Electronic Journal of Statistics
Vol. 10 (2016) 2420–2434
ISSN: 1935-7524
DOI: 10.1214/16-EJS1176

Spatial-sign based high-dimensional location test

Long Feng and Fasheng Sun*

Key Laboratory for Applied Statistics of MOE and School of Mathematics and Statistics, Northeast Normal University, Changchun, China
e-mail: flnankai@126.com; sunfs359@nenu.edu.cn

Abstract: In this paper, we consider the problem of testing the mean vector in the high-dimensional settings. We proposed a new robust scalar transform invariant test based on spatial sign. The proposed test statistic is asymptotically normal under elliptical distributions. Simulation studies show that our test is very robust and efficient in a wide range of distributions.

MSC 2010 subject classifications: Primary 62H15; secondary 62H11, 62G35.
Keywords and phrases: Asymptotic normality, high-dimensional data, large p, small n, spatial median, spatial-sign test, scalar-invariance.

Received January 2015.

1. Introduction

Assume $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ is an independent sample from p-variate distribution $F(\mathbf{x}-\boldsymbol{\theta})$ located at $p = p_n$ variate center $\boldsymbol{\theta} = \boldsymbol{\theta}_n$. We consider the following one sample testing problem

$$H_0 : \boldsymbol{\theta}_n = \mathbf{0} \quad \text{versus} \quad H_1 : \boldsymbol{\theta}_n \neq \mathbf{0}.$$

One typical test statistic is Hotelling's T^2. However, it can not be applied when $p_n > n-1$ because of the singularity of the sample covariance matrix. Recently, many efforts have been devoted to solve the problem, such as [1], [19], [18], [2], [13], [4] and [5]. They established the asymptotic normality of their test statistics under the assumption of diverging factor model [1]. Even this data structure generates a rich collection of $\boldsymbol{X}$, it is not easily met in practice. Moreover, multivariate t distribution or mixtures of multivariate normal distributions does not satisfy the diverging factor model. This motivates us to construct a robust test procedure.

Multivariate sign or rank is often used to construct robust test statistics in the multivariate setting [16, 9, 11, 17, 10, 7, 8]. Especially, multivariate sign tests enjoy many desirable properties. First, those test statistics are distribution-free under mild assumptions, or asymptotically so. Second, they do not require stringent parametric assumptions, nor any moment conditions. Third, they have high

*Corresponding author.

asymptotic relative efficiency with respect to the classic Hotelling's T^2 test, especially under the heavy-tailed distributions. However, the classic spatial-sign test also can not work in the high-dimensional settings because the scatter matrix [12] is unable to be estimated. Recently, without estimating the scatter matrix, [20] and [14] proposed a high-dimensional nonparametric test based on the direction of $\boldsymbol{X}_i$, i.e. $\boldsymbol{X}_i/||\boldsymbol{X}_i||$. Even it is workable and robust in high-dimensional settings, it loses all the information of the scalar of different variables and then is not scalar-invariant. In practice, different components may have completely different physical or biological readings and thus certainly their scales would not be identical. [18] and [13] proposed two scalar-invariant tests under different assumption of correlation matrix. As shown above, they are not robust for the heavy-tailed distributions. In this paper, we proposed a new robust test based on spatial sign. We show that it is scalar invariant and asymptotic normal under some mild conditions. The asymptotic relative efficiency of our test with respect to [13]'s test is the same as the classic spatial-sign test with respect to the Hotelling's T^2 test. Simulation comparisons show that our procedure has good size and power for a wide range of dimensions, sample sizes and distributions. All the proofs are given in the appendix.

2. Robust high-dimensional test

2.1. The proposed test statistic

The spatial sign function is defined as $U(\mathbf{x}) = ||\mathbf{x}||^{-1}\mathbf{x}I(\mathbf{x} \neq \mathbf{0})$. In traditional fixed p circumstance, the following so-called "inner centering and inner standardization" sign-based procedure is usually used (cf., Chapter 6 of [12])

$$Q_n^2 = np\bar{\boldsymbol{U}}^T\bar{\boldsymbol{U}}, \tag{2.1}$$

where $\bar{\boldsymbol{U}} = \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{U}}_i$, $\hat{\boldsymbol{U}}_i = U(\mathbf{S}^{-1/2}\boldsymbol{X}_{ij})$, $\mathbf{S}^{-1/2}$ are Tyler's scatter matrix (cf., Section 6.1.3 of [12]). Q_n^2 is affine-invariant and can be regarded as a nonparametric counterpart of Hotelling's T^2 test statistic by using the spatial-signs instead of the original observations $\boldsymbol{X}_{ij}$'s. However, when $p > n$, Q_n^2 is not defined as the matrix $\mathbf{S}^{-1/2}$ is not available in high-dimensional settings.

Motivated by [10], we suggest to find a pair of diagonal matrix $\mathbf{D}$ and vector $\boldsymbol{\theta}$ for each sample that simultaneously satisfy

$$\frac{1}{n}\sum_{i=1}^n U(\boldsymbol{\epsilon}_i) = 0 \ \text{ and } \ \frac{p_n}{n}\mathrm{diag}\left\{\sum_{i=1}^n U(\boldsymbol{\epsilon}_i)U(\boldsymbol{\epsilon}_i)^T\right\} = \mathbf{I}_{p_n}, \tag{2.2}$$

where $\boldsymbol{\epsilon}_i = \mathbf{D}^{-1/2}(\boldsymbol{X}_i - \boldsymbol{\theta})$. $(\mathbf{D}, \boldsymbol{\theta})$ can be viewed as a simplified version of Hettmansperger-Randles (HR) estimator without considering the off-diagonal elements of $\mathbf{S}$. We can adapt the recursive algorithm of [10] to solve (2.2). That is, repeat the following three steps until convergence:

(i) $\boldsymbol{\epsilon}_i \leftarrow \mathbf{D}^{-1/2}(\boldsymbol{X}_i - \boldsymbol{\theta}), \quad j = 1, \cdots, n_i;$

(ii) $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \frac{\mathbf{D}^{1/2}\sum_{j=1}^{n} U(\boldsymbol{\epsilon}_i)}{\sum_{j=1}^{n} ||\boldsymbol{\epsilon}_i||^{-1}}$;
(iii) $\mathbf{D} \leftarrow p_n \mathbf{D}^{1/2}\mathrm{diag}\{n^{-1}\sum_{j=1}^{n} U(\boldsymbol{\epsilon}_i)U(\boldsymbol{\epsilon}_i)^T\}\mathbf{D}^{1/2}$.

The resulting estimators of location and diagonal matrix are denoted as $\hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{D}}$. We may use the sample mean and sample variances as the initial estimators. Unfortunately, there is no proof so far for the convergence of the above algorithm, even for the low-dimensional cases, although it always seems to work in practice. There is no proof for the existence or uniqueness of the above HR estimate either. Some further research are deserved for this topic.

