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 matrix is assumed to be spiked, which often arises in practice. We derive the asymptotic distribution of Chen and Qin (2010)'s test statistic under spiked covariance. Also, a new test procedure is proposed 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 matrix is spiked.

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

1. Introduction

Suppose $X_{k,1},\dots,X_{k,n_k}$ are independent identically distributed (i.i.d.) p-dimensional normal random vectors with unknown mean vector μ_k and covariance matrix $\mathbf{\Sigma}$, $k=1,2$. We consider the hypothesis testing problem

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

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

In this paper, the high dimensional setting is adopted, that is, the dimension p varies as n increases, where $n = n_1 + n_2$ is the total sample size. Testing hypotheses (1) is important in many fields, including biology, finance and economics.

A classical test statistic for hypotheses (1) is Hotelling's T^2 test statistic $(\bar{X}_1 - \bar{X}_2)^T \mathbf{S}^{-1}(\bar{X}_1 - \bar{X}_2)$, where $\bar{X}_k = n_k^{-1}\sum_{i=1}^{n_k} X_{k,i}$ is the mean vector of sample k, $k = 1, 2$, and $\mathbf{S} = (n-2)^{-1}\sum_{k=1}^{2}\sum_{i=1}^{n_k}(X_{k,i} - \bar{X}_k)(X_{k,i} - \bar{X}_k)^T$ is the pooled sample covariance matrix. However, Hotelling's test statistic 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 the low power of Hotelling's test is that S is a poor estimator of $\mathbf{\Sigma}$ is large compared with n. See Chen and Qin (2010) and the references therein. For testing hypotheses (1) in high dimensional settings, many test statistics are based on the estimation of $(\mu_1 - \mu_2)^T \mathbf{A}(\mu_1 - \mu_2)$ for a positive definite matrix $\mathbf{A}$. Bai and Saranadasa (1996) proposed a test based on

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

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

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

where $\mathbf{S}_k = (n_k - 1)^{-1}\sum_{i=1}^{n_k}(X_{k,i} - \bar{X}_k)(X_{k,i} - \bar{X}_k)^T$, $k = 1, 2$. As an estimator of $\|\mu_1 - \mu_2\|^2$, T_{CQ} is unbiased even if the covariances of two samples are different, while T_{BS} is unbiased only when the covariances are the same or $n_1 = n_2$. Srivastava and Du (2008) proposed a test based on

$$T_{SD} = (\bar{X}_1 - \bar{X}_2)^T [\operatorname{diag}(\mathbf{S})]^{-1}(\bar{X}_1 - \bar{X}_2),$$

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

As Ma et al. (2015) pointed out, however, the asymptotic properties of these test procedures may not be valid if strong correlations exist. For example, the condition

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

adopted by Chen and Qin (2010) is violated when $\mathbf{\Sigma}$ has a uniform correlation structure, that is, $\mathbf{\Sigma} = (1-\rho)\mathbf{I}_p + \rho \mathbf{1}_p \mathbf{1}_p^T$ where $0 < \rho < 1$, $\mathbf{I}_p$ is the p dimensional identity matrix and $\mathbf{1}_p$ is the p dimensional vector with elements 1. In this case, $\mathbf{\Sigma}$ has eigenvalues $1+\rho(p-1)$ and $1-\rho$ with multiplicities 1 and $p-1$ respectively. Then (2) is violated since

$$\frac{\mathrm{tr}(\mathbf{\Sigma}^4)}{\mathrm{tr}^2(\mathbf{\Sigma}^2)} = \frac{\big(1+\rho(p-1)\big)^4 + (1-\rho)^4(p-1)}{\big[\big(1+\rho(p-1)\big)^2 + (1-\rho)^2(p-1)\big]^2} \to 1$$

as $p \to \infty$. Under uniform correlation structure, the leading eigenvalue of $\mathbf{\Sigma}$ is significantly larger than the rest of eigenvalues. This is a special case of the spiked covariance model

$$\mathbf{\Sigma} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T + \sigma^2 \mathbf{I}_p, \tag{3}$$

where $\mathbf{\Lambda} = \mathrm{diag}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r)$, $\boldsymbol{\lambda}_1 \geq \cdots \geq \boldsymbol{\lambda}_r > 0$, $r \geq 1$, $\mathbf{V}$ is a $p \times r$ orthonormal matrix and $\sigma^2 > 0$. The spiked covariance model (3) is adopted by many theoretical studies, see Cai et al. (2013), Birnbaum et al. (2013), Passemier et al. (2017) and the references therein. The spiked covariance arises when variables are strongly correlated and the correlations are determined by a small number of factors.

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. In section 2, it will be seen that the asymptotic normality of T_{CQ} is not valid when λ_i's in (3) are large. Generally, the asymptotic distribution of T_{CQ} is the distribution of a weighted sum of chi-squared random variables. In a special case, the asymptotic distribution is the distribution of a weighted sum of chi-squared random variables and a normal random variable.

Recently, a class of test procedures are proposed through random projection. See Lopes et al. (2011), Thulin (2014) and Srivastava et al. (2016). The idea is to project data on some random lower-dimensional subspaces. It has been shown that these procedures perform well under strong correlations. 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 null distribution of the test statistic is derived and asymptotic power is also given. We will see that the test is more powerful than T_{CQ}.

The rest of the paper is organized as follows. In Section 2, we revisit Chen and Qin (2010)'s test. In Section 3, we propose a test procedure and exploit properties of the test. In Section 4, simulations are carried out and a real data example is given. Section 5 contains some discussion. All the technical details are in appendix.

2. Asymptotic properties of Chen and Qin (2010)'s test

Throughout the paper, we assume $p \to \infty$ as $n \to \infty$ and $n_1/n_2 \to c \in (0, +\infty)$, that is, we consider high dimensional and balanced data.

In Chen and Qin (2010), the asymptotic normality of T_{CQ} is derived under the condition (2). We shall show that under the null hypothesis, the condition (2) is essential for the asymptotic normality of T_{CQ}. We note that under the null hypothesis, T_{CQ} is a quadratic form of a standard normal random vector. To see this, let $Z_{k,i} = \mathbf{\Sigma}^{-1/2} X_{k,i}$, $k = 1, 2$, $i = 1, \ldots, n_k$. It can be seen that $Z_{k,i}$ is $N_p(0, \mathbf{I}_p)$ distributed under the null hypothesis. Write $Z = (Z_{1,1}^T, \ldots, Z_{1,n_1}^T, Z_{2,1}^T, \ldots, Z_{2,n_2}^T)^T$. Then $T_{CQ} = Z^T\big(\mathbf{B}_n \otimes \mathbf{\Sigma}\big)Z$, where $\otimes$ is the Kronecker product and

$$\mathbf{B}_n = \begin{pmatrix} \frac{1}{n_1(n_1-1)}(\mathbf{1}_{n_1}\mathbf{1}_{n_1}^T - \mathbf{I}_{n_1}) & -\frac{1}{n_1 n_2}\mathbf{1}_{n_1}\mathbf{1}_{n_2}^T \\ -\frac{1}{n_1 n_2}\mathbf{1}_{n_2}\mathbf{1}_{n_1}^T & \frac{1}{n_2(n_2-1)}(\mathbf{1}_{n_2}\mathbf{1}_{n_2}^T - \mathbf{I}_{n_2}) \end{pmatrix}.$$

Using characteristic function method, one can prove the following result which gives a necessary and sufficient condition for the asymptotic normality of the quadratic form of a standard normal random vector.

Lemma 1. *Suppose Y_n is a k_n dimensional standard normal random vector and $\mathbf{A}_n$ is a $k_n \times k_n$ symmetric matrix. Then as $n \to \infty$, a necessary and sufficient condition for*

$$\frac{Y_n^T \mathbf{A}_n Y_n - \mathrm{E}\, Y_n^T \mathbf{A}_n Y_n}{[\mathrm{Var}(Y_n^T \mathbf{A}_n Y_n)]^{1/2}} \xrightarrow{\mathcal{L}} N(0,1) \tag{4}$$

is that

$$\frac{\lambda_1(\mathbf{A}_n^2)}{\mathrm{tr}(\mathbf{A}_n^2)} \to 0, \tag{5}$$

where "$\xrightarrow{\mathcal{L}}$" means convergence of a sequence of random variables in law and $\lambda_i(\cdot)$ means the ith largest eigenvalue.

To apply Lemma 1 to T_{CQ}, one needs to calculate the eigenvalues of $\mathbf{B}_n \otimes \mathbf{\Sigma}$. Note that the eigenvalues of $\mathbf{B}_n$ 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. Thus,

$$\mathrm{tr}\,(\mathbf{B}_n \otimes \mathbf{\Sigma})^2 = \mathrm{tr}(\mathbf{B}_n^2)\,\mathrm{tr}\,\mathbf{\Sigma}^2 = (\frac{1}{n_1(n_1-1)} + \frac{1}{n_2(n_2-1)} + \frac{2}{n_1n_2})\,\mathrm{tr}\,\mathbf{\Sigma}^2,$$

and

$$\lambda_1\Big((\mathbf{B}_n \otimes \mathbf{\Sigma})^2\Big) = \lambda_1(\mathbf{B}_n^2)\lambda_1(\mathbf{\Sigma}^2) = \Big(\frac{1}{n_1} + \frac{1}{n_2}\Big)^2 \lambda_1(\mathbf{\Sigma}^2).$$

Because $n_1/n_2 \to c$, the condition

$$\frac{\lambda_1\Big((\mathbf{B}_n \otimes \mathbf{\Sigma})^2\Big)}{\mathrm{tr}\,\Big(\mathbf{B}_n \otimes \mathbf{\Sigma}\Big)^2} \to 0$$

is equivalent to $\lambda_1(\mathbf{\Sigma}^2)/\,\mathrm{tr}\,\mathbf{\Sigma}^2 \to 0$. From

$$\frac{\lambda_1(\mathbf{\Sigma})^4}{(\sum_{i=1}^p \lambda_i(\mathbf{\Sigma})^2)^2} \le \frac{\sum_{i=1}^p \lambda_i(\mathbf{\Sigma})^4}{(\sum_{i=1}^p \lambda_i(\mathbf{\Sigma})^2)^2} \le \frac{\lambda_1(\mathbf{\Sigma})^2 \sum_{i=1}^p \lambda_i(\mathbf{\Sigma})^2}{(\sum_{i=1}^p \lambda_i(\mathbf{\Sigma})^2)^2} = \frac{\lambda_1(\mathbf{\Sigma})^2}{\sum_{i=1}^p \lambda_i(\mathbf{\Sigma})^2},$$

we can see that $\lambda_1^2(\mathbf{\Sigma})/\mathrm{tr}(\mathbf{\Sigma}^2) \to 0$ is equivalent to (2). Then Lemma 1 implies that under the null hypothesis, the condition (2) is a necessary and sufficient condition for

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

The above result implies that Chen and Qin (2010)'s test procedure can be used only when the eigenvalues of $\mathbf{\Sigma}$ are concentrated around their average. In a class of applications, however, the correlations between variables are mainly driven by several common factors, and consequently, $\mathbf{\Sigma}$ has a few eigenvalues which are much larger than the others. See, for example, Jung and Marron (2009), Cai et al. (2013) and Fan and Wang (2015). To characterize such correlations between variables, we consider the spiked covariance structure (3). For $p \geq q$, let $\mathbb{O}_{p\times q}$ denote the collection of all $p \times q$ column orthogonal matrices. We make the following assumption for the covariance matrix $\mathbf{\Sigma}$.

Assumption 1. *The covariance matrix $\mathbf{\Sigma}$ has structure $\mathbf{\Sigma} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T + \sigma^2\mathbf{I}_p$, where $\mathbf{V} \in \mathbb{O}_{p\times r}$, r is a known number and $\mathbf{\Lambda} = \mathrm{diag}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r)$, $\boldsymbol{\lambda}_1 \geq \cdots \geq \boldsymbol{\lambda}_r > 0$. As n, p tend to infinity, the parameters r, σ^2 are fixed and $\mathbf{\Lambda}$ satisfies*

$$\kappa p^\beta \geq \boldsymbol{\lambda}_1 \geq \cdots \geq \boldsymbol{\lambda}_r \geq \kappa^{-1} p^\beta,$$

where $\kappa > 1$ and $\beta \geq 1/2$ are constants.

