High-dimensional two-sample test under spiked covariance

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

[a]*School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081,China*

[b]*Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing 100081,China*

Abstract

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

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

1. Introduction

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

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

*Corresponding author

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

In this paper, 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 Σ 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

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

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

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

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

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

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

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

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

adopted by Chen and Qin (2010) can be violated when $\Sigma = (1-c)I_p + c\mathbf{1}_p\mathbf{1}_p^T$ where $-1/(p-1) < c < 1$, I_p is the p dimensional identity matrix and $\mathbf{1}_p$ is the p dimensional vector with elements 1. 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 Σ is spiked (see Cai et al. (2013)). The random projection methods imply that test procedures are improved when data are projected on certain subspaces. We will see that the ideal subspace is the orthogonal complement of the principal space. Fortunately, the principal space can be estimated consistently even in high dimensional setting by the theory of principal component analysis (PCA). With the assumption of spiked covariance model, we propose a new test procedure through projection on the (estimated) ideal subspace. The asymptotic distribution of the test statistic is derived and hence asymptotic power is given. We will see that the test is more powerful than T_{CQ}. Moreover, even there's no strong correlation showing up, we prove that the new test performs equally well as T_{CQ} does. The idea is also generalized to the unequal variance setting and similar results still hold.

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

2. Model and assumptions

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

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

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

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

Assumption 2. *Let $X_{k,1}, \ldots, X_{k,n_k}$ be i.i.d. samples with common distribution $N(\mu_k, \Sigma_k)$, $k = 1, 2$. Assume Σ_k has structure $\Sigma_k = V_k \Lambda_k V_k^T + \sigma_k^2 I_p$, where $\Lambda_k = \operatorname{diag}(\lambda_{k,1}, \ldots, \lambda_{k,r_k})$, $\lambda_{k,1} \geq \cdots \geq \lambda_{k,r_k} > 0$, V_k is a $p \times r_k$ orthonormal matrix and $\sigma_k^2 > 0$, $k = 1, 2$. The factor number r_k and σ_k^2 are fixed as n_1, n_2, p vary.*

Under 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 $\tilde{V}_k \tilde{V}_k^T = I - V_k V_k^T$.

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

$$\hat{r} = \mathrm{argmax}_{l \le R} \frac{\lambda_l(S)}{\lambda_{l+1}(S)}, \tag{3}$$

where R is a hyperparameter. See Ahn and Horenstein (2013) for detail. Thus, without loss of generality, we will assume that r is known throughout the paper.

Assumption 3. *Assume that there exists $\kappa > 0$ and $\beta \ge 1/2$ such that*

$$\kappa p^{\beta} \ge \lambda_{k1} \ge \cdots \ge \lambda_{kr_k} \ge \kappa^{-1} p^{\beta}.$$

Remark 1. Proposition 1 will show that condition (2) is necessary for Chen and Qin (2010)'s method. Note that the condition (2) is equivalent to $\beta < 1/2$ in the current context. Hence $\beta \ge 1/2$ is the exact complement of Chen and Qin (2010)'s scope. The special case $\beta = 1$ corresponds to the factor model in paper Ma et al. (2015) with some restrictions.

In the rest of this section, we introduce some notations that will be used. Let $\tau = (n_1 + n_2)/(n_1 n_2)$, S be the pooled sample covariance:

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

where $S_k = (n_k - 1)^{-1} \sum_{i=1}^{n_k} (X_{k,i} - \bar{X}_k)(X_{k,i} - \bar{X}_k)^T$ is the sample covariance of the sample k, $k = 1, 2$. Denote by $\mathrm{Wishart}_p(m, \Psi)$ the p dimensional Wishart distribution with parameter Ψ and m degrees of freedom.

For random variable ξ and η, we write $\xi \sim \eta$ to denote they have the same distribution. Let $\mathcal{L}(\xi)$ be the distribution of ξ and $\mathcal{L}(\xi|\eta)$ be the conditional distribution of ξ given η. We denote by "$\xrightarrow{a.s.}$", "$\xrightarrow{P}$" and "$\xrightarrow{\mathcal{L}}$" the almost surely convergence, convergence in probability and weak convergence.

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 $n \to \infty$.

We denote by $\|\cdot\|$ and $\|\cdot\|_F$ the operator and Frobenius norm of matrix, separately. For $p \ge q$, define

$$\mathbb{O}_{p \times q} = \{O |\, O \text{ is } p \times q \text{ column orthonormal matrix }\}.$$

3. Methodology

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

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

To simplify notations, the subscript k of Σ_k, V_k, Λ_k, σ_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 $X_{k,i} \sim N_p(0, \Sigma)$, $k = 1, 2$, $i = 1, 2, \ldots, n_k$. Suppose Assumption 1 holds. Then (2) is a necessary and sufficient condition for*

$$\frac{T_{CQ} - \mathrm{E}T_{CQ}}{\left[\mathrm{Var}(T_{CQ})\right]^{1/2}} \xrightarrow{L} N(0, 1). \tag{4}$$

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, Σ'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

$$\tau\|V^T(\bar{X}_1 - \bar{X}_2)\|^2 + \tau\|\tilde{V}^T(\bar{X}_1 - \bar{X}_2)\|^2. \tag{5}$$

