

# High-dimensional two-sample test under spiked covariance

Rui Wang<sup>a</sup>, Xingzhong Xu<sup>a,b,\*</sup>

<sup>a</sup>*School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, China*

<sup>b</sup>*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  $\mu_k$  and  $\Sigma_k$  are unknown,  $k = 1, 2$ . We consider the hypothesis testing problem:

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

---

\*Corresponding author

Email address: `xuzz@bit.edu.cn` (Xingzhong Xu)

In this paper, high dimensional setting is adopted, i.e., the dimension  $p$  varies as  $n$  increase, where  $n = n_1 + n_2$ . 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  $p \geq n - 1$ . Moreover, Bai and Saranadasa (1996) showed that even if  $p < n - 1$ , 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  $\Sigma$  when  $p$  is large compared with  $n$ . 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

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

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

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

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  $A = [\text{diag}(\Sigma)]^{-1}$ , Srivastava and Du (2008) proposed a test based on

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

where  $\text{diag}(A)$  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.,  $\Sigma$  is far away from diagonal matrix. For example, the assumption

$$\text{tr}(\Sigma^4) = o[\text{tr}^2\{(\Sigma)^2\}] \quad (2)$$

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. Ma et al. (2015) considered a factor model and proposed a asymptotical parameter bootstrap procedure to adjust Chen and Qin (2010)'s critical value.

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  $\Sigma$  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  $\{X_{k1}, \dots, X_{kn_k}\}$ ,  $k = 1, 2$  be two independent random samples from  $p$  dimensional normal distribution with means  $\mu_1$  and  $\mu_2$  respectively.

**Assumption 1.** Assume  $p \rightarrow \infty$  as  $n \rightarrow \infty$ . Furthermore, assume two samples are balanced, that is,

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

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  $X_{ki}$ ,  $i = 1, 2, \dots, n_k$  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,  $D_k = \text{diag}(\lambda_{k1}^{\frac{1}{2}}, \dots, \lambda_{kr_k}^{\frac{1}{2}})$  with  $\lambda_{k1} \geq \dots \geq \lambda_{kr_k} > 0$ ,  $V_k$  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  $\Sigma_k = \text{Var}(X_{ki}) = V_k \Lambda_k V_k^T + \sigma_k^2 I_p$ ,  $\Lambda_k = D_k^2 = \text{diag}(\lambda_{k1}, \dots, \lambda_{kr_k})$ . 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  $\kappa$  such that

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

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 met 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. The case  $\beta = 1$  corresponds to the factor model in paper Ma et al. (2015) with some restrictions of parameters.

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

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

where

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

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  $\Sigma_k$ ,  $V_k$ ,  $D_k$ ,  $\Lambda_k$ ,  $\sigma_k$  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  $ET = 0$  under null hypothesis and  $ET > 0$  under alternative. To determine the critical value, the asymptotic distribution of  $T$  need to be derived, say

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

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

$$\Phi\left(\frac{ET}{\sqrt{\text{Var}(T)}} - z_{1-\alpha}\right).$$

Thus, a general idea to boost the power of test is to reduce the variance  $\text{Var}(T)$  while the mean  $E(T)$  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  $\Sigma$  by  $\Sigma = \sum_{i=1}^p \lambda_i p_i p_i^T$  with  $\lambda_1 \geq \dots \geq \lambda_p$ , where  $p_i$ ,  $i = 1, \dots, p$ , are orthonormal  $p$  dimensional vectors. The main body of both  $T_{BS}$  and  $T_{CQ}$  is

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

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  $\Sigma$  is known, the LRT is based on

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

The difference between (5) and (6) is the weights  $\lambda_i^{-1}$ . Unfortunately,  $\lambda_i$ 's are hard to precisely estimate in high dimensional setting. See Bai and Silverstein (2010) for detail. Nevertheless, it's possible to identify which  $\lambda_i$ '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  $\Sigma$  are  $\lambda_1 + \sigma^2, \dots, \lambda_r + \sigma^2, \sigma^2, \dots, \sigma^2$ . 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

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

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

**Proposition 1.** *Under Assumptions 1-4 and local alternative, that means,  $\frac{n}{p}\|\mu_1 - \mu_2\|^2 \rightarrow 0$ , 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

We denote by  $\hat{V}$  and  $\hat{\hat{V}}$  the first  $r$  and last  $p-r$  eigenvectors of  $S$  respectively. Similarly, we denote by  $\hat{V}_i$  and  $\hat{\hat{V}}_i$  the first  $r$  and last  $p-r$  eigenvectors of  $S_i$  respectively,  $i = 1, 2$ . As the estimators of their population counterparts, these simple statistics actually reach the optimal convergence rate. See Cai et al. (2013).

Since  $T_1$  depends on subspace  $\tilde{V}\tilde{V}^T$  which is unknown, we must estimate it. The first part of  $T_1$  is  $\|\tilde{V}^T(\bar{X}_1 - \bar{X}_2)\|^2$ . We estimate it directly by  $\|\hat{\tilde{V}}^T(\bar{X}_1 - \bar{X}_2)\|^2$ . Note that the second part of  $T_1$  is  $\frac{1}{n_1}\text{tr}(\tilde{V}^T S_1 \tilde{V})$ . Since it only involves sample one, we estimate it by  $\frac{1}{n_1}\text{tr}(\hat{\tilde{V}}_1^T S_1 \hat{\tilde{V}}_1)$ . Similarly, we estimate the third part of  $T_1$  by  $\frac{1}{n_2}\text{tr}(\hat{\tilde{V}}_2^T S_2 \hat{\tilde{V}}_2)$ .

Define

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

We propose our new test statistic as

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

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

**Remark 2.** Compared with random projection method, our projection is determined by the structure of  $S_1$ ,  $S_2$  and  $S$ . We don't project multiple times as random projection method did, which leads to reproducibility in practice.

**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  $X_{2j} \mapsto X_{2j} + \mu$ ,  $i = 1, \dots, n_1$ ,  $j = 1, \dots, n_2$ .

