Statistica Sinica **27** (2017), 1419-1436
doi:https://doi.org/10.5705/ss.202015.0199

COMPOSITE T^2 TEST FOR HIGH-DIMENSIONAL DATA

Long Feng[1], Changliang Zou[1], Zhaojun Wang[1] and Lixing Zhu[2]

[1]*Nankai University and* [2]*Hong Kong Baptist University*

Abstract: We consider high-dimensional location test problems in which the number of variables p may exceed the sample size n. The classical T^2 test does not work well because the contamination bias in estimating the covariance matrix grows rapidly with p. Unlike most existing remedies abandoning all the correlation information, the composite T^2 test developed here makes use of them in a practical and efficient way. Under mild conditions, the proposed test statistic is asymptotically normal, and allows the dimensionality to almost exponentially increase in n. The test inherits certain appealing features of the classical T^2 test and does not suffer from large bias contamination. Due to incorporating much correlation information, the proposed test can deliver more robust performance than existing methods in many cases. Simulation studies demonstrate the validity of asymptotic analysis.

Key words and phrases: Asymptotic normality, composite T^2 test, high-dimensional data, large-p-small-n.

1. Introduction

Assume that $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ are independent and identically distributed random p-vectors from a distribution $F(\boldsymbol{x} - \boldsymbol{\mu})$ located at p-variate center $\boldsymbol{\mu}$. The classic one-sample testing problem is

$$H_0 : \boldsymbol{\mu} = \mathbf{0} \quad \text{versus} \quad H_1 : \boldsymbol{\mu} \neq \mathbf{0}. \tag{1.1}$$

The classic test statistic is the Hotelling's $T^2 = n\bar{\boldsymbol{X}}^T\hat{\boldsymbol{\Sigma}}^{-1}\bar{\boldsymbol{X}}$ where $\bar{\boldsymbol{X}}$ is the sample mean vector and $\hat{\boldsymbol{\Sigma}}$ is the sample covariance matrix, but it cannot be applied to the so-called large-p-small-n paradigm ($p > n-1$) due to the singularity of $\hat{\boldsymbol{\Sigma}}$. One could replace $\hat{\boldsymbol{\Sigma}}$ with its nonsingular diagonal matrix (Srivastava (2009); Park and Ayyala (2013)) or an identity matrix (Bai and Saranadasa (1996); Chen and Qin (2010)), but these tests lose all the information of the correlations between those variables. One could replace $\hat{\boldsymbol{\Sigma}}$ by a sparse matrix estimator (Bickel and Levina (2008); Cai and Liu (2011)), but it is difficult to maintain the significant level for such modified test statistics (Feng, Zou and Wang (2015)) because of the contamination bias that grows rapidly with p. Chen et al. (2011) propose a regularized Hotelling's T^2 test, $n\bar{\boldsymbol{X}}^T(\hat{\boldsymbol{\Sigma}}+\lambda\mathbf{I}_p)^{-1}\bar{\boldsymbol{X}}$, $\lambda > 0$, by stabilizing the inverse of $\hat{\mathbf{\Sigma}}$. But, the size and power of their test are deeply impacted by the choice of λ and the sparsity of $\mathbf{\Sigma}$.

We propose another test, called the composite T^2 test. Its first step is to sequentially select the K variables that have the largest correlation among all combinations of K elements from the remaining variables. We group the variables in many blocks and let the correlation between those blocks be rather small. Then we construct p/K Hotelling T^2 test statistics and combine them. The asymptotic normality of the proposed test can be derived under some mild conditions. We allows the dimensionality to increase almost exponentially with n. We derive the asymptotic relative efficiency of our test with the Park and Ayyala (2013) test. Our test performs better in most cases and simulation support this.

The remainder of the paper is organized as follows. In the next section, the test statistic is constructed and its asymptotic normality is established. We extend our method to the two-sample problem in Section 3. Simulations are represented on Section 4. Technical details are provided in the Appendix.

2. One Sample Problem

2.1. The test statistic

In high-dimensional settings, the classic Hotelling T^2 cannot work because the sample covariance matrix $\hat{\mathbf{\Sigma}}$ is not invertible. However, we can divide the p variables into several small parts for which the covariance matrix is invertible and then sum the Hotelling T^2 test statistics.

$$W_n = \sum_{i=1}^{N} T_{A_i}^2 = \sum_{i=1}^{N} n\bar{\boldsymbol{X}}_{A_i}^T \mathbf{S}_{A_i}^{-1} \bar{\boldsymbol{X}}_{A_i},$$

where $A_1 \cup \cdots A_N = \{1, \cdots, p\}$, $A_i \cap A_j = \emptyset$ and $\bar{X}_{A_i}$, $\mathbf{S}_{A_i}$ are the sample mean vector and covariance matrix of $X_{st}, t \in A_i$, $s = 1, \cdots, n$. We might choose those subsets from some available prior information. For example, in multi-sensor detection problem, the sensors located in the same spatial point could be naturally grouped together. When no preference is given, we suggest fixing the subsets with the same sizes, $|A_i| = K = [p/N]$, $i = 1, \cdots, N-1$, and $|A_N| = p - (N-1)K$, and strong correlated in the subset with correlations between subsets as weak as possible. We propose an algorithm to divide the variables.

For any symmetric matrix $\mathbf{B} = (b_{ij}) \in \mathbb{R}^{q\times q}$, $\|\mathbf{B}\|_{l_1} = \sum_{1\le i,j\le q} |b_{ij}|$. For a subset $A \subset \{1, \cdots, q\}$, let $\mathbf{B}_A = (a_{ij}) \in \mathbb{R}^{q\times q}$ with $a_{ij} = b_{ij}$ if $i, j \in A$ and

$a_{ij} = 0$ if i or $j \notin A$. For a set of subsets $\mathcal{C} = \{C_1, \cdots, C_s\}$, $\mathbf{B}_{\mathcal{C}} = (c_{ij}) \in \mathbb{R}^{q\times q}$ denotes the "submatrix" of $\mathbf{B}$ with $c_{ij} = b_{ij}$ if $i, j \in C_k, k = 1, \cdots, s$, and 0 otherwise.

Consider the matrix $\mathbf{R}^0 \in \mathbb{R}^{p\times p}$.

Algorithm 1

Step 1. Find the initial subset $A_1 = \underset{A\subset\{1,\cdots,p\},|A|=K}{\text{argmax}} \|\mathbf{R}^0_A\|_{l_1}$.

Step 2. Suppose $A_1, \cdots, A_i$ have been selected and let $A_{-i} = \{1, \cdots, p\}\backslash \bigcup_{k=1}^{i} A_k$. Find $A_{i+1} = \underset{A\subset A_{-i},|A|=K}{\text{argmax}} \|\mathbf{R}^0_A\|_{l_1}$, $i = 1, \cdots, N-2$.

Step 3. Take $A_N = A_{-(N-1)}$.

Remark 1. If we search all the submatrices with size K, the computation burden of Step 1 would be $O(p^K)$. In practice, we suggest the following algorithm. First, find $\{a_1, a_2\} = \underset{1\le i<j\le p}{\text{argmax}} |\text{cor}(X_{li}, X_{lj})|$. Then, find the k-th variable with the largest correlation with $\{a_1, \cdots, a_{k-1}\}$ in the remaining subsets $\{1, \cdots, p\}\backslash\{a_1, \cdots, a_{k-1}\}$, $a_k = \underset{i\in\{1,\cdots,p\}\backslash\{a_1,\cdots,a_{k-1}\}}{\text{argmax}} \sum_{j=1}^{k-1} |\text{cor}(X_{li}, X_{la_j})|$. Denote the result subset by A_1'. We also use the same algorithm in Step 2, and have good performance in practice.

Let $A_{n1}, \cdots, A_{nN}$ be the selected sets by the algorithm based on the sample correlation matrix $\hat{\mathbf{R}}$. Then we can write the test statistic W_n as $W_n = n\bar{\boldsymbol{X}}^T\hat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{O}_n^K}\bar{\boldsymbol{X}}$, where $\mathcal{O}_n^K = \{A_{n1}, \cdots, A_{nN}\}$. There is no explicit form of the expectation of W_n under the null hypothesis. When p gets larger, there is a non-negligible bias term because $\hat{\boldsymbol{\Sigma}}_{\mathcal{O}_n^K}$ is not independent of $\bar{\boldsymbol{X}}$ and the sample mean and variance is only root-n consistent (Feng et al. (2015)).

