Statistics and Its Interface Volume 12 (2019) 93–106

An adaptive spatial-sign-based test for mean vectors of elliptically distributed high-dimensional data

Bu Zhou, Jia Guo, Jianwei Chen, and Jin-Ting Zhang*

Recently, a nonparametric test for mean vectors of elliptically distributed high-dimensional data has been proposed in the literature. The asymptotic normality of the test statistic under some strong assumptions is established. In practice, however, these strong assumptions may not be satisfied or hardly be checked so that the above test may not perform well in terms of size control. In this paper, we propose an adaptive spatial-sign-based test for mean vectors of elliptically distributed high-dimensional data without imposing strong assumptions. The null distribution of the proposed test statistic is shown to be a chi-squared mixture which is generally skewed. We propose to approximate the null distribution using the well-known Welch–Satterthwaite χ^2-approximation. The resulting approximate distribution is able to adapt to the shape of the underlying null distribution of the proposed test statistic. Simulation studies and three real data examples demonstrate that the proposed test has a better size control than the existing nonparametric test while both tests enjoy about the same powers.

AMS 2000 subject classifications: Primary 62H15; secondary 62G10.
Keywords and phrases: High-dimensional data, Spatial-sign-based test, Welch–Satterthwaite χ^2-approximation.

1. INTRODUCTION

Suppose we have n independently and identically distributed (i.i.d.) observed vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ from a p-variate elliptical distribution (Fang et al. 1990, Anderson 2003) with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, we are interested in testing the following hypotheses:

$$H_0 : \boldsymbol{\mu} = \boldsymbol{0}, \ vs \ H_1 : \boldsymbol{\mu} \neq \boldsymbol{0}. \tag{1.1}$$

Let $\bar{\boldsymbol{x}} = n^{-1} \sum_{i=1}^n \boldsymbol{x}_i$ and $\boldsymbol{S} = (n-1)^{-1} \sum_{i=1}^n (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\top$ denote the usual sample mean vector and sample covariance matrix, respectively, and let $\|\boldsymbol{x}\| = (\boldsymbol{x}^\top \boldsymbol{x})^{1/2}$ denote the usual L^2-norm of $\boldsymbol{x}$. When $p \geq n$, the classical Hotelling T^2 test cannot be used because the sample covariance $\boldsymbol{S}$ is not invertible. One way to solve this problem is to replace the sample covariance matrix $\boldsymbol{S}$ in the Hotelling T^2 test statistic with the identity matrix $\boldsymbol{I}_p$. For the corresponding problem of testing the equality of mean vectors of two high-dimensional samples, the tests proposed by Bai and Saranadasa (1996) and Chen and Qin (2010), denoted as BS and CQ, respectively, are two examples adopting this idea. Although not explicitly discussed by their authors, most high-dimensional two-sample tests can be modified to test the one-sample problem (1.1). For example, to test (1.1), the BS test can be modified as

$$T_{BS} = \frac{n\|\bar{\boldsymbol{x}}\|^2 - \operatorname{tr}(\boldsymbol{S})}{\sqrt{\frac{2n(n-1)}{(n-2)(n+1)}\left\{\operatorname{tr}(\boldsymbol{S}^2) - \frac{\operatorname{tr}^2(\boldsymbol{S})}{n-1}\right\}}}, \tag{1.2}$$

where and throughout this paper, we denote the trace of a matrix $\boldsymbol{A}$ as $\operatorname{tr}(\boldsymbol{A})$, and for some integer s, we write $\{\operatorname{tr}(\boldsymbol{A})\}^s$ as $\operatorname{tr}^s(\boldsymbol{A})$ for notational ease. Similarly, for testing (1.1), the CQ test can be modified as

$$T_{CQ} = \frac{\sum_{i \neq j} \boldsymbol{x}_i^\top \boldsymbol{x}_j}{\sqrt{2n(n-1)\widehat{\operatorname{tr}(\boldsymbol{\Sigma}^2)}}}, \tag{1.3}$$

where

$$\widehat{\operatorname{tr}(\boldsymbol{\Sigma}^2)} = \frac{1}{n(n-1)} \operatorname{tr}\left\{\sum_{j \neq k}^n (\boldsymbol{x}_j - \bar{\boldsymbol{x}}_{(j,k)})\boldsymbol{x}_j^\top (\boldsymbol{x}_k - \bar{\boldsymbol{x}}_{(j,k)})\boldsymbol{x}_k^\top\right\} \tag{1.4}$$