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

$$\hat{r} = \text{argmax}_{l \leq R} \frac{\lambda_l(S)}{\lambda_{l+1}(S)}, \quad (8)$$

where  $R$  is a hyperparameter. 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



$\|\tilde{V}(\mu_1 - \mu_2)\|^2$ . Then the superiority of the new test will be established if

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

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  $\|I_p - \tilde{V}\tilde{V}^T\|_F^2 / \|I_p\|_F^2 = r/p \rightarrow 0$ . In bayesian framework, if we assume that the elements of  $\mu_k$  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  $\mu_k$  will be used in Theorem 1.

In order to formulate a test procedure,  $\sigma^2$  needs to be estimated. Note that  $\sigma^2$  can be written as

$$\sigma^2 = \sum_{i=r+1}^p \lambda_i(\Sigma). \quad (10)$$

We estimate it by sample version:

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

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

#### 4. Theoretical results

In this section, we derive some properties of the new test statistic.

First we consider the case when the eigenvalues of  $\Sigma$  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). Similar to bayesian models, we assume a normal prior distribution for  $\mu_k$  to characterize ‘dense’ alternative. When the eigenvalues of  $\Sigma$  is bounded, spike variance model is not valid. Hence

some estimators in our test procedure make no sense. Particularly, if  $\hat{r}$  is estimated by (8). the  $\hat{r}$  is nothing but a random integer not greater than  $R$  and  $\hat{V}\hat{V}^T$  is just a random projection. In this case, the difference of our test statistic and Chen and Qin (2010)'s is small. The 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  $X_{ki} \sim N(\mu_k, \Sigma)$ ,  $i = 1, \dots, n_k$ ,  $k = 1, 2$ . 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  $\mu_k$  is independently generated by  $N(0, (n_k \sqrt{p})^{-1} \psi)$  for  $k = 1, 2$ ,  $\psi$  is a constant and  $\hat{r}$  is estimated by (8). If  $p \rightarrow \infty$  as  $n \rightarrow \infty$  and  $p = o(n^2)$ , then we have*

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

Now we establish the asymptotic normality of the new test statistic under spiked covariance model. We first give a result of the convergence rate of  $\hat{\sigma}^2$ .

**Proposition 2.** *Under Assumptions 1-4, we have that*

$$\hat{\sigma}^2 = \sigma^2 + O_P\left(\frac{\max(n, p)}{np}\right).$$

Proposition 2 implies that  $\hat{\sigma}^2$  is a consistent estimator of  $\sigma^2$ .

To derive the asymptotic normality of the new test statistic, we require the following relationship of  $n$  and  $p$ .

**Assumption 5.** *Assume  $\sqrt{p}/n \rightarrow 0$ .*

The assumption 5 looks restrictive. However, it may not be possible to relaxed since the asymptotic normality of  $\|\hat{V}^T(\bar{X}_1 - \bar{X}_2)\|^2$  requires

$$\frac{\lambda_1((\hat{V}^T \Sigma \hat{V})^2)}{\text{tr}((\hat{V}^T \Sigma \hat{V})^2)} \xrightarrow{P} 0, \quad (11)$$

see Example 5.1 of Jiang (1996). And (11) is equivalent to Assumption 5 by Lemma 2 in appendix.

Our first step is to prove the asymptotic normality under null hypothesis.

**Theorem 2.** *Under Assumptions 1-5, if the null hypothesis holds, then*

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

Theorem 2 can be generalized to the case of local alternatives.

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

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

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.

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

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

The power of  $T_{CQ}$  is of the form

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

The relative efficiency of our test with respect to Chen's test is

$$\sqrt{\frac{\text{tr}\Sigma^2}{(p-r)\sigma^4}} \frac{\|\tilde{V}(\mu_1 - \mu_2)\|^2}{\|\mu_1 - \mu_2\|^2} \sim p^{\beta-1/2} \frac{\|\tilde{V}(\mu_1 - \mu_2)\|^2}{\|\mu_1 - \mu_2\|^2},$$

which is large when  $\beta > 1/2$  and  $\|\tilde{V}(\mu_1 - \mu_2)\|/\|\mu_1 - \mu_2\|$  is close to 1.

## 5. Unequal Variance

In this section, we concern the situation with unequal covariance matrices. Assume  $\{X_{11}, \dots, X_{1n_1}\}$  and  $\{X_{21}, \dots, X_{2n_2}\}$  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,

$$VV^T = (V_1, V_2)((V_1, V_2)^T(V_1, V_2))^{+}(V_1, V_2)^T.$$

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

$$T_3 = \|\hat{\tilde{V}}^T(\bar{X}_1 - \bar{X}_2)\|^2 - \frac{1}{n_1}\text{tr}(\hat{\tilde{V}}^T S_1 \hat{\tilde{V}}) - \frac{1}{n_2}\text{tr}(\hat{\tilde{V}}^T S_2 \hat{\tilde{V}}).$$

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

**Theorem 4.** *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 5.** 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 5.** *Under Assumptions 1-3 and 5, if the local alternative holds, that is,*

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

then

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

$\sigma_n^2$  can be estimated by ratio consistent estimator of  $\sigma_k^2$  for  $k = 1, 2$ . Thus, if  $n$  and  $p$  are large and  $\sqrt{p}/n$  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  $\lambda_i$  as  $p^\beta$  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 is not satisfied. Level of the new test is a little inflated compared with oracle level and it performs better when  $n$  is larger.