Under the null hypotheses, we have

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

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 ∞ as $p \to \infty$ 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 (5), the first term has larger variance and tends to contain much weaker signal. This motivates us to drop the first part of (5) and only use the second part. After adjustment of expectation, we define the following statistic

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

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

Proposition 2. *Under Assumptions 1-4 and local alternative, that means,* $\frac{n}{p}\|\mu_1 - \mu_2\|^2 \to 0$, *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 $X_{k,i}$ to $\tilde{V}^T X_{k,i}$ ($i = 1, \ldots, n_k$, $k = 1, 2$) and then invoking the statistic of Chen and Qin (2010). A class of 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). It is known that random projection based methods offer higher power when the variables are dependent. However, these test procedures are randomized, which is undesirable in practice. This raise the question: is there an optimal projection which is nonrandomized?

It can be seen that

$$\tilde{V} = \underset{O \in \mathbb{O}_{p\times(p-r)}}{\arg\min} \ \mathrm{Var}\left(\|O^T(\bar{X}_1 - \bar{X}_2)\|^2\right).$$

Thus, transformation by $\tilde{V}$ is optimal in the sense of variance reduction. 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 $\hat{\tilde{V}}$ the first r and last $p-r$ eigenvectors of S respectively. Similarly, we denote by $\hat{V}_k$ and $\hat{\tilde{V}}_k$ the first r and last $p-r$ eigenvectors of S_k

respectively, $k = 1, 2$. As estimators of their population counterparts, these simple statistics actually reach the optimal convergence rate (See Cai et al. (2013)).

Note that T_1 relies on the subspace $\tilde{V}\tilde{V}^T$ which is unknown and thus should be estimated. The first part of T_1, $\|\tilde{V}^T(\bar{X}_1 - \bar{X}_2)\|^2$, can be directly estimated by $\|\hat{\tilde{V}}^T(\bar{X}_1 - \bar{X}_2)\|^2$. Note that $n_1^{-1}\mathrm{tr}(\tilde{V}^T S_1 \tilde{V})$, the second part of T_1, only involves sample one. We estimate it by $n_1^{-1}\mathrm{tr}(\hat{\tilde{V}}_1^T S_1 \hat{\tilde{V}}_1)$. Similarly, we estimate the third part of T_1 by $n_2^{-1}\mathrm{tr}(\hat{\tilde{V}}_2^T S_2 \hat{\tilde{V}}_2)$. Define

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

The asymptotic result of Proposition 2 involves σ^2. In order to formulate a test procedure by asymptotic distribution, σ^2 needs to be consistently estimated. Note that σ^2 can be written as $\sigma^2 = (p-r)^{-1}\sum_{i=r+1}^{p}\lambda_i(\Sigma)$. Thus 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

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

In next section, it will be proved that the asymptotic null distribution of Q is $N(0,1)$. Thus, the null hypothesis is rejected when Q is larger than the upper α quantile of $N(0,1)$.

Remark 2. When both samples are simultaneously transformed by shift and orthogonal transformation, the statistic T_2 is invariant. More precisely, T_2 is invariant under the following transformation:

$$X_{1,i} \mapsto OX_{1,i} + \mu \text{ and } X_{2,j} \mapsto OX_{2,j} + \mu,\ i = 1, \ldots, n_1,\ j = 1, \ldots, n_2,$$

where $\mu \in \mathbb{R}^p$ and $O \in \mathbb{O}_{p\times p}$.

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

$$\frac{\|\tilde{V}^T(\mu_1-\mu_2)\|}{\|\mu_1-\mu_2\|}\approx 1. \tag{7}$$

Obviously, (7) 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, (7) is reasonable since $\tilde{V}\tilde{V}^T$ is nearly an identity matrix in the sense that $\|I_p-\tilde{V}\tilde{V}^T\|_F^2/\|I_p\|_F^2=r/p\to 0$. In bayesian framework, if we assume that the elements of μ_k are independently generated from certain prior distribution, it can be established that $\|\tilde{V}(\mu_1-\mu_2)\|/\|\mu_1-\mu_2\|\xrightarrow{P}1$. Such assumption for μ_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 σ^2. Our proof relies on the Weyl's inequality.

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

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

In the construction of T_2, we replace $\tilde{V}$ by $\hat{\tilde{V}}$. It is important to know the convergence property of $\hat{\tilde{V}}\hat{\tilde{V}}^T$ as an estimator of $\tilde{V}\tilde{V}^T$. However, $\hat{\tilde{V}}\hat{\tilde{V}}^T$ can not always consistently estimate $\tilde{V}\tilde{V}^T$ in high dimensional setting. The asymptotic normality of the new test statistic requires the following relationship between n and p:

Assumption 5. *Assume $p/n^2\to 0$.*

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

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

The Assumption 5 is undesirable in Theorem 1. However, it may not be able to be relaxed. In fact, conditioning on $\hat{\tilde{V}}$, the asymptotic normality of $\|\hat{\tilde{V}}^T(\bar{X}_1 - \bar{X}_2)\|^2$, the major part of T_2, requires

$$\frac{\lambda_1\big((\hat{\tilde{V}}^T\Sigma\hat{\tilde{V}})^2\big)}{\mathrm{tr}\big((\hat{\tilde{V}}^T\Sigma\hat{\tilde{V}})^2\big)} \xrightarrow{P} 0. \tag{8}$$

See Lemma 4 in appendix. By Lemma 2 in appendix, (8) is equivalent to Assumption 5.

