{(2)}^{-1}\tilde{V}^T\hat{V}\big) \asymp \frac{n}{p}\lambda_r\text{tr}\big(\hat{V}^T\tilde{V}\tilde{V}^T\hat{V}\big) = \frac{n}{p}\lambda_r\frac{1}{2}\|VV^T - \hat{V}\hat{V}^T\|_F^2.$$

Therefore $\|VV^T - \hat{V}\hat{V}^T\|_F^2 \asymp \frac{p}{n\lambda_r}$ almost surely.

Then we prove (24). It follows from (25) that

$$\Lambda_{(1)}^{-1/2} V^T \hat{U}\hat{\Lambda}\hat{U}^T V \Lambda_{(1)}^{-1/2} = \frac{1}{n}\tilde{Z}_{(1)}(I - \frac{1}{n}J)\tilde{Z}_{(1)}^T \xrightarrow{a.s.} I_r. \tag{29}$$

But

$$\begin{aligned} \mathrm{tr}\big(\Lambda_{(1)}^{-1/2} V^T \hat{U}\hat{\Lambda}\hat{U}^T V \Lambda_{(1)}^{-1/2}\big) &\geq \mathrm{tr}\big(\Lambda_{(1)}^{-1/2} V^T \hat{V}\hat{\Lambda}_{(1)}\hat{V}^T V \Lambda_{(1)}^{-1/2}\big) \\ &\geq \frac{\hat{\lambda}_r}{\lambda_1}\Big(r - \frac{1}{2}\|\hat{V}\hat{V}^T - VV^T\|_F^2\Big). \end{aligned} \tag{30}$$

When $\frac{p}{n\lambda_r} \to \infty$, $\hat{\lambda}_r \asymp p/n$.Then (24) holds.

□

Suprisingly, from our proof we can see that the error of PCA can be estimated well!

9. The new theory

Lemma 6. *if*

$$\frac{\lambda_1(\Sigma_i)}{[\mathrm{tr}(\Sigma_i^2)]^{1/2}} \to 0, \tag{31}$$

$i = 1, 2$. *Then*

$$\mathrm{tr}(\Sigma_i\Sigma_j\Sigma_l\Sigma_h) = o[\mathrm{tr}^2\{(\Sigma_1 + \Sigma_2)^2\}] \quad \textit{for } i, j, l, h = 1 \textit{ or } 2. \tag{32}$$

Proof. There are totally 6 possibilities of $\mathrm{tr}(\Sigma_i\Sigma_j\Sigma_l\Sigma_h)$:

$$\begin{array}{ccc} \mathrm{tr}(\Sigma_1\Sigma_1\Sigma_1\Sigma_1) & \mathrm{tr}(\Sigma_1\Sigma_1\Sigma_1\Sigma_2) & \mathrm{tr}(\Sigma_1\Sigma_1\Sigma_2\Sigma_2) \\ \mathrm{tr}(\Sigma_1\Sigma_2\Sigma_1\Sigma_2) & \mathrm{tr}(\Sigma_1\Sigma_2\Sigma_2\Sigma_2) & \mathrm{tr}(\Sigma_2\Sigma_2\Sigma_2\Sigma_2) \end{array}$$

For $\mathrm{tr}(\Sigma_1\Sigma_2\Sigma_1\Sigma_2)$ we have

$$\begin{aligned} \frac{\mathrm{tr}(\Sigma_1\Sigma_2\Sigma_1\Sigma_2)}{\mathrm{tr}^2\{(\Sigma_1 + \Sigma_2)^2\}} &= \frac{\mathrm{tr}(\Sigma_1^{1/2}\Sigma_2\Sigma_1\Sigma_2\Sigma_1^{1/2})}{\mathrm{tr}^2\{(\Sigma_1 + \Sigma_2)^2\}} \\ &\leq \frac{\lambda_1(\Sigma_1)\mathrm{tr}(\Sigma_1^{1/2}\Sigma_2^2\Sigma_1^{1/2})}{\mathrm{tr}^2\{(\Sigma_1 + \Sigma_2)^2\}} \\ &\leq \frac{\lambda_1^2(\Sigma_1)\mathrm{tr}(\Sigma_2^2)}{\mathrm{tr}^2\{(\Sigma_1 + \Sigma_2)^2\}} \\ &\leq \frac{\lambda_1^2(\Sigma_1)\mathrm{tr}(\Sigma_2^2)}{\mathrm{tr}(\Sigma_1^2)\mathrm{tr}(\Sigma_2^2)} \to 0. \end{aligned}$$

The other cases can be proved similarly. $\square$

Define

$$T(A) = \frac{\sum_{i\neq j}^{n_1} X_{1i}^T AA^T X_{1j}}{n_1(n_1-1)} + \frac{\sum_{i\neq j}^{n_2} X_{2i}^T AA^T X_{2j}}{n_2(n_2-1)} - 2\frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} X_{1i}^T AA^T X_{2j}}{n_1 n_2}. \tag{33}$$

Let

$$W(A) = \frac{2}{n_1(n_1-1)}\mathrm{tr}(A^T\Sigma_1 A)^2 + \frac{2}{n_2(n_2-1)}\mathrm{tr}(A^T\Sigma_2 A)^2 + \frac{4}{n_1 n_2}\mathrm{tr}(A^T\Sigma_1 AA^T\Sigma_2 A). \tag{34}$$

Theorem 11. *Suppose $\hat{\hat{V}}$ is a random $(p-\hat{r})\times p$ orthgonal matrix which is independent of the data, $\hat{r}\geq \max(r_1,r_2)$ and is bounded. $n_1/n_2\to\psi$. Then a sufficient condition of*

$$\frac{T(\hat{\hat{V}}) - \|\hat{\hat{V}}^T(\mu_1-\mu_2)\|^2}{\sqrt{W(\hat{\hat{V}})}} \xrightarrow{\mathcal{L}} N(0,1) \tag{35}$$

is that

$$\lambda_1(\Sigma_i)\mathrm{tr}(\hat{\hat{V}}^T V_i V_i^T \hat{\hat{V}}) = o_P(\sqrt{p}), \tag{36}$$

and

$$(\mu_1-\mu_2)^T\hat{\hat{V}}\hat{\hat{V}}^T\Sigma_i\hat{\hat{V}}\hat{\hat{V}}^T(\mu_1-\mu_2) = o_P(\frac{p}{n}) \tag{37}$$

$i=1,2$. If we further assume $\lambda_1(\Sigma)\asymp\lambda_r(\Sigma)$, $\Sigma_1=\Sigma_2$, $\mu_1=\mu_2$ and data are from normal distribution, then (36) is also neccessary.

Proof. We only need to prove that (35) holds conditioning on $\hat{\hat{V}}$. Notice that $T(\hat{\hat{V}})$ can be regard as the T_{CQ} of data $\hat{\hat{V}}^T X_{i,j}$. Since $\hat{\hat{V}}^T X_{i,j} = \hat{\hat{V}}^T U_i\Lambda_i^{1/2} Z_{i,j}$. Denote $\Sigma_i^* = \hat{\hat{V}}^T\Sigma_i\hat{\hat{V}}$, $i=1,2$. By Chen Qin's theorem, we only need to check condition (3.6) (for Σ_i^*) and (3.4) of Chen Qin's paper. Since

$$\begin{aligned}\Sigma_i^* &= \hat{\hat{V}}^T V_i\Lambda_{i,(1)}V_i^T\hat{\hat{V}} + \hat{\hat{V}}^T\tilde{V}_i\Lambda_{i,(2)}\tilde{V}_i^T\hat{\hat{V}}\\ &\stackrel{def}{=} A_1 + A_2\end{aligned} \tag{38}$$

First step: deal with A_2.

For $j = 1, \ldots, p - \hat{r}$, we have

$$\begin{aligned} \lambda_j(A_2) &= \lambda_j(\Lambda_{i,(2)}^{1/2} \tilde{V}_i^T \hat{\tilde{V}} \hat{\tilde{V}}^T \tilde{V}_i \Lambda_{i,(2)}^{1/2}) \\ &= \lambda_j(\Lambda_{i,(2)}^{1/2} \tilde{V}_i^T \tilde{V}_i \Lambda_{i,(2)}^{1/2} - \Lambda_{i,(2)}^{1/2} \tilde{V}_i^T \hat{V} \hat{V}^T \tilde{V}_i \Lambda_{i,(2)}^{1/2}) \end{aligned} \tag{39}$$

Since the rank of $\Lambda_{i,(2)}^{1/2} \tilde{V}_i^T \hat{V} \hat{V}^T \tilde{V}_i \Lambda_{i,(2)}^{1/2}$ is at most $\hat{r}$, we have for $j = 1, \ldots, p-2\hat{r}$ that

$$\lambda_{j+\hat{r}}(\Lambda_{i,(2)}^{1/2} \tilde{V}_i^T \tilde{V}_i \Lambda_{i,(2)}^{1/2}) \leq \lambda_j(A_2) \leq \lambda_j(\Lambda_{i,(2)}^{1/2} \tilde{V}_i^T \tilde{V}_i \Lambda_{i,(2)}^{1/2}). \tag{40}$$

Note that $\lambda_j(\Lambda_{i,(2)}^{1/2} \tilde{V}_i^T \tilde{V}_i \Lambda_{i,(2)}^{1/2}) = \lambda_j(\tilde{V}_i \Lambda_{i,(2)} \tilde{V}_i^T) = \lambda_{j+r_i}(\Sigma_i)$, $j = 1, \ldots, p - r_i$. Hence we have

$$\lambda_{j+\hat{r}+r_i}(\Sigma_i) \leq \lambda_j(A_2) \leq \lambda_{j+r_i}(\Sigma_i), \tag{41}$$

for $j = 1, \ldots, p - 2\hat{r}$. Since $\text{rank}(A_1) \leq r_i$, we have for $j = r_i + 1, \ldots, p - 2\hat{r}$ that

$$\lambda_{j+\hat{r}+r_i}(\Sigma_i) \leq \lambda_j(A_2) \leq \lambda_j(\Sigma_i^*) \leq \lambda_{j-r_i}(A_2) \leq \lambda_j(\Sigma_i). \tag{42}$$

Specially, $\lambda_j(\Sigma_i^*) \asymp 1$, $j = r_i + 1, \ldots, p - 2\hat{r}$.