Table 1: Test level simulation

$n$	$p$	$\beta=0.5$		$\beta=1$		$\beta=2$	
		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

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  $\Sigma$  is spiked, the new test outperforms  $T_{CQ}$  substantially; when  $\Sigma$  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 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 eigenvalue 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

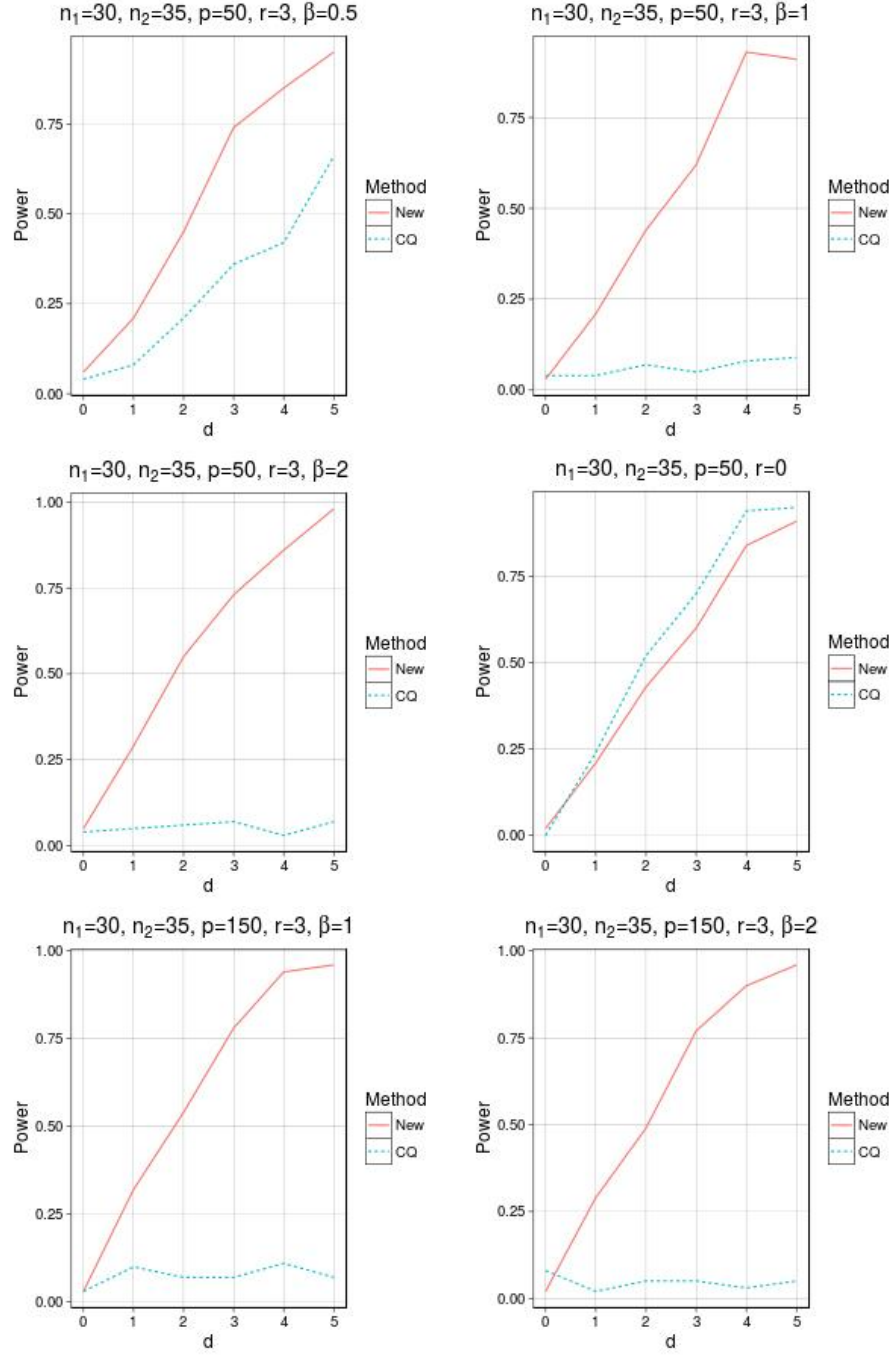


Figure 1: Empirical power simulation.  $\alpha$  is set to be 0.05.  $d$  is proportional to  $\|\mu_1 - \mu_2\|^2$ . For each simulation, we do 50 permutations to determine critical value. We generate 100 independent samples to compute empirical power.

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  $\sqrt{p}/n \rightarrow 0$ . 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 study the asymptotic behavior of permutation based test 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 2.** *Let  $H$  and  $P$  be two symmetric matrices and  $M = H + P$ . If  $\text{rank}(P) < k$ , then*

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

**Lemma 2** (Convergence rate of principal space estimation). *Under the Assumption 1-4, we have*

$$E\|\hat{V}\hat{V}^T - VV^T\|_F^2 = O\left(\frac{p}{p^\beta n}\right).$$

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

$$E\|\hat{V}\hat{V}^T - VV^T\|_F^2 \asymp r \wedge (p - r) \wedge \frac{r(p - r)}{(n_1 + n_2 - 2)h(\lambda)} \quad (12)$$



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 (12) is of order  $p^{1-\beta}/n$ . We note that it is assumed  $\beta \geq \frac{1}{2}$  in Assumption 3, together with  $\sqrt{p}/n \rightarrow 0$  we have  $p^{1-\beta}/n \rightarrow 0$ . Hence  $\hat{V}\hat{V}^T$  reaches the convergence rate.  $\square$

**Lemma 3** (Bai-Yin's law). *Suppose  $B_n = \frac{1}{q}ZZ^T$  where  $Z$  is  $p \times q$  random matrix composed of i.i.d. random variables with zero mean, unit variance and finite fourth moment. As  $q \rightarrow \infty$  and  $\frac{p}{q} \rightarrow c \in [0, \infty)$ , the largest and smallest non-zero eigenvalues of  $B_n$  converge almost surely to  $(1 + \sqrt{c})^2$  and  $(1 - \sqrt{c})^2$ , respectively.*

**Remark 6.** Lemma 3 is known as the Bai-Yin's law (Bai and Yin (1993)). As in Remark 1 of Bai and Yin (1993), the smallest non-zero eigenvalue is the  $p - q + 1$  smallest eigenvalue of  $B$  for  $c > 1$ .