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*

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

Note that the power of T_{CQ} is of the form

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

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

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

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 hold, the asymptotic normality is not valid. The next theorem gives an asymptotic result in this case.

Theorem 2. *Suppose Assumptions 1-4 hold and $\lambda_1 = \cdots = \lambda_r = \kappa p^\beta$. Suppose $\mu_1 = \mu_2$, $\beta > 1/2$ and $p/n^2 \to \infty$. We have*

$$\tau^{-1}\big\|\hat{\tilde{V}}^T\big(\bar{X}_1 - \bar{X}_2\big)\big\|^2$$

When Assumption 5 doesn't hold, permutation method can be used to determine the critical value. We will see from simulation results that the new test has good power behavior even if p is much large than n.

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 r is unknown and is estimated by (3), then $\hat{r}$ is nothing but a random integer which does not exceed R. And $\hat{V}\hat{V}^T$ is just a random projection. It is a natural question how the new test procedure behaves in this case. We study the asymptotic behavior of the new test procedure in two non-spiked setting.

First we consider the case when the eigenvalues of Σ is bounded. Similar to bayesian models, we assume a normal prior distribution for μ_k to characterize 'dense' alternative. The next theorem shows that the asymptotical power function of the new test is equal to Chen and Qin (2010)'s test in this case.

Theorem 3. *Assume $X_{k,i} \sim N(\mu_k, \Sigma)$, $i = 1, \ldots, n_k$, $k = 1, 2$. Suppose that Assumptions 1 and 5 holds. Suppose $0 < c \leq \lambda_p(\Sigma) \leq \lambda_1(\Sigma) \leq C < \infty$ where c and C are constants. Suppose the prior distribution of μ_k is $N(0, (n_k\sqrt{p})^{-1}\psi I_p)$, $k = 1, 2$, where ψ is a constant and $\hat{r} \leq R$ for a positive constant R. We have*

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

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 μ_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 when the covariance matrices are not spiked, the new test procedure also has good power performance.

5. Unequal variance

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

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

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

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

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

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 3. 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.

σ_n^2 can be estimated by ratio consistent estimators of σ_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 λ_i as p^β plus a random error from $U(0, 1)$.

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

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

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. Permutation method can produce an exact test procedure, see Lehmann and Romano (2005)'s Example 15.2.2. For T_{CQ}, note that permutation method only need it's main part, $\|\bar{X}_1 - \bar{X}_2\|^2$, which is also the main part of Bai and Saranadasa (1996) and Ma et al. (2015)'s method. These methods all produce the same permutation test. We plot the empirical power versus $\|\mu_1 - \mu_2\|$ when other parameters hold constant. The results are illustrated in figure 1. From the results, we can find that when Σ is spiked, the new test outperforms T_{CQ} substantially; when Σ is not spiked, the new test and T_{CQ} are comparable.

