

# 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  $n = n_1 + n_2$  is the total sample size. 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. This 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. To characterize strong correlations, Ma et al. (2015) considered a factor model and proposed a 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  $\kappa > 0$  and  $\beta \geq \frac{1}{2}$  such that

$$\kappa p^\beta \geq \lambda_{k1} \geq \dots \geq \lambda_{kr_k} \geq \kappa^{-1} p^\beta.$$

The restriction  $\beta \geq 1/2$  breaks down the Condition (2). 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.

Throughout the paper, let  $\tau = (n_1 + n_2)/(n_1 n_2)$ ,  $S$  be the pooled sample covariance:

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

where

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

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

We write  $\xi \sim \eta$  to denote the random variable  $\xi$  and  $\eta$  have the same distribution. For nonrandom positive sequence  $\{a_n\}$  and  $\{b_n\}$ ,  $a_n \asymp b_n$  represents  $a_n = O(b_n)$  and  $b_n = O(a_n)$  as  $n \rightarrow \infty$ .

### 3. Methodology

In this section, we describe our new test procedure for hypotheses (1). For simplicity, we now work on equal covariance setting and unequal covariance setting will be considered latter.

**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 Chen and Qin (2010), to prove the asymptotic normality of  $T_{CQ}$ , they assumed condition (2). The following proposition tells that (2) is also necessary for the normality of  $T_{CQ}$ .

**Proposition 1.** Suppose  $X_{ki} \sim N_p(0, \Sigma)$ ,  $k = 1, 2$ ,  $i = 1, 2, \dots, n_k$ . Suppose Assumption 1 holds. Then (2) is a necessary and sufficient condition for

$$\frac{T_{CQ} - \mathbb{E}T_{CQ}}{[\text{Var}(T_{CQ})]^{1/2}} \xrightarrow{L} N(0, 1). \quad (3)$$

The Proposition 1 implies that under spiked covariance, Chen and Qin (2010)'s method can not guarantee the test level. To this end, Ma et al. (2015) proposed a test procedure which is based on  $T_{CQ}$  and has the correct asymptotic test level. In their paper,  $\Sigma$ 's first few eigenvalues are assumed to be of order  $p$ .

Note that  $T_{BS}$ ,  $T_{CQ}$  and Ma et al. (2015)'s method are all based on  $\tau\|\bar{X}_1 - \bar{X}_2\|^2$ , which can be written as the sum of two parts

$$\tau\|V^T(\bar{X}_1 - \bar{X}_2)\|^2 + \tau\|\tilde{V}^T(\bar{X}_1 - \bar{X}_2)\|^2. \quad (4)$$

Under the null hypotheses, we have

$$\text{Var}(\tau\|V^T(\bar{X}_1 - \bar{X}_2)\|^2) = \sum_{i=1}^r 2(\lambda_i + \sigma^2)^2, \quad \text{Var}(\|\tilde{V}^T(\bar{X}_1 - \bar{X}_2)\|^2) = 2\sigma^4(p-r).$$

The ratio of the two variance is

$$\frac{\sum_{i=1}^r 2(\lambda_i + \sigma^2)^2}{2\sigma^4(p-r)} \asymp p^{2\beta-1},$$

which tends to  $\infty$  as  $p \rightarrow \infty$  for  $\beta > 1/2$ . On the other hand,  $\tau\|V^T(\bar{X}_1 - \bar{X}_2)\|^2$  only involves the signal from  $r$  dimension. Thus, compared with the second

term of (4), the first term has larger variance and tends to contain much weaker signal. This motivates us to drop the first part of (4) and only use the second part. After adjustment of expectation, we define the following statistic

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

Proposition 2 shows that the asymptotic distribution of  $T_1$  is normal.

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

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

In another point of view,  $T_1$  is obtained by transforming  $X_{ki}$  to  $\tilde{V}^T X_{ki}$  ( $i = 1, \dots, n_k, k = 1, 2$ ) and then invoking the statistic of Chen and Qin (2010). Several test procedures have been proposed through random projection to lower dimensional space, for example, Lopes et al. (2011), Thulin (2014) and Srivastava et al. (2015). For any  $p \times (p-r)$  column orthonormal matrix  $O_{p \times (p-r)}$ , we have

$$\text{Var}(\|O_{p \times (p-r)}^T(\bar{X}_1 - \bar{X}_2)\|^2) \geq \text{Var}(\|\tilde{V}^T(\bar{X}_1 - \bar{X}_2)\|^2).$$

Hence  $\tilde{V}$  is the optimal transformation in the sense of reducing variance. Based on  $\tilde{V}^T X_{ki}$ , the likelihood ratio test statistic for hypothesis (1) is  $\|\tilde{V}^T(\bar{X}_1 - \bar{X}_2)\|^2$  which coincides with our proposal. In this view,  $T_1$  can be regarded as a restricted likelihood ratio test.

### 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 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  $\|\hat{\hat{V}}^T(\bar{X}_1 -$

$\bar{X}_2)\|^2$ . 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

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

The asymptotic result of Proposition 2 involves  $\sigma^2$ . In order to formulate a test procedure by asymptotic distribution,  $\sigma^2$  needs to be consistently estimated. Note that  $\sigma^2$  can be written as

$$\sigma^2 = \sum_{i=r+1}^p \lambda_i(\Sigma). \quad (5)$$

It can be estimated by

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

Now we propose our new test statistic as

$$Q = \frac{T_2}{\hat{\sigma}^2 \sqrt{2\tau^2 p}}. \quad (6)$$

In next section, it will be proved that the asymptotic distribution of  $Q$  is  $N(0, 1)$  under null hypotheses. Thus, we reject the null hypothesis when  $Q$  is larger than the upper  $\alpha$  quantile of  $N(0, 1)$ .

**Remark 1.** 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 2.** 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 3.** If  $r$  is an unknown positive number, a consistent estimator of  $r$  is

$$\hat{r} = \text{argmax}_{l \leq R} \frac{\lambda_l(S)}{\lambda_{l+1}(S)}, \quad (7)$$

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

$$\frac{\|\tilde{V}^T(\mu_1 - \mu_2)\|}{\|\mu_1 - \mu_2\|} \approx 1. \quad (8)$$

Unfortunately, (8) is not always the case since there always exists some  $\tilde{V}$  and  $\mu_1 - \mu_2$  such that  $\|\tilde{V}^T(\mu_1 - \mu_2)\| = 0$ . However, (8) 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 for  $\mu_k$  will be used in Theorem 3.