Then, we define the following test statistic

$$T_n = \frac{2}{n(n-1)}\sum\sum_{i<j} U\left(\hat{\mathbf{D}}_{ij}^{-1/2}\boldsymbol{X}_i\right)^T U\left(\hat{\mathbf{D}}_{ij}^{-1/2}\boldsymbol{X}_j\right)$$

where $\hat{\mathbf{D}}_{ij}$ are the corresponding diagonal matrix estimator using leave-two-out sample $\{\boldsymbol{X}_k\}_{k\neq i,j}^n$. The leave-two-out estimates $\hat{\mathbf{D}}_{ij}$ is independent of $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ and then there is no bias term of T_n if $p = O(n^2)$. Otherwise, if we use $\hat{\mathbf{D}}$ in T_n, there is a non-negligible bias term of T_n if p has the same order as n^2. See more information in [4]. The tests statistics proposed by [1, 2, 14, 20] are invariant under orthogonal transformations, $\boldsymbol{X}_i \to \mathbf{P}\boldsymbol{X}_i$ where $\mathbf{P}$ is an orthogonal matrix. In contrast, T_n is not invariant under the orthogonal transformations, but it is invariant under scalar transformations $\boldsymbol{X}_i \to \mathbf{B}\boldsymbol{X}_i$ where $\mathbf{B} = \mathrm{diag}\{b_1^2, \cdots, b_{p_n}^2\}$ and $b_1, \cdots, b_{p_n}$ are non-zero constants.

2.2. Asymptotic results

First, we state the assumption of the distribution of $\boldsymbol{X}$:

(A1) Variables $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$ in the n-th row are independently and identically distributed (i.i.d.) from $p = p_n$-variate elliptical distribution with density functions $\det(\boldsymbol{\Sigma})^{-1/2} g_n(||\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\theta})||)$ where $\boldsymbol{\theta} = \boldsymbol{\theta}_n$'s are the symmetry centers and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_n$'s are the positive definite symmetric $p_n \times p_n$ scatter matrices.

We also need the following conditions for asymptotic analysis:

(C1) $\mathrm{tr}(\mathbf{R}_n^4) = o(\mathrm{tr}^2(\mathbf{R}_n^2))$, where $\mathbf{R}_n = \mathbf{D}_n{}^{-1/2}\boldsymbol{\Sigma}_n\mathbf{D}_n{}^{-1/2}$;
(C2) $n^{-2}p_n^2/\mathrm{tr}(\mathbf{R}_n^2) = O(1)$ and $\log p_n = o(n)$;
(C3) $(\mathrm{tr}(\mathbf{R}_n^2) - p_n) = o(n^{-1}p_n^2)$.

To appreciate these conditions, define the p_n eigenvalues of $\mathbf{R}_n$ are $\lambda_{n,1}, \cdots, \lambda_{n,p_n}$ and $\nu_{n,k} = \sum_{i=1}^{p_n}\lambda_{n,i}^k$, $k = 2, 4$. Then, the above three conditions become

(C1$'$) $\nu_{n,4} = o(\nu_{n,2}^2)$;
(C2$'$) $n^{-2}p_n^2/\nu_{n,2} = O(1)$ and $\log p_n = o(n)$;
(C3$'$) $(\nu_{n,2} - p_n) = o(n^{-1}p_n^2)$.

If $\lambda_{n,1}, \cdots, \lambda_{n,p_n}$ are all bounded, $\nu_{n,4} = O(p_n)$ and $\nu_{n,2} = O(p_n)$. So Condition (C1) holds. Moreover, in this case, Condition (C2) and (C3) become $p_n = O(n^2)$ and $p_n/n \to \infty$. Thus, we could allow the dimension being the square of the sample size. To get the consistency of the diagonal matrix, the dimension must diverging faster than the sample sizes.

The following theorem establishes the asymptotic null distribution of T_n.

Theorem 1. *Under Assumption (A1), Conditions (C1)–(C3) and H_0, as $\min(p_n, n) \to \infty$, $T_n/\sigma_n \xrightarrow{d} N(0,1)$, where $\sigma_n^2 = \frac{2}{n(n-1)p_n^2}\mathrm{tr}(\mathbf{R}_n^2)$.*

We propose the following estimator to estimate the trace term in σ_n^2

$$\widehat{\mathrm{tr}(\mathbf{R}_n^2)} = \frac{p_n^2}{n(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \left(U(\hat{\mathbf{D}}_{ij}^{-1/2}(\boldsymbol{X}_i - \hat{\boldsymbol{\theta}}_{ij}))^T U(\hat{\mathbf{D}}_{ij}^{-1/2}(\boldsymbol{X}_j - \hat{\boldsymbol{\theta}}_{ij})) \right)^2$$

where $(\hat{\boldsymbol{\theta}}_{ij}, \hat{\mathbf{D}}_{ij})$ are the corresponding spatial median and diagonal matrix estimators using leave-two-out sample $\{\boldsymbol{X}_k\}_{k\neq i,j}^n$. By Proposition 2 in [3], $\widehat{\mathrm{tr}(\mathbf{R}_n^2)}/\mathrm{tr}(\mathbf{R}_n^2) \xrightarrow{p} 1$ as $p_n, n \to \infty$. Consequently, a ratio-consistent estimator of σ_n^2 under H_0 is $\hat{\sigma}_n^2 = \frac{2}{n(n-1)p_n^2}\widehat{\mathrm{tr}(\mathbf{R}_n^2)}$. And then we reject the null hypothesis with α level of significance if $T_n/\hat{\sigma}_n > z_\alpha$, where z_α is the upper α quantile of $N(0,1)$.

Next, we consider the asymptotic distribution of R_n under the alternative hypothesis

(C4) $\boldsymbol{\theta}_n^T \mathbf{D}_n{}^{-1} \boldsymbol{\theta}_n = O(c_0^{-2}\sigma_n)$ and $\boldsymbol{\theta}_n^T \mathbf{D}_n{}^{-1} \boldsymbol{\Sigma}_n \mathbf{D}_n{}^{-1} \boldsymbol{\theta}_n = o(c_0^{-2} n p_n \sigma_n^2)$ where $c_0 = E(||\mathbf{D}_n{}^{-1/2}(\boldsymbol{X}_i - \boldsymbol{\theta}_n)||^{-1})$.

Suppose $\lambda_{n,1}, \cdots, \lambda_{n,p_n}$ are all bounded, Condition (C4) becomes $\boldsymbol{\theta}_n^T \mathbf{D}_n{}^{-1} \boldsymbol{\theta}_n = O(n^{-1}p_n^{1/2})$ and $\boldsymbol{\theta}_n^T \mathbf{D}_n{}^{-1} \boldsymbol{\Sigma}_n \mathbf{D}_n{}^{-1} \boldsymbol{\theta}_n = O(n^{-1}p_n^3)$, i.e. $||\boldsymbol{\theta}_n||^2 = O(n^{-1}p_n^{1/2})$. If $\boldsymbol{\theta}_n = (\theta_{n,1}, \cdots, \theta_{n,p_n})$, $\theta_{n,i} = \delta_n$, $i = 1, \cdots, p_n$, we require that $\delta_n = O(n^{-1/2}p_n^{-1/4})$, which can be viewed as a high-dimensional version of the local alternative hypotheses.

Theorem 2. *Under Assumption (A1) and Conditions (C1)–(C4), as $\min(n, p_n) \to \infty$,*

$$\frac{T_n - c_0^2 \boldsymbol{\theta}_n^T \mathbf{D}_n^{-1} \boldsymbol{\theta}_n}{\sigma_n} \xrightarrow{d} N(0,1)$$

Theorem 1 and 2 allow us to compare the proposed test with some existing work in terms of limiting efficiency. The asymptotic power of our proposed test (abbreviated as SS) under the local alternative is

$$\beta_{SS}(\boldsymbol{\theta}_n) = \Phi\left(-z_\alpha + \frac{c_0^2 n p \boldsymbol{\theta}_n^T \mathbf{D}_n^{-1} \boldsymbol{\theta}_n}{\sqrt{2\mathrm{tr}(\mathbf{R}_n^2)}} \right).$$

Table 1
ARE(R_n, PA) with different ν.

