Acta Mathematicae Applicatae Sinica, English Series
Vol. 38, No. 1 (2022) 78–86
https://doi.org/10.1007/s10255-022-1070-z
http://www.ApplMath.com.cn & www.SpringerLink.com

Acta Mathematicae Applicatae Sinica, English Series

High-dimensional Tests for Mean Vector: Approaches without Estimating the Mean Vector Directly

Bo CHEN†, Hai-meng WANG

School of Mathematics and Information Technology, Jiangsu Second Normal University, Nanjing 211200, China
(†E-mail address: pbchb@163.com)

Abstract Several tests for multivariate mean vector have been proposed in the recent literature. Generally, these tests are directly concerned with the mean vector of a high-dimensional distribution. The paper presents two new test procedures for testing mean vector in large dimension and small samples. We do not focus on the mean vector directly, which is a different framework from the existing choices. The first test procedure is based on the asymptotic distribution of the test statistic, where the dimension increases with the sample size. The second test procedure is based on the permutation distribution of the test statistic, where the sample size is fixed and the dimension grows to infinity. Simulations are carried out to examine the finite-sample performance of the tests and to compare them with some popular nonparametric tests available in the literature.

Keywords asymptotic distribution; high-dimensional data; permutation test; U-statistic; testing mean vector
2000 MR Subject Classification 62H15; 62G32

1 Introduction

Suppose that $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are n independent realizations of a p-dimensional random vector $\mathbf{x}$, which is symmetric about $\boldsymbol{\mu} \in R^p$, i.e., $\mathbf{x} - \boldsymbol{\mu} \overset{\mathrm{d}}{=} \boldsymbol{\mu} - \mathbf{x}$. In multivariate analysis, a canonical testing problem is that of testing the mean vector. With the rapid development of information technologies, analyses involving a large number of variables p are becoming more prevalent in statistical applications. High dimensionality poses significant challenge to hypothesis testing. The main challenge of high-dimensional data is that the dimension p is much larger than the sample sizes n. When this happens, many traditional statistical methods and theories may not necessarily work since they assume that p keeps unchanged as n increases. In this article, we shall test in large dimension and small samples the hypothesis

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

To cope with the high dimensionality, several alternative approaches have been suggested in the recent literature. These methods include, but are not limited to [2, 5–8, 12, 15, 16] and [18]. These tests are very useful. Generally, these tests are directly concerned with the mean vector of a high-dimensional distribution. Under a different framework where we do not focus on the mean vector directly, we propose two novel test procedures for testing mean vector. Our proposals are based on an energy statistic and can be conveniently used in high dimension low sample size situations.

There are other tests for high dimensional data based on energy statistic; see [14] and [19] in the context of testing dependence. In this paper, we exploit the same principle in the context of testing mean vector. As pointed out by [13], energy statistics are extremely useful and are

Manuscript received December 5, 2019. Accepted on October 26, 2021.
†Corresponding author.

typically more general and often more powerful against general alternatives than classical (non-energy type) statistics. The simulation results also demonstrate that our proposals are quite competitive in comparison with some popular nonparametric tests available in the literature.

The rest of this paper is organized as follows. The two new test procedures are presented in Section 2. Specifically, the first test procedure is introduced in Section 2.1 where we derive the asymptotic distribution of the test statistic when the dimension increases with the sample size. The second test procedure is based on the permutation distribution of the test statistic and is investigated in Section 2.2 where we assume that the sample size is fixed and the dimension grows to infinity. Simulation studies are carried out in Section 3 to investigate the numerical performance of the test procedures. The article concludes with a short discussion in Section 4. All the technical proofs are gatheblack in the Appendix.

2 Main Results

Let $\mathbf{x}$ be R^p-valued random vector with finite expectation. From [4] and [13], we know that H_0 holds if and only if $\mathfrak{M}(\mathbf{x}) = 0$, where $\mathfrak{M}(\mathbf{x}) = E||\mathbf{x}_1+\mathbf{x}_2|| - E||\mathbf{x}_1-\mathbf{x}_2||$ is nonnegative and $||\mathbf{x}||$ is the Eucledean norm of $\mathbf{x}$. According to [13], one can refer to $\mathfrak{M}(\mathbf{x})$ as an energy distance. Based on the energy distance, it is natural to consider the energy statistic

$$T_n = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} (||\mathbf{x}_i + \mathbf{x}_j|| - ||\mathbf{x}_i - \mathbf{x}_j||). \tag{2.1}$$

Based on T_n, we give two new test procedures for testing mean vector in high dimension low sample size situations. One focuses on the asymptotic and high-dimensional distribution of T_n; the other focuses on the high-dimensional permutation distribution of T_n.