The covariance structure in Assumption 1 is commonly adopted in PCA study. See Cai et al. (2013), Birnbaum et al. (2013), Passemier et al. (2017) and the references therein. This covariance structure is also connected with the factor model. In fact, the model in Assumption 1 with $\beta = 1$ corresponds to the factor model in Ma et al. (2015) with homoscedastic noise.

In Assumption 1, the column space of $\mathbf{V}$ is the eigenspace of $\mathbf{\Sigma}$ associated with the r leading eigenvalues, and is therefore called principal space. Since $\mathbf{V}$ is a column orthogonal matrix, $\mathbf{V}\mathbf{V}^T$ is the orthogonal projection onto the principal space. Let $\tilde{\mathbf{V}}$ be a member of $\mathbb{O}_{p\times(p-r)}$ such that the columns of $\tilde{\mathbf{V}}$ are orthogonal to the columns of $\mathbf{V}$. Although such $\tilde{\mathbf{V}}$ is not unique, the orthogonal projection $\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T = \mathbf{I}_p - \mathbf{V}\mathbf{V}^T$ is unique and is equal to the orthogonal projection onto the orthogonal complement of principal space.

For positive sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \asymp b_n$ to denote $a_n = O(b_n)$ and $b_n = O(a_n)$ as $n \to \infty$. Under Assumption 1, we have

$$\frac{\operatorname{tr}(\mathbf{\Sigma}^4)}{\operatorname{tr}^2(\mathbf{\Sigma}^2)} = \frac{\sum_{i=1}^r (\boldsymbol{\lambda}_i + \sigma^2)^4 + (p-r)\sigma^8}{\left(\sum_{i=1}^r (\boldsymbol{\lambda}_i + \sigma^2)^2 + (p-r)\sigma^4\right)^2} \asymp \frac{p^{4\beta} + p}{(p^{2\beta} + p)^2}.$$

The right hand side tends to 0 if and only if $\beta < 1/2$. Our previous arguments assert that the asymptotic distribution of T_{CQ} won't be normal for $\beta \geq 1/2$. To derive the asymptotic distribution of T_{CQ} for $\beta \geq 1/2$, note that the variation of T_{CQ} is mainly due to $\|\bar{X}_1 - \bar{X}_2\|^2$. Let $\tau = 1/n_1 + 1/n_2$. Under the null hypothesis, we have

$$\operatorname{Var}(\|\bar{X}_1 - \bar{X}_2\|^2) = 2\tau^2 \operatorname{tr}(\mathbf{\Sigma}^2) = 2\tau^2 \sum_{i=1}^r (\boldsymbol{\lambda}_i + \sigma^2)^2 + 2\tau^2 (p-r)\sigma^4,$$

where the first term of the right hand side is of order $p^{2\beta}/n^2$ and the second term is of order p/n^2. If $\beta = 1/2$, the two terms are of the same order. If $\beta > 1/2$, however, the second term is dominated by the first term. This implies that the asymptotic distributions of T_{CQ} are different for $\beta = 1/2$ and $\beta > 1/2$. Since the variance of $(\tau p^\beta)^{-1}\|\bar{X}_1 - \bar{X}_2\|^2$ is bounded under the null hypothesis, we use τp^β to standardize T_{CQ}. The following two theorems give the asymptotic distributions of $(\tau p^\beta)^{-1} T_{CQ}$ when $\beta = 1/2$ and $\beta > 1/2$, respectively.

Theorem 1. *Under Assumption 1, suppose $\beta = 1/2$ and $\lambda_i / p^\beta \to \omega_i \in (0, +\infty)$, $i = 1, \ldots, r$. Let $Z_0, Z_1, \ldots, Z_r$ be i.i.d. $N(0,1)$ random variables, then the following results hold:*

(a) *If $\mu_1 = \mu_2$, then*

$$\frac{1}{\tau p^\beta} T_{CQ} \xrightarrow{w} \sqrt{2}\sigma^2 Z_0 + \sum_{i=1}^r \omega_i Z_i^2 - \sum_{i=1}^r \omega_i,$$

where "$\xrightarrow{w}$" denotes weak convergence.

(b) *If $(\tau p^\beta)^{-1/2}\left(\mathbf{V}^T(\mu_1 - \mu_2)\right)_i \to \zeta_i \in (-\infty, +\infty)$, $i = 1, \ldots, r$, and $(\tau p^\beta)^{-1}\|\tilde{\mathbf{V}}^T(\mu_1 - \mu_2)\|^2 \to \zeta^* \in [0, +\infty)$, then*

$$\frac{1}{\tau p^\beta} T_{CQ} \xrightarrow{w} \sqrt{2}\sigma^2 Z_0 + \sum_{i=1}^r (\sqrt{\omega_i} Z_i + \zeta_i)^2 + \zeta^* - \sum_{i=1}^r \omega_i.$$

Theorem 2. *Under Assumption 1, suppose $\beta > 1/2$ and $\lambda_i/p^\beta \to \omega_i \in (0, +\infty)$, $i = 1, \dots, r$. Let $Z_1, \dots, Z_r$ be i.i.d. $N(0, 1)$ random variables, then the following results hold:*

(a) *If $\mu_1 = \mu_2$, then*

$$\frac{1}{\tau p^\beta} T_{CQ} \xrightarrow{w} \sum_{i=1}^{r} \omega_i Z_i^2 - \sum_{i=1}^{r} \omega_i.$$

(b) *If $(\tau p^\beta)^{-1/2} \big(\mathbf{V}^T(\mu_1 - \mu_2)\big)_i \to \zeta_i \in (-\infty, +\infty)$, $i = 1, \dots, r$, and $(\tau p^\beta)^{-1} \|\tilde{\mathbf{V}}^T(\mu_1 - \mu_2)\|^2 \to \zeta^* \in [0, +\infty)$, then*

$$\frac{1}{\tau p^\beta} T_{CQ} \xrightarrow{w} \sum_{i=1}^{r} (\sqrt{\omega_i} Z_i + \zeta_i)^2 + \zeta^* - \sum_{i=1}^{r} \omega_i.$$

Remark 1. By the definitions of ζ_i and ζ^*, we have

$$\frac{1}{\tau p^\beta} \|\mu_1 - \mu_2\|^2 = \frac{1}{\tau p^\beta} \|\mathbf{V}^T(\mu_1 - \mu_2)\|^2 + \frac{1}{\tau p^\beta} \|\tilde{\mathbf{V}}^T(\mu_1 - \mu_2)\|^2 \to \sum_{i=1}^{r} \zeta_i^2 + \zeta^*.$$

Thus, $\sum_{i=1}^{r} \zeta_i^2$ and ζ^* characterize the signal strength in the principal space and the complement of the principal space, respectively. Under the conditions of Theorem 1 or Theorem 2, the following statements are equivalent:

(1) $\zeta_1 = \cdots = \zeta_r = \zeta^* = 0$.

(2) $\|\mu_1 - \mu_2\|^2 = o(\tau p^\beta)$.

(3) The asymptotic distributions of $(\tau p^\beta)^{-1} T_{CQ}$ are the same under the null hypothesis and the alternative hypothesis.

(4) Any test procedure based on T_{CQ} has trivial power asymptotically.

It is implied by Theorem 1 and Theorem 2 that the original critical value of T_{CQ} can not be used when $\beta \geq 1/2$. Now we adjust the critical value of T_{CQ} such that the resulting test has correct level asymptotically. Consider the random variable $W = \sqrt{2p}\sigma^2 Z_0 + \sum_{i=1}^{r} \boldsymbol{\lambda}_i Z_i^2 - \sum_{i=1}^{r} \boldsymbol{\lambda}_i$, where $Z_0, Z_1, \dots, Z_r$ are i.i.d. $N(0, 1)$ random variables. Let $F(x; \boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_r, \sigma^2)$ be the cumulative distribution function of W. Under the conditions of Theorem 1, we have

$$\frac{W}{p^\beta} \xrightarrow{w} \sqrt{2}\sigma^2 Z_0 + \sum_{i=1}^{r} \omega_i Z_i^2 - \sum_{i=1}^{r} \omega_i.$$

Under the conditions of Theorem 2, we have

$$\frac{W}{p^\beta} \xrightarrow{w} \sum_{i=1}^r \omega_i Z_i^2 - \sum_{i=1}^r \omega_i.$$

Hence in both case, we have

$$\sup_{x\in\mathbb{R}} \big| \Pr\big(\frac{1}{\tau} T_{CQ} \le x\big) - \Pr\big(W \le x\big)\big| = o(1).$$

Thus, if we reject the null hypothesis when

$$\frac{1}{\tau} T_{CQ} > F^{-1}(1-\alpha; \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r, \sigma^2),$$

then the resulting test has level α asymptotically for $\beta \ge 1/2$. However, the distribution $F(x; \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r, \sigma^2)$ involves some unknown parameters. In order to consistently estimate $F(x; \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r, \sigma^2)$, we need to give ratio consistent estimators of $\boldsymbol{\lambda}_i$, $i = 1, \ldots, r$, and σ^2. The following proposition shows that $\boldsymbol{\lambda}_i(S)$ can consistently estimate $\boldsymbol{\lambda}_i$, $i = 1, \ldots, r$.

Proposition 1. *Under Assumption 1, suppose $p^{1-\beta} = o(n)$, then*

$$\frac{\lambda_i(\mathbf{S})}{\boldsymbol{\lambda}_i} \xrightarrow{P} 1, \qquad i = 1, \ldots, r,$$

where "$\xrightarrow{P}$" means convergence in probability.

Remark 2. Proposition (1) requires the condition $p^{1-\beta} = o(n)$. If $\beta = 1/2$, this condition becomes $p/n^2 \to 0$. If $\beta \ge 1$, this condition is trivially fulfilled.

In section 3, we will give an estimator $\hat{\sigma}^2$ of σ^2.Proposition 3 asserts that $\hat{\sigma}^2$ is consistent. Now we propose a corrected T_{CQ} test procedure which reject the null hypothesis if

$$\tau^{-1} T_{CQ} > F^{-1}(1-\alpha; \hat{\boldsymbol{\lambda}}_1, \ldots, \hat{\boldsymbol{\lambda}}_r, \hat{\sigma}^2).$$

Then under the conditions of Proposition 1 and the conditions of either Theorem 1 or Theorem 2, the corrected T_{CQ} test procedure has level α asymptotically.

As we have seen in Remark 1, the corrected T_{CQ} test procedure has trivial power if $\|\mu_1 - \mu_2\|^2 = o(\tau p^\beta)$. Then as β increases, the corrected T_{CQ} test procedure becomes less powerful. This implies that the power of the corrected T_{CQ} test procedure is negatively affected by the large eigenvalues of $\mathbf{\Sigma}$.

3. A projection test

In section 2, we adjusted the critical value of T_{CQ} under Assumption 1. However, the power of the corrected T_{CQ} test procedure is negatively affected by the large eigenvalues of $\mathbf{\Sigma}$. This motivates us to propose a new test for hypotheses (1) under Assumption 1.

Recently, a class of test procedures have been proposed through random projection to lower dimensional space. See, for example, Lopes et al. (2011), Thulin (2014) and Srivastava et al. (2016). 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. Then, is there an optimal projection which is nonrandomized?