is a ratio-consistent estimator of $\operatorname{tr}(\boldsymbol{\Sigma}^2)$, and $\bar{\boldsymbol{x}}_{(j,k)}$ denotes the usual sample mean vector of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ with the observed vectors $\boldsymbol{x}_j$ and $\boldsymbol{x}_k$ excluded. The asymptotic normality of the BS and CQ tests are established by the respective authors. Wang et al. (2015) showed that while the BS and CQ tests work well for high-dimensional data with light-tailed distributions, e.g., the multivariate normal distribution, they are less powerful for high-dimensional data with heavy-tailed distributions. A generalization of the multivariate normal distribution, which includes many heavy-tailed distributions, such as normal mixture and multivariate-t distributions, is the elliptical distributions. Elliptically distributed data have the following representation (Fang et al. 1990):

$$\boldsymbol{x}_i = \boldsymbol{\mu} + \boldsymbol{\epsilon}_i, \ \boldsymbol{\epsilon}_i = r_i \Gamma \boldsymbol{u}_i, \ i = 1, \ldots, n, \tag{1.5}$$

*Corresponding author.

Table 1. Empirical sizes (%) of the four tests in Simulation 1

			$\rho = 0.1$				$\rho = 0.5$				$\rho = 0.9$			
Model	p	n	NEW	WPL	BS	CQ	NEW	WPL	BS	CQ	NEW	WPL	BS	CQ
1	50	30	5.52	6.91	1.48	6.88	6.64	7.55	6.54	7.70	5.83	8.21	7.92	8.44
		60	5.21	6.46	2.81	6.59	5.99	7.15	6.37	6.99	5.99	7.32	6.98	7.31
		120	5.36	6.07	3.74	6.05	5.65	7.01	6.62	7.00	5.15	7.42	7.54	7.79
	500	30	7.17	7.07	1.04	7.28	6.38	7.92	7.14	8.60	6.31	7.75	7.97	8.39
		60	6.73	6.73	1.71	7.41	6.46	7.11	6.75	7.45	5.37	7.00	7.27	7.57
		120	6.28	6.35	2.99	6.91	5.93	7.07	6.77	7.21	5.36	7.24	7.28	7.42
	1000	30	7.39	7.26	1.04	7.88	6.81	7.86	6.70	8.14	5.76	7.21	6.97	7.59
		60	6.85	7.31	1.66	7.46	5.95	7.14	6.30	7.09	5.85	7.12	7.30	7.67
		120	6.70	7.17	3.04	7.19	5.75	7.05	6.99	7.54	5.34	6.76	7.20	7.42
2	50	30	6.02	6.66	1.59	6.75	6.15	7.66	6.46	8.55	6.25	7.71	7.48	8.18
		60	5.66	6.07	2.18	6.64	5.71	7.24	6.46	7.86	5.89	7.11	6.87	7.52
		120	5.17	6.21	2.57	6.50	5.71	6.63	5.98	7.01	5.50	7.48	6.94	7.25
	500	30	6.90	7.60	1.00	7.75	6.59	7.77	6.59	8.46	6.03	8.42	8.03	8.97
		60	6.29	6.28	1.39	6.83	6.00	7.28	6.27	7.82	5.46	7.34	7.28	7.98
		120	6.10	7.08	2.03	6.94	6.08	6.56	6.11	7.24	5.40	6.79	6.66	7.16
	1000	30	6.80	7.74	0.97	7.62	6.89	8.05	6.16	8.43	6.03	7.93	7.29	8.22
		60	6.60	7.15	1.48	7.78	6.17	7.73	6.70	8.17	5.62	7.41	6.79	7.49
		120	6.79	6.97	1.89	7.18	6.13	7.13	6.03	6.98	5.50	7.33	7.05	7.60
ARE			26.16	36.77	61.54	41.82	23.32	46.57	29.93	53.60	14.04	48.39	45.36	55.52

where $\Gamma : p \times p$ is a constant matrix, $\boldsymbol{u}_i$ is a random vector uniformly distributed on the unit sphere in $\mathbb{R}^p$, $r_i \geq 0$ is a random variable independent of $\boldsymbol{u}_i$, and $p^{-1}\,\mathrm{E}(r_i^2)\Gamma\Gamma^\top = \boldsymbol{\Sigma}$.