Similar to Feng and Sun (2015), we consider a test statistic based on the leave out method (abbreviated as CT hereafter):

$$T_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j\ne i}^{n} \boldsymbol{X}_i^T \widehat{\boldsymbol{\Sigma}^{(i,j)}_{\mathcal{O}_{ij}^K}}^{-1} \boldsymbol{X}_j, \tag{2.1}$$

where $\hat{\boldsymbol{\Sigma}}^{(i,j)}$, $\hat{\mathbf{R}}^{(i,j)}$ are the corresponding sample covariance and correlation matrixs of $\{\boldsymbol{X}_k\}_{k\ne i,j}$, respectively, $\mathcal{O}_{ij}^K$ are the corresponding selected sets based on $\hat{\mathbf{R}}^{(i,j)}$. Now $\boldsymbol{X}_i$, $\widehat{\boldsymbol{\Sigma}^{(i,j)}_{\mathcal{O}_{ij}^K}}$, and $\boldsymbol{X}_j$ are independent from each other, and the expectation of T_n is zero under the null hypothesis.

2.2. Results

Following Bai and Saranadasa (1996) and Chen and Qin (2010) model:

$$\boldsymbol{X}_i = \boldsymbol{\Gamma}\mathbf{z}_i + \boldsymbol{\mu} \quad \text{for} \quad i = 1, \cdots, n, \tag{2.2}$$

where each $\mathbf{\Gamma}$ is a $p\times m$ matrix for some $m\geq p$ such that $\mathbf{\Gamma}\mathbf{\Gamma}^T=\mathbf{\Sigma}$, and $\{\mathbf{z}_i\}_{i=1}^n$ are m-variate independent and identically distributed random vectors with

$$E(\mathbf{z}_i)=0,\ \mathrm{var}(\mathbf{z}_i)=\mathbf{I}_m,\ E(z_{il}^4)=3+\Delta,\ E(z_{il}^8)=m_8\in(0,\infty),$$
$$E(z_{ik_1}^{\alpha_1}z_{ik_2}^{\alpha_2}\cdots z_{ikq}^{\alpha_q})=E(z_{ik_1}^{\alpha_1})E(z_{ik_2}^{\alpha_2})\cdots E(z_{ikq}^{\alpha_q}), \tag{2.3}$$

whenever $\sum_{k=1}^q\alpha_k\leq 8$ and $k_1\neq k_2\cdots\neq k_q$. The data structure (2.3) generates a rich collection of $\boldsymbol{X}_i$ from $\mathbf{z}_i$ with a given covariance. We need the some conditions when as $n,p\to\infty$,

(C1) $\varpi_{\min}=\min_{1\leq k\leq N}\varpi_k>\omega$, $\varpi_k=\min\limits_{\substack{A\subset\{1,\cdots,p\}\backslash\{A_1^o\cup\cdots\cup A_{k-1}^o\}\\ |A|=K,A\neq A_k^o}}(\lambda_{A_k^o}-\lambda_A)/(\sigma_A+\sigma_{A_k^o})$, where ω is a positive constant and $\lambda_A=\|\mathbf{R}_A\|_{l_1}$. σ_A^2 is the asymptotic variance of $\sqrt{n}\|\hat{\mathbf{R}}_A\|_{l_1}$.

(C2) $\mathrm{tr}(\mathbf{\Lambda}_K^4)=o(\mathrm{tr}^2(\mathbf{\Lambda}_K^2))$, where $\mathbf{\Lambda}_K=\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}_{\mathcal{O}^K}^{-1}\mathbf{\Sigma}^{1/2}$ and $\mathcal{O}^K=\{A_1^o,\cdots,A_N^o\}$ are the selected sets based on the true correlation matrix $\mathbf{R}$.

(C3) $\boldsymbol{\mu}^T\mathbf{\Sigma}_{\mathcal{O}^K}^{-1}\mathbf{\Sigma}\mathbf{\Sigma}_{\mathcal{O}^K}^{-1}\boldsymbol{\mu}=o(n^{-1}\mathrm{tr}(\mathbf{\Lambda}_K^2))$ and $(\boldsymbol{\mu}^T\mathbf{\Sigma}_{\mathcal{O}^K}^{-1}\boldsymbol{\mu})^2=o((\log p)^{-1/2}n^{-3/2}\mathrm{tr}(\mathbf{\Lambda}_K^2))$.

(C4) $\log p=o(n)$.

Condition (C1) is a technical condition to make the partition in Algorithm 1 identifiable. To appreciate Condition (C2), when $K=1$. (C2) then becomes $\mathrm{tr}(\mathbf{R}^4)=o\{\mathrm{tr}^2(\mathbf{R}^2)\}$, which is similar to Condition (3.7) in Chen and Qin (2010). If $\lambda_1\leq\cdots\leq\lambda_p$ are the eigenvalues of $\mathbf{\Lambda}_K$ and $\nu_k=\sum_{i=1}^p\lambda_i^k$, (C2) is $\nu_4=o(\nu_2^2)$. If all eigenvalues of $\mathbf{\Lambda}_K$ are bounded, $\nu_4=O(p)$ and $\nu_2=O(p)$, (C2) is trivially true; (C3) is $\|\boldsymbol{\mu}\|^2=O(n^{-1}p^{1/2})$, which can be viewed as a high-dimensional version of the local alternative hypotheses.

Proposition 1. *Under* (C1)-(C4)*, we have*

$$P\left(\bigcap_{1\leq i<j\leq n}\left\{\mathcal{O}_{ij}^K=\mathcal{O}^K\right\}\right)=1-O(n^{3/2}p^{K+1}e^{-n\omega^2/2}).$$

Theorem 1. *Under* (C1)-(C4)*, we have*

$$\frac{T_n-\boldsymbol{\mu}^T\mathbf{\Sigma}_{\mathcal{O}^K}^{-1}\boldsymbol{\mu}}{\sqrt{2n^{-2}\mathrm{tr}(\mathbf{\Lambda}_K^2)}}\xrightarrow{\mathcal{L}}N(0,1).$$

To construct a test procedure, we propose a ratio-consistent estimator of $\mathrm{tr}(\mathbf{\Lambda}_K^2)$,

$$\widehat{\mathrm{tr}(\mathbf{\Lambda}_K^2)} = \frac{1}{2P_n^4}\sum^{*}(\boldsymbol{X}_{i_1}-\boldsymbol{X}_{i_2})^T \widehat{\mathbf{\Sigma}_{\mathcal{O}^K_{i_1,i_2,i_3,i_4}}^{(i_1,i_2,i_3,i_4)}}^{-1}(\boldsymbol{X}_{i_3}-\boldsymbol{X}_{i_4}) \times(\boldsymbol{X}_{i_1}-\boldsymbol{X}_{i_4})^T \widehat{\mathbf{\Sigma}_{\mathcal{O}^K_{i_1,i_2,i_3,i_4}}^{(i_1,i_2,i_3,i_4)}}^{-1}(\boldsymbol{X}_{i_3}-\boldsymbol{X}_{i_2}), \tag{2.4}$$

where $\mathcal{O}^K_{i_1,i_2,i_3,i_4}$ are the selected sets based on $\mathbf{R}^{\widehat{(i_1,i_2,i_3,i_4)}}$, and $\mathbf{\Sigma}^{\widehat{(i_1,i_2,i_3,i_4)}}$, $\mathbf{R}^{\widehat{(i_1,i_2,i_3,i_4)}}$ are the corresponding sample covariance and correlation matrix of $\{\boldsymbol{X}_k\}_{k\neq i_1,i_2,i_3,i_4}$, respectively. Throughout, we use $\sum^*$ to denote summations over distinct indexes. For example, in $\widehat{\mathrm{tr}(\mathbf{\Lambda}_K^2)}$, the summation is over the set $\{i_1\neq i_2\neq i_3\neq i_4\}$, for all $i_1,i_2,i_3,i_4\in\{1,\cdots,n\}$ and $P_n^m=n!/(n-m)!$.