For $\mathbf{O} \in \mathbb{O}_{p\times k}$ $(k \le p)$, define statistic

$$T(\mathbf{O}) = \|\mathbf{O}^T(\bar{X}_1 - \bar{X}_2)\|^2 - \frac{1}{n_1}\operatorname{tr}(\mathbf{O}^T\mathbf{S}_1\mathbf{O}) - \frac{1}{n_2}\operatorname{tr}(\mathbf{O}^T\mathbf{S}_2\mathbf{O}).$$

Then $T(\mathbf{O})$ is Chen and Qin (2010)'s statistic on the transformed data $\mathbf{O}^T X_{k,i}$. Denote by $\Phi(\cdot)$ the cumulative distribution function of the standard normal random variable. Under the condition (2), Chen and Qin (2010) proved that the asymptotic power of T_{CQ} under the local alternative is

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

Hence the power of T_{CQ} is largely impacted by $\|\mu_1 - \mu_2\|^2/\sqrt{2\tau^2 \operatorname{tr}(\mathbf{\Sigma}^2)}$, which may be viewed as a signal to noise ratio (SNR). Consequently, $\|\mathbf{O}^T(\mu_1 - \mu_2)\|^2/\sqrt{2\tau^2 \operatorname{tr}(\mathbf{O}^T\mathbf{\Sigma}^2\mathbf{O})}$ measures the power of $T(\mathbf{O})$. To consider an average-case scenario, like Lopes et al. (2011), we temporarily place a prior on $\mu_1 - \mu_2$ and assume that $\mu_1 - \mu_2$ is from the uniform distribution on the unit sphere. In this case, an average SNR can be defined as

$$\mathrm{E}\left(\frac{\|\mathbf{O}^T(\mu_1 - \mu_2)\|^2}{\sqrt{2\tau^2 \operatorname{tr}(\mathbf{O}^T\mathbf{\Sigma}^2\mathbf{O})}}\right) = \frac{k/p}{\sqrt{2\tau^2 \operatorname{tr}(\mathbf{O}^T\mathbf{\Sigma}^2\mathbf{O})}}. \tag{6}$$

It can be expected that the $T(\mathbf{O})$ maximizing the average SNR has the best average power behavior among $\{T(\mathbf{O}) : \mathbf{O} \in \mathbb{O}_{p\times k},\ k \le p\}$.

Note that for fixed k, (6) is maximized when the columns of $\mathbf{O}$ are equal to the last k eigenvectors of $\mathbf{\Sigma}$. Thus, it remains to maximize

$$\frac{k/p}{\sqrt{2\tau^2 \sum_{i=p-k+1}^{p} \lambda_i^2(\mathbf{\Sigma})}} \tag{7}$$

over k. If $k \leq p-r$, (7) is equal to $\sqrt{k}/(\sqrt{2}\sigma^2\tau p)$ which is an increasing function of k. If $k > p-r$, we have

$$\begin{aligned}
&\frac{k/p}{\sqrt{2\tau^2(\sum_{i=p-k+1}^{r} \lambda_i^2(\mathbf{\Sigma}) + k\sigma^4)}} \leq \frac{1}{\sqrt{2\tau^2(\kappa^{-2}p^{2\beta} + (p-r)\sigma^4)}} \\
=&\frac{p/(p-r)}{\sqrt{\kappa^{-2}p^{2\beta}/((p-r)\sigma^4) + 1}} \frac{(p-r)/p}{\sqrt{2\tau^2((p-r)\sigma^4)}}.
\end{aligned}$$

Hence for sufficiently large p, we have

$$\frac{k/p}{\sqrt{2\tau^2(\sum_{i=p-k+1}^{r} \lambda_i^2(\mathbf{\Sigma}) + k\sigma^4)}} < \frac{(p-r)/p}{\sqrt{2\tau^2((p-r)\sigma^4)}}.$$

Thus, for sufficiently large p, (7) is maximized when $k = p - r$.

As we have shown, (6) is maximized when $\mathbf{O} = \tilde{V}$. We define the following variable

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

Note that based on $\tilde{\mathbf{V}}^T X_{ki}$, the likelihood ratio test statistic for hypothesis (1) is $\|\tilde{\mathbf{V}}^T(\bar{X}_1 - \bar{X}_2)\|^2$. In this view, T_1 can be regarded as a restricted likelihood ratio statistic. It can be shown that T_1 is asymptotically normal.

Proposition 2. *Under Assumption 1, suppose $\frac{n}{p}\|\mu_1 - \mu_2\|^2 = o(1)$, we have*

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

Note that T_1 relies on the subspace $\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T$ which is unknown. Thus, we estimate $\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T$ by its sample counterpart. We denote by $\hat{\mathbf{V}}$ and $\hat{\tilde{\mathbf{V}}}$ the first r and last $p-r$ eigenvectors of S respectively. Similarly, we denote by $\hat{\mathbf{V}}_k$ and $\hat{\tilde{\mathbf{V}}}_k$ the first r and last $p-r$ eigenvectors of S_k respectively, $k = 1, 2$. As the main part of T_1, $\|\tilde{\mathbf{V}}^T(\bar{X}_1 - \bar{X}_2)\|^2$ can be directly estimated by $\|\hat{\tilde{\mathbf{V}}}^T(\bar{X}_1 - \bar{X}_2)\|^2$. While $n_k^{-1}\mathrm{tr}(\tilde{\mathbf{V}}^T S_k \tilde{\mathbf{V}})$ can be estimated by $n_k^{-1}\mathrm{tr}(\hat{\tilde{\mathbf{V}}}_k^T S_k \hat{\tilde{\mathbf{V}}}_k)$, $k = 1, 2$. Define

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

The asymptotic property of T_2 is closely related to the consistency rate of $\hat{\tilde{\mathbf{V}}}\hat{\tilde{\mathbf{V}}}^T$ as an estimator of $\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T$. However, $\hat{\tilde{\mathbf{V}}}\hat{\tilde{\mathbf{V}}}^T$ can not always consistently estimate $\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T$ in high dimensional setting. In fact, Cai et al. (2013)'s Theorem 5 implies that it is possible only when $p^{1-\beta}/n \to 0$, see Lemma 5 in appendix. The asymptotic normality of T_2 requires a stronger condition.

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

The following theorem establishes the asymptotic normality of T_2.

Theorem 3. *Under Assumptions 1 and 2, suppose*

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

we have

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

The proof of Theorem 3 implies that the conclusion of Theorem 3 does not hold if Assumption 2 is violated.

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 $\sigma^2 = (p-r)^{-1}\sum_{i=r+1}^{p}\lambda_i(\mathbf{\Sigma})$, where $\lambda_i(\mathbf{\Sigma})$ is the ith largest eigenvalue of $\mathbf{\Sigma}$. So σ^2 can be estimated by

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

Using Weyl's inequality, we can derive the consistency rate of $\hat{\sigma}^2$.

Proposition 3. *Under Assumptions 1, we have*

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

Now we propose our new test statistic as

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

By Theorem 3 and Proposition 3, Q is asymptotically distributed as $N(0,1)$ under the null hypothesis. Thus, we reject the null hypothesis when Q is larger than the upper α quantile of $N(0,1)$. The asymptotic power function of the new test can be obtained immediately.

Corollary 1. *Under the conditions of Theorem 3, the asymptotic power function of the new test is*

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

In Section 2, we have seen that the test procedure T_{CQ} has trivial power if $\|\mu_1 - \mu_2\|^2 = o(\tau p^\beta)$. Corollary 1 implies that the asymptotic power function of the new test is not affected by β. As a result, when $\beta > 1/2$, the new test tends to be much more powerful.

4. Numerical studies

4.1. Simulation results

In this section, we consider the simulation performance of the proposed test and compare it with several other tests, including the tests in Chen and Qin (2010). These tests are denoted respectively by CQ in the rest of this section. The data generation mechanism is as follow. We randomly choose a $\mathbf{U} \in \mathbb{O}_{p\times p}$ from Haar invariant distribution. Let d_i equal to p^β plus a random error from $U(0,1)$ (Uniform distribution between 0 and 1), $i = 1, \ldots, r$. Construct $p \times p$ diagonal matrix $\mathbf{D} = \mathrm{diag}(\sqrt{d_1}, \ldots, \sqrt{d_r}, 1, \ldots, 1)$. Then, we independently generate data by the formula

$$X_{k,i} = \mu_k + \mathbf{UD}Y_{k,i} \quad i = 1, \ldots, n_k \text{ and } k = 1, 2,$$

where $Y_{k,i}$ is a p dimensional random vector whose entries are i.i.d. random variables with common distribution F. We will consider three different distributions of F.

- Normal: $F \sim N(0,1)$.
- Chi-squared: $F \sim (\chi_4^2 - 4)/\sqrt{8}$, where χ_4^2 is a chi-squared random variable with degree of freedom 4.
- Student's t: $F \sim t_4/\sqrt{2}$, where t_4 is a Student's t random variable with degree of freedom 4.

We take nominal level $\alpha = 0.05$.

First, we simulate the level of the new test. We set factor number $r = 1$. Samples are repeatedly generated 2000 times to calculate empirical level. For comparison, we also give the corresponding 'oracle' level which is calculated by variable $T_1/(\sigma^2\sqrt{2p\tau^2})$. The result is listed in Table 3. Level of the new test is a little inflated compared with oracle level.

Next, we simulate the empirical power of the new test. The results in Section 2 have showed that the level of the Chen and Qin (2010)'s test can't be guaranteed when $\beta \geq 1/2$. To be fair, critical values are all determined by permutation method. We permute the sample 100 times to determine the critical value. The test procedure is repeated 500 times to obtain empirical power. We plot the empirical power versus signal-to-noise ratio (SNR) which is defined as SNR $= \|\mu_1 - \mu_2\|^2/(\sigma^2\sqrt{2\tau^2 p})$. The results are illustrated in figure 1, where 'NEW', 'CQ' and 'SD' represent the new test, Chen and Qin (2010)'s test and Srivastava and Du (2008)'s test respectively. From the results, we can find that when $\mathbf{\Sigma}$ is spiked, the new test outperforms T_{CQ} substantially; when $\mathbf{\Sigma}$ is not spiked, all three tests have similar performance.

4.2. 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

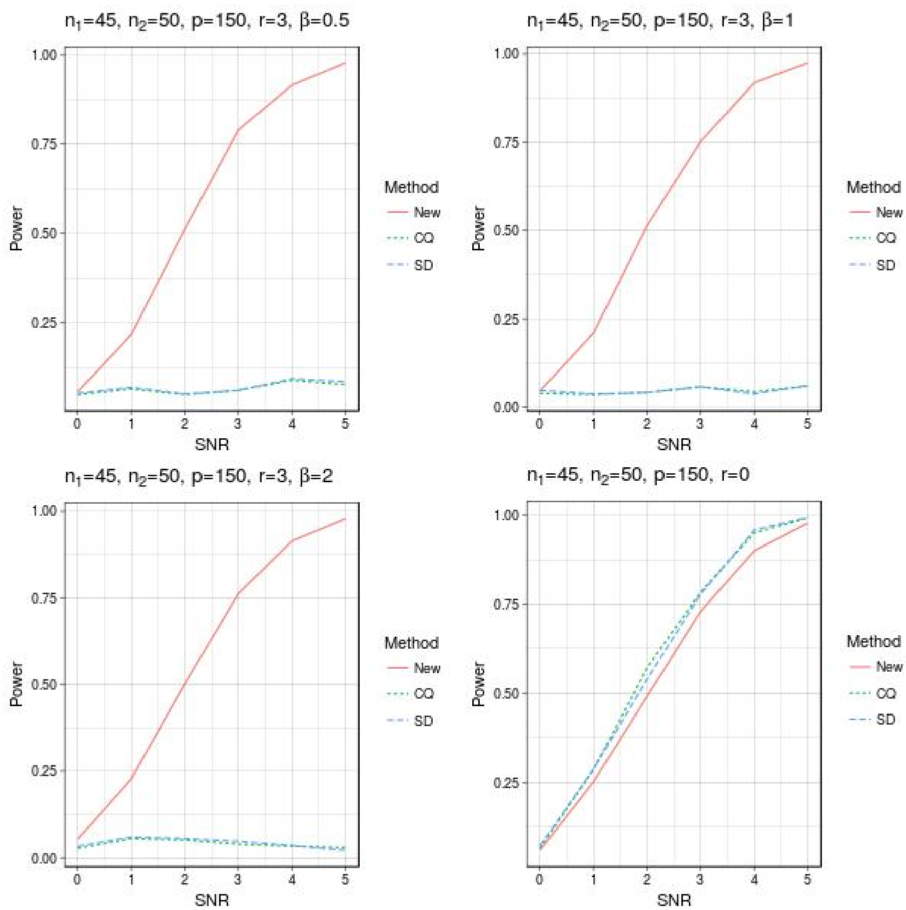

Figure 1: Empirical power simulation.