#### 4. Theoretical results

In this section, we study the asymptotic behavior of the new test procedure.

We first give a result of the convergence rate of  $\hat{\sigma}^2$ . In particular, it can be seen that  $\hat{\sigma}^2$  is a consistent estimator of  $\sigma^2$ . Our proof relies on the Weyl's inequality.

**Proposition 3.** *Under Assumptions 1-4, we have that*

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

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

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

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

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

*then*

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

**Remark 4.** The Assumption 5 is a strong condition. However, it may not be able to be relaxed. In fact, the asymptotic normality of  $\|\hat{\tilde{V}}^T(\bar{X}_1 - \bar{X}_2)\|^2$ , the major part of  $T_2$ , requires

$$\frac{\lambda_1((\hat{\tilde{V}}^T \Sigma \hat{\tilde{V}})^2)}{\text{tr}((\hat{\tilde{V}}^T \Sigma \hat{\tilde{V}})^2)} \xrightarrow{P} 0. \quad (9)$$

See Lemma 4 in appendix. And (9) is equivalent to Assumption 5 by Lemma 2 in appendix.

By Proposition 3 and Theorem 1, the power function of the 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 the new test is*

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

Note that 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.

When Assumption 5 doesn't met, the asymptotic normality are not valid. The next theorem gives an asymptotic result in this case.

**Theorem 2.** *Under Assumptions 1-4 and  $p/n^2 \rightarrow \infty$ , if the local alternative holds, that is,*

$$\|\mu_1 - \mu_2\|^2 \dots$$

then

$$\frac{T_2 - \|\tilde{V}^T(\mu_1 - \mu_2)\|^2}{\dots}$$

However, this does not mean the new statistic can not be used. In fact, since the samples are exchangeable under null hypothesis, we can always use permutation method to determine the critical value. We will see from simulation results that the new test has good power behavior even in large  $p$  small  $n$  case.

In practice, it may not be an easy task to check if the covariance matrices are spiked, especially in high dimension setting. When the spiked covariance model is not valid, some estimators in our test procedure make no sense. In particular, if  $\hat{r}$  is estimated by (7), the  $\hat{r}$  is nothing but a random integer not greater than  $R$  and  $\hat{V}\hat{V}^T$  is just a random projection. Hence it is a natural question how the new test procedure behaves when the spiked covariance model breaks down. We study the asymptotic behavior of the new test procedure in two non-spiked setting.

First we consider the case when the eigenvalues of  $\Sigma$  is bounded. Similar to bayesian models, we assume a normal prior distribution for  $\mu_k$  to characterize ‘dense’ alternative. 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 3.** *Assume  $X_{ki} \sim N(\mu_k, \Sigma)$ ,  $i = 1, \dots, n_k$ ,  $k = 1, 2$ . Suppose that Assumptions 1 and 5 holds,  $0 < c \leq \lambda_p(\Sigma) \leq \lambda_1(\Sigma) \leq C < \infty$  where  $c$  and  $C$  are constants, each element of  $\mu_k$  is independently generated by  $N(0, (n_k \sqrt{p})^{-1} \psi)$  for  $k = 1, 2$ , where  $\psi$  is a constant and  $\hat{r} \leq R$  for a positive constant  $R$ . Then we have*

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

The second setting we consider is the model in Assumption 2 with  $r = 0$ . In this case, the Assumption 5 can be dropped and we don’t need to assume a random  $\mu_k$ .

**Theorem 4.** *Under Assumptions 1-4 with factor number  $r = 0$ , if*

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

and  $\hat{r} \leq R$  for a positive constant  $R$ , then

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

These results show that the new test procedure is robust against the invalidity of spiked covariance model.

## 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

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

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

**Theorem 5.** *Under Assumptions 1-3 and 5, if*

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

then we have

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

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. However, the centering term should be  $\|\hat{\tilde{V}}^T(\mu_1 - \mu_2)\|^2$  and can not be replaced by  $\|\tilde{V}^T(\mu_1 - \mu_2)\|^2$ .

$\sigma_n^2$  can be estimated by ratio consistent estimators 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.

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.

