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

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

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

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

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

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

Assumption 3. *Assume that there is some constant $\kappa > 0$ and $\beta \geq \frac{1}{2}$ such that*

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

The restriction $\beta \geq 1/2$ breaks down the Condition (**??**). If $\beta < 1/2$, Condition (**??**) is meet and Chen and Qin (2010)'s method is valid. Hence $\beta = 1/2$ is the boundary of the scope between T_{CQ} and our new test. The case

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

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

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

where

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

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

We write $\xi \sim \eta$ to denote the random variable ξ and η have the same distribution. For nonrandom positive sequence $\{a_n\}$ and $\{b_n\}$, $a_n \asymp b_n$ represents $a_n \geq c b_n$ and $a_n \leq C b_n$ for some positive c, C for every n.

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

For a symmetric matrix A, we define $\lambda_i(A)$ to be the ith largest eigenvalue of A and $\lambda_{\max}(A)$, $\lambda_{\min}(A)$ to be the maximal and minimal eigenvalues respectively. We denote by $\operatorname{tr}(A)$ the trace of A.

The notations $\xrightarrow{P}$ and $\xrightarrow{\mathcal{L}}$ are used to denote convergence in probability and weak convergence respectively.

Let $[m] = \{1, \ldots, m\}$.

2. PCA Theory

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

The following lemma is from Davidson and Szarek (2001):

Lemma 1 (Davidson-Szarek bound)**.** *Let Z be an $N \times n$ matrix whose entries are independent standard normal random variables. Then for every $t > 0$, with probability at least $1 - 2\exp(-t^2/2)$ one has*

$$(\sqrt{N} - \sqrt{n} - t)^2 \leq \lambda_{\min(N,n)}(ZZ^T) \leq \lambda_1(ZZ^T) \leq (\sqrt{N} + \sqrt{n} + t)^2.$$

By the Cramer-Chernoff method, we have the following lemma.

Lemma 2. *Under the assumption of Lemma 1, then for every $t > 0$, with probability at least $1 - \exp\big(-t^2/2\big)$ one has*

$$\mathrm{tr}(ZZ^T) \geq Nn\Big(1 - \sqrt{\frac{2}{Nn}}t\Big).$$

Proof. Note that $\mathrm{tr}(ZZ^T) \sim \chi^2_{Nn}$. Then for $t > 0$, we have

$$\Pr\big(-\mathrm{tr}(ZZ^T) + Nn \geq t\big) = \Pr\big(\exp(-\lambda\chi^2_{Nn} + Nn\lambda) \geq \exp(t\lambda)\big)$$
$$\leq \exp(\big(Nn - t\big)\lambda)\,\mathrm{E}\exp(-\lambda\chi^2_{Nn}) = \exp\Big(\big(Nn - t\big)\lambda - \frac{Nn}{2}\log(1 + 2\lambda)\Big),$$

where $\lambda > 0$ can be arbitrary. If $0 < t < Nn$, let $\lambda = \frac{t}{2(Nn-t)}$ and we get

$$\Pr\big(-\mathrm{tr}(ZZ^T) + Nn \geq t\big) \leq \exp\Big(\frac{t}{2} + \frac{Nn}{2}\log(1 - \frac{t}{Nn})\Big).$$

Since for $0 < x < 1$, $\log(1-x) \leq -x - \frac{x^2}{2}$, we have that

$$\Pr\big(-\mathrm{tr}(ZZ^T) + Nn \geq t\big) \leq \exp\Big(-\frac{t^2}{4Nn}\Big). \qquad (1)$$

If $t \geq Nn$ The left hand side of (1) is 0 for trivial reason. Hence (1) holds for all $t > 0$. The conclusion follows by substituting t by $\sqrt{2Nnt}$ in (1). $\square$