	$\nu=3$	$\nu=4$	$\nu=5$	$\nu=6$	$\nu=\infty$
ARE	2.54	1.76	1.51	1.38	1.00

In comparison, [13] showed that the asymptotic power of their proposed test (abbreviated as PA hereafter) is

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

where $\tilde{\mathbf{D}}_n$ and $\tilde{\mathbf{R}}_n$ are the variance and correlation matrix of $\boldsymbol{X}_i$, respectively. Note that Park and Ayyala (2013) needed the diverging factor model for their asymptotic results. Direct power comparison for these two tests maybe not appropriate. So we reproof their results under assumption (A1) in the supplemental material. Consequently, the asymptotic relative efficiency (ARE) of T_n with PA test is

$$\mathrm{ARE}(T_n,\mathrm{PA})=\frac{c_0^2p_n\boldsymbol{\theta}_n^T\mathbf{D}_n^{-1}\boldsymbol{\theta}_n}{\boldsymbol{\theta}_n^T\tilde{\mathbf{D}}_n^{-1}\boldsymbol{\theta}_n}\sqrt{\frac{\mathrm{tr}(\tilde{\mathbf{R}}_n^2)}{\mathrm{tr}(\mathbf{R}_n^2)}}=c_0^2E(||\boldsymbol{\epsilon}||^2).$$

where the last equality is followed by $\mathrm{tr}(\tilde{\mathbf{R}}_n^2)=\mathrm{tr}(\mathbf{R}_n^2)$ and $\tilde{\mathbf{D}}_n=p_n^{-1}E(||\boldsymbol{\epsilon}||^2)\mathbf{D}_n$. Similar to the proof of Theorem 2, we can show that $c_0=E(||\boldsymbol{\epsilon}||^{-1})(1+o(1))$ by Condition (C3). Thus,

$$\mathrm{ARE}(T_n,\mathrm{PA})=E^2(||\boldsymbol{\epsilon}||^{-1})E(||\boldsymbol{\epsilon}||^2).$$

If $\boldsymbol{X}_i$ are generated from multivariate t-distribution with ν degrees of freedom ($\nu>2$),

$$\mathrm{ARE}(T_n,\mathrm{PA})=\frac{2}{\nu-2}\left(\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\right)^2.$$

Table 1 reports the ARE with different ν. Under the multivariate normal distribution ($\nu=\infty$), our SS test is the same powerful as PA test. However, our SS test is much more powerful than PA test under the heavy-tailed distributions.

In contrast, [20] showed that the power of their test (abbreviated as WPL) is

$$\beta_{WPL}(\boldsymbol{\theta}_n)=\Phi\left(-z_\alpha+\frac{n\boldsymbol{\theta}_n^T\mathbf{A}^2\boldsymbol{\theta}_n}{\sqrt{2\mathrm{tr}(\mathbf{B}^2)}}\right)$$

where $\mathbf{A}=E(||\boldsymbol{\varepsilon}_i||^{-1}(\mathbf{I}_{p_n}-U(\boldsymbol{\varepsilon}_i)U(\boldsymbol{\varepsilon}_i)^T))$, $\mathbf{B}=E(U(\boldsymbol{\varepsilon}_i)U(\boldsymbol{\varepsilon}_i)^T)$ and $\boldsymbol{\varepsilon}_i=\boldsymbol{X}_i-\boldsymbol{\theta}_n$. First, if all the diagonal elements of $\boldsymbol{\Sigma}_n$ are equal, i.e. $\mathbf{D}_n=(\delta,\cdots,\delta)$, we can show that $\mathbf{A} = c_0\mathbf{I}_{p_n}(1+o(1))$ and $\text{tr}(\mathbf{B}^2) = p_n^{-2}\delta^2\text{tr}(\mathbf{R}_n^2)$ by Condition (C3). Then,

$$\beta_{WPL}(\boldsymbol{\theta}_n) = \Phi\left(-z_\alpha + \frac{c_0^2 np\boldsymbol{\theta}_n^T\mathbf{D}_n^{-1}\boldsymbol{\theta}_n}{\sqrt{2\text{tr}(\mathbf{R}_n^2)}}\right).$$

Thus, our SS test has the same power as WPL test in this case. However, their test is not scalar-invariant. To appreciate the effect of scalar-invariance, we consider the following representative cases. Let $\boldsymbol{\Sigma}_n$ be a diagonal matrix. The first half diagonal elements of $\boldsymbol{\Sigma}_n$ are all τ_1^2 and the rest diagonal elements are all τ_2^2. The mean only shift on the first half components, i.e. $\mu_i = \zeta, i = 1, \cdots, p_n/2$ and the others are zeros. Thus,

$$\beta_{SS}(\boldsymbol{\theta}_n) = \Phi\left(-z_\alpha + \frac{nE^2(||\boldsymbol{\epsilon}||^{-1})\zeta^2}{2\sqrt{2p_n}\tau_1^2}\right).$$

However, it is difficult to calculate the explicit form of β_{WPL} for arbitrary τ_1^2, τ_2^2. We only consider two special cases. If $\tau_1^2 \gg \tau_2^2$,

$$\beta_{WPL}(\boldsymbol{\theta}_n) \approx \Phi\left(-z_\alpha + \frac{nE^2(||\boldsymbol{\epsilon}||^{-1})\zeta^2}{2\sqrt{p_n}\tau_1^2}\right).$$

Thus, ARE(T_n,WPL) has a positive lower bound of $1/\sqrt{2}$. However, if $\tau_2^2 \gg \tau_1^2$,

$$\beta_{WPL}(\boldsymbol{\theta}_n) \approx \Phi\left(-z_\alpha + \frac{nE^2(||\boldsymbol{\epsilon}||^{-1})\zeta^2}{2\sqrt{p_n}\tau_2^2}\right).$$

Then, ARE(T_n,WPL)$=\tau_2^2/(\sqrt{2}\tau_1^2)$ could be very large. This property shows the necessity of a test with the scale-invariance property.

3. Simulation

Here we report a simulation study designed to evaluate the performance of the proposed SS test. All the simulation results are based on 2,500 replications. The number of variety of multivariate distributions and parameters are too large to allow a comprehensive, all-encompassing comparison. We choose certain representative examples for illustration. The following scenarios are firstly considered.

(I) Multivariate normal distribution. $\boldsymbol{X}_i \sim N(\boldsymbol{\theta}, \mathbf{R}_n)$.

(II) Multivariate normal distribution with different component variances. $\boldsymbol{X}_i \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma}_n)$, where $\boldsymbol{\Sigma}_n = \mathbf{D}_n^{1/2}\mathbf{R}_n\mathbf{D}_n^{1/2}$ and $\mathbf{D}_n = \text{diag}\{d_1^2, \cdots, d_{p_n}^2\}$, $d_j^2 = 3$, $j \le p_n/2$ and $d_j^2 = 1$, $j > p_n/2$.

(III) Multivariate t-distribution $t_{p_n,4}$. $\boldsymbol{X}_i$'s are generated from $t_{p_n,4}$ with $\boldsymbol{\Sigma}_n = \mathbf{R}_n$.

(IV) Multivariate t-distribution with different component variances. $\boldsymbol{X}_i$'s are generated from $t_{p_n,4}$ with $\boldsymbol{\Sigma}_n = \mathbf{D}_n^{1/2}\mathbf{R}_n\mathbf{D}_n^{1/2}$ and d_j^2's are generated from χ_4^2.