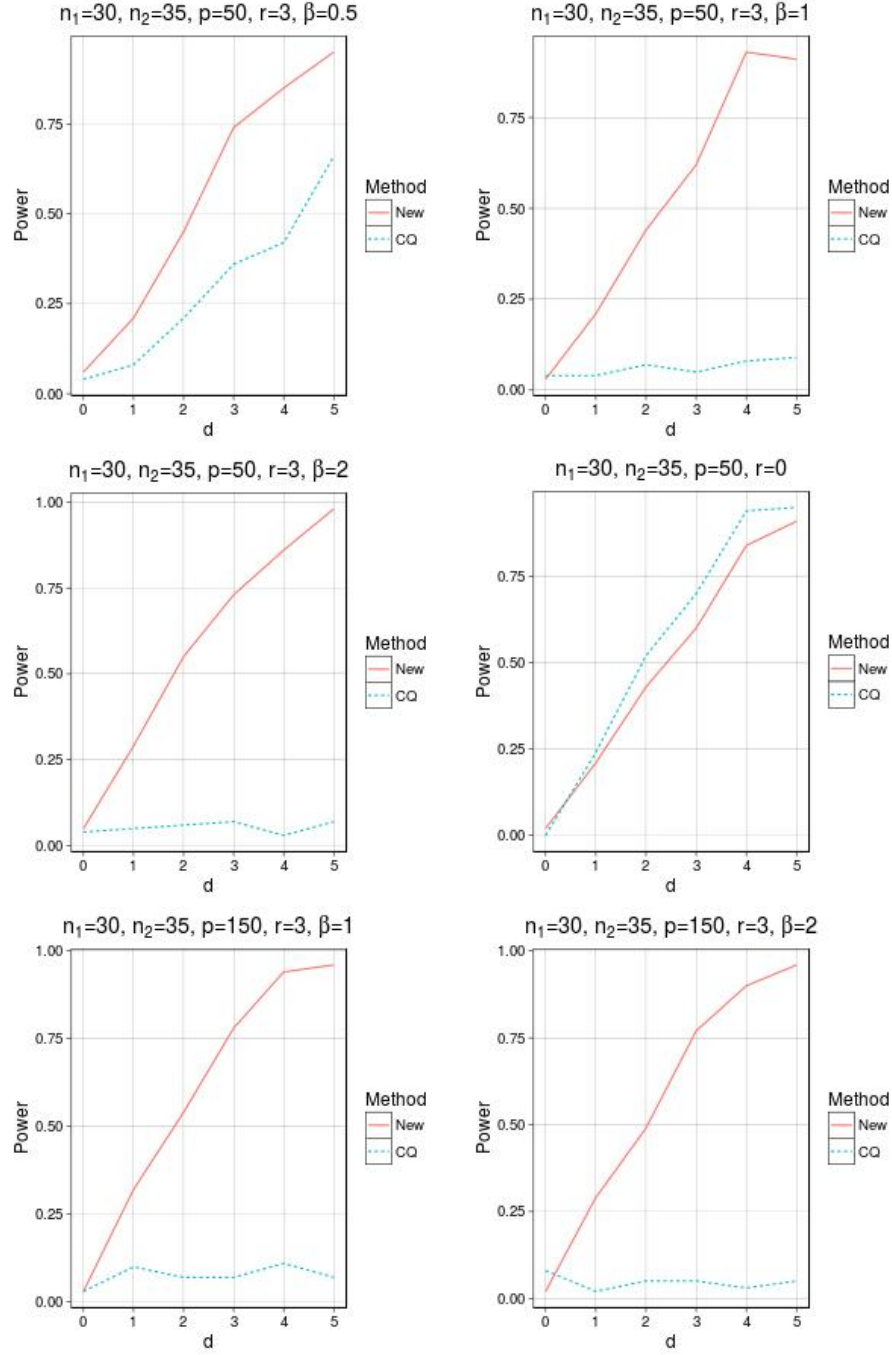


Figure 1: Empirical power simulation.  $\alpha$  is set to be 0.05.  $d$  is proportional to  $\|\mu_1 - \mu_2\|^2$ . For each simulation, we do 50 permutations to determine critical value. We generate 100 independent samples to compute empirical power.

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

### 6.2. Real data analysis

In this section, we study the practical problem considered in Ma et al. (2015). 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 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

$$E\|\hat{V}\hat{V}^T - VV^T\|_F^2 \asymp r \wedge (p - r) \wedge \frac{r(p - r)}{(n_1 + n_2 - 2)h(\lambda)} \quad (10)$$

as long as the right hand side tends to 0. Here  $h(\lambda) = \frac{\lambda^2}{\lambda+1}$ . In model of Assumption 2,  $r$  is fixed,  $\lambda = cp^\beta$ . It's obvious that the right hand side of (10) 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.  $\square$

**Lemma 4.** *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*

$$\frac{X_n^T A_n X_n - \mathbb{E} X_n^T A_n X_n}{[\text{Var}(X_n^T A_n X_n)]^{1/2}} \xrightarrow{\mathcal{L}} N(0, 1) \quad (11)$$

*is that*

$$\frac{\lambda_{\max}(A_n^2)}{\text{tr}(A_n^2)} \rightarrow 0. \quad (12)$$

**Remark 7.** This lemma is from the Example 5.1 of Jiang (1996). Here we give a proof by characteristic function.

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

$$\frac{X_n^T A_n X_n - \mathbb{E} X_n^T A_n X_n}{[\text{Var}(X_n^T A_n X_n)]^{1/2}} = \sum_{i=1}^{k_n} \frac{\lambda_i(A_n)}{[2\text{tr}(A_n^2)]^{1/2}} (Z_{ni}^2 - 1), \quad (13)$$

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

If 12 holds, then

$$\begin{aligned} & \sum_{i=1}^{k_n} \mathbb{E} \left[ \frac{\lambda_i^2(A_n)}{2\text{tr}(A_n^2)} (Z_{ni}^2 - 1)^2 \left\{ \frac{\lambda_i^2(A_n)}{2\text{tr}(A_n^2)} (Z_{ni}^2 - 1)^2 \geq \epsilon \right\} \right] \\ & \leq \sum_{i=1}^{k_n} \frac{\lambda_i^2(A_n)}{2\text{tr}(A_n^2)} \mathbb{E} \left[ (Z_{n1}^2 - 1)^2 \left\{ \frac{\lambda_{\max}(A_n^2)}{2\text{tr}(A_n^2)} (Z_{n1}^2 - 1)^2 \geq \epsilon \right\} \right] \\ & = \frac{1}{2} \mathbb{E} \left[ (Z_{n1}^2 - 1)^2 \left\{ \frac{\lambda_{\max}(A_n^2)}{2\text{tr}(A_n^2)} (Z_{n1}^2 - 1)^2 \geq \epsilon \right\} \right] \rightarrow 0. \end{aligned}$$

Hence 11 follows by Lindeberg's central limit theorem.

Conversely, if 11 holds, we will prove that there is a subsequence of  $\{n\}$  along which 12 holds. Then 12 will hold by a standard contradiction argument.

Denote  $c_{ni} = \lambda_i(A_n)/[2\text{tr}(A_n^2)]^{1/2}$  ( $i = 1, \dots, k_n$ ), we have  $c_{ni} \in [-\sqrt{2}/2, \sqrt{2}/2]$ . Since 11 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

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

Denote  $b_{nl} = \sum_{j=1}^{k_n} (c_{nj})^l$ ,  $n = 1, 2, \dots$  and  $l = 3, 4, \dots$ . For  $l \geq 3$ ,  $|\sum_{j=1}^{k_n} (c_{nj})^l| \leq |\sum_{j=1}^{k_n} (c_{nj})^2| = 1/2$ . By Helly's selection theorem, there's a subsequence of  $\{n\}$  along which  $\lim_{n \rightarrow \infty} b_{nl} = b_l$  exists for every  $l$ . Apply dominated convergence theorem to this subsequence we have  $\log \mathbb{E} \exp \left( it \sum_{j=1}^{k_n} c_{nj}(Z_{nj}^2 - 1) \right) \rightarrow -\frac{1}{2}t^2 + \frac{1}{2} \sum_{l=3}^{+\infty} b_l \frac{1}{l} (i2t)^l$  for  $t \in (-1/2, 1/2)$ . By the property of power series, we have  $b_l = 0$  for  $l \geq 3$ . Then 12 follows by noting that  $b_{n4} \geq \max_j (c_{nj})^4$ .  $\square$

