Multivariate-sign-based high-dimensional tests for the two-sample location problem

Long Feng, Changliang Zou and Zhaojun Wang

LPMC and Institute of Statistics, Nankai University, Tianjin, China

Abstract

This article concerns tests for the two-sample location problem when data dimension is larger than the sample size. Existing multivariate-sign-based procedures are not robust against high dimensionality, producing tests with type I error rates far away from nominal levels. This is mainly due to the biases from estimating location parameters. We propose a novel test to overcome this issue by using the "leave-one-out" idea. The proposed test statistic is scalar-invariant and thus is particularly useful when different components have different scales in high-dimensional data. Asymptotic properties of the test statistic are studied. Compared with other existing approaches, simulation studies show that the proposed method behaves well in terms of sizes and power.

Keywords: Asymptotic normality; High-dimensional data; Large p, small n; Spatial median; Spatial-sign test; Scalar-invariance.

1 Introduction

Assume that $\{\mathbf{X}_{i1}, \cdots, \mathbf{X}_{in_i}\}$ for $i = 1, 2$ are two independent random samples with sample sizes n_1 and n_2, from p-variate distributions $F(\mathbf{x} - \boldsymbol{\theta}_1)$ and $G(\mathbf{x} - \boldsymbol{\theta}_2)$ located at p-variate centers $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. We wish to test

$$H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 \quad \text{versus} \quad H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2. \tag{1}$$

Such a hypothesis test plays an important role in a number of statistical problems. A classical method to deal with this problem is the famous Hotelling's T^2 test statistic $T_n^2 = \frac{n_1 n_2}{n}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)^T \mathbf{S}_n^{-1}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)$, where $n = n_1 + n_2$, $\bar{\mathbf{X}}_1$ and $\bar{\mathbf{X}}_2$ are the two sample means and $\mathbf{S}_n$ is the pooled sample covariance. However, T_n^2 is undefined when the dimension of data is greater than the within sample degrees of freedom, say a so-called "large p, small n" case which is largely motivated by the identification of significant genes in microarray and genetic sequence studies (e.g., see Kosorok and Ma 2007 and the references therein).

The "large p, small n" situation refers to high-dimensional data whose dimension p increases to infinity as the number of observations $n \to \infty$. With the rapid development of technology, various types of high-dimensional data have been generated in many areas, such as hyperspectral imagery, internet portals, microarray analysis and DNA. Like the Hotelling's T^2 test mentioned above, traditional methods may not work any more in this situation since they assume that p keeps unchanged as n increases. This challenge calls for new statistical tests to deal with high-dimensional data, see Dempster (1958), Bai and Saranadasa (1996), Srivastava (2009) and Chen and Qin (2010) for two-sample tests for means, Ledoit and Wolf (2002), Schott (2005), Srivastava (2005), Chen et al. (2010) and Zou et al. (2014) for testing a specific covariance structure, Goeman et al. (2006), Zhong et al. (2011) and Feng et al. (2013) for high-dimensional regression coefficients.

Under the equality of two covariance matrices, say $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}$, Bai and Saranadasa (1996) proposed a test statistic based on the squared Euclidean distance, $||\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2||^2$. The key feature of Bai and Saranadasa's proposal is to use the Euclidian norm to replace the Mahalanobis norm since having $\mathbf{S}_n^{-1}$ is no longer beneficial when $p/n \to c > 0$. Chen and Qin (2010) considered removing $\sum_{j=1}^{n_i} \mathbf{X}_{ij}^T \mathbf{X}_{ij}$ for $i = 1, 2$ from $||\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2||^2$ because these terms impose demands on the dimensionality. However, both these two methods are not scalar-invariant. In practice, different components may have completely different physical or biological readings and thus certainly their scales would not be identical. For example, in Nettleton et al. (2008), the authors pointed out that "it is well known that different genes exhibit different levels of variation. If this heterogeneity of variance is not accounted for, genes with larger variability can dominate the results of the proposed test." As a result, prior to analysis the authors applied scaling, which is equivalent to standardizing that data for each gene to a common variance.

Note that both of Bai and Saranadasa's (1996) and Chen and Qin's (2010) tests take the sum of all p squared mean differences without using the information from the diagonal elements of the sample covariance, i.e., the variances of variables, and thus their test power would heavily depend on the underlying variance magnitudes. To address this issue, Srivastava and Du (2008) proposed a scalar-transformation-invariant test under the assumption of the equality of the two covariance matrices. Srivastava et al. (2013) further extended this approach to unequal covariance matrices and revealed that their test has certain advantages over Chen and Qin's (2010) test by asymptotic and numerical analysis. Park and Ayyala (2013) considered a similar setting with the idea of leave-one-out cross validation as employed in Chen and Qin (2010). Recently, Feng et al. (2014) and Gregory et al. (2014) also considered cases where heteroscedasticity is present. Both two papers addressed the issues of non-negligible bias-terms due to the plug-in of variance estimators, while the latter one considered a higher-order expansion for bias-correction.

Although essentially nonparametric in spirit, the statistical performance of the moment-based tests mentioned above would be degraded when the non-normality is severe, especially for heavy-tailed distributions. This motivates us to consider using multivariate sign-and/or-rank-based approaches to construct robust tests for (1). Many nonparametric methods have been developed, as a reaction to the Gaussian approach of Hotelling's test, with the objective of extending to the multivariate context the classical univariate rank and signed-rank techniques. Essentially, these methods belong to three main groups. The first of these groups relies on componentwise rankings (see, e.g., the monograph by Puri and Sen 1971), but suffers from a severe lack of invariance with respect to affine transformations, which has been the main motivation for the other two approaches. The second group uses the concept of interdirections (Peters and Randles 1990; Randles 1992), while the third one (Hettmansperger and Oja 1994; Möttönen and Oja 1995; Randles 2000) is closely related with the spatial signs and ranks along with the use of the so-called spatial median. See Oja and Randles (2004) and Oja (2010). This work belongs to the third group with emphasis on the applicability and effectiveness in high-dimensional environment.

Most of the tests proposed in the works mentioned above are based on spatial-signs and ranks of the norms of observations centered at $\boldsymbol{\theta}$ (an estimate in practice), with test statistics that have structures similar to T_n^2. Those statistics are distribution-free under mild assumptions, or asymptotically so. Please refer to Chapter 11 of Oja (2010) for a nice overview. Among them, due to its simplicity and effectiveness, the test entirely based on spatial-signs is of particular interest and has been detailedly discussed. In this paper, we focus on this type of test and show that some "off-the-shelf" modifications for high-dimensional data are not robust against high dimensionality in the sense that they would produce tests with type I errors far away from nominal levels. This is mainly due to additional biases yielded by using the estimation of location parameter to replace the true one. In the next section, we develop a novel remedy that is robust against high dimensionality. We show that the proposed test statistic is asymptotically normal under elliptical distributions. Simulation comparisons show that our procedure performs reasonably well in terms of sizes and power for a wide range of dimensions, sample sizes and distributions. Recently, Wang et al. (2014) proposed a high-dimensional nonparametric multivariate test for mean vector based on spatial-signs. However, their focus is on one-sample problem and thus their method is significantly different from our proposal as we will explain in a later section.

2 Multivariate-sign-based high-dimensional tests

2.1 The proposed test statistic

We develop the test for (1) under the elliptically symmetric assumption which is commonly adopted in the literature of multivariate-sign-based approaches (Oja 2010). Assume $\{\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in_i}\}$, $i = 1, 2$ be two independently and identically distributed (i.i.d.) random samples from p-variate elliptical distribution with density functions $\det(\mathbf{\Sigma}_i)^{-1/2} g_i(||\mathbf{\Sigma}_i^{-1/2}(\mathbf{x} - \boldsymbol{\theta}_i)||)$, $i = 1, 2$, where $\boldsymbol{\theta}_i$'s are the symmetry centers and $\mathbf{\Sigma}_i$'s are two positive definite symmetric $p \times p$ scatter matrices. The spatial sign function is defined as $U(\mathbf{x}) = ||\mathbf{x}||^{-1}\mathbf{x}I(\mathbf{x} \neq \mathbf{0})$. Denote $\boldsymbol{\varepsilon}_{ij} = \mathbf{\Sigma}_i^{-1/2}(\mathbf{X}_{ij} - \boldsymbol{\theta}_i)$. The modulus $||\boldsymbol{\varepsilon}_{ij}||$ and the direction $\mathbf{u}_{ij} = U(\boldsymbol{\varepsilon}_{ij})$ are independent, and the direction vector $\mathbf{u}_{ij}$ is uniformly distributed on the p-dimensional unit sphere. It is then well known that $E(\mathbf{u}_{ij}) = 0$ and $\text{cov}(\mathbf{u}_{ij}) = p^{-1}\mathbf{I}_p$.

In traditional fixed p circumstance, the following so-called "inner centering and inner standardization" sign-based procedure is usually used (cf., Section 11.3 of Oja 2010)

$$Q_n^2 = p \sum_{i=1}^{2} n_i \hat{\mathbf{U}}_i^T \hat{\mathbf{U}}_i, \tag{2}$$

where $\hat{\mathbf{U}}_i = n_i^{-1} \sum_{j=1}^{n_i} \hat{\mathbf{U}}_{ij}$, $\hat{\mathbf{U}}_{ij} = U(\mathbf{S}^{-1/2}(\mathbf{X}_{ij} - \hat{\boldsymbol{\theta}}))$, $\hat{\boldsymbol{\theta}}$ and $\mathbf{S}^{-1/2}$ are Hettmansperger and Randles's (2002) estimates (HRE) of location and scatter matrix so that

$$\sum_{i=1}^{2} \sum_{j=1}^{n_i} \hat{\mathbf{U}}_{ij} = \mathbf{0} \quad \text{and} \quad pn^{-1} \sum_{i=1}^{2} \sum_{j=1}^{n_i} \hat{\mathbf{U}}_{ij} \hat{\mathbf{U}}_{ij}^T = \mathbf{I}_p.$$

Q_n^2 is affine-invariant and can be regarded as a nonparametric counterpart of T_n^2 by using the spatial-signs instead of the original observations $\mathbf{X}_{ij}$'s. However, when $p > n$, Q_n^2 is not defined as the matrix $\mathbf{S}^{-1/2}$ is is not available in high-dimensional settings. Though there has been a growing body of research in large-scale covariance matrix estimation under certain assumptions of sparseness (e.g., Bickel and Levina 2008; Cai and Liu 2011), obtaining a sparse estimator of scatter matrix in the robust context appears to be even more complicated and has not been thoroughly addressed in the literature. More importantly, the effectiveness of Mahalanobis distance-based tests is adversely impacted by an increased dimension even $p < n$, reflecting a reduced degree of freedom in estimation when the dimensionality is close to the sample size. The contamination bias, which grows rapidly with p, would make the Mahalanobis distance-based tests unreliable for a large p. Bai and Saranadasa (1996) provided certain asymptotic justifications and Goeman et al. (2006) contained some numerical evidence. Please refer to numerical comparison in Section 3 and some technical discussion in Appendix F of the Supplemental Material.

Alternatively, we develop a scalar-transformation-invariant test on the line of Srivastava et al. (2013), which is able to integrate all the individual information in a relatively "fair" way. To this end, we suggest to find a pair of diagonal matrix $\mathbf{D}_i$ and vector $\boldsymbol{\theta}_i$ for each sample that simultaneously satisfy

$$\frac{1}{n_i} \sum_{j=1}^{n_i} U(\boldsymbol{\epsilon}_{ij}) = 0 \quad \text{and} \quad \frac{p}{n_i} \text{diag} \left\{ \sum_{j=1}^{n_i} U(\boldsymbol{\epsilon}_{ij}) U(\boldsymbol{\epsilon}_{ij})^T \right\} = \mathbf{I}_p, \tag{3}$$

where $\boldsymbol{\epsilon}_{ij} = \mathbf{D}_i^{-1/2}(\mathbf{X}_{ij} - \boldsymbol{\theta}_i)$. $(\mathbf{D}_i, \boldsymbol{\theta}_i)$ can be viewed as a simplified version of HRE without considering the off-diagonal elements of $\mathbf{S}$. We can adapt the recursive algorithm of

Hettmansperger and Randles (2002) to solve (3). That is, repeat the following three steps until convergence:

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

(ii) $\boldsymbol{\theta}_i \leftarrow \boldsymbol{\theta}_i + \frac{\mathbf{D}_i^{1/2} \sum_{j=1}^{n_i} U(\boldsymbol{\epsilon}_{ij})}{\sum_{j=1}^{n} ||\boldsymbol{\epsilon}_{ij}||^{-1}};$

(iii) $\mathbf{D}_i \leftarrow p\mathbf{D}_i^{1/2} \mathrm{diag}\{n_i^{-1} \sum_{j=1}^{n_i} U(\boldsymbol{\epsilon}_{ij}) U(\boldsymbol{\epsilon}_{ij})^T\} \mathbf{D}_i^{1/2}.$

The resulting estimators of location and diagonal matrix are denoted as $\hat{\boldsymbol{\theta}}_i$ and $\hat{\mathbf{D}}_i$, $i = 1, 2$, respectively. We may use the sample mean and sample variances as the initial estimators.

It appears that we can construct Q_n^2 with a pooled sample estimate, $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{D}})$, obtained by using (3). However, this would yield a bias-term which is not negligible (with respect to the standard deviation) when $n/p = O(1)$ because we replace the true spatial median with its estimate. It seems infeasible to develop a bias-correction procedure as done in Zou et al. (2014) because the bias term depends on the unknown quantities $\boldsymbol{\Sigma}_i$'s. Please refer to Appendix C in the Supplemental Material for the closed-form of this bias and associated asymptotic analysis. In fact, the test statistic proposed by Wang et al. (2014) is essentially in a similar fashion to Q_n^2. However, their method does not suffer from additional biases because in a one-sample problem we do not need the estimate of spatial median.

To overcome this difficulty, we propose the following test statistic

$$R_n = -\frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} U^T(\hat{\mathbf{D}}_{1,i}^{-1/2}(\mathbf{X}_{1i} - \hat{\boldsymbol{\theta}}_{2,j})) U(\hat{\mathbf{D}}_{2,j}^{-1/2}(\mathbf{X}_{2j} - \hat{\boldsymbol{\theta}}_{1,i})),$$

where $\hat{\boldsymbol{\theta}}_{i,j}$ and $\hat{\mathbf{D}}_{i,j}$ are the corresponding location vectors and scatter matrices using "leave-one-out" samples $\{\mathbf{X}_{ik}\}_{k \neq j}$. Intuitively, if $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, both $U(\hat{\mathbf{D}}_{1,i}^{-1/2}(\mathbf{X}_{1i} - \hat{\boldsymbol{\theta}}_{2,j}))$ and $U(\hat{\mathbf{D}}_{2,j}^{-1/2}(\mathbf{X}_{2j} - \hat{\boldsymbol{\theta}}_{1,i}))$ would deviate from $\mathbf{0}$ to certain degree and thus a large value of R_n leads to reject the null hypothesis. As shown later, $E(R_n) \propto (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)^T \mathbf{D}_1^{-1/2} \mathbf{D}_2^{-1/2} (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)$ (approximately speaking), and hence R_n can well reflect the difference between two locations and is basically all we need for testing. A derivation in Appendix A shows that under H_0 the expectation of R_n is asymptotically negligible compared to its standard deviation. This feature is particular useful in the construction of the test because we do not need to estimate its expectation. The following proposition shows that the proposed test statistic R_n is invariant under location shifts and the group of scalar transformations.