By (42),

$$\sum_{j=\hat{r}+2r_i+1}^{p-\hat{r}+r_i} \lambda_j^2(\Sigma_i) \leq \sum_{j=r_i+1}^{p-2\hat{r}} \lambda_j^2(\Sigma_i^*) \leq \sum_{j=r_i+1}^{p-2\hat{r}} \lambda_j^2(\Sigma_i). \tag{43}$$

Hence

$$\frac{\sum_{j=r_i+1}^{p-2\hat{r}} \lambda_j^2(\Sigma_i^*)}{\sum_{j=r_i+1}^{p} \lambda_j^2(\Sigma_i)} \xrightarrow{P} 1. \tag{44}$$

Below we will use the assumption of PCA rate.

$$\lambda_1(A_1) \leq \lambda_1(\Sigma_i) \text{tr}(\hat{\tilde{V}}^T V_i V_i^T \hat{\tilde{V}}). \tag{45}$$

And

$$\lambda_1(\Sigma_i^*) \leq \lambda_1(A_1) + \lambda_1(A_2) \leq o_P(\sqrt{p}) + \lambda_{r_i+1}(\Sigma_i) = o_P(\sqrt{p}). \tag{46}$$

Since

$$\sum_{j=r_i+1}^{p-2\hat{r}} \lambda_j^2(\Sigma_i^*) \leq \text{tr}(\Sigma_i^{*2}) \leq r_i \lambda_1^2(\Sigma_i^*) + \sum_{j=r_i+1}^{p-2\hat{r}} \lambda_j^2(\Sigma_i^*) + 2\hat{r} \lambda_{p-2\hat{r}+1}^2(\Sigma_i^*).$$

But $\sum_{j=r_i+1}^{p-2\hat{r}} \lambda_j^2(\Sigma_i^*) \asymp p$, $\lambda_1^2(\Sigma_i^*) = o_P(p)$, and $2\hat{r}\lambda_{p-2\hat{r}+1}^2(\Sigma_i^*) = O_P(1)$ since $\hat{r}$ is bounded. It follows that

$$\frac{\operatorname{tr}(\Sigma_i^{*2})}{\sum_{j=r_i+1}^{p-2\hat{r}} \lambda_j^2(\Sigma_i^*)} \xrightarrow{P} 1. \tag{47}$$

It follows that

$$\frac{\operatorname{tr}(\Sigma_i^{*2})}{\sum_{j=r_i+1}^{p} \lambda_j^2(\Sigma_i)} \xrightarrow{P} 1. \tag{48}$$

Notice that

$$\frac{\lambda_1(\Sigma_i^*)}{[\operatorname{tr}(\Sigma_i^{*2})]^{1/2}} \leq \frac{o_P(\sqrt{p})}{\left[\sum_{j=r_i+1}^{p-2\hat{r}} \lambda_j^2(\Sigma_i^*)\right]^{1/2}} \asymp \frac{o_P(\sqrt{p})}{\sqrt{p}} \to 0. \tag{49}$$

By lemma, the first condition of Chen and Qin is satisfied.

$$\begin{aligned}\frac{(\mu_1-\mu_2)^T\hat{\hat{V}}\Sigma_i^*\hat{\hat{V}}^T(\mu_1-\mu_2)}{n^{-1}\operatorname{tr}\{(\Sigma_1^*+\Sigma_2^*)^2\}} &\leq \frac{(\mu_1-\mu_2)^T\hat{\hat{V}}\Sigma_i^*\hat{\hat{V}}^T(\mu_1-\mu_2)}{n^{-1}\operatorname{tr}(\Sigma_1^{*2})} \\ &\asymp \frac{(\mu_1-\mu_2)^T\hat{\hat{V}}\Sigma_i^*\hat{\hat{V}}^T(\mu_1-\mu_2)}{n^{-1}p}\end{aligned}$$

Neccessaity is in the conditional sense. Suppose

$$\frac{T(\hat{\hat{V}}) - \|\hat{\hat{V}}^T(\mu_1-\mu_2)\|^2}{\sqrt{W(\hat{\hat{V}})}} \xrightarrow{\mathcal{L}} N(0,1) \tag{50}$$

Write $T(\hat{\hat{V}}) = T^{(1)} + T^{(2)}$, where

$$\begin{aligned}T^{(1)} =&\frac{\sum_{i\neq j}^{n_1}(X_{1i}-\mu_1)^T\hat{\hat{V}}\hat{\hat{V}}^T(X_{1j}-\mu_1)}{n_1(n_1-1)} + \frac{\sum_{i\neq j}^{n_2}(X_{2i}-\mu_2)^T\hat{\hat{V}}\hat{\hat{V}}^T(X_{2j}-\mu_2)}{n_2(n_2-1)} \\ &- 2\frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(X_{1i}-\mu_1)^T\hat{\hat{V}}\hat{\hat{V}}^T(X_{2j}-\mu_2)}{n_1n_2}\end{aligned}$$

and

$$\begin{aligned}T^{(2)} =&2(\bar{X}_1-\mu_1)^T\hat{\hat{V}}\hat{\hat{V}}^T(\mu_1-\mu_2) + 2(\bar{X}_2-\mu_2)^T\hat{\hat{V}}\hat{\hat{V}}^T(\mu_2-\mu_1) \\ &+ \|\hat{\hat{V}}(\mu_1-\mu_2)\|^2.\end{aligned}$$

Since $W(\hat{\tilde{V}}) \geq \frac{2}{n_1(n_1-1)}\mathrm{tr}(\Sigma_1^{*2}) \asymp \frac{p}{n^2}$. We have

$$\begin{aligned}\mathrm{Var}\Big(\frac{T^{(2)} - \|\hat{\tilde{V}}(\mu_1-\mu_2)\|^2}{\sqrt{W(\hat{\tilde{V}})}}\Big) =& \frac{4}{W(\hat{\tilde{V}})}\Big((\mu_1-\mu_2)^T\hat{\tilde{V}}\hat{\tilde{V}}^T\frac{1}{n_1}\Sigma_1\hat{\tilde{V}}\hat{\tilde{V}}^T(\mu_1-\mu_2) \\ &+ (\mu_1-\mu_2)^T\hat{\tilde{V}}\hat{\tilde{V}}^T\frac{1}{n_2}\Sigma_2\hat{\tilde{V}}\hat{\tilde{V}}^T(\mu_1-\mu_2)\Big) \\ =& o_P(1)\frac{n^2}{p}\frac{1}{n}\frac{p}{n} \xrightarrow{P} 0.\end{aligned}$$

And $\frac{T^{(2)} - \|\hat{\tilde{V}}(\mu_1-\mu_2)\|^2}{\sqrt{W(\hat{\tilde{V}})}}$ has mean zero, hence converges to 0 in probability. By Slutsky's theorem,

$$\frac{T^{(1)}}{\sqrt{W(\hat{\tilde{V}})}} \xrightarrow{\mathcal{L}} N(0,1). \tag{51}$$

It follows by previous theorem that $\frac{\lambda_1(\Sigma^*)}{[\mathrm{tr}\Sigma^{*2}]^{1/2}} \to 0$.Then

$$\frac{\sum_{i=1}^r \lambda_i^2(\Sigma^*)}{\mathrm{tr}\Sigma^{*2}} \leq \frac{r\lambda_1^2(\Sigma^*)}{\mathrm{tr}\Sigma^{*2}} \to 0. \tag{52}$$

Then

$$\frac{\sum_{i=1}^r \lambda_i^2(\Sigma^*)}{\sum_{i=r+1}^p \lambda_i^2(\Sigma^*)} \to 0, \tag{53}$$

which is equivalent to λ. By (44), $\sum_{i=1}^r \lambda_i^2(\Sigma^*) = o(p)$. Then

$$\lambda_1(A_1) \leq \lambda_1^2(\Sigma^*) \leq \sum_{i=1}^r \lambda_i^2(\Sigma^*) = o(p) \tag{54}$$

But

$$\lambda_1(A_1) \asymp \lambda_1(\Sigma_i)\lambda_1(\hat{\tilde{V}}^T V V^T \hat{\tilde{V}}) \asymp \lambda_1(\Sigma_i)\lambda_1(\hat{\tilde{V}}^T V V^T \hat{\tilde{V}}) \tag{55}$$

The first equivalence of above holds by assumption. The second equivalence holds because $\mathrm{rank}(\hat{\tilde{V}}^T V V^T \hat{\tilde{V}}) \leq r$. The conclusion follows. □

The rest of the Appendix is devoted to the proof of propositions and theorems in the paper.