2.1 The Test Based on the Asymptotic and High-dimensional Distribution

In this section, we present a test via investigating the asymptotic distribution of T_n when the dimension increases with the sample size. To this end, we first assume that the random vector $\mathbf{x}_i$ is generated from the symmetric independent component models ([11]). That is,

$$\mathbf{x}_i = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{z}_i, \tag{2.2}$$

where $\boldsymbol{\Gamma}$ is a full rank $p\times p$ matrix; $\mathbf{z}_i = (Z_{i1},\cdots,Z_{ip})^{\mathrm{T}}$ has independent components Z_{ij}s and Z_{ij} is symmetric about zero. We denote $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathrm{T}}$ for easy future reference. To implement the analysis, we further assume that each Z_{ij} has finite 8th moment. A similar condition was used by [7, 17] and [20].

Additional, we need the following conditions to regulate for the "large p, small n" is,

(C1) $p\to\infty$, as $n\to\infty$;

(C2) $\mathrm{tr}(\boldsymbol{\Sigma}^4) = o\{\mathrm{tr}^2(\boldsymbol{\Sigma}^2)\}$;

(C3) $\frac{n\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^2)}}{\mathrm{tr}(\boldsymbol{\Sigma})} = o(1)$.

We here remark that Conditions (C1) and (C2) are rather mild and are similar to Assumption 2 in [7] and condition (2.8) in [20]. Condition (C2) also implies that $\mathrm{tr}(\boldsymbol{\Sigma}^2) = o\{\mathrm{tr}^2(\boldsymbol{\Sigma})\}$ and $\mathrm{tr}(\boldsymbol{\Sigma}^3) = o\{\mathrm{tr}(\boldsymbol{\Sigma})\mathrm{tr}(\boldsymbol{\Sigma}^2)\}$. These facts are need to show (5.1) in Appendix, although we omit the technical details. Since there would be a little bias term when dealing with $\big(\frac{||\mathbf{x}_i-\mathbf{x}_j||}{\sqrt{2\mathrm{tr}(\boldsymbol{\Sigma})}}+1\big)^{-1}$, we need Condition (C3) to eliminate the bias term in (5.2). To appreciate condition (C3), consider that all the eigenvalues of $\boldsymbol{\Sigma}$ are bounded. Thus, the condition becomes $n^2 = o(p)$, which is true when p gets larger.

Theorem 1. *Let the test statistic T_n be defined as in (2.1). Assume that Model (2.2) and Conditions (C1)–(C3). Then under H_0, as $n, p \to \infty$,*

$$\frac{n\sqrt{tr(\mathbf{\Sigma})}}{2\sqrt{tr(\mathbf{\Sigma}^2)}}T_n \xrightarrow{d} N(0,1).$$

However, $\mathrm{tr}(\mathbf{\Sigma})$ and $\mathrm{tr}(\mathbf{\Sigma}^2)$ are appeablack in the asymptotic variance, so we need to estimate them. From (2.3) of [7], we can estimate $\mathrm{tr}(\mathbf{\Sigma})$ by $\widehat{\mathrm{tr}}(\mathbf{\Sigma}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_i - \frac{1}{n(n-1)}\sum_{i \neq j} \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j$. To estimate $\mathrm{tr}(\mathbf{\Sigma}^2)$, define

$$A_{1n} = \frac{1}{n(n-1)}\sum_{\substack{i,j=1\\ i\neq j}}^{n} (\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j)^2, \qquad A_{2n} = \frac{1}{n(n-1)(n-2)}\sum_{\substack{i,j,k=1\\ i\neq j, j\neq k, k\neq i}}^{n} \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{x}_k,$$

and

$$A_{3n} = \frac{1}{n(n-1)(n-2)(n-3)}\sum_{\substack{i,j,k,l=1\\ i\neq j, j\neq k, k\neq l,}}^{n} \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{x}_l.$$