Proposition 1 *Define* $\tilde{\mathbf{X}}_{ij} = \mathbf{D}^{1/2}\mathbf{X}_{ij} + \boldsymbol{c}$, *where* $\boldsymbol{c}$ *is a constant vector,* $\mathbf{D} = \mathrm{diag}\{d_1^2, \cdots, d_p^2\}$, *and* $d_1, \cdots, d_p$ *are non-zero constants. Denote the corresponding test statistic with* $\tilde{\mathbf{X}}_{ij}$ *as* $\tilde{R}_n$. *Then, we have* $\tilde{R}_n = R_n$.

2.2 Asymptotic results

Next, we study the asymptotic behavior of R_n under the null and local alternative hypotheses. Let $\mathbf{R}_i = \mathbf{D}_i^{-1/2}\boldsymbol{\Sigma}_i\mathbf{D}_i^{-1/2}$, $c_i = E(||\mathbf{D}_i^{-1/2}(\mathbf{X}_{ij} - \boldsymbol{\theta}_i)||^{-1})$,

$$\mathbf{A}_1 = c_2c_1^{-1}\boldsymbol{\Sigma}_1^{1/2}\mathbf{D}_2^{-1/2}\mathbf{D}_1^{-1/2}\boldsymbol{\Sigma}_1^{1/2},$$

$$\mathbf{A}_2 = c_1c_2^{-1}\boldsymbol{\Sigma}_2^{1/2}\mathbf{D}_1^{-1/2}\mathbf{D}_2^{-1/2}\boldsymbol{\Sigma}_2^{1/2},$$

$$\mathbf{A}_3 = \boldsymbol{\Sigma}_1^{1/2}\mathbf{D}_1^{-1/2}\mathbf{D}_2^{-1/2}\boldsymbol{\Sigma}_2^{1/2},$$

and $\sigma_n^2 = \left(\frac{2}{n_1(n_1-1)p^2}\mathrm{tr}(\mathbf{A}_1^2) + \frac{2}{n_2(n_2-1)p^2}\mathrm{tr}(\mathbf{A}_2^2) + \frac{4}{n_1n_2p^2}\mathrm{tr}(\mathbf{A}_3^T\mathbf{A}_3)\right)$. From the proof of Theorem 1, we can see that $\mathrm{var}(R_n) = \sigma_n^2(1 + o(1))$. We need the following conditions for asymptotic analysis:

(C1) $n_1/(n_1 + n_2) \to \kappa \in (0, 1)$ as $n \to \infty$;

(C2) $\mathrm{tr}(\mathbf{A}_i^T\mathbf{A}_j\mathbf{A}_l^T\mathbf{A}_k) = o(\mathrm{tr}^2\{(\mathbf{A}_1 + \mathbf{A}_2 + \mathbf{A}_3)^T(\mathbf{A}_1 + \mathbf{A}_2 + \mathbf{A}_3)\})$ for $i, j, l, k = 1, 2, 3$;

(C3) $n^{-2}/\sigma_n = O(1)$ and $\log p = o(n)$;

(C4) $(\mathrm{tr}(\mathbf{R}_i^2) - p) = o(n^{-1}p^2)$.

Remark 1 To appreciate Condition (C2), consider the simple case $\mathbf{D}_1 = \mathbf{D}_2 = \mathbf{I}_p$, $c_1 = c_2$, thus the condition becomes $\mathrm{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_k\boldsymbol{\Sigma}_l) = o(\mathrm{tr}^2((\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^2))$, $i, j, k, l = 1, 2$, which is the same as condition (3.8) in Chen and Qin (2010). To better understand Condition (C3), consider $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ with bounded eigenvalues which leads to $\sigma_n^2 = O(n^{-2}p)$. Thus, the condition becomes $p = O(n^2)$, which allows dimensionality to increase as the square of sample size. Certainly, when p is larger than n^2, there would be another bias-term in R_n which is difficult to calculate and deserves a future research.

The Condition (C4) is used to get the consistency of the diagonal matrix estimators. If $\mathbf{R}_i = \mathbf{I}_p$, $i = 1, 2$, the module and the direction of $\mathbf{D}_i^{-1/2}(\mathbf{X}_{ij} - \boldsymbol{\theta}_i)$ are independent and accordingly it is easy to obtain the consistence of the diagonal matrix. However, the module and the direction of $\mathbf{D}_i^{-1/2}(\mathbf{X}_{ij}-\boldsymbol{\theta}_i)$ are not independent in general case. Consider a simple setting, $\mathrm{tr}(\mathbf{R}_i^2)=O(p)$ (Srivastava and Du 2008). This condition reduces to $p/n\to\infty$. When the correlation between each components becomes larger, the dimension is required to be higher to reduce the difference between the module $||\boldsymbol{\varepsilon}_{ij}||$ and $\boldsymbol{\varepsilon}_{ij}^T\mathbf{R}_i\boldsymbol{\varepsilon}_{ij}$. See more information in the proof of Theorem 1 in Appendix A. □

The following theorem establishes the asymptotic null distribution of R_n.

Theorem 1 *Under Conditions (C1)-(C4) and H_0, as $(p,n)\to\infty$, $R_n/\sigma_n \xrightarrow{d} N(0,1)$.*

We propose the following estimators to estimate the trace terms in σ_n^2

$$\widehat{\mathrm{tr}(\mathbf{A}_1^2)} = \frac{p^2\hat{c}_2^2\hat{c}_1^{-2}}{n_1(n_1-1)}\sum_{k=1}^{n_1}\sum_{l\neq k}^{n_1}\left(\tilde{\mathbf{U}}_{1l}^T\hat{\mathbf{D}}_2^{-1/2}\hat{\mathbf{D}}_1^{1/2}\tilde{\mathbf{U}}_{1k}\right)^2,$$
$$\widehat{\mathrm{tr}(\mathbf{A}_2^2)} = \frac{p^2\hat{c}_1^2\hat{c}_2^{-2}}{n_2(n_2-1)}\sum_{k=1}^{n_1}\sum_{l\neq k}^{n_2}\left(\tilde{\mathbf{U}}_{2l}^T\hat{\mathbf{D}}_1^{-1/2}\hat{\mathbf{D}}_2^{1/2}\tilde{\mathbf{U}}_{2k}\right)^2,$$
$$\mathrm{tr}\widehat{(\mathbf{A}_3^T\mathbf{A}_3)} = \frac{p^2}{n_1n_2}\sum_{l=1}^{n_1}\sum_{k=1}^{n_2}\left(\tilde{\mathbf{U}}_{1l}^T\tilde{\mathbf{U}}_{2k}\right)^2,$$

where $\tilde{\mathbf{U}}_{ij}=U(\hat{\mathbf{D}}_{i,j}^{-1/2}(\mathbf{X}_{ij}-\hat{\boldsymbol{\theta}}_{i,j}))$ and $\hat{c}_i=n_i^{-1}\sum_{j=1}^{n_i}||\hat{\mathbf{D}}_{i,j}^{-1/2}(\mathbf{X}_{ij}-\hat{\boldsymbol{\theta}}_{i,j})||$.

Proposition 2 *Suppose Conditions (C1)-(C4) hold. Then, we have*

$$\frac{\mathrm{tr}\widehat{(\mathbf{A}_i^T\mathbf{A}_i)}}{\mathrm{tr}(\mathbf{A}_i^T\mathbf{A}_i)} \xrightarrow{p} 1, i=1,2,3, \quad as \quad (p,n)\to\infty.$$

As a consequence, a ratio-consistent estimator of σ_n^2 under H_0 is

$$\hat{\sigma}_n^2=\frac{2}{n_1(n_1-1)p^2}\widehat{\mathrm{tr}(\mathbf{A}_1^2)}+\frac{2}{n_2(n_2-1)p^2}\widehat{\mathrm{tr}(\mathbf{A}_2^2)}+\frac{4}{n_1n_2p^2}\mathrm{tr}\widehat{(\mathbf{A}_3^T\mathbf{A}_3)}.$$

This result suggests rejecting H_0 with α level of significance if $R_n/\hat{\sigma}_n>z_\alpha$, 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

(C5) $(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)^T\mathbf{D}_1^{-1/2}\mathbf{D}_2^{-1/2}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)=O(c_1^{-1}c_2^{-1}\sigma_n)$.

This assumption implies that the difference between $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ is not large relative to $c_1^{-1}c_2^{-1}\sigma_n$ so that a workable expression for the variance of R_n can be derived and thus leads to an explicit power expression for the proposed test. It can be viewed as a high-dimensional version of the local alternative hypotheses.

Theorem 2 *Under Conditions (C1)-(C5), as* $(p,n)\to\infty$, $(R_n-\delta_n)/\tilde{\sigma}_n \xrightarrow{d} N(0,1)$, *where* $\delta_n = c_1c_2(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)^T\mathbf{D}_1^{-1/2}\mathbf{D}_2^{-1/2}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)$ *and*

$$\tilde{\sigma}_n^2 = \sigma_n^2 + \frac{c_2^2}{n_1p}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)^T\mathbf{D}_2^{-1/2}\mathbf{R}_1\mathbf{D}_2^{-1/2}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2) + \frac{c_1^2}{n_2p}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)^T\mathbf{D}_1^{-1/2}\mathbf{R}_2\mathbf{D}_1^{-1/2}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2).$$

Theorems 1 and 2 allow us to compare the proposed test with some existing work, such as Chen and Qin (2010) and Srivastava et al. (2013), in terms of the limiting efficiency. We consider the following local alternatives

$$H_1 : c_1c_2(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)^T\mathbf{D}_1^{-1/2}\mathbf{D}_2^{-1/2}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2) = \varpi\sigma_n$$

with some constant $\varpi > 0$. Accordingly, the asymptotic power of our proposed spatial-sign-based test (abbreviated SS) under this local alternative is

$$\beta_{\mathrm{SS}}(||\boldsymbol{\theta}_1-\boldsymbol{\theta}_2||) = \Phi(-z_\alpha\sigma_n/\tilde{\sigma}_n + \delta_n/\tilde{\sigma}_n),$$

where Φ is the standard normal distribution function. In order to obtain an explicit expression for comparison use, we assume that $F = G$ and $\lambda_{\max}(p^{-1}\mathbf{R}_i) = o(n^{-1})$. Then, $\mathbf{D}_1 = \mathbf{D}_2 = \mathbf{D}$, $\mathbf{R}_1 = \mathbf{R}_2 = \mathbf{R}$ and $c_1 = c_2 = c_0$. As a consequence, $\tilde{\sigma}_n = \sigma_n(1+o(1))$ and the asymptotic power becomes

$$\beta_{\mathrm{SS}}(||\boldsymbol{\theta}_1-\boldsymbol{\theta}_2||) = \Phi\left(-z_\alpha + \frac{c_0^2pn\kappa(1-\kappa)(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)^T\mathbf{D}^{-1}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)}{\sqrt{2\mathrm{tr}(\mathbf{R}^2)}}\right).$$

In comparison, by Srivastava et al. (2013) we can show that the asymptotic power of their proposed test (abbreviated as SKK hereafter) is

$$\beta_{\mathrm{SKK}}(||\boldsymbol{\theta}_1-\boldsymbol{\theta}_2||) = \Phi\left(-z_\alpha + \frac{np\kappa(1-\kappa)(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)^T\mathbf{D}^{-1}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)}{E(||\boldsymbol{\varepsilon}_{ij}||^2)\sqrt{2\mathrm{tr}(\mathbf{R}^2)}}\right).$$

Thus, the asymptotic relative efficiency (ARE) of SS with respect to SKK is

$$\mathrm{ARE}(\mathrm{SS},\mathrm{SKK}) = c_0^2E(||\boldsymbol{\varepsilon}_{ij}||^2) \approx \{E(||\boldsymbol{\varepsilon}_{ij}||^{-1})\}^2E(||\boldsymbol{\varepsilon}_{ij}||^2),$$

where we use the fact that $c_0 = E(||\boldsymbol{\varepsilon}_{ij}||^{-1})(1 + o(1))$ under Condition (C4) (see the proof of Theorem 1).

It can be also shown that for multivariate normal distributions ARE(SS, SKK) $\to 1$ as $p \to \infty$ (see Appendix E in the Supplemental Material). Hence, the SS test is asymptotically as efficient as Srivastava et al.'s (2013) test in such settings. When the dimension p is not very large, it can be expected that the proposed test, using only the direction of an observation from the origin, should be outperformed by the test constructed with original observations like that of Srivastava et al. (2013). However, as $p \to \infty$, the disadvantage diminishes.

On the other hand, if $\mathbf{X}_{ij}$'s are generated from the multivariate t-distribution with ν degrees of freedom ($\nu > 2$),

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

The ARE values for $\nu = 3, 4, 5, 6$ are 2.54, 1.76, 1.51, and 1.38, respectively. Clearly, the SS test is more powerful than SKK when the distributions are heavy-tailed (ν is small), which is verified by simulation studies in Section 3.

In contrast, Chen and Qin (2010) showed that the power of their proposed test (abbreviated as CQ) is

$$\beta_{\mathrm{CQ}}(||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2||) = \Phi\left(-z_\alpha + \frac{np\kappa(1-\kappa)||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2||^2}{E(||\boldsymbol{\varepsilon}_{ij}||^2)\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma}^2)}}\right).$$

As shown by Srivastava et al. (2013), this quantity, β_{CQ}, can be much smaller than β_{SKK} in the cases that different components have different scales because the CQ test is not scalar-invariant. To appreciate the effect of scalar-invariance, we consider the multivariate normality assumption and $\boldsymbol{\Sigma}_i$ be diagonal matrices. The variances of the first $p/2$ components are τ_1^2 and the variances of the other components are τ_2^2. Assume $\theta_{1k} - \theta_{2k} = \zeta$, $k = 1, \cdots, \lfloor p/2 \rfloor$, where θ_{ik} denotes the k the component of $\boldsymbol{\theta}_i, i = 1, 2$. In this setting,

$$\beta_{\mathrm{SS}}(||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2||) = \beta_{\mathrm{SKK}}(||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2||) = \Phi\left(-z_\alpha + \frac{n\sqrt{p}\kappa(1-\kappa)\zeta^2}{2\sqrt{2}\tau_1^2}\right),$$

$$\beta_{\mathrm{CQ}}(||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2||) = \Phi\left(-z_\alpha + \frac{n\sqrt{p}\kappa(1-\kappa)\zeta^2}{2\sqrt{\tau_1^4 + \tau_2^4}}\right).$$

Thus, the ARE of the proposed test (so as SKK) with respect to the CQ test is $\sqrt{\tau_1^4 + \tau_2^4}/(\sqrt{2}\tau_1^2)$. It is clear that the SS and SKK are more powerful than CQ if $\tau_1^2 < \tau_2^2$ and vice versa. This

ARE has a positive lower bound of $1/\sqrt{2}$ when $\tau_1^2 >> \tau_2^2$, and it can be arbitrarily large if τ_1^2/τ_2^2 is close to zero. This property shows the necessity of a test with the scale-invariance property.