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

**Proof Of Proposition 1.** We first note that (2) is equivalent to

$$\frac{\lambda_{\max}(\Sigma)}{(\text{tr}\Sigma^2)^{1/2}} \rightarrow 0. \quad (14)$$

Let  $X_{ki} = \Sigma^{1/2} Z_{ki}$  with  $Z_{ki}$  distributed as  $N_p(0, I_p)$ . Denote  $Z = (Z_{11}, \dots, Z_{1n_1}, Z_{21}, \dots, Z_{2n_2})^T$ . Then  $T_{CQ}$  is a quadratic form of  $Z$  and we can write

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

where

$$B_n = \begin{pmatrix} \frac{1}{n_1(n_1-1)}(n_1\gamma_1\gamma_1^T - I_{n_1}) & -\frac{1}{\sqrt{n_1n_2}}\gamma_1\gamma_2^T \\ -\frac{1}{\sqrt{n_1n_2}}\gamma_2\gamma_1^T & \frac{1}{n_2(n_2-1)}(n_2\gamma_2\gamma_2^T - I_{n_2}) \end{pmatrix},$$

$\gamma_1$  is an  $n_1$  dimensional vector with all elements equal to  $1/\sqrt{n_1}$  and  $\gamma_2$  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  $-1/n_1(n_1-1)$ ,  $-1/n_2(n_2-1)$ ,  $(n_1+n_2)/n_1n_2$  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

$$\{(\eta^T, \underbrace{0, \dots, 0}_{n_2})^T \mid \eta \in \mathbb{R}^{n_1} \text{ and } \eta^T \gamma_1 = 0\}.$$

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

$$\{(\underbrace{0, \dots, 0}_{n_1}, \eta^T)^T \mid \eta \in \mathbb{R}^{n_2} \text{ and } \eta^T \gamma_2 = 0\}.$$

The eigenvector corresponding to  $(n_1+n_2)/n_1n_2$  is

$$\left(-\sqrt{\frac{n_2}{n_1+n_2}}\gamma_1^T, \sqrt{\frac{n_1}{n_1+n_2}}\gamma_2^T\right)^T.$$

The eigenvector corresponding to 0 is

$$\left(\sqrt{\frac{n_1}{n_1+n_2}}\gamma_1^T, \sqrt{\frac{n_2}{n_1+n_2}}\gamma_2^T\right)^T.$$

It follows that

$$\text{tr}(B_n \otimes \Sigma)^2 = \text{tr}(B_n^2) \text{tr} \Sigma^2 = \left(\frac{1}{n_1(n_1-1)} + \frac{1}{n_1(n_1-1)} + \frac{2}{n_1n_2}\right) \text{tr} \Sigma^2.$$

And

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

The theorem follows by Lemma 4.  $\square$

**Proof Of Proposition 2.** Since  $V$  and  $\tilde{V}$  are orthogonal, we have

$$\tilde{V}^T X_{ki} = \tilde{V}^T \mu_i + \tilde{V}^T Z_{ki} \sim N(\tilde{V}^T \mu_k, \sigma^2 I_{p-r}) \quad k = 1, 2 \text{ and } i = 1, \dots, n_k.$$

Let  $\bar{Z}_1$  and  $\bar{Z}_2$  be the sample mean of  $\{Z_{1i}\}$  and  $\{Z_{2i}\}$  respectively. Then

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

But

$$\begin{aligned} 2(\mu_1 - \mu_2)^T \tilde{V} \tilde{V}^T (\bar{Z}_1 - \bar{Z}_2) &\sim N(0, 4\sigma^2 \tau \|\tilde{V}^T (\mu_1 - \mu_2)\|^2) \\ &= O_P(\sqrt{\tau} \|\tilde{V}^T (\mu_1 - \mu_2)\|) = o_P\left(\frac{\sqrt{p}}{n}\right). \end{aligned}$$

Then

$$\|\tilde{V}^T (\bar{X}_1 - \bar{X}_2)\|^2 = \|\tilde{V}^T (\mu_1 - \mu_2)\|^2 + \|\tilde{V}^T (\bar{Z}_1 - \bar{Z}_2)\|^2 + o_P\left(\frac{\sqrt{p}}{n}\right). \quad (15)$$

Note that  $\frac{1}{n_i} \tilde{V}^T S_i \tilde{V} \sim \frac{\sigma^2}{n_i(n_i-1)} \text{Wishart}_{p-r}(n_i-1, I_{p-r})$ ,  $i = 1, 2$ . Then

$$\begin{aligned} \frac{1}{n_i} \text{tr}(\tilde{V}^T S_i \tilde{V}) &\sim \frac{\sigma^2}{n_i(n_i-1)} \chi_{(p-r)(n_i-1)}^2 \\ &= \sigma^2 \frac{p-r}{n_i} (1 + O_P(\frac{1}{\sqrt{(p-r)(n_i-1)}})), \end{aligned}$$

where the second line holds by central limit theorem. It follows that

$$\frac{1}{n_1} \text{tr}(\tilde{V}^T S_1 \tilde{V}) + \frac{1}{n_2} \text{tr}(\tilde{V}^T S_2 \tilde{V}) = \sigma^2 \tau (p-r) + o_P\left(\frac{\sqrt{p}}{n}\right). \quad (16)$$

By (15) and (16), we have

$$\frac{T_1 - \|\tilde{V}^T (\mu_1 - \mu_2)\|^2}{\sigma^2 \sqrt{2\tau^2 p}} = \frac{\|\tilde{V}^T (\bar{Z}_1 - \bar{Z}_2)\|^2 - \sigma^2 \tau (p-r)}{\sigma^2 \sqrt{2\tau^2 p}} + o_P(1). \quad (17)$$

Note that  $\|\tilde{V}^T (\bar{Z}_1 - \bar{Z}_2)\|^2 \sim \sigma^2 \tau \chi_{p-r}^2$ . The proposition follows by central limit theorem.  $\square$