Table 2
Empirical sizes and power (%) comparison at 5% significance under Scenarios (I)-(V)

		Size				Dense				Sparse			
n	p_n	SS	PA	WPL	SD	SS	PA	WPL	SD	SS	PA	WPL	SD
Scenario I													
50	200	5.4	6.2	5.2	5.2	29	31	29	28	31	33	31	29
50	400	6.5	7.3	6.6	6.8	29	33	30	29	31	34	31	32
50	1000	4.7	6.9	5.7	8.9	25	33	29	38	25	32	29	38
100	200	6.2	6.5	6.3	5.4	61	63	61	60	66	68	66	64
100	400	5.9	6.2	5.1	5.1	63	64	63	60	67	68	68	65
100	1000	5.3	6.2	5.3	4.8	63	65	64	62	66	67	67	64
Scenario II													
50	200	5.6	6.3	5.7	5.1	63	66	25	63	70	72	29	70
50	400	6.5	7.3	5.4	6.7	69	72	28	70	69	72	30	70
50	1000	4.6	6.9	6.1	8.7	67	74	28	79	66	73	28	79
100	200	5.9	6.3	5.0	5.4	95	96	62	95	98	97	70	97
100	400	5.9	6.1	6.2	5.2	97	97	66	96	98	98	71	98
100	1000	5.3	6.2	6.3	4.9	98	98	67	98	98	99	68	99
Scenario III													
50	200	5.3	4.3	5.2	2.4	51	36	51	12	58	40	58	12
50	400	6.4	6.9	6.6	1.1	54	39	55	10	57	41	58	10
50	1000	4.4	7.1	5.7	0.9	51	37	56	5.3	51	39	57	4.8
100	200	6.0	7.8	6.3	2.3	89	69	89	42	93	72	93	45
100	400	5.9	6.6	5.1	1.2	91	66	91	27	93	71	93	29
100	1000	5.4	6.1	5.3	1.0	93	69	93	7.5	94	70	95	7.7
Scenario IV													
50	200	5.3	4.3	6.5	2.4	87	67	50	32	92	74	58	41
50	400	6.4	6.9	5.9	1.1	89	74	56	21	97	88	60	42
50	1000	4.4	7.1	6.1	0.8	99	88	57	12	99	92	58	17
100	200	6.0	7.8	6.0	2.2	100	94	90	80	100	93	94	77
100	400	5.9	6.6	6.2	1.3	100	94	92	67	100	95	94	67
100	1000	5.4	6.1	5.9	0.9	100	99	94	51	100	97	96	42
Scenario V													
50	200	5.5	6.0	5.2	1.2	44	34	44	10	51	38	50	8.9
50	400	6.5	6.8	6.6	0.5	49	39	50	5.8	50	40	52	6.8
50	1000	4.6	7.1	5.7	0.8	44	40	50	4.5	45	39	50	5.2
100	200	6.2	6.6	6.3	1.6	84	67	84	38	89	70	89	44
100	400	5.9	5.1	5.1	1.1	87	66	87	19	89	71	90	21
100	1000	5.3	5.8	5.3	0.9	89	68	89	5.2	90	71	91	6.3

(V) Multivariate mixture normal distribution $\mathrm{MN}_{p_n,\gamma,9}$. $\boldsymbol{X}_i$'s are generated from $\gamma f_{p_n}(\boldsymbol{\theta},\mathbf{R}_n)+(1-\gamma)f_{p_n}(\boldsymbol{\theta},9\mathbf{R}_n)$, denoted by $\mathrm{MN}_{p_n,\gamma,9}$, where $f_{p_n}(\cdot;\cdot)$ is the density function of p_n-variate multivariate normal distribution. γ is chosen to be 0.9.

Here we consider the correlation matrix $\mathbf{R}_n=(0.5^{|i-j|})_{1\le i,j\le p_n}$. Two sample sizes $n=50,100$ and three dimensions $p_n=200,400,1000$ are considered. For power comparison, under H_1, we consider two patterns of allocation for $\boldsymbol{\theta}_n$. One is dense case, i.e. the first 50% components of $\boldsymbol{\theta}_n$ are zeros. The other is sparse case, i.e. the first 95% components of $\boldsymbol{\theta}_n$ are zeros. To make the power comparable among the configurations of H_1, we set $\eta =: ||\boldsymbol{\theta}_n||^2/\sqrt{\mathrm{tr}^2(\boldsymbol{\Sigma}_n)}=0.03$ throughout the simulation. And the nonzeros components of $\boldsymbol{\theta}_n$ are all equal. Table 2 reports the empirical sizes and power of SS, PA, WPL and [18]'s test (abbreviated as SD hereafter) for multivariate normal (Scenario I and II) and non-normal (Scenario III, IV and V) distributions, respectively. From Table 2, we observe that our SS test can control the empirical sizes very well in all cases. WPL test can also maintain the significant level very well. However, the empirical sizes of the PA tests are a little larger than the nominal level in many cases, especially for the non-normal distributions. And the empirical sizes of the SD tests are smaller than the nominal lever for the non-normal distributions. Under Scenario I and II, PA and SD test has some advantages over SS as we would expert because the underlying distribution is multivariate normal. However, under the non-normal distributions, our SS test performs significantly better than PA and SD test. It is consistent with the theoretical results in Section 2. When the component variances are same (Scenario I, III and V), the power of our SS test is similar to WPL test. Even we need to estimate the scalar matrix, we do not lose much efficiency in these cases. However, when the component variances are not equal (Scenarios (II) and (IV)), our SS test, even PA test, are much more powerful than WPL test, which further shows that a scalar-invariant test is necessary. All these results show that our SS test is very powerful and robust in a wide range of distributions.

Appendix

Because Theorem 1 is a special case of Theorem 2 with $\boldsymbol{\theta}_n = \mathbf{0}$, we only need to proof Theorem 2. First, we restate the Lemma 4 in [21] here.

Lemma 1. *Suppose $\boldsymbol{u}$ are independent identically distributed uniform on the unit p sphere. For any $p \times p$ symmetric matrix $\mathbf{M}$, we have*

$$
\begin{aligned}
E(\boldsymbol{u}^T\mathbf{M}\boldsymbol{u})^2 =&\{\mathrm{tr}^2(\mathbf{M}) + 2\mathrm{tr}(\mathbf{M}^2)\}/(p^2+2p),\\
E(\boldsymbol{u}^T\mathbf{M}\boldsymbol{u})^4 =&\{3\mathrm{tr}^2(\mathbf{M}^2) + 6\mathrm{tr}(\mathbf{M}^4)\}/\{p(p+2)(p+4)(p+6)\}.
\end{aligned}
$$