An important class of tests for elliptical distributions are based on the multivariate spatial sign or rank function, see Oja (2010) for an introduction to these multivariate nonparametric tests. The multivariate spatial signs of the original data are given by

$$\boldsymbol{z}_i = U(\boldsymbol{x}_i) = \begin{cases} \frac{\boldsymbol{x}_i}{\|\boldsymbol{x}_i\|}, & \boldsymbol{x}_i \neq \boldsymbol{0}, \\ \boldsymbol{0}, & \boldsymbol{x}_i = \boldsymbol{0}, \end{cases} \quad i = 1, \ldots, n. \tag{1.6}$$

Note that when $\boldsymbol{x}_i \neq \boldsymbol{0}$, we have $\|\boldsymbol{z}_i\|^2 = \boldsymbol{z}_i^\top \boldsymbol{z}_i = 1$, so that the original nonzero observations are transformed into vectors on the unit sphere in $\mathbb{R}^p$, whose L^2-norms are always 1. Inspired by the CQ test, Wang et al. (2015) proposed a nonparametric one-sample test (denoted as the WPL test) based on the transformed data (1.6) for elliptically distributed high-dimensional data, as briefly described below.

Set

$$\boldsymbol{V}_p = \mathrm{Cov}\{U(\boldsymbol{\epsilon}_1)\} = \mathrm{E}\,\frac{\boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_1^\top}{\|\boldsymbol{\epsilon}_1\|^2}, \tag{1.7}$$

which equals to the covariance matrix of $\boldsymbol{z}_1$ under the null hypothesis. That is, under the null hypothesis, we have $\mathrm{E}(\boldsymbol{z}_1\boldsymbol{z}_1^\top) = \boldsymbol{V}_p$. By (1.7), it is easy to see that

$$\mathrm{tr}(\boldsymbol{V}_p) = 1. \tag{1.8}$$

By imitating the CQ test statistic (1.3), Wang et al. (2015) defined their test statistic for testing (1.1) as

$$T_{WPL} = \frac{\sum_{i<j} \boldsymbol{z}_i^\top \boldsymbol{z}_j}{\sqrt{\frac{n(n-1)}{2}\widehat{\mathrm{tr}(\boldsymbol{V}_p^2)}}}, \tag{1.9}$$

where

$$\widehat{\mathrm{tr}(\boldsymbol{V}_p^2)} = \frac{1}{n(n-1)}\,\mathrm{tr}\left\{\sum_{j\neq k}(\boldsymbol{z}_j - \bar{\boldsymbol{z}}_{(j,k)})\boldsymbol{z}_j^\top(\boldsymbol{z}_k - \bar{\boldsymbol{z}}_{(j,k)})\boldsymbol{z}_k^\top\right\} \tag{1.10}$$

is an estimator of $\mathrm{tr}(\boldsymbol{V}_p^2)$.

Wang et al. (2015) showed that compared with the classical nonparametric test with finite fixed p, the WPL test has a substantial power gain and is more powerful than the CQ test for high-dimensional data with heavy-tailed distributions. The WPL test is conducted via a normal approximation to its null distribution. However, strong conditions (see Conditions (C1) and (C2) of Wang et al. 2015) are needed for the WPL test statistic (1.9) to have a normal limit distribution under the null hypothesis. These conditions usually assume that the underlying covariance matrix is sparse in the sense that the p component variables of the data are nearly independent. This is, however, unrealistic for many highly correlated high-dimensional data. In fact, the simulation studies presented in Section 3 indicate that the normal approximation to the null distribution of the WPL test statistic is not adequate for highly correlated high-dimensional data so that the WPL test tends to have inflated empirical sizes. For example, from Table 1, it is seen that the empirical sizes of the WPL test can be as large as