Proof Of Proposition 1. Since V and $\tilde{V}$ are orthogonal, we have

$$\tilde{V}^T X_{ki} \sim N(\tilde{V}\mu_k, \sigma^2 I_{p-r}).$$

By Chen and Qin (2010)'s section 6.1, we have

$$Var(T_1) = 2\tau^2 p\sigma^4(1+o(1)).$$

Random sequences $\tilde{V}^T X_{ki}$ fulfill the condition of Chen and Qin (2010)'s theorem 1, hence the conclusion follows. $\square$

Proof Of Proposition 2. By a standard orthogonal transformation, $\hat{\sigma}^2_{(1)}$ has the same law with respect to $\frac{1}{(n_1+n_2-2)(p-r)}\mathrm{tr}\sum_{k=1}^{2}\sum_{i=1}^{n_i-1}\hat{\tilde{V}}^T Y_{ki}Y_{ki}^T\hat{\tilde{V}}$, where $Y_{ki} = VDU^*_{ki} + Z^*_{ki}$. Here U^*_{ki}'s are i.i.d. random vectors with r dimensional standard normal distribution and Z^*_{ki}'s are i.i.d. random vectors distributed as $N_p(0,\sigma^2 I_p)$ which are independent of U^*_{ki}'s. Denote $U^* = (U^*_{11},\ldots,U^*_{1(n_1-1)},U^*_{21},\ldots,U^*_{2(n_2-1)})^T$ and $Z^* = (Z^*_{11},\ldots,Z^*_{1(n_1-1)},Z^*_{21},\ldots,Z^*_{2(n_2-1)})^T$. Hence we have

$$\begin{aligned}\hat{\sigma}^2_{(1)} =&\frac{1}{(n_1+n_2-2)(p-r)}\|U^*DV^T\hat{\tilde{V}}\|_F^2 + \frac{1}{(n_1+n_2-2)(p-r)}\|Z^*\hat{\tilde{V}}\|_F^2\\ &+\frac{2}{(n_1+n_2-2)(p-r)}\mathrm{tr}\hat{\tilde{V}}^T Z^{*T}U^*DV^T\hat{\tilde{V}}\\ \overset{\text{def}}{=}&R_1+R_2+R_3.\end{aligned}$$

By Lemma 4, we have $\lambda_1(U^{*T}U^*) = O_P(n_1+n_2-2)$. Therefore,

$$\begin{aligned}R_1 &= \frac{1}{(n_1+n_2-2)(p-r)}\mathrm{tr}\hat{\tilde{V}}^T VDU^{*T}U^*DV^T\hat{\tilde{V}}\\ &= O_P(1)\frac{1}{p-r}\mathrm{tr}\hat{\tilde{V}}^T VD^2V^T\hat{\tilde{V}}\\ &= O_P(1)\frac{1}{p-r}\mathrm{tr}D^2V^T(I-\hat{V}\hat{V}^T)V\\ &\le O_P(1)\frac{1}{p}\sqrt{\mathrm{tr}D^4}\sqrt{\mathrm{tr}(V^T(I-\hat{V}\hat{V}^T)V)^2}.\end{aligned}\tag{56}$$

And

$$\mathrm{tr}(V^T(I-\hat{V}\hat{V}^T)V)^2 \le (\mathrm{tr}V^T(I-\hat{V}\hat{V}^T)V)^2 \tag{57}$$

$$= \frac{1}{4}\|VV^T-\hat{V}\hat{V}^T\|_F^4 \tag{58}$$

$$= O(\frac{p^2}{p^{2\beta}(n_1+n_2)^2}).$$

The inequality (57) holds because $V^T(I-\hat{V}\hat{V}^T)V$ is positive semi-definite and equality (58) is by the fact that $\mathrm{tr}V^T(I-\hat{V}\hat{V}^T)V = \frac{1}{2}\|VV^T-\hat{V}\hat{V}^T\|_F^2$. Since $\mathrm{tr}D^4 = O_P(p^{2\beta})$, it follows that $R_1 = O_P(\frac{1}{n_1+n_2}) \xrightarrow{P} 0$.

We note that

$$
\begin{aligned}
&|R_2 - \frac{1}{(n_1+n_2-2)(p-r)}\|Z^*\tilde{V}\|_F^2| \\
&= \frac{1}{(n_1+n_2-2)(p-r)}|\mathrm{tr}Z^{*T}Z^*(\hat{\tilde{V}}\hat{\tilde{V}}^T - \tilde{V}\tilde{V}^T)| \\
&= \frac{1}{(n_1+n_2-2)(p-r)}|\mathrm{tr}Z^{*T}Z^*(\hat{V}\hat{V}^T - VV^T)| \\
&\le \frac{1}{(n_1+n_2-2)(p-r)}\|Z^{*T}Z^*\|_F\|\hat{V}\hat{V}^T - VV^T\|_F.
\end{aligned}
\tag{59}
$$

By directly calculating expectation, it's easy to check that $\|Z^{*T}Z^*\|_F = O_P((n_1+n_2-2)p)$. Since $\beta \ge 1/2$ and $\frac{\sqrt{p}}{n} \to 0$, we have

$$
\|VV^T - \hat{V}\hat{V}^T\|_F^2 = O_P(\frac{p}{p^\beta(n_1+n_2)}) = o(1). \tag{60}
$$

Therefore, $|R_2 - \frac{1}{(n_1+n_2-2)(p-r)}\|Z^*\tilde{V}\|_F^2| \xrightarrow{P} 0$. But $\frac{1}{(n_1+n_2-2)(p-r)}\|Z^*\hat{V}\|_F^2 \xrightarrow{P} \sigma^2$ by law of large numbers, which yields $R_2 \xrightarrow{P} \sigma^2$.

Finally, as $R_3 \le 2\sqrt{R_1}\sqrt{R_2}$ we have $R_3 \xrightarrow{P} 0$. It follows that $\hat{\sigma}^2_{(1)}$ is consistent.

Next we proof the consistency of $\hat{\sigma}^2_{(2)}$.

We denote

$$
\begin{aligned}
S_{UU} &= \frac{1}{n_1+n_2-2}\sum_{k=1}^{2}\sum_{i=1}^{n_k}(U_{ki}-\bar{U}_k)(U_{ki}-\bar{U}_k)^T, \\
S_{ZZ} &= \frac{1}{n_1+n_2-2}\sum_{k=1}^{2}\sum_{i=1}^{n_k}(Z_{ki}-\bar{Z}_k)(Z_{ki}-\bar{Z}_k)^T, \\
S_{UZ} &= \frac{1}{n_1+n_2-2}\sum_{k=1}^{2}\sum_{i=1}^{n_k}(U_{ki}-\bar{U}_k)(Z_{ki}-\bar{Z}_k)^T, \\
S_{ZU} &= \frac{1}{n_1+n_2-2}\sum_{k=1}^{2}\sum_{i=1}^{n_k}(Z_{ki}-\bar{Z}_k)(U_{ki}-\bar{U}_k)^T.
\end{aligned}
\tag{61}
$$

Then

$$
\begin{aligned}
S &= \frac{1}{n_1+n_2-2}\sum_{k=1}^{2}\sum_{i=1}^{n_k}(X_{ki}-\bar{X}_k)(X_{ki}-\bar{X}_k)^T \\
&= \frac{1}{n_1+n_2-2}\sum_{k=1}^{2}\sum_{i=1}^{n_k}(VD(U_{ki}-\bar{U}_k)+Z_{ki}-\bar{Z}_k)(VD(U_{ki}-\bar{U}_k)+Z_{ki}-\bar{Z}_k)^T \\
&= VDS_{UU}DV^T + VDS_{UZ} + S_{ZU}DV^T + S_{ZZ} \\
&= VD(S_{UU}DV^T + S_{UZ}) + S_{ZU}DV^T + S_{ZZ} \\
&\stackrel{\text{def}}{=} F_1 + F_2 + F_3.
\end{aligned} \tag{62}
$$

By Weyl's inequality, we have

$$\lambda_{p-2r}(F_1+F_2) + \lambda_{2r+k}(S_{ZZ}) \le \lambda_k(S) \le \lambda_{2r+1}(F_1+F_2) + \lambda_{k-2r}(S_{ZZ}) \tag{63}$$

for $2r+1 \le k \le p-2r$. But $rank(F_1+F_2) \le 2r$, since $rank(F_k) \le rank(D) = r$ for $k = 1, 2$. Thus the kth eigenvalue of $F_1 + F_2$ equals to zero, where $2r+1 \le k \le p-2r$. In this way,

$$\lambda_{2r+k}(S_{ZZ}) \le \lambda_k(S) \le \lambda_{k-2r}(S_{ZZ}). \tag{64}$$

By above argument, we have upper bound

$$\frac{1}{p-4r}\sum_{k=2r+1}^{p-2r}\lambda_k(S) \le \frac{1}{p-4r}\sum_{k=1}^{p-4r}\lambda_k(S_{ZZ}) \le \frac{1}{p-4r}\mathrm{tr}S_{ZZ} \tag{65}$$

and lower bound

$$\frac{1}{p-4r}\sum_{k=2r+1}^{p-2r}\lambda_k(S) \ge \frac{1}{p-4r}\sum_{k=4r+1}^{p}\lambda_k(S_{ZZ}) \ge \frac{1}{p-4r}\mathrm{tr}S_{ZZ} - \frac{4r}{p-4r}\lambda_1(S_{ZZ}). \tag{66}$$