A.1. Proof of Theorem 2

Define $\boldsymbol{U}_i = U(\mathbf{D}_n^{-1/2}(\boldsymbol{X}_i - \boldsymbol{\theta}_n))$, $r_i = ||\mathbf{D}_n^{-1/2}(\boldsymbol{X}_i - \boldsymbol{\theta}_n)||$ and $\boldsymbol{u}_i = U(\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{X}_i - \boldsymbol{\theta}_n))$. By the definition of $U(\hat{\mathbf{D}}_{ij}^{-1/2}\boldsymbol{X}_i)$,

$$
\begin{aligned}
&U(\hat{\mathbf{D}}_{ij}^{-1/2}\boldsymbol{X}_i)\\
&\quad= (\mathbf{D}_n^{-1/2}(\boldsymbol{X}_i - \boldsymbol{\theta}_n) + \mathbf{D}_n^{-1/2}\boldsymbol{\theta}_n + (\hat{\mathbf{D}}_{ij}^{-1/2} - \mathbf{D}_n^{-1/2})\boldsymbol{X}_i)\\
&\qquad \times ||\mathbf{D}_n^{-1/2}(\boldsymbol{X}_i - \boldsymbol{\theta}_n) + \mathbf{D}_n^{-1/2}\boldsymbol{\theta}_n + (\hat{\mathbf{D}}_{ij}^{-1/2} - \mathbf{D}_n^{-1/2})\boldsymbol{X}_i||^{-1}\\
&\quad= (\boldsymbol{U}_i + r_i^{-1}\mathbf{D}_n^{-1/2}\boldsymbol{\theta}_n + (\hat{\mathbf{D}}_{ij}^{-1/2}\mathbf{D}_n^{1/2} - \mathbf{I}_{p_n})\boldsymbol{U}_i)\\
&\qquad \times (1 + 2r_i^{-1}\boldsymbol{U}_i^T(\hat{\mathbf{D}}_{ij}^{-1/2} - \mathbf{D}_n^{-1/2})\boldsymbol{X}_i + r_i^{-2}||(\hat{\mathbf{D}}_{ij}^{-1/2} - \mathbf{D}_n^{-1/2})\boldsymbol{X}_i||^2\\
&\qquad + 2r_i^{-1}\mathbf{D}_n^{-1/2}\boldsymbol{\theta}_n + r_i^{-2}\boldsymbol{\theta}_n^T\mathbf{D}_n^{-1}\boldsymbol{\theta}_n)^{-1/2}.
\end{aligned}
$$

And

$$
\begin{aligned}
&\frac{2}{n(n-1)}\sum\sum_{i<j} U(\hat{\mathbf{D}}_{ij}^{-1/2}\boldsymbol{X}_i)^T U(\hat{\mathbf{D}}_{ij}^{-1/2}\boldsymbol{X}_j)\\
&= \frac{2}{n(n-1)}\sum\sum_{i<j}(\boldsymbol{U}_i + r_i^{-1}\mathbf{D}_n^{-1/2}\boldsymbol{\theta}_n + (\hat{\mathbf{D}}_{ij}^{-1/2}\mathbf{D}_n^{1/2} - \mathbf{I}_{p_n})\boldsymbol{U}_i)^T\\
&\quad \times (\boldsymbol{U}_j + r_j^{-1}\mathbf{D}_n^{-1/2}\boldsymbol{\theta}_n + (\hat{\mathbf{D}}_{ij}^{-1/2}\mathbf{D}_n^{1/2} - \mathbf{I}_{p_n})\boldsymbol{U}_j)(1+\alpha_{ij})^{-1/2}(1+\alpha_{ji})^{-1/2}\\
&= \frac{2}{n(n-1)}\sum\sum_{i<j}\boldsymbol{U}_i^T\boldsymbol{U}_j + \frac{2}{n(n-1)}\sum\sum_{i<j} r_i^{-1}r_j^{-1}\boldsymbol{\theta}_n^T\mathbf{D}_n^{-1}\boldsymbol{\theta}_n\\
&\quad + \frac{2}{n(n-1)}\sum\sum_{i<j}\boldsymbol{U}_i^T\boldsymbol{U}_j[(1+\alpha_{ij})^{-1/2}(1+\alpha_{ji})^{-1/2} - 1]\\
&\quad + \frac{4}{n(n-1)}\sum\sum_{i<j}\boldsymbol{U}_i^T(\hat{\mathbf{D}}_{ij}^{-1/2}\mathbf{D}_n^{1/2} - \mathbf{I}_{p_n})\boldsymbol{U}_j(1+\alpha_{ij})^{-1/2}(1+\alpha_{ji})^{-1/2}\\
&\quad + \frac{2}{n(n-1)}\sum\sum_{i<j}\boldsymbol{U}_i^T(\hat{\mathbf{D}}_{ij}^{-1/2}\mathbf{D}_n^{1/2} - \mathbf{I}_{p_n})^2\boldsymbol{U}_j(1+\alpha_{ij})^{-1/2}(1+\alpha_{ji})^{-1/2}\\
&\quad + \frac{2}{n(n-1)}\sum\sum_{i<j} r_i^{-1}r_j^{-1}\boldsymbol{\theta}_n^T\mathbf{D}_n^{-1}\boldsymbol{\theta}_n[(1+\alpha_{ij})^{-1/2}(1+\alpha_{ji})^{-1/2} - 1]\\
&\quad + \frac{4}{n(n-1)}\sum\sum_{i<j} r_j^{-1}\boldsymbol{U}_i^T\mathbf{D}_n^{-1/2}\boldsymbol{\theta}_n(1+\alpha_{ij})^{-1/2}(1+\alpha_{ji})^{-1/2}\\
&\quad + \frac{2}{n(n-1)}\sum\sum_{i<j}\boldsymbol{U}_i^T(\hat{\mathbf{D}}_{ij}^{-1/2}\mathbf{D}_n^{1/2} - \mathbf{I}_{p_n})\mathbf{D}_n^{-1/2}\boldsymbol{\theta}_n(1+\alpha_{ij})^{-1/2}(1+\alpha_{ji})^{-1/2}\\
&\doteq \frac{2}{n(n-1)}\sum\sum_{i<j}\boldsymbol{U}_i^T\boldsymbol{U}_j + \frac{2}{n(n-1)}\sum\sum_{i<j} r_i^{-1}r_j^{-1}\boldsymbol{\theta}_n^T\mathbf{D}_n^{-1}\boldsymbol{\theta}_n\\
&\quad + A_{n1} + A_{n2} + A_{n3} + A_{n4} + A_{n5} + A_{n6}
\end{aligned}
$$

where $\alpha_{ij} = 2r_i^{-1}\boldsymbol{U}_i^T(\hat{\mathbf{D}}_{ij}^{-1/2} - \mathbf{D}_n^{-1/2})\boldsymbol{X}_i + r_i^{-2}||(\hat{\mathbf{D}}_{ij}^{-1/2} - \mathbf{D}_n^{-1/2})\boldsymbol{X}_i||^2 + 2r_i^{-1}\mathbf{D}_n^{-1/2}\boldsymbol{\theta}_n + r_i^{-2}\boldsymbol{\theta}_n^T\mathbf{D}_n^{-1}\boldsymbol{\theta}_n)^{-1/2}$. Note that $r_i^{-1}\boldsymbol{U}_i^T(\hat{\mathbf{D}}_{ij}^{-1/2} - \mathbf{D}_n^{-1/2})\boldsymbol{X}_i = \boldsymbol{U}_i^T(\hat{\mathbf{D}}_{ij}^{-1/2}\mathbf{D}_n^{1/2} - \mathbf{I}_{p_n})\boldsymbol{U}_i + r_i^{-1}\boldsymbol{U}_i^T(\hat{\mathbf{D}}_{ij}^{-1/2} - \mathbf{D}_n^{-1/2})\boldsymbol{\theta}_n = O_p(n^{-1/2}(\log p_n)^{1/2})$ and $r_i^{-2}||(\hat{\mathbf{D}}_{ij}^{-1/2} - \mathbf{D}_n^{-1/2})\boldsymbol{X}_i||^2 = O_p(n^{-1}\log p_n)$ by Lemma 2 in [3]. By Condition (C2) and (C4), $r_i^{-1}\mathbf{D}_n^{-1/2}\boldsymbol{\theta}_n = O_p(\sigma_n^{1/2}) = O_p(n^{-1})$ and $r_i^{-2}\boldsymbol{\theta}_n^T\mathbf{D}_n^{-1}\boldsymbol{\theta}_n = O_p(\sigma_n) = O_p(n^{-2})$ where $\sigma_n^2 = \frac{2}{n(n-1){p_n}^2}\mathrm{tr}(\mathbf{R}_n^2)$. So $\alpha_{ij} = O_p(n^{-1/2}(\log p_n)^{1/2})$. First, we will show that $A_{n1} = o_p(\sigma_n)$. By the Cauchy inequality,