2.3 Bootstrap procedure and computational issue

The proposed SS test is based on the asymptotic normality; a good approximation requires that both n and p are large. As shown in Section 3, when p is not large enough (as in Table 1), the empirical size is a little larger than the nominal one. In contrast, when p is too large compared to n, our proposed test is somewhat conservative. The reason is that Condition (C3) basically prevents us from using the asymptotic normality for the case that p grows in a too fast rate of n. To address this issue, we propose the application of a bootstrap procedure for finite-sample cases.

The procedure is implemented as follows. We firstly calculate the based samples $\hat{\mathbf{X}}_{ij} = \mathbf{X}_{ij} - \hat{\boldsymbol{\theta}}_{i,j}$, $i = 1, 2$. Then two bootstrap samples $\{\mathbf{X}^*_{ij}\}_{j=1}^{n_i}$ are drawn from $\{\hat{\mathbf{X}}_{ij}\}_{j=1}^{n_i}$, $i = 1, 2$, respectively. A bootstrap test statistic R_n^* is accordingly built from the bootstrap sample $\left(\{\mathbf{X}^*_{1j}\}_{j=1}^{n_1}, \{\mathbf{X}^*_{2j}\}_{j=1}^{n_2}\right)$. When this procedure is repeated many times, the bootstrap critical value z_α^* is the empirical $1-\alpha$ quantile of the bootstrap test statistic. The test with rejection region $R_n \geq z_\alpha^*$ is our proposal. We recommend to use this bootstrap method when either p is not large (say, $p \leq 50$) or p is very large compared to n (in a rate of $O(n^2)$ or faster). We will study the effectiveness of this bootstrap method by simulation in the next section.

The leave-one-out procedure seems complex but basically computes fast. Today's computing power has improved dramatically and it is computationally feasible to implement the SS test. For example, it takes 1.5s to calculate $R_n/\tilde{\sigma}_n$ in FORTRAN using Inter Core 2.2 MHz CPU with $n_1 = n_2 = 50$ and $p = 1440$. The leave-one-out estimator essentially requires $O(pn)$ computation and thus the calculation of R_n is of order $O(pn^3)$. Also note that computing the estimates of the trace terms needs $O(pn^2)$ computation and thus the complexity of the entire procedure is $O(pn^3)$. In contrast, the SKK test requires $O(p^2n)$ computation. When p is large but n is small, such as $n_1 = n_2 = 50, p = 1440$, our method is even faster than Srivastava et al.'s (2013) test. The FORTRAN code for implementing the procedure is available in the Supplemental Material.

3 Numerical studies

3.1 Monte Carlo simulations

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. $\mathbf{X}_{ij} \sim N(\boldsymbol{\theta}_i, \mathbf{R}_i)$.

(II) Multivariate normal distribution with different component variances. $\mathbf{X}_{ij} \sim N(\boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i)$, where $\boldsymbol{\Sigma}_i = \mathbf{D}_i^{1/2}\mathbf{R}_i\mathbf{D}_i^{1/2}$ and $\mathbf{D}_i = \text{diag}\{d_{i1}^2, \cdots, d_{ip}^2\}$, $d_{ij}^2 = 3$, $j \le p/2$ and $d_{ij}^2 = 1$, $j > p/2$.

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

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

(V) Multivariate mixture normal distribution $\text{MN}_{p,\gamma,9}$. $\mathbf{X}_{ij}$'s are generated from $\gamma f_p(\boldsymbol{\theta}_i, \mathbf{R}_i) + (1-\gamma) f_p(\boldsymbol{\theta}_i, 9\mathbf{R}_i)$, denoted by $\text{MN}_{p,\gamma,9}$, where $f_p(\cdot;\cdot)$ is the density function of p-variate multivariate normal distribution. γ is chosen to be 0.8.

First, we consider the low dimensional case $p < n$ and compare the SS test with the traditional spatial-sign-based test (abbreviated as TS). We choose $\mathbf{R}_1 = \mathbf{R}_2 = (\rho_{jk})$, $\rho_{jk} = 0.5^{|j-k|}$. Without loss of generality, under H_1, we fix $\boldsymbol{\theta}_1 = 0$ and choose $\boldsymbol{\theta}_2$ as follows. The percentage of $\theta_{1l} = \theta_{2l}$ for $l = 1, \cdots, p$ are chosen to be 95% and 50%, respectively. At each percentage level, two patterns of allocation are explored for the nonzero θ_{2l}: the equal allocation and linear allocation where all the nonzero θ_{2l} are linearly increasing allocations. To make the power comparable among the configurations of H_1, we set $\eta =: ||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2||^2/\sqrt{\text{tr}(\boldsymbol{\Sigma}_1^2)} = 0.1$ throughout the simulation. Two combinations of (n, p) are considered: $(n_i, p) = (50, 40)$ and $(n_i, p) = (75, 60)$. Tables 1 reports the empirical sizes and power comparison at a 5%

nominal significance level under various scenarios. From Table 1, we observe that the sizes of the SS test are generally close to the nominal level under all the scenarios, though in some cases the SS appears to be a little liberal. In contrast, the sizes of the TS test are much smaller than 5%, i.e., too conservative. Our SS test is clearly more powerful than the TS test in most cases. Such finding is consistent with that in Bai and Saranadasa (1996) that demonstrated classical Mahalanobis distance may not work well because the contamination bias in estimating the covariance matrix grows rapidly with p. When p and n are comparable in certain sense, having the inverse of the estimate of the scatter matrix in constructing tests would be no longer beneficial.

Table 1: Empirical size and power (%) comparison of the proposed SS test and the traditional spatial sign test (TS) at 5% significance under Scenarios (I)-(V) when $p < n$

	Size				Power								
						Equal Allocation				Linear Allocation			
(n_i, p)	$(50, 40)$		$(75, 60)$			$(50, 40)$		$(75, 60)$		$(50, 40)$		$(75, 60)$	
	SS	TS	SS	TS		SS	TS	SS	TS	SS	TS	SS	TS
Scenario					%								
(I)	6.5	0.8	4.8	1.7	50%	45	6.6	69	10	45	5.5	64	10
	–	–	–	–	95%	69	16	86	21	46	17	72	25
(II)	6.4	1.5	4.6	1.8	50%	93	18	99	35	92	18	100	33
	–	–	–	–	95%	100	55	100	72	100	60	100	78
(III)	6.3	1.1	4.8	1.1	50%	72	11	92	21	70	11	91	25
	–	–	–	–	95%	93	39	99	56	79	41	96	57
(IV)	4.5	2.6	5.5	1.4	50%	96	31	99	59	97	36	99	58
	–	–	–	–	95%	100	58	100	78	100	63	100	77
(V)	6.4	1.7	4.9	1.6	50%	78	14	95	28	77	12	94	28
	–	–	–	–	95%	96	43	100	63	83	46	98	67

Next, we consider the high-dimensional cases, $p > n$, and compare the SS with the tests proposed by Chen and Qin (2010) (CQ), Srivastava et al. (2013) (SKK) and Gregory et al.

(2014) (GCT). Due to the fact that the SKK procedure uses the estimate of $\text{tr}(\mathbf{R}^2)$ under normality assumption which has a considerable bias under non-normal distribution as shown by Srivastava et al. (2011), we replace it by the the estimator proposed by Srivastava et al. (2014). GCT is implemented using the function *GCT.test* in the R package *"highD2pop"* . We consider the cases with unequal correlation matrices: $\mathbf{R}_1 = (0.5^{|j-k|})$ and $\mathbf{R}_2 = \mathbf{I}_p$. The sample size n_i is chosen as 25, 50 and 100. Three dimensions for each sample size $p = 120, 480$ and 1440 are considered. Table 2 reports the empirical sizes at a 5% nominal significance level. The empirical sizes of SS, SKK and CQ tests are converging to the nominal level as both p and n increase together under all the scenarios. There is some slight size distortion (tends to be larger than 5%) for the CQ test when p is small or moderate. In contrast, our proposed SS test seems to be somewhat conservative when p/n is very large, such as $p = 480$ or 1440 and $n_i = 25$. This is because the ratio of $\hat{\sigma}_n^2/\sigma_n^2$ tends to be slightly larger than one in such cases. The empirical sizes of GCT tends to be a little smaller than the nominal level when p/n is large, especially under the non-normal cases.

To get a broader picture of goodness-of-fit of using the asymptotic normality for $R_n/\hat{\sigma}_n$, Figure 1 displays the normal Q-Q plots with various combinations of sample size and dimension. Here we only present the results of Scenarios (I), (III) and (V) since the results for the other scenarios are similar. There is a general convergence of our test statistic to $N(0,1)$ as n and p increase simultaneously.

For power comparison, we consider the same configurations of H_1 as before, except that $\eta =: ||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2||^2/\sqrt{\text{tr}(\boldsymbol{\Sigma}_1^2) + \text{tr}(\boldsymbol{\Sigma}_2^2)} = 0.1$. Three combinations of (n_i, p) are used: (25,120), (50,480) and (50,1440). The results of empirical power are given in Table 3. The results of comparison with Feng et al. (2014) are all reported in the Supplemental Material. Generally, under Scenarios (I) and (II), SKK has certain advantages over SS as we would expect, since the underlying distribution is multivariate normal. The SS also offers quite satisfactory performance though its sizes are often a little smaller than SKK as shown in Table 2. Under Scenario (I), the variances of components are identical and thus the superior efficiency of CQ is obvious. However, under Scenario (II), both SKK and SS outperform CQ uniformly by a quite large margin of power, which again concurs with the asymptotic comparison in Section 2.2.

Table 2: Empirical size comparison at 5% significance under Scenarios (I)-(V) when $p > n$

Scenario		p											
		120				480				1440			
	n_i	SS	SKK	CQ	GCT	SS	SKK	CQ	GCT	SS	SKK	CQ	GCT
(I)	25	4.0	9.1	5.6	3.6	3.0	6.8	5.7	4.3	1.8	1.4	4.6	2.0
	50	5.3	8.2	5.4	3.8	4.9	7.4	5.6	4.2	4.4	5.0	7.0	3.4
	100	4.8	6.2	7.0	4.2	5.1	5.3	5.4	4.2	4.8	6.7	3.8	3.5
(II)	25	4.2	9.1	5.5	2.6	3.0	6.8	5.8	4.3	1.8	1.5	6.8	5.7
	50	5.7	8.1	5.2	3.4	3.9	7.3	5.4	5.1	4.4	5.0	7.2	4.6
	100	4.8	6.2	6.5	4.1	4.9	5.1	6.4	4.4	3.8	6.7	4.4	4.1
	25	3.8	5.0	5.7	2.1	2.8	1.7	6.8	1.0	1.6	0.0	4.4	0.2
(III)	50	5.8	5.7	5.1	2.6	4.7	3.9	6.0	2.6	4.4	1.6	6.0	2.5
	100	4.0	4.8	6.1	3.0	4.6	4.9	6.3	3.4	4.2	3.6	4.6	3.9
	25	3.4	4.9	4.2	1.6	2.0	2.0	5.4	1.6	1.3	0.0	6.3	1.7
(IV)	50	4.8	5.8	5.2	2.5	4.8	5.0	6.6	3.5	3.2	1.4	6.0	4.0
	100	4.2	5.2	6.6	3.0	5.2	6.2	4.2	4.0	3.6	3.7	4.7	3.2
	25	3.2	3.7	6.0	1.3	2.7	1.4	4.5	1.7	1.1	0.0	5.4	1.6
(V)	50	5.6	7.0	7.1	3.3	4.9	4.4	6.2	2.6	4.2	1.0	5.9	2.5
	100	4.5	5.9	7.0	3.5	5.1	4.7	5.7	3.1	4.9	3.2	5.4	4.2

Under the other elliptical scenarios (III), (IV) and (V), the SS test is clearly more efficient than the CQ, SKK and GCT tests, and the difference is quite remarkable. Certainly, this is not surprising as neither $t_{p,4}$ nor $\mathrm{MN}_{p,\gamma,9}$ distribution belongs to the linear transformation model on which the validity of CQ depends much. Because the variance estimator in GCT usually leads to an overestimation under the alternative, especially for the sparse cases, the power of GCT is not as good as the other scalar-invariant tests. In addition, the power of the four tests is mainly dependent on $||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2||$ as analyzed in Section 2 and thus should be invariant (roughly speaking) for different patterns of allocation and different percentage levels of true null.