**Proof Of Proposition 3.** 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.** Note that  $\text{tr}(\hat{\tilde{V}}_i^T S_i \hat{\tilde{V}}_i) = \sum_{i=r+1}^p \lambda_i(S_i)$ ,  $i = 1, 2$ . Similar to Proposition 3, we have that  $\text{tr}(\hat{\tilde{V}}_i^T S_i \hat{\tilde{V}}_i) = (p-r)\sigma^2 + O_P(\frac{\max(n,p)}{n})$ ,  $i = 1, 2$ . Hence

$$\frac{T_2 - \|\tilde{V}^T(\mu_1 - \mu_2)\|^2}{\sigma^2 \sqrt{2\tau^2 p}} = \frac{\|\hat{\tilde{V}}^T(\bar{X}_1 - \bar{X}_2)\|^2 - \|\tilde{V}^T(\mu_1 - \mu_2)\|^2 - \sigma^2 \tau(p-r)}{\sigma^2 \sqrt{2\tau^2 p}} + O_P(\frac{\max(n,p)}{n\sqrt{p}}).$$

By Assumption 5,  $\frac{\max(n,p)}{n\sqrt{p}} = \max(\frac{1}{\sqrt{p}}, \frac{\sqrt{p}}{n}) \rightarrow 0$ . And

$$\begin{aligned} & \frac{\|\hat{\tilde{V}}^T(\bar{X}_1 - \bar{X}_2)\|^2 - \|\tilde{V}^T(\mu_1 - \mu_2)\|^2 - \sigma^2 \tau(p-r)}{\sigma^2 \sqrt{2\tau^2 p}} \\ &= \frac{1}{\sigma^2 \sqrt{2\tau^2 p}} \left( \|\hat{\tilde{V}}^T((\bar{X}_1 - \mu_1) - (\bar{X}_2 - \mu_2))\|^2 - \sigma^2 \tau(p-r) + \right. \\ & \quad \left. 2(\mu_1 - \mu_2)^T \hat{\tilde{V}} \hat{\tilde{V}}^T ((\bar{X}_1 - \mu_1) - (\bar{X}_2 - \mu_2)) + \|\hat{\tilde{V}}^T(\mu_1 - \mu_2)\|^2 - \|\tilde{V}^T(\mu_1 - \mu_2)\|^2 \right). \end{aligned}$$

Let

$$\begin{aligned} P_1 &= \|\hat{\tilde{V}}^T((\bar{X}_1 - \mu_1) - (\bar{X}_2 - \mu_2))\|^2 - \sigma^2 \tau(p-r), \\ P_2 &= 2(\mu_1 - \mu_2)^T \hat{\tilde{V}} \hat{\tilde{V}}^T ((\bar{X}_1 - \mu_1) - (\bar{X}_2 - \mu_2)), \\ P_3 &= \|\hat{\tilde{V}}^T(\mu_1 - \mu_2)\|^2 - \|\tilde{V}^T(\mu_1 - \mu_2)\|^2. \end{aligned}$$

To prove the theorem, we only need to show that

$$\frac{P_1}{\sigma^2 \sqrt{2\tau^2 p}} \xrightarrow{\mathcal{L}} N(0, 1), \quad \frac{P_2}{\sigma^2 \sqrt{2\tau^2 p}} \xrightarrow{P} 0 \quad \text{and} \quad \frac{P_3}{\sigma^2 \sqrt{2\tau^2 p}} \xrightarrow{P} 0.$$

We first deal with  $P_2$ . To prove the convergence in probability, we only need to prove the convergence in  $L^2$ . Note that  $\bar{X}_1$ ,  $\bar{X}_2$  and  $S$  are mutually independent. And  $\hat{\tilde{V}} \hat{\tilde{V}}^T$  only depends on  $S$ , thus

$$\begin{aligned} \mathbb{E}P_2^2 &= \mathbb{E}[\mathbb{E}P_2^2 | S] = 4\tau \mathbb{E}[(\mu_1 - \mu_2)^T \hat{\tilde{V}} \hat{\tilde{V}}^T \Sigma \hat{\tilde{V}} \hat{\tilde{V}}^T (\mu_1 - \mu_2)] \\ &\leq 4\tau \mathbb{E}[\lambda_1(\hat{\tilde{V}}^T \Sigma \hat{\tilde{V}}) (\mu_1 - \mu_2)^T \hat{\tilde{V}} \hat{\tilde{V}}^T (\mu_1 - \mu_2)] \leq 4\tau \|\mu_1 - \mu_2\|^2 \mathbb{E}[\lambda_1(\hat{\tilde{V}}^T \Sigma \hat{\tilde{V}})] \\ &= O(\frac{\sqrt{p}}{n^2}) \mathbb{E}[\lambda_1(\hat{\tilde{V}}^T (V D^2 V^T + \sigma^2 I_p) \hat{\tilde{V}})] \leq O(\frac{\sqrt{p}}{n^2}) (\kappa p^\beta \mathbb{E}[\lambda_1(\hat{\tilde{V}}^T V V^T \hat{\tilde{V}})] + \sigma^2). \end{aligned}$$

By the following useful relationship

$$\lambda_1(\hat{\tilde{V}}^T V V^T \hat{\tilde{V}}) \leq \text{tr}(\hat{\tilde{V}}^T V V^T \hat{\tilde{V}}) = \frac{1}{2} \|V V^T - \hat{\tilde{V}} \hat{\tilde{V}}^T\|_F^2$$

and Lemma 2, we have that

$$\mathbb{E}P_2^2 = O\left(\frac{\sqrt{p}}{n^2}\right)\left(O\left(\frac{p}{n}\right) + \sigma^2\right) = o\left(\frac{p}{n^2}\right).$$

As for  $P_3$ . To prove the convergence in probability, here we prove the convergence in  $L^1$ :

$$\begin{aligned} \mathbb{E}|P_3| &= \mathbb{E}|(\mu_1 - \mu_2)^T(\hat{\tilde{V}}\hat{\tilde{V}}^T - \tilde{V}\tilde{V}^T)(\mu_1 - \mu_2)| \leq \|\mu_1 - \mu_2\|^2 \mathbb{E}\|\hat{\tilde{V}}\hat{\tilde{V}}^T - \tilde{V}\tilde{V}^T\| \\ &= \|\mu_1 - \mu_2\|^2 \mathbb{E}\|\hat{\tilde{V}}\hat{\tilde{V}}^T - VV^T\| \leq \|\mu_1 - \mu_2\|^2 \sqrt{\mathbb{E}\|\hat{\tilde{V}}\hat{\tilde{V}}^T - VV^T\|^2} \\ &\leq \|\mu_1 - \mu_2\|^2 \sqrt{\mathbb{E}\|\hat{\tilde{V}}\hat{\tilde{V}}^T - VV^T\|_F^2} = O\left(\frac{\sqrt{p}}{n}\right) \sqrt{O\left(\frac{p}{p^\beta n}\right)} = o\left(\frac{\sqrt{p}}{n}\right). \end{aligned}$$