Proposition 2. *Under* (C1), (C2) *and* (C4), *as* $n,p\to\infty$,

$$\frac{\widehat{\mathrm{tr}(\mathbf{\Lambda}_K^2)}}{\mathrm{tr}(\mathbf{\Lambda}_K^2)}\xrightarrow{p}1.$$

This result suggests rejecting H_0 with α level of significance if $T_n/\sqrt{2n^{-2}\widehat{\mathrm{tr}(\mathbf{\Lambda}_K^2)}}>z_\alpha$, where z_α is the upper α quantile of $N(0,1)$.

According to Theorem 1, the power under the local alternative (C3) is

$$\beta_{CT}(\boldsymbol{\mu})=\Phi\left(-z_\alpha+\frac{\boldsymbol{\mu}^T\mathbf{\Sigma}_{\mathcal{O}^K}^{-1}\boldsymbol{\mu}}{\sqrt{2n^{-2}\mathrm{tr}(\mathbf{\Lambda}_K^2)}}\right),$$

where $\Phi(\cdot)$ is the standard normal distribution function. The performance of the proposed test relies upon the choice of K. The optimal choice of K is the maximizer of β_{CT}, but $\boldsymbol{\mu}$ is unknown. For simplicity, we only illustrate the procedure with $K=2$.

Remark 2. In practice, if we know of the correlation between variables, the should be combined. For example, if we know some genes express a trait together, we should combine them in a subset. The choice of K in practice deserves attention. When the correlations between those variables are strong, we need to use a large K, but generally a small K is preferable. More information in the simulation studies.

Park and Ayyala (2013) showed that the power of their test (abbreviated as PA hereafter) is

$$\beta_{PA}(\boldsymbol{\mu})=\Phi\left(-z_\alpha+\frac{\boldsymbol{\mu}^T\mathbf{D}^{-1}\boldsymbol{\mu}}{\sqrt{2n^{-2}\mathrm{tr}(\mathbf{R}^2)}}\right),$$

where $\mathbf{D}$ is the diagonal matrix of $\mathbf{\Sigma}$. It is difficult to propose compare our test with that of Park and Ayyala under general settings.

3. Two Sample Problem

In this section, we extend our proposed test to the two-sample case (Chen and Qin (2010); Cai, Liu and Xia (2014); Feng et al. (2015); Gregory et al. (2015)). Let $\boldsymbol{X}_{ij}$, $j = 1, \cdots, n_i$, $i = 1, 2$, be independent p-dimensional multivariate random vectors from the diverging factor model as (2.3) with mean $\boldsymbol{\mu}_i$ and unknown common covariance matrix $\boldsymbol{\Sigma}$.

We extend the test statistic T_n in (2.1) to the two-sample case

$$Q_n = \frac{1}{n_1 n_2 (n_1 - 1)(n_2 - 1)} \sum_{1 \le i_1 \ne i_2 \le n_1} \sum_{1 \le j_1 \ne j_2 \le n_2} (\boldsymbol{X}_{1i_1} - \boldsymbol{X}_{2j_1})^T \boldsymbol{\Sigma}_{\mathcal{O}^K_{i_1,i_2,j_1,j_2}}^{\widehat{(i_1,i_2,j_1,j_2)}^{-1}} (\boldsymbol{X}_{1i_2} - \boldsymbol{X}_{2j_2}), \quad (3.1)$$

where $\boldsymbol{\Sigma}^{\widehat{(i_1,i_2,j_1,j_2)}}, \mathbf{R}^{\widehat{(i_1,i_2,j_1,j_2)}}$ are the corresponding pooled sample covariance matrix and correlation matrix of the sample $\{\boldsymbol{X}_{1k}\}_{k \ne i_1,i_2}$ and $\{\boldsymbol{X}_{2l}\}_{l \ne j_1,j_2}$, respectively, and $\mathcal{O}^K_{i_1,i_2,j_1,j_2}$ are the selected sets based on $\mathbf{R}^{\widehat{(i_1,i_2,j_1,j_2)}}$ by Algorithm 1.

For $n = n_1 + n_2$ and $n_1/n \to \kappa \in (0, 1)$, we consider the alternative hypothesis.

(C5) $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = o(n^{-1} \mathrm{tr}(\boldsymbol{\Lambda}_K^2))$ and $((\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))^2 = o((\log p)^{-1/2} n^{-3/2} \mathrm{tr}(\boldsymbol{\Lambda}_K^2))$.

Theorem 2. *Under* (C1), (C2), (C4), *and* (C5), *as* $n, p \to \infty$, *we have*

$$\frac{Q_n - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{\sqrt{2(n_1^{-1} + n_2^{-1})^2 \mathrm{tr}(\boldsymbol{\Lambda}_K^2)}} \xrightarrow{\mathcal{L}} N(0, 1).$$

For simplicity, we only use the first sample to estimate $\mathrm{tr}(\boldsymbol{\Lambda}_K^2)$ by (2.4), and then we reject H_0 with α level of significance if $Q_n / \sqrt{2(n_1^{-1} + n_2^{-1})^2 \widehat{\mathrm{tr}(\boldsymbol{\Lambda}_K^2)}} > z_\alpha$.

4. Simulation

4.1. One sample problem

4.1.1. Large-p-small-n case

Here we report a simulation study designed to evaluate the performance of our proposed test (abbreviated as CT_2 with $K = 2$ and CT_5 with $K = 5$). We compared our test with the methods proposed by Chen et al. (2011) (abbreviated as RHT) and Park and Ayyala (2013) and various covariance matrices.

(I) $\boldsymbol{\Sigma} = (\sigma_{ij})$, $\sigma_{ii} \sim U(0, 1)$, $i = 1, \cdots, p$ and $\sigma_{ij} = 0$ for $i \ne j$;

(II) $\boldsymbol{\Sigma} = (\sigma_{ij})$, $\sigma_{2k-1,2k} = \sigma_{2k,2k-1} = 0.8$, $k = 1, \cdots, p/2$ and $\sigma_{ii} = 1$, $i =$

$1, \cdots, p$;

(III) $\mathbf{\Sigma} = (\sigma_{ij})$, $\sigma_{2k-1,2k} = \sigma_{2k,2k-1} = -0.8$, $k = 1, \cdots, p/2$ and $\sigma_{ii} = 1$, $i = 1, \cdots, p$;

(IV) $\mathbf{\Sigma} = (\sigma_{ij})$, $\sigma_{ij} = 0.8^{|i-j|}$;

(V) $\mathbf{\Sigma} = (\sigma_{ij})$, $\sigma_{ij} = (-0.8)^{|i-j|}$;

(VI) $\mathbf{\Sigma} = (\sigma_{ij})$, $\sigma_{ij} = 0.2$ for $i \neq j$ and $\sigma_{ii} = 1$, $i = 1, \cdots, p$;

(VII) $\mathbf{\Sigma} = (\sigma_{ij})$, $\sigma_{ij} = 0.9$ for $i \neq j$ and $\sigma_{ii} = 1$, $i = 1, \cdots, p$.

We considered that $\boldsymbol{X}$ was (a) $N(\boldsymbol{\mu}, \mathbf{\Sigma})$; (b) multivariate t-distribution $MT(\boldsymbol{\mu}, \mathbf{\Sigma}, 5)$; (c) multivariate chisquare, $\boldsymbol{X} = \boldsymbol{\mu} + \mathbf{\Sigma}^{-1/2}\mathbf{Z}$, where $\mathbf{Z} = (Z_{ij})_{1 \leq i,j \leq p}$, Z_{ij} was centered χ^2_4. For the alternative hypothesis, we considered two patterns for $\boldsymbol{\mu} = \kappa(\mu_1, \cdots, \mu_p)$. As random cases: (i) $\mu_i \sim N(0,1)$, $i = 1, \cdots, p$; (ii) randomly half of the μ_i were $N(0,1)$ and the others zero; (iii) randomly $[0.05p]$ of the μ_i were $N(0,1)$ and the others zero. As fixed cases, we took (iv) $\mu_i = 1$, $i = 1, \cdots, p$; (v) $\mu_{2k-1} = 1, \mu_{2k} = -1$, $k = 1, \cdots, p/2$; (vi) $\mu_{2k-1} = 1, \mu_{2k} = 0$, $k = 1, \cdots, p/2$; (vii)] $\mu_i = 1$, $i = 1, \cdots, [0.05p]$ and the others zero.