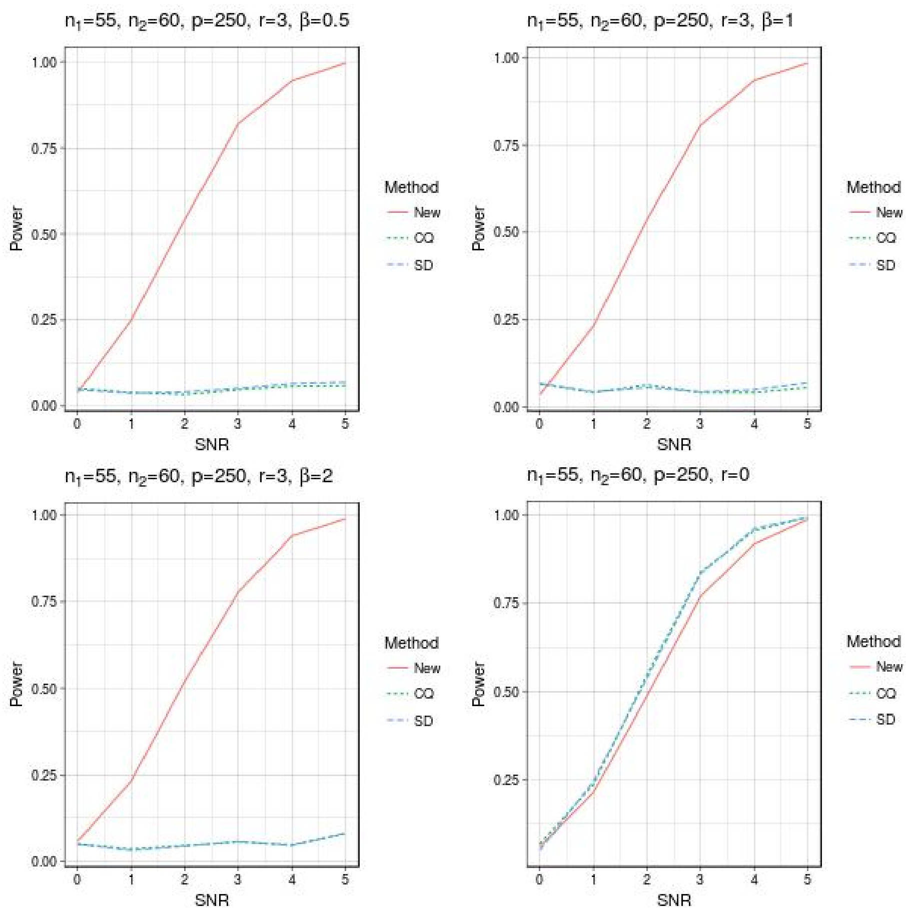

Figure 2: Empirical power simulation.

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 $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$. 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).

5. 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 derived the asymptotic distribution of Chen and Qin (2010)'s test statistic. To reduce the variance of T_{CQ}, we dropped big variance terms 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.

In our paper, we have assumed r is known. If r is an unknown positive number, a consistent estimator of r is

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

where R is a hyperparameter. See Ahn and Horenstein (2013) for detail.

The asymptotic normality of the new test statistic relies on the assumption $\sqrt{p}/n \to 0$. In the situation of small n 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.

Non normality

Appendix

Lemma 2 (Weyl's inequality)**.** *Let H and P be two symmetric $n \times n$ matrices and $M = H + P$. If $r + s - 1 \le i \le j + k - n$, we have*

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

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

Lemma 3 (Cai et al. (2015), Proposition 1)**.** *Let A_1 and A_2 be $p \times p$ symmetric matrices. Let $r < p$ be arbitrary and let $\mathbf{V}_1, \mathbf{V}_2 \in \mathbb{O}_{p,r}$ be formed by the r leading singular vectors of A_1 and A_2, respectively. Then*

$$\|A_1 - A_2\| \ge \frac{1}{2}(\lambda_r(A_1) - \lambda_{r+1}(A_2))\|\mathbf{V}_1\mathbf{V}_1^T - \mathbf{V}_2\mathbf{V}_2^T\|.$$

Lemma 4 (Davidson and Szarek (2001), Theorem II.7)**.** *Let Z be a $p \times n$ random matrix with i.i.d. $N(0,1)$ entries. Then for any $t > 0$,*

$$\Pr(\sqrt{\lambda_1(ZZ^T)} > \sqrt{n} + \sqrt{p} + t) \le e^{-t^2/2},$$
$$\Pr(\sqrt{\lambda_{\min(n,p)}(ZZ^T)} < \sqrt{n} - \sqrt{p} - t) \le e^{-t^2/2}.$$

We give two useful corollaries of Lemma 4.

Corollary 2. *Suppose that W_n is a $p \times p$ random matrix distributed as* $\mathrm{Wishart}_p(n, \mathbf{I}_p)$, *the p dimensional Wishart distribution with parameter Ψ and m degrees of freedom. Then as $n, p \to \infty$,*

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

Proof. The result follows from the inequality

$$\Pr\Big(\frac{\lambda_1(W_n)}{\max(n,p)} > 16\Big) \le \Pr\Big(\lambda_1(W_n) > 8(n+p)\Big) \le \Pr\Big(\lambda_1(W_n) > 4(\sqrt{n}+\sqrt{p})^2\Big)$$
$$= \Pr\Big(\sqrt{\lambda_1(W_n)} > 2(\sqrt{n}+\sqrt{p})\Big) \le \Pr\Big(\sqrt{\lambda_1(W_n)} > 2\sqrt{n}+\sqrt{p}\Big) \le e^{-n/2},$$

where the last inequality follows from Lemma 4 with $t=\sqrt{n}$. $\square$

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

$$\|\frac{1}{n}W_n - \mathbf{I}_p\| = O_P\Big(\max\big(\sqrt{\frac{p}{n}},\frac{p}{n}\big)\Big).$$

Proof. Since the eigenvalues of $\frac{1}{n}W_n - \mathbf{I}_p$ are $\frac{1}{n}\lambda_1(W_n)-1 \ge \cdots \ge \frac{1}{n}\lambda_p(W_n)-1$, we have

$$\|\frac{1}{n}W_n - \mathbf{I}_p\| = \max\Big(\frac{1}{n}\lambda_1(W_n)-1, 1-\frac{1}{n}\lambda_p(W_n)\Big).$$

This, combined with union bound, yields

$$\Pr\Big(\|\frac{1}{n}W_n-\mathbf{I}_p\| > 4\big(\sqrt{\frac{p}{n}}+\frac{p}{n}\big)\Big) \le \Pr\Big(\lambda_1(W_n) > \big(\sqrt{n}+2\sqrt{p}\big)^2\Big)+\Pr\Big(\lambda_p(W_n) < n-4\sqrt{np}-4p\Big).$$

The first term can be bounded by Lemma 4 with $t=\sqrt{p}$.

$$\Pr\Big(\lambda_1(W_n) > \big(\sqrt{n}+2\sqrt{p}\big)^2\Big) = \Pr\Big(\sqrt{\lambda_1(W_n)} > \sqrt{n}+2\sqrt{p}\Big) \le e^{-p^2/2}.$$

We now show that the second term is also bounded by $e^{-p^2/2}$. To see this, note that If $p > n/4$, then $n-4\sqrt{np}-4p \le n-4p < 0$. In this case, $\Pr\Big(\lambda_p(W_n) < n-4\sqrt{np}-4p\Big) = 0$. If $p\le n/4$, we have

$$\Pr\Big(\lambda_p(W_n) < n-4\sqrt{np}-4p\Big) \le \Pr\Big(\lambda_p(W_n) < n-4\sqrt{np}+4p\Big)$$
$$= \Pr\Big(\sqrt{\lambda_p(W_n)} < \sqrt{n}-\sqrt{2p}\Big) \le e^{-p^2/2},$$

where the last inequality follows from Lemma 4 with $t=\sqrt{p}$.

Now we have the bound

$$\Pr\Big(\|\frac{1}{n}W_n - \mathbf{I}_p\| > 4\big(\sqrt{\frac{p}{n}}+\frac{p}{n}\big)\Big) \le 2e^{-p^2/2}.$$

Then

$$\|\frac{1}{n}W_n - \mathbf{I}_p\| = O_P\Big(\sqrt{\frac{p}{n}}+\frac{p}{n}\Big) = O_P\Big(\max\big(\sqrt{\frac{p}{n}},\frac{p}{n}\big)\Big).$$

$\square$

Proof of Lemma 1. By a standard orthogonal transformation, we can write

$$\frac{Y_n^T \mathbf{A}_n Y_n - \mathrm{E}\, Y_n^T \mathbf{A}_n Y_n}{[\mathrm{Var}(Y_n^T \mathbf{A}_n Y_n)]^{1/2}} = \sum_{i=1}^{k_n} \frac{\lambda_i(\mathbf{A}_n)}{\big[2\,\mathrm{tr}(\mathbf{A}_n^2)\big]^{1/2}} (Z_{ni}^2 - 1), \tag{9}$$

where $Z_{n1}, \ldots, Z_{nk_n}$ are independent standard normal random variables.

If 5 holds, then

$$\begin{aligned}
&\sum_{i=1}^{k_n} \mathrm{E}\Big[\frac{\lambda_i^2(\mathbf{A}_n)}{2\mathrm{tr}(\mathbf{A}_n^2)}(Z_{ni}^2-1)^2\Big\{\frac{\lambda_i^2(\mathbf{A}_n)}{2\mathrm{tr}(\mathbf{A}_n^2)}(Z_{ni}^2-1)^2 \geq \epsilon\Big\}\Big] \\
\leq &\sum_{i=1}^{k_n} \frac{\lambda_i^2(\mathbf{A}_n)}{2\,\mathrm{tr}(\mathbf{A}_n^2)}\, \mathrm{E}\Big[(Z_{n1}^2-1)^2\Big\{\frac{\lambda_1(\mathbf{A}_n^2)}{2\,\mathrm{tr}(\mathbf{A}_n^2)}(Z_{n1}^2-1)^2 \geq \epsilon\Big\}\Big] \\
= &\frac{1}{2}\,\mathrm{E}\Big[(Z_{n1}^2-1)^2\Big\{\frac{\lambda_1(\mathbf{A}_n^2)}{2\,\mathrm{tr}(\mathbf{A}_n^2)}(Z_{n1}^2-1)^2 \geq \epsilon\Big\}\Big] \to 0.
\end{aligned}$$

Hence 4 follows by Lindeberg's central limit theorem.

Conversely, if 4 holds, we will prove that there is a subsequence of $\{n\}$ along which 5 holds. Then 5 follows by a standard contradiction argument.

Denote $c_{ni} = \lambda_i(\mathbf{A}_n)/\big[2\,\mathrm{tr}(\mathbf{A}_n^2)\big]^{1/2}$, $i = 1, \ldots, k_n$. Since 4 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. Denote by $\log z$ ($z \in \mathbb{C}$) the principal branch of the complex logarithm. For $t \in (-1/2, 1/2)$, we have

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

where the second equality holds since $0 \leq c_{ni} \leq \sqrt{2}/2$ by definition. Let $b_{nl} = \sum_{j=1}^{k_n} (c_{nj})^l$, $n = 1, 2, \cdots$ and $l = 3, 4, \cdots$. Note that for $l \geq 3$, we have

$$|b_{nl}| = \big|\sum_{j=1}^{k_n}(c_{nj})^l\big| \leq \big|\sum_{j=1}^{k_n}(c_{nj})^2\big| = 1/2.$$

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. For this subsequence, applying dominated convergence theorem yields

$$\mathrm{E}\left[\exp\left(it\sum_{j=1}^{k_n}c_{nj}(Z_{nj}^2-1)\right)\right]\to\exp\Big(-\frac{1}{2}t^2+\frac{1}{2}\sum_{l=3}^{+\infty}b_l\frac{1}{l}(i2t)^l\Big),\quad t\in\big(-\frac{1}{2},\frac{1}{2}\big).$$

But the left hand side converges to $\exp(-t^2/2)$. It follows that

$$-\frac{1}{2}t^2+\frac{1}{2}\sum_{l=3}^{+\infty}b_l\frac{1}{l}(i2t)^l=-\frac{1}{2}t^2+2\pi mi,\quad t\in\big(-\frac{1}{2},\frac{1}{2}\big),$$

for some integer m. By the uniqueness of power series, we mulst have $m=0$ and $b_l=0$ for $l\geq 3$. Then 5 follows by noting that $b_{n4}\geq\max_j(c_{nj})^4$. $\square$

Proves of Theorem 1 and Theorem 2. In both Theorem 1 and Theorem 2, (a) is a corrolary of (b). We shall prove (b) of Theorem 1 and Theorem 2 simultaneously.