$$
\begin{aligned}
E(A_{n1}^2) &= O(n^{-4})\sum\sum_{i<j} E\{\boldsymbol{U}_i^T\boldsymbol{U}_j[(1+\alpha_{ij})^{-1/2}(1+\alpha_{ji})^{-1/2} - 1]\}^2\\
&\le O(n^{-2})E(\boldsymbol{U}_i^T\boldsymbol{U}_j)^2 E[(1+\alpha_{ij})^{-1/2}(1+\alpha_{ji})^{-1/2} - 1]^2\\
&= O(n^{-3}p_n^{-2}\log p_n \mathrm{tr}(\mathbf{R}_n^2)) = o(\sigma_n^2).
\end{aligned}
$$

where $E(\boldsymbol{U}_i^T\boldsymbol{U}_j)^2 = O(p_n^{-2}\mathrm{tr}(\mathbf{R}_n^2))$ follows from the upcoming statements. And

$$\begin{aligned}
&A_{n2}\\
&= \frac{4}{n(n-1)}\sum\sum_{i<j}\boldsymbol{U}_i^T(\hat{\mathbf{D}}_{ij}^{-1/2}\mathbf{D}_n^{1/2}-\mathbf{I}_{p_n})\boldsymbol{U}_j\\
&\quad + \frac{4}{n(n-1)}\sum\sum_{i<j}\boldsymbol{U}_i^T(\hat{\mathbf{D}}_{ij}^{-1/2}\mathbf{D}_n^{1/2}-\mathbf{I}_{p_n})\boldsymbol{U}_j[(1+\alpha_{ij})^{-1/2}(1+\alpha_{ji})^{-1/2}-1]\\
&\doteq G_{n1}+G_{n2}.
\end{aligned}$$

Next we will show that $E(G_{n1}^2) = o(\sigma_n^2)$.

$$\begin{aligned}
&E\left(G_{n1}^2\right)\\
&= O(n^{-4})\sum\sum_{i<j}E\left(\left(\boldsymbol{U}_i^T(\hat{\mathbf{D}}_{ij}^{-1/2}\mathbf{D}_n^{1/2}-\mathbf{I}_{p_n})\boldsymbol{U}_j\right)^2\right)\\
&= O(n^{-4})\sum\sum_{i<j}E\left(\frac{\left(\boldsymbol{u}_i^T{\boldsymbol{\Sigma}_n}^{1/2}\mathbf{D}_n^{-1/2}(\hat{\mathbf{D}}_{ij}^{-1/2}\mathbf{D}_n^{1/2}-\mathbf{I}_{p_n})\mathbf{D}_n^{-1/2}{\boldsymbol{\Sigma}_n}^{1/2}\boldsymbol{u}_j\right)^2}{(1+\boldsymbol{u}_i^T(\mathbf{R}_n-\mathbf{I}_{p_n})\boldsymbol{u}_i)(1+\boldsymbol{u}_j^T(\mathbf{R}_n-\mathbf{I}_{p_n})\boldsymbol{u}_j)}\right)\\
&\le O(n^{-4})\sum\sum_{i<j}\left\{E\left(\boldsymbol{u}_i^T{\boldsymbol{\Sigma}_n}^{1/2}\mathbf{D}_n^{-1/2}(\hat{\mathbf{D}}_{ij}^{-1/2}\mathbf{D}_n^{1/2}-\mathbf{I}_{p_n})\mathbf{D}_n^{-1/2}{\boldsymbol{\Sigma}_n}^{1/2}\boldsymbol{u}_j\right)^2\right.\\
&\quad + CE\left(\left(\boldsymbol{u}_i^T{\boldsymbol{\Sigma}_n}^{1/2}\mathbf{D}_n^{-1/2}(\hat{\mathbf{D}}_{ij}^{-1/2}\mathbf{D}_n^{1/2}-\mathbf{I}_{p_n})\mathbf{D}_n^{-1/2}{\boldsymbol{\Sigma}_n}^{1/2}\boldsymbol{u}_j\right)^2\right.\\
&\quad \left.\left.\times\, \boldsymbol{u}_i^T(\mathbf{R}_n-\mathbf{I}_{p_n})\boldsymbol{u}_i\right)\right\},
\end{aligned}$$

where the last inequality follows by the Taylor expansion and C is a constant. Define $\mathbf{H} = {\boldsymbol{\Sigma}_n}^{1/2}\mathbf{D}_n^{-1/2}(\hat{\mathbf{D}}_{ij}^{-1/2}\mathbf{D}_n^{1/2}-\mathbf{I}_{p_n})\mathbf{D}_n^{-1/2}{\boldsymbol{\Sigma}_n}^{1/2}$ and then according to Lemma 2 in [3], $\mathrm{tr}(E(\mathbf{H}^2)) = o(\mathrm{tr}(\mathbf{R}_n^2))$ and $\mathrm{tr}(E(\mathbf{H}^4)) = o(\mathrm{tr}(\mathbf{R}_n^4)) = o(\mathrm{tr}^2(\mathbf{R}_n^2))$ by Condition (C1). By the Cauchy inequality, we have

$$\begin{aligned}
E\big(\boldsymbol{u}_i^T{\boldsymbol{\Sigma}_n}^{1/2}\mathbf{D}_n^{-1/2}(\hat{\mathbf{D}}_{ij}^{-1/2}&\mathbf{D}_n^{1/2}-\mathbf{I}_{p_n})\mathbf{D}_n^{-1/2}{\boldsymbol{\Sigma}_n}^{1/2}\boldsymbol{u}_j\big)^2\\
&= {p_n}^{-2}E(\mathrm{tr}(\mathbf{H}^2)) = o({p_n}^{-2}\mathrm{tr}(\mathbf{R}_n^2)),\\
E((\boldsymbol{u}_i^T\mathbf{H}\boldsymbol{u}_j)^2\boldsymbol{u}_i^T(\mathbf{R}_n-\mathbf{I}_{p_n})\boldsymbol{u}_i) &\le (E(\boldsymbol{u}_i^T\mathbf{H}\boldsymbol{u}_j)^4E((\boldsymbol{u}_i^T(\mathbf{R}_n-\mathbf{I}_{p_n})\boldsymbol{u}_i)^2)^{1/2}\\
&\le ({p_n}^{-4}\mathrm{tr}(E(\mathbf{H}^4)){p_n}^{-2}(\mathrm{tr}(\mathbf{R}_n-\mathbf{I}_{p_n})^2))^{1/2}\\
&= o({p_n}^{-2}\mathrm{tr}(\mathbf{R}_n^2)),\\
E((\boldsymbol{u}_i^T\mathbf{H}\boldsymbol{u}_j)^2\boldsymbol{u}_j^T(\mathbf{R}_n-\mathbf{I}_{p_n})\boldsymbol{u}_j) &\le (E(\boldsymbol{u}_i^T\mathbf{H}\boldsymbol{u}_j)^4E((\boldsymbol{u}_j^T(\mathbf{R}_n-\mathbf{I}_{p_n})\boldsymbol{u}_j)^2)^{1/2}\\
&\le ({p_n}^{-4}\mathrm{tr}(E(\mathbf{H}^4)){p_n}^{-2}(\mathrm{tr}(\mathbf{R}_n-\mathbf{I}_{p_n})^2))^{1/2}\\
&= o({p_n}^{-2}\mathrm{tr}(\mathbf{R}_n^2)).
\end{aligned}$$