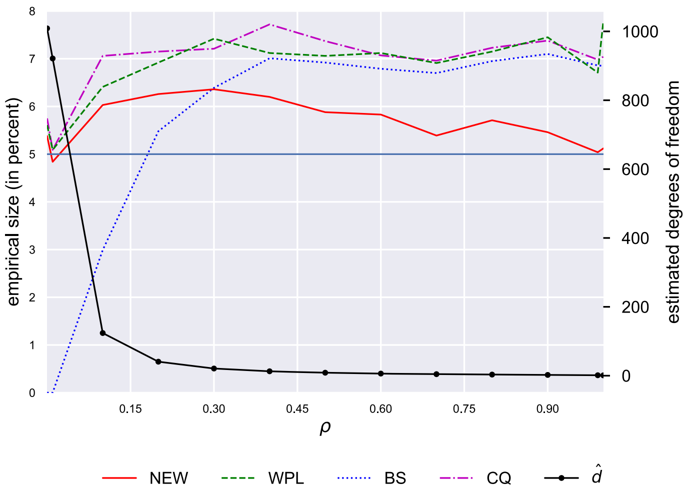

Figure 1. Empirical size curves (%) of the four tests and the estimated approximate degrees of freedom curve of the NEW test for a set of ρ's values simulated under Model 1 of Simulation 1 with $n = 120$ and $p = 1000$.

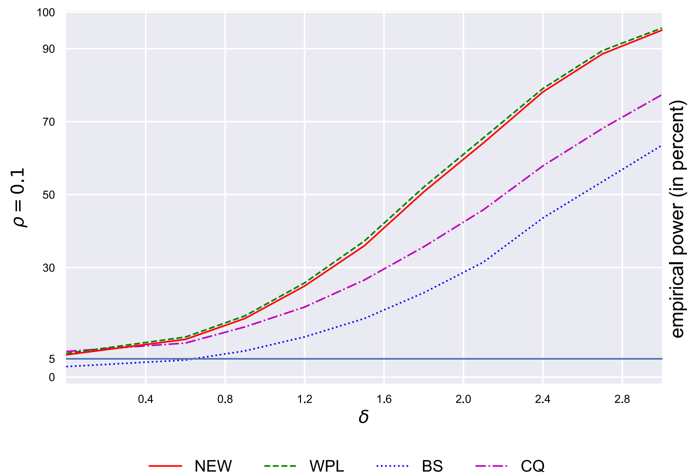

Figure 2. Empirical power curves (%) of the four tests for a set of δ's values with $\rho = 0.1$, simulated under Model 1 of Simulation 1 with $n = 120$ and $p = 1000$.

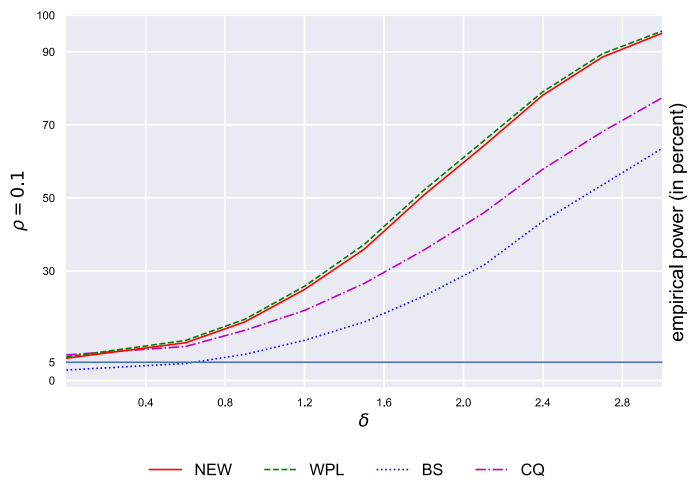

Figure 3. Empirical power curves (%) of the four tests for a set of δ's values with $\rho = 0.9$, simulated under Model 1 of Simulation 1 with $n = 120$ and $p = 1000$.

8.42% for highly correlated high-dimensional data when the nominal size is 5%, meaning a 68.4% relative error. More details are presented in Figure 1. This is an unacceptable size control problem since the resulting conclusion can be misleading and not reliable. It also artificially enlarges the power of the WPL test in some degree; see Figures 2 and 3 for some details. A possible reason for this undesired property of the WPL test is that although the underlying distribution of T_{WPL} can be skewed, the approximate normal distribution used by the WPL test is always symmetric and bell-shaped and hence it is not flexible to adapt to the underlying null distribution of T_{WPL}.