Like [7], we propose the following ratio consistent estimator of $\mathrm{tr}(\mathbf{\Sigma}^2)$,

$$\widehat{\mathrm{tr}}(\mathbf{\Sigma}^2) = A_{1n} - 2A_{2n} + A_{3n}.$$

By combining Slutsky's theorem, the first test procedure rejects H_0 at a significant level α if

$$nT_n \geq \frac{2\sqrt{\widehat{\mathrm{tr}}(\mathbf{\Sigma}^2)}}{\sqrt{\widehat{\mathrm{tr}}(\mathbf{\Sigma})}} z_\alpha. \tag{2.3}$$

2.2 The Test Based on the High-dimensional Permutation Distribution

Note that

$$T_n = \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}(||\mathbf{x}_i - \mathbf{x}_j^*|| - ||\mathbf{x}_i + \mathbf{x}_j^*||), \tag{2.4}$$

where $\mathbf{x}_i^* = -\mathbf{x}_i$ for $i = 1, \cdots, n$. Under H_0, $\mathbf{x}_1, \cdots, \mathbf{x}_n$ and $\mathbf{x}_1^*, \cdots, \mathbf{x}_n^*$ have the same distribution, while under H_1 they differ in their locations. Consider the pooled sample $\{\mathbf{z}_1, \cdots, \mathbf{z}_n, \mathbf{z}_{n+1}, \cdots, \mathbf{z}_{2n}\}$ with $\mathbf{z}_i = \mathbf{x}_i$ and $\mathbf{z}_{n+i} = \mathbf{x}_i^*$. We first give the descriptions of this procedure.

(S1) We generate a permutation from the set $\{1, \cdots, n, n+1, \cdots, 2n\}$, namely, $\{k_1, \cdots, k_{2n}\}$. Let the set $\mathcal{I} = \{k_j : n < k_j, 1 \leq j \leq n\}$. For any $i \in \mathcal{I}$, we exchange $\mathbf{z}_i$ and $\mathbf{z}_{i-n}$. We then obtain the bootstrap sample $\{\mathbf{z}_1^b, \cdots, \mathbf{z}_n^b, \mathbf{z}_{n+1}^b, \cdots, \mathbf{z}_{2n}^b\}$.

(S2) Compute $T_n^{(1)}$ by using the bootstrap sample with $\mathbf{x}_i = \mathbf{z}_i^b$ and $\mathbf{x}_i^* = \mathbf{z}_{i+n}^b$ for $i = 1, \cdots, n$.

(S3) Repeat (S1) and (S2) for B times, and obtain $T_n^{(l)}$, $l = 1, \cdots, B$. And then the bootstrap estimate of $P(T_n \leq t)$ is $\frac{1}{B}\sum_{l=1}^{B} I\{T_n^{(l)} \leq t\}$. Based on the bootstrap estimate, we can compute the corresponding p-value.

To enhance readability, we here restate that $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ are the mean vector and the dispersion matrix. In what follows, we investigate the limiting behavior of the permutation test when the sample size is fixed and the dimension grows to infinity. In order to carry out our investigation, we make the following assumptions, which are different from Section 2.1.

(D1) There exist $\sigma^2 > 0$ and v such that (i) $\mathrm{tr}(\mathbf{\Sigma})/ \to \sigma^2$ and (ii) $||\boldsymbol{\mu}||^2/p \to v^2$ as $p \to \infty$.

(D2) Fourth moments of the components of $\mathbf{x}$ are uniformly bounded.

(D3) Let $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ be three independent copies of $\mathbf{x}$ and let $\mathbf{y}_3 = -\mathbf{x}_3$. Let $\mathbf{x}_i = (X_{i1}, \cdots, X_{ip})$ and $\mathbf{y}_i = (Y_{i1}, \cdots, Y_{ip})$. For $(U_i, V_i) = (X_{1i}, X_{2i})$ or (X_{1i}, Y_{3i}), $\sum_{i \neq j} \left| \mathrm{Corr}\{(U_i - V_i)^2, (U_j - V_j)^2\} \right|$ is of the order $o(p^2)$.

We here remark that the conditions (D1)–(D3) are common assumptions under the framework where the sample size is fixed and the dimension grows to infinity. See [10] and [3] for more details. In fact, we need the fourth moment condition (D2) and the weak dependence among the component variables (D3) to have the weak law of large numbers (WLLN) for the sequence of dependent and non-identically distributed random variables. Furthermore, note that if the components of $\mathbf{x}$ are i.i.d. with finite second moments, (D1) and (D3) get automatically satisfied. Recall that [1] derived the WLLN for mixingales. Therefore, under (D1), (D2) and (D3), we can show that

(a) $||\mathbf{x}_i - \mathbf{x}_j||/\sqrt{p} \to \sqrt{2}\sigma$ for $1 \le i < j \le n$;

(b) $||\mathbf{x}_i + \mathbf{x}_j||/\sqrt{p} \to \sqrt{2\sigma^2 + 4v^2}$ for $1 \le i < j \le n$.