Lemma 3. *Under the assumption of Lemma 1, then for every $t > 0$, with probability at least $1 - 3\exp(-t^2/2)$, for every i such that $1 \le i \le \min(N, n)$, we have*

$$\lambda_i(ZZ^T) \ge \max(N,n)\Big(1 - \sqrt{\frac{2}{Nn}}t - (i-1)(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{n}} + \frac{t}{\sqrt{Nn}})^2\Big).$$

Proof. By Lemma 2, for every $t > 0$, with probability at least $1 - \exp(-t^2/2)$, we have that

$$\sum_{j=1}^{i-1} \lambda_j(ZZ^T) + \sum_{j=i}^{\min(N,n)} \lambda_j(ZZ^T) = \mathrm{tr}(ZZ^T) \ge Nn(1 - \sqrt{\frac{2}{Nn}}t),$$

where $1 \le i \le \min(N, n)$. Thus, with probability at least $1 - \exp(-t^2/2)$, for every i such that $1 \le i \le \min(N, n)$, we have

$$\begin{aligned}
\lambda_i(ZZ^T) \ge & \frac{1}{\min(N,n)} \sum_{j=i}^{\min(N,n)} \lambda_j(ZZ^T) \\
\ge & \frac{1}{\min(N,n)}\Big(Nn(1 - \sqrt{\frac{2}{Nn}}t) - \sum_{j=1}^{i-1} \lambda_j(ZZ^T)\Big) \\
\ge & \frac{1}{\min(N,n)}\Big(Nn(1 - \sqrt{\frac{2}{Nn}}t) - (i-1)\lambda_1(ZZ^T)\Big).
\end{aligned}$$

By the above inequality and Lemma 1, with probability at least $1-3\exp(-t^2/2)$, for every i such that $1 \le i \le \min(N, n)$, we have

$$\begin{aligned}
\lambda_i(ZZ^T) \ge & \frac{1}{\min(N,n)}\Big(Nn(1 - \sqrt{\frac{2}{Nn}}t) - (i-1)(\sqrt{N} + \sqrt{n} + t)^2\Big) \\
= & \max(N,n)\Big(1 - \sqrt{\frac{2}{Nn}}t - (i-1)(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{n}} + \frac{t}{\sqrt{Nn}})^2\Big).
\end{aligned}$$

The last equality holds since $\max(N, n) = Nn/\min(N, n)$.

□

Assumption 4. *Suppose that $Z = (Z_1, \ldots, Z_n)$ is an $p \times n$ random matrix whose entries Z_{ij}'s are i.i.d. standard normal random variables, $i = 1, \ldots, p$, $j = 1, \ldots, n$. Let the sample matrix be $X = (X_1, \ldots, X_n) = U\Lambda^{1/2}Z$, where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ with $\lambda_1 \ge \cdots \ge \lambda_p$ and U is a $p \times p$ orthogonal matrix. Suppose $c \le \lambda_p \le \lambda_{r+1} \le C$, where $c > 0$ and $C > 0$ are absolute constants.*

The sample covariance matrix is $\frac{1}{n}XX^T$. We denote by $\frac{1}{n}XX^T = \hat{U}\hat{\Lambda}\hat{U}^T$ the spectral decomposition of S where $\hat{\Lambda} = \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_p)$ and $\hat{U}$ is a orthogonal matrix.

Let u_i be the ith column of U, $i = 1, \ldots, p$. Denote $U = (V, \tilde{V})$, where V and $\tilde{V}$ are the first r and last $p - r$ columns of U respectively. Similarly, we define the corresponding part of $\hat{U}$ by $\hat{u}_i$, $\hat{V}$ and $\hat{\tilde{V}}$.

The PCA theory is mainly focus on the convergence properties of $\hat{u}_i$ to it's population counterpart u_i. See Jung and Marron (2009), Shen et al. (2012), Shen et al. (2013) and Fan and Wang (2017) for some recent developements for PCA theory. Here we are mainly interested in the asymptotic properties of $\hat{V}$. Compared to existing results, the consistency results of $\hat{V}$ require less assumptions on the order of $\lambda_1, \ldots, \lambda_r$.