Since $(n_k-1)\mathbf{S}_k\sim\mathrm{Wishart}_p(n_k-1,\mathbf{\Sigma})$, $k=1,2$, we have

$$\mathrm{E}\Big(\frac{1}{n_1}\operatorname{tr}\mathbf{S}_1+\frac{1}{n_2}\operatorname{tr}\mathbf{S}_2\Big)=\tau\operatorname{tr}\mathbf{\Sigma},$$

and

$$\begin{aligned}&\mathrm{Var}\Big(\frac{1}{n_1}\operatorname{tr}\mathbf{S}_1+\frac{1}{n_2}\operatorname{tr}\mathbf{S}_2\Big)=\Big(\frac{2}{n_1^2(n_1-1)}+\frac{2}{n_2^2(n_2-1)}\Big)\operatorname{tr}\mathbf{\Sigma}^2\\=&O\Big(\frac{1}{n^3}(p^{2\beta}+p)\Big)=O\Big(\frac{p^{2\beta}}{n^3}\Big).\end{aligned}$$

It follows that

$$\begin{aligned}&\frac{1}{n_1}\operatorname{tr}\mathbf{S}_1+\frac{1}{n_2}\operatorname{tr}\mathbf{S}_2=\tau\operatorname{tr}\mathbf{\Sigma}+O_P\Big(\frac{1}{n\sqrt{n}}p^\beta\Big)\\=&\tau\sum_{i=1}^r(\boldsymbol{\lambda}_i+\sigma^2)+\tau(p-r)\sigma^2+O_P\Big(\frac{1}{n\sqrt{n}}p^\beta\Big)\\=&\tau p^\beta\sum_{i=1}^r\omega_i+\tau(p-r)\sigma^2+o_P\Big(\frac{1}{n}p^\beta\Big).\end{aligned}$$

Thus,

$$\frac{1}{\tau p^\beta}\big(\frac{1}{n_1}\operatorname{tr}\mathbf{S}_1+\frac{1}{n_2}\operatorname{tr}\mathbf{S}_2\big)=\sum_{i=1}^r\omega_i+p^{1-\beta}\sigma^2+o_P(1).\tag{10}$$

Next we deal with $\|\bar{X}_1-\bar{X}_2\|^2$. Note that we have

$$\|\bar{X}_1-\bar{X}_2\|^2=\|\mathbf{V}^T(\bar{X}_1-\bar{X}_2)\|^2+\|\tilde{\mathbf{V}}^T(\bar{X}_1-\bar{X}_2)\|^2.$$

These two terms are independent. For the first term, note that $\mathbf{V}^T(\bar{X}_1 - \bar{X}_2) \sim N_r\big(\mathbf{V}^T(\mu_1 - \mu_2), \tau(\mathbf{\Lambda} + \sigma^2\mathbf{I}_r)\big)$, we have

$$\begin{aligned}\|\mathbf{V}^T(\bar{X}_1 - \bar{X}_2)\|^2 &\sim \sum_{i=1}^{r}\Big(\sqrt{\tau(\lambda_i + \sigma^2)}Z_i + \big(\mathbf{V}^T(\mu_1 - \mu_2)\big)_i\Big)^2 \\ &= \tau p^{\beta}\sum_{i=1}^{r}\Big(\sqrt{p^{-\beta}(\lambda_i + \sigma^2)}Z_i + \frac{1}{\sqrt{\tau p^{\beta}}}\big(\mathbf{V}^T(\mu_1 - \mu_2)\big)_i\Big)^2.\end{aligned}$$

By the assumptions of the theorem, we have that

$$\frac{1}{\tau p^{\beta}}\|\mathbf{V}^T(\bar{X}_1 - \bar{X}_2)\|^2 \xrightarrow{w} \sum_{i=1}^{r}(\sqrt{\omega_i}Z_i + \zeta_i)^2. \tag{11}$$

As for $\|\tilde{\mathbf{V}}^T(\bar{X}_1 - \bar{X}_2)\|^2$, we have that

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

Since $\tilde{\mathbf{V}}^T(\bar{X}_1 - \bar{X}_2) \sim N_{p-r}(\tilde{\mathbf{V}}^T(\mu_1 - \mu_2), \sigma^2\tau\mathbf{I}_{p-r})$, by central limit theorem, we have

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

For the intersection term, we have

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

It follows that

$$\frac{1}{\tau p^{\beta}}\big(\big\|\tilde{\mathbf{V}}^T(\bar{X}_1 - \bar{X}_2)\big\|^2 - \sigma^2\tau(p-r) - \big\|\tilde{\mathbf{V}}^T(\mu_1 - \mu_2)\big\|^2\big) \xrightarrow{\mathcal{L}} \sqrt{2}\sigma^2\delta_{\{\frac{1}{2}\}}(\beta)Z_0, \tag{12}$$

where $\delta_{\frac{1}{2}}(\beta)$ equals 1 if $\beta = 1/2$ and equals 0 otherwise.

Combining (10) (11) and (12) leads to

$$
\begin{aligned}
&\frac{1}{\tau p^\beta} T_{CQ} = \frac{1}{\tau p^\beta}\big(\|\bar{X}_1 - \bar{X}_2\|^2 - \frac{1}{n_1}\operatorname{tr}\mathbf{S}_1 - \frac{1}{n_2}\operatorname{tr}\mathbf{S}_2\big)\\
=&\frac{1}{\tau p^\beta}\|\mathbf{V}^T(\bar{X}_1 - \bar{X}_2)\|^2 + \frac{1}{\tau p^\beta}\big(\|\tilde{\mathbf{V}}^T(\bar{X}_1 - \bar{X}_2)\|^2 - \sigma^2\tau(p-r) - \|\tilde{\mathbf{V}}^T(\mu_1 - \mu_2)\|^2\big)\\
&-\frac{1}{\tau p^\beta}\Big(\frac{1}{n_1}\operatorname{tr}\mathbf{S}_1 + \frac{1}{n_2}\operatorname{tr}\mathbf{S}_2\Big) + \frac{\sigma^2(p-r)}{p^\beta} + \frac{1}{\tau p^\beta}\|\tilde{\mathbf{V}}^T(\mu_1 - \mu_2)\|^2\\
=&\sum_{i=1}^r(\sqrt{\omega_i}Z_i + \zeta_i)^2 + \sqrt{2}\sigma^2\delta_{\{\frac{1}{2}\}}(\beta)Z_0 - (\sum_{i=1}^r\omega_i + p^{1-\beta}\sigma^2) + \frac{\sigma^2(p-r)}{p^\beta} + \zeta^* + o_P(1)\\
\xrightarrow{\mathcal{L}}&\sum_{i=1}^r(\sqrt{\omega_i}Z_i + \zeta_i)^2 + \zeta^* + \sqrt{2}\sigma^2\delta_{\{\frac{1}{2}\}}(\beta)Z_0 - \sum_{i=1}^r\omega_i.
\end{aligned}
$$

This implies the conclusions of Theorem 1 and Theorem 2.

□

Proof of Proposition 1. Let $\mathbf{\Sigma} = \mathbf{U}\mathbf{E}\mathbf{U}^T$ denote the spectral decomposition of $\mathbf{\Sigma}$, where $\mathbf{U} = (\mathbf{V}, \tilde{\mathbf{V}})$ and $\mathbf{E} = \operatorname{diag}(\boldsymbol{\lambda}_1 + \sigma^2, \ldots, \boldsymbol{\lambda}_r + \sigma^2, \sigma^2, \ldots, \sigma^2)$. Denote by $\mathbf{S} = \hat{\mathbf{U}}\hat{\mathbf{E}}\hat{\mathbf{U}}^T$ the spectral decomposition of $\mathbf{S}$, where $\hat{\mathbf{U}} = (\hat{\mathbf{V}}, \hat{\tilde{\mathbf{V}}})$ and $\hat{\mathbf{E}} = \operatorname{diag}(\hat{\boldsymbol{\lambda}}_1, \ldots, \hat{\boldsymbol{\lambda}}_p)$. Let $\mathbf{Z}$ be a $p \times (n-2)$ random matrix with i.i.d. $N(0,1)$ entries. Denote $\mathbf{Z} = (\mathbf{Z}_{(1)}^T, \mathbf{Z}_{(2)}^T)^T$, where $\mathbf{Z}_{(1)}$ and $\mathbf{Z}_{(2)}$ are the first r rows and last $p-r$ rows of $\mathbf{Z}$.

The sample covariance matrix S has the same distribution as $(n-2)^{-1}\mathbf{U}\mathbf{E}^{1/2}\mathbf{Z}\mathbf{Z}^T\mathbf{E}^{1/2}\mathbf{U}^T$. This implies that $\hat{\lambda}_i = \lambda_i(S) \sim (n-2)^{-1}\lambda_i(\mathbf{Z}^T\mathbf{E}\mathbf{Z})$, $i = 1, \ldots, r$. Hence we only need to deal with the asymptotic property of $(n-2)^{-1}\lambda_i(\mathbf{Z}^T\mathbf{E}\mathbf{Z})$. For $i = 1, \ldots, r$, we have

$$
\begin{aligned}
&|\lambda_i(\mathbf{Z}^T\mathbf{E}\mathbf{Z}) - (n-2)(\lambda_i + \sigma^2)|\\
\le&|\lambda_i(\mathbf{Z}^T\mathbf{E}\mathbf{Z}) - \lambda_i\big(\mathbf{Z}_{(1)}^T(\mathbf{\Lambda} + \sigma^2\mathbf{I}_r)\mathbf{Z}_{(1)}\big)| + |\lambda_i\big(\mathbf{Z}_{(1)}^T(\mathbf{\Lambda} + \sigma^2\mathbf{I}_r)\mathbf{Z}_{(1)}\big) - (n-2)(\lambda_i + \sigma^2)|
\end{aligned}
$$

By the equality $\mathbf{Z}^T\mathbf{E}\mathbf{Z} = \mathbf{Z}_{(1)}^T(\mathbf{\Lambda} + \sigma^2\mathbf{I}_r)\mathbf{Z}_{(1)} + \sigma^2\mathbf{Z}_{(2)}^T\mathbf{Z}_{(2)}$ and Weyl's inequality, the first term satisfies

$$
|\lambda_i(\mathbf{Z}^T\mathbf{E}\mathbf{Z}) - \lambda_i\big(\mathbf{Z}_{(1)}^T(\mathbf{\Lambda} + \sigma^2\mathbf{I}_r)\mathbf{Z}_{(1)}\big)| \le \|\mathbf{Z}^T\mathbf{E}\mathbf{Z} - \mathbf{Z}_{(1)}^T(\mathbf{\Lambda} + \sigma^2\mathbf{I}_r)\mathbf{Z}_{(1)}\| = \sigma^2\|\mathbf{Z}_{(2)}^T\mathbf{Z}_{(2)}\|
$$