Table 3: Empirical power comparison at 5% significance under Scenarios (I)-(V) when $p > n$

Scenario	%	$p = 120$, $n_i = 25$				$p = 480$, $n_i = 50$				$p = 1440$, $n_i = 50$			
		SS	SKK	CQ	GCT	SS	SKK	CQ	GCT	SS	SKK	CQ	GCT
Equal Allocation													
(I)	50%	31	43	38	26	73	81	79	73	77	82	84	80
	95%	38	51	45	16	76	83	79	62	78	83	84	75
(II)	50%	74	85	38	66	100	100	80	100	99	100	85	100
	95%	83	91	44	29	100	100	84	98	100	100	86	100
(III)	50%	59	41	41	26	97	75	78	67	97	60	85	69
	95%	67	48	49	13	100	79	82	56	98	63	85	61
(IV)	50%	76	69	43	49	100	96	82	90	100	92	87	88
	95%	83	84	50	15	100	97	84	78	100	94	87	85
(V)	50%	65	39	40	23	99	76	76	73	99	63	84	75
	95%	74	47	47	12	100	80	81	61	100	67	85	68
Linear Allocation													
(I)	50%	32	42	39	28	72	81	77	72	77	82	84	79
	95%	33	43	38	12	73	80	77	52	76	82	84	71
(II)	50%	75	84	39	64	100	100	79	100	99	100	86	100
	95%	76	85	39	18	100	100	81	93	100	100	85	100
(III)	50%	57	40	40	23	97	75	79	69	98	61	86	67
	95%	60	42	44	9.1	99	75	80	49	98	62	85	62
(IV)	50%	82	71	45	41	100	96	81	89	98	92	87	91
	95%	93	62	44	10	100	95	83	63	100	95	84	82
(V)	50%	65	39	41	21	99	77	76	67	98	64	85	71
	95%	66	39	41	8.5	99	77	78	50	98	65	85	67

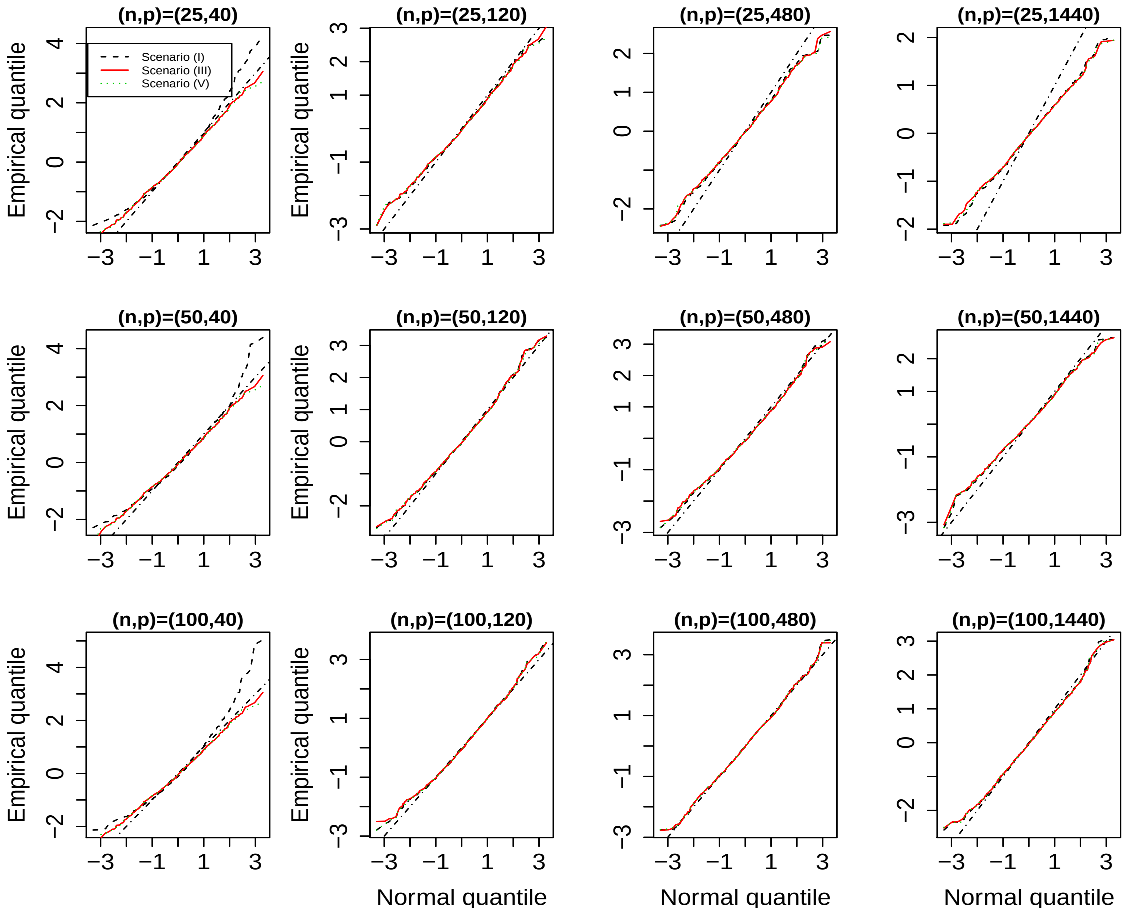

Figure 1: Normal Q-Q plots of our SS test statistics under Scenarios (I), (III) and (V).

Next, to study the effect of correlation matrix on the proposed test and to further discuss the application scope of our method, we explore another four scenarios with different correlations and distributions. The following moving average model is used:

$$X_{ijk} = ||\boldsymbol{\rho}_i||^{-1}(\rho_{i1}Z_{ij} + \rho_{i2}Z_{i(j+1)} + \cdots + \rho_{iT_i}Z_{i(j+T_i-1)}) + \theta_{ij}$$

for $i = 1, 2$, $j = 1, \cdots, n_i$ and $k = 1, \cdots, p$ where $\boldsymbol{\rho}_i = (\rho_{i1}, \ldots, \rho_{iT_i})^T$ and $\{Z_{ijk}\}$ are i.i.d. random variables. Consider four scenarios for the innovation $\{Z_{ijk}\}$:

(VI) All the $\{Z_{ijk}\}$'s are from $N(0, 1)$;

(VII) the first $p/2$ components of $\{Z_{ijk}\}_{k=1}^{p}$ are from centralized Gamma(8,1), and the others are from $N(0, 1)$.

(VIII) All the $\{Z_{ijk}\}$'s are from t_3;

(IX) All the $\{Z_{ijk}\}$'s are from $0.8N(0,1) + 0.2N(0,9)$.

The coefficients $\{\rho_{il}\}_{l=1}^{T_i}$ are generated independently from $U(2,3)$ and are kept fixed once generated through our simulations. The correlations among X_{ijk} and X_{ijl} are determined by $|k-l|$ and T_i. We consider the "full dependence" for the first sample and the "2-dependence" for the second sample, i.e. $T_1 = p$ and $T_2 = 3$, to generate different covariances of $\mathbf{X}_{ij}$. For simplicity, set $\eta =: ||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2||^2/\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_1^2) + \mathrm{tr}(\boldsymbol{\Sigma}_2^2)} = 0.05$ and $(n_i, p) = (50, 480)$.

Table 4 reports the empirical sizes and power of SS, SKK, CQ and GCT. Even under the last three non-elliptical scenarios, the SS test can maintain the empirical sizes reasonably well. Again, the empirical sizes of CQ tend to be larger than the nominal level. The empirical sizes of GCT deviate much from the nominal level, making its power unnecessarily high. In general, the SS test performs better than the CQ and SKK test in terms of power for the three non-normal distributions, especially under the 95% pattern. This may be explained as the proposed test, using only the direction of an observation from the origin but not its distance from the origin, would be more robust in certain degrees for the considered heavy-tailed distributions.

Next, we compare the SS test with a nonparametric method proposed by Biswas and Ghosh (2014) (abbreviated as BG). The samples are generated from Scenarios (I)-(V) with $\mathbf{R}_1 = (0.5^{|j-k|})$ and $\mathbf{R}_2 = \sigma^2\mathbf{R}_1$. Note that the null hypothesis of the BG test is the equality of two distributions rather than the equality of two location parameters. Thus, under H_0, we set $\sigma^2 = 1$. Under H_1, consider $\boldsymbol{\theta}_1 = \mathbf{0}$, $\boldsymbol{\theta}_2 = (\theta, \ldots, \theta)$. Table 5 reports the empirical size and power comparison with $(\theta, \sigma^2) = (2.5, 1)$ or (1,1.2) when $(n_i, p) = (50, 240)$. The empirical sizes of BG appears to be a little larger than the nominal level. The general conclusion is that our SS test significantly outperforms the BG test when the two location parameters are different. This is not surprising to us because the BG test is an omnibus one which is also effective for the difference in scale or shape parameters. When both location and scale parameters are different, the BG test performs better than our test under normality assumption (Scenarios (I) and (II)), while our SS test is more efficient under the other three non-normal scenarios.

We also study how our method is compared with some variants of the Hotelling's T^2 test. One variant is Chen et al.'s (2011) regularized Hotelling's T^2 test (abbreviated as RHT) which

Table 4: Empirical size and power comparison at 5% significance under Scenarios (VI)-(IX) with $p = 480$, $n_1 = n_2 = 50$

	Size				Power								
						Equal Allocation				Linear Allocation			
	SS	SKK	CQ	GCT		SS	SKK	CQ	GCT	SS	SKK	CQ	GCT
Scenario					%								
(VI)	6.2	4.7	5.0	24.8	50%	42	44	43	74	44	46	44	79
	–	–	–	–	95%	61	48	47	83	59	45	43	66
(VII)	6.1	7.2	7.8	23.4	50%	64	56	45	87	79	67	46	93
	–	–	–	–	95%	99	90	50	98	100	96	47	85
(VIII)	5.3	6.0	6.5	23.3	50%	42	45	45	77	44	47	45	80
	–	–	–	–	95%	58	53	49	83	55	49	46	66
(IX)	5.1	5.3	8.2	25.9	50%	44	44	45	77	45	44	44	78
	–	–	–	–	95%	61	51	49	82	58	48	46	67

uses $\mathbf{S} + \zeta \mathbf{I}_p$ instead of $\mathbf{S}$ with a sufficiently small value of ζ. Chen et al.'s (2011) suggested a resampling procedure to implement the RHT. This method is implemented using the *"RHT"* R package with the false alarm rate 0.05. Another natural variant is to use a sparse estimate of $\mathbf{\Sigma}$ to replace $\mathbf{S}$. Here we consider the banding estimator proposed by Bickel and Levina (2008) and denote the resulting test as the sparse Hotelling's T^2 test (abbreviated as SHT). It is easy to see that the normalized $\frac{n_1 n_2}{n}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)^T \mathbf{\Sigma}^{-1}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)$ is asymptotically normal as $(n, p) \to \infty$. As shown in Appendix F, this asymptotic normality does not hold well even when an optimal estimator of $\mathbf{\Sigma}$ is used in high dimensional settings. Thus, we also consider to use the bootstrap procedure for the SHT (denoted as SHTB). For simplicity, the correlation matrices are fixed as $\mathbf{R}_1 = \mathbf{R}_2 = (0.5^{|j-k|})_{1 \le j,k \le p}$. We choose the banding parameter as five, resulting in a nearly "oracle" estimator of $\mathbf{\Sigma}$. Table 6 reports the empirical sizes of SS, RHT, SHT and SHTB when $(n_i, p) = (50, 240)$. We observe that the RHT is rather conservative in this high-dimensional setting even with a bootstrap procedure, while the empirical sizes of SHT are much larger than the nominal level as we have explained before.

Table 5: Empirical size and power comparison with of SS and Biswas and Ghosh's (2014) test (BG) at 5% significance under Scenarios (I)-(V) when $n_1 = n_2 = 50, p = 240$.

Scenario	$(\mu, \sigma^2) = (0, 1)$		$(\mu, \sigma^2) = (2.5, 1)$		$(\mu, \sigma^2) = (1, 1.2)$	
	SS	BG	SS	BG	SS	BG
(I)	5.7	8.0	100	96	54	100
(II)	5.7	7.0	100	25	35	100
(III)	5.3	8.1	100	8.7	45	19
(IV)	5.3	9.0	99	9.0	22	19
(V)	5.5	7.2	100	8.4	40	13

Fortunately, the SHTB can well maintain the significant level. For power comparison, we consider two cases for $\boldsymbol{\theta}_2$. Half random: the last $p/2$ components of $\boldsymbol{\theta}_2$ are generated from $N(0, 1)$ and the others are zero; Sparse random: the last 5% components of $\boldsymbol{\theta}_2$ are generated from $N(0, 1)$ and the others are zeros. The power results with $\eta = ||\boldsymbol{\theta}_2||^2/\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_1^2)} = 0.1$ are also tabulated in Table 6. Our SS test has superior efficiency under all the scenarios, which further concurs with our previous claim that having the inverse of the estimate of the scatter matrix in high-dimensional settings may not be beneficial.

Finally, we evaluate the performance of the bootstrap procedure (denoted as SSB) suggested in Section 2.3. Table 7 reports the empirical sizes of the SSB. The data generation mechanism is the same as that in Table 2. Clearly, this bootstrap procedure is able to improve the SS test with asymptotic in terms of size-control in the sense that its empirical sizes are closer to the nominal level. In this table, the power values of SS and SSB with $(n_i, p) = (25, 480)$ are also presented, from which we can see that SSB performs better than SS in all cases because the latter is conservative in this setting of (n_i, p) .

We conducted some other simulations with various alternative hypotheses, p and nominal size, to check whether the above conclusions would change in other cases. These simulation results, not reported here but available from the authors, show that the SS test works well for other cases as well in terms of its sizes, and its good power performance still holds for other choices of alternatives.

Table 6: Empirical size and power comparison of SS, Chen et al.'s (2011) regularized Hotelling's T^2 test (RHT) and a variant of RHT (SHT) at 5% significance under Scenarios (I)-(V) when $n_1 = n_2 = 50, p = 240$. SHTB denotes the SHT using the bootstrap

	Size				Power					
					Half Random			Sparse Random		
Scenario	SS	RHT	SHT	SHTB	SS	RHT	SHTB	SS	RHT	SHTB
(I)	5.7	1.4	54.6	5.4	49	15	10	49	19	7.4
(II)	5.7	0.3	54.5	5.2	97	15	17	96	25	10
(III)	5.3	0.0	67.3	5.4	84	2.3	7.9	85	6.9	8.3
(IV)	5.3	0.0	61.0	6.2	96	4.1	10	93	9.2	8.9
(V)	5.5	0.0	75.9	6.1	92	1.0	10	91	5.3	9.3

3.2 A real-data example

Here we demonstrate the proposed methodology by applying it to a real dateset. The *colon data* (Alon, et al. 1999) contains the expression of the 2000 genes with highest minimal intensity across the 62 tissues. (http://microarray.princeton.edu/oncology/affydata/ index.html) There are 22 normal colon tissues and 40 tumor colon tissues. We want to test the hypothesis that the tumor group have the same gene expression levels as the normal group. Figure 2 shows the p-values of normality test and standard deviations of each variables of these two samples. From the above two figures of Figure 2, we observe that most variables of the tumor colon issues are not normal distributed. Thus, we could expect that SS test would be more robust than CQ and SKK tests. Furthermore, from the bottom panels of Figure 2, the standard deviations of each variables of these two samples are range from 15.9 to 4655, which illustrates that a scalar-invariant test is needed. Thus, we apply SS and SKK tests to this data sets. The p-values of our SS test with asymptotic normality and the bootstrap procedure are 0.0093 and 0.001, respectively. These results suggest the rejection of the null hypothesis; The gene expression levels of tumor group are significantly different from the normal group. However, the p-value of SKK test is about 0.18 for this dataset and thus cannot detect the difference between these two groups at 5% significance level. It is again

Table 7: Empirical size and power of the SS test using the bootstrap at 5% significance under Scenarios (I)-(V)

	Size								Power			
(n_i, p)	(25,20)		(50,40)		(25,240)		(25,480)		$(25, 480)$			
Scenario	SS	SSB	SS	SSB	SS	SSB	SS	SSB	SS	SSB	SS	SSB
(I)	6.0	4.4	6.5	4.2	3.6	4.2	3.0	4.6	24	30	28	31
(II)	6.0	4.6	6.4	5.3	2.1	4.2	3.0	4.5	75	80	78	81
(III)	5.7	5.2	6.3	5.5	3.2	4.1	2.8	4.5	55	80	52	67
(IV)	4.7	5.1	4.5	4.8	3.4	4.1	2.0	4.7	80	88	91	95
(V)	5.8	4.7	6.4	4.7	3.2	4.5	2.7	5.0	64	88	64	72

consistent with our preceding theoretical and numerical analysis.