To overcome the above problems, in this paper, based on the spatial signs (1.6) of the original data, we propose to use the following test statistic

$$T_{n,p} = n\|\bar{\boldsymbol{z}}\|^2, \tag{1.11}$$

which is connected with the statistic of the WPL test in the following way:

$$T_{n,p} = \sqrt{\frac{2(n-1)}{n}\widehat{\operatorname{tr}(\boldsymbol{V}_p^2)}}T_{WPL} + 1. \tag{1.12}$$

Note that $T_{n,p}$ and T_{WPL} are essentially equivalent since $\{2n^{-1}(n-1)\widehat{\operatorname{tr}(\boldsymbol{V}_p^2)}\}^{1/2}$ is nearly a constant when n is large. Thus, their distributions are similar in shapes, i.e., both distributions are skewed, symmetric or normal. However, it is seen from (1.11) that $T_{n,p}$ is always nonnegative while T_{WPL} takes both positive and negative values.

The main contributions of this paper can be summarized as follows. First of all, we show in Theorem 1 that the limit null distribution of $T_{n,p}$ is in general a chi-squared mixture with non-negative coefficients. This means that the limit null distribution of $T_{n,p}$ is generally skewed and often not normally distributed. This result is very different from the normal limit distribution of T_{WPL} obtained with strong conditions imposed in Wang et al. (2015). Secondly, we propose to approximate the distribution of $T_{n,p}$ using a scaled chi-squared random variable of form $R \sim \chi_d^2/d$ which is always nonnegative and generally skewed. The parameter d is usually called the approximate degrees of freedom of $T_{n,p}$. We determine the value of d via matching the variances of $T_{n,p}$ and R under the null hypothesis. By doing so, the distribution of R is adaptive to the shape of the underlying null distribution of $T_{n,p}$ in the sense that when the distribution of $T_{n,p}$ is skewed, the value of d is small; when the distribution of $T_{n,p}$ is symmetric, the value of d is large; and when the null distribution of $T_{n,p}$ is asymptotically normal, the value of d will tend to infinity. The above method is essentially the well-known Welch–Satterthwaite (W–S) χ^2-approximation (a two-moment matched method, see Welch 1947, Satterthwaite 1946, and Zhang et al. 2015 among others) to the distribution of $T_{n,p}$ since we always have $\mathrm{E}(R) = 1$ and by Lemma 1, under the null hypothesis, we have $\mathrm{E}(T_{n,p}) = 1$. Note also that $T_{n,p}$ and R have the same range $[0, \infty)$. The adaptivity of the above χ^2-approximation is obviously not shared by the normal approximation used in the WPL test although the normal approximation is also a two-moment matched method. Thirdly, based on Theorem 1, we present a sufficient and necessary condition such that the limit null distribution of $T_{n,p}$ is normal. This condition allows us to determine if and when a normal approximation to the underlying distribution of $T_{n,p}$ is adequate in practice. Fourthly, based on the theoretical results of Zhang (2005) and Zhang et al. (2015), we show that the density error bound of the W–S χ^2-approximation to the limit null distribution of $T_{n,p}$ is generally of smaller order than that of the normal approximation. This hence shows that the former is generally preferred to the latter theoretically in terms of size control. This is actually demonstrated by several simulation studies under various settings, presented in Section 3 and in the supplementary material (http://intlpress.com/site/pub/pages/journals/items/sii/content/vols/0012/0001/s002), and two real data examples presented in Section 4. Fifthly, we show that under some mild conditions, the estimator (1.10) is a ratio-consistent estimator of $\mathrm{tr}(\boldsymbol{V}_p^2)$ which has not been shown in Wang et al. (2015) but we shall use it in estimating the approximate degrees of freedom d. Note that the ratio-consistency of the estimator (1.4) of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ as shown in Chen and Qin (2010) does not imply the ratio-consistency of the estimator (1.10) of $\mathrm{tr}(\boldsymbol{V}_p^2)$ since the factor model and the "pseudo independence" condition required by Chen and Qin (2010) are no longer satisfied by the transformed data (1.6).

The rest of the article is organized as follows. The main results are presented in Section 2. Several simulation studies and two real data examples are presented in Sections 3 and 4, respectively. Two additional simulation studies are presented in the supplementary material. All the technical proofs of the main results are given in the Appendix.