Both (a) and (b) are very useful to show the consistency result on the permutation test when the sample size is fixed and the dimension grows to infinity. Define $C(n, m)$ to be the number of m-combinations from a set of n elements.

Theorem 2. *Suppose that we have n independent observations, which satisfy (D1), (D2) and (D3). Also assume that $v^2 > 0$. Then, unless n is very small (i.e., $C(2n, n) \le 2/\alpha$), the power of the proposed permutation test with level α converges to 1 as p tends to infinity.*

Due to the proof of Theorem 2, it is clear that for fixed n, the limiting p-value of the permutation test, i.e., the limiting value (as $p \to \infty$) of $P(\frac{1}{\sqrt{p}} T_n \ge \gamma_0)$ under the permutation distribution is $2/\binom{2n}{n}$, where $\gamma_0 = \sqrt{2\sigma^2 + 4v^2} - \sqrt{2}\sigma$ which is positive unless $v^2 = 0$. In particular, for a test of level 0.05 (or 0.01), it is enough to have 4 (or 5) observations from each class for the convergence of the power to unity.

3 Numerical Study

3.1 Simulation Study

In this section, we conduct simulations to demonstrate the performance of the proposed test procedures. We refer to the test based on the asymptotic and high-dimensional distribution and the test based on the high-dimensional permutation distribution as the AH test and the PH test, respectively. For comparison purposes, we also conducted the test proposed by [7] and the scalar transformation invariant test proposed by [12]. We refer to them as the CQ test and the PA test, respectively. The proposed test, the CQ test and the PA test are sum-of-squares-based tests and designed for dense alternatives. For the sake of fairness, we exclude the supremum-based tests developed in [5], [6] [8] and [18]. The attained significance level to the nominal value $\alpha = 0.05$ and the power are investigated in finite samples by simulation. The number of experiments is 500. The number of permutations is 500.

Consider that $\mathbf{x}$ is taken to be the multivariate t-distribution with v degrees of freedom. We take $v = 3$ and $v = \infty$, respectively. The $v = \infty$ corresponds to the multivariate normal distribution. The distribution with $v = 3$ is heavy-tailed. The mean vector of $\mathbf{x}$ and the scatter matrix of $\mathbf{x}$ are taken to be $\boldsymbol{\mu} = \delta \mathbf{1}_p$ and $\mathbf{\Sigma} = (\rho^{|i-j|})$, where $\mathbf{1}_p$ denotes the p-dimensional vector with all elements unity. For comparison, we take $\rho = 0$ and $\rho = 0.5$, respectively. For normally distributed data, we only consider the case of $\rho = 0.5$ to conserve space because all the tests perform well and the simulation results are not significantly improved in this case. The $\delta = 0$ corresponds to the null hypothesis. To assess the power of the tests, we take $\delta = 0.1, 0.2$ when $v = \infty$ and $\delta = 0.15, 0.2$ when $v = 3$. To demonstrate the effectiveness of the proposed permutation test, we consider one relatively small sample size ($n = 20$). We compute powers of these tests for nine different values of p, i.e., $p = 25, 50, 75, 100, 125, 150, 175, 200, 225$.

Table 1. Empirical size and power of the tests for different (δ, p) when $v = \infty$ and $\rho = 0.5$

	p								
	25	50	75	100	125	150	175	200	225
$\delta = 0$									
PH	0.04	0.05	0.06	0.05	0.06	0.05	0.05	0.06	0.06
AH	0.05	0.05	0.06	0.05	0.06	0.06	0.05	0.06	0.05
CQ	0.05	0.06	0.06	0.05	0.06	0.07	0.06	0.06	0.05
PA	0.06	0.06	0.07	0.06	0.07	0.07	0.06	0.07	0.06
$\delta = 0.1$									
PH	0.16	0.20	0.24	0.30	0.32	0.37	0.41	0.45	0.48
AH	0.16	0.22	0.24	0.27	0.33	0.38	0.42	0.44	0.47
CQ	0.17	0.21	0.24	0.28	0.33	0.39	0.42	0.45	0.48
PA	0.17	0.23	0.28	0.31	0.35	0.40	0.45	0.48	0.50
$\delta = 0.2$									
PH	0.53	0.75	0.83	0.94	0.95	0.97	0.98	1.00	1.00
AH	0.56	0.71	0.81	0.93	0.96	0.98	1.00	1.00	1.00
CQ	0.54	0.70	0.82	0.93	0.96	0.98	1.00	1.00	1.00
PA	0.55	0.73	0.85	0.94	0.96	0.98	1.00	1.00	1.00