4 Discussion

Our asymptotic and numerical results together suggest that the proposed spatial-sign test is quite robust and efficient in testing the equality of locations, especially for heavy-tailed or skewed distributions. It should be pointed out that when the data come from some light-tailed distributions, the SS is expected to be outperformed by the SKK test. This drawback is certainly inherited from the spatial-sign-based nature of SS and shared by all the sign-or rank- based procedures.

Our analysis in this paper shows that the spatial-sign-based test combined with the data transformation via the estimated diagonal matrix leads to a powerful test procedure. The analysis also shows that the data transformation is quite crucial in high-dimensional data. This confirms the benefit of the transformation discovered by Srivastava et al. (2013) for L_2-norm-based tests. In a significant development in another direction that using the max-norm rather than the L_2-norm, Cai et al. (2014) proposed a test based on the max-norm of marginal t-statistics. See also Zhong et al. (2013) for a related discussion. Generally speaking, the max-norm test is for more sparse and stronger signals whereas the L_2 norm

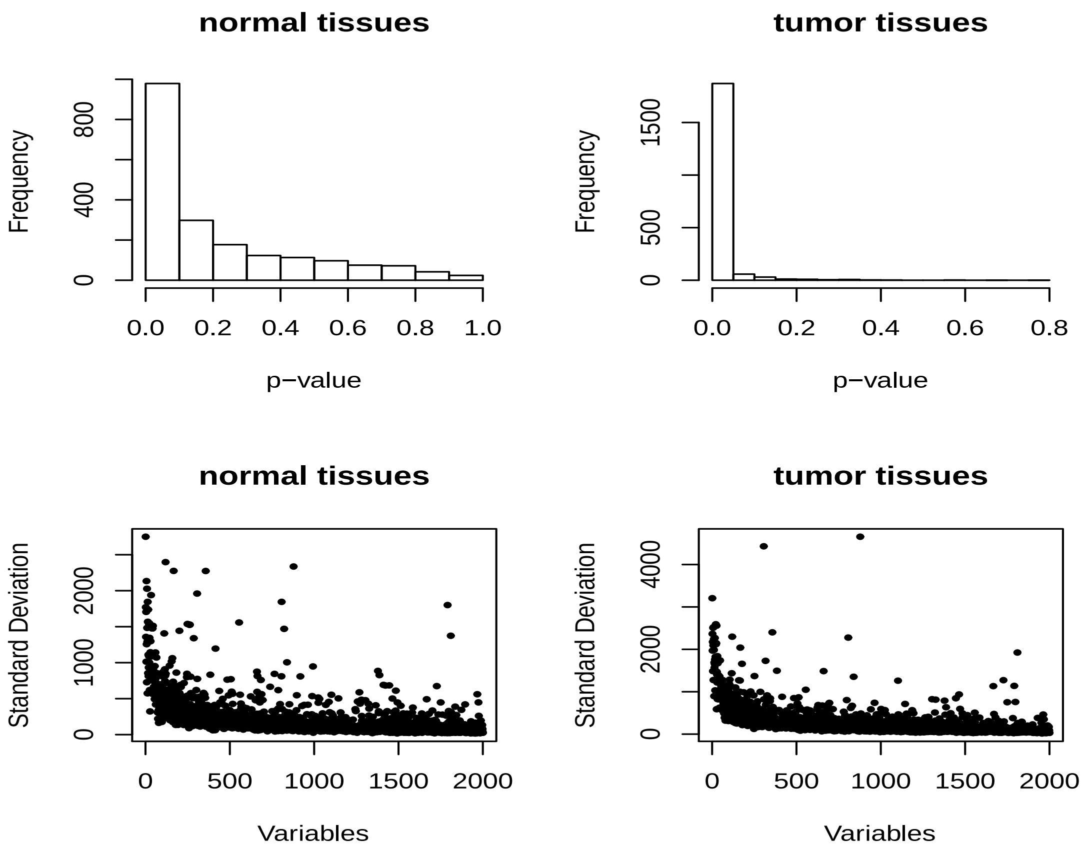

Figure 2: The p-values of normality test and standard deviation of each variables of the two samples.

test is for denser but fainter signals. Developing a spatial-sign-based test for sparse signals is of interest in the future study. Besides the spatial-sign-based sphericity test, there are some other tests based on spatial-signed-rank or spatial-rank in the literatures, such as Hallin and Paindaveine (2006). Deriving similar procedures for those tests are highly nontrivial due to their complicated construction and deserves some future research.

In our tests, we standardize the data for each variable to a common scale to account for heterogeneity of variance. On the other hand, Dudoit et al. (2002) standardized the gene expression data so that the observations (arrays) have unit variance across variables (genes). They pointed out that scale adjustments would be desirable in some cases to prevent the expression levels in one particular array from dominating the average expression levels across arrays. This way of standardization in high-dimensional settings warrants future study.

Acknowledgement

The authors would like to thank the Editor, Associate Editor and two anonymous referees for their constructive comments that lead to substantial improvement for the paper. In particular, we are grateful to Associate Editor for the suggestion of using resampling procedures.

Appendix

This appendix contains six parts but we only present here the Appendix A which gives the succinct proofs of Theorems 1-2. Some additional simulation results, technical arguments in Section 2 and the proofs of Lemmas 1,2,4,6 and 7 can be found in the Appendices B-F in the Supplemental Material. The proofs of Lemmas 3 and 5 are given below because they are the core of the technical arguments and may be interesting in their own rights.

Appendix A: The proofs of Theorems 1-2

Before proving the two main theorems in Section 2, we present several necessary lemmas.

Lemma 1 *For any matrix* $\mathbf{M}$*, we have* $E(\mathbf{u}_{ij}^T\mathbf{M}\mathbf{u}_{ij})^2 = O\left(p^{-2}\mathrm{tr}(\mathbf{M}^T\mathbf{M})\right), i = 1, 2, j = 1, \cdots, n_i$.

Define $\mathbf{D}_i = \mathrm{diag}\{d_{i1}^2, d_{i2}^2, \cdots, d_{ip}^2\}$, $i = 1, 2$ and $\boldsymbol{d}_i = (d_{i1}, d_{i2}, \cdots, d_{ip})$, $\boldsymbol{\eta}_i = (\boldsymbol{\theta}_i^T, \boldsymbol{d}_i)^T$. Let the corresponding estimator be $\hat{\boldsymbol{\eta}}_i = (\hat{\boldsymbol{\theta}}_i^T, \hat{\boldsymbol{d}}_i)^T$.

Lemma 2 *Under Condition (C4), we have* $\max_{1\le j\le p}(\hat{d}_{ij} - d_{ij}) = O_p(n_i^{-1/2}(\log p)^{1/2})$.

We denote $\mathbf{U}_{ij} = U(\mathbf{D}_i^{-1/2}(\mathbf{X}_{ij} - \boldsymbol{\theta}_i))$, $\hat{\mathbf{U}}_{ij} = U(\hat{\mathbf{D}}_i^{-1/2}(\mathbf{X}_{ij} - \boldsymbol{\theta}_i))$, $r_{ij} = ||\mathbf{D}_i^{-1/2}(\mathbf{X}_{ij} - \boldsymbol{\theta}_i)|| = ||\mathbf{R}_i^{-1/2}\boldsymbol{\varepsilon}_{ij}||$, $\hat{r}_{ij} = ||\hat{\mathbf{D}}_i^{-1/2}(\mathbf{X}_{ij} - \boldsymbol{\theta}_i)||$, and $r_{ij}^* = ||\boldsymbol{\varepsilon}_{ij}||$ for $j = 1, \ldots, n_i, i = 1, 2$. Define $\hat{\boldsymbol{\mu}}_i = \hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i$ and $\hat{\boldsymbol{\mu}}_{i,j} = \hat{\boldsymbol{\theta}}_{i,j} - \boldsymbol{\theta}_i, i = 1, 2$. The following lemma provides an asymptotic expansion for $\hat{\boldsymbol{\theta}}_i$. Note that given $\hat{\mathbf{D}}_i$, the estimator $\hat{\boldsymbol{\mu}}_i$ is the minimizer of the following objective function

$$L(\boldsymbol{\mu}) = ||\hat{\mathbf{D}}_i^{-1/2}(\mathbf{X}_{ij} - \boldsymbol{\theta}_i - \boldsymbol{\mu})||$$

or the estimator $\hat{\boldsymbol{\theta}}_i$ is equivalent to solve the equation

$$\sum_{k=1}^{n_i} U(\hat{\mathbf{D}}_i^{-1/2}(\mathbf{X}_{ik} - \boldsymbol{\theta})) = 0. \tag{A.1}$$

Lemma 3 *Under Conditions (C1), (C3) and (C4), $\hat{\boldsymbol{\mu}}_i$, $i = 1, 2$ admits the following asymptotic representation*

$$\hat{\boldsymbol{\mu}}_i = \frac{1}{n_i} c_i^{-1} \mathbf{D}_i^{1/2} \sum_{j=1}^{n_i} \mathbf{U}_{ij} + o_p(b_{n,p,i}),$$

where $c_i = E(r_{ij}^{-1})$ and $b_{n,p,i} = c_i^{-1} n^{-1/2}$.

Proof. We first show that $||\hat{\boldsymbol{\mu}}_i|| = O_p(b_{n,p,i})$. Note that the objective function $L(\boldsymbol{\mu})$ is a strictly convex function in $\boldsymbol{\mu}$. Thus as long as we can show that it has $b_{n,p,i}^{-1}$-consistent local minimizer, it must be $b_{n,p,i}^{-1}$-consistent global minimizer. The existence of a $b_{n,p,i}^{-1}$-consistent local minimizer is implied by that fact that for an arbitrarily small $\epsilon > 0$, there exist a sufficiently large constant C, which does no depend on n or p, such that

$$\liminf_n P\left(\inf_{\mathbf{u} \in \mathbb{R}^p, ||\mathbf{u}|| = C} L(b_{n,p,i}\mathbf{u}) > L(\mathbf{0}) \right) > 1 - \epsilon. \tag{A.2}$$

Next, we prove (A.2). Consider the expansion of $||\hat{\mathbf{D}}_i^{-1/2}(\mathbf{X}_{ij} - \boldsymbol{\theta}_i - b_{n,p,i}\mathbf{u})||$,

$$||\hat{\mathbf{D}}_i^{-1/2}(\mathbf{X}_{ij} - \boldsymbol{\theta}_i - b_{n,p,i}\mathbf{u})|| = ||\hat{\mathbf{D}}_i^{-1/2}(\mathbf{X}_{ij} - \boldsymbol{\theta}_i)||\{1 - 2b_{n,p,i}\hat{r}_{ij}^{-1}\mathbf{u}^T\hat{\mathbf{D}}_i^{-1/2}\hat{\mathbf{U}}_{ij} + b_{n,p,i}^2\hat{r}_{ij}^{-2}\mathbf{u}^T\hat{\mathbf{D}}_i^{-1}\mathbf{u}\}^{1/2}.$$

Because $b_{n,p,i}\hat{r}_{ij}^{-1}\mathbf{u}^T\hat{\mathbf{D}}_i^{-1/2}\hat{\mathbf{U}}_{ij} = O_p(n^{-1/2})$ and $b_{n,p,i}^2\hat{r}_{ij}^{-2}\mathbf{u}^T\hat{\mathbf{D}}_i^{-1}\mathbf{u} = O_p(n^{-1})$, we can see that

$$\begin{aligned}
&||\hat{\mathbf{D}}_i^{-1/2}(\mathbf{X}_{ij} - \boldsymbol{\theta}_i - b_{n,p,i}\mathbf{u})|| \\
&= ||\hat{\mathbf{D}}_i^{-1/2}(\mathbf{X}_{ij} - \boldsymbol{\theta}_i)|| - b_{n,p}\mathbf{u}^T\hat{\mathbf{D}}_i^{-1/2}\hat{\mathbf{U}}_{ij} + b_{n,p}^2(2\hat{r}_{ij})^{-1}\mathbf{u}^T\hat{\mathbf{D}}_i^{-1/2}(\mathbf{I}_p - \hat{\mathbf{U}}_{ij}\hat{\mathbf{U}}_{ij}^T)\hat{\mathbf{D}}_i^{-1/2}\mathbf{u} + O_p(c_i^{-1}n^{-3/2}).
\end{aligned}$$

So, it can be easily seen

$$
\begin{aligned}
&c_i(L(b_{n,p,i}\mathbf{u}) - L(\mathbf{0})) \\
&\quad = c_i \sum_{j=1}^{n_i} \{||\hat{\mathbf{D}}_i^{-1/2}(\mathbf{X}_{ij} - \boldsymbol{\theta}_i - b_{n,p,i}\mathbf{u})|| - ||\hat{\mathbf{D}}_i^{-1/2}(\mathbf{X}_{ij} - \boldsymbol{\theta}_i)||\} \\
&\quad = -n^{-1/2}\mathbf{u}^T\hat{\mathbf{D}}_i^{-1/2}\sum_{i=1}^{n}\hat{\mathbf{U}}_{ij} + c_i^{-1}n_i^{-1}\mathbf{u}^T\hat{\mathbf{D}}_i^{-1/2}\left\{\sum_{j=1}^{n_i}(2\hat{r}_{ij})^{-1}\left(\mathbf{I}_p - \hat{\mathbf{U}}_{ij}\hat{\mathbf{U}}_{ij}^T\right)\right\}\hat{\mathbf{D}}_i^{-1/2}\mathbf{u} + O_p(n^{-1/2}).
\end{aligned} \tag{A.3}
$$

First, as $E(||n_i^{-1/2}\sum_{i=1}^n \hat{\mathbf{U}}_{ij}||^2) = 1$ and $\mathrm{var}(||n_i^{-1/2}\sum_{i=1}^n \hat{\mathbf{U}}_{ij}||^2) = O(1)$, we know that $||n_i^{-1/2}\sum_{i=1}^n \hat{\mathbf{U}}_{ij}|| = O_p(1)$, and accordingly

$$
\left| -n^{-1/2}\mathbf{u}^T\hat{\mathbf{D}}_i^{-1/2}\sum_{i=1}^{n}\hat{\mathbf{U}}_{ij} \right| \le ||\hat{\mathbf{D}}_i^{-1/2}\mathbf{u}||||n_i^{-1/2}\sum_{i=1}^{n}\hat{\mathbf{U}}_{ij}|| = O_p(1).
$$

Define $\mathbf{A} = n_i^{-1}\sum_{j=1}^{n_i}(2\hat{r}_{ij})^{-1}\hat{\mathbf{U}}_{ij}\hat{\mathbf{U}}_{ij}^T$. After some tedious calculation, we can obtain that $E(\mathrm{tr}(\mathbf{A}^2)) = O(c_i^2 n^{-1}))$. Then $E(\mathbf{u}^T\hat{\mathbf{D}}_i^{-1/2}\mathbf{A}\hat{\mathbf{D}}_i^{-1/2}\mathbf{u})^2 \le E((\mathbf{u}^T\hat{\mathbf{D}}_i^{-1}\mathbf{u})^2\mathrm{tr}(\mathbf{A}^2)) = O(c_i^2 n^{-1})$, which leads to $\mathbf{u}^T\hat{\mathbf{D}}_i^{-1/2}\mathbf{A}\hat{\mathbf{D}}_i^{-1/2}\mathbf{u} = O_p(c_i n^{-1/2})$. Thus, we have

$$
n_i^{-1}\mathbf{u}^T\hat{\mathbf{D}}_i^{-1/2}\left\{\sum_{j=1}^{n_i}\hat{r}_{ij}^{-1}\left(\mathbf{I}_p - \hat{\mathbf{U}}_{ij}\hat{\mathbf{U}}_{ij}^T\right)\right\}\hat{\mathbf{D}}_i^{-1/2}\mathbf{u} = n_i^{-1}\mathbf{u}^T\hat{\mathbf{D}}_i^{-1/2}\sum_{j=1}^{n_i}\hat{r}_{ij}^{-1}\hat{\mathbf{D}}_i^{-1/2}\mathbf{u} + O_p(c_i n^{-1/2}),
$$

where we use the fact that $n_i^{-1}\sum_{j=1}^{n_i}\hat{r}_{ij}^{-1} = c_i + O_p(c_i n^{-1/2})$. By choosing a sufficiently large C, the second term in (A.3) dominates the first term uniformly in $||\mathbf{u}|| = C$. Hence, (A.2) holds and accordingly $\hat{\boldsymbol{\mu}}_i = O_p(b_{n,p,i})$.