6.2. Real data analysis

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

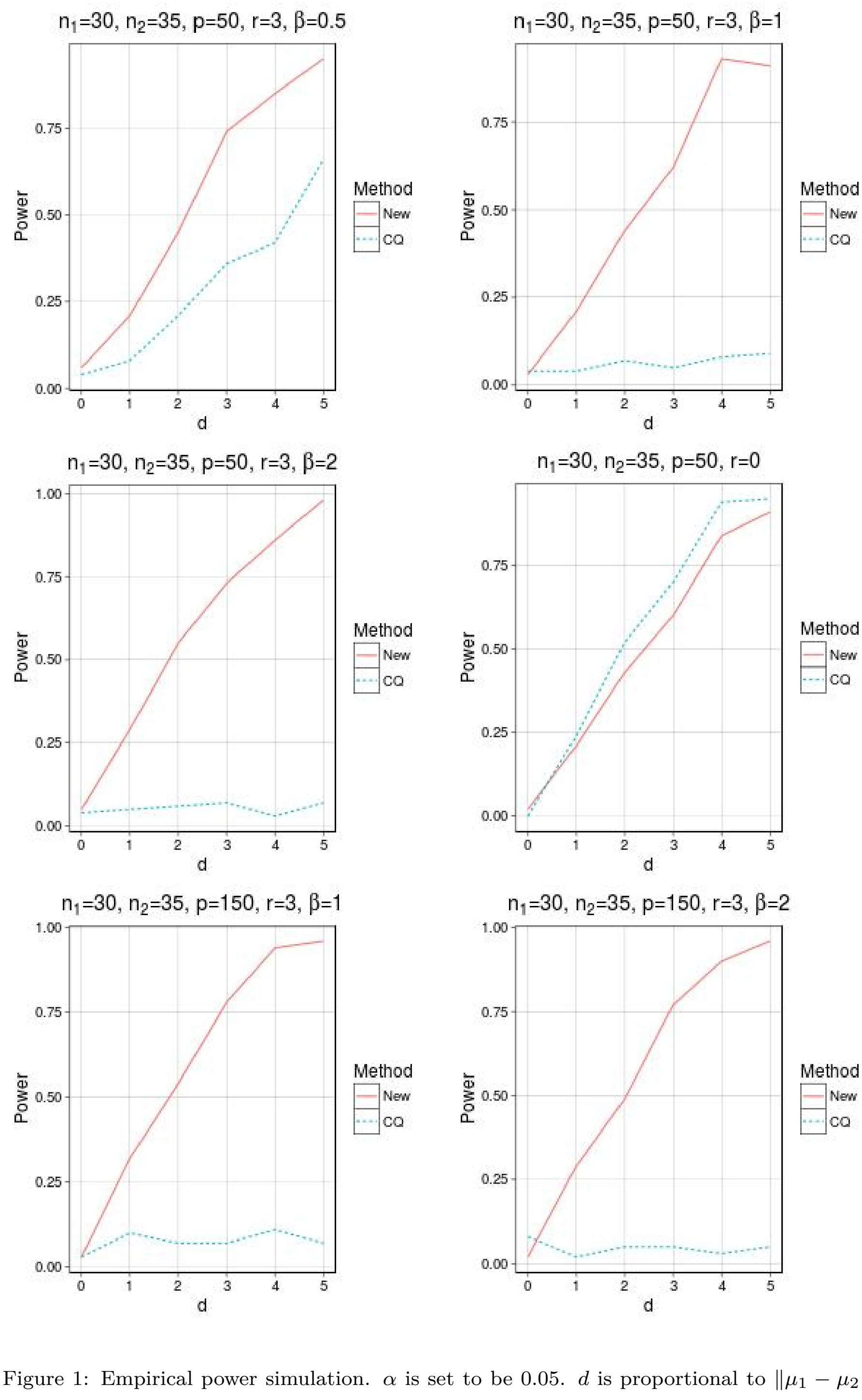

Figure 1: Empirical power simulation. α is set to be 0.05. d is proportional to $\|\mu_1 - \mu_2\|^2$. For each simulation, we do 50 permutations to determine critical value. We generate 100 independent samples to compute empirical power.

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 drop big variance terms from T_{CQ} and obtain a new test statistic. The asymptotic normality of the new statistic is proved and the asymptotic power is given.

In another paper, Zhao and Xu (2016) proved that 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, in some circumstances, as our work have shown, the complement of principal subspace is more useful.

Our theoretical results rely on the assumption $\sqrt{p}/n \to 0$. In the situation of small sample or very large p, the critical value of the new test can be determined by permutation method. Our simulation shows that the new test still performs well. It remains a theoretical interest to study the asymptotic behavior of permutation based test in these situations.

Appendix

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* $\operatorname{rank}(P) < k$*, then*

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

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

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

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

$$E\|\hat{V}\hat{V}^T - VV^T\|_F^2 \asymp r \wedge (p-r) \wedge \frac{r(p-r)}{(n_1+n_2-2)h(\lambda)} \tag{9}$$

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

□

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 $q \to \infty$ and $\frac{p}{q} \to c \in [0,\infty)$, 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 4. 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 $n \to \infty$,*

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

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

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 subsequance. 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 subsequance. Now the conclusion holds by a standard subsequance 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*

$$\frac{X_n^T A_n X_n - \mathrm{E} X_n^T A_n X_n}{[\mathrm{Var}(X_n^T A_n X_n)]^{1/2}} \xrightarrow{\mathcal{L}} N(0,1) \tag{10}$$

is that

$$\frac{\lambda_{\max}(A_n^2)}{\mathrm{tr}(A_n^2)} \to 0. \tag{11}$$

Remark 5. This lemma is from the Example 5.1 of Jiang (1996). Here we give a proof by characteristic function.

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

$$\frac{X_n^T A_n X_n - \mathrm{E} X_n^T A_n X_n}{[\mathrm{Var}(X_n^T A_n X_n)]^{1/2}} = \sum_{i=1}^{k_n} \frac{\lambda_i(A_n)}{\big[2\mathrm{tr}(A_n^2)\big]^{1/2}} (Z_{ni}^2 - 1), \tag{12}$$

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

If 11 holds, then

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

Hence 10 follows by Lindeberg's central limit theorem.

Conversely, if 10 holds, we will prove that there is a subsequence of $\{n\}$ along which 11 holds. Then 11 will hold by a standard contradiction argument.

Denote $c_{ni} = \lambda_i(A_n)/\big[2\mathrm{tr}(A_n^2)\big]^{1/2}$ ($i = 1, \ldots, k_n$), we have $c_{ni} \in [-\sqrt{2}/2, \sqrt{2}/2]$. Since 10 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

$$\log \mathrm{E} \exp \big(it \sum_{j=1}^{k_n} c_{nj}(Z_{nj}^2 - 1)\big) = -i(\sum_{j=1}^{k_n} c_{nj})t - \frac{1}{2}\sum_{j=1}^{k_n} \log(1 - i2c_{nj}t)$$

$$= -i(\sum_{j=1}^{k_n} c_{nj})t + \frac{1}{2}\sum_{j=1}^{k_n}\sum_{l=1}^{+\infty} \frac{1}{l}(i2c_{nj}t)^l = -i(\sum_{j=1}^{k_n} c_{nj})t + \frac{1}{2}\sum_{l=1}^{+\infty}\Big[\sum_{j=1}^{k_n}(c_{nj})^l\Big]\frac{1}{l}(i2t)^l$$

$$= -\frac{1}{2}t^2 + \frac{1}{2}\sum_{l=3}^{+\infty}\Big[\sum_{j=1}^{k_n}(c_{nj})^l\Big]\frac{1}{l}(i2t)^l.$$