To make the power comparable among the configurations of H_1, the coefficient κ was selected so that the signal-to-noise was $\boldsymbol{\mu}^T \mathbf{\Sigma}^{-1} \boldsymbol{\mu} = 1.5$ throughout. We took $(n, p) = (30, 100)$ or $(40, 200)$.

Tables 1-3 report the results of the three tests under different distributions and choices of $\boldsymbol{\mu}$ in the one-sample case. The PA test and ours have reasonable sizes in most cases. The RHT test did not control the empirical size very well, especially when the correlations between the variables were large. Chen et al. (2011) use the shrinkage estimator $(\hat{\mathbf{\Sigma}} + \lambda \mathbf{I}_p)^{-1}$ to estimate the inverse of covariance matrix $\mathbf{\Sigma}^{-1}$ in their test statistic. Thus, if the difference between $\mathbf{\Sigma}$ and $\lambda \mathbf{I}_p$ is very small, RHT performs very well, such as Case (I); if $\mathbf{\Sigma}$ is not very sparse, the power of the RHT test is smaller than the other tests.

The results together suggest that the CT test is quite robust and efficient in testing the shift of locations, especially when there are strong correlations between the variables. If the correlations between the variables are not large, our test outperforms the PA test when the direction of location shift is contrary to the correlation between the variables, and vice versa. If the direction of location shift is random, our test is more efficient than the PA test.

4.1.2. Large-n-small-p case

We considered the large-n-small-p case to compare our test, CT_2, with the

Table 1. Empirical sizes and power (%) comparisons under the multivariate normal distribution in the one-sample case.

	PA	RHT	CT_2	CT_5	PA	RHT	CT_2	CT_5	PA	RHT	CT_2	CT_5	PA	RHT	CT_2	CT_5
	Random Cases															
Model	size				(i)				(ii)				(iii)			
	$n=30,p=100$															
(I)	5.1	5.2	5.1	5.3	70	65	72	75	73	69	72	69	69	63	72	67
(II)	4.7	4.3	4.9	6.2	20	15	79	48	20	16	77	51	20	16	76	41
(III)	4.3	4.5	4.4	3.9	20	18	79	51	21	18	79	57	21	17	78	57
(IV)	7.1	1.2	6.1	6.2	10	2.5	31	42	10	2.3	30	55	10	2.6	29	45
(V)	4.7	1.4	4.7	5.4	10	4.1	30	51	10	3.7	29	52	9.5	3.1	31	48
(VI)	7.1	1.3	6.3	3.8	21	23	43	61	23	17	44	57	22	23	45	56
(VI)	5.6	0.4	4.6	4.6	6.4	22	7.3	29	6.1	20	6.5	33	5.3	31	5.6	31
	$n=40,p=200$															
(I)	6.1	4.5	5.3	5.2	71	61	71	77	72	65	72	69	71	61	72	73
(II)	6.3	2.1	4.2	3.9	20	13	79	55	19	13	79	53	18	13	78	55
(III)	6.2	4.2	4.1	4.1	20	14	80	54	20	14	77	47	18	12	79	45
(IV)	7.4	1.3	6.1	4.8	10	1.3	30	51	9.3	2.2	29	37	10	1.2	29	50
(V)	4.3	2.4	5.4	4.7	9.6	2.1	29	45	9.2	3.1	30	46	11	2.3	30	48
(VI)	7.1	1.1	6.6	6.2	15	14	31	55	16	11	33	37	15	1.7	33	52
(VI)	4.6	0.1	4.3	4.9	8.3	12	9.5	20	4.6	0.4	5.0	19	6.1	15	6.9	24
	Fixed Cases															
	(iv)				(v)				(vi)				(vii)			
	$n=30,p=100$															
(I)	70	5.3	69	70	75	67	73	66	68	28	72	70	72	62	73	72
(II)	94	2.1	79	81	13	10	78	53	21	8.3	79	47	63	44	76	68
(III)	10	5.5	78	49	94	85	78	79	20	10	78	56	10	11	77	43
(IV)	100	0.4	100	99	7.2	1.3	24	35	10	1.3	31	45	59	24	65	57
(V)	7.3	1.3	24	47	100	93	100	99	9.1	2.2	27	45	8.3	3.7	28	41
(VI)	100	1.2	100	100	22	23	45	58	48	13	63	75	24	26	45	64
(VII)	41.6	0.5	41.3	100	5.0	20	6.0	26	5.3	12	6.7	29	6.0	25	7.9	30
	$n=40,p=200$															
(I)	69	4.7	72	62	73	64	73	71	73	23	75	75	67	65	69	74
(II)	94	1.5	76	77	11	7.4	81	42	19	6.6	78	53	95	84	77	87
(III)	12	4.4	80	56	97	88	79	85	20	10	79	49	11	10	80	39
(IV)	100	1.2	100	98	6.5	1.1	24	43	8.2	1.2	27	50	88	45	85	83
(V)	7.1	2.3	23	45	100	99	100	99	12	2.4	30	53	9.6	2.2	25	43
(VI)	100	1.1	100	100	15	11	32	51	36	5.1	52	65	18	13	34	49
(VII)	54	0.0	53.3	100	6.3	0.7	7.7	19	3.3	0.8	4.8	21	5.2	11	6.5	19

classic Hotelling's T^2 test (abbreviated HT hereafter) and the PA test. The settings were the same as those in Section 4.1.1, except $n = 50, p = 4$. We considered only the multivariate normal. Table 4 reports the results. For models (I)-(III), our test CT_2 performs similar to the HT test since $\mathbf{\Sigma}_{\mathcal{O}^K} = \mathbf{\Sigma}$. For the other models, the HT test is more powerful than CT_2 because of the loss of the information of some correlation of variables. The CT_2 test outperforms the PA

Table 2. Empirical sizes and power (%) comparisons under the multivariate t distribution in the one-sample case.

Model	PA	RHT	CT_2	CT_5	PA	RHT	CT_2	CT_5	PA	RHT	CT_2	CT_5	PA	RHT	CT_2	CT_5
	Random Cases															
Model	size				(i)				(ii)				(iii)			
	$n=30, p=100$															
(I)	5.1	2.2	5.7	4.3	70	53	72	77	69	47	72	69	70	48	72	70
(II)	5.2	0.2	4.4	5.1	20	4.3	74	42	22	5.1	76	47	19	1.8	73	52
(III)	5.4	1.8	4.7	5.3	20	8.1	75	53	23	13	77	49	19	9.1	74	45
(IV)	7.1	0.1	5.3	5.4	11	0.1	30	41	10	0.6	29	39	9	0.7	32	35
(V)	6.3	0.0	5.2	4.7	11	1.5	33	43	9	2.1	32	40	10	0.8	32	45
(VI)	7.2	0.0	5.9	5.9	24	15	47	64	26	12	49	61	26	13	53	57
(VII)	6.3	0.0	6.7	3.9	6.3	3.1	6.3	28	6.0	5.6	5.0	36	3.7	5.3	4.7	29
	$n=40, p=200$															
(I)	6.1	0.2	4.7	3.9	70	37	73	73	69	34	71	70	67	46	72	73
(II)	6.6	1.3	5.1	4.1	20	3.3	76	49	17	3.2	75	48	19	2.3	75	49
(III)	5.2	0.0	4.7	5.6	19	2.7	73	53	18	3.5	77	54	18	2.8	73	42
(IV)	5.8	0.1	5.8	6.3	9	1.1	33	51	10	1.7	31	35	9	1.1	26	41
(V)	5.7	0.1	5.5	5.9	9	1.5	26	48	10	1.1	29	47	8	1.3	29	40
(VI)	6.5	0.1	6.2	4.1	17	5.3	37	58	17	7.1	41	49	18	5.6	40	65
(VII)	9.7	0.1	9.3	4.1	10	1.5	11	19	5.3	1.8	5.3	18	5.7	1.1	6.0	20
	Fixed Cases															
	(iv)				(v)				(vi)				(vii)			
	$n=30, p=100$															
(I)	69	1.1	72	70	72	57	74	73	71	16	73	69	62	56	69	68
(II)	90	2.3	74	77	11	2.3	75	37	21	2.1	74	53	58	38	72	68
(III)	12	1.7	77	43	89	86	77	81	17	1.4	74	48	12	4.3	73	48
(IV)	99	0.6	98	98	8.4	0.3	27	38	11	0.3	32	39	59	31	64	61
(V)	8.3	1.0	25	35	99	96	99	97	12	1.1	31	48	7.4	2.2	25	38
(VI)	100	0.2	100	99	25	16	50	61	52	7.2	66	73	27	20	51	65
(VII)	42	0.2	41	100	4.7	5.3	4.7	45	8.3	4.8	7.3	38	7.7	4.1	7.7	37
	$n=40, p=200$															
(I)	69	0.0	71	79	70	35	72	79	69	9.6	73	70	71	29	69	75
(II)	91	0.1	75	78	10	3.6	74	41	19	2.7	75	54	92	70	75	85
(III)	11	0.3	74	43	91	83	75	82	19	1.1	76	47	12	0.0	76	50
(IV)	100	0.4	99	99	6.8	1.1	25	40	9	0.6	29	37	83	48	83	83
(V)	8.5	0.1	26	41	99	100	100	99	10	0.7	29	35	7	0.0	27	39
(VI)	100	0.0	100	100	16	6.2	39	56	36	3.2	51	60	16	5.6	37	58
(VII)	55	0.0	54	100	6.1	1.3	6.9	19	5.3	0.8	5.7	22	4.8	0.5	4.4	21