So we obtain that $G_{n1} = o_p(\sigma_n)$. Similar to A_{n1}, we can show that $G_{n2} = o_p(\sigma_n)$ and then $A_{n2} = o_p(\sigma_n)$. Taking the same procedure as A_{n2}, we can also obtain

$A_{n3} = o_p(\sigma_n)$. Moreover, by taking the same procedure to $\boldsymbol{u}_i^T(\mathbf{R}_n - \mathbf{I}_{p_n})\boldsymbol{u}_i$ as G_{n1},

$$\begin{aligned}
&\frac{2}{n(n-1)}\sum\sum_{i<j} \boldsymbol{U}_i^T \boldsymbol{U}_j \\
&= \frac{2}{n(n-1)}\sum\sum_{i<j} \frac{\boldsymbol{u}_i^T \boldsymbol{\Sigma}_n{}^{1/2}\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n{}^{1/2}\boldsymbol{u}_j}{\sqrt{1+\boldsymbol{u}_i^T(\mathbf{R}_n - \mathbf{I}_{p_n})\boldsymbol{u}_i}\sqrt{1+\boldsymbol{u}_j^T(\mathbf{R}_n - \mathbf{I}_{p_n})\boldsymbol{u}_j}} \\
&= \frac{2}{n(n-1)}\sum\sum_{i<j} \boldsymbol{u}_i^T \boldsymbol{\Sigma}_n{}^{1/2}\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n{}^{1/2}\boldsymbol{u}_j + o_p(\sigma_n).
\end{aligned}$$

Next, we will show that

$$\sqrt{\frac{n(n-1)p_n^2}{2\mathrm{tr}(\mathbf{R}_n^2)}}\frac{2}{n(n-1)}\sum\sum_{i<j} \boldsymbol{u}_i^T \boldsymbol{\Sigma}_n^{1/2}\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n^{1/2}\boldsymbol{u}_j \xrightarrow{d} N(0,1)$$

Define $W_{nk} = \sum_{i=2}^k Z_{ni}$ where $Z_{ni} = \sum_{j=1}^{i-1} \frac{2}{n(n-1)}\boldsymbol{u}_i^T \boldsymbol{\Sigma}_n^{1/2}\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n^{1/2}\boldsymbol{u}_j$. Let $\mathcal{F}_{n,i} = \sigma\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_i\}$ be the σ-field generated by $\{\boldsymbol{u}_j, j \le i\}$. Obviously, $E(Z_{ni}|\mathcal{F}_{n,i-1}) = 0$ and it follows that $\{W_{nk}, \mathcal{F}_{n,k}; 2 \le k \le n\}$ is a zero mean martingale. The central limit theorem [6] will hold if we can show

$$\frac{\sum_{j=2}^n E[Z_{nj}^2|\mathcal{F}_{n,j-1}]}{\sigma_n^2} \xrightarrow{p} 1. \tag{A.1}$$

and for any $\epsilon > 0$,

$$\sigma_n^{-2}\sum_{j=2}^n E[Z_{nj}^2 I(|Z_{nj}| > \epsilon\sigma_n|)|\mathcal{F}_{n,j-1}] \xrightarrow{p} 0. \tag{A.2}$$

It can be shown that

$$\begin{aligned}
&\sum_{j=2}^n E(Z_{nj}^2|\mathcal{F}_{n,j-1}) \\
&= \frac{4}{n^2(n-1)^2}\sum_{j=2}^n\sum_{i=1}^{j-1} \boldsymbol{u}_i^T \boldsymbol{\Sigma}_n^{1/2}\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n^{1/2}\boldsymbol{u}_i \\
&\quad + \frac{4}{n^2(n-1)^2}\sum_{j=2}^n\sum_{i_1<i_2}^{j-1}\sum^{j-1} \boldsymbol{u}_{i_1}^T \boldsymbol{\Sigma}_n^{1/2}\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n^{1/2}\boldsymbol{u}_{i_2} \\
&\doteq C_{n1} + C_{n2}.
\end{aligned}$$

Simple algebras lead to

$$E(C_{n1}) = \sigma_n^2,$$

$$\mathrm{var}(C_{n1}) = \frac{16}{n^4(n-1)^4}\sum_{j=1}^{n-1} j^2[E\{(\boldsymbol{u}_j^T\boldsymbol{\Sigma}_n^{1/2}\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n^{1/2}\boldsymbol{u}_j)^2\} - {p_n}^{-2}\mathrm{tr}^2(\mathbf{R}_n^2)].$$

By Lemma 1, $E((\boldsymbol{u}_j^T\boldsymbol{\Sigma}_n^{1/2}\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n^{1/2}\boldsymbol{u}_j)^2) = O(p_n^{-2}\mathrm{tr}^2(\mathbf{R}_n^2))$. Thus, $\mathrm{var}(C_{n2}) = o(\sigma_n^4)$. Then, $C_{n1}/\sigma_n^2 \xrightarrow{p} 1$. Similarly, $E(C_{n2}) = 0$ and

$$\frac{\mathrm{var}(C_{n2})}{\sigma_n^2} = \frac{32}{n^4(n-1)^4}\sum_{i=3}^{n}\frac{i(n-i+1)(i-1)}{2}\frac{\mathrm{tr}(\mathbf{R}_n^4)}{\mathrm{tr}^2(\mathbf{R}_n^2)} \xrightarrow{p} 0$$

implies $C_{n2} = o_p(\sigma_n^2)$. Thus, (A.1) holds. It remains to show (A.2). Note that

$$\sigma_n^{-2}\sum_{j=2}^{n} E[Z_{nj}^2 I(|Z_{nj}| > \epsilon\sigma_n)|\mathcal{F}_{n,j-1}] \le \sigma_n^{-4}\epsilon^{-2}\sum_{j=2}^{n} E[Z_{nj}^4|\mathcal{F}_{n,j-1}].$$

Accordingly, the assertion of this lemma is true if we can show

$$E\left\{\sum_{j=2}^{n} E[Z_{nj}^4|\mathcal{F}_{n,j-1}]\right\} = o(\sigma_n^4).$$

Note that

$$\begin{aligned} E\left\{\sum_{j=2}^{n} E[Z_{nj}^4|\mathcal{F}_{n,j-1}]\right\} &= \sum_{j=2}^{n} E(Z_{nj}^4) \\ &= O(n^{-8})\sum_{j=2}^{n} E\left(\sum_{i=1}^{j-1}\boldsymbol{u}_j^T\boldsymbol{\Sigma}_n^{1/2}\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n^{1/2}\boldsymbol{u}_i\right)^4. \end{aligned}$$

which can be decomposed as $3Q + P$ where

$$\begin{aligned} Q =& O(n^{-8})\sum_{j=2}^{n}\sum_{s<t}^{j-1}\sum^{j-1} E(\boldsymbol{u}_j^T\boldsymbol{\Sigma}_n^{1/2}\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n^{1/2}\boldsymbol{u}_s\boldsymbol{u}_s^T\boldsymbol{\Sigma}_n^{1/2}\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n^{1/2}\boldsymbol{u}_j \\ &\times \boldsymbol{u}_j^T\boldsymbol{\Sigma}_n^{1/2}\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n^{1/2}\boldsymbol{u}_t\boldsymbol{u}_t^T\boldsymbol{\Sigma}_n^{1/2}\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n^{1/2}\boldsymbol{u}_j) \\ P =& O(n^{-8})\sum_{j=2}^{n}\sum_{i=1}^{j-1} E((\boldsymbol{u}_j^T\boldsymbol{\Sigma}_n^{1/2}\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n^{1/2}\boldsymbol{u}_i)^4) \end{aligned}$$