Denote $b_{nl} = \sum_{j=1}^{k_n}(c_{nj})^l$, $n = 1,2,\cdots$ and $l = 3,4,\cdots$. For $l \geq 3$, $\big|\sum_{j=1}^{k_n}(c_{nj})^l\big| \leq \big|\sum_{j=1}^{k_n}(c_{nj})^2\big| = 1/2$. By Helly's selection theorem, there's a subsequence of $\{n\}$ along which $\lim_{n\to\infty} b_{nl} = b_l$ exists for every l. Apply dominated convergence theorem to this subsequence we have $\log \mathrm{E}\exp\big(it\sum_{j=1}^{k_n} c_{nj}(Z_{nj}^2-1)\big) \to -\frac{1}{2}t^2 + \frac{1}{2}\sum_{l=3}^{+\infty} b_l \frac{1}{l}(i2t)^l$ for $t \in (-1/2, 1/2)$. By the property of power series, we have $b_l = 0$ for $l \geq 3$. Then 11 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

$$\frac{\lambda_{\max}^2(\Sigma)}{\mathrm{tr}\Sigma^2} \to 0. \tag{13}$$

Let $Z_{k,i} = \Sigma^{-1/2}X_{k,i}$. Then $Z_{k,i} \sim N_p(0, I_p)$. Denote

$$Z = (Z_{1,1}^T, \ldots, Z_{1,n_1}^T, Z_{2,1}^T, \ldots, Z_{2,n_2}^T)^T.$$

Note that T_{CQ} is a quadratic form of Z and can be written as

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

where

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

γ_1 is an n_1 dimensional vector with all elements equal to $1/\sqrt{n_1}$ and γ_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

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

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

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

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

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

The eigenvector corresponding to 0 is

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

It follows that

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

And

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

The theorem follows by Lemma 4. □

Proof Of Proposition 2. Let $Y_{k,i} = \tilde{V}^T(X_{k,i} - \mu_k)$, $i = 1, \ldots, n_k$, $k = 1, 2$. Then $Y_{k,i} \sim N(\tilde{V}^T\mu_k, \sigma^2 I_{p-r})$. Let $\bar{Y}_1$ and $\bar{Y}_2$ be the sample means of $\{Y_{1,i}\}_{i=1}^{n_1}$ and $\{Y_{2,i}\}_{i=1}^{n_2}$ respectively. Then

$$\begin{aligned}
&\|\tilde{V}^T(\bar{X}_1 - \bar{X}_2)\|^2 = \|\tilde{V}^T(\mu_1 - \mu_2) + (\bar{Y}_1 - \bar{Y}_2)\|^2 \\
=&\|\tilde{V}^T(\mu_1 - \mu_2)\|^2 + \|\bar{Y}_1 - \bar{Y}_2\|^2 + 2(\mu_1 - \mu_2)^T\tilde{V}(\bar{Y}_1 - \bar{Y}_2) \\
=&\|\tilde{V}^T(\mu_1 - \mu_2)\|^2 + \|\bar{Y}_1 - \bar{Y}_2\|^2 + o_P(\frac{\sqrt{p}}{n}).
\end{aligned} \tag{14}$$

The last equality holds since

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

For $k=1,2$, we have

$$
\begin{aligned}
&\frac{1}{n_k}\mathrm{tr}(\tilde{V}^TS_k\tilde{V})\sim\frac{\sigma^2}{n_k(n_k-1)}\chi^2_{(p-r)(n_k-1)}\\
=&\sigma^2\frac{p-r}{n_k}(1+O_P(\frac{1}{\sqrt{(p-r)(n_k-1)}})),
\end{aligned}
$$

where the last equality comes from central limit theorem. It follows that

$$
\frac{1}{n_1}\mathrm{tr}(\tilde{V}^TS_1\tilde{V})+\frac{1}{n_2}\mathrm{tr}(\tilde{V}^TS_2\tilde{V})=\sigma^2\tau(p-r)+o_P(\frac{\sqrt{p}}{n}). \tag{15}
$$

Equation (14) and (15) imply that

$$
\frac{T_1-\|\tilde{V}^T(\mu_1-\mu_2)\|^2}{\sigma^2\sqrt{2\tau^2p}}=\frac{\|\bar{Y}_1-\bar{Y}_2\|^2-\sigma^2\tau(p-r)}{\sigma^2\sqrt{2\tau^2p}}+o_P(1). \tag{16}
$$

Since $\|\tilde{V}^T(\bar{Z}_1-\bar{Z}_2)\|^2\sim\sigma^2\tau\chi^2_{p-r}$, the proposition follows by central limit theorem. □