Finally, by a first-order Taylor expansion of (A.1), we have

$$
\sum_{j=1}^{n_i}(\hat{\mathbf{U}}_{ij} - \hat{r}_{ij}^{-1}\hat{\mathbf{D}}_i^{-1/2}\hat{\boldsymbol{\mu}}_i))\left\{1 + \hat{r}_{ij}^{-1}\hat{\mathbf{U}}_{ij}^T\hat{\mathbf{D}}_i^{-1/2}\hat{\boldsymbol{\mu}}_i + O_p(n^{-1})\right\} = 0,
$$

and then

$$
\begin{aligned}
&\left\{n_i^{-1}\sum_{j=1}^{n_i}\hat{r}_{ij} + O_p(c_i n^{-1/2})\right\}\hat{\mathbf{D}}_i^{-1/2}\hat{\boldsymbol{\mu}}_i - \left(n_i^{-1}\sum_{j=1}^{n_i}\hat{r}_{ij}\hat{\mathbf{U}}_{ij}\hat{\mathbf{U}}_{ij}^T\right)\hat{\mathbf{D}}_i^{-1/2}\hat{\boldsymbol{\mu}}_i \\
&= \left(n_i^{-1}\sum_{j=1}^{n_i}\hat{\mathbf{U}}_{ij}\right)(1 + O_p(n^{-1})).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\hat{\boldsymbol{\mu}}_i =&(c_i + O_p(c_i n^{-1/2}))^{-1}\hat{\mathbf{D}}_i^{-1/2}\left(n_i^{-1}\sum_{j=1}^{n_i}\hat{\mathbf{U}}_{ij}\right)(1+O_p(n^{-1}))\\
=&c_i^{-1}\hat{\mathbf{D}}_i^{1/2}\left(n_i^{-1}\sum_{j=1}^{n_i}\hat{\mathbf{U}}_{ij}\right)+o_p(b_{n,p,i})\\
=&c_i^{-1}\mathbf{D}_i^{1/2}\left(n_i^{-1}\sum_{j=1}^{n_i}\mathbf{U}_{ij}\right)+c_i^{-1}\hat{\mathbf{D}}_i^{1/2}\left(n_i^{-1}\sum_{j=1}^{n_i}\hat{\mathbf{U}}_{ij}-n_i^{-1}\sum_{j=1}^{n_i}\mathbf{U}_{ij}\right)\\
&+c_i^{-1}(\hat{\mathbf{D}}_i^{1/2}-\mathbf{D}_i^{1/2})\left(n_i^{-1}\sum_{j=1}^{n_i}\mathbf{U}_{ij}\right)+o_p(b_{n,p,i})\\
\doteq&c_i^{-1}\mathbf{D}_i^{1/2}\left(n_i^{-1}\sum_{j=1}^{n_i}\mathbf{U}_{ij}\right)+B_1+B_2+o_p(b_{n,p,i}).
\end{aligned}
$$

Next, we will show that $B_1 = o_p(b_{n,p,i})$ and $B_2 = o_p(b_{n,p,i})$. By the Taylor expansion,

$$
\begin{aligned}
U(\hat{\mathbf{D}}_i^{-1/2}(\mathbf{X}_{ij}-\boldsymbol{\theta}_i)) =&\mathbf{U}_{ij}+\{\mathbf{I}_p-\mathbf{U}_{ij}\mathbf{U}_{ij}^T\}(\hat{\mathbf{D}}_i^{-1/2}-\mathbf{D}_i^{-1/2})\mathbf{D}_i^{1/2}\mathbf{U}_{ij}\\
&+\frac{1}{2r_{ij}^2}||(\hat{\mathbf{D}}_i^{-1/2}-\mathbf{D}_i^{-1/2})(\mathbf{X}_{ij}-\boldsymbol{\theta}_i)||\mathbf{U}_{ij}+o_p(n^{-1}).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
n_i^{-1}\sum_{j=1}^{n_i}\hat{\mathbf{U}}_{ij}-n_i^{-1}\sum_{j=1}^{n_i}\mathbf{U}_{ij} =&n_i^{-1}\sum_{j=1}^{n_i}\{\mathbf{I}_p-\mathbf{U}_{ij}\mathbf{U}_{ij}^T\}(\hat{\mathbf{D}}_i^{-1/2}-\mathbf{D}_i^{-1/2})\mathbf{D}_i^{1/2}\mathbf{U}_{ij}\\
&+n_i^{-1}\sum_{j=1}^{n_i}\frac{1}{2r_{ij}^2}||(\hat{\mathbf{D}}_i^{-1/2}-\mathbf{D}_i^{-1/2})(\mathbf{X}_{ij}-\boldsymbol{\theta}_i)||^2\mathbf{U}_{ij}+o_p(n^{-1})\\
\doteq&B_{11}+B_{12}+o_p(n^{-1}),
\end{aligned}
$$

and according to Lemma 2,

$$
\begin{aligned}
E(||B_{11}||^2)\le C(\log p/n)^{1/2}E||n_i^{-1}\sum_{j=1}^{n_i}\mathbf{U}_{ij}||^2 = O((\log p)^{1/2}n^{-3/2}) = o(n^{-1}),\\
E(||B_{12}||^2)\le C(\log p/n)^{1/2}E||n_i^{-1}\sum_{j=1}^{n_i}\mathbf{U}_{ij}||^2 = O((\log p)^{1/2}n^{-3/2}) = o(n^{-1}),
\end{aligned}
$$

by Condition (C3). So, $B_1 = o_p(b_{n,p,i})$. Similar to B_1, we can also show that $B_2 = o_p(b_{n,p,i})$, from which the assertion in this lemma immediately follows. $\square$

Similar to Lemma 3, we also have

$$\hat{\boldsymbol{\theta}}_{i,l} = \boldsymbol{\theta}_i + \frac{1}{n_i - 1} c_i^{-1} \mathbf{D}_i^{1/2} \sum_{j \neq l}^{n_i} \mathbf{U}_{ij} + o_p(b_{n,p,i}), \quad i = 1, 2.$$

The next lemma measures the asymptotic difference between $U(\hat{\mathbf{D}}_{k,i}^{-1/2}(\mathbf{X}_{ki} - \hat{\boldsymbol{\theta}}_{3-k,j}))$ and $\mathbf{U}_{ki}$, $k = 1, 2$.

Lemma 4 *Under H_0 and Condition (C1), for $k = 1, 2$,*

$$\begin{aligned} U(\hat{\mathbf{D}}_{k,i}^{-1/2}(\mathbf{X}_{ki} - \hat{\boldsymbol{\theta}}_{3-k,j})) =& \mathbf{U}_{ki} - \frac{1}{r_{ki}} [\mathbf{I}_p - \mathbf{U}_{ki} \mathbf{U}_{ki}^T] \mathbf{D}_k^{-1/2} \hat{\boldsymbol{\mu}}_{3-k,j} + [\mathbf{I}_p - \mathbf{U}_{ki} \mathbf{U}_{ki}^T](\hat{\mathbf{D}}_{k,i}^{-1/2} \mathbf{D}_k^{1/2} - \mathbf{I}_p) \mathbf{U}_{ki} \\ & - \frac{1}{r_{ki}} [\mathbf{I}_p - \mathbf{U}_{ki} \mathbf{U}_{ki}^T](\hat{\mathbf{D}}_{k,i}^{-1/2} - \mathbf{D}_k^{-1/2}) \hat{\boldsymbol{\mu}}_{3-k,j} \\ & + \frac{1}{2 r_{ki}^2} ||(\hat{\mathbf{D}}_{k,i}^{-1/2} - \mathbf{D}_k^{-1/2})(\mathbf{X}_{ki} - \boldsymbol{\theta}) - \hat{\mathbf{D}}_{k,i}^{-1/2} \hat{\boldsymbol{\mu}}_{3-k,j}||^2 \mathbf{U}_{ki} + o_p(n^{-1}). \end{aligned}$$

Next, we will give an asymptotic equivalence to R_n under H_0.

Lemma 5 *Under H_0 and Conditions (C1)-(C4), $R_n = Z_n + o_p(n^{-2})$, where*

$$Z_n = \frac{\sum_{i=1}^{n_1} \sum_{i \neq j}^{n_1} \mathbf{u}_{1i}^T \mathbf{A}_1 \mathbf{u}_{1j}}{n_1(n_1 - 1)} + \frac{\sum_{i=1}^{n_2} \sum_{i \neq j}^{n_2} \mathbf{u}_{2i}^T \mathbf{A}_2 \mathbf{u}_{2j}}{n_2(n_2 - 1)} - 2 \frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \mathbf{u}_{1i}^T \mathbf{A}_3 \mathbf{u}_{2j}}{n_1 n_2}.$$

Proof. By Lemma 2,

$$\begin{aligned}
&U(\hat{\mathbf{D}}_{1,i}^{-1/2}(\mathbf{X}_{1i}-\hat{\boldsymbol{\theta}}_{2,j}))^T U(\hat{\mathbf{D}}_{2,j}^{-1/2}(\mathbf{X}_{2j}-\hat{\boldsymbol{\theta}}_{1,i}))\\
&\quad=\mathbf{U}_{1i}^T\mathbf{U}_{2j}-\frac{1}{r_{1i}}\hat{\boldsymbol{\mu}}_{2,j}^T[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T]\mathbf{D}_1^{-1/2}\mathbf{U}_{2j}-\frac{1}{r_{1i}}\hat{\boldsymbol{\mu}}_{2,j}^T[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T](\hat{\mathbf{D}}_{1,i}^{-1/2}-\mathbf{D}_1^{-1/2})\mathbf{U}_{2j}\\
&\qquad+\mathbf{U}_{1i}^T(\hat{\mathbf{D}}_{1,i}^{-1/2}\mathbf{D}_1^{1/2}-\mathbf{I}_p)[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T]\mathbf{U}_{2j}\\
&\qquad+\frac{1}{2r_{1i}^2}||(\hat{\mathbf{D}}_{1,i}^{-1/2}-\mathbf{D}_1^{-1/2})(\mathbf{X}_{1i}-\boldsymbol{\theta})-\hat{\mathbf{D}}_{1,i}^{-1/2}\hat{\boldsymbol{\mu}}_{2,j}||^2\mathbf{U}_{1i}^T\mathbf{U}_{2j}\\
&\qquad-\frac{1}{r_{2j}}\mathbf{U}_{1i}^T[\mathbf{I}_p-\mathbf{U}_{2j}\mathbf{U}_{2j}^T]\mathbf{D}_2^{-1/2}\hat{\boldsymbol{\mu}}_{1,i}+\frac{1}{r_{1i}r_{2j}}\hat{\boldsymbol{\mu}}_{2,j}^T[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T][\mathbf{I}_p-\mathbf{U}_{2j}\mathbf{U}_{2j}^T]\mathbf{D}_1^{-1/2}\mathbf{D}_2^{-1/2}\hat{\boldsymbol{\mu}}_{1,i}\\
&\qquad+\frac{1}{r_{2j}}\mathbf{U}_{1i}^T(\hat{\mathbf{D}}_{1,i}^{-1/2}\mathbf{D}_1^{1/2}-\mathbf{I}_p)[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T][\mathbf{I}_p-\mathbf{U}_{2j}\mathbf{U}_{2j}^T]\mathbf{D}_2^{-1/2}\hat{\boldsymbol{\mu}}_{1,i}\\
&\qquad+\frac{1}{r_{1i}r_{2j}}\hat{\boldsymbol{\mu}}_{2,j}^T[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T](\hat{\mathbf{D}}_{1,i}^{-1/2}-\mathbf{D}_1^{-1/2})[\mathbf{I}_p-\mathbf{U}_{2j}\mathbf{U}_{2j}^T]\mathbf{D}_2^{-1/2}\hat{\boldsymbol{\mu}}_{1,i}\\
&\qquad-\frac{1}{2r_{1i}^2r_{2j}}||(\hat{\mathbf{D}}_{1,i}^{-1/2}-\mathbf{D}_1^{-1/2})(\mathbf{X}_{1i}-\boldsymbol{\theta})-\hat{\mathbf{D}}_{1,i}^{-1/2}\hat{\boldsymbol{\mu}}_{2,j}||^2\mathbf{U}_{1i}^T[\mathbf{I}_p-\mathbf{U}_{2j}\mathbf{U}_{2j}^T]\mathbf{D}_2^{-1/2}\hat{\boldsymbol{\mu}}_{1,i}\\
&\qquad-\frac{1}{r_{2j}}U(\hat{\mathbf{D}}_{1,i}^{-1/2}(\mathbf{X}_{1i}-\hat{\boldsymbol{\theta}}_{2,j}))^T[\mathbf{I}_p-\mathbf{U}_{2j}\mathbf{U}_{2j}^T](\hat{\mathbf{D}}_{2,j}^{-1/2}-\mathbf{D}_2^{-1/2})\hat{\boldsymbol{\mu}}_{1,i}\\
&\qquad+\frac{1}{2r_{2j}^2}||(\hat{\mathbf{D}}_{2,j}^{-1/2}-\mathbf{D}_2^{-1/2})(\mathbf{X}_{2i}-\boldsymbol{\theta})-\hat{\mathbf{D}}_{2,j}^{-1/2}\hat{\boldsymbol{\mu}}_{1,i}||^2U(\hat{\mathbf{D}}_1^{-1/2}(\mathbf{X}_{1i}-\hat{\boldsymbol{\theta}}_{2,j}))^T\mathbf{U}_{2j}+o_p(n^{-2}),
\end{aligned}$$

which implies that

$$\begin{aligned}
R_n=&\frac{1}{n_1n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\frac{1}{r_{1i}}\hat{\boldsymbol{\mu}}_{2,j}^T[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T]\mathbf{D}_1^{-1/2}\mathbf{U}_{2j}+\frac{1}{n_1n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\frac{1}{r_{2j}}\mathbf{U}_{1i}^T[\mathbf{I}_p-\mathbf{U}_{2j}\mathbf{U}_{2j}^T]\mathbf{D}_2^{-1/2}\hat{\boldsymbol{\mu}}_{1,i}\\
&-\frac{1}{n_1n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left(\mathbf{U}_{1i}^T\mathbf{U}_{2j}+\frac{1}{r_{1i}r_{2j}}\hat{\boldsymbol{\mu}}_{2,j}^T[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T][\mathbf{I}_p-\mathbf{U}_{2j}\mathbf{U}_{2j}^T]\mathbf{D}_1^{-1/2}\mathbf{D}_2^{-1/2}\hat{\boldsymbol{\mu}}_{1,i}\right)\\
&-\frac{1}{n_1n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\mathbf{U}_{1i}^T[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T](\hat{\mathbf{D}}_1^{-1/2}\mathbf{D}_1^{1/2}-\mathbf{I}_p)\mathbf{U}_{2j}+Q_n+o_p(n^{-2}),
\end{aligned}$$

where Q_n denote the rest parts of R_n. For simplicity, we only show that

$$\frac{1}{n_1n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\mathbf{U}_{1i}^T(\hat{\mathbf{D}}_{1,i}^{-1/2}\mathbf{D}_1^{1/2}-\mathbf{I}_p)[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T]\mathbf{U}_{2j}=o_p(\sigma_n),$$

and we can show that the other parts in Q_n are all $o_p(\sigma_n)$ by using similar arguments.