test for the models (II)-(V) in most cases.

4.2. Two sample problem

We compared our test CT_2 with the PA test, the RHT test, Cai, Liu and Xia (2014)'s test (abbreviated as CLX test) and Gregory et al. (2015)'s test (abbreviated as GCT test) in the two-sample case. Here we considered $X_{1i} \sim N(\mathbf{0}, \mathbf{\Sigma})$, $i = 1, \cdots, n_1$ and $X_{2j} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma})$, $j = 1, \cdots, n_2$. We only considered

Table 3. Empirical sizes and power (%) comparisons under the multivariate chisquare distribution in the one-sample case.

	PA	RHT	CT_2	CT_5	PA	RHT	CT_2	CT_5	PA	RHT	CT_2	CT_5	PA	RHT	CT_2	CT_5
	Random Cases															
Model	size				(i)				(ii)				(iii)			
	$n=30, p=100$															
(I)	5.8	3.9	6.8	5.9	67	67	67	59	71	69	69	58	67	66	66	60
(II)	5.1	4.1	6.1	6.3	20	18	72	51	19	12	75	47	18	20	70	44
(III)	4.3	3.3	6.9	4.7	20	17	72	47	18	20	72	43	19	21	71	49
(IV)	5.2	1.1	6.5	6.2	9.3	2.2	30	46	8.6	1.1	28	44	10	3.4	31	51
(V)	5.0	1.0	6.0	6.4	11	3.0	28	43	10	2.6	32	42	10	3.5	29	39
(VII)	6.9	0.2	5.3	5.1	26	23	47	59	23	24	48	54	25	30	45	54
(VI)	5.0	1.3	5.8	4.3	5.0	26	6.2	33	3.1	28	4.3	31	8.3	24	8.7	32
	$n=40, p=200$															
(I)	5.9	4.7	6.1	6.6	69	58	67	68	72	63	71	73	67	63	66	68
(II)	5.7	5.2	6.9	5.2	19	14	77	47	16	16	75	56	19	16	72	49
(III)	6.0	7.1	6.0	6.8	20	19	73	53	18	19	74	53	19	27	72	46
(IV)	6.1	0.6	6.5	4.3	9.6	2.6	28	52	7.1	2.9	29	47	10	1.9	31	45
(V)	6.0	2.3	6.3	5.2	7.0	1.2	28	41	8.2	2.3	28	45	10	5.1	29	49
(VI)	6.3	1.7	5.8	6.1	14	11	35	46	16	12	34	56	13	11	34	47
(VII)	6.7	0.1	6.3	5.1	5.3	14	5.3	20	7.7	14	9.0	14	4.0	11	4.3	17
	Fixed Cases															
	(iv)				(v)				(vi)				(vii)			
	$n=30, p=100$															
(I)	70	4.3	54	57	69	62	69	61	67	30	63	58	64	70	70	67
(II)	93	2.7	81	77	10	7.3	71	45	22	6.3	77	45	57	48	75	57
(III)	12	3.5	79	37	92	73	71	74	18	4.1	81	47	11	11	76	46
(IV)	100	0.2	100	99	8	1.9	24	39	10	0.6	32	44	60	24	67	59
(V)	9.7	1.9	24	47	100	94	100	96	10	1.7	28	40	7.6	2.1	27	49
(VI)	100	0.6	100	100	25	19	47	60	47	12	63	75	28	28	50	67
(VII)	42	0.1	44	100	5.0	21	6.5	27	3.5	13	4.7	39	5.0	20	6.2	35
	$n=40, p=200$															
(I)	69	5.1	53	49	71	65	70	72	69	30	62	64	67	63	71	62
(II)	94	3.4	81	74	11	8.4	73	49	20	7.2	77	51	90	85	76	81
(III)	12	8.5	79	43	95	79	75	84	21	6.9	82	47	12	12	75	49
(IV)	100	0.6	100	100	6	0.0	23	38	11	0.4	28	47	82	50	87	84
(V)	5.3	2.1	21	39	100	100	100	100	11	2.1	27	47	7.3	2.7	24	41
(VI)	100	0.8	100	100	16	14	34	51	34	14	51	68	15	14	36	45
(VII)	57	0.3	56	100	5.7	9.5	5.3	22	7.7	11	8.4	20	6.1	7.6	7.2	17

the model (IV) and cases (ii) and (vi) for $\boldsymbol{\mu} = \kappa(\mu_1, \cdots, \mu_p)$. The coefficient κ was selected so that the signal-to-noise $\|\boldsymbol{\mu}\|^2/\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)} = 0.1$. We took $n_1 = n_2 = 15, 20, 30$, and $p = 224$.

Table 5 reports the results. The sizes of the PA test and our test are close to the nominal level. Our test is more powerful than the PA test in both (ii) and (vi). The sizes of RHT test are still smaller than the nominal level, and the tests

Table 4. Empirical sizes and power (%) comparisons under the multivariate normal distribution in the one-sample case.

	PA	HT	CT_2	PA	HT	CT_2	PA	HT	CT_2	PA	HT	CT_2
	Random Cases											
Model	size			(i)			(ii)			(iii)		
(I)	6.0	6.0	4.0	33	39	38	38	42	40	36	42	41
(II)	7.0	6.7	6.3	17	43	46	20	42	42	12	43	44
(III)	8.7	6.3	5.7	19	47	49	16	42	42	14	46	46
(IV)	3.0	6.7	3.7	9.6	37	27	12	42	29	8.9	41	28
(V)	8.3	5.0	5.1	9.1	39	27	11	38	26	10	42	28
(VI)	4.7	4.0	4.0	37	40	38	39	42	41	32	40	40
(VII)	6.2	5.2	6.3	6.0	42	25	8.0	41	26	8.3	39	25
	Fixed Cases											
	(iv)			(v)			(vi)			(vii)		
(I)	41	43	42	42	42	42	47	46	46	23	44	43
(II)	51	43	43	9.2	35	40	17	41	44	16	41	42
(III)	9.0	37	37	52	41	40	15	41	42	13	46	47
(IV)	58	38	48	6.7	40	25	9.6	41	25	11	43	31
(V)	8.2	41	29	61	40	53	11	41	27	11	37	26
(VI)	58	43	50	33	40	38	38	38	38	45	47	47
(VII)	63	42	59	9.1	45	29	13	42	30	8.4	35	23

Table 5. Empirical sizes and power (%) comparisons under the multivariate normal distribution in the two sample case.

n_i	size					(ii)					(vi)				
	PA	RHT	CT_2	GCT	CLX	PA	RHT	CT_2	GCT	CLX	PA	RHT	CT_2	GCT	CLX
15	5.3	0.0	4.7	9.6	36	15	3.3	59	24	56	15	2.6	46	24	57
20	6.2	2.1	5.1	8.0	23	17	5.5	73	31	51	18	2.4	72	28	48
30	5.7	1.2	5.6	11	10	27	3.7	97	41	53	26	2.1	96	42	45

are not effective under the alternative hypothesis. The sizes of the GCT test are a little larger than the nominal level. And our test CT_2 also outperforms GCT test. The CLX test did not control the empirical sizes very well in these cases, especially when the sample size was small.