Now we prove the asymptotic normality of  $P_1$ . To make clear the sense of convergence, we need a metric for weak convergence. For two distribution function  $F$  and  $G$ , the Levy metric  $\rho$  of  $F$  and  $G$  is defined as

$$\rho(F, G) = \inf\{\epsilon : F(x - \epsilon) - \epsilon \leq G(x) \leq F(x + \epsilon) + \epsilon \text{ for all } x\}.$$

It's well known that  $\rho(F_n, F) \rightarrow 0$  if and only if  $F_n \xrightarrow{\mathcal{L}} F$ .

The conditional distribution of  $\hat{\tilde{V}}^T((\bar{X}_1 - \mu_1) - (\bar{X}_2 - \mu_2))$  given  $S$  is  $N(0, \tau \hat{\tilde{V}}^T \Sigma \hat{\tilde{V}})$ .

As we have shown,

$$\lambda_1(\hat{\tilde{V}}^T \Sigma \hat{\tilde{V}}) \leq \frac{1}{2} \kappa p^\beta \|VV^T - \hat{\tilde{V}}\hat{\tilde{V}}^T\|_F^2 + \sigma^2 = O_P\left(\frac{p}{n} + 1\right).$$

On the other hand,  $\lambda_i(\hat{\tilde{V}}^T \Sigma \hat{\tilde{V}}) = \sigma^2$  for  $i = r + 1, \dots, p - r$ . Then

$$(p - 2r)\sigma^4 \leq \text{tr}(\hat{\tilde{V}}^T \Sigma \hat{\tilde{V}})^2 \leq \left(\frac{p}{n} + 1\right)^2 O_P(1) + (p - 2r)\sigma^4,$$

or

$$\text{tr}(\hat{\tilde{V}}^T \Sigma \hat{\tilde{V}})^2 = p\sigma^4(1 + o_P(1)). \quad (18)$$

It follows that

$$\frac{\lambda_1^2(\hat{\tilde{V}}^T \Sigma \hat{\tilde{V}})}{\text{tr}(\hat{\tilde{V}}^T \Sigma \hat{\tilde{V}})^2} = O_P\left(\frac{(p/n + 1)^2}{p}\right) = o_P(1). \quad (19)$$

Then for every subsequence of  $\{n\}$ , there's a further subsequence along which (19) holds almost surely. By Lemma 4, for every subsequence of  $\{n\}$ , there's a further



subsequence along which we have

$$\rho\left(\mathcal{L}\left(\frac{\|\hat{V}^T((\bar{X}_1 - \mu_1) - (\bar{X}_2 - \mu_2))\|^2 - \tau \text{tr}(\hat{V}^T \Sigma \hat{V})}{\sqrt{2\tau^2 \text{tr}(\hat{V}^T \Sigma \hat{V})^2}} \middle| S\right), N(0, 1)\right) \xrightarrow{a.s.} 0.$$

It means that

$$\rho\left(\mathcal{L}\left(\frac{\|\hat{V}^T((\bar{X}_1 - \mu_1) - (\bar{X}_2 - \mu_2))\|^2 - \tau \text{tr}(\hat{V}^T \Sigma \hat{V})}{\sqrt{2\tau^2 \text{tr}(\hat{V}^T \Sigma \hat{V})^2}} \middle| S\right), N(0, 1)\right) \xrightarrow{P} 0.$$

Thus the weak convergence also holds unconditionally:

$$\frac{\|\hat{V}^T((\bar{X}_1 - \mu_1) - (\bar{X}_2 - \mu_2))\|^2 - \tau \text{tr}(\hat{V}^T \Sigma \hat{V})}{\sqrt{2\tau^2 \text{tr}(\hat{V}^T \Sigma \hat{V})^2}} \xrightarrow{\mathcal{L}} N(0, 1).$$

Similar to (18) we have

$$\text{tr}(\hat{V}^T \Sigma \hat{V}) = (p - r)\sigma^2\left(1 + O_P\left(\frac{1}{n} + \frac{1}{p}\right)\right). \quad (20)$$

By (18), (20) and Slutsky's theorem,

$$\frac{\|\hat{V}^T((\bar{X}_1 - \mu_1) - (\bar{X}_2 - \mu_2))\|^2 - \sigma^2\tau(p - r)}{\sigma^2\sqrt{2\tau^2 p}} \xrightarrow{\mathcal{L}} N(0, 1).$$

Now the desired asymptotic properties of  $P_1$ ,  $P_2$  and  $P_3$  are established, the theorem follows.  $\square$

**Proof Of Theorem 3.** By assumption,  $\hat{r} \leq R$  for some constant  $R$ . Similar to the proof of Proposition 3, in the current context we have that  $\text{tr}(\hat{V}_i S_i \hat{V}_i) = \text{tr}\Sigma + P_P\left(\frac{\max(n, p)}{n}\right)$ ,  $i = 1, 2$ . It follows that

$$\frac{T_2 - \|\mu_1 - \mu_2\|^2}{\sqrt{2\tau^2 \text{tr}\Sigma^2}} = \frac{\|\hat{V}^T(\bar{X}_1 - \bar{X}_2)\|^2 - \|\mu_1 - \mu_2\|^2 - \tau \text{tr}\Sigma}{\sqrt{2\tau^2 \text{tr}\Sigma^2}} + o_P(1).$$