Let $Z_{(1)}$ and $Z_{(2)}$ be the first r rows and the last $p-r$ rows of Z respectively. Then $Z_{(1)}$ is an $r \times n$ matrix and $Z_{(2)}$ is an $(p - r) \times n$ matrix. Let $\Lambda_{(1)} = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$ and $\Lambda_{(2)} = \mathrm{diag}(\lambda_{r+1}, \ldots, \lambda_p)$. Define $\hat{\Lambda}_{(1)}$ and $\hat{\Lambda}_{(2)}$ in a similar way.

Theorem 1. *Suppose Assumption 4 holds. Let i be a fixed number such that $1 \le i \le r$, then for every $t > 0$, with probability at least $1 - 9\exp(-t^2/2)$, we have*

$$\frac{\hat{\lambda}_i}{\lambda_i} \ge 1 - \frac{2}{\sqrt{n}}(\sqrt{r} + t) +$$
$$\frac{c\max(p - r, n)}{n\lambda_i}\Big(1 - \sqrt{\frac{2}{(p - r)n}}t - (i - 1)\big(\frac{1}{\sqrt{p - r}} + \frac{1}{\sqrt{n}} + \frac{t}{\sqrt{(p - r)n}}\big)^2\Big),$$

and

$$\frac{\hat{\lambda}_i}{\lambda_i} \le 1 + \frac{2}{\sqrt{n}}(\sqrt{r} + t) + \frac{1}{n}(\sqrt{r} + t)^2 + \frac{C}{n\lambda_i}(\sqrt{p - r} + \sqrt{n} + t)^2.$$

Proof. The non-zero eigenvalues of $\frac{1}{n}XX^T$ are equal to that of $\frac{1}{n}X^TX$. And $\frac{1}{n}X^TX$ can be further written as the sum of two quantities

$$\frac{1}{n}X^TX = \frac{1}{n}Z^T\Lambda Z = \frac{1}{n}Z_{(1)}^T\Lambda_{(1)}Z_{(1)} + \frac{1}{n}Z_{(2)}^T\Lambda_{(2)}Z_{(2)} \stackrel{def}{=} A + B.$$

By Weyl's inequality,

$$\frac{\max(\lambda_i(A), \lambda_i(B))}{\lambda_i} \le \frac{\hat{\lambda}_i}{\lambda_i} \le \frac{\lambda_i(A)}{\lambda_i} + \frac{\lambda_{\max}(B)}{\lambda_i}, \tag{2}$$

where $i = 1, \dots, r$. We deal with $\lambda_i(A)$ and $\lambda_i(B)$ separately.

First we deal with $\lambda_i(A)$, $i = 1, \dots, r$. By Corollary 1, we have that

$$\frac{\lambda_i(A)}{\lambda_i} \le \frac{1}{n\lambda_i} \lambda_{\max}\big(Z_{(1)}^T \operatorname{diag}(\underbrace{0, \dots, 0}_{i-1}, \underbrace{\lambda_i, \dots, \lambda_i}_{r-i+1}) Z_{(1)}\big).$$

Then by Lemma 1, with probability at least $1 - 2\exp(-t^2/2)$ we have

$$\frac{\lambda_i(A)}{\lambda_i} \le \frac{1}{n}(\sqrt{n} + \sqrt{r-i+1} + t)^2 \le 1 + \frac{2}{\sqrt{n}}(\sqrt{r} + t) + \frac{1}{n}(\sqrt{r} + t)^2. \tag{3}$$

On the other hand, by Weyl's inequility, we have that

$$\frac{\lambda_i(A)}{\lambda_i} \ge \frac{1}{n\lambda_i} \lambda_i\big(Z_{(1)}^T \operatorname{diag}(\underbrace{\lambda_i, \dots, \lambda_i}_{i}, \underbrace{0, \dots, 0}_{r-i}) Z_{(1)}\big).$$