Thus it suffices to prove that $\frac{1}{p}\mathrm{tr}S_{ZZ} \xrightarrow{P} \sigma^2$ and $\frac{1}{p}\lambda_1(S_{ZZ}) \xrightarrow{P} 0$. We note that $(n_1+n_2-2)S_{ZZ}$ is distributed as $\text{Wishart}_p(n_1+n_2-2, \sigma^2 I_p)$. Hence $\frac{1}{p}\mathrm{tr}(S_{ZZ}) \xrightarrow{P} \sigma^2$ by law of large numbers. And $\frac{1}{p}\lambda_1(S_{ZZ}) = o_P(1)$ by Lemma 4. Therefore the consistency of $\hat{\sigma}^2_{(2)}$ is proved. □

Proof Of Theorem 1. Our proof starts with the observation that the elements of $\mu_1 - \mu_2$ is distributed as $N(0, \tau p^{-\frac{1}{2}}\psi)$. Hence

$$(\mu_1-\mu_2)^T\Sigma(\mu_1-\mu_2) = O(\|\mu_1-\mu_2\|^2) = O_P(\tau p^{\frac{1}{2}}) = o_P(\tau\,\mathrm{tr}\Sigma^2).$$

Here the second equality holds by law of large number, the first and third equalities are due to boundedness of the eigenvalues of Σ. It follows that every subsequence has a further subsequence along which we have

$$(\mu_1-\mu_2)^T\Sigma(\mu_1-\mu_2)=o(\tau\,\mathrm{tr}\Sigma^2)$$

almost surely (a.s.). Let

$$\eta_n=\frac{T_{CQ}-\|\mu_1-\mu_2\|^2}{\sqrt{2\tau^2\mathrm{tr}\Sigma^2}},$$

then by Theorem 1 in Chen and Qin (2010),

$$P(\eta_n\le x|\mu_1,\mu_2)\to\Phi(x)\quad\text{a.s.}$$

along the further subsequence. Therefore,

$$P(\eta_n\le x|\mu_1,\mu_2)\xrightarrow{P}\Phi(x).$$

We conclude from dominated convergence theorem that $\eta_n\xrightarrow{\mathcal{L}}N(0,1)$. What is left is to show that

$$\frac{T_{CQ}-T_2}{\sqrt{2\tau^2\mathrm{tr}\Sigma^2}}\xrightarrow{P}0. \tag{67}$$

We note that

$$\begin{aligned}T_{CQ}-T_2&=\frac{\sum_{i\neq j}^{n_1}X_{1i}^T\hat V\hat V^TX_{1j}}{n_1(n_1-1)}+\frac{\sum_{i\neq j}^{n_2}X_{2i}^T\hat V\hat V^TX_{2j}}{n_2(n_2-1)}-2\frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}X_{1i}^T\hat V\hat V^TX_{2j}}{n_1n_2}\\&\stackrel{\text{def}}{=}P_1+P_2-2P_3.\end{aligned}$$

And

$$\frac{P_1}{\sqrt{2\tau^2\mathrm{tr}\Sigma^2}}=O(1)\frac{\sum_{i\neq j}^{n_1}X_{1i}^T\hat V\hat V^TX_{1j}}{n_1\sqrt{p}},$$

which can be further written by

$$\begin{aligned}\frac{\sum_{i\neq j}^{n_1}X_{1i}^T\hat V\hat V^TX_{1j}}{n_1\sqrt{p}}&=\frac{n_1(n_1-1)\bar X_1^T\hat V\hat V^T\bar X_1}{n_1\sqrt{p}}-\frac{\sum_{i=1}^{n_1}(X_{1i}-\bar X_1)^T\hat V\hat V^T(X_{1i}-\bar X_1)}{n_1\sqrt{p}}\\&\stackrel{\text{def}}{=}R_1-R_2.\end{aligned}$$

Now we deal with R_1. Since $\bar X_1|\mu_1\sim N(\mu_1,\frac{1}{n}\Sigma)$ and $\mu_1\sim N(0,\frac{\psi}{n_1\sqrt{p}}I_p)$, we have $\bar X_1\sim N(0,\frac{1}{n_1}(\Sigma+\frac{1}{\sqrt{p}}\psi I_p))$. Hence we have $\hat V^T\bar X_1|S\sim N(0,\frac{1}{n}\hat V^T(\Sigma+$

$\frac{1}{\sqrt{p}}\psi I_p)\hat{V})$ by the independence of S and $(\mu_1, \bar{X}_1)$. Therefore,

$$\begin{aligned} E[\bar{X}_1^T \hat{V}\hat{V}^T \bar{X}_1] &= EE[\bar{X}_1^T \hat{V}\hat{V}^T \bar{X}_1 | S] \\ &= E[\frac{1}{n_1}\mathrm{tr}\hat{V}^T(\Sigma + \frac{1}{\sqrt{p}}\psi I_p)\hat{V}] \\ &= O(\frac{1}{n_1}). \end{aligned}$$

The last equality holds because the rank of $\hat{V}$ is at most R which is fixed. It follows that $R_1 \xrightarrow{P} 0$.

$$\begin{aligned} R_2 &= \frac{\mathrm{tr}[\hat{V}^T \sum_{i=1}^{n_1}(X_{1i} - \bar{X}_1)(X_{1i} - \bar{X}_1)^T \hat{V}]}{n_1\sqrt{p}} \\ &\le R\frac{\lambda_1(\sum_{i=1}^{n_1}(X_{1i} - \bar{X}_1)(X_{1i} - \bar{X}_1)^T)}{n_1\sqrt{p}}. \end{aligned}$$

Lemma 4 implies that $\lambda_1(\sum_{i=1}^{n_1}(X_{1i} - \bar{X}_1)(X_{1i} - \bar{X}_1)^T) = O_P(\max(n_1, p))$. Therefore, by noting $p = o(n_1^2)$ we have $R_2 \xrightarrow{P} 0$. It follows that $\frac{P_1}{\sqrt{2\tau^2 \mathrm{tr}\Sigma^2}} \xrightarrow{P} 0$. Similar arguments lead to $\frac{P_2}{\sqrt{2\tau^2 \mathrm{tr}\Sigma^2}} \xrightarrow{P} 0$.

$$\begin{aligned} \frac{P_3}{\sqrt{2\tau^2 \mathrm{tr}\Sigma^2}} &= O(1)\frac{\sqrt{n_1 n_2}\bar{X}_1^T \hat{V}\hat{V}^T \bar{X}_2}{\sqrt{p}} \\ &\le O(1)\sqrt{\frac{n_1\|\hat{V}^T\bar{X}_1\|^2}{\sqrt{p}}}\sqrt{\frac{n_2\|\hat{V}^T\bar{X}_2\|^2}{\sqrt{p}}}, \end{aligned}$$

where the inequality is due to Cauchy inequality. By noting the relationship with R_1, the right hand side converges to 0 in probability. And the proof is completed.

□

Proof Of Theorem 2. By Chen and Qin (2010)'s Theorem 1, we have

$$\frac{n_1 n_2 T_1}{\sqrt{2p}(n_1+n_2)\sigma^2} \xrightarrow{\mathcal{L}} N(0,1).$$

It suffices to prove

$$\frac{n_1 n_2 (T_1 - T_2)}{\sqrt{2p}(n_1+n_2)\sigma^2} \xrightarrow{P} 0.$$

Since the test statistic is invariant under transformation $X_{1i} \mapsto X_{1i} + \mu$, $X_{2j} \mapsto X_{2j} + \mu$, without loss of generality, we assume $\mu_1 = \mu_2 = 0$. By noting that $\hat{\tilde{V}}\hat{\tilde{V}}^T - \tilde{V}\tilde{V}^T = VV^T - \hat{V}\hat{V}^T$ we have $\frac{n_1 n_2 (T_1 - T_2)}{\sqrt{2p}(n_1+n_2)\sigma^2} = P_1 + P_2 - 2P_3$, where

$$P_1 = \frac{n_2 \sum_{i \neq j} X_{1i}^T (VV^T - \hat{V}\hat{V}^T) X_{1j}}{\sqrt{2p}(n_1 + n_2)(n_1 - 1)\sigma^2},$$

$$P_2 = \frac{n_1 \sum_{i \neq j} X_{2i}^T (VV^T - \hat{V}\hat{V}^T) X_{2j}}{\sqrt{2p}(n_1 + n_2)(n_2 - 1)\sigma^2},$$

$$P_3 = \frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} X_{1i}^T (VV^T - \hat{V}\hat{V}^T) X_{2j}}{\sqrt{2p}(n_1 + n_2)\sigma^2}.$$

Write

$$\begin{aligned} P_1 &= O(1) \frac{\sum_{i \neq j} X_{1i}^T (VV^T - \hat{V}\hat{V}^T) X_{1j}}{n_1 \sqrt{p}} \\ &= O(1)(R_1 - R_2), \end{aligned}$$

where

$$R_1 = \frac{n_1}{\sqrt{p}} (\bar{X}_1^T VV^T \bar{X}_1 - \bar{X}_1^T \hat{V}\hat{V}^T \bar{X}_1)$$

and

$$R_2 = \frac{1}{n_1 \sqrt{p}} \sum_{i=1}^{n_1} (X_{1i}^T VV^T X_{1i} - X_{1i}^T \hat{V}\hat{V}^T X_{1i}).$$