Since  $\bar{X}_i | \mu_i \sim N(\mu_i, \frac{1}{n_i} \Sigma)$  and  $\mu_i \sim N(0, \frac{\psi}{n_i \sqrt{p}} I_p)$ , we have  $\bar{X}_i \sim N(0, \frac{1}{n_i} (\Sigma + \frac{1}{\sqrt{p}} \psi I_p))$ ,  $i = 1, 2$ . Hence we have that  $\hat{V}^T(\bar{X}_1 - \bar{X}_2) | S \sim N(0, \tau \hat{V}^T (\Sigma + \frac{1}{\sqrt{p}} \psi I_p) \hat{V})$  by the independence of  $S$  and  $(\mu_1, \mu_2, \bar{X}_1, \bar{X}_2)$ . Note that

$$c + \frac{1}{\sqrt{p}} \psi \leq \lambda_{\min}(\hat{V}^T (\Sigma + \frac{1}{\sqrt{p}} \psi I_p) \hat{V}) \leq \lambda_{\max}(\hat{V}^T (\Sigma + \frac{1}{\sqrt{p}} \psi I_p) \hat{V}) \leq C + \frac{1}{\sqrt{p}} \psi.$$

Then by Lemma 4,

$$\frac{\|\hat{V}^T(\bar{X}_1 - \bar{X}_2)\|^2 - \tau \text{tr}(\hat{V}^T(\Sigma + \frac{1}{\sqrt{p}}\psi I_p)\hat{V})}{\sqrt{2\tau^2 \text{tr}(\hat{V}^T(\Sigma + \frac{1}{\sqrt{p}}\psi I_p)\hat{V})^2}} \xrightarrow{\mathcal{L}} N(0, 1). \quad (21)$$

It can be easily shown that

$$\frac{\text{tr}(\hat{V}^T(\Sigma + \frac{1}{\sqrt{p}}\psi I_p)\hat{V})^2}{\text{tr}(\hat{V}^T \Sigma \hat{V})^2} \xrightarrow{P} 1. \quad (22)$$

Next we will show that

$$\frac{\text{tr}(\hat{V}^T \Sigma \hat{V})^2}{\text{tr} \Sigma^2} \xrightarrow{P} 1. \quad (23)$$

In fact, for  $i = 1, \dots, p$  we have

$$\lambda_i(\hat{V}^T \Sigma \hat{V}) = \lambda_i(\Sigma^{1/2} \hat{V} \hat{V}^T \Sigma^{1/2}) \leq \lambda_i(\Sigma). \quad (24)$$

On the other hand, for  $i = 1, \dots, p - \hat{r}$  we have that

$$\lambda_i(\hat{V}^T \Sigma \hat{V}) = \lambda_i(\Sigma^{1/2}(I_p - \hat{V} \hat{V}^T) \Sigma^{1/2}) = \lambda_i(\Sigma - \Sigma^{1/2} \hat{V} \hat{V}^T \Sigma^{1/2}) \geq \lambda_{i+\hat{r}}(\Sigma), \quad (25)$$

where the last inequality holds by Weyl's inequality and the fact that the rank of  $\Sigma^{1/2} \hat{V} \hat{V}^T \Sigma^{1/2}$  is at most  $\hat{r}$ . By (24) and (25),

$$\sum_{i=\hat{r}+1}^p \lambda_i^2(\Sigma) \leq \text{tr}(\hat{V}^T \Sigma \hat{V})^2 \leq \text{tr} \Sigma^2.$$

Then  $|\text{tr}(\hat{V}^T \Sigma \hat{V})^2 - \text{tr} \Sigma^2| \leq \sum_{i=1}^{\hat{r}} \lambda_i^2(\Sigma) \leq RC^2$ . Hence (23) holds. By (21), (22), (23) and Slutsky's theorem,

$$\frac{\|\hat{V}^T(\bar{X}_1 - \bar{X}_2)\|^2 - \tau \text{tr}(\hat{V}^T \Sigma \hat{V}) - \frac{p-\hat{r}}{\sqrt{p}} \tau \psi}{\sqrt{2\tau^2 \text{tr} \Sigma^2}} \xrightarrow{\mathcal{L}} N(0, 1).$$

Note that

$$\begin{aligned} & \frac{\|\hat{V}^T(\bar{X}_1 - \bar{X}_2)\|^2 - \|\mu_1 - \mu_2\|^2 - \tau \text{tr} \Sigma^2}{\sqrt{2\tau^2 \text{tr} \Sigma}} \\ &= \frac{\|\hat{V}^T(\bar{X}_1 - \bar{X}_2)\|^2 - \tau \text{tr}(\hat{V}^T \Sigma \hat{V}) - \frac{p-\hat{r}}{\sqrt{p}} \tau \psi}{\sqrt{2\tau^2 \text{tr} \Sigma^2}} + \frac{\frac{p-\hat{r}}{\sqrt{p}} \psi - \frac{1}{\tau} \|\mu_1 - \mu_2\|^2}{\sqrt{2\text{tr} \Sigma^2}} + \frac{\text{tr}(\hat{V} \Sigma \hat{V}) - \text{tr} \Sigma^2}{\sqrt{2\text{tr} \Sigma}}. \end{aligned}$$

We only need to show the last two terms are negligible. But  $\frac{1}{\tau}\|\mu_1 - \mu_2\|^2 \sim \frac{\psi}{\sqrt{p}}\chi_p^2 = \sqrt{p}\psi + O_P(1)$  by central limit theorem, then

$$\frac{\frac{p-\hat{r}}{\sqrt{p}}\psi - \frac{1}{\tau}\|\mu_1 - \mu_2\|^2}{\sqrt{2\text{tr}\Sigma^2}} = o_P(1).$$

And

$$\frac{\text{tr}(\hat{V}\Sigma\hat{V}) - \text{tr}\Sigma^2}{\sqrt{2\text{tr}\Sigma}} = o_P(1)$$

by (24) and (25). The proof is completed.  $\square$

**Proof Of Theorem 4.** Since  $r = 0$ ,  $X_{ki} = \mu_k + Z_{ki}$ ,  $i = 1, \dots, n_k$  and  $k = 1, 2$ .

As in the proof of Theorem 3, we only need to prove

$$\frac{\|\hat{V}^T(\bar{Z}_1 - \bar{Z}_2) + \hat{V}^T(\mu_1 - \mu_2)\|^2 - \|\mu_1 - \mu_2\|^2 - \tau p \sigma^2}{\sigma^2 \sqrt{2\tau^2 p}} \xrightarrow{\mathcal{L}} N(0, 1).$$

Independent of data, generate a  $p \times p$  random orthogonal matrix with Haar invariant distribution. It can be seen that  $(O(\bar{Z}_1 - \bar{Z}_2), OSO^T) \sim ((\bar{Z}_1 - \bar{Z}_2), S)$  and are independent of  $O$ . But the eigenvectors of  $OSO^T$  are  $(O\hat{V}, O\hat{V})$ , thus  $(O(\bar{Z}_1 - \bar{Z}_2), O\hat{V}) \sim ((\bar{Z}_1 - \bar{Z}_2), \hat{V})$ . It follows that

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

Note that  $O(\mu_1 - \mu_2)/\|\mu_1 - \mu_2\|$  is uniformly distributed on the unit ball in  $\mathbb{R}^p$ .