$$\begin{aligned}
&\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\mathbf{U}_{1i}^T(\hat{\mathbf{D}}_{1,i}^{-1/2}\mathbf{D}_1^{1/2}-\mathbf{I}_p)[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T]\mathbf{U}_{2j}\\
&\qquad=\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\mathbf{U}_{1i}^T(\hat{\mathbf{D}}_{1,i}^{-1/2}\mathbf{D}_1^{1/2}-\mathbf{I}_p)\mathbf{U}_{2j}+\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\mathbf{U}_{1i}^T(\hat{\mathbf{D}}_{1,i}^{-1/2}\mathbf{D}_1^{1/2}-\mathbf{I}_p)\mathbf{U}_{1i}\mathbf{U}_{1i}^T\mathbf{U}_{2j}\\
&\qquad\doteq G_{n1}+G_{n2}.
\end{aligned}$$

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

$$\begin{aligned}
E\left(G_{n1}^2\right)=&\frac{1}{n_1^2 n_2^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}E\left(\left(\mathbf{U}_{1i}^T(\hat{\mathbf{D}}_{1,i}^{-1/2}\mathbf{D}_1^{1/2}-\mathbf{I}_p)\mathbf{U}_{2j}\right)^2\right)\\
=&\frac{1}{n_1^2 n_2^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}E\left(\frac{\left(\mathbf{u}_{1i}^T\boldsymbol{\Sigma}_1^{1/2}\mathbf{D}_1^{-1/2}(\hat{\mathbf{D}}_{1,i}^{-1/2}\mathbf{D}_1^{1/2}-\mathbf{I}_p)\mathbf{D}_2^{-1/2}\boldsymbol{\Sigma}_2^{1/2}\mathbf{u}_{2j}\right)^2}{(1+\mathbf{u}_{1i}^T(\mathbf{R}_1-\mathbf{I}_p)\mathbf{u}_{1i})(1+\mathbf{u}_{2j}^T(\mathbf{R}_2-\mathbf{I}_p)\mathbf{u}_{2j})}\right)\\
\le&\frac{1}{n_1^2 n_2^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left\{E\left(\mathbf{u}_{1i}^T\boldsymbol{\Sigma}_1^{1/2}\mathbf{D}_1^{-1/2}(\hat{\mathbf{D}}_{1,i}^{-1/2}\mathbf{D}_1^{1/2}-\mathbf{I}_p)\mathbf{D}_2^{-1/2}\boldsymbol{\Sigma}_2^{1/2}\mathbf{u}_{2j}\right)^2\right.\\
&+CE\left(\left(\mathbf{u}_{1i}^T\boldsymbol{\Sigma}_1^{1/2}\mathbf{D}_1^{-1/2}(\hat{\mathbf{D}}_{1,i}^{-1/2}\mathbf{D}_1^{1/2}-\mathbf{I}_p)\mathbf{D}_2^{-1/2}\boldsymbol{\Sigma}_2^{1/2}\mathbf{u}_{2j}\right)^2\mathbf{u}_{1i}^T(\mathbf{R}_1-\mathbf{I}_p)\mathbf{u}_{1i}\right)\\
&\left.+CE\left(\left(\mathbf{u}_{1i}^T\boldsymbol{\Sigma}_1^{1/2}\mathbf{D}_1^{-1/2}(\hat{\mathbf{D}}_{1,i}^{-1/2}\mathbf{D}_1^{1/2}-\mathbf{I}_p)\mathbf{D}_2^{-1/2}\boldsymbol{\Sigma}_2^{1/2}\mathbf{u}_{2j}\right)^2\mathbf{u}_{2j}^T(\mathbf{R}_2-\mathbf{I}_p)\mathbf{u}_{2j}\right)\right\},
\end{aligned}$$

where the last inequality follows by the Taylor expansion. Define $\mathbf{H}=\boldsymbol{\Sigma}_1^{1/2}\mathbf{D}_1^{-1/2}(\hat{\mathbf{D}}_{1,i}^{-1/2}\mathbf{D}_1^{1/2}-\mathbf{I}_p)\mathbf{D}_2^{-1/2}\boldsymbol{\Sigma}_2^{1/2}$ and then according to Lemma 2, $\mathrm{tr}(E(\mathbf{H}^2))=o(\mathrm{tr}(\mathbf{A}_3^T\mathbf{A}_3))$ and $\mathrm{tr}(E(\mathbf{H}^4))=o(\mathrm{tr}((\mathbf{A}_3^T\mathbf{A}_3)^2))=o(\mathrm{tr}^2(\mathbf{A}_3^T\mathbf{A}_3))$ by Condition (C2). By the Cauchy inequality, we have

$$\begin{aligned}
E\left(\mathbf{u}_{1i}^T\boldsymbol{\Sigma}_1^{1/2}\mathbf{D}_1^{-1/2}(\hat{\mathbf{D}}_{1,i}^{-1/2}\mathbf{D}_1^{1/2}-\mathbf{I}_p)\mathbf{D}_2^{-1/2}\boldsymbol{\Sigma}_2^{1/2}\mathbf{u}_{2j}\right)^2&=p^{-2}\mathrm{tr}(\mathbf{H}^2)=o(p^{-2}\mathrm{tr}(\mathbf{R}_1\mathbf{R}_2)),\\
E((\mathbf{u}_{1i}^T\mathbf{H}\mathbf{u}_{2j})^2\mathbf{u}_{1i}^T(\mathbf{R}_1-\mathbf{I}_p)\mathbf{u}_{1i})&\le(E(\mathbf{u}_{1i}^T\mathbf{H}\mathbf{u}_{2j})^4E((\mathbf{u}_{1i}^T(\mathbf{R}_1-\mathbf{I}_p)\mathbf{u}_{1i})^2)^{1/2}\\
&\le(p^{-4}\mathrm{tr}(E(\mathbf{H}^4))p^{-2}(\mathrm{tr}(\mathbf{R}_1-\mathbf{I}_p)))^{1/2},\\
E((\mathbf{u}_{1i}^T\mathbf{H}\mathbf{u}_{2j})^2\mathbf{u}_{2j}^T(\mathbf{R}_2-\mathbf{I}_p)\mathbf{u}_{2j})&\le(E(\mathbf{u}_{1i}^T\mathbf{H}\mathbf{u}_{2j})^4E((\mathbf{u}_{2j}^T(\mathbf{R}_2-\mathbf{I}_p)\mathbf{u}_{2j})^2)^{1/2}\\
&\le(p^{-4}\mathrm{tr}(E(\mathbf{H}^4))p^{-2}(\mathrm{tr}(\mathbf{R}_2-\mathbf{I}_p)))^{1/2}.
\end{aligned}$$

So we obtain that $G_{n1} = o_p(\sigma_n)$ by Condition (C4). Taking the same procedure, we can also show that $G_{n2} = o_p(\sigma_n)$. Moreover,

$$\begin{aligned}
&\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} r_{2j}^{-1}\mathbf{U}_{1i}^T[\mathbf{I}_p - \mathbf{U}_{2j}\mathbf{U}_{2j}^T]\mathbf{D}_2^{-1/2}\hat{\boldsymbol{\mu}}_{1,i} \\
&\quad = \frac{1}{n_1 n_2 (n_1-1)}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} r_{2j}^{-1}\mathbf{U}_{1i}^T[\mathbf{I}_p - \mathbf{U}_{2j}\mathbf{U}_{2j}^T]\mathbf{D}_2^{-1/2}\mathbf{D}_1^{1/2}c_1^{-1}\sum_{l\neq i}^{n_1}\mathbf{U}_{1l}(1+o_p(1)) \\
&\quad = \frac{1}{n_1(n_1-1)}\sum_{i=1}^{n_1}\sum_{l\neq i}^{n_1}\mathbf{U}_{1i}^T\left(\frac{1}{n_2}\sum_{j=1}^{n_2} r_{2j}^{-1}[\mathbf{I}_p - \mathbf{U}_{2j}\mathbf{U}_{2j}^T]\right)\mathbf{D}_2^{-1/2}\mathbf{D}_1^{1/2}c_1^{-1}\mathbf{U}_{1l}(1+o_p(1)) \\
&\quad = \frac{1}{n_1(n_1-1)}\sum_{i=1}^{n_1}\sum_{l\neq i}^{n_1} c_2 c_1^{-1}\mathbf{U}_{1i}^T\mathbf{D}_2^{-1/2}\mathbf{D}_1^{1/2}\mathbf{U}_{1l}(1+o_p(1)),
\end{aligned}$$

and

$$\begin{aligned}
J_{n1} =& \frac{1}{n_1(n_1-1)}\sum_{i=1}^{n_1}\sum_{l\neq i}^{n_1} c_2 c_1^{-1}\frac{\mathbf{u}_{1i}^T\boldsymbol{\Sigma}_1^{1/2}\mathbf{D}_1^{-1/2}\mathbf{D}_2^{-1/2}\boldsymbol{\Sigma}_1^{1/2}\mathbf{u}_{1l}}{(1+\mathbf{u}_{1i}^T(\mathbf{R}_1-\mathbf{I}_p)\mathbf{u}_{1i})^{1/2}(1+\mathbf{u}_{1l}^T(\mathbf{R}_1-\mathbf{I}_p)\mathbf{u}_{1l})^{1/2}} \\
=& \frac{1}{n_1(n_1-1)}\sum_{i=1}^{n_1}\sum_{l\neq i}^{n_1} c_2 c_1^{-1}\mathbf{u}_{1i}^T\boldsymbol{\Sigma}_1^{1/2}\mathbf{D}_1^{-1/2}\mathbf{D}_2^{-1/2}\boldsymbol{\Sigma}_1^{1/2}\mathbf{u}_{1l} \\
&+ \frac{1}{n_1(n_1-1)}\sum_{i=1}^{n_1}\sum_{l\neq i}^{n_1} c_2 c_1^{-1}\mathbf{u}_{1i}^T C_{ni}\boldsymbol{\Sigma}_1^{1/2}\mathbf{D}_1^{-1/2}\mathbf{D}_2^{-1/2}\boldsymbol{\Sigma}_1^{1/2}\mathbf{u}_{1l}\mathbf{u}_{1i}^T(\mathbf{R}_1-\mathbf{I}_p)\mathbf{u}_{1i} \\
\doteq& J_{n11} + J_{n12},
\end{aligned}$$

where C_{ni} is a bounded random variable between -1 and $-(\mathbf{u}_{1i}^T\mathbf{R}_1\mathbf{u}_{1i})^2$. By the same arguments as G_{n1}, we can show that $J_{n12} = o_p(\sigma_n)$. Thus,

$$\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} r_{2j}^{-1}\mathbf{U}_{1i}^T[\mathbf{I}_p - \mathbf{U}_{2j}\mathbf{U}_{2j}^T]\mathbf{D}_2^{-1/2}\hat{\boldsymbol{\mu}}_{1,i} = \frac{1}{n_1(n_1-1)}\sum_{i=1}^{n_1}\sum_{l\neq i}^{n_1}\mathbf{u}_{1i}^T\mathbf{A}_1\mathbf{u}_{1l} + o_p(\sigma_n).$$

Similarly,

$$\frac{1}{n_1n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\frac{1}{r_{1i}}\hat{\boldsymbol{\mu}}_{2,j}^T[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T]\mathbf{D}_1^{-1/2}\mathbf{U}_{2j}=\frac{1}{n_2(n_2-1)}\sum_{i=1}^{n_2}\sum_{l\neq i}^{n_2}\mathbf{u}_{2i}^T\mathbf{A}_2\mathbf{u}_{2l}+o_p(\sigma_n),$$

$$\begin{aligned}\frac{1}{n_1n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\frac{1}{r_{1i}r_{2j}}\hat{\boldsymbol{\mu}}_{2,j}^T[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T][\mathbf{I}_p-\mathbf{U}_{2j}\mathbf{U}_{2j}^T]\mathbf{D}_1^{-1/2}\mathbf{D}_2^{-1/2}\hat{\boldsymbol{\mu}}_{1,i}&\\ =\frac{1}{n_1n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\mathbf{u}_{1i}^T\mathbf{A}_3\mathbf{u}_{2j}+o_p(\sigma_n),&\end{aligned}$$

$$\frac{1}{n_1n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\mathbf{U}_{1i}^T\mathbf{U}_{2j}=\frac{1}{n_1n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\mathbf{u}_{1i}^T\mathbf{A}_3\mathbf{u}_{2j}+o_p(\sigma_n).$$

Finally, under H_0,

$$R_n=\frac{\sum_{i\neq j}^{n_1}\mathbf{u}_{1i}^T\mathbf{A}_1\mathbf{u}_{1j}}{n_1(n_1-1)}+\frac{\sum_{i\neq j}^{n_2}\mathbf{u}_{2i}^T\mathbf{A}_2\mathbf{u}_{2j}}{n_2(n_2-1)}-2\frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\mathbf{u}_{1i}^T\mathbf{A}_3\mathbf{u}_{2j}}{n_1n_2}+o_p(\sigma_n).$$

It can be easily verified that $E(Z_n)=0$ and

$$\mathrm{var}(Z_n)=\frac{2}{n_1(n_1-1)p^2}\mathrm{tr}(\mathbf{A}_1^2)+\frac{2}{n_2(n_2-1)p^2}\mathrm{tr}(\mathbf{A}_2^2)+\frac{4}{n_1n_2p^2}\mathrm{tr}(\mathbf{A}_3^T\mathbf{A}_3).$$

By Condition (C3), $\mathrm{var}(Z_n)=O(n^{-2})$. Consequently, we have $R_n=Z_n+o_p(n^{-2})$. □

Now, to prove Theorem 1, it remains to show that Z_n is asymptotically normal. Clearly, Z_n is essentially a two-sample U-statistic with order two. However, the standard CLT for U-statistics (Serfling 1980) is not directly applicable because the conditional variance of kernel function is zero. However, the martingale central limit theorem (Hall and Hyde 1980) can be used here. We require some additional lemmas stated as follows.