Again by Lemma 1, with probability at least $1 - 2\exp(-t^2/2)$ we have

$$\frac{\lambda_i(A)}{\lambda_i} \ge \frac{1}{n}(\sqrt{n} - \sqrt{i} - t)^2 \ge 1 - \frac{2}{\sqrt{n}}(\sqrt{r} + t). \tag{4}$$

Now we deal with $\lambda_i(B)$. Since $\lambda_i(B) \ge \frac{c}{n}\lambda_i(Z_{(2)}^T Z_{(2)})$, by Lemma 3, with probability at least $1 - 3\exp(-t^2/2)$ we have

$$\lambda_i(B) \ge \frac{c\max(p-r, n)}{n}\Big(1 - \sqrt{\frac{2}{(p-r)n}}t - (i-1)(\frac{1}{\sqrt{p-r}} + \frac{1}{\sqrt{n}} + \frac{t}{\sqrt{(p-r)n}})^2\Big). \tag{5}$$

Since $\lambda_1(B) \le \frac{C}{n}\lambda_1(Z_{(2)}^T Z_{(2)})$, by Lemma 1, with probability at least $1 - 2\exp(-t^2/2)$ we have

$$\lambda_{\max}(B) \le \frac{C}{n}(\sqrt{p-r} + \sqrt{n} + t)^2. \tag{6}$$

The theorem follows by (2), (3), (4), (5), and (6). $\square$

Theorem 2. *Suppose Assumption 4 holds and $p/n \to \infty$. If $\frac{p}{n\lambda_r} \to 0$, then almost surely we have*

$$\|\hat{V}\hat{V}^T - VV^T\|_F^2 \asymp \frac{p}{n\lambda_r}. \tag{7}$$

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

$$r - \frac{1}{2}\|\hat{V}\hat{V}^T - VV^T\|_F^2 = O_{a.s.}(\frac{n\lambda_1}{p}). \tag{8}$$

Proof. Since

$$\frac{1}{n}XX^T = \hat{U}\hat{\Lambda}\hat{U}^T = \frac{1}{n}U\Lambda^{1/2}ZZ^T\Lambda^{1/2}U^T,$$

we have

$$\Lambda^{-1/2}U^T\hat{U}\hat{\Lambda}\hat{U}^T U\Lambda^{-1/2} = \frac{1}{n}ZZ^T. \tag{9}$$

First, we prove (7). It follows from (9) that

$$\Lambda_{(2)}^{-1/2}\tilde{V}^T\hat{U}\hat{\Lambda}\hat{U}^T\tilde{V}\Lambda_{(2)}^{-1/2} = \frac{1}{n}\tilde{Z}_{(2)}\tilde{Z}_{(2)}^T. \tag{10}$$

The left hand side of (10) equals to $C+D$, where $C = \Lambda_{(2)}^{-1/2}\tilde{V}^T\hat{V}\hat{\Lambda}_{(1)}\hat{V}^T\tilde{V}\Lambda_{(2)}^{-1/2}$ and $D = \Lambda_{(2)}^{-1/2}\tilde{V}^T\hat{\tilde{V}}\hat{\Lambda}_{(2)}\hat{\tilde{V}}^T\tilde{V}\Lambda_{(2)}^{-1/2}$.