**Corollary 3.** *Suppose that  $W_n$  is a  $p \times p$  matrix distributed as  $\text{Wishart}_p(n, I_p)$ . Then as  $n \rightarrow \infty$ ,*

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

**Proof.** Since  $[0, +\infty]$  is compact, for every subsequence  $\{n_k\}$  of  $\{n\}$ , there is a further subsequence  $\{n_{k_l}\}$  along which  $p/n \rightarrow c \in [0, +\infty]$ .

If  $c \in [0, +\infty)$ , by Lemma 3, we have that

$$\frac{\lambda_1(W_{n_{k_l}})}{n_{k_l}} \xrightarrow{P} (1 + c)^2.$$

Hence the conclusion holds along this subsequence. If  $c = +\infty$ , suppose  $W_n = Z_n Z_n^T$  where  $Z_n$  is a  $p \times n$  matrix with all elements distributed as  $N(0, 1)$ . Then

$$\frac{\lambda_1(W_{n_{k_l}})}{p} = \frac{Z_{n_{k_l}}^T Z_{n_{k_l}}}{p} \xrightarrow{P} 1,$$

by Lemma 3, which proves the conclusion along the subsequence. Now the conclusion holds by a standard subsequence argument.  $\square$

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

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

Note that

$$\begin{aligned} (\mu_1 - \mu_2)^T \tilde{V} (\tilde{V}^T \Sigma \tilde{V}) \tilde{V}^T (\mu_1 - \mu_2) &= \sigma^2 (\mu_1 - \mu_2)^T \tilde{V} \tilde{V}^T (\mu_1 - \mu_2) \\ &\leq \sigma^2 (\mu_1 - \mu_2)^T (\mu_1 - \mu_2) = o\left(\frac{p}{n}\right) = o\left(\frac{1}{n} \text{tr}(\tilde{V}^T \Sigma \tilde{V})^2\right). \end{aligned}$$

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

**Proof Of Proposition 2.** Note that  $(n-2)S \sim \text{Wishart}_p(n-2, \Sigma)$ . Denote by  $\Sigma = OEO^T$  the spectral decomposition of  $\Sigma$ , where  $O$  is an orthogonal matrix and  $E = \text{diag}(\lambda_1 + \sigma^2, \dots, \lambda_r + \sigma^2, \sigma^2, \dots, \sigma^2)$ . Let  $Z$  be a  $p \times (n-2)$  random matrix with all elements i.i.d. distributed as  $N(0, 1)$ , then

$$S \sim \frac{1}{n-2} O E^{1/2} Z Z^T E^{1/2} O^T.$$

Thus,

$$\begin{aligned} \hat{\sigma}^2 &\sim \frac{1}{(p-r)(n-2)} \sum_{i=r+1}^p \lambda_i (O E^{1/2} Z Z^T E^{1/2} O^T) \\ &= \frac{1}{(p-r)(n-2)} \sum_{i=r+1}^{n-2} \lambda_i (Z^T E Z). \end{aligned}$$

Denote  $Z^T = (Z_{(1)}^T, Z_{(2)}^T)^T$ , where  $Z_{(1)}$  is the first  $r$  rows of  $Z$  and  $Z_{(2)}$  is the rest rows. We have

$$Z^T E Z = Z_{(1)}^T E_1 Z_{(1)} + \sigma^2 Z_{(2)}^T Z_{(2)},$$

where  $E_1 = \text{diag}(\lambda_1 + \sigma^2, \dots, \lambda_r + \sigma^2)$ . The first term is of rank  $r$ . By Weyl's inequality, we have for  $i = r+1, \dots, n-2$  that

$$\sigma^2 \lambda_i (Z_{(2)}^T Z_{(2)}) \leq \lambda_i (Z^T E Z) \leq \sigma^2 \lambda_{i-r} (Z_{(2)}^T Z_{(2)}).$$

It follows that

$$\sigma^2 \sum_{i=r+1}^{n-2} \lambda_i(Z_{(2)}^T Z_{(2)}) \leq \sum_{i=r+1}^{n-2} \lambda_i(Z^T E Z) \leq \sigma^2 \sum_{i=1}^{n-r-2} \lambda_i(Z_{(2)}^T Z_{(2)}).$$

Hence we have

$$\begin{aligned} & \left| \frac{1}{(p-r)(n-2)} \sum_{i=r+1}^{n-2} \lambda_i(Z^T E Z) - \frac{1}{(p-r)(n-2)} \sigma^2 \sum_{i=1}^{n-2} \lambda_i(Z_{(2)}^T Z_{(2)}) \right| \\ & \leq r \sigma^2 \frac{1}{(p-r)(n-2)} \lambda_1(Z_{(2)}^T Z_{(2)}). \end{aligned}$$

By Corollary 3,  $\lambda_1(Z_{(2)}^T Z_{(2)}) = O_P(\max(n, p))$ . Hence

$$\begin{aligned} & \frac{1}{(p-r)(n-2)} \sum_{i=r+1}^{n-2} \lambda_i(Z^T E Z) \\ & = \frac{1}{(p-r)(n-2)} \sigma^2 \sum_{i=1}^{n-2} \lambda_i(Z_{(2)}^T Z_{(2)}) + O_P\left(\frac{\max(n, p)}{np}\right) \\ & = \frac{1}{(p-r)(n-2)} \sigma^2 \text{tr}(Z_{(2)}^T Z_{(2)}) + O_P\left(\frac{\max(n, p)}{np}\right) \\ & = \sigma^2 + O_P\left(\frac{1}{\sqrt{np}}\right) + O_P\left(\frac{\max(n, p)}{np}\right). \end{aligned}$$

The last line of the above equality holds since  $\text{tr}(Z_{(2)}^T Z_{(2)})$  is the sum of square of the elements of  $Z_{(2)}$  and thus central limit theorem can be invoked. The theorem follows by noting that

$$O_P\left(\frac{1}{\sqrt{np}}\right) = O_P\left(\frac{\sqrt{np}}{np}\right) = O_P\left(\frac{\max(n, p)}{np}\right).$$