For the second term, we have

$$
\begin{aligned}
&|\lambda_i\big(\mathbf{Z}_{(1)}^T(\mathbf{\Lambda}+\sigma^2\mathbf{I}_r)\mathbf{Z}_{(1)}\big)-(n-2)(\lambda_i+\sigma^2)| \\
=&\big|\lambda_i\big((\mathbf{\Lambda}+\sigma^2\mathbf{I}_r)^{1/2}\mathbf{Z}_{(1)}\mathbf{Z}_{(1)}^T(\mathbf{\Lambda}+\sigma^2\mathbf{I}_r)^{1/2}\big)-\lambda_i\big((n-2)(\mathbf{\Lambda}+\sigma^2\mathbf{I}_r)\big)\big| \\
\le&\|(\mathbf{\Lambda}+\sigma^2\mathbf{I}_r)^{1/2}\mathbf{Z}_{(1)}\mathbf{Z}_{(1)}^T(\mathbf{\Lambda}+\sigma^2\mathbf{I}_r)^{1/2}-(n-2)(\mathbf{\Lambda}+\sigma^2\mathbf{I}_r)\| \\
\le&(n-2)(\lambda_1+\sigma^2)\big\|\frac{1}{n-2}\mathbf{Z}_{(1)}\mathbf{Z}_{(1)}^T-\mathbf{I}_r\big\|,
\end{aligned}
$$

where the first inequality follows from Weyl's inequality. Hence,

$$
\begin{aligned}
&\Big|\frac{(n-2)^{-1}\lambda_i(\mathbf{Z}^T\mathbf{E}\mathbf{Z})}{\lambda_i}-1\Big|\le\frac{1}{(n-2)\lambda_i}\Big|\lambda_i(\mathbf{Z}^T\mathbf{E}\mathbf{Z})-(n-2)(\lambda_i+\sigma^2)\Big|+\frac{\sigma^2}{\lambda_i} \\
\le&\frac{\sigma^2}{(n-2)\lambda_i}\|\mathbf{Z}_{(2)}^T\mathbf{Z}_{(2)}\|+\frac{\lambda_1+\sigma^2}{\lambda_i}\big\|\frac{1}{n-2}\mathbf{Z}_{(1)}\mathbf{Z}_{(1)}^T-\mathbf{I}_r\big\|\Big)+\frac{\sigma^2}{\lambda_i}.
\end{aligned}
$$

By Corollary 2, the first term satisfies

$$
\frac{\sigma^2}{(n-2)\lambda_i}\|\mathbf{Z}_{(2)}^T\mathbf{Z}_{(2)}\|=O_P\Big(\max\big(\frac{\sigma^2}{\lambda_i},\frac{\sigma^2 p}{(n-2)\lambda_i}\big)\Big)=o_P(1).
$$

By law of large numbers, $\big\|\frac{1}{n-2}\mathbf{Z}_{(1)}\mathbf{Z}_{(1)}^T-\mathbf{I}_r\big\|=o_P(1)$. Hence

$$
\Big|\frac{(n-2)^{-1}\lambda_i(\mathbf{Z}^T\mathbf{E}\mathbf{Z})}{\lambda_i}-1\Big|=o_P(1).
$$

□

Lemma 5. *Under Assumption 1, we have*

$$
\|\hat{\mathbf{V}}\hat{\mathbf{V}}^T-\mathbf{V}\mathbf{V}^T\|^2=O_P(\frac{p}{p^\beta n}).
$$

The convergence rate $p/(p^\beta n)$ is optimal, see Cai et al. (2013), Theorem 5.

Proof. By Lemma 3,

$$
\|\hat{\mathbf{V}}\hat{\mathbf{V}}^T-\mathbf{V}\mathbf{V}^T\|\le\frac{2}{\lambda_r}\|S-\mathbf{\Sigma}\|.
$$

We only need to bound the right hand side Define $\mathbf{U}$, $\mathbf{E}$, $\mathbf{Z}$, $\mathbf{Z}_{(1)}$ and $\mathbf{Z}_{(2)}$ as in the proof of Proposition 1. Since $S \sim (n-2)^{-1}\mathbf{U}\mathbf{E}^{1/2}\mathbf{Z}\mathbf{Z}^T\mathbf{E}^{1/2}\mathbf{U}^T$, we have

$$\begin{aligned}
&\|S-\mathbf{\Sigma}\| = \|(\mathbf{V}\mathbf{V}^T + \tilde{\mathbf{V}}\tilde{\mathbf{V}}^T)(S-\mathbf{\Sigma})(\mathbf{V}\mathbf{V}^T + \tilde{\mathbf{V}}\tilde{\mathbf{V}}^T)\| \\
\le&\|\mathbf{V}\mathbf{V}^T(S-\mathbf{\Sigma})\mathbf{V}\mathbf{V}^T\| + 2\|\mathbf{V}\mathbf{V}^T(S-\mathbf{\Sigma})\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\| + \|\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T(S-\mathbf{\Sigma})\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\| \\
\le&\|\mathbf{V}^T(S-\mathbf{\Sigma})\mathbf{V}\| + 2\|\mathbf{V}^T(S-\mathbf{\Sigma})\tilde{\mathbf{V}}\| + \|\tilde{\mathbf{V}}^T(S-\mathbf{\Sigma})\tilde{\mathbf{V}}\| \\
\sim&\Big\|\frac{1}{n-2}(\mathbf{\Lambda}+\sigma^2\mathbf{I}_r)^{1/2}\mathbf{Z}_{(1)}\mathbf{Z}_{(1)}^T(\mathbf{\Lambda}+\sigma^2\mathbf{I}_r)^{1/2} - (\mathbf{\Lambda}+\sigma^2\mathbf{I}_r)\Big\| \\
&+\Big\|\frac{1}{n-2}\sigma(\mathbf{\Lambda}+\sigma^2\mathbf{I}_r)^{1/2}\mathbf{Z}_{(1)}\mathbf{Z}_{(2)}^T\Big\| + \sigma^2\Big\|\frac{1}{n-2}\mathbf{Z}_{(2)}\mathbf{Z}_{(2)}^T - \mathbf{I}_{p-r}\Big\| \\
\le&(\lambda_1+\sigma^2)\Big\|\frac{1}{n-2}\mathbf{Z}_{(1)}\mathbf{Z}_{(1)}^T - \mathbf{I}_r\Big\| + \frac{\sqrt{(\lambda_1+\sigma^2)\sigma^2}}{n-2}\|\mathbf{Z}_{(1)}\mathbf{Z}_{(2)}^T\| + \sigma^2\Big\|\frac{1}{n-2}\mathbf{Z}_{(2)}\mathbf{Z}_{(2)}^T - \mathbf{I}_{p-r}\Big\|
\end{aligned}$$

By law of large numbers, $\|\frac{1}{n-2}\mathbf{Z}_{(1)}\mathbf{Z}_{(1)}^T - \mathbf{I}_r\| = O_P(1/\sqrt{n})$. By Lemma 3, $\big\|\frac{1}{n-2}\mathbf{Z}_{(2)}\mathbf{Z}_{(2)}^T - \mathbf{I}_{p-r}\big\| = O_p(\max(\sqrt{p/n}, p/n))$. By the independence of $\mathbf{Z}_{(1)}$ and $\mathbf{Z}_{(2)}$, we have

$$\mathrm{E}\,\|\mathbf{Z}_{(1)}\mathbf{Z}_{(2)}^T\|^2 \le \mathrm{E}\,\|\mathbf{Z}_{(1)}\mathbf{Z}_{(2)}^T\|_F^2 = \mathrm{E}\,\mathrm{tr}\left(\mathbf{Z}_{(1)}\mathbf{Z}_{(2)}^T\mathbf{Z}_{(2)}\mathbf{Z}_{(1)}^T\right) = (p-r)\,\mathrm{E}\,\mathrm{tr}\left(\mathbf{Z}_{(1)}\mathbf{Z}_{(1)}^T\right) = rn(p-r).$$

Hence $\|\mathbf{Z}_{(1)}\mathbf{Z}_{(2)}^T\| = O_P(\sqrt{np})$. Combining these bounds leads to

$$\|S-\mathbf{\Sigma}\| = O_P(\frac{\lambda_1}{\sqrt{n}}) + O_P(\sqrt{\frac{\lambda_1 p}{n}}) + O_P\Big(\max\big(\sqrt{\frac{p}{n}}, \frac{p}{n}\big)\Big) = O_P(\sqrt{\frac{\lambda_1 p}{n}}) + O_P(\frac{p}{n}).$$

Thus,

$$\|\hat{\mathbf{V}}\hat{\mathbf{V}}^T - \mathbf{V}\mathbf{V}^T\| \le \frac{2}{\lambda_r}\|S-\mathbf{\Sigma}\| = O_P(\sqrt{\frac{p}{n\lambda_r}}) + O_P(\frac{p}{n\lambda_r}) = O_P(\sqrt{\frac{p}{n\lambda_r}}).$$

□

Proof of Proposition 2. Note that

$$\begin{aligned}
&\|\tilde{\mathbf{V}}^T(\bar{X}_1-\bar{X}_2)\|^2 = \big\|\tilde{\mathbf{V}}^T(\mu_1-\mu_2) + \tilde{\mathbf{V}}^T\big((\bar{X}_1-\mu_1)-(\bar{X}_2-\mu_2)\big)\big\|^2 \\
=&\|\tilde{\mathbf{V}}^T(\mu_1-\mu_2)\|^2 + \big\|\tilde{\mathbf{V}}^T\big((\bar{X}_1-\mu_1)-(\bar{X}_2-\mu_2)\big)\big\|^2 + 2(\mu_1-\mu_2)^T\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\big((\bar{X}_1-\mu_1)-(\bar{X}_2-\mu_2)\big) \\
=&\|\tilde{\mathbf{V}}^T(\mu_1-\mu_2)\|^2 + \big\|\tilde{\mathbf{V}}^T\big((\bar{X}_1-\mu_1)-(\bar{X}_2-\mu_2)\big)\big\|^2 + o_P(\frac{\sqrt{p}}{n}).
\end{aligned} \tag{13}$$

The last equality holds since

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

For $k = 1, 2$, we have

$$\frac{1}{n_k}\mathrm{tr}(\tilde{\mathbf{V}}^T S_k \tilde{\mathbf{V}}) \sim \frac{\sigma^2}{n_k(n_k-1)}\chi^2_{(p-r)(n_k-1)} = \sigma^2 \frac{p-r}{n_k}\Big(1 + O_P\big(\frac{1}{\sqrt{(p-r)(n_k-1)}}\big)\Big),$$

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

$$\frac{1}{n_1}\mathrm{tr}(\tilde{\mathbf{V}}^T S_1 \tilde{\mathbf{V}}) + \frac{1}{n_2}\mathrm{tr}(\tilde{\mathbf{V}}^T S_2 \tilde{\mathbf{V}}) = \sigma^2\tau(p-r) + o_P(\frac{\sqrt{p}}{n}). \tag{14}$$

Equation (13) and (14) imply that

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

Since $\|\tilde{\mathbf{V}}^T\big((\bar{X}_1 - \mu_1) - (\bar{X}_2 - \mu_2)\big)\|^2 \sim \sigma^2\tau\chi^2_{p-r}$, the proposition follows by central limit theorem. □

Proof of Proposition 3. Note that $(n-2)S \sim \mathrm{Wishart}_p(n-2, \mathbf{\Sigma})$. Define $\mathbf{U}$, $\mathbf{E}$, $\mathbf{Z}$, $\mathbf{Z}_{(1)}$ and $\mathbf{Z}_{(2)}$ as in the proof of Proposition 1. We have

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

Hence

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

We note that

$$\mathbf{Z}^T\mathbf{E}\mathbf{Z} = \mathbf{Z}_{(1)}^T(\mathbf{\Lambda} + \sigma^2\mathbf{I}_r)\mathbf{Z}_{(1)} + \sigma^2\mathbf{Z}_{(2)}^T\mathbf{Z}_{(2)},$$

where the first term is of rank r. Applying Weyl's inequality yields

$$\sigma^2\lambda_i(\mathbf{Z}_{(2)}^T\mathbf{Z}_{(2)}) \le \lambda_i(\mathbf{Z}^T\mathbf{E}\mathbf{Z}) \le \sigma^2\lambda_{i-r}(\mathbf{Z}_{(2)}^T\mathbf{Z}_{(2)}), \quad i = r+1, \dots, n-2.$$