Table 2. Empirical size and power of the tests for different (δ, p) when $v = 3$ and $\rho = 0$

	p								
	25	50	75	100	125	150	175	200	225
$\delta = 0$									
PH	0.06	0.06	0.04	0.07	0.06	0.05	0.07	0.06	0.04
AH	0.05	0.04	0.04	0.05	0.04	0.05	0.04	0.04	0.05
CQ	0.07	0.05	0.06	0.05	0.05	0.06	0.06	0.06	0.05
PA	0.04	0.03	0.02	0.02	0.03	0.02	0.03	0.03	0.02
$\delta = 0.15$									
PH	0.30	0.41	0.50	0.59	0.72	0.73	0.79	0.82	0.84
AH	0.22	0.35	0.40	0.52	0.58	0.67	0.72	0.75	0.78
CQ	0.23	0.30	0.36	0.42	0.47	0.53	0.54	0.61	0.64
PA	0.17	0.22	0.24	0.24	0.27	0.29	0.33	0.34	0.38
$\delta = 0.2$									
PH	0.52	0.71	0.80	0.89	0.95	0.96	0.98	1.00	1.00
AH	0.42	0.64	0.73	0.85	0.89	0.94	0.94	0.96	0.98
CQ	0.37	0.52	0.60	0.73	0.76	0.83	0.84	0.86	0.87
PA	0.29	0.30	0.31	0.33	0.34	0.37	0.37	0.39	0.43

A well-behaved test should have an empirical size around 0.05. From Tables 1–3, we see that most of the methods perform quite well except that the PA tends to be smaller than the nominal 5% level when $v = 3$. This may imply that the PA test is adversely affected by heavy-tailed distribution. For the sake of comparison, we further report the power of the tests. Tables 1–3 also report the empirical powers of the tests for different (δ, p, ρ, v). It is easy to observe that when $v = \infty$, all the tests have good performance and the differences among them are very small. However, it is interesting to see that when $v = 3$, the proposed test procedures PH and AH significantly outperform the selected tests CQ and PA. That is, the proposed test performs better when the error is heavy-tailed. This may agree with [13]'s assertion that energy statistics are are typically more general and often more powerful against general alternatives than classical (non-energy type) statistics. In particular, we further observe that the PH is better than the AH in such a case. Consequently, we tend to suggest using the PH in practice.

Table 3. Empirical size and power of the tests for different (δ, p) when $v = 3$ and $\rho = 0.5$

	p								
	25	50	75	100	125	150	175	200	225
$\delta = 0$									
PH	0.06	0.06	0.05	0.06	0.07	0.06	0.07	0.06	0.06
AH	0.04	0.04	0.05	0.04	0.06	0.04	0.04	0.04	0.05
CQ	0.07	0.06	0.06	0.06	0.07	0.06	0.05	0.05	0.05
PA	0.04	0.03	0.03	0.04	0.03	0.02	0.02	0.03	0.03
$\delta = 0.15$									
PH	0.16	0.19	0.22	0.27	0.36	0.39	0.43	0.46	0.48
AH	0.11	0.12	0.17	0.23	0.26	0.30	0.32	0.35	0.39
CQ	0.15	0.14	0.18	0.22	0.24	0.24	0.25	0.31	0.32
PA	0.15	0.16	0.16	0.18	0.19	0.21	0.23	0.26	0.28
$\delta = 0.2$									
PH	0.25	0.37	0.46	0.56	0.71	0.76	0.82	0.85	0.88
AH	0.18	0.28	0.38	0.51	0.59	0.62	0.64	0.74	0.76
CQ	0.20	0.26	0.31	0.41	0.46	0.50	0.51	0.57	0.61
PA	0.21	0.23	0.25	0.25	0.28	0.29	0.33	0.36	0.41

3.2 A Real Example

In this section,we consider amicroarray data set consisting of a large p and small n. The data set is a leukemia data set, which is available at http://www.broad.mit.edu/cgi-bin/cancer/datasets.cgi. There are $p = 7129$ genes and samples generated from the acute lymphocytic leukemia group consist of $n = 47$. As an illustration of our approach, we ask if there is evidence of mean zero. The p-value of energy statistic test is 1.6×10^{-8}. There is evidence that the data distribution has nonzero mean.