□

**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

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

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

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

almost surely (a.s.). Let

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

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

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

along the further subsequence. Therefore,

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

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

$$\frac{T_{CQ} - T_2}{\sqrt{2\tau^2 \text{tr}\Sigma^2}} \xrightarrow{P} 0. \quad (13)$$

We note that

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

And

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

which can be further written by

$$\begin{aligned} \frac{\sum_{i \neq j}^{n_1} X_{1i}^T \hat{V} \hat{V}^T X_{1j}}{n_1 \sqrt{p}} &= \frac{n_1(n_1 - 1) \bar{X}_1^T \hat{V} \hat{V}^T \bar{X}_1}{n_1 \sqrt{p}} - \frac{\sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^T \hat{V} \hat{V}^T (X_{1i} - \bar{X}_1)}{n_1 \sqrt{p}} \\ &\stackrel{\text{def}}{=} R_1 - R_2. \end{aligned}$$

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,

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

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

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

Lemma 3 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  $\frac{P_1}{\sqrt{2\tau^2 \text{tr}\Sigma^2}} \xrightarrow{P} 0$ . Similar arguments lead to  $\frac{P_2}{\sqrt{2\tau^2 \text{tr}\Sigma^2}} \xrightarrow{P} 0$ .

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

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

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

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

It suffices to prove

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

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  $\hat{\hat{V}}\hat{\hat{V}}^T - \tilde{V}\tilde{V}^T = VV^T - \hat{V}\hat{V}^T$  we have  $n_1 n_2 (T_1 - T_2) / (\sqrt{2pn}\sigma^2) = P_1 + P_2 - 2P_3$ , where

$$\begin{aligned} P_1 &= \frac{n_2 \sum_{i \neq j} X_{1i}^T (VV^T - \hat{V}\hat{V}^T) X_{1j}}{\sqrt{2p}(n_1 + n_2)(n_1 - 1)\sigma^2}, \\ P_2 &= \frac{n_1 \sum_{i \neq j} X_{2i}^T (VV^T - \hat{V}\hat{V}^T) X_{2j}}{\sqrt{2p}(n_1 + n_2)(n_2 - 1)\sigma^2}, \end{aligned}$$

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

Write

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

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

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

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

$$\begin{aligned} Q_1 &= \frac{n_1}{\sqrt{p}} \bar{X}_1^T V(I_r - V^T \hat{V}\hat{V}^T V) V^T \bar{X}_1, \\ Q_2 &= \frac{n_1}{\sqrt{p}} \bar{X}_1^T \tilde{V}\tilde{V}^T \hat{V}\hat{V}^T VV^T \bar{X}_1, \\ Q_3 &= \frac{n_1}{\sqrt{p}} \bar{X}_1^T \tilde{V}\tilde{V}^T \hat{V}\hat{V}^T \tilde{V}\tilde{V}^T \bar{X}_1. \end{aligned}$$

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.  $\|V^T \bar{X}_1\|^2 = O_P(\frac{p^\beta}{n_1})$ .  $I_r - V^T \hat{V}\hat{V}^T V$  is a positive semi-definite matrix which only relies on  $S$ . Combining these observations, we have

$$\begin{aligned} |Q_1| &\leq \frac{n_1}{\sqrt{p}} \|I_r - V^T \hat{V}\hat{V}^T V\| \|V^T \bar{X}_1\|^2 \\ &\leq \frac{n_1}{\sqrt{p}} \text{tr}(I_r - V^T \hat{V}\hat{V}^T V) \|V^T \bar{X}_1\|^2 \\ &= \frac{n_1}{2\sqrt{p}} \|VV^T - \hat{V}\hat{V}^T\|_F^2 \|V^T \bar{X}_1\|^2 \\ &= O_P(1) \frac{n_1}{\sqrt{p}} \frac{p}{p^\beta n_1} \frac{p^\beta}{n_1} \xrightarrow{P} 0. \end{aligned} \tag{14}$$

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

$$\begin{aligned}
E(Q_2^2) &= \frac{n_1^2}{p} E(\bar{X}_1^T \tilde{V} \tilde{V}^T \hat{V} \hat{V}^T V V^T \bar{X}_1 \bar{X}_1^T V V^T \hat{V} \hat{V}^T \tilde{V} \tilde{V}^T \bar{X}_1) \quad (15) \\
&= \frac{n_1^2}{p} E(\bar{X}_1^T \tilde{V} \tilde{V}^T \hat{V} \hat{V}^T V \frac{1}{n_1} (\Lambda + \sigma^2 I_r) V^T \hat{V} \hat{V}^T \tilde{V} \tilde{V}^T \bar{X}_1) \\
&= O(1) \frac{n_1 p^\beta}{p} E(\bar{X}_1^T \tilde{V} \tilde{V}^T \hat{V} \hat{V}^T V V^T \hat{V} \hat{V}^T \tilde{V} \tilde{V}^T \bar{X}_1) \\
&= O(1) \frac{n_1 p^\beta}{p} E \operatorname{tr}(\tilde{V}^T \hat{V} \hat{V}^T V V^T \hat{V} \hat{V}^T \tilde{V} \tilde{V}^T \bar{X}_1 \bar{X}_1^T \tilde{V}) \\
&= O(1) \frac{n_1 p^\beta}{p} \frac{\sigma^2}{n_1} E \operatorname{tr}(\tilde{V}^T \hat{V} \hat{V}^T V V^T \hat{V} \hat{V}^T \tilde{V}) \\
&\leq O(1) \frac{p^\beta}{p} E \operatorname{tr}(\tilde{V}^T \hat{V} \hat{V}^T I_p \hat{V} \hat{V}^T \tilde{V}) \\
&= O(1) \frac{p^\beta}{p} E \operatorname{tr}(\tilde{V}^T \hat{V} \hat{V}^T \tilde{V}) \\
&= O(1) \frac{p^\beta}{p} E \operatorname{tr}(\hat{V} \hat{V}^T (I_p - V V^T)) \\
&= O(1) \frac{p^\beta}{p} \frac{1}{2} E \|V V^T - \hat{V} \hat{V}^T\|_F^2 \\
&= O\left(\frac{1}{n_1}\right) \rightarrow 0.
\end{aligned}$$