Obviously, we have $\lambda_1(C) \leq n^{-1}\lambda_1(\tilde{Z}_{(2)}\tilde{Z}_{(2)}^T)$ and $\lambda_1(C) \geq C^{-1}\hat{\lambda}_r\lambda_1(\tilde{V}^T\hat{V}\hat{V}^T\tilde{V})$. It follows that

$$\lambda_1(\tilde{V}^T\hat{V}\hat{V}^T\tilde{V}) \leq \frac{C}{n\hat{\lambda}_r}\lambda_1(\tilde{Z}_{(2)}\tilde{Z}_{(2)}^T).$$

If $p - r \geq n$, then $\text{Rank}(C) = r$, $\text{Rank}(D) = n - r$ and $\text{Rank}(C + D) = n$. By Weyl's inequality,

$$\lambda_n(C + D) \leq \lambda_r(C) + \lambda_{n-r+1}(D) = \lambda_r(C).$$

Thus,

$$\frac{1}{n}\lambda_n(\tilde{Z}_{(2)}\tilde{Z}_{(2)}^T) \leq \lambda_r(C) \leq \frac{\hat{\lambda}_1}{c}\lambda_r(\tilde{V}^T\hat{V}\hat{V}^T\tilde{V}),$$

or

$$\lambda_r(\tilde{V}^T\hat{V}\hat{V}^T\tilde{V}) \geq \frac{c}{n\hat{\lambda}_1}\lambda_n(\tilde{Z}_{(2)}\tilde{Z}_{(2)}^T).$$

By Bai Yin's law, we have that

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

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

$$\frac{n}{p}\hat{\Lambda}_{(1)}^{1/2}\hat{V}^T\tilde{V}\Lambda_{(2)}^{-1}\tilde{V}^T\hat{V}\hat{\Lambda}_{(1)}^{1/2} \xrightarrow{a.s.} I_r. \tag{11}$$

When $\frac{p}{n\lambda_r} \to 0$, $\hat\lambda_i$'s are ratio consistent for $1 \le i \le r$. That is, $\Lambda_{(1)}^{-1}\hat\Lambda_{(1)} \to I_r$ almost surely. Then it follows from (11) that

$$\frac{n}{p}\Lambda_{(1)}^{1/2}\hat V^T\tilde V\Lambda_{(2)}^{-1}\tilde V^T\hat V\Lambda_{(1)}^{1/2} \xrightarrow{a.s.} I_r. \tag{12}$$

Notice that

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

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

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

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

Then we prove (8). It follows from (9) that

$$\Lambda_{(1)}^{-1/2}V^T\hat U\hat\Lambda\hat U^TV\Lambda_{(1)}^{-1/2} = \frac{1}{n}\tilde Z_{(1)}(I - \frac{1}{n}J)\tilde Z_{(1)}^T \xrightarrow{a.s.} I_r. \tag{13}$$

But

$$\begin{aligned}\mathrm{tr}\big(\Lambda_{(1)}^{-1/2}V^T\hat U\hat\Lambda\hat U^TV\Lambda_{(1)}^{-1/2}\big) &\ge \mathrm{tr}\big(\Lambda_{(1)}^{-1/2}V^T\hat V\hat\Lambda_{(1)}\hat V^TV\Lambda_{(1)}^{-1/2}\big)\\ &\ge \frac{\hat\lambda_r}{\lambda_1}\Big(r - \frac{1}{2}\|\hat V\hat V^T - VV^T\|_F^2\Big).\end{aligned} \tag{14}$$

When $\frac{p}{n\lambda_r} \to \infty$, $\hat\lambda_r \asymp p/n$.Then (8) holds.

$\square$

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

Appendix

Lemma 4 (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 1. *Let H and P be two symmetric matrices and $M = H + P$. If* $\mathrm{rank}(P) < k$*, then*

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

Lemma 5 (Convergence rate of principal space estimation)**.** *Under the Assumption 1-**??**, 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

$$E\|\hat{V}\hat{V}^T - VV^T\|_F^2 \asymp r \wedge (p - r) \wedge \frac{r(p - r)}{(n_1 + n_2 - 2)h(\lambda)} \tag{15}$$

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 (15) 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. $\square$

Lemma 6 (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 1. Lemma 6 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 2. *Suppose that W_n is a $p \times p$ matrix distributed as* $\mathrm{Wishart}_p(n, I_p)$*. 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 6, 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 6, which proves the conclusion along the subsequance. Now the conclusion holds by a standard subsequance argument. □