2. MAIN RESULTS

2.1 Asymptotic null distribution

Denote $T_{n,p,0}$ as $T_{n,p}$ under the null hypothesis. Let $\xrightarrow{\mathcal{L}}$, $\xrightarrow{P}$ and $\stackrel{d}{=}$ denote convergence in distribution, in probability and equal in distribution, respectively. The following theorem shows that the limit null distribution of $T_{n,p}$ is a chi-squared mixture, i.e., a linear combination of a series of independent chi-squared random variables.

Theorem 1. *For any fixed finite p, as $n \to \infty$, we have $T_{n,p,0} \xrightarrow{\mathcal{L}} T_{p,0}$ where $T_{p,0} \stackrel{d}{=} \sum_{r=1}^p \lambda_{p,r} A_r$ with $\lambda_{p,r}$'s being the eigenvalues of $\boldsymbol{V}_p$ and $A_1, \ldots, A_r, \ldots$ being i.i.d. χ_1^2 random variables. The above expression also holds for $p = \infty$ provided that $\lim_{p\to\infty} \boldsymbol{V}_p = \boldsymbol{V}_\infty$ and $\lim_{p\to\infty} \lambda_{p,r} = \lambda_{\infty,r}$ for all $r = 1, 2, \ldots$ uniformly where $\lambda_{\infty,r}$'s are the eigenvalues of $\boldsymbol{V}_\infty$.*

Remark 1. *Theorem 1 shows that $T_{n,p,0}$ is asymptotically a chi-squared mixture which is generally skewed and is often not normally distributed unless some strong assumptions are imposed as in Wang et al. (2015). In practice, it is often not easy to check if the assumptions imposed in Wang et al. (2015) are satisfied.*

2.2 Approximate the distribution of $T_{p,0}$

In practice, p is always finite but it can be very large. Further, the eigenvalues $\lambda_{p,r}$'s of $\boldsymbol{V}_p$ are generally unknown and it is often rather challenging to estimate them consistently. Therefore, it is unrealistic to compute the distribution of $T_{p,0}$ directly via replacing the eigenvalues $\lambda_{p,r}$'s by their estimates. Fortunately, we can approximate the distribution of $T_{p,0}$ by the W–S χ^2-approximation. Since $\mathrm{E}(T_{p,0}) = \sum_{r=1}^p \lambda_{p,r} = \mathrm{tr}(\boldsymbol{V}_p) = 1$, we can approximate the distribution of $T_{p,0}$ using that of a random variable of form

$$R \sim \chi_d^2/d, \tag{2.1}$$

where χ_d^2 denotes a chi-squared random variable with d degrees of freedom. The parameter d may be called the approximate degrees of freedom of $T_{p,0}$. It can be determined via matching the variances of $T_{p,0}$ and R. By (2.1), we have $\mathrm{Var}(R) = 2/d$ and by Theorem 1, we have $\mathrm{Var}(T_{p,0}) = \sum_{r=1}^p \lambda_{p,r}^2 = \mathrm{tr}(\boldsymbol{V}_p^2)$. Equating the variances of $T_{p,0}$ and R leads to

$$d = 1/\mathrm{tr}(\boldsymbol{V}_p^2). \tag{2.2}$$

Under some conditions, the distribution of $T_{p,0}$ can tend to normal as shown by Theorem 1 (c) of Zhang (2005). A question arises naturally, under the same conditions, will the distribution of R also tend to normal? To answer this question, we introduce the following notations. Let $\lambda_{\max}$ denote the largest eigenvalue of $\boldsymbol{V}_p$, and let

$$\Delta = \frac{\lambda_{\max}^2}{\mathrm{tr}(\boldsymbol{V}_p^2)}, \quad d^* = \frac{\mathrm{tr}^3(\boldsymbol{V}_p^2)}{\mathrm{tr}^2(\boldsymbol{V}_p^3)}, \quad \text{and} \quad M = \frac{\mathrm{tr}(\boldsymbol{V}_p^4)}{\mathrm{tr}^2(\boldsymbol{V}_p^2)}. \tag{2.3}$$

Then we can easily show that the skewness and kurtosis of $T_{p,0}$ can be expressed as $\sqrt{8/d^*}$ and $12M$, respectively. In addition, by Lemma 1 (c) and (d) of Zhang (2005), we have

$$1/d^* \le \Delta \le (1/d^*)^{1/3}, \quad \text{and} \quad 1/d^* \le M \le (\Delta/d^*)^{1/2}. \tag{2.4}$$

Furthermore, by some simple algebra, we have (Zhang et al. 2015, Theorem 3)

$$1 \le d^* \le d \le p. \tag{2.5}$$

We have the following useful theorem.