To deal with R_1, we further decompose $\bar{X}_1^T \hat{V}\hat{V}^T \bar{X}_1$ into 3 parts

$$\begin{aligned} \bar{X}_1^T \hat{V}\hat{V}^T \bar{X}_1 =& \bar{X}_1^T (VV^T + \tilde{V}\tilde{V}^T) \hat{V}\hat{V}^T (VV^T + \tilde{V}\tilde{V}^T) \bar{X}_1 \\ =& \bar{X}_1^T VV^T \hat{V}\hat{V}^T VV^T \bar{X}_1 + 2\bar{X}_1^T \tilde{V}\tilde{V}^T \hat{V}\hat{V}^T VV^T \bar{X}_1 \\ &+ \bar{X}_1^T VV^T \hat{V}\hat{V}^T VV^T \bar{X}_1. \end{aligned}$$

The above decomposition is a technique we will use many times. R_1 can thus be written by $R_1 = Q_1 - 2Q_2 - Q_3$, where

$$Q_1 = \frac{n_1}{\sqrt{p}} \bar{X}_1^T V (I_r - V^T \hat{V}\hat{V}^T V) V^T \bar{X}_1,$$

$$Q_2 = \frac{n_1}{\sqrt{p}} \bar{X}_1^T \tilde{V}\tilde{V}^T \hat{V}\hat{V}^T VV^T \bar{X}_1,$$

$$Q_3 = \frac{n_1}{\sqrt{p}} \bar{X}_1^T \tilde{V}\tilde{V}^T \hat{V}\hat{V}^T \tilde{V}\tilde{V}^T \bar{X}_1.$$

It's clear that $V^T\bar{X}_1 \sim N_r(0, \frac{1}{n}(\Lambda + \sigma^2 I_p))$, $\tilde{V}^T\bar{X}_1 \sim N_{p-r}(0, \frac{\sigma^2}{n} I_p)$ and S are mutually independent. $\|V^T\bar{X}_1\|^2 = O_P(\frac{p^\beta}{n_1})$. $I_r - V^T\hat{V}\hat{V}^TV$ is a positive semi-difinite matrix which only relies on S. Combining these observations, we have

$$
\begin{aligned}
|Q_1| &\leq \frac{n_1}{\sqrt{p}}\|I_r - V^T\hat{V}\hat{V}^TV\|\|V^T\bar{X}_1\|^2 \\
&\leq \frac{n_1}{\sqrt{p}}\operatorname{tr}(I_r - V^T\hat{V}\hat{V}^TV)\|V^T\bar{X}_1\|^2 \\
&= \frac{n_1}{2\sqrt{p}}\|VV^T - \hat{V}\hat{V}^T\|_F^2\|V^T\bar{X}_1\|^2 \\
&= O_P(1)\frac{n_1}{\sqrt{p}}\frac{p}{p^\beta n_1}\frac{p^\beta}{n_1} \xrightarrow{P} 0.
\end{aligned}
\tag{68}
$$

We next prove $Q_2 \xrightarrow{L^2} 0$.

$$
\begin{aligned}
E(Q_2^2) &= \frac{n_1^2}{p}E(\bar{X}_1^T\tilde{V}\tilde{V}^T\hat{V}\hat{V}^TVV^T\bar{X}_1\bar{X}_1^TVV^T\hat{V}\hat{V}^T\tilde{V}\tilde{V}^T\bar{X}_1) \qquad (69) \\
&= \frac{n_1^2}{p}E(\bar{X}_1^T\tilde{V}\tilde{V}^T\hat{V}\hat{V}^TV\frac{1}{n_1}(\Lambda + \sigma^2 I_r)V^T\hat{V}\hat{V}^T\tilde{V}\tilde{V}^T\bar{X}_1) \\
&= O(1)\frac{n_1p^\beta}{p}\mathrm{E}(\bar{X}_1^T\tilde{V}\tilde{V}^T\hat{V}\hat{V}^TVV^T\hat{V}\hat{V}^T\tilde{V}\tilde{V}^T\bar{X}_1) \\
&= O(1)\frac{n_1p^\beta}{p}\mathrm{E}\operatorname{tr}(\tilde{V}^T\hat{V}\hat{V}^TVV^T\hat{V}\hat{V}^T\tilde{V}\tilde{V}^T\bar{X}_1\bar{X}_1^T\tilde{V}) \\
&= O(1)\frac{n_1p^\beta}{p}\frac{\sigma^2}{n_1}\mathrm{E}\operatorname{tr}(\tilde{V}^T\hat{V}\hat{V}^TVV^T\hat{V}\hat{V}^T\tilde{V}) \\
&\leq O(1)\frac{p^\beta}{p}\mathrm{E}\operatorname{tr}(\tilde{V}^T\hat{V}\hat{V}^TI_p\hat{V}\hat{V}^T\tilde{V}) \\
&= O(1)\frac{p^\beta}{p}\mathrm{E}\operatorname{tr}(\tilde{V}^T\hat{V}\hat{V}^T\tilde{V}) \\
&= O(1)\frac{p^\beta}{p}\mathrm{E}\operatorname{tr}(\hat{V}\hat{V}^T(I_p - VV^T)) \\
&= O(1)\frac{p^\beta}{p}\frac{1}{2}\mathrm{E}\|VV^T - \hat{V}\hat{V}^T\|_F^2 \\
&= O(\frac{1}{n_1}) \to 0.
\end{aligned}
$$

Similarly,

$$E(Q_3) = \frac{n_1}{\sqrt{p}} \frac{\sigma^2}{n_1} E\,\mathrm{tr}(\tilde{V}^T \hat{V} \hat{V}^T \tilde{V}) \le \frac{r\sigma^2}{\sqrt{p}} \to 0. \tag{70}$$

By (68), (69) and (70), $R_1 \xrightarrow{P} 0$. Next we deal with R_2. Write $R_2 = W_1 - 2W_2 - W_3$, where

$$W_1 = \frac{1}{n_1\sqrt{p}} \sum_{i=1}^{n_1} X_{1i}^T V (I_r - V^T \hat{V} \hat{V}^T V) V^T X_{1i},$$

$$W_2 = \frac{1}{n_1\sqrt{p}} \sum_{i=1}^{n_1} X_{1i}^T \tilde{V} \tilde{V}^T \hat{V} \hat{V}^T V V^T X_{1i},$$

$$W_3 = \frac{1}{n_1\sqrt{p}} \sum_{i=1}^{n_1} X_{1i}^T \tilde{V} \tilde{V}^T \hat{V} \hat{V}^T \tilde{V} \tilde{V}^T X_{1i}.$$

It is seen that independence property does not holds anymore compared with the case we deal with R_1. Nevertheless, the form is of sum, which makes it possible to apply law of large numbers.

We note that

$$W_1 = \frac{1}{n_1\sqrt{p}} \mathrm{tr}((I_r - V^T \hat{V} \hat{V}^T V) \sum_{i=1}^{n_1} V^T X_{1i} X_{1i}^T V). \tag{71}$$

By law of large numbers,

$$\frac{1}{n_1} (\Lambda + \sigma^2 I_r)^{-\frac{1}{2}} \sum_{i=1}^{n_1} V^T X_{1i} X_{1i}^T V (\Lambda + \sigma^2 I_r)^{-\frac{1}{2}} \xrightarrow{P} I_r.$$

Hence $\lambda_1(\sum_{i=1}^{n_1} V^T X_{1i} X_{1i}^T V) = O_P(n_1 p^\beta)$. Substituting it into (71) yields

$$\begin{aligned} W_1 &\le O_P(1) \frac{p^\beta}{\sqrt{p}} \mathrm{tr}(I_r - V^T \hat{V} \hat{V}^T V) \\ &= O_P(1) \frac{p^\beta}{\sqrt{p}} \|V V^T - \hat{V} \hat{V}^T\|_F^2 \\ &= O_P(1) \frac{p^\beta}{\sqrt{p}} \frac{p}{p^\beta n_1} \\ &= O_P(\frac{\sqrt{p}}{n_1}) \xrightarrow{P} 0. \end{aligned} \tag{72}$$

Next we deal with W_2. By cauchy inequality, we have

$$
\begin{aligned}
W_2 &= \frac{1}{n_2\sqrt{p}} \mathrm{tr} \tilde{V}^T \hat{V} \hat{V}^T V (\sum_{i=1}^{n_1} V^T X_{1i} X_{1i}^T \tilde{V}) \\
&\leq \frac{1}{n_2\sqrt{p}} \sqrt{\mathrm{tr}(\tilde{V}^T \hat{V} \hat{V}^T V V^T \hat{V} \hat{V}^T \tilde{V})} \sqrt{\mathrm{tr}(V Z Z^T \tilde{V} \tilde{V}^T Z Z^T V^T)},
\end{aligned}
$$

where $Z = (X_{11}, \ldots, X_{1n_1})$. Note that

$$
\mathrm{tr}(\tilde{V}^T \hat{V} \hat{V}^T V V^T \hat{V} \hat{V}^T \tilde{V}) \leq \mathrm{tr}(\tilde{V}^T \hat{V} \hat{V}^T \tilde{V}) = \frac{1}{2} \| V V^T - \hat{V} \hat{V}^T \|_F^2
$$

and

$$
\mathrm{tr}(V Z Z^T \tilde{V} \tilde{V}^T Z Z^T V^T) \leq \lambda_1 (Z^T \tilde{V} \tilde{V}^T Z) \mathrm{tr}(V Z Z^T V^T).
$$