Acknowledgment

The authors thank the Editor, an associate editor, and two referees for their many helpful comments that have resulted in significant improvements in the article. Feng was supported by the National Natural Science Foundation of China Grants 11501092 and the Fundamental Research Funds for the Central Universities. Zou and Wang was supported by the NNSF of China Grants 11431006, 11131002, 11371202, the Foundation for the Author of National Excellent Doctoral Dissertation of PR China 201232. Zhu was partly supported by a grant from the Research Grants Council of Hong Kong, Hong Kong, China.

Appendix

A.1. Proof of Proposition 1

Proof. Let $\lambda_A = \|\mathbf{R}_A\|_{l_1}$ with $\hat{\lambda}_A^{ij}$ the corresponding estimator based on the sample $\{\boldsymbol{X}_k\}_{k\neq i,j}$. By the Central Limited Theorem, $\sqrt{n}(\hat{\lambda}_A^{ij}-\lambda_A) \xrightarrow{\mathcal{L}} N(0,\sigma_A^2)$. If $\epsilon = (\lambda_{A_1^o} - \lambda_A)\sigma_{A_1^o}/(\sigma_A+\sigma_{A_1^o})$, we have

$$\begin{aligned}
P\left(\hat{\lambda}_{A_1^o} < \hat{\lambda}_A^{ij}\right) &= P\left(\hat{\lambda}_{A_1^o} - \lambda_{A_1^o} < \hat{\lambda}_A^{ij} - \lambda_{A_1^o}, \hat{\lambda}_{A_1^o} - \lambda_{A_1^o} > -\epsilon\right) \\
&\quad + P\left(\hat{\lambda}_{A_1^o} - \lambda_{A_1^o} < \hat{\lambda}_A^{ij} - \lambda_{A_1^o}, \hat{\lambda}_{A_1^o} - \lambda_{A_1^o} < -\epsilon\right) \\
&\le P\left(\hat{\lambda}_A^{ij} - \lambda_{A_1^o} > -\epsilon\right) + P\left(\hat{\lambda}_{A_1^o} - \lambda_{A_1^o} < -\epsilon\right) \\
&= \Phi\left(\frac{\sqrt{n}(\lambda_{A_1^o} - \lambda_A - \epsilon)}{\sigma_A}\right) + \Phi\left(\frac{-\sqrt{n}\epsilon}{\sigma_{A_1^o}}\right) \\
&= 2\Phi\left(\frac{\sqrt{n}(\lambda_{A_1^o} - \lambda_A)}{\sigma_A + \sigma_{A_1^o}}\right) \le \frac{2}{\sqrt{2\pi n \varpi_1}} e^{-n\varpi_1^2/2}.
\end{aligned}$$

If $\mathcal{O}_{ij}^K = (A_1^{ij}, \cdots, A_N^{ij})$,

$$\begin{aligned}
P(A_1^{ij} \neq A_1^o) &= P\left(\bigcup_{A\in\{1,\cdots,p\},|A|=K,A\neq A_1^o} \left\{\hat{\lambda}_{A_1^o} < \hat{\lambda}_A^{ij}\right\}\right) \\
&\le \mathrm{C}_p^K P\left(\hat{\lambda}_{A_1^o} < \hat{\lambda}_A^{ij}\right) \le \frac{2\mathrm{C}_p^K}{\sqrt{2\pi n\varpi_1^2}} e^{-n\varpi_1^2/2}.
\end{aligned}$$

Similarly, we can show that $P(A_k^{ij} \neq A_k^o) \le \mathrm{C}_p^K/\sqrt{2\pi n\varpi_k^2} e^{-n\varpi_k^2/2}$. Then

$$\begin{aligned}
&P\left(\bigcap_{1\le i<j\le n} \{\mathcal{O}_{ij}^K = \mathcal{O}^K\}\right) = 1 - P\left(\bigcup_{1\le i<j\le n} \{\mathcal{O}_{ij}^K \neq \mathcal{O}^K\}\right) \\
&= 1 - P\left(\bigcup_{1\le i<j\le n} \bigcup_{1\le k\le N} \left\{A_k^{ij} \neq A_k^o\right\}\right) \\
&\le 1 - \frac{n^2 N\mathrm{C}_p^K}{\sqrt{2\pi n\varpi_{\min}^2}} e^{-n\varpi_{\min}^2/2} = 1 - O(n^{3/2}p^{K+1}e^{-n\omega^2/2}),
\end{aligned}$$

by (C1).

A.2. Proof of Theorem 1

Proof. According to Proposition 1, we only need consider the asymptotic property of $\tilde{T}_n$,

$$\begin{aligned}\tilde{T}_n &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\boldsymbol{X}_i^T\widehat{\boldsymbol{\Sigma}_{\mathcal{O}^K}^{(i,j)}}^{-1}\boldsymbol{X}_j\\ &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\boldsymbol{X}_i^T\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}\boldsymbol{X}_j+\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\boldsymbol{X}_i^T\left(\widehat{\boldsymbol{\Sigma}_{\mathcal{O}^K}^{(i,j)}}^{-1}-\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}\right)\boldsymbol{X}_j\\ &\doteq \tilde{T}_{n1}+\tilde{T}_{n2}.\end{aligned}$$

We show that

$$\frac{\tilde{T}_{n1}-\boldsymbol{\mu}^T\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}\boldsymbol{\mu}}{\sqrt{2n^{-2}\mathrm{tr}(\boldsymbol{\Lambda}_K^2)}}\xrightarrow{\mathcal{L}}N(0,1). \tag{A.1}$$

and $\tilde{T}_{n2}=o_p\left(\sqrt{2n^{-2}\mathrm{tr}(\boldsymbol{\Lambda}_K^2)}\right)$, where

$$\begin{aligned}\tilde{T}_{n1} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(\boldsymbol{X}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}(\boldsymbol{X}_j-\boldsymbol{\mu})+\frac{2}{n}\sum_{i=1}^{n}\boldsymbol{\mu}^T\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}(\boldsymbol{X}_i-\boldsymbol{\mu})\\ &\quad+\boldsymbol{\mu}^T\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}\boldsymbol{\mu}\doteq\tilde{T}_{n11}+\tilde{T}_{n12}+\boldsymbol{\mu}^T\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}\boldsymbol{\mu}.\end{aligned}$$

It is easy to show that $E(\tilde{T}_{n12})=0$ and $\mathrm{var}(\tilde{T}_{n12})=4n^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}\boldsymbol{\mu}=o(2n^{-2}\mathrm{tr}(\boldsymbol{\Lambda}_K^2))$. So $\tilde{T}_{n12}=o_p(\sqrt{2n^{-2}\mathrm{tr}(\boldsymbol{\Lambda}_K^2)})$. We need only show the asymptotic normality of $\tilde{T}_{n11}$. Without lose of generality, we take $\boldsymbol{\mu}=0$.

Take $V_{nj}=n^{-1}(n-1)^{-1}\sum_{i=1}^{j-1}\boldsymbol{X}_i^T\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}\boldsymbol{X}_j$, $j=2,\cdots,n$, and $W_{nk}=\sum_{i=2}^{k}V_{ni}$, $k=2,\cdots,n$. Let $\mathcal{F}_i=\sigma\{\boldsymbol{X}_1,\cdots,\boldsymbol{X}_i\}$ be the σ-field generated by $\{\boldsymbol{X}_j\}_{j\leq i}$. It is easy to show that $E(V_{ni}|\mathcal{F}_{i-1})=0$ and it follows that $\{W_{nk},\mathcal{F}_k;\ 2\leq k\leq n\}$ is a zero mean martingale. Let $v_{ni}=E(V_{ni}^2|\mathcal{F}_{i-1})$, $2\leq i\leq n$ and $V_n=\sum_{i=2}^{n}v_{ni}$. The central limit theorem (Hall and Hyde (1980)) will hold if we can show

$$\frac{V_n}{\mathrm{var}(W_{nn})}\xrightarrow{p}1, \tag{A.2}$$

and, for any $\epsilon>0$,

$$\sum_{i=2}^{n}n^2\mathrm{tr}^{-1}(\boldsymbol{\Lambda}_K^2)E\left[V_{ni}^2I(|V_{ni}|>\epsilon\sqrt{n^{-2}\mathrm{tr}(\boldsymbol{\Lambda}_K^2)})|\mathcal{F}_{i-1}\right]\xrightarrow{p}0. \tag{A.3}$$