Theorem 2. *As $p \to \infty$, the distribution of $T_{p,0}$ tends to normal if and only if $d^* \to \infty$. In addition, when $d^* \to \infty$, we have*

$$\frac{T_{p,0}-1}{\sqrt{2/d}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \text{ and } \frac{R-1}{\sqrt{2/d}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Remark 2. *Theorem 2 indicates that when $d^* \to \infty$, both the distributions of $T_{p,0}$ and R tend to normal, and when d is finite, both the distributions of $T_{p,0}$ and R will not tend to normal. In practice, p is always finite and hence by (2.5), both d^* and d are finite so that both the distributions of $T_{p,0}$ and R will not tend to normal.*

2.3 Accuracy of the χ^2-approximation and the normal approximation

Let the probability density function and the normalized version of a random variable X be denoted by $f_X(x)$ and $\tilde{X} = \{X - \mathrm{E}(X)\}/\sqrt{\mathrm{Var}(X)}$, respectively. Then by Theorem 4 of Zhang et al. (2015), when $\Delta < 1/10$ and $d > 10$ we have

$$\begin{aligned} &\sup_x |f_{\tilde{T}_{p,0}}(x) - f_{\tilde{R}}(x)| \\ &\le 0.1403\left[\{3 + \tfrac{3.8578}{(1-10\Delta)^{5/2}}\}M + \{3 + \tfrac{3.8578}{(1-10/d)^{5/2}}\}d^{-1}\right] \\ &\quad + 0.7040\{(d^*)^{-1/2} - d^{-1/2}\}. \end{aligned} \tag{2.6}$$

That is, the density approximate error bound of the W–S χ^2-approximation to the distribution of $T_{p,0}$ is determined by the variance, skewness, and kurtosis of $T_{p,0}$. Let $\phi(x)$ denote the probability density function of the standard normal distribution. By Theorem 1 (a) of Zhang (2005), when $\Delta < 1/8$, we have

$$\sup_x |f_{\tilde{T}_{p,0}}(x) - \phi(x)| \le 0.1323\left\{4 + \frac{2.3617}{(1-8\Delta)^2}\right\}(d^*)^{-1/2}. \tag{2.7}$$

Remark 3. *By (2.6), the density error bound of the W–S χ^2-approximation R to $T_{p,0}$ is $O(M) + O(d^{-1}) + O\{(d^*)^{-1/2} - d^{-1/2}\}$ while by (2.7), the density error bound of the normal approximation is $O\{(d^*)^{-1/2}\}$. By (2.4) and (2.5), $O(M)$, $O(d^{-1})$ and $O\{(d^*)^{-1/2} - d^{-1/2}\}$ are of smaller orders or generally smaller than $O\{(d^*)^{-1/2}\}$. Thus we theoretically justify that the W–S χ^2-approximation to $T_{p,0}$ is generally preferred to the normal approximation. These conclusions are actually verified by simulation results presented in Section 3 and in the supplementary material and explain why our test has a much better size control than other approaches generally as demonstrated in Section 3 and in the supplementary material.*

2.4 Implementation

For a given high-dimensional sample, both n and p are finite. To take this fact into account, by Theorem 1, one may approximate the distribution of $T_{n,p,0}$ directly using that of R (2.1) via matching the variances of $T_{n,p,0}$ and R to determine the approximate degrees of freedom d. By some calculation, we have the following lemma.

Lemma 1. *We have* $\mathrm{E}(T_{n,p,0}) = 1$ *and* $\mathrm{Var}(T_{n,p,0}) = \frac{2(n-1)}{n}\,\mathrm{tr}(\boldsymbol{V}_p^2)$.

Therefore, equating the variances of $T_{n,p,0}$ and R leads to

$$d = n/\left\{(n-1)\,\mathrm{tr}(\boldsymbol{V}_p^2)\right\}. \tag{2.8}$$

For any fixed finite p, as $n \to \infty$, it is easy to see that the expression (2.8) will tend to the expression (2.2).