$Z^T \tilde{V} \tilde{V}^T Z$ is distributed as $\mathrm{Wishart}_n(p - r, \sigma^2 I_{n_1})$, hence $\lambda_1(Z^T \tilde{V} \tilde{V}^T Z) = O_P(\max(n_1, p))$ by Lemma 4. Again by law of large numbers $\mathrm{tr}(V Z Z^T V^T) = O(p^\beta n_1)$. Combining the argument above yields

$$
\begin{aligned}
W_2 &= O_P(1) \frac{1}{n_1 \sqrt{p}} \sqrt{\frac{p}{p^\beta n_1}} \sqrt{\max(n_1, p) p^\beta n_1} \\
&= O_P(\frac{\sqrt{\max(n_1, p)}}{n_1}) \xrightarrow{P} 0.
\end{aligned} \tag{73}
$$

To deal with W_3, note that $\tilde{V}^T \sum_{i=1}^{n_1} X_{1i} X_{1i}^T \tilde{V}$ is of distribution $\mathrm{Wishart}_{p-r}(n, \sigma^2 I_{p-r})$. By Lemma 4, $\lambda_1(\tilde{V}^T \sum_{i=1}^{n_1} X_{1i} X_{1i}^T \tilde{V}) = O_P(\max(p, n_1))$. Hence

$$
\begin{aligned}
W_3 &= \frac{1}{n_1 \sqrt{p}} \mathrm{tr}(\hat{V}^T \tilde{V} \tilde{V}^T \sum_{i=1}^{n_1} X_{1i} X_{1i}^T \tilde{V} \tilde{V}^T \hat{V}) \\
&\leq \frac{O_P(\max(n_1, p))}{n_1 \sqrt{p}} \mathrm{tr} \hat{V}^T \tilde{V} \tilde{V}^T \hat{V} \xrightarrow{P} 0,
\end{aligned} \tag{74}
$$

due to $\mathrm{tr} \hat{V}^T \tilde{V} \tilde{V}^T \hat{V} = O_P(1)$. Combining (72), (73) and (74) gives $R_2 \xrightarrow{P} 0$. So far, it has been proved that $P_1 \xrightarrow{P} 0$. By symmetry, $P_2 \xrightarrow{P} 0$.

We proceed to deal with P_3. Note that $P_3 = O(1)(M_1 - M_2 - M_3 - M_4)$, where

$$
M_1 = \frac{n_1}{\sqrt{p}} \bar{X}_1^T V (I_r - V^T \hat{V} \hat{V}^T V) V^T \bar{X}_2,
$$

$$
M_2 = \frac{n_1}{\sqrt{p}} \bar{X}_1^T \tilde{V} \tilde{V}^T \hat{V} \hat{V}^T V V^T \bar{X}_2,
$$

$$M_3 = \frac{n_1}{\sqrt{p}} \bar{X}_1^T V V^T \hat{V} \hat{V}^T \tilde{V} \tilde{V}^T \bar{X}_2,$$
$$M_4 = \frac{n_1}{\sqrt{p}} \bar{X}_1^T \tilde{V} \tilde{V}^T \hat{V} \hat{V}^T \tilde{V} \tilde{V}^T \bar{X}_2.$$

Since $\|V^T \bar{X}_i\|^2 = O_P(\frac{p^\beta}{n_1})$ for $i = 1, 2$, we have

$$\begin{aligned} |M_1| &\leq \frac{n_1}{\sqrt{p}} \|I_r - V^T \hat{V} \hat{V}^T V\| \|V^T \bar{X}_1\| \|V^T \bar{X}_2\| \\ &\leq \frac{n_1}{2\sqrt{p}} \|V V^T - \hat{V} \hat{V}^T\|_F^2 \|V^T \bar{X}_1\| \|V^T \bar{X}_2\| \\ &= O_P(1) \frac{n_1}{\sqrt{p}} \frac{p}{p^\beta n_1} \frac{p^\beta}{n_1} \xrightarrow{P} 0. \end{aligned} \tag{75}$$

The joint distribution of $(V^T \bar{X}_2, \tilde{V}^T \bar{X}_1, S)$ is identity to that of $(V^T \bar{X}_1, \tilde{V}^T \bar{X}_1, S)$. Therefore, M_2 has the same distribution with Q_2 and converges to 0 in probability. The same reasoning yields $M_3 \xrightarrow{P} 0$.

Applying cauchy inequality gives

$$M_4 \leq \sqrt{\frac{n_1}{\sqrt{p}} \bar{X}_1^T \tilde{V} \tilde{V}^T \hat{V} \hat{V}^T \tilde{V} \tilde{V}^T \bar{X}_1} \sqrt{\frac{n_1}{\sqrt{p}} \bar{X}_2^T \tilde{V} \tilde{V}^T \hat{V} \hat{V}^T \tilde{V} \tilde{V}^T \bar{X}_2}.$$

Then $M_4 \xrightarrow{P} 0$ by the same reason as $Q_3 \xrightarrow{P} 0$, and consequently $P_3 \xrightarrow{P} 0$. This finishes the proof.

□

Proof Of Theorem 3. By Chen and Qin (2010)'s Theorem 1, we have

$$\frac{n_1 n_2 (T_1 - \|\tilde{V}(\mu_1 - \mu_2)\|^2)}{\sqrt{2p}(n_1 + n_2)\sigma^2} \xrightarrow{\mathcal{L}} N(0, 1).$$

It suffices to prove

$$\frac{n_1 n_2 (T_1 - T_2)}{\sqrt{2p}(n_1 + n_2)\sigma^2} \xrightarrow{P} 0.$$

Write $T_1 = T_1^{(1)} + T_1^{(2)}$, where

$$\begin{aligned} T_1^{(1)} =& \frac{\sum_{i \neq j}^{n_1} (X_{1i} - \mu_1)^T \tilde{V} \tilde{V}^T (X_{1j} - \mu_1)}{n_1(n_1 - 1)} + \frac{\sum_{i \neq j}^{n_2} (X_{2i} - \mu_2)^T \tilde{V} \tilde{V}^T (X_{2j} - \mu_2)}{n_2(n_2 - 1)} \\ &- 2 \frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (X_{1i} - \mu_1)^T \tilde{V} \tilde{V}^T (X_{2j} - \mu_2)}{n_1 n_2} \end{aligned}$$

and

$$
\begin{aligned}
T_1^{(2)} =& 2(\bar{X}_1-\mu_1)^T\tilde{V}\tilde{V}^T(\mu_1-\mu_2)+(\bar{X}_2-\mu_2)^T\tilde{V}\tilde{V}^T(\mu_2-\mu_1)\\
&+\|\tilde{V}(\mu_1-\mu_2)\|^2.
\end{aligned}
$$

Similarly, $T_2 = T_2^{(1)} + T_2^{(2)}$. By Theorem 2, it holds that

$$
\frac{n_1n_2(T_2^{(1)}-T_1^{(1)})}{\sqrt{2p}(n_1+n_2)\sigma^2} \xrightarrow{P} 0. \tag{76}
$$

We are left with the task of dealing with $T_1^{(2)}$ and $T_2^{(2)}$. Note that

$$
\begin{aligned}
\frac{n_1n_2(T_2^{(2)}-T_1^{(2)})}{\sqrt{2p}(n_1+n_2)\sigma^2} =& O(1)\frac{n_1}{\sqrt{p}}(\bar{X}_1-\mu_1)^T(VV^T-\hat{V}\hat{V}^T)(\mu_1-\mu_2)+ &(77)\\
& O(1)\frac{n_2}{\sqrt{p}}(\bar{X}_2-\mu_2)^T(VV^T-\hat{V}\hat{V}^T)(\mu_2-\mu_1)+ &(78)\\
& O(1)\frac{n_1+n_2}{\sqrt{p}}(\mu_1-\mu_2)^T(VV^T-\hat{V}\hat{V}^T)(\mu_1-\mu_2). &(79)
\end{aligned}
$$

We proceed to deal with the (77).

$$
\begin{aligned}
&E|\frac{n_1}{\sqrt{p}}(\bar{X}_1-\mu_1)^T(VV^T-\hat{V}\hat{V}^T)(\mu_1-\mu_2)|^2 &(80)\\
&=\frac{n_1^2}{p}E(\mu_1-\mu_2)^T(VV^T-\hat{V}\hat{V}^T)\frac{\Sigma}{n_1}(VV^T-\hat{V}\hat{V}^T)(\mu_1-\mu_2)\\
&=O(n_1p^{\beta-1})E(\mu_1-\mu_2)^T(VV^T-\hat{V}\hat{V}^T)^2(\mu_1-\mu_2)\\
&\le O(n_1p^{\beta-1})\|\mu_1-\mu_2\|^2E\|VV^T-\hat{V}\hat{V}^T\|_F^2\\
&=O(n_1p^{\beta-1})\frac{p}{p^\beta n_1}\|\mu_1-\mu_2\|^2\\
&=O(\|\mu_1-\mu_2\|^2).
\end{aligned}
$$

Combining with conditions $\frac{(n_1+n_2)}{\sqrt{p}}\|\mu_1-\mu_2\|^2 = O(1)$ and $\frac{\sqrt{p}}{n_1+n_2} \to 0$, it follows that (80) converges to 0 in probability. By symmetry, (78) also converges to 0 in probability. It remains to deal with (79). But

$$
\begin{aligned}
&E|\frac{n_1+n_2}{\sqrt{p}}(\mu_1-\mu_2)^T(VV^T-\hat{V}\hat{V}^T)(\mu_1-\mu_2)| &(81)\\
&\le \frac{n_1+n_2}{\sqrt{p}}\|\mu_1-\mu_2\|^2E\|VV^T-\hat{V}\hat{V}^T\|\\
&\le \frac{n_1+n_2}{\sqrt{p}}\|\mu_1-\mu_2\|^2\sqrt{E\|VV^T-\hat{V}\hat{V}^T\|^2}\\
&\le O(1)\sqrt{E\|VV^T-\hat{V}\hat{V}^T\|_F^2}.
\end{aligned}
$$

Since $E\|VV^T - \hat{V}\hat{V}^T\|_F^2 = O(\frac{p}{p^\beta n_1})$ and $\beta \geq 1/2$, (81) converges to 0.