Proof Of Proposition 3. Note that $(n-2)S\sim\mathrm{Wishart}_p(n-2,\Sigma)$. Denote by $\Sigma=UEU^T$ the spectral decomposition of Σ, where $U=(V,\tilde{V})$ is an orthogonal matrix and $E=\mathrm{diag}(\lambda_1+\sigma^2,\ldots,\lambda_r+\sigma^2,\sigma^2,\ldots,\sigma^2)$. Let Z be a $p\times(n-2)$ random matrix with all elements i.i.d. distributed as $N(0,1)$, then

$$
S\sim\frac{1}{n-2}UE^{1/2}ZZ^TE^{1/2}U^T.
$$

Thus,

$$
\hat{\sigma}^2\sim\frac{1}{(p-r)(n-2)}\sum_{i=r+1}^{p}\lambda_i(UE^{1/2}ZZ^TE^{1/2}U^T)=\frac{1}{(p-r)(n-2)}\sum_{i=r+1}^{n-2}\lambda_i(Z^TEZ).
$$

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

$$
Z^TEZ=Z_{(1)}^T(\Lambda+\sigma^2I_r)Z_{(1)}+\sigma^2Z_{(2)}^TZ_{(2)}.
$$

The first term is of rank r. By Weyl's inequality, we have

$$\sigma^2\lambda_i(Z_{(2)}^T Z_{(2)}) \le \lambda_i(Z^T EZ) \le \sigma^2\lambda_{i-r}(Z_{(2)}^T Z_{(2)}), \quad i = r+1, \ldots, n-2.$$

Thus,

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

It follows that

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

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

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

The last equality comes from central limit theorem. The theorem follows by noting that

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

$\square$

Proof Of Theorem 1. Note that $\mathrm{tr}(\hat{\tilde{V}}_k^T S_k \hat{\tilde{V}}_k) = \sum_{i=r+1}^{p} \lambda_i(S_i)$, $k = 1, 2$. Similar to Proposition 3, we have $\mathrm{tr}(\hat{\tilde{V}}_k^T S_k \hat{\tilde{V}}_k) = (p-r)\sigma^2 + O_P(\max(n,p)/n)$, $k = 1, 2$. Then

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

By Assumption 5, $n^{-1}p^{-1/2}\max(n,p) = \max(p^{-1/2}, p^{1/2}/n) \to 0$. Note that

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

Let

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

To prove the theorem, it suffices 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 proves 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

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

By the relationship

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

and Lemma 2, we have that

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

Next we deal with P_3. To prove the convergence in probability, we prove the convergence in L^1.

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

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 ρ of F and G is defined as

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

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

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

$$\tau^{-1}\big\|\hat{\hat{V}}^T\big((\bar{X}_1-\mu_1)-(\bar{X}_2-\mu_2)\big)\big\|^2 \sim \sum_{i=1}^{p-r}\lambda_i(\hat{\hat{V}}^T\Sigma\hat{\hat{V}})\xi_i^2,$$

where $\{\xi_i\}_{i=1}^{p-r}$ are i.i.d. standard normal random variables which are independent of $\hat{\hat{V}}$. Note that

$$\lambda_1(\hat{\hat{V}}^T\Sigma\hat{\hat{V}}) \le \frac{1}{2}\kappa p^\beta\|VV^T-\hat{V}\hat{V}^T\|_F^2+\sigma^2.$$

Hence $\lambda_i(\hat{\hat{V}}^T\Sigma\hat{\hat{V}})=O_P(p/n+1)$, $i=1,\ldots,r$. Moreover, by Weyl's inequality, $\lambda_i(\hat{\hat{V}}^T\Sigma\hat{\hat{V}})=\sigma^2$, $i=r+1,\ldots,p-r$. Therefore

$$\mathrm{tr}(\hat{\hat{V}}^T\Sigma\hat{\hat{V}})^2=\big(\frac{p}{n}+1\big)^2O_P(1)+(p-2r)\sigma^4=p\sigma^4(1+o_P(1)). \tag{17}$$

It follows that

$$\frac{\lambda_1^2(\hat{\hat{V}}^T\Sigma\hat{\hat{V}})}{\mathrm{tr}(\hat{\hat{V}}^T\Sigma\hat{\hat{V}})^2}=O_P\Big(\frac{(p/n+1)^2}{p}\Big)=o_P(1). \tag{18}$$

Then for every subsequence of $\{n\}$, there's a further subsequence along which (18) holds almost surely. By Lemma 4, for every subsequence of $\{n\}$, there's a further subsequence along which we have

$$\rho\Big(\mathcal{L}\Big(\frac{\|\hat{\hat{V}}^T\big((\bar{X}_1-\mu_1)-(\bar{X}_2-\mu_2)\big)\|^2-\tau\mathrm{tr}(\hat{\hat{V}}^T\Sigma\hat{\hat{V}})}{\sqrt{2\tau^2\mathrm{tr}(\hat{\hat{V}}^T\Sigma\hat{\hat{V}})^2}}\Big|S\Big),N(0,1)\Big) \xrightarrow{a.s.} 0. \tag{19}$$

It means that (19) tends to 0 in probability. It can be seen that the weak convergence also holds unconditionally.