Lemma 7. *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*

$$\frac{X_n^T A_n X_n - \mathrm{E} X_n^T A_n X_n}{[\mathrm{Var}(X_n^T A_n X_n)]^{1/2}} \xrightarrow{\mathcal{L}} N(0, 1) \tag{16}$$

is that

$$\frac{\lambda_{\max}(A_n^2)}{\mathrm{tr}(A_n^2)} \to 0. \tag{17}$$

Remark 2. 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

$$\frac{X_n^T A_n X_n - \mathrm{E} X_n^T A_n X_n}{[\mathrm{Var}(X_n^T A_n X_n)]^{1/2}} = \sum_{i=1}^{k_n} \frac{\lambda_i(A_n)}{\left[2\mathrm{tr}(A_n^2)\right]^{1/2}} (Z_{ni}^2 - 1), \tag{18}$$

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

If 17 holds, then

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

Hence 16 follows by Lindeberg's central limit theorem.

Conversely, if 16 holds, we will prove that there is a subsequence of $\{n\}$ along which 17 holds. Then 17 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 16 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 17 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 ??. Since V and $\tilde{V}$ are orthogonal, we have

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

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

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

But

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

Then

$$
\|\tilde{V}^T(\bar{X}_1 - \bar{X}_2)\|^2 = \|\tilde{V}^T(\mu_1 - \mu_2)\|^2 + \|\tilde{V}^T(\bar{Z}_1 - \bar{Z}_2)\|^2 + o_P(\frac{\sqrt{p}}{n}). \tag{19}
$$

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

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

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

$$
\frac{1}{n_1}\mathrm{tr}(\tilde{V}^T S_1 \tilde{V}) + \frac{1}{n_2}\mathrm{tr}(\tilde{V}^T S_2 \tilde{V}) = \sigma^2\tau(p - r) + o_P(\frac{\sqrt{p}}{n}). \tag{20}
$$

By (19) and (20), we have

$$
\frac{T_1 - \|\tilde{V}^T(\mu_1 - \mu_2)\|^2}{\sigma^2\sqrt{2\tau^2 p}} = \frac{\|\tilde{V}^T(\bar{Z}_1 - \bar{Z}_2)\|^2 - \sigma^2\tau(p - r)}{\sigma^2\sqrt{2\tau^2 p}} + o_P(1). \tag{21}
$$

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

Proof Of Proposition ??. Note that $(n-2)S \sim \mathrm{Wishart}_p(n - 2, \Sigma)$. Denote by $\Sigma = OEO^T$ the spectral decomposition of Σ, where O 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

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

Thus,

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

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

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

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

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

It follows that

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

Hence we have

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

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

$$\frac{1}{(p-r)(n-2)} \sum_{i=r+1}^{n-2} \lambda_i(Z^TEZ)$$
$$= \frac{1}{(p-r)(n-2)} \sigma^2 \sum_{i=1}^{n-2} \lambda_i(Z_{(2)}^T Z_{(2)}) + O_P(\frac{\max(n,p)}{np})$$
$$= \frac{1}{(p-r)(n-2)} \sigma^2 \text{tr}(Z_{(2)}^T Z_{(2)}) + O_P(\frac{\max(n,p)}{np})$$
$$= \sigma^2 + + O_P(\frac{1}{\sqrt{np}}) + O_P(\frac{\max(n,p)}{np}).$$

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

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

$\square$

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

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

By Assumption **??**, $\frac{\max(n,p)}{n\sqrt{p}} = \max(\frac{1}{\sqrt{p}}, \frac{\sqrt{p}}{n}) \to 0$. And

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

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, we only need to show that

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

We first deal with P_2. To 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

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

By the following useful relationship

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

and Lemma 5, we have that

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

As for P_3. To prove the convergence in probability, here 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 $N(0, \tau\hat{\hat{V}}^T\Sigma\hat{\hat{V}})$. As we have shown,