It follows that
$$\frac{n_1 n_2 (T_2^{(2)} - T_1^{(2)})}{\sqrt{2p}(n_1+n_2)\sigma^2} \xrightarrow{P} 0.$$
Together with (76), the theorem follows. □

Proof Of Theorem 4. The theorem follows by Theorem 1, Theorem 3 and Theorem 2. □

Proof Of Theorem 5. The proof is based on Theorem 2 and procedure runs almost the same. The notation is parallel to Theorem 2. Denote by $O_{m\times n}$ the $m \times n$ matrix with all elements equal to 0. First we consider term
$$R_1' = \frac{n_1}{\sqrt{p}}(\bar{X}_1^T VV^T \bar{X}_1 - \bar{X}_1^T \hat{V}\hat{V}^T \bar{X}_1),$$
and
$$R_2' = \frac{1}{n_1\sqrt{p}}\sum_{i=1}^{n_1}(X_{1i}^T VV^T X_{1i} - X_{1i}^T \hat{V}\hat{V}^T X_{1i}).$$
We note that $VV^T \geq V_1V_1^T$. Define $V_{2\ominus 1}V_{2\ominus 1}^T = VV^T - V_1V_1^T$. From the fact that $rank(V) \leq r_1 + r_2$ and $V_{2\ominus 1}^T V_1 = O_{r_2\times r_1}$, it can easily deduced that $\frac{n_1}{\sqrt{p}}\bar{X}_1^T VV^T\bar{X}_1 - \frac{n_1}{\sqrt{p}}\bar{X}_1^T V_1V_1^T\bar{X}_1 \xrightarrow{P} 0$. Hence to proof $R_1' \xrightarrow{P} 0$, we only need to consider
$$\frac{n_1}{\sqrt{p}}(\bar{X}_1^T V_1V_1^T \bar{X}_1 - \bar{X}_1^T \hat{V}\hat{V}^T \bar{X}_1),$$
which can be further decomposed into three parts

$$Q_1' = \frac{n_1}{\sqrt{p}}\bar{X}_1^T V_1(I_{r_1} - V_1^T\hat{V}\hat{V}^T V_1)V_1^T\bar{X}_1,$$
$$Q_2' = \frac{n_1}{\sqrt{p}}\bar{X}_1^T \tilde{V}_1\tilde{V}_1^T\hat{V}\hat{V}^T V_1V_1^T\bar{X}_1,$$
$$Q_3' = \frac{n_1}{\sqrt{p}}\bar{X}_1^T \tilde{V}_1\tilde{V}_1^T\hat{V}\hat{V}^T \tilde{V}_1\tilde{V}_1^T\bar{X}_1.$$
We note that $Q_1' \leq n_1\bar{X}_1^T V_1(I_{r_1} - V_1^T\hat{V}_1\hat{V}_1^T V_1)V_1^T\bar{X}_1/\sqrt{p}$ which converges to 0 by Theorem 2. Q_2' can be written as

$$Q_2' = \frac{n_1}{\sqrt{p}}\bar{X}_1^T \tilde{V}_1\tilde{V}_1^T\hat{V}_1\hat{V}_1^T V_1V_1^T\bar{X}_1 + \frac{n_1}{\sqrt{p}}\bar{X}_1^T \tilde{V}_1\tilde{V}_1^T\hat{V}_{2\ominus 1}\hat{V}_{2\ominus 1}^T V_1V_1^T\bar{X}_1.$$

The first term converges to 0 by Theorem 2. For the second term we have

$$\begin{aligned}
&\frac{n_1}{\sqrt{p}}\bar{X}_1^T\tilde{V}_1\tilde{V}_1^T\hat{V}_{2\ominus 1}\hat{V}_{2\ominus 1}^T V_1 V_1^T\bar{X}_1 \\
&\leq \sqrt{\frac{n_1}{\sqrt{p}}\bar{X}_1^T\tilde{V}_1\tilde{V}_1^T\hat{V}_{2\ominus 1}\hat{V}_{2\ominus 1}^T\tilde{V}_1\tilde{V}_1^T\bar{X}_1}\sqrt{\frac{n_1}{\sqrt{p}}\bar{X}_1^T V_1 V_1^T\hat{V}_{2\ominus 1}\hat{V}_{2\ominus 1}^T V_1 V_1^T\bar{X}_1}.
\end{aligned}$$

The first term converges to 0 because $\hat{V}_{2\ominus 1}^T\tilde{V}_1\tilde{V}_1^T\bar{X}_1|(S_1, S_2) \sim N(0, \frac{\sigma_1^2}{n_1}\hat{V}_{2\ominus 1}^T\tilde{V}_1\tilde{V}_1^T\hat{V}_{2\ominus 1})$ whose conditional variance is dominated by $\frac{\sigma_1^2}{n_1}$. And for the second term, we have

$$\begin{aligned}
\frac{n_1}{\sqrt{p}}\bar{X}_1^T V_1 V_1^T\hat{V}_{2\ominus 1}\hat{V}_{2\ominus 1}^T V_1 V_1^T\bar{X}_1 &\leq \frac{n_1}{\sqrt{p}}\bar{X}_1^T V_1 V_1^T\hat{\hat{V}}_1\hat{\hat{V}}_1^T V_1 V_1^T\bar{X}_1 \\
&\leq \frac{n_1}{\sqrt{p}}\|V_1^T\hat{\hat{V}}_1\hat{\hat{V}}_1^T V_1\|\|V_1^T\bar{X}_1\|^2 \\
&\leq O(1)\frac{n_1}{\sqrt{p}}\|V_1V_1^T - \hat{V}_1\hat{V}_1^T\|_F^2\frac{p^\beta}{n_1} \\
&= O(\frac{\sqrt{p}}{n_1}) \xrightarrow{P} 0.
\end{aligned}$$

$Q_3' = \frac{n_1}{\sqrt{p}}\bar{X}_1^T\tilde{V}_1\tilde{V}_1^T\hat{V}_{2\ominus 1}\hat{V}_{2\ominus 1}^T\tilde{V}_1\tilde{V}_1^T\bar{X}_1 + \frac{n_1}{\sqrt{p}}\bar{X}_1^T\tilde{V}_1\tilde{V}_1^T\hat{V}_1\hat{V}_1^T\tilde{V}_1\tilde{V}_1^T\bar{X}_1$ converges to 0 in probability by combining Theorem 2 and the argument above. Hence $R_1' \xrightarrow{P} 0$.

To prove $R_2' \xrightarrow{P} 0$, we note that $\frac{1}{n_1\sqrt{p}}\sum_{i=1}^{n_1}X_{1i}^T VV^T X_{1i} - \frac{1}{n_1\sqrt{p}}\sum_{i=1}^{n_1}X_{1i}^T V_1V_1^T X_{1i} \xrightarrow{P} 0$. Hence it suffices to consider the following three parts

$$W_1' = \frac{1}{n_1\sqrt{p}}\sum_{i=1}^{n_1}X_{1i}^T V_1(I_{r_1} - V_1^T\hat{V}\hat{V}^T V_1)V_1^T X_{1i},$$

$$W_2' = \frac{1}{n_1\sqrt{p}}\sum_{i=1}^{n_1}X_{1i}^T\tilde{V}_1\tilde{V}_1^T\hat{V}\hat{V}^T V_1V_1^T X_{1i}$$

and

$$W_3' = \frac{1}{n_1\sqrt{p}}\sum_{i=1}^{n_1}X_{1i}^T\tilde{V}_1\tilde{V}_1^T\hat{V}\hat{V}^T\tilde{V}_1\tilde{V}_1^T X_{1i}.$$

$W_1' \xrightarrow{P} 0$ by $\hat{V}\hat{V}^T \leq \hat{V}_1\hat{V}_1^T$ and Theorem 2.

$$W_2' = \frac{1}{n_1\sqrt{p}}\sum_{i=1}^{n_1}X_{1i}^T\tilde{V}_1\tilde{V}_1^T\hat{V}_1\hat{V}_1^T V_1V_1^T X_{1i} + \frac{1}{n_1\sqrt{p}}\sum_{i=1}^{n_1}X_{1i}^T\tilde{V}_1\tilde{V}_1^T\hat{V}_{2\ominus 1}\hat{V}_{2\ominus 1}^T V_1V_1^T X_{1i}.$$

The first term converges to 0 by Theorem 2. As for the second term, we can use the similar technique as we deal with W_2 and the fact $\hat{V}_{2\ominus 1}\hat{V}_{2\ominus 1}^T \leq \hat{\hat{V}}\hat{\hat{V}}^T$. The proof of $W_3' \xrightarrow{P} 0$ runs the same as the case of W_3.