4 Discussion

In this paper, we present some high-dimensional tests for mean vector based on an energy distance, which leads to an energy statistic. Based on the energy statistic, we propose two new test procedures for testing the mean vector. One is based on the asymptotic and high-dimensional distribution of the statistic; the other is based on the high-dimensional permutation distribution of the statistic. Monte Carlo simulations demonstrate that the new test procedures are quite competitive.

5 Appendix: Proof of Theorems

This appendix gives the proofs of Theorems 1 and 2.

Appendix A: Proof of Theorem 1

For any $i \neq j$, we reformulate $||\mathbf{x}_i - \mathbf{x}_j|| - \sqrt{2\text{tr}(\boldsymbol{\Sigma})}$ as follows,

$$||\mathbf{x}_i - \mathbf{x}_j|| - \sqrt{2\text{tr}(\boldsymbol{\Sigma})} = \frac{\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{\sqrt{2\text{tr}(\boldsymbol{\Sigma})}} - \sqrt{2\text{tr}(\boldsymbol{\Sigma})}}{\frac{||\mathbf{x}_i - \mathbf{x}_j||}{\sqrt{2\text{tr}(\boldsymbol{\Sigma})}} + 1}.$$

Similar to Proposition 5.1 of [7], we can show that under Model (2.2), $\text{Var}\,(||\mathbf{x}_i - \mathbf{x}_j||^2) = 2\text{Var}(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_i) + 4\text{Var}(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j) = O\{\text{tr}(\boldsymbol{\Sigma}^2)\}$. Noting that $E\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{2\text{tr}(\boldsymbol{\Sigma})} = 1$, we have $\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{2\text{tr}(\boldsymbol{\Sigma})} - 1 = O_p\{\sqrt{\frac{\text{tr}(\boldsymbol{\Sigma}^2)}{\text{tr}^2(\boldsymbol{\Sigma})}}\}$, where we use the condition (C2). This also implies $\frac{||\mathbf{x}_i - \mathbf{x}_j||}{\sqrt{2\text{tr}(\boldsymbol{\Sigma})}}$ converges to one in probability. Therefore, by Taylor's Theorem, we have $\big(\frac{||\mathbf{x}_i - \mathbf{x}_j||}{\sqrt{2\text{tr}(\boldsymbol{\Sigma})}} + 1\big)^{-1} = \frac{1}{2} + O_p\{\sqrt{\frac{\text{tr}(\boldsymbol{\Sigma}^2)}{\text{tr}^2(\boldsymbol{\Sigma})}}\}$ Consequently, it follows that

$$\begin{aligned}||\mathbf{x}_i - \mathbf{x}_j|| - \sqrt{2\text{tr}(\boldsymbol{\Sigma})} =& \frac{1}{2\sqrt{2\text{tr}(\boldsymbol{\Sigma})}}\{||\mathbf{x}_i - \mathbf{x}_j||^2 - 2\text{tr}(\boldsymbol{\Sigma})\} \\ &+ \frac{1}{\sqrt{2\text{tr}(\boldsymbol{\Sigma})}}\{||\mathbf{x}_i - \mathbf{x}_j||^2 - 2\text{tr}(\boldsymbol{\Sigma})\}O_p\Big\{\sqrt{\frac{\text{tr}(\boldsymbol{\Sigma}^2)}{\text{tr}^2(\boldsymbol{\Sigma})}}\Big\} \\ =& \frac{1}{2\sqrt{2\text{tr}(\boldsymbol{\Sigma})}}\{||\mathbf{x}_i - \mathbf{x}_j||^2 - 2\text{tr}(\boldsymbol{\Sigma})\} + O_p\Big\{\frac{\text{tr}(\boldsymbol{\Sigma}^2)}{\text{tr}^{3/2}(\boldsymbol{\Sigma})}\Big\}.\end{aligned}$$