It can be shown that

$$v_{ni} = \frac{1}{n^2(n-1)^2}\left(\sum_{j=1}^{i-1} \boldsymbol{X}_j^T \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{X}_j + 2\sum_{1\le j<k<i} \boldsymbol{X}_j^T \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{X}_k\right).$$

Then,

$$\frac{V_n}{\mathrm{var}(W_{nn})} = \frac{2}{n(n-1)\mathrm{tr}(\boldsymbol{\Lambda}_K^2)}\left(\sum_{j=1}^{n-1} j\boldsymbol{X}_j^T \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{X}_j \right.$$
$$\left. +2\sum_{1\le j<k\le n} \boldsymbol{X}_j^T \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{X}_k\right) \doteq C_{n1} + C_{n2}.$$

Simple algebras lead to

$$E(C_{n1}) = 1,$$
$$\mathrm{var}(C_{n1}) = \frac{4}{n^2(n-1)^2\mathrm{tr}^2(\boldsymbol{\Lambda}_K^2)} E\left(\sum_{j=1}^{n-1} j^2(\boldsymbol{X}_j^T \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{X}_j)^2 - \mathrm{tr}^2(\boldsymbol{\Lambda}_K^2)\right).$$

Let $\Gamma^T \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \Gamma = (\omega_{kl})_{1\le k\le l\le m}$.Under the diverging factor model,

$$\begin{aligned}
&E((\boldsymbol{X}_j^T \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{X}_j)^2) \\
&= E((\mathbf{z}_j^T \boldsymbol{\Gamma}^T \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{\Gamma} \mathbf{z}_j)^2) = E\left(\left(\sum_{k=1}^m \sum_{l=1}^m \omega_{kl} z_{jk} z_{jl}\right)^2\right) \\
&= \sum_{k=1}^m \sum_{l=1}^m \sum_{s=1}^m \sum_{t=1}^m \omega_{kl}\omega_{st} E(z_{jk}z_{jl}z_{js}z_{jt}) = (3+\Delta)\sum_{k=1}^m \omega_{kk}^2 + \sum_{k\ne l} \omega_{kl}^2 \\
&= (2+\Delta)\sum_{k=1}^m \omega_{kk}^2 + \mathrm{tr}(\boldsymbol{\Lambda}_K^4) \le (3+\Delta)\mathrm{tr}(\boldsymbol{\Lambda}_K^4). \qquad \text{(A.4)}
\end{aligned}$$

Under (C2), $E((\boldsymbol{X}_j^T \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{X}_j)^2) = o(\mathrm{tr}^2(\boldsymbol{\Lambda}_K^2))$. Hence, $\mathrm{var}(C_{n1}) \to 0$ and then $C_{n1} \xrightarrow{p} 1$. Similarly, $E(C_{n2}) = 0$ and

$$\mathrm{var}(C_{n2}) = \frac{16}{n(n-1)} \frac{\mathrm{tr}(\boldsymbol{\Lambda}_K^4)}{\mathrm{tr}^2(\boldsymbol{\Lambda}_K^2)} \to 0$$

implies $C_{n2} \xrightarrow{p} 0$. Thus, (A.2) holds. It remains to show (A.3). Since

$$E\left[Z_{ni}^2 I\left(|Z_{ni}| > \epsilon\sqrt{n^{-2}\mathrm{tr}(\boldsymbol{\Lambda}_K^2)}\right)|\mathcal{F}_{i-1}\right] \le E(Z_{ni}^4|\mathcal{F}_{i-1})/(\epsilon^2 n^{-2}\mathrm{tr}(\boldsymbol{\Lambda}_K^2))$$

we need only show that

$$\sum_{i=2}^n E(Z_{ni}^4) = o(n^{-4}\mathrm{tr}^2(\boldsymbol{\Lambda}_K^2)).$$

Note that

$$\sum_{i=2}^{n} E(Z_{ni}^4) = O(n^{-4}) \sum_{i=2}^{n} E\left(\left(\sum_{j=1}^{i-1} \eta_i \eta_j \boldsymbol{X}_i^T \boldsymbol{X}_j\right)^4\right)$$

and this can be decomposed as $3Q + P$, where

$$Q = O(n^{-8}) \sum_{i=2}^{n} \sum_{s \neq t}^{i-1} E\left(\boldsymbol{X}_i^T \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{X}_s \boldsymbol{X}_s^T \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{X}_i \boldsymbol{X}_i^T \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{X}_t \boldsymbol{X}_t^T \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{X}_i\right),$$

$$P = O(n^{-8}) \sum_{i=2}^{n} \sum_{s=1}^{i-1} E\left((\boldsymbol{X}_i^T \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{X}_s)^4\right).$$

Here $Q = O(n^{-4}) E((\boldsymbol{X}_i^T \boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{X}_i)^2) = o(\mathrm{tr}^2(\boldsymbol{\Lambda}_K^2))$ by similar arguments in (A.4). For P, with $\boldsymbol{\Gamma}^T \boldsymbol{\Gamma} = (\nu_{kl})_{1 \leq k,l \leq m}$.

$$P = O(n^{-8}) \sum_{i=2}^{n} \sum_{s=1}^{i-1} E\left((\mathbf{z}_i^T \boldsymbol{\Gamma}^T \boldsymbol{\Gamma} \mathbf{z}_s)^4\right) = O(n^{-8}) \sum_{i \neq j} E\left(\left(\sum_{k,l=1}^{m} \nu_{kl} z_{ik} z_{jl}\right)^4\right)$$

$$= O(n^{-6}) \Bigg(\sum_{k,l=1}^{m} \nu_{kl}^4 E(z_{ik}^4) E(z_{jl}^4) + \sum_{k \neq l}^{m} \sum_{s \neq t}^{m} v_{kl}^2 v_{st}^2 E(z_{ik}^2) E(z_{is}^2) E(z_{jl}^2) E(z_{jt}^2)$$

$$+2 \sum_{k=1}^{m} \sum_{s \neq t}^{m} v_{ks}^2 v_{kt}^2 E(z_{ik}^4) E(z_{js}^2 z_{jt}^2) + \sum_{k \neq l}^{m} \sum_{s \neq t}^{m} v_{kl} v_{kt} v_{st} v_{sl} E(z_{ik}^2) E(z_{jl}^2) E(z_{is}^2) E(z_{jt}^2) \Bigg).$$

Note that $\mathrm{tr}^2(\boldsymbol{\Lambda}_K^2) = (\sum_{s,t} \nu_{st}^2)^2 = \sum_{k,l,s,t} \nu_{st}^2 \nu_{kl}^2$ and

$$\sum_{k,l=1}^{m} \nu_{kl}^4 \leq \left(\sum_{k,l} \nu_{kl}^2\right)^2, \sum_{k=1}^{m} \sum_{s \neq t}^{m} v_{ks}^2 v_{kt}^2 \leq \left(\sum_{k,l} \nu_{kl}^2\right)^2,$$

$$\sum_{k \neq l}^{m} \sum_{s \neq t}^{m} v_{kl}^2 v_{st}^2 \leq \sum_{k,l,s,t} \nu_{st}^2 \nu_{kl}^2, \sum_{k \neq l}^{m} \sum_{s \neq t}^{m} v_{kl} v_{kt} v_{st} v_{sl} \leq \sum_{k \neq l} \omega_{kl}^2 \leq \sum_{k,l} \omega_{kl}^2 = \mathrm{tr}(\boldsymbol{\Lambda}_K^4).$$

Thus, under (C2), $P = o(n^{-4} \mathrm{tr}^2(\boldsymbol{\Lambda}_K^2))$ and then (A.3) follows. This complete the proof of (A.1).