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

Similar to (17) we have

$$\mathrm{tr}(\hat{\tilde{V}}^T\Sigma\hat{\tilde{V}})=(p-r)\sigma^2\big(1+O_P\big(\frac{1}{n}+\frac{1}{p}\big)\big). \tag{20}$$

By (17), (20) and Slulsk's theorem,

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

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

Proof of Theorem 2. From the proof of Theorem 1 we know that

$$\tau^{-1}\big\|\hat{\tilde{V}}^T\big(\bar{X}_1-\bar{X}_2\big)\big\|^2\sim\sum_{i=1}^{p-r}\lambda_i(\hat{\tilde{V}}^T\Sigma\hat{\tilde{V}})\xi_i^2,$$

where $\{\xi_i\}_{i=1}^{p-r}$ are i.i.d. standard normal random variables which are independent of $\hat{\tilde{V}}$. Note that $\lambda_i(\hat{\tilde{V}}^T\Sigma\hat{\tilde{V}})=\kappa p^\beta\lambda_i(\hat{\tilde{V}}^TVV^T\hat{\tilde{V}})+\sigma^2$, $i=1,\dots,p-r$, we have

$$\sum_{i=1}^{p-r}\lambda_i(\hat{\tilde{V}}^T\Sigma\hat{\tilde{V}})\xi_i^2=\kappa p^\beta\sum_{i=1}^{r}\lambda_i(\hat{\tilde{V}}^TVV^T\hat{\tilde{V}})\xi_i^2+\sigma^2\sum_{i=1}^{p-r}\xi_i^2.$$

Existing PCA results for $\hat{\tilde{V}}^TVV^T\hat{\tilde{V}}$ are not enough for current purpose. Denote by $S=\hat{U}\hat{E}\hat{U}^T$ the spectral decomposition of S, where $\hat{U}=(\hat{V},\hat{\tilde{V}})$ and $\hat{E}=\mathrm{diag}(\hat{\lambda}_1,\dots,\hat{\lambda}_p)$. To study the property $\hat{\tilde{V}}^TVV^T\hat{\tilde{V}}$, we need to consider $\hat{\lambda}_i$, the ith eigenvalue of S, $i=1,\dots,r$.

Denote by $\Sigma=UEU^T$ the spectral decomposition of Σ, where $U=(V,\tilde{V})$ is an orthogonal matrix and $E=\mathrm{diag}(\lambda_1+\sigma^2,\dots,\lambda_r+\sigma^2,\sigma^2,\dots,\sigma^2)$. Let Z be a $p\times(n-2)$ random matrix with all elements i.i.d. distributed as $N(0,1)$. Denote $Z=(Z_{(1)}^T,Z_{(2)}^T)^T$, where $Z_{(1)}$ and $Z_{(2)}$ are the first r rows and last $p-r$ rows of Z. We have

$$S\sim\frac{1}{n-2}UE^{1/2}ZZ^TE^{1/2}U^T. \tag{21}$$

It can be seen that $\hat{\lambda}_i = \lambda_i(S) \sim (n-2)^{-1}\lambda_i(Z^T E Z)$, $i = 1, \dots, r$. Note that

$$Z^T E Z = (\kappa p^\beta + \sigma^2) Z_{(1)}^T Z_{(1)} + \sigma^2 Z_{(2)}^T Z_{(2)}.$$

By Bai-Yin's law, $\lambda_{\max}(Z_{(2)}^T Z_{(2)}) = p(1 + o_P(1))$ and $\lambda_{\min}(Z_{(2)}^T Z_{(2)}) = p(1 + o_P(1))$. By law of large numbers, $\lambda_i(Z_{(1)}^T Z_{(1)}) = \lambda_i(Z_{(1)} Z_{(1)}^T) = n(1 + o_P(1))$, $i = 1, \dots, r$. Since

$$\lambda_i(Z_{(1)}^T Z_{(1)}) + \lambda_{\min}(Z_{(2)}^T Z_{(2)}) \le \lambda_i(Z^T E Z) \le \lambda_i(Z_{(1)}^T Z_{(1)}) + \lambda_{\max}(Z_{(2)}^T Z_{(2)}),$$

we can deduce that

$$\hat{\lambda}_i = (\kappa n p^\beta + p\sigma^2)(1 + o_P(1)), \quad i = 1, \dots, r. \tag{22}$$

Next, note that (21) implies $\sigma^{-2}\tilde{V}^T S \tilde{V} \sim (n-2)^{-1} Z_{(2)} Z_{(2)}^T$. Let $\hat{E}_1 = (\hat{\lambda}_1, \dots, \hat{\lambda}_r)$, $\hat{E}_2 = (\hat{\lambda}_{r+1}, \dots, \hat{\lambda}_p)$, then $S = \hat{V}\hat{E}_1\hat{V}^T + \hat{\hat{V}}\hat{E}_2\hat{\hat{V}}^T$. We have

$$\tilde{V}^T \hat{V} \hat{E}_1 \hat{V}^T \tilde{V} + \tilde{V}^T \hat{\hat{V}} \hat{E}_2 \hat{\hat{V}}^T \tilde{V} \sim \frac{\sigma^2}{n-2} Z_{(2)} Z_{(2)}^T$$

By Bai-Yin's law, the right hand side. By asymptotic Cochran's theorem.

$$\lambda_1 \lambda_{n-2}$$

$\square$

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

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

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

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

Then by Lemma 4,

$$\frac{\|\hat{\tilde{V}}^T(\bar{X}_1-\bar{X}_2)\|^2-\tau\mathrm{tr}(\hat{\tilde{V}}^T(\Sigma+\frac{1}{\sqrt{p}}\psi I_p)\hat{\tilde{V}})}{\sqrt{2\tau^2\mathrm{tr}(\hat{\tilde{V}}^T(\Sigma+\frac{1}{\sqrt{p}}\psi I_p)\hat{\tilde{V}})^2}}\xrightarrow{\mathcal{L}}N(0,1). \tag{23}$$