Similarly,

$$\begin{aligned}
E(Q_3) &= \frac{n_1}{\sqrt{p}} \frac{\sigma^2}{n_1} E \operatorname{tr}(\tilde{V}^T \hat{V} \hat{V}^T \tilde{V}) \quad (16) \\
&\leq \frac{r \sigma^2}{\sqrt{p}} \rightarrow 0.
\end{aligned}$$

By (14), (15) and (16),  $R_1 \xrightarrow{P} 0$ . Next we deal with  $R_2$ . Write  $R_2 = W_1 - 2W_2 - W_3$ , where

$$\begin{aligned}
W_1 &= \frac{1}{n_1 \sqrt{p}} \sum_{i=1}^{n_1} X_{1i}^T V (I_r - V^T \hat{V} \hat{V}^T V) V^T X_{1i}, \\
W_2 &= \frac{1}{n_1 \sqrt{p}} \sum_{i=1}^{n_1} X_{1i}^T \tilde{V} \tilde{V}^T \hat{V} \hat{V}^T V V^T X_{1i}, \\
W_3 &= \frac{1}{n_1 \sqrt{p}} \sum_{i=1}^{n_1} X_{1i}^T \tilde{V} \tilde{V}^T \hat{V} \hat{V}^T \tilde{V} \tilde{V}^T X_{1i}.
\end{aligned}$$

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

$$W_1 = \frac{1}{n_1 \sqrt{p}} \text{tr}((I_r - V^T \hat{V} \hat{V}^T V) \sum_{i=1}^{n_1} V^T X_{1i} X_{1i}^T V). \quad (17)$$

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 (17) yields

$$\begin{aligned} W_1 &\leq O_P(1) \frac{p^\beta}{\sqrt{p}} \text{tr}(I_r - V^T \hat{V} \hat{V}^T V) \\ &= O_P(1) \frac{p^\beta}{\sqrt{p}} \|V V^T - \hat{V} \hat{V}^T\|_F^2 \\ &= O_P(1) \frac{p^\beta}{\sqrt{p}} \frac{p}{p^\beta n_1} \\ &= O_P\left(\frac{\sqrt{p}}{n_1}\right) \xrightarrow{P} 0. \end{aligned} \quad (18)$$

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

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

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

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

and

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

$Z^T \tilde{V} \tilde{V}^T Z$  is distributed as  $\text{Wishart}_n(p - r, \sigma^2 I_{n_1})$ , hence  $\lambda_1(Z^T \tilde{V} \tilde{V}^T Z) = O_P(\max(n_1, p))$  by Lemma 3. Again by law of large numbers  $\text{tr}(V Z Z^T V^T) =$



$O(p^\beta n_1)$ . Combining the argument above yields

$$\begin{aligned} W_2 &= O_P(1) \frac{1}{n_1 \sqrt{p}} \sqrt{\frac{p}{p^\beta n_1}} \sqrt{\max(n_1, p) p^\beta n_1} \\ &= O_P\left(\frac{\sqrt{\max(n_1, p)}}{n_1}\right) \xrightarrow{P} 0. \end{aligned} \quad (19)$$

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  $\text{Wishart}_{p-r}(n, \sigma^2 I_{p-r})$ . By Lemma 3,  $\lambda_1(\tilde{V}^T \sum_{i=1}^{n_1} X_{1i} X_{1i}^T \tilde{V}) = O_P(\max(p, n_1))$ . Hence

$$\begin{aligned} W_3 &= \frac{1}{n_1 \sqrt{p}} \text{tr}(\hat{V}^T \tilde{V} \tilde{V}^T \sum_{i=1}^{n_1} X_{1i} X_{1i}^T \tilde{V} \tilde{V}^T \hat{V}) \\ &\leq \frac{O_P(\max(n_1, p))}{n_1 \sqrt{p}} \text{tr} \hat{V}^T \tilde{V} \tilde{V}^T \hat{V} \xrightarrow{P} 0, \end{aligned} \quad (20)$$

due to  $\text{tr} \hat{V}^T \tilde{V} \tilde{V}^T \hat{V} = O_P(1)$ . Combining (18), (19) and (20) 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

$$\begin{aligned} M_1 &= \frac{n_1}{\sqrt{p}} \bar{X}_1^T V (I_r - V^T \hat{V} \hat{V}^T V) V^T \bar{X}_2, \\ M_2 &= \frac{n_1}{\sqrt{p}} \bar{X}_1^T \tilde{V} \tilde{V}^T \hat{V} \hat{V}^T V V^T \bar{X}_2, \\ M_3 &= \frac{n_1}{\sqrt{p}} \bar{X}_1^T V V^T \hat{V} \hat{V}^T \tilde{V} \tilde{V}^T \bar{X}_2, \\ M_4 &= \frac{n_1}{\sqrt{p}} \bar{X}_1^T \tilde{V} \tilde{V}^T \hat{V} \hat{V}^T \tilde{V} \tilde{V}^T \bar{X}_2. \end{aligned}$$

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

$$\begin{aligned} |M_1| &\leq \frac{n_1}{\sqrt{p}} \|I_r - V^T \hat{V} \hat{V}^T V\| \|V^T \bar{X}_1\| \|V^T \bar{X}_2\| \\ &\leq \frac{n_1}{2\sqrt{p}} \|V V^T - \hat{V} \hat{V}^T\|_F^2 \|V^T \bar{X}_1\| \|V^T \bar{X}_2\| \\ &= O_P(1) \frac{n_1}{\sqrt{p}} \frac{p}{p^\beta n_1} \frac{p^\beta}{n_1} \xrightarrow{P} 0. \end{aligned} \quad (21)$$

The joint distribution of  $(V^T \bar{X}_2, \tilde{V}^T \bar{X}_1, S)$  is identity to that of  $(V^T \bar{X}_1, \tilde{V}^T \bar{X}_2, S)$ .