Similarly, we have that under H_0,

$$||\mathbf{x}_i + \mathbf{x}_j|| - \sqrt{2\text{tr}(\boldsymbol{\Sigma})} = \frac{1}{2\sqrt{2\text{tr}(\boldsymbol{\Sigma})}}\{||\mathbf{x}_i + \mathbf{x}_j||^2 - 2\text{tr}(\boldsymbol{\Sigma})\} + O_p\Big\{\frac{\text{tr}(\boldsymbol{\Sigma}^2)}{\text{tr}^{3/2}(\boldsymbol{\Sigma})}\Big\}.$$

Therefore,

$$\begin{aligned}T_n =& \frac{2}{n(n-1)}\sum_{1\le i<j\le n}\big[\{||\mathbf{x}_i + \mathbf{x}_j|| - \sqrt{2\text{tr}(\boldsymbol{\Sigma})}\} - \{||\mathbf{x}_i - \mathbf{x}_j|| - \sqrt{2\text{tr}(\boldsymbol{\Sigma})}\}\big] \\ =& \frac{4}{n(n-1)\sqrt{2\text{tr}(\boldsymbol{\Sigma})}}\sum_{1\le i<j\le n}\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j + O_p\Big\{\frac{\text{tr}(\boldsymbol{\Sigma}^2)}{\text{tr}^{3/2}(\boldsymbol{\Sigma})}\Big\}.\end{aligned}$$

Due to martingale central limit theorem[9] and exactly similar to the proof of (5.12) in [20], it follows that under Model (2.2) and conditions (C1)–(C3), as $n, p \to \infty$,

$$\frac{1}{\sqrt{n(n-1)\text{tr}(\boldsymbol{\Sigma}^2)/2}}\sum_{1\le i<j\le n}\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j \xrightarrow{d} N(0,1). \tag{5.1}$$

Recall that

$$\frac{\sqrt{n(n-1)\text{tr}(\boldsymbol{\Sigma})}}{2\sqrt{\text{tr}(\boldsymbol{\Sigma}^2)}}T_n = \frac{\sqrt{2}}{\sqrt{n(n-1)\text{tr}(\boldsymbol{\Sigma}^2)}}\sum_{1\le i<j\le n}\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j + O_p\Big\{\frac{n\sqrt{\text{tr}(\boldsymbol{\Sigma}^2)}}{\text{tr}(\boldsymbol{\Sigma})}\Big\}. \tag{5.2}$$

Therefore, due to Slutsky's theorem and condition (C3), we get the desired result. □

Appendix B: Proof of Theorem 2

If (D1)–(D3) hold, using the results (a) and (b) stated in Section 2.2, for fixed n and $p \to \infty$, $T_n/\sqrt{p} = \gamma_{01} - \gamma_{02} \stackrel{\text{def}}{=} \gamma_0$ in probability, where $\gamma_{01} = \sqrt{2\sigma^2 + 4v^2}$ and $\gamma_{02} = \sqrt{2}\sigma$. Next, let us consider the permutation distribution of T_n.

Let $\sharp(\cdot)$ be the counting measure. For the pooled sample $\{\mathbf{z}_1, \cdots, \mathbf{z}_n, \mathbf{z}_{n+1}, \cdots, \mathbf{z}_{2n}\}$ with $\mathbf{z}_i = \mathbf{x}_i$ and $\mathbf{z}_{n+i} = \mathbf{x}_i^*$, suppose that $s = \sharp(\mathcal{I})$ observations ($s = 1, \cdots, n$) from $\{\mathbf{z}_1, \cdots, \mathbf{z}_n\}$ and s observations from $\{\mathbf{z}_{n+1}, \cdots, \mathbf{z}_{2n}\}$ are exchanged according to the step (S1) described in Section 2.2. This gives the bootstrap sample $\{\mathbf{z}_1^b, \cdots, \mathbf{z}_n^b, \mathbf{z}_{n+1}^b, \cdots, \mathbf{z}_{2n}^b\}$. Letting $\mathbf{x}_i = \mathbf{z}_i^b$ and $\mathbf{x}_i^* = \mathbf{z}_{n+i}^b$ for $i = 1, \cdots, n$, we have that $\sum_{1 \le i < j \le n} ||\mathbf{x}_i + \mathbf{x}_j^*|| = \{C(s,2) + C(n-s,2)\}\gamma_{02} + s(n-s)\gamma_{01}$ and $\sum_{1 \le i < j \le n} ||\mathbf{x}_i - \mathbf{x}_j^*|| = \{C(s,2) + C(n-s,2)\}\gamma_{01} + s(n-s)\gamma_{02}$ in probability. Thus, we can show that the value of the test statistic converges to

$$\gamma_s \stackrel{\text{def}}{=} \frac{\{C(s,2) + C(n-s,2)\}\gamma_{01} + s(n-s)\gamma_{02}}{C(n,2)} - \frac{\{C(s,2) + C(n-s,2)\}\gamma_{02} + s(n-s)\gamma_{01}}{C(n,2)}.$$