Summing over i gives

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

Then

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

Note that $\lambda_i(\mathbf{Z}_{(2)}^T\mathbf{Z}_{(2)})$ is bounded above by $\lambda_1(\mathbf{Z}_{(2)}^T\mathbf{Z}_{(2)})$ and by Corollary 2, $\lambda_1(\mathbf{Z}_{(2)}^T\mathbf{Z}_{(2)}) = O_P(\max(n,p))$. It follows that

$$\begin{aligned}
&\Big|\frac{1}{(p-r)(n-2)}\sum_{i=r+1}^{n-2}\lambda_i(\mathbf{Z}^T\mathbf{E}\mathbf{Z}) - \frac{1}{(p-r)(n-2)}\sigma^2\sum_{i=1}^{n-2}\lambda_i(\mathbf{Z}_{(2)}^T\mathbf{Z}_{(2)})\Big| \\
\le& r\sigma^2\frac{1}{(p-r)(n-2)}\lambda_1(\mathbf{Z}_{(2)}^T\mathbf{Z}_{(2)}) = O_P\Big(\frac{\max(n,p)}{np}\Big).
\end{aligned}$$

Hence

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

Note that $\mathrm{tr}(\mathbf{Z}_{(2)}^T\mathbf{Z}_{(2)})$ is a sum of $(p-r)(n-2)$ i.i.d. χ_1^2 random variables. By central limit theorem,

$$\frac{1}{(p-r)(n-2)}\sigma^2\,\mathrm{tr}(\mathbf{Z}_{(2)}^T\mathbf{Z}_{(2)}) = \sigma^2 + O_P\Big(\frac{1}{\sqrt{np}}\Big).$$

Therefore,

$$\frac{1}{(p-r)(n-2)}\sum_{i=r+1}^{n-2}\lambda_i(\mathbf{Z}^T\mathbf{E}\mathbf{Z}) = \sigma^2 + O_P\Big(\frac{1}{\sqrt{np}}\Big) + O_P\Big(\frac{\max(n,p)}{np}\Big) = \sigma^2 + O_P\Big(\frac{\max(n,p)}{np}\Big),$$

where the last equality holds since

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

$\square$

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

$k = 1, 2$. Hence,

$$
\begin{aligned}
&\frac{T_2 - \|\tilde{\mathbf{V}}^T(\mu_1 - \mu_2)\|^2}{\sigma^2\sqrt{2\tau^2 p}} \\
=&\frac{\|\hat{\tilde{\mathbf{V}}}^T(\bar{X}_1 - \bar{X}_2)\|^2 - \|\tilde{\mathbf{V}}^T(\mu_1 - \mu_2)\|^2 - \sigma^2\tau(p-r)}{\sigma^2\sqrt{2\tau^2 p}} \\
&- \frac{\frac{1}{n_1}(\mathrm{tr}(\hat{\tilde{\mathbf{V}}}_1^T S_1 \hat{\tilde{\mathbf{V}}}_1) - (p-r)\sigma^2)}{\sigma^2\sqrt{2\tau^2 p}} - \frac{\frac{1}{n_2}(\mathrm{tr}(\hat{\tilde{\mathbf{V}}}_2^T S_2 \hat{\tilde{\mathbf{V}}}_2) - (p-r)\sigma^2)}{\sigma^2\sqrt{2\tau^2 p}} \\
=&\frac{\|\hat{\tilde{\mathbf{V}}}^T(\bar{X}_1 - \bar{X}_2)\|^2 - \|\tilde{\mathbf{V}}^T(\mu_1 - \mu_2)\|^2 - \sigma^2\tau(p-r)}{\sigma^2\sqrt{2\tau^2 p}} + O_P\Big(\frac{\max(n,p)}{n\sqrt{p}}\Big) \\
=&\frac{\|\hat{\tilde{\mathbf{V}}}^T(\bar{X}_1 - \bar{X}_2)\|^2 - \|\tilde{\mathbf{V}}^T(\mu_1 - \mu_2)\|^2 - \sigma^2\tau(p-r)}{\sigma^2\sqrt{2\tau^2 p}} + o_P(1),
\end{aligned}
$$

where the last equality holds since

$$\frac{\max(n,p)}{n\sqrt{p}} = \max\Big(\frac{1}{\sqrt{p}}, \frac{\sqrt{p}}{n}\Big) \to 0.$$

We write

$$
\begin{aligned}
&\frac{\|\hat{\tilde{\mathbf{V}}}^T(\bar{X}_1 - \bar{X}_2)\|^2 - \|\tilde{\mathbf{V}}^T(\mu_1 - \mu_2)\|^2 - \sigma^2\tau(p-r)}{\sigma^2\sqrt{2\tau^2 p}} \\
=&\frac{1}{\sigma^2\sqrt{2\tau^2 p}}(P_1 + P_2 + P_3),
\end{aligned}
$$

where

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

To prove the theorem, it suffices to show that

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

First we deal with P_2. Let ϵ be any fixed positive number. We have

$$\Pr\Big(\frac{P_2}{\sigma^2\sqrt{2\tau^2 p}} > \epsilon\Big) = \mathrm{E}[\Pr(P_2 > \epsilon\sigma^2\sqrt{2\tau^2 p}\big|S)].$$

Since the conditional probability $\Pr(P_2 > \epsilon\sigma^2\sqrt{2\tau^2 p}\big|S)$ is bounded, by dominated convergence theorem, we only need to prove $\Pr(P_2 > \epsilon\sigma^2\sqrt{2\tau^2 p}\big|S) \xrightarrow{P} 0$. Note that $\bar{X}_1$, $\bar{X}_2$, and S are mutually independent and $\hat{\tilde{\mathbf{V}}}\hat{\tilde{\mathbf{V}}}^T$ only depends on S. We have

$$
\begin{aligned}
&\Pr(P_2 > \epsilon\sigma^2\sqrt{2\tau^2 p}\big|S) \le \frac{1}{2\epsilon^2\sigma^4\tau^2 p}\,\mathrm{E}(P_2^2|S)\\
=&\frac{1}{2\epsilon^2\sigma^4\tau^2 p}4\tau(\mu_1-\mu_2)^T\hat{\tilde{\mathbf{V}}}\hat{\tilde{\mathbf{V}}}^T\mathbf{\Sigma}\hat{\tilde{\mathbf{V}}}\hat{\tilde{\mathbf{V}}}^T(\mu_1-\mu_2)\\
\le&\frac{2}{\epsilon^2\sigma^4\tau p}\lambda_1(\hat{\tilde{\mathbf{V}}}^T\mathbf{\Sigma}\hat{\tilde{\mathbf{V}}})(\mu_1-\mu_2)^T\hat{\tilde{\mathbf{V}}}\hat{\tilde{\mathbf{V}}}^T(\mu_1-\mu_2)\\
\le&\frac{2}{\epsilon^2\sigma^4\tau p}\|\mu_1-\mu_2\|^2\lambda_1(\hat{\tilde{\mathbf{V}}}^T\mathbf{\Sigma}\hat{\tilde{\mathbf{V}}})\\
=&O(\frac{1}{\sqrt{p}})\lambda_1(\hat{\tilde{\mathbf{V}}}^T(\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T+\sigma^2\mathbf{I}_p)\hat{\tilde{\mathbf{V}}})\\
\le&O(\frac{1}{\sqrt{p}})\big(\kappa p^\beta\lambda_1(\hat{\tilde{\mathbf{V}}}^T\mathbf{V}\mathbf{V}^T\hat{\tilde{\mathbf{V}}})+\sigma^2\big).
\end{aligned}
$$

But

$$
\lambda_1(\hat{\tilde{\mathbf{V}}}^T\mathbf{V}\mathbf{V}^T\hat{\tilde{\mathbf{V}}}) = \|\mathbf{V}^T\hat{\tilde{\mathbf{V}}}\|^2 = \|\mathbf{V}\mathbf{V}^T - \hat{\mathbf{V}}\hat{\mathbf{V}}^T\|^2 = O_P\Big(\frac{p}{p^\beta n}\Big),
$$

where the last two equality follows from Golub and Van Loan (2013), Theorem 2.5.1 and the last equality follows from Lemma 5. Thus,

$$
\Pr(P_2 > \epsilon\sigma^2\sqrt{2\tau^2 p}\big|S) = O(\frac{1}{\sqrt{p}})\big(O_P(\frac{p}{n})+\sigma^2\big) = O(1)(O_P(\frac{\sqrt{p}}{n})+\frac{\sigma^2}{\sqrt{p}}) = o_P(1).
$$

Next we deal with P_3. Note that

$$
\begin{aligned}
&|P_3| = \big|(\mu_1-\mu_2)^T(\hat{\tilde{\mathbf{V}}}\hat{\tilde{\mathbf{V}}}^T - \tilde{\mathbf{V}}\tilde{\mathbf{V}}^T)(\mu_1-\mu_2)\big| \le \|\mu_1-\mu_2\|^2\|\hat{\tilde{\mathbf{V}}}\hat{\tilde{\mathbf{V}}}^T - \tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\|\\
=&\|\mu_1-\mu_2\|^2\|\hat{\mathbf{V}}\hat{\mathbf{V}}^T - \mathbf{V}\mathbf{V}^T\| = O(\frac{\sqrt{p}}{n})\sqrt{O_P(\frac{p}{p^\beta n})} = o_P(\frac{\sqrt{p}}{n}).
\end{aligned}
$$

Hence

$$
\frac{P_3}{\sigma^2\sqrt{2\tau^2 p}} = O(\frac{n}{\sqrt{p}})P_3 = o_P(1).
$$

Now we prove the asymptotic normality of P_1. To make clear the mode 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 $F_n \xrightarrow{\mathcal{L}} F$.

Since the conditional distribution of $\hat{\hat{\mathbf{V}}}^T\big((\bar{X}_1 - \mu_1) - (\bar{X}_2 - \mu_2)\big)$ given S is $N(0, \tau\hat{\hat{\mathbf{V}}}^T\mathbf{\Sigma}\hat{\hat{\mathbf{V}}})$, we have that

$$\tau^{-1}\big\|\hat{\hat{\mathbf{V}}}^T\big((\bar{X}_1 - \mu_1) - (\bar{X}_2 - \mu_2)\big)\big\|^2 \sim \sum_{i=1}^{p-r} \lambda_i(\hat{\hat{\mathbf{V}}}^T\mathbf{\Sigma}\hat{\hat{\mathbf{V}}})\xi_i^2, \tag{15}$$

where $\{\xi_i\}_{i=1}^{p-r}$ are i.i.d. standard normal random variables which are independent of $\hat{\hat{\mathbf{V}}}$. So the asymptotic distribution of P_1 relies on the asymptotic behavior of $\lambda_i(\hat{\hat{\mathbf{V}}}^T\mathbf{\Sigma}\hat{\hat{\mathbf{V}}})$. As we have shown,

$$\lambda_1(\hat{\hat{\mathbf{V}}}^T\mathbf{\Sigma}\hat{\hat{\mathbf{V}}}) \le \kappa p^\beta \lambda_1(\hat{\hat{\mathbf{V}}}^T\mathbf{V}\mathbf{V}^T\hat{\hat{\mathbf{V}}}) + \sigma^2 = \kappa p^\beta \|\mathbf{V}\mathbf{V}^T - \hat{\mathbf{V}}\hat{\mathbf{V}}^T\|^2 + \sigma^2. \tag{16}$$

Hence $\lambda_i(\hat{\hat{\mathbf{V}}}^T\mathbf{\Sigma}\hat{\hat{\mathbf{V}}}) = O_P(p/n + 1)$, $i = 1, \dots, r$. On the other hand, for $i = r+1, \dots, p-r$, we have

$$\lambda_i(\hat{\hat{\mathbf{V}}}^T\mathbf{\Sigma}\hat{\hat{\mathbf{V}}}) = \lambda_i(\hat{\hat{\mathbf{V}}}^T\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T\hat{\hat{\mathbf{V}}}) + \sigma^2 = \sigma^2, \tag{17}$$

where the last equality follows from $\mathrm{Rank}(\hat{\hat{\mathbf{V}}}^T\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T\hat{\hat{\mathbf{V}}}) \le \mathrm{Rank}(\mathbf{V}) = r$. This, combined with (16), yields

$$\mathrm{tr}(\hat{\hat{\mathbf{V}}}^T\mathbf{\Sigma}\hat{\hat{\mathbf{V}}})^2 = \big(\frac{p}{n} + 1\big)^2 O_P(1) + (p - 2r)\sigma^4 = p\sigma^4(1 + o_P(1)). \tag{18}$$