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.  $\square$

**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

$$\frac{n_1 n_2 (T_2^{(1)} - T_1^{(1)})}{\sqrt{2p}(n_1 + n_2)\sigma^2} \xrightarrow{P} 0. \quad (22)$$

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

$$\frac{n_1 n_2 (T_2^{(2)} - T_1^{(2)})}{\sqrt{2p}(n_1 + n_2)\sigma^2} = O(1) \frac{n_1}{\sqrt{p}} (\bar{X}_1 - \mu_1)^T (V V^T - \hat{V} \hat{V}^T) (\mu_1 - \mu_2) + \quad (23)$$

$$O(1) \frac{n_2}{\sqrt{p}} (\bar{X}_2 - \mu_2)^T (V V^T - \hat{V} \hat{V}^T) (\mu_2 - \mu_1) + \quad (24)$$

$$O(1) \frac{n_1 + n_2}{\sqrt{p}} (\mu_1 - \mu_2)^T (V V^T - \hat{V} \hat{V}^T) (\mu_1 - \mu_2). \quad (25)$$

We proceed to deal with the (23).

$$\begin{aligned}
& E \left| \frac{n_1}{\sqrt{p}} (\bar{X}_1 - \mu_1)^T (VV^T - \hat{V}\hat{V}^T)(\mu_1 - \mu_2) \right|^2 \\
&= \frac{n_1^2}{p} E(\mu_1 - \mu_2)^T (VV^T - \hat{V}\hat{V}^T) \frac{\Sigma}{n_1} (VV^T - \hat{V}\hat{V}^T)(\mu_1 - \mu_2) \\
&= O(n_1 p^{\beta-1}) E(\mu_1 - \mu_2)^T (VV^T - \hat{V}\hat{V}^T)^2 (\mu_1 - \mu_2) \\
&\leq O(n_1 p^{\beta-1}) \|\mu_1 - \mu_2\|^2 E \|VV^T - \hat{V}\hat{V}^T\|_F^2 \\
&= O(n_1 p^{\beta-1}) \frac{p}{p^\beta n_1} \|\mu_1 - \mu_2\|^2 \\
&= O(\|\mu_1 - \mu_2\|^2).
\end{aligned} \tag{26}$$

Combining with conditions  $\frac{(n_1+n_2)}{\sqrt{p}} \|\mu_1 - \mu_2\|^2 = O(1)$  and  $\frac{\sqrt{p}}{n_1+n_2} \rightarrow 0$ , it follows that (26) converges to 0 in probability. By symmetry, (24) also converges to 0 in probability. It remains to deal with (25). But

$$\begin{aligned}
& E \left| \frac{n_1 + n_2}{\sqrt{p}} (\mu_1 - \mu_2)^T (VV^T - \hat{V}\hat{V}^T)(\mu_1 - \mu_2) \right| \\
&\leq \frac{n_1 + n_2}{\sqrt{p}} \|\mu_1 - \mu_2\|^2 E \|VV^T - \hat{V}\hat{V}^T\| \\
&\leq \frac{n_1 + n_2}{\sqrt{p}} \|\mu_1 - \mu_2\|^2 \sqrt{E \|VV^T - \hat{V}\hat{V}^T\|^2} \\
&\leq O(1) \sqrt{E \|VV^T - \hat{V}\hat{V}^T\|_F^2}.
\end{aligned} \tag{27}$$

Since  $E \|VV^T - \hat{V}\hat{V}^T\|_F^2 = O(\frac{p}{p^\beta n_1})$  and  $\beta \geq 1/2$ , (27) 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 (22), the theorem follows.  $\square$

**Proof Of Theorem 4.** 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  $\text{rank}(V) \leq r_1 + r_2$  and  $V_{2\ominus 1}V_1 = O_{r_2 \times r_1}$ , 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

$$\begin{aligned} Q'_1 &= \frac{n_1}{\sqrt{p}}\bar{X}_1^T V_1(I_{r_1} - V_1^T \hat{V}\hat{V}^T V_1)V_1^T \bar{X}_1, \\ Q'_2 &= \frac{n_1}{\sqrt{p}}\bar{X}_1^T \tilde{V}_1\tilde{V}_1^T \hat{V}\hat{V}^T V_1V_1^T \bar{X}_1, \\ Q'_3 &= \frac{n_1}{\sqrt{p}}\bar{X}_1^T \tilde{V}_1\tilde{V}_1^T \hat{V}\hat{V}^T \tilde{V}_1\tilde{V}_1^T \bar{X}_1. \end{aligned}$$

We note that  $Q'_1 \leq n_1\bar{X}_1^T V_1(I_{r_1} - V_1^T \hat{V}\hat{V}^T V_1)V_1^T \bar{X}_1/\sqrt{p}$  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

$$\begin{aligned} &\frac{n_1}{\sqrt{p}}\bar{X}_1^T \tilde{V}_1\tilde{V}_1^T \hat{V}_{2\ominus 1}\hat{V}_{2\ominus 1}^T V_1V_1^T \bar{X}_1 \\ &\leq \sqrt{\frac{n_1}{\sqrt{p}}\bar{X}_1^T \tilde{V}_1\tilde{V}_1^T \hat{V}_{2\ominus 1}\hat{V}_{2\ominus 1}^T \tilde{V}_1\tilde{V}_1^T \bar{X}_1} \sqrt{\frac{n_1}{\sqrt{p}}\bar{X}_1^T V_1V_1^T \hat{V}_{2\ominus 1}\hat{V}_{2\ominus 1}^T V_1V_1^T \bar{X}_1}. \end{aligned}$$