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

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

$$(p-2r)\sigma^4 \le \mathrm{tr}(\hat{\hat{V}}^T\Sigma\hat{\hat{V}})^2 \le \big(\frac{p}{n}+1\big)^2 O_P(1) + (p-2r)\sigma^4,$$

or

$$\mathrm{tr}(\hat{\hat{V}}^T\Sigma\hat{\hat{V}})^2 = p\sigma^4(1+o_P(1)). \tag{22}$$

It follows that

$$\frac{\lambda_1^2(\hat{\hat{V}}^T\Sigma\hat{\hat{V}})}{\mathrm{tr}(\hat{\hat{V}}^T\Sigma\hat{\hat{V}})^2} = O_P\Big(\frac{(p/n+1)^2}{p}\Big) = o_P(1). \tag{23}$$

Then for every subsequence of $\{n\}$, there's a further subsequence along which (23) holds almost surely. By Lemma 7, for every subsequence of $\{n\}$, there's a further subsequence along which we have

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

It means that

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

Thus the weak convergence also holds unconditionally:

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

Similar to (22) we have

$$\mathrm{tr}(\hat{\hat{V}}^T\Sigma\hat{\hat{V}})=(p-r)\sigma^2\big(1+O_P\big(\frac{1}{n}+\frac{1}{p}\big)\big). \tag{24}$$

By (22), (24) and Slulsk's theorem,

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

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

Proof Of Theorem **??**. By assumption, $\hat{r}\le R$ for some constant R. Similar to the proof of Proposition **??**, in the current context we have that $\mathrm{tr}(\hat{\hat{V}}_iS_i\hat{\hat{V}}_i)=\mathrm{tr}\Sigma+P_P(\frac{\max(n,p)}{n})$, $i=1,2$. 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 7,

$$\frac{\|\hat{\hat{V}}^T(\bar{X}_1-\bar{X}_2)\|^2-\tau\mathrm{tr}(\hat{\hat{V}}^T(\Sigma+\frac{1}{\sqrt{p}}\psi I_p)\hat{\hat{V}})}{\sqrt{2\tau^2\mathrm{tr}(\hat{\hat{V}}^T(\Sigma+\frac{1}{\sqrt{p}}\psi I_p)\hat{\hat{V}})^2}}\xrightarrow{\mathcal{L}}N(0,1). \tag{25}$$

It can be easily shown that

$$\frac{\mathrm{tr}(\hat{\hat{V}}^T(\Sigma+\frac{1}{\sqrt{p}}\psi I_p)\hat{\hat{V}})^2}{\mathrm{tr}(\hat{\hat{V}}^T\Sigma\hat{\hat{V}})^2}\xrightarrow{P}1. \tag{26}$$

Next we will show that

$$\frac{\mathrm{tr}(\hat{\hat{V}}^T\Sigma\hat{\hat{V}})^2}{\mathrm{tr}\Sigma^2}\xrightarrow{P}1. \tag{27}$$

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

$$\lambda_i(\hat{\hat{V}}^T\Sigma\hat{\hat{V}})=\lambda_i(\Sigma^{1/2}\hat{\hat{V}}\hat{\hat{V}}^T\Sigma^{1/2})\leq\lambda_i(\Sigma). \tag{28}$$

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

$$\lambda_i(\hat{\hat{V}}^T\Sigma\hat{\hat{V}})=\lambda_i(\Sigma^{1/2}(I_p-\hat{V}\hat{V}^T)\Sigma^{1/2})=\lambda_i(\Sigma-\Sigma^{1/2}\hat{V}\hat{V}^T\Sigma^{1/2})\geq\lambda_{i+\hat{r}}(\Sigma), \tag{29}$$

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 (28) and (29),

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

Then $|\mathrm{tr}(\hat{\hat{V}}^T\Sigma\hat{\hat{V}})^2-\mathrm{tr}\Sigma^2|\leq\sum_{i=1}^{\hat{r}}\lambda_i^2(\Sigma)\leq RC^2$. Hence (27) holds. By (25), (26), (27) and Slutsky's theorem,

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

Note that

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

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

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

And

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

by (28) and (29). The proof is completed.