Consequently,

$$\frac{\lambda_1^2(\hat{\hat{\mathbf{V}}}^T\mathbf{\Sigma}\hat{\hat{\mathbf{V}}})}{\mathrm{tr}(\hat{\hat{\mathbf{V}}}^T\mathbf{\Sigma}\hat{\hat{\mathbf{V}}})^2} = O_P\Big(\frac{(p/n+1)^2}{p}\Big) = o_P(1). \tag{19}$$

Then for every subsequence of $\{n\}$, there's a further subsequence along which (19) holds almost surely. This, combined with (15) and Lemma 1, implies that for every subsequence of $\{n\}$, there's a further subsequence along which

$$\rho\big(\mathcal{L}(Y_n|S), N(0,1)\big) \xrightarrow{a.s.} 0, \tag{20}$$

where

$$Y_n = \frac{\|\hat{\hat{\mathbf{V}}}^T\big((\bar{X}_1 - \mu_1) - (\bar{X}_2 - \mu_2)\big)\|^2 - \tau\mathrm{tr}(\hat{\hat{\mathbf{V}}}^T\mathbf{\Sigma}\hat{\hat{\mathbf{V}}})}{\sqrt{2\tau^2\mathrm{tr}(\hat{\hat{\mathbf{V}}}^T\mathbf{\Sigma}\hat{\hat{\mathbf{V}}})^2}},$$

and $\mathcal{L}(Y_n|S)$ is the conditional distribution of Y_n given S. By the definition of weak convergence, if (20) holds along some subsequence $\{n_k\}$, then for every continuous bounded function $f(\cdot)$, $\mathrm{E}[f(Y_n)|S] \xrightarrow{a.s.} \mathrm{E}[f(\epsilon)]$ along $\{n_k\}$, where ϵ is a random variable with standard normal distribution. By dominated convergence theorem, $\mathrm{E}[f(Y_n)] \to \mathrm{E}[f(\epsilon)]$ along $\{n_k\}$. This implies that $Y_n \xrightarrow{\mathcal{L}} N(0,1)$ along $\{n_k\}$. Thus, for every subsequence of n, there is a further subsequence along which $Y_n \xrightarrow{\mathcal{L}} N(0,1)$ along $\{n_k\}$. This means $Y_n \xrightarrow{\mathcal{L}} N(0,1)$, or

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

By (16) and (17), we have

$$\begin{aligned}\mathrm{tr}(\hat{\hat{\mathbf{V}}}^T\mathbf{\Sigma}\hat{\hat{\mathbf{V}}}) &= \sum_{i=1}^{r}\lambda_i(\hat{\hat{\mathbf{V}}}^T\mathbf{\Sigma}\hat{\hat{\mathbf{V}}}) + \sum_{i=r+1}^{p-r}\lambda_i(\hat{\hat{\mathbf{V}}}^T\mathbf{\Sigma}\hat{\hat{\mathbf{V}}}) \\ &= O_P(\frac{p}{n}+1)+(p-2r)\sigma^2 = (p-r)\sigma^2 + o_P(\sqrt{p}).\end{aligned} \tag{21}$$

By (18), (21) and Slutsky's theorem, we have

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

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

Acknowledgements

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

References

Ahn, S.C., Horenstein, A.R., 2013. Eigenvalue ratio test for the number of factors. Econometrica 81, 1203–1227. doi:10.3982/ecta8968.

Bai, Z., Saranadasa, H., 1996. Effect of high dimension: by an example of a two sample problem. Statistica Sinica 6, 311–329.

Birnbaum, A., Johnstone, I.M., Nadler, B., Paul, D., 2013. Minimax bounds for sparse PCA with noisy high-dimensional data. The Annals of Statistics 41, 1055–1084. doi:10.1214/12-aos1014.

Cai, T., Ma, Z., Wu, Y., 2015. Optimal estimation and rank detection for sparse spiked covariance matrices. Probability Theory & Related Fields 161, 781–815.

Cai, T.T., Ma, Z., Wu, Y., 2013. Sparse PCA: Optimal rates and adaptive estimation. The Annals of Statistics 41, 3074–3110. doi:10.1214/13-aos1178.

Chen, L.S., Paul, D., Prentice, R.L., Wang, P., 2011. A regularized hotelling'sT2test for pathway analysis in proteomic studies. Journal of the American Statistical Association 106, 1345–1360. doi:10.1198/jasa.2011.ap10599.

Chen, S.X., Qin, Y.L., 2010. A two-sample test for high-dimensional data with applications to gene-set testing. The Annals of Statistics 38, 808–835. doi:10.1214/09-aos716.

Davidson, K.R., Szarek, S.J., 2001. Handbook of the Geometry of Banach Spaces. volume 1. North-Holland, Amsterdam. URL: http://www.sciencedirect.com/science/article/pii/S1874584901800103, doi:http://dx.doi.org/10.1016/S1874-5849(01)80010-3. handbook of the Geometry of Banach Spaces.

Fan, J., Wang, W., 2015. Asymptotics of empirical eigen-structure for ultra-high dimensional spiked covariance model arXiv:arXiv:1502.04733.

Golub, G.H., Van Loan, C.F., 2013. Matrix Computations. Fourth ed., The Johns Hopkins University Press.

Horn, R.A., Johnson, C.R., 2012. Matrix Analysis. 2nd ed., Cambridge University Press, New York.

Jung, S., Marron, J.S., 2009. PCA consistency in high dimension, low sample size context. The Annals of Statistics 37, 4104–4130. doi:10.1214/09-aos709.

Lopes, M., Jacob, L., Wainwright, M.J., 2011. A more powerful two-sample test in high dimensions using random projection, in: Shawe-Taylor, J., Zemel, R.S., Bartlett, P.L., Pereira, F., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 24. Curran Associates, Inc., pp. 1206–1214.

Ma, Y., Lan, W., Wang, H., 2015. A high dimensional two-sample test under a low dimensional factor structure. Journal of Multivariate Analysis 140, 162–170. doi:10.1016/j.jmva.2015.05.005.

Passemier, D., Li, Z., Yao, J., 2017. On estimation of the noise variance in high dimensional probabilistic principal component analysis. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79, 51–67. doi:10.1111/rssb.12153.

Srivastava, M.S., Du, M., 2008. A test for the mean vector with fewer observations than the dimension. Journal of Multivariate Analysis 99, 386–402. doi:10.1016/j.jmva.2006.11.002.

Srivastava, R., Li, P., Ruppert, D., 2016. RAPTT: An exact two-sample test in high dimensions using random projections. Journal of Computational and Graphical Statistics 25, 954–970. doi:10.1080/10618600.2015.1062771.

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

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

Table 1: $n_1 = n_2 = 60$

	Normal				Chi-squared				Student's t			
p	200	400	600	800	200	400	600	800	200	400	600	800
$\beta = 0.5$												
New1	0.092	0.109	0.121	0.132	0.082	0.111	0.118	0.130	0.093	0.119	0.123	0.136
New2	0.057	0.064	0.067	0.070	0.050	0.065	0.070	0.081	0.060	0.068	0.067	0.073
oracle	0.048	0.051	0.049	0.051	0.040	0.046	0.051	0.052	0.052	0.050	0.047	0.051
chi	0.046	0.045	0.044	0.035	0.043	0.045	0.041	0.042	0.042	0.040	0.036	0.040
fast	0.067	0.061	0.059	0.052	0.065	0.071	0.057	0.057	0.066	0.063	0.059	0.057
CQ	0.060	0.057	0.059	0.054	0.059	0.068	0.059	0.061	0.061	0.058	0.057	0.058
$\beta = 1$												
New1	0.094	0.107	0.134	0.148	0.099	0.114	0.134	0.136	0.081	0.113	0.137	0.140
New2	0.061	0.071	0.081	0.090	0.070	0.072	0.074	0.067	0.056	0.068	0.086	0.076
oracle	0.051	0.056	0.063	0.059	0.054	0.059	0.051	0.046	0.046	0.051	0.059	0.052
chi	0.048	0.059	0.052	0.052	0.062	0.055	0.054	0.045	0.055	0.058	0.057	0.064
fast	0.050	0.053	0.050	0.056	0.060	0.054	0.056	0.043	0.058	0.058	0.058	0.061
CQ	0.067	0.074	0.068	0.072	0.084	0.067	0.072	0.065	0.075	0.081	0.080	0.086
$\beta = 2$												
New1	0.076	0.112	0.117	0.127	0.095	0.110	0.115	0.135	0.090	0.105	0.128	0.131
New2	0.049	0.070	0.071	0.066	0.067	0.066	0.068	0.075	0.062	0.061	0.079	0.072
oracle	0.041	0.050	0.048	0.040	0.056	0.044	0.044	0.050	0.050	0.050	0.050	0.046
chi	0.056	0.057	0.055	0.058	0.056	0.059	0.061	0.046	0.048	0.048	0.058	0.058
fast	0.057	0.057	0.051	0.056	0.059	0.059	0.063	0.051	0.051	0.050	0.059	0.057
CQ	0.076	0.072	0.067	0.079	0.072	0.077	0.082	0.067	0.070	0.065	0.079	0.080

Table 2: $n_1 = n_2 = 120$

	Normal				Chi-squared				Student's t			
p	200	400	600	800	200	400	600	800	200	400	600	800
$\beta = 0.5$												
New1	0.087	0.078	0.083	0.085	0.080	0.071	0.085	0.094	0.065	0.072	0.088	0.081
New2	0.068	0.059	0.060	0.066	0.064	0.053	0.062	0.062	0.051	0.056	0.059	0.057
oracle	0.054	0.046	0.051	0.051	0.058	0.048	0.053	0.052	0.043	0.048	0.051	0.044
chi	0.048	0.053	0.051	0.039	0.042	0.049	0.048	0.038	0.052	0.046	0.042	0.046
fast	0.070	0.073	0.075	0.063	0.070	0.074	0.078	0.064	0.078	0.071	0.070	0.071
CQ	0.056	0.064	0.063	0.054	0.052	0.064	0.066	0.054	0.062	0.059	0.059	0.062
$\beta = 1$												
New1	0.084	0.081	0.082	0.087	0.069	0.084	0.080	0.089	0.072	0.080	0.086	0.080
New2	0.068	0.061	0.059	0.062	0.057	0.061	0.066	0.058	0.060	0.059	0.061	0.060
oracle	0.062	0.053	0.052	0.051	0.052	0.056	0.057	0.046	0.054	0.050	0.053	0.053
chi	0.046	0.056	0.053	0.054	0.043	0.053	0.057	0.052	0.053	0.043	0.057	0.058
fast	0.046	0.058	0.052	0.051	0.043	0.054	0.053	0.052	0.055	0.045	0.053	0.059
CQ	0.062	0.074	0.069	0.070	0.058	0.069	0.072	0.070	0.067	0.062	0.070	0.081
$\beta = 2$												
New1	0.074	0.068	0.080	0.098	0.070	0.070	0.083	0.099	0.064	0.072	0.070	0.080
New2	0.058	0.054	0.057	0.072	0.051	0.054	0.064	0.070	0.048	0.051	0.049	0.058
oracle	0.051	0.047	0.047	0.061	0.046	0.046	0.052	0.059	0.042	0.048	0.041	0.048
chi	0.060	0.058	0.046	0.051	0.060	0.051	0.043	0.053	0.054	0.054	0.054	0.048
fast	0.061	0.056	0.045	0.053	0.060	0.047	0.041	0.052	0.054	0.056	0.053	0.046
CQ	0.082	0.076	0.064	0.065	0.075	0.061	0.057	0.073	0.071	0.075	0.072	0.065

Table 3: 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