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

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

$Q'_3 = \frac{n_1}{\sqrt{p}} \bar{X}_1^T \tilde{V}_1 \tilde{V}_1^T \hat{V}_{2\ominus 1} \hat{V}_{2\ominus 1}^T \tilde{V}_1 \tilde{V}_1^T \bar{X}_1 + \frac{n_1}{\sqrt{p}} \bar{X}_1^T \tilde{V}_1 \tilde{V}_1^T \hat{V}_1 \hat{V}_1^T \tilde{V}_1 \tilde{V}_1^T \bar{X}_1$  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  $\frac{1}{n_1 \sqrt{p}} \sum_{i=1}^{n_1} X_{1i}^T V V^T X_{1i} - \frac{1}{n_1 \sqrt{p}} \sum_{i=1}^{n_1} X_{1i}^T V_1 V_1^T X_{1i} \xrightarrow{P} 0$ . Hence it suffices to consider the following three parts

$$W'_1 = \frac{1}{n_1 \sqrt{p}} \sum_{i=1}^{n_1} X_{1i}^T V_1 (I_{r_1} - V_1^T \hat{V} \hat{V}^T V_1) V_1^T X_{1i},$$

$$W'_2 = \frac{1}{n_1 \sqrt{p}} \sum_{i=1}^{n_1} X_{1i}^T \tilde{V}_1 \tilde{V}_1^T \hat{V} \hat{V}^T V_1 V_1^T X_{1i}$$

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.

$$W'_2 = \frac{1}{n_1 \sqrt{p}} \sum_{i=1}^{n_1} X_{1i}^T \tilde{V}_1 \tilde{V}_1^T \hat{V}_1 \hat{V}_1^T V_1 V_1^T X_{1i} + \frac{1}{n_1 \sqrt{p}} \sum_{i=1}^{n_1} X_{1i}^T \tilde{V}_1 \tilde{V}_1^T \hat{V}_{2\ominus 1} \hat{V}_{2\ominus 1}^T V_1 V_1^T X_{1i}.$$

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  $\hat{V}_{2\ominus 1} \hat{V}_{2\ominus 1}^T \leq \hat{\tilde{V}} \hat{\tilde{V}}^T$ . 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

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

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

$$M'_3 = \frac{n_1}{\sqrt{p}} \bar{X}_1^T V V^T \hat{V} \hat{V}^T \tilde{V} \tilde{V}^T \bar{X}_2,$$

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

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

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

0. We note that

$$\begin{aligned}
& \frac{n_1}{\sqrt{p}} \bar{X}_1^T V (I_{r_1+r_2} - V^T \hat{V} \hat{V}^T V) V^T \bar{X}_1 \\
&= \left( \frac{n_1}{\sqrt{p}} \bar{X}_1^T V_1 V_1^T \bar{X}_1 - \frac{n_1}{\sqrt{p}} \bar{X}_1^T V_1 V_1^T \hat{V} \hat{V}^T V_1 V_1^T \bar{X}_1 \right) \\
&+ \frac{n_1}{\sqrt{p}} \bar{X}_1^T V_{2\ominus 1} V_{2\ominus 1}^T \bar{X}_1 - 2 \frac{n_1}{\sqrt{p}} \bar{X}_1^T V_{2\ominus 1} V_{2\ominus 1}^T \hat{V} \hat{V}^T V_1 V_1^T \bar{X}_1 \\
&- \frac{n_1}{\sqrt{p}} \bar{X}_1^T V_{2\ominus 1} V_{2\ominus 1}^T \hat{V} \hat{V}^T V_{2\ominus 1} V_{2\ominus 1}^T \bar{X}_1 \\
&= L_1 + L_2 - 2L_3 - L_4.
\end{aligned}$$

Check that  $L_1 = Q'_1$ . It's also clear that  $L_2 \xrightarrow{P} 0$ . To deal with  $L_3$ , we note that

$$\begin{aligned}
& \left| \frac{n_1}{\sqrt{p}} \bar{X}_1^T V_{2\ominus 1} V_{2\ominus 1}^T \hat{V}_1 \hat{V}_1^T V_1 V_1^T \bar{X}_1 \right| \\
&\leq \frac{n_1}{\sqrt{p}} \|V_{2\ominus 1}^T \bar{X}_1\| \|V_{2\ominus 1}^T \hat{V}_1\|_F \|\hat{V}_1^T V_1 V_1^T \bar{X}_1\| \\
&\leq \frac{n_1}{\sqrt{p}} O\left(\frac{1}{\sqrt{n_1}}\right) \sqrt{\text{tr}(\hat{V}_1^T V_{2\ominus 1} V_{2\ominus 1}^T \hat{V}_1)} O\left(\frac{\sqrt{p^\beta}}{\sqrt{n_1}}\right).
\end{aligned} \tag{28}$$

But

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

And

$$\begin{aligned}
\left| \frac{n_1}{\sqrt{p}} \bar{X}_1^T V_{2\ominus 1} V_{2\ominus 1}^T \hat{V}_{2\ominus 1} \hat{V}_{2\ominus 1}^T V_1 V_1^T \bar{X}_1 \right| &\leq \frac{n_1}{\sqrt{p}} \|V_{2\ominus 1}^T \hat{V}_{2\ominus 1} V_{2\ominus 1}^T \bar{X}_1\| \|\hat{V}_{2\ominus 1}^T V_1\|_F \|V_1^T \bar{X}_1\| \\
&\leq \frac{n_1}{\sqrt{p}} O\left(\frac{1}{\sqrt{n_1}}\right) \sqrt{\text{tr}(\hat{V}_{2\ominus 1}^T V_1 V_1^T \hat{V}_{2\ominus 1})} O\left(\frac{\sqrt{p^\beta}}{\sqrt{n_1}}\right).
\end{aligned} \tag{29}$$

But

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

Therefore  $L_3 \xrightarrow{P} 0$ . And  $L_4 \xrightarrow{P} 0$  for trival reason.  $M'_2$  can be similarly treated by technique (28) and (29). 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 5.*** 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.
- Bai ZD, Yin Y. Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Annals of Probability* 1993;21(3):1275–94.
- 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.

- Jiang J. Reml estimation: asymptotic behavior and related topics. *Annals of Statistics* 1996;24(1):255–86.
- 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.