Independent of data and  $O$ , generate a random variable  $R > 0$  with  $R^2 \sim \chi_p^2$ .

Then

$$\xi \stackrel{\text{def}}{=} R \frac{O(\mu_1 - \mu_2)}{\|\mu_1 - \mu_2\|} \sim N_p(0_p, I_p).$$

Now we have

$$\begin{aligned} &\|\hat{V}^T(\bar{Z}_1 - \bar{Z}_2) + \hat{V}^T O(\mu_1 - \mu_2)\|^2 \\ &= \|\hat{V}^T(\bar{Z}_1 - \bar{Z}_2)\|^2 + \frac{\|\hat{V}^T \xi\|^2}{R^2} \|\mu_1 - \mu_2\|^2 + \frac{\|\mu_1 - \mu_2\|}{R} \xi^T \hat{V} \hat{V}^T (\bar{Z}_1 - \bar{Z}_2). \end{aligned} \tag{26}$$

Since  $\hat{V}^T(\bar{Z}_1 - \bar{Z}_2)|\hat{V} \sim N_{p-\hat{r}}(0_{p-\hat{r}}, \tau\sigma^2 I_{p-\hat{r}})$ , the asymptotic normality of the first term of (26) follows by central limit theorem:

$$\frac{\|\hat{V}^T(\bar{Z}_1 - \bar{Z}_2)\|^2 - \tau(p - \hat{r})\sigma^2}{\sigma^2 \sqrt{2\tau^2(p - \hat{r})}} \xrightarrow{\mathcal{L}} N(0, 1). \tag{27}$$

By the fact that  $\hat{V}^T \xi | \hat{V} \sim N_{p-\hat{r}}(0_{p-\hat{r}}, I_{p-\hat{r}})$  and central limit theorem, we have

$$\|\hat{V}^T \xi\|^2 = (p - \hat{r})(1 + O_P(\frac{1}{\sqrt{p - \hat{r}}})) = p(1 + O_P(\frac{1}{\sqrt{p}})).$$

Also by central limit theorem,  $R^2 = p(1 + O_P(\frac{1}{\sqrt{p}}))$ . Thus for the second term of (26), we have

$$\frac{\|\hat{V}^T \xi\|^2}{R^2} \|\mu_1 - \mu_2\|^2 = \|\mu_1 - \mu_2\|^2 + O_P(\frac{1}{\sqrt{p}}) \|\mu_1 - \mu_2\|^2 = \|\mu_1 - \mu_2\|^2 + o_P(\sigma^2 \sqrt{2\tau^2 p}). \quad (28)$$

Now we deal with the second term of (26). Note that  $\xi^T \hat{V} \hat{V}^T (\bar{Z}_1 - \bar{Z}_2) | (\hat{V}, (\bar{Z}_1 - \bar{Z}_2)) \sim N(0, \|\hat{V}^T (\bar{Z}_1 - \bar{Z}_2)\|^2)$ , which implies that

$$\xi^T \hat{V} \hat{V}^T (\bar{Z}_1 - \bar{Z}_2) = O_P(1) \|\hat{V}^T (\bar{Z}_1 - \bar{Z}_2)\| = O_P(\sqrt{\tau p}).$$

It follows that

$$\frac{\|\mu_1 - \mu_2\|}{R} \xi^T \hat{V} \hat{V}^T (\bar{Z}_1 - \bar{Z}_2) = O_P(\sqrt{\tau}) \|\mu_1 - \mu_2\| = o_P(\sigma^2 \sqrt{2\tau^2 p}). \quad (29)$$

By (26), (27), (28), (29) and Slutsky's theorem, we have the conclusion

$$\frac{\|\hat{V}^T (\bar{Z}_1 - \bar{Z}_2) + \hat{V}^T O(\mu_1 - \mu_2)\|^2 - \|\mu_1 - \mu_2\|^2 - \tau p \sigma^2}{\sigma^2 \sqrt{2\tau^2 p}} \xrightarrow{\mathcal{L}} N(0, 1).$$

□

**Proof Of Theorem 5.** The method of Theorem 1's proof can still work here with some modifications. The term  $P_3$  in Theorem 1's proof disappears in the current circumstance. The other two terms can be treated as before if we can show that

$$\lambda_1(\hat{V}^T \Sigma_k \hat{V}) = O_P(\frac{p}{n}) \quad k=1,2.$$

In fact,

$$\lambda_1(\hat{V}^T \Sigma_k \hat{V}) = \lambda_1(\hat{V}^T V_k D_k^2 V_k^T \hat{V}) + \sigma^2 \leq \kappa p^\beta \lambda_1(\hat{V}^T V_k V_k^T \hat{V}) + \sigma^2.$$

But

$$\lambda_1(\hat{V}^T V_k V_k^T \hat{V}) = \lambda_1(V_k^T (I_p - \hat{V} \hat{V}^T) V_k) \leq \lambda_1(V_k^T (I_p - \hat{V}_k \hat{V}_k^T) V_k).$$

The last inequality holds since  $\hat{V}\hat{V}^T$  is the projection on the sum space of  $\hat{V}_1\hat{V}_1^T$  and  $\hat{V}_2\hat{V}_2^T$  and hence  $\hat{V}\hat{V}^T \geq \hat{V}_1\hat{V}_1^T$ . Thus,

$$\lambda_1(\hat{V}^T V_k V_k^T \hat{V}) \leq \frac{1}{2} \|V_k V_k^T - \hat{V}_k \hat{V}_k^T\|_F^2 = O_P\left(\frac{p}{np^\beta}\right).$$

Therefore,  $\lambda_1(\hat{V}^T \Sigma_k \hat{V}) = O_P(\frac{p}{n})$ . □

### 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. Repl 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. Rappt: 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.