So $Q = O(n^{-5}{p_n}^{-2}E((\boldsymbol{u}_j^T\boldsymbol{\Sigma}_n^{1/2}\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n^{1/2}\boldsymbol{u}_j)^2)) = O(n^{-5}{p_n}^{-4}\mathrm{tr}^2(\mathbf{R}_n^2)) = o(\sigma_n^4)$. Define $\boldsymbol{\Sigma}_n^{1/2}\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n^{1/2} = (v_{ij})_{1\le i,j\le p_n}$.

$$E(\boldsymbol{u}_s^T\mathbf{A}_3\boldsymbol{u}_t)^4 = E\left(\sum_{i=1}^{p_n}\sum_{j=1}^{p_n} v_{ij}u_{si}u_{tj}\right)^4$$

$$= \sum_{i_1,\ldots,i_4=1}^{p_n} \sum_{j_1,\ldots,j_4=1}^{p_n} v_{i_1j_1}v_{i_2j_2}v_{i_3j_3}v_{i_4j_4}E(u_{si_1}u_{si_2}u_{si_3}u_{si_4})E(u_{tj_1}u_{tj_2}u_{tj_3}u_{tj_4})$$
$$= O({p_n}^{-4}) \sum_{i_1,\ldots,i_4=1}^{p_n} \sum_{j_1,\ldots,j_4=1}^{p_n} v_{i_1j_1}v_{i_2j_2}v_{i_3j_3}v_{i_4j_4}.$$

By the Cauchy inequality, we have

$$\begin{aligned}
&\sum_{i_1,i_2,i_3,i_4=1}^{p_n} \sum_{j_1,j_2,j_3,j_4=1}^{p_n} v_{i_1j_1}v_{i_2j_2}v_{i_3j_3}v_{i_4j_4} \\
&\qquad \le \frac{1}{4} \sum_{i_1,i_2,i_3,i_4=1}^{p_n} \sum_{j_1,j_2,j_3,j_4=1}^{p_n} (v_{i_1j_1}^2 + v_{i_2j_2}^2)(v_{i_3j_3}^2 + v_{i_4j_4}^2) \\
&\qquad = \sum_{i_1,i_2,j_1,j_2=1}^{p_n} v_{i_1j_1}^2 v_{i_2j_2}^2 = \left(\sum_{i_1,j_1} v_{i_1j_1}^2\right)^2 = \mathrm{tr}^2(\mathbf{R}_n^2).
\end{aligned}$$

Thus, $P = O(n^{-6}{p_n}^{-4}\mathrm{tr}(\mathbf{R}_n^2)) = o(\sigma_n^4)$. So, we obtain that

$$\frac{2}{n(n-1)}\sum\sum_{i<j} \boldsymbol{U}_i^T\boldsymbol{U}_j \xrightarrow{d} N(0, \sigma_n^2).$$

Obviously,

$$\frac{2}{n(n-1)}\sum\sum_{i<j} r_i^{-1}r_j^{-1}\boldsymbol{\theta}_n^T\mathbf{D}_n^{-1}\boldsymbol{\theta}_n = c_0^2\boldsymbol{\theta}_n^T\mathbf{D}_n^{-1}\boldsymbol{\theta}_n + o_p(\sigma_n),$$

and $A_{n4} = o_p(\sigma_n)$ by the same arguments as A_{n1}. Similarly,

$$\begin{aligned}
A_{n5} &= O_p((c_0^2 n^{-1}p_n^{-1}\boldsymbol{\theta}_n^T\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n\mathbf{D}_n^{-1}\boldsymbol{\theta}_n)^{1/2}) = o_p(\sigma_n), \\
A_{n6} &= O_p((c_0^2 n^{-3/2}p_n^{-1}\log p_n\boldsymbol{\theta}_n^T\mathbf{D}_n^{-1}\boldsymbol{\Sigma}_n\mathbf{D}_n^{-1}\boldsymbol{\theta}_n)^{1/2}) = o_p(\sigma_n).
\end{aligned}$$

So

$$\frac{T_n - c_0^2\boldsymbol{\theta}_n^T\mathbf{D}_n^{-1}\boldsymbol{\theta}_n}{\sigma_n} \xrightarrow{d} N(0,1).$$

Here we complete the proof. □

Acknowledgements

This work is supported by the Fundamental Research Funds for the Central Universities and the NNSF of China grants 11501092, 11101074, 11271205 and 11471069.

References

[1] Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, **6**, 311–29. MR1399305

[2] 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. MR2604697

[3] Feng, L., Zou, C. and Wang, Z. (2016). Multivariate-sign-based high-dimensional tests for the two-sample location problem. *Journal of the American Statistical Association*, To appear. MR3538700

[4] Feng, L., Zou, C., Wang, Z. and Zhu, L. X. (2016). Two sample behrens-fisher problem for high-dimensional data. *Statistica Sinica*, **25**, 1297–1312. MR3409068

[5] 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. MR3367268

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

[7] Hallin, M. and Paindaveine, D. (2002). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *The Annals of Statistics*, **30**, 1103–1133. MR1926170

[8] Hallin, M. and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape. I: optimal rank-based tests for sphericity. *The Annals of Statistics*, **34**, 2707–2756. MR2329465

[9] Hettmansperger, T. P. and Oja, H. (1994). Affine invariant multivariate multisample sign tests. *Journal of the Royal Statistical Society, Series B,* **56**, 235–249. MR1257810

[10] Hettmansperger, T. P. and Randles, R. H. (2002). A practical affine equivariant multivariate median. *Biometrika*, **89**, 851–860. MR1946515

[11] Möttönen J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *Journal of Nonparametric Statistics,* **5** 201–213. MR1346895

[12] Oja, H. (2010). *Multivariate Nonparametric Methods with R*. New York: Springer. MR2598854

[13] 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. MR3011304

[14] Paindaveine, D. and Verdebout, T. (2016). On high-dimensional sign tests. *Bernoulli*, **22**, 1745–1769. MR3474832

[15] Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. New York: Wiley. MR0298844

[16] Randles, R. H. (1992). A two sample extension of the multivariate interdirection sign test. *L1-Statistical Analysis and Related Methods (ed. Y. Dodge)*, Elsevier, Amsterdam, pp. 295–302. MR1214840

[17] Randles, R. H. (2000). A simpler, affine invariant, multivariate,

distribution-free sign test. *Journal of the American Statistical Association*, **95**, 1263–1268. MR1792189

[18] Srivastava, M.S. (2009). A test of the mean vector with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis*, **100**, 518–532. MR2483435

[19] Srivastava, M. S. and Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis*, **99**, 386–402. MR2396970

[20] Wang, L., Peng, B. and Li, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *Journal of the American Statistical Association*, **110**, 1658–1669. MR3449062

[21] Zou, C., Peng, L, Feng, L. and Wang, Z. (2014). Multivariate-sign-based high-dimensional tests for sphericity. *Biometrika*, **101**, 229–236. MR3180668