Finally it suffices to proof following four terms converges to 0 in probability

$$M_1' = \frac{n_1}{\sqrt{p}}\bar{X}_1^T V(I_{r_1+r_2} - V^T\hat{V}\hat{V}^T V)V^T\bar{X}_2,$$
$$M_2' = \frac{n_1}{\sqrt{p}}\bar{X}_1^T \tilde{V}\tilde{V}^T\hat{V}\hat{V}^T VV^T\bar{X}_2,$$
$$M_3' = \frac{n_1}{\sqrt{p}}\bar{X}_1^T VV^T\hat{V}\hat{V}^T\tilde{V}\tilde{V}^T\bar{X}_2,$$

and

$$M_4' = \frac{n_1}{\sqrt{p}}\bar{X}_1^T \tilde{V}\tilde{V}^T\hat{V}\hat{V}^T\tilde{V}\tilde{V}^T\bar{X}_2.$$

We note that

$$M_1' \le \sqrt{\frac{n_1}{\sqrt{p}}\bar{X}_1^T V(I_{r_1+r_2} - V^T\hat{V}\hat{V}^T V)V^T\bar{X}_1}\sqrt{\frac{n_1}{\sqrt{p}}\bar{X}_1^T V(I_{r_1+r_2} - V^T\hat{V}\hat{V}^T V)V^T\bar{X}_2}.$$

Hence to prove $M_1' \xrightarrow{P} 0$, it suffices to prove $\frac{n_1}{\sqrt{p}}\bar{X}_1^T V(I_{r_1+r_2} - V^T\hat{V}\hat{V}^T V)V^T\bar{X}_1 \xrightarrow{P} 0$. We note that

$$\begin{aligned}
&\frac{n_1}{\sqrt{p}}\bar{X}_1^T V(I_{r_1+r_2} - V^T\hat{V}\hat{V}^T V)V^T\bar{X}_1\\
&= (\frac{n_1}{\sqrt{p}}\bar{X}_1^T V_1V_1^T\bar{X}_1 - \frac{n_1}{\sqrt{p}}\bar{X}_1^T V_1V_1^T\hat{V}\hat{V}^T V_1V_1^T\bar{X}_1)\\
&+ \frac{n_1}{\sqrt{p}}\bar{X}_1^T V_{2\ominus 1}V_{2\ominus 1}^T\bar{X}_1 - 2\frac{n_1}{\sqrt{p}}\bar{X}_1^T V_{2\ominus 1}V_{2\ominus 1}^T\hat{V}\hat{V}^T V_1V_1^T\bar{X}_1\\
&- \frac{n_1}{\sqrt{p}}\bar{X}_1^T V_{2\ominus 1}V_{2\ominus 1}^T\hat{V}\hat{V}^T V_{2\ominus 1}V_{2\ominus 1}^T\bar{X}_1\\
&= L_1 + L_2 - 2L_3 - L_4.
\end{aligned}$$

Check that $L_1 = Q_1'$. It's also clear that $L_2 \xrightarrow{P} 0$. To deal with L_3, we note that

$$\begin{aligned}
&|\frac{n_1}{\sqrt{p}}\bar{X}_1^T V_{2\ominus 1}V_{2\ominus 1}^T\hat{V}_1\hat{V}_1^T V_1V_1^T\bar{X}_1|\\
&\le \frac{n_1}{\sqrt{p}}\|V_{2\ominus 1}^T\bar{X}_1\|\|V_{2\ominus 1}^T\hat{V}_1\|_F\|\hat{V}_1^T V_1V_1^T\bar{X}_1\| \qquad (82)\\
&\le \frac{n_1}{\sqrt{p}}O(\frac{1}{\sqrt{n_1}})\sqrt{tr(\hat{V}_1^T V_{2\ominus 1}V_{2\ominus 1}^T\hat{V}_1)}O(\frac{\sqrt{p^\beta}}{\sqrt{n_1}}).
\end{aligned}$$

But

$$tr(\hat{V}_1^T V_{2\ominus 1}V_{2\ominus 1}^T\hat{V}_1) \le tr(\hat{V}_1^T\tilde{V}_1\tilde{V}_1^T\hat{V}_1) = \|V_1V_1^T - \hat{V}_1\hat{V}_1^T\|_F^2 = O_P(\frac{p}{p^\beta n_1}).$$

And

$$|\frac{n_1}{\sqrt{p}}\bar{X}_1^T V_{2\ominus 1} V_{2\ominus 1}^T \hat{V}_{2\ominus 1} \hat{V}_{2\ominus 1}^T V_1 V_1^T \bar{X}_1| \leq \frac{n_1}{\sqrt{p}} \|V_{2\ominus 1}^T \hat{V}_{2\ominus 1} V_{2\ominus 1}^T \bar{X}_1\| \|\hat{V}_{2\ominus 1}^T V_1\|_F \|V_1^T \bar{X}_1\|$$
$$\leq \frac{n_1}{\sqrt{p}} O(\frac{1}{\sqrt{n_1}}) \sqrt{tr(\hat{V}_{2\ominus 1}^T V_1 V_1^T \hat{V}_{2\ominus 1})} O(\frac{\sqrt{p^\beta}}{\sqrt{n_1}}). \tag{83}$$

But

$$tr(\hat{V}_{2\ominus 1}^T V_1 V_1^T \hat{V}_{2\ominus 1}) \leq tr(\hat{\hat{V}}_1^T V_1 V_1^T \hat{\hat{V}}_1) = \|V_1 V_1^T - \hat{V}_1 \hat{V}_1^T\|_F^2 = O_P(\frac{p}{p^\beta n_1}).$$

Therefore $L_3 \xrightarrow{P} 0$. And $L_4 \xrightarrow{P} 0$ for trival reason. M_2' can be similarly treated by technique (82) and (83). Since $\sqrt{n_1}\hat{V}^T \tilde{V} \tilde{V}^T \bar{X}_2$ is bounded in probability, we have $M_4' \xrightarrow{P} 0$, which completes the proof.

□

Proof Of Theorem 6. The theorem follows by the same method as Theorem 3. □

Acknowledgements

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

References

Ahn SC, Horenstein AR. Eigenvalue ratio test for the number of factors. Econometrica 2013;81(3):1203–27. doi:10.3982/ECTA8968.

Bai Z, Saranadasa H. Effect of high dimension: by an example of a two sample problem. Statist Sinica 1996;6(2):311–29.

Bai Z, Silverstein JW. Spectral analysis of large dimensional random matrices. 2nd ed. Springer Series in Statistics. Springer, New York, 2010. doi:10.1007/978-1-4419-0661-8.

Cai TT, Liu W, Xia Y. Two-sample test of high dimensional means under dependence. J R Stat Soc Ser B Stat Methodol 2014;76(2):349–72. doi:10.1111/rssb.12034.

Cai TT, Ma Z, Wu Y. Sparse PCA: optimal rates and adaptive estimation. Ann Statist 2013;41(6):3074–110. doi:10.1214/13-AOS1178.

Chen LS, Paul D, Prentice RL, Wang P. A regularized Hotelling's T^2 test for pathway analysis in proteomic studies. J Amer Statist Assoc 2011;106(496):1345–60. doi:10.1198/jasa.2011.ap10599.

Chen SX, Qin YL. A two-sample test for high-dimensional data with applications to gene-set testing. Ann Statist 2010;38(2):808–35. doi:10.1214/09-AOS716.

El Karoui N. On the largest eigenvalue of Wishart matrices with identity covariance when n, p and p/n tend to infinity. ArXiv Mathematics e-prints 2003;arXiv:math/0309355.

Johnstone IM. On the distribution of the largest eigenvalue in principal components analysis. Ann Statist 2001;29(2):295–327. doi:10.1214/aos/1009210544.

Lehmann EL, Romano JP. Testing statistical hypotheses. 3rd ed. Springer Texts in Statistics. Springer, New York, 2005.

Lopes M, Jacob L, Wainwright MJ. A more powerful two-sample test in high dimensions using random projection. In: Shawe-Taylor J, Zemel RS, Bartlett PL, Pereira F, Weinberger KQ, editors. Advances in Neural Information Processing Systems 24. Curran Associates, Inc.; 2011. p. 1206–14.

Ma Y, Lan W, Wang H. A high dimensional two-sample test under a low dimensional factor structure. J Multivariate Anal 2015;140:162–70. doi:10.1016/j.jmva.2015.05.005.

Srivastava MS, Du M. A test for the mean vector with fewer observations than the dimension. Journal of Multivariate Analysis 2008;99(3):386 − 402. doi:http://dx.doi.org/10.1016/j.jmva.2006.11.002.

Srivastava R, Li P, Ruppert D. Raptt: An exact two-sample test in high dimensions using random projections. Journal of Computational and Graphical Statistics 2015;doi:10.1080/10618600.2015.1062771.

Thulin M. A high-dimensional two-sample test for the mean using random subspaces. Computational Statistics & Data Analysis 2014;74:26 − 38. doi:http://dx.doi.org/10.1016/j.csda.2013.12.003.

Zhao J, Xu X. A generalized likelihood ratio test for normal mean when p is greater than n. Comput Statist Data Anal 2016;99:91–104. doi:10.1016/j.csda.2016.01.006.