Let $\mathbf{Y}_i=\mathbf{u}_{1i}$ for $i=1,\dots,n_1$ and $\mathbf{Y}_{j+n1}=\mathbf{u}_{2j}$ for $j=1,\dots,n_2$ and for $i\neq j$,

$$\phi_{ij}=\begin{cases}n_1^{-1}(n_1-1)^{-1}\mathbf{Y}_i^T\mathbf{A}_1\mathbf{Y}_j, & i,j\in\{1,2,\dots,n_1\},\\ -n_1^{-1}n_2^{-1}\mathbf{Y}_i^T\mathbf{A}_3\mathbf{Y}_j, & i\in\{1,2,\dots,n_1\},j\in\{n_1+1,\dots,n\},\\ n_2^{-1}(n_2-1)^{-1}\mathbf{Y}_i^T\mathbf{A}_2\mathbf{Y}_j, & i,j\in\{n_1+1,n_1+2,\dots,n\}.\end{cases}$$

Define $Z_{nj}=\sum_{i=1}^{j-1}\phi_{ij}$ for $j=2,3,\dots,n$, $S_{nm}=\sum_{j=1}^{m}Z_{nj}$ and $\mathcal{F}_{nm}=\sigma\{\mathbf{Y}_1,\mathbf{Y}_2,\dots,\mathbf{Y}_m\}$ which is the σ-algebra generated by $\{\mathbf{Y}_1,\mathbf{Y}_2,\dots,\mathbf{Y}_m\}$. Now

$$Z_n=2\sum_{j=2}^{n}Z_{nj}.$$

We can verify that for each n, $\{S_{nm}, \mathcal{F}_{nm}\}_{m=1}^n$ is the sequence of zero mean and a square integrable martingale. In order to prove the normality of Z_n, it suffices to show the following two lemmas. The proofs of these two lemmas are straightforward but the calculation involved is rather tedious, and thus they are included in Appendix D of the Supplemental Material.

Lemma 6 *Suppose the conditions given in Theorem 1 all hold. Then,*

$$\frac{\sum_{j=2}^n E[Z_{nj}^2|\mathcal{F}_{n,j-1}]}{\sigma_n^2} \xrightarrow{p} \frac{1}{4}.$$

Lemma 7 *Suppose the conditions given in Theorem 1 all hold. Then,*

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

Proof of Theorem 1: Based on Corollary 3.1 of Hall and Heyde (1980), Lemmas 6-7, it can be concluded that $Z_n/\sigma_n \xrightarrow{d} N(0,1)$. By combining Lemma 5, we can obtain the assertion of this theorem immediately. $\square$

Proof of Theorem 2 Similar to Lemma 4, we can obtain that

$$\begin{aligned}
&U(\hat{\mathbf{D}}_{1,i}^{-1/2}(\mathbf{X}_{1i}-\hat{\boldsymbol{\theta}}_{2,j}))\\
=&\mathbf{U}_{1i} - \frac{1}{r_{1i}}[\mathbf{I}_p - \mathbf{U}_{1i}\mathbf{U}_{1i}^T]\mathbf{D}_1^{-1/2}\hat{\boldsymbol{\mu}}_{2,j} + \frac{1}{r_{1i}}[\mathbf{I}_p - \mathbf{U}_{1i}\mathbf{U}_{1i}^T]\mathbf{D}_1^{-1/2}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)\\
&-\frac{1}{r_{1i}}[\mathbf{I}_p - \mathbf{U}_{1i}\mathbf{U}_{1i}^T](\hat{\mathbf{D}}_{1,i}^{-1/2}-\mathbf{D}_1^{-1/2})\hat{\boldsymbol{\mu}}_{2,j} + \frac{1}{r_{1i}}[\mathbf{I}_p - \mathbf{U}_{1i}\mathbf{U}_{1i}^T](\hat{\mathbf{D}}_{1,i}^{-1/2}-\mathbf{D}_1^{-1/2})(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)\\
&+\frac{1}{2r_{1i}^2}||(\hat{\mathbf{D}}_{1,i}^{-1/2}-\mathbf{D}_1^{-1/2})(\mathbf{X}_{1i}-\boldsymbol{\theta}_2) - \hat{\mathbf{D}}_1^{-1/2}\hat{\boldsymbol{\mu}}_{2,j}||^2\mathbf{U}_{1i} + o_p(n^{-1}).
\end{aligned}$$

Thus, taking the same procedure as Lemma 5, we obtain that

$$\begin{aligned}
R_n =&\frac{1}{n_1n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\frac{1}{r_{1i}r_{2j}}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)^T\mathbf{D}_1^{-1/2}[\mathbf{I}_p - \mathbf{U}_{1i}\mathbf{U}_{1i}^T][\mathbf{I}_p - \mathbf{U}_{2j}\mathbf{U}_{2j}^T]\mathbf{D}_2^{-1/2}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2) + Z_n\\
&+\frac{1}{n_1n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\frac{1}{r_{2j}}\mathbf{U}_{1i}^T[\mathbf{I}_p - \mathbf{U}_{2j}\mathbf{U}_{2j}^T]\mathbf{D}_2^{-1/2}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)\\
&+\frac{1}{n_1n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\frac{1}{r_{1i}}\mathbf{U}_{2j}^T[\mathbf{I}_p - \mathbf{U}_{1i}\mathbf{U}_{1i}^T]\mathbf{D}_1^{-1/2}(\boldsymbol{\theta}_2-\boldsymbol{\theta}_1) + o_p(n^{-2}).
\end{aligned}$$

By the same arguments as Lemma 5, we can show that

$$\begin{aligned}
&\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\frac{1}{r_{1i}r_{2j}}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)^T\mathbf{D}_1^{-1/2}[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T][\mathbf{I}_p-\mathbf{U}_{2j}\mathbf{U}_{2j}^T]\mathbf{D}_2^{-1/2}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)\\
&\qquad\qquad= c_1c_2(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)^T\mathbf{D}_1^{-1/2}\mathbf{D}_2^{-1/2}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)+o_p(\sigma_n),\\
&\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\frac{1}{r_{2j}}\mathbf{U}_{1i}^T[\mathbf{I}_p-\mathbf{U}_{2j}\mathbf{U}_{2j}^T]\mathbf{D}_2^{-1/2}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)\\
&\qquad\qquad=\frac{1}{n_1}\sum_{i=1}^{n_1}c_2\mathbf{u}_{1i}^T\boldsymbol{\Sigma}_1^{1/2}\mathbf{D}_1^{-1/2}\mathbf{D}_2^{-1/2}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)+o_p(\sigma_n),\\
&\frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\frac{1}{r_{1i}}\mathbf{U}_{2j}^T[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T]\mathbf{D}_1^{-1/2}(\boldsymbol{\theta}_2-\boldsymbol{\theta}_1)\\
&\qquad\qquad=\frac{1}{n_2}\sum_{i=1}^{n_2}c_1\mathbf{u}_{2i}^T\boldsymbol{\Sigma}_2^{1/2}\mathbf{D}_2^{-1/2}\mathbf{D}_1^{-1/2}(\boldsymbol{\theta}_2-\boldsymbol{\theta}_1)+o_p(\sigma_n).
\end{aligned}$$

Thus, we can rewrite R_n as follows

$$\begin{aligned}
R_n =&Z_n + c_1c_2(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)^T\mathbf{D}_1^{-1/2}\mathbf{D}_2^{-1/2}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)+\frac{1}{n_1}\sum_{i=1}^{n_1}c_2\mathbf{u}_{1i}^T\boldsymbol{\Sigma}_1^{1/2}\mathbf{D}_1^{-1/2}\mathbf{D}_2^{-1/2}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)\\
&+\frac{1}{n_2}\sum_{i=1}^{n_2}c_1\mathbf{u}_{2i}^T\boldsymbol{\Sigma}_2^{1/2}\mathbf{D}_2^{-1/2}\mathbf{D}_1^{-1/2}(\boldsymbol{\theta}_2-\boldsymbol{\theta}_1)+o_p(\sigma_n),
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{var}(R_n) =&\bigg(\sigma_n^2+\frac{c_2^2}{n_1p}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)^T\mathbf{D}_2^{-1/2}\mathbf{R}_1\mathbf{D}_2^{-1/2}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)\\
&+\frac{c_1^2}{n_2p}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)^T\mathbf{D}_1^{-1/2}\mathbf{R}_2\mathbf{D}_1^{-1/2}(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)\bigg)(1+o(1)).
\end{aligned}$$

Next, taking the same procedure as in the proof of Theorem 1, we can prove the assertion.
□

References

Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D., Levine, A. J. (1999), "Broad Patterns of Gene Expression Revealed by Clustering Analysis of Tumor and Normal Colon Tissues Probed by Oligonucleotide Arrays," *Proceedings of the National Academy of Sciences of U.S.A.*, 96, 6745–6750.

Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis,* Hoboken, NJ: Wiley, .

Bai, Z. and Saranadasa, H. (1996), "Effect of High Dimension: by an Example of a Two Sample Problem," *Statistica Sinica,* 6, 311–329.

Bickel, P. and Levina, E. (2008), "Regularized Estimation of Large Covariance Matrices," *The Annals of Statistics*, 36, 199–227.

Biswas, M. and Ghosh. A. K. (2014), "A Nonparametric Two-Sample Test Applicable to High Dimensional Data," *Journal of Multivariate Analysis*, 123, 160–171.

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: Series 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, 810–815.

Dempster, A. P. (1958), "A High Dimensional Two Sample Significance Test," *The Annals of Mathematical Statistics*, 29, 995–1010.

Dudoit, S., Fridlyand, J. and Speed, T. P. (2002), "Comparison of Discrimination Methods for the Classification of Tumors using Gene Expression Data," *Journal of the American statistical association*, 97, 77–87.

Feng, L., Zou, C., Wang, Z. and Chen, B. (2013), "Rank-based Score Tests for High-Dimensional Regression Coefficients," *Electronic Journal of Statistics*, 7, 2131–2149.

Feng, L., Zou, C., Wang, Z. and Zhu, L. X. (2014), "Two Sample Behrens-Fisher Problem for High-Dimensional Data," *Statistica Sinica*, To appear.

Goeman, J., Van De Geer, S. A. and Houwelingen, V. (2006), "Testing Against a High-Dimensional Alternative," *Journal of the Royal Statistical Society, Series B*, 68, 477–493.

Gregory, K. B., Carroll, R. J., Baladandayuthapani, V. and Lahiri, S. N. (2014), "A Two-Sample Test For Equality of Means in High Dimension,", *Journal of the American Statistical Association*, To appear.

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

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.

Hettmansperger, T. P. and Oja, H. (1994), "Affine Invariant Multivariate Multisample Sign Tests," *Journal of the Royal Statistical Society, Series B,* 56, 235–249.

Hettmansperger, T. P. and Randles, R. H. (2002), "A Practical Affine Equivariant Multivariate Median," *Biometrika*, 89, 851–860.

Kosorok, M. and Ma, S. (2007), "Marginal Asymptotics for the 'Large p, Small n' Paradigm: with Applications to Microarray Data," *The Annals of Statistics,* 35, 1456-1486.

Ledoit, O. and Wolf, M. (2002), "Some Hypothesis Tests for the Covariance Matrix when the Dimension is Large Compared to the Sample Size," *The Annals of Statistics,* 30, 1081–1102.

Möttönen J. and Oja, H. (1995), "Multivariate Spatial Sign and Rank Methods," *Journal of Nonparametric Statistics,* 5 201–213.

Nettleton, D., Recknor, J. and Reecy, J. M. (2008), "Identification of Differentially Expressed Gene Categories in Microarray Studies using Nonparametric Multivariate Analysis," *Bioinformatics*, 24, 192–201.

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

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.

Peters, D. and Randles, R. H. (1990), "A Bivariate Signed Rank Test for the Two-Sample Location Problem," *Journal of the Royal Statistical Society, B*, 53, 493–504.

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

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.

Randles, R. H. (2000), "A Simpler, Affine Invariant, Multivariate, Distribution-Free Sign Test," *Journal of the American Statistical Association*, 95, 1263–1268.

Schott, J. R. (2005), "Testing for Complete Independence in High Dimensions," *Biometrika,* 92, 951–956.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics,* New York: John Wiley & Sons.

Srivastava, M. S. (2005), "Some Tests Concerning the Covariance Matrix in High Dimensional Data," *Journal of Japan Statistical Society*, 35, 251–272.

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.

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.

Srivastava, M. S., Katayama, S. and Kano, Y. (2013), "A Two Sample Test in High Dimensional Data," *Journal of Multivariate Analysis*, 114, 349–358.

Srivastava, M. S., Kollo, T. and von Rosen, D. (2011), "Some Tests for the Covariance Matrix with Fewer Observations Than the Dimension under Non-normality," *Journal of Multivariate Analysis*, 102, 1090–1103.

Srivastava, M. S., Yanagihara, H. and Kubokawa, T. (2014), "Tests for Covariances matrices in High Dimension with Less Sample Size," *Journal of Multivariate Analysis*, 130, 289–309.

Wang, L., Peng, B. and Li, R. (2013). "A High-Dimensional Nonparametric Multivariate Test for Mean Vector," Technical report, the Pennsylvania State University.

Zhong, P.-S. and Chen, S. X. (2011). "Tests for High Dimensional Regression Coefficients with Factorial Designs," *Journal of the American Statistical Association*, 106, 260–274.

Zhong, P., Chen, S. X. and Xu, M. (2013), "Tests Alternative to Higher Criticism for High Dimensional Means under Sparsity and Column-Wise Dependence," *The Annals of Statistics*, 41, 2703–3110.

Zou, C., Peng, L, Feng, L. and Wang, Z. (2014), "Multivariate-Sign-Based High-Dimensional Tests for Sphericity," *Biometrika*, 101, 229–236.