$\square$

***Proof Of Theorem* ??.** Since $r=0$, $X_{ki} = \mu_k + Z_{ki}$, $i=1,\ldots,n_k$ and $k=1,2$. As in the proof of Theorem **??**, 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

$$\begin{aligned}&\|\hat{\hat{V}}^T(\bar{Z}_1-\bar{Z}_2) + \hat{\hat{V}}^T O(\mu_1-\mu_2)\|^2 \\ =&\|\hat{\hat{V}}^T(\bar{Z}_1-\bar{Z}_2)\|^2 + \frac{\|\hat{\hat{V}}^T\xi\|^2}{R^2}\|\mu_1-\mu_2\|^2 + \frac{\|\mu_1-\mu_2\|}{R}\xi^T\hat{\hat{V}}\hat{\hat{V}}^T(\bar{Z}_1-\bar{Z}_2).\end{aligned} \tag{30}$$

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 (30) follows by central limit theorem:

$$\frac{\|\hat{\hat{V}}^T(\bar{Z}_1-\bar{Z}_2)\|^2 - \tau(p-\hat{r})\sigma^2}{\sigma^2\sqrt{2\tau^2(p-\hat{r})}} \xrightarrow{\mathcal{L}} N(0,1). \tag{31}$$

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 (30), we have

$$\frac{\|\hat{\hat{V}}^T\xi\|^2}{R^2}\|\mu_1-\mu_2\|^2 = \|\mu_1-\mu_2\|^2+O_P(\frac{1}{\sqrt{p}})\|\mu_1-\mu_2\|^2 = \|\mu_1-\mu_2\|^2+o_P(\sigma^2\sqrt{2\tau^2 p}). \tag{32}$$

Now we deal wih the second term of (30). Note that $\xi^T\hat{\hat{V}}\hat{\hat{V}}^T(\bar{Z}_1-\bar{Z}_2)\big|\big(\hat{\hat{V}},(\bar{Z}_1-\bar{Z}_2)\big) \sim N(0,\|\hat{\hat{V}}^T(\bar{Z}_1-\bar{Z}_2)\|^2)$, 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

$$\frac{\|\mu_1-\mu_2\|}{R}\xi^T\hat{\hat{V}}\hat{\hat{V}}^T(\bar{Z}_1-\bar{Z}_2) = O_P(\sqrt{\tau})\|\mu_1-\mu_2\| = o_P(\sigma^2\sqrt{2\tau^2 p}). \tag{33}$$

By (30), (31), (32), (33) and Slutsky's theorem, we have the conclusion

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

□

Proof Of Theorem ??. The method of Theorem **??**'s proof can still work here with some modifications. The term P_3 in THeorem **??**'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,

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

But

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

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

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

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.

Davidson KR, Szarek SJ. Local operator theory, random matrices and banach spaces. "handbook in Banach Spaces" Vol 2001;:317–66.

Fan J, Wang W. Asymptotics of empirical eigen-structure for ultra-high dimensional spiked covariance model. Annals of Statistics 2017;.

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

Jung S, Marron JS. Pca consistency in high dimension, low sample size context. Annals of Statistics 2009;37(6):4104–30.

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.

Shen D, Shen H, Marron JS. A general framework for consistency of principal component analysis. ArXiv e-prints 2012;arXiv:1211.2671.

Shen D, Shen H, Marron JS. Consistency of sparse pca in high dimension, low sample size contexts. Journal of Multivariate Analysis 2013;115(1):317–333.