Next, we show that $\tilde{T}_{n2} = o_p(\sqrt{n^{-2} \mathrm{tr}(\boldsymbol{\Lambda}_K^2)})$. Obviously, $E(\tilde{T}_{n2}) = 0$. Here, we need only show that $E(\tilde{T}_{n2}^2) = o(n^{-2} \mathrm{tr}(\boldsymbol{\Lambda}_K^2))$. Take $\boldsymbol{\Sigma}_{\mathcal{O}^K} \widehat{\boldsymbol{\Sigma}_{\mathcal{O}}^{(i,j)}}^{-1} = (\hat{d}_{st})_{1 \leq s \leq t \leq p}$ and $\mathbf{I}_p = (d_{st})_{1 \leq s \leq t \leq p}$. By the Central Limit Theorem, $\sqrt{n}(\hat{d}_{st} - d_{st}) \xrightarrow{\mathcal{L}} N(0, \zeta_{st}^2)$. Let $\sigma_{\max}^2 = \max_{1 \leq s \leq t \leq p} \zeta_{st}^2$. As $n, p \to \infty$,

$$P\left(\max_{1 \leq s \leq t \leq p} (\hat{d}_{st} - d_{st}) > 2\sigma_{\max} n^{-1/2} (\log p)^{1/2}\right)$$

$$\begin{aligned}
&\leq \sum_{s=1}^{p}\sum_{t=1}^{p} P\left(\sqrt{n}(\hat{d}_{st}-d_{st}) > 2\sigma_{\max}(\log p)^{1/2}\right)\\
&= \sum_{s=1}^{p}\sum_{t=1}^{p}\left(1-\Phi(2\sigma_{\max}\zeta_{st}^{-1}(\log p)^{1/2})\right) \leq p^2(1-\Phi((4\log p)^{1/2}))\\
&\leq \frac{p^2}{\sqrt{8\pi\log p}}e^{-2\log p} \to 0.
\end{aligned}$$

Thus, $\max_{1\leq s\leq t\leq p}(\hat{d}_{st}-d_{st}) = O_p(n^{-1/2}(\log p)^{1/2})$, and then by (C3),

$$\begin{aligned}
E(\tilde{T}_{n2}^2) &\leq C(\log p)^{1/2}n^{-1/2}E(\tilde{T}_{n1}^2)\\
&\leq C(\log p)^{1/2}n^{-1/2}\left((\boldsymbol{\mu}^T\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}\boldsymbol{\mu})^2 + n^{-2}\mathrm{tr}(\boldsymbol{\Lambda}_K^2)\right) = o(n^{-2}\mathrm{tr}(\boldsymbol{\Lambda}_K^2)).
\end{aligned}$$

A.3. Proof of Proposition 2

Proof. Similar to Proposition 1, we can show that

$$P\left(\bigcap_{i_1,i_2,i_3,i_4}\{\mathcal{O}_{i_1,i_2,i_3,i_4}^K = \mathcal{O}^K\}\right) = 1 - O(n^{7/2}p^{K+1}e^{-n\omega^2/2}).$$

And similar to the argument for $\tilde{T}_{n2}$ in the proof of Theorem 1, we can show that

$$\begin{aligned}
\widehat{\mathrm{tr}(\boldsymbol{\Lambda}_K^2)} &= \frac{1}{2P_n^4}\sum^{*}(\boldsymbol{X}_{i_1}-\boldsymbol{X}_{i_2})^T\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}(\boldsymbol{X}_{i_3}-\boldsymbol{X}_{i_4})(\boldsymbol{X}_{i_1}-\boldsymbol{X}_{i_4})^T\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}(\boldsymbol{X}_{i_3}-\boldsymbol{X}_{i_2})\\
&\quad + o_p(\mathrm{tr}(\boldsymbol{\Lambda}_K^2))\\
&= \frac{1}{P_n^2}\sum^{*}\left(\boldsymbol{X}_{i_1}^T\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}\boldsymbol{X}_{i_2}\right)^2 - \frac{2}{P_n^3}\sum^{*}\boldsymbol{X}_{i_1}^T\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}\boldsymbol{X}_{i_2}\boldsymbol{X}_{i_2}^T\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}\boldsymbol{X}_{i_3}\\
&\quad + \frac{1}{P_n^4}\sum^{*}\boldsymbol{X}_{i_1}^T\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}\boldsymbol{X}_{i_2}\boldsymbol{X}_{i_3}^T\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}\boldsymbol{X}_{i_4} + o_p(\mathrm{tr}(\boldsymbol{\Lambda}_K^2)).
\end{aligned}$$

Then, with Theorem 2 in Chen, Zhang and Zhong (2010), we can easily obtain the result.

A.4. Proof of Theorem 2

Proof. Similar to Proposition 1, we can show that

$$P\left(\bigcap_{i_1,i_2,j_1,j_2}\{\mathcal{O}_{i_1,i_2,j_1,j_2}^K = \mathcal{O}^K\}\right) = 1 - O(n^{7/2}p^{K+1}e^{-n\omega^2/2}).$$

Then

$$Q_n = \frac{1}{n_1n_2(n_1-1)(n_2-1)}\sum_{1\leq i_1\neq i_2\leq n_1}\sum_{1\leq j_1\neq j_2\leq n_2}(\boldsymbol{X}_{1i_1}-\boldsymbol{X}_{2j_1})^T\boldsymbol{\Sigma}_{\mathcal{O}^K}^{-1}(\boldsymbol{X}_{1i_2}-\boldsymbol{X}_{2j_2})$$

$$
\begin{aligned}
&+ o_p(\sqrt{n^{-2}\mathrm{tr}(\mathbf{\Lambda}_K^2)}) \\
= &\frac{1}{n_1(n_1-1)} \sum_{1\le i_1\ne i_2\le n_1} \boldsymbol{X}_{1i_1}^T \mathbf{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{X}_{1i_2} + \frac{1}{n_2(n_2-1)} \sum_{1\le j_1\ne j_2\le n_2} \boldsymbol{X}_{2j_1}^T \mathbf{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{X}_{2j_2} \\
&- \frac{2}{n_1 n_2} \sum_{i_1=1}^{n_1} \sum_{j=1}^{n_2} \boldsymbol{X}_{1i_1}^T \mathbf{\Sigma}_{\mathcal{O}^K}^{-1} \boldsymbol{X}_{2j_1} + o_p(\sqrt{n^{-2}\mathrm{tr}(\mathbf{\Lambda}_K^2)}).
\end{aligned}
$$

As before, Chen and Qin (2010), we can obtain the result.

References

Bai Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica* **6**, 311–29.

Bickel, P. and Levina, E. (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics* **36**, 199–227.

Cai, T. T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association* **106**, 672–684.

Cai, T. T., Liu, W. and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society B* **76**, 349–372.

Chen, L. S., Paul, D., Prentice, R. L. and Wang, P. (2011) A regularized Hotelling's T^2 test for pathway analysis in proteomic studies. *Journal of the American Statistical Association* **106**, 1345–1360.

Chen, S. X. and 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.

Chen, S. X., Zhang, L. X. and Zhong, P. S. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association* **105**, 801–815.

Feng, L. and Sun, F. (2015). A note on high-dimensional two sample test. *Statistics and Probability Letters*, **105**, 29–36.

Feng, L., Zou, C. and Wang, Z. (2015). Multivariate-sign-based high-dimensional tests for the two-sample location problem. *Journal of the American Statistical Association* **111**, 721–735.

Feng, L., Zou, C., Wang, Z. and Zhu, L. X. (2015). Two sample Behrens-Fisher problem for high-dimensional data. *Statistica Sinica*, **25**, 1297–1312.

Gregory, K. B., Carroll, R. J., Baladandayuthapani, V. and Lahiri, S. N. (2015). A two-sample test for equality of means in high dimension. *Journal of the American Statistical Association* **110**, 837–849.

Hall, P. G. and Hyde, C. C. (1980), *Martingale Central Limit Theory and its Applications.* Academic Press, New York.

Park, J. and Ayyala, D. N. (2013). A test for the mean vector in large dimension and small samples. *Journal of Statistical Planning and Inference* **143**, 929-943.

Srivastava, M. S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis* **100**, 518–532.

Institute of Statistics and LPMC, Nankai University, Tianjin, 300071, China
E-mail: flnankai@126.com
Institute of Statistics and LPMC, Nankai University, Tianjin, 300071, China
E-mail: nk.chlzou@gmail.com
Institute of Statistics and LPMC, Nankai University, Tianjin, 300071, China
E-mail: zjwang@nankai.edu.cn
Department of Mathematics, Hong Kong Baptist University, Hong Kong, China
E-mail: lzhu@hkbu.edu.hk

(Received June 2015; accepted July 2016)