After simple calculations, we further have

$$\gamma_s = \Big\{1 - \frac{2s(n-s)}{C(n,2)}\Big\}(\gamma_{01} - \gamma_{02}) = \Big\{1 - \frac{2s(n-s)}{C(n,2)}\Big\}\gamma_0.$$

Note that under H_1, $\gamma_0 > 0$ and $1 - \frac{2s(n-s)}{C(n,2)} \le 1$ for all choices of s, where the equality holds if and only if $s = 0$ and $s = n$. That is, $\gamma_s \le \gamma_0$ for all choices of s and the equality holds if and only if $s = 0$ and $s = n$. On the other hand, as $p \to \infty$, the permutation distribution tend to have $n + 1$ mass points $\gamma_0, \gamma_1, \cdots, \gamma_n$ with probabilities $\frac{C(n,0)}{C(2n,n)}, \frac{C(n,1)}{C(2n,n)}, \cdots, \frac{C(n,n)}{C(2n,n)}$ respectively. Therefore, as $p \to \infty$, under the permutation distribution, the test statistic takes the value γ_0 or higher with probability tending to $2/C(2n,n)$. □

Acknowledgments. The authors thank the editor, the AE, and the reviewers for their constructive comments, which have led to a dramatic improvement of the earlier version of this article.

References

[1] Andrews, D. Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory*, 4: 458–467 (1988).

[2] Bai, Z., Sarandasa, H. Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, 6: 311–329 (1996)

[3] Biswas, M., Mukhopadhyay, M., Ghosh, A. A distribution-free two-sample run test applicable to high-dimensional data. *Biometrika*, 101: 913–926 (2014)

[4] Buja, A., Logan, B., Reeds, J. Inequalities and positive definite functions arising from a problem in multi dimensional scaling. *Annals of Statistics*, 22: 406–438 (1994)

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

[6] Chang, J., Zheng, C., Zhou, W. X., Zhou, W. Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *Biometrics*, 73: 1300–1310 (2017)

[7] Chen, S., Qin, Y. A two sample test for high dimensional data with applications to gene-set testing. *Annals of Statistics*, 38: 808–835 (2010)

[8] Chen, S., Li, J., Zhong, P. Two-Sample and ANOVA Tests for High Dimensional Means. *Annals of Statistics*, 47: 1443–1474

[9] Hall, P., Heyde, C. Martingale Limit Theory and Its Application. Academic Press, New York, 1980

[10] Hall, P., Marron, J., Neeman, A. Geometric representation of high dimension, low sample size data. *Journal of the Royal Statistical Society, Series B*, 67: 427–444 (2005)

[11] Ilmonen, P., Paindaveine, D. Semiparametrically efficient inference based on signed ranks in symmetric independent component models. *Annals of Statistics*, 39: 2448–2476 (2011)

[12] Park, J., Nag Ayyala, D. A test for the mean vector in large dimension and small samples. *Journal of Statistical Planning and Inference*, 143: 929–943 (2013)
[13] Székely, G., Rizzo, M. Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference,* 143: 1249–1272
[14] Székely, G., Rizzo, M. The distance correlation t-test of independence in high dimension. *Journal of Multivariate Analysis,* 117: 193–213 (2013)
[15] Srivastava, M. A test for the mean vector with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis,* 100: 518–532 (2009)
[16] Srivastava, M., Du, M. A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis,* 99: 386–402
[17] Xu, K., Hao, X. A nonparametric test for block-diagonal covariance structure in high dimension and small samples. *Journal of Multivariate Analysis*, 173: 551–567 (2019)
[18] Xue, K., Yao, F. Distribution and correlation-free two-sample test of high-dimensional means. *The Annals of Statistics*, 48: 1304–1328
[19] Yao, S., Zhang, X., Shao, X. Testing mutual independence in high dimension via distance covariance. *Journal of Royal Statistical Society, Series B*, 80: 455–480 (2018)
[20] Zhong, P., Chen, S. Tests for high dimensional regression coefficients with factorial designs. *Journal of the American Statistical Association* , 106: 260–274 (2011)