It can be easily shown that

$$\frac{\mathrm{tr}(\hat{\tilde{V}}^T(\Sigma+\frac{1}{\sqrt{p}}\psi I_p)\hat{\tilde{V}})^2}{\mathrm{tr}(\hat{\tilde{V}}^T\Sigma\hat{\tilde{V}})^2}\xrightarrow{P}1. \tag{24}$$

Next we will show that

$$\frac{\mathrm{tr}(\hat{\tilde{V}}^T\Sigma\hat{\tilde{V}})^2}{\mathrm{tr}\Sigma^2}\xrightarrow{P}1. \tag{25}$$

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

$$\lambda_i(\hat{\tilde{V}}^T\Sigma\hat{\tilde{V}})=\lambda_i(\Sigma^{1/2}\hat{\tilde{V}}\hat{\tilde{V}}^T\Sigma^{1/2})\leq\lambda_i(\Sigma). \tag{26}$$

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

$$\lambda_i(\hat{\tilde{V}}^T\Sigma\hat{\tilde{V}})=\lambda_i(\Sigma^{1/2}(I_p-\hat{V}\hat{V}^T)\Sigma^{1/2})=\lambda_i(\Sigma-\Sigma^{1/2}\hat{V}\hat{V}^T\Sigma^{1/2})\geq\lambda_{i+\hat{r}}(\Sigma), \tag{27}$$

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 (26) and (27),

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

Then $|\mathrm{tr}(\hat{\tilde{V}}^T\Sigma\hat{\tilde{V}})^2-\mathrm{tr}\Sigma^2|\leq\sum_{i=1}^{\hat{r}}\lambda_i^2(\Sigma)\leq RC^2$. Hence (25) holds. By (23), (24), (25) and Slutsky's theorem,

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

Note that

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

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

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

by (26) and (27). The proof is completed.

$\square$

Proof Of Theorem 4. Since $r = 0$, $X_{ki} = \mu_k + Z_{ki}$, $i = 1, \ldots, n_k$ and $k = 1, 2$. As in the proof of Theorem 3, we only need to prove

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

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

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

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

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

Now we have

$$\begin{aligned}&\|\hat{\hat{V}}^T(\bar{Z}_1 - \bar{Z}_2) + \hat{\hat{V}}^T O(\mu_1 - \mu_2)\|^2 \\ =&\|\hat{\hat{V}}^T(\bar{Z}_1 - \bar{Z}_2)\|^2 + \frac{\|\hat{\hat{V}}^T\xi\|^2}{R^2}\|\mu_1 - \mu_2\|^2 + \frac{\|\mu_1 - \mu_2\|}{R}\xi^T\hat{\hat{V}}\hat{\hat{V}}^T(\bar{Z}_1 - \bar{Z}_2).\end{aligned} \tag{28}$$

Since $\hat{\hat{V}}^T(\bar{Z}_1 - \bar{Z}_2)|\hat{\hat{V}} \sim N_{p-\hat{r}}(0_{p-\hat{r}}, \tau\sigma^2 I_{p-\hat{r}})$, the asymptotic normality of the first term of (28) follows by central limit theorem:

$$\frac{\|\hat{\hat{V}}^T(\bar{Z}_1 - \bar{Z}_2)\|^2 - \tau(p - \hat{r})\sigma^2}{\sigma^2\sqrt{2\tau^2(p - \hat{r})}} \xrightarrow{\mathcal{L}} N(0, 1). \tag{29}$$

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

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

Also by central limit theorem, $R^2 = p(1+O_P(\frac{1}{\sqrt{p}}))$. Thus for the second term of (28), we have

$$\frac{\|\hat{\hat{V}}^T\xi\|^2}{R^2}\|\mu_1-\mu_2\|^2 = \|\mu_1-\mu_2\|^2+O_P(\frac{1}{\sqrt{p}})\|\mu_1-\mu_2\|^2 = \|\mu_1-\mu_2\|^2+o_P(\sigma^2\sqrt{2\tau^2p}). \tag{30}$$

Now we deal wih the second term of (28). Note that $\xi^T\hat{\hat{V}}\hat{\hat{V}}^T(\bar{Z}_1-\bar{Z}_2)\big|(\hat{\hat{V}},(\bar{Z}_1-\bar{Z}_2)) \sim N(0,\|\hat{\hat{V}}^T(\bar{Z}_1-\bar{Z}_2)\|^2)$, which implies that

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

It follows that

$$\frac{\|\mu_1-\mu_2\|}{R}\xi^T\hat{\hat{V}}\hat{\hat{V}}^T(\bar{Z}_1-\bar{Z}_2) = O_P(\sqrt{\tau})\|\mu_1-\mu_2\| = o_P(\sigma^2\sqrt{2\tau^2p}). \tag{31}$$

By (28), (29), (30), (31) and Slutsky's theorem, we have the conclusion

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

□

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

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

In fact,

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

But

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

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,

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

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

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. 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.