To apply the proposed test, we need to estimate d or $\mathrm{tr}(\boldsymbol{V}_p^2)$ consistently. An estimator of $\mathrm{tr}(\boldsymbol{V}_p^2)$ is given in (1.10) by Wang et al. (2015). A computational efficient expression of $\widehat{\mathrm{tr}(\boldsymbol{V}_p^2)}$ is given in Eq. (8) of Wang et al. (2015). The estimator (1.10) is inspired by the estimator $\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}$ (1.4) of Chen and Qin (2010) under a general factor data model, which was shown to be ratio-consistent by Chen and Qin (2010) under some assumptions including a "pseudo independence" condition. However, Wang et al. (2015) did not show the ratio-consistency of $\widehat{\mathrm{tr}(\boldsymbol{V}_p^2)}$ given in (1.10) for high-dimensional elliptically distributed data considered here while the original ratio-consistency result obtained by Chen and Qin (2010) is not directly applicable because the factor data model (3.1) and the "pseudo independence" condition (3.2) in Chen and Qin (2010) are no longer satisfied by the transformed data (1.6). Here we would like to close this gap via showing the ratio-consistency of $\widehat{\mathrm{tr}(\boldsymbol{V}_p^2)}$. To this end, we impose the following Condition A:

$$\begin{aligned} &(1) \quad n^{-1}\,\mathrm{E}(\boldsymbol{z}_1^\top \boldsymbol{V}_p \boldsymbol{z}_1)^2 = o\left\{\mathrm{tr}^2(\boldsymbol{V}_p^2)\right\}, \\ &(2) \quad n^{-2}\,\mathrm{E}(\boldsymbol{z}_1^\top \boldsymbol{z}_2)^4 = o\left\{\mathrm{tr}^2(\boldsymbol{V}_p^2)\right\}. \end{aligned}$$

Condition A is rather general and much milder than the respective conditions imposed by Chen and Qin (2010) for the ratio-consistency of their estimator $\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}$. It can be shown (see the proof of Lemma 1 in Wang et al. 2015) that Condition A is satisfied under the conditions (C1) and (C2) of Wang et al. (2015). The following theorem presents the unbiasedness and the ratio-consistency of $\widehat{\mathrm{tr}(\boldsymbol{V}_p^2)}$ under Condition A and the null hypothesis.

Theorem 3. *Under H_0, we have* $\mathrm{E}\left\{\widehat{\mathrm{tr}(\boldsymbol{V}_p^2)}\right\} = \mathrm{tr}(\boldsymbol{V}_p^2)$. *Further, under H_0 and Condition A, we have* $\widehat{\mathrm{tr}(\boldsymbol{V}_p^2)}/\,\mathrm{tr}(\boldsymbol{V}_p^2) \xrightarrow{P} 1$.

With $\widehat{\mathrm{tr}(\boldsymbol{V}_p^2)}$ and by (2.8), a ratio-consistent estimator of d is then given by $\hat{d} = n/\{(n-1)\widehat{\mathrm{tr}(\boldsymbol{V}_p^2)}\}$. The proposed test can then be conducted via using the critical value $\chi^2_{\hat{d}}(\alpha)/\hat{d}$ or the approximate p-value $\Pr\left(\chi^2_{\hat{d}} > \hat{d}\, T_{n,p}\right)$ where $\chi^2_{\hat{d}}(\alpha)$ denotes the $100(1-\alpha)$th percentile of the chi-squared distribution with $\hat{d}$ degrees of freedom. The empirical performance of the proposed test will be demonstrated by the simulation studies presented in the next section and in the supplementary material.

3. SIMULATION STUDIES

In this section, we conduct two simulation studies (some additional simulation studies are presented in the supplementary material) to compare the proposed test (denoted as NEW) against the WPL, BS and CQ tests. We compare them under various simulation settings in terms of size control and power, aiming to see how the proposed test performs compared with the existing competitors.

In each run, we generate the data using the following factor model

$$\boldsymbol{x}_i = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{v}_i,\ i = 1, \ldots, n, \tag{3.1}$$

where $\boldsymbol{\mu} = \delta\boldsymbol{h}$ and $\boldsymbol{\Sigma}$ is a $p \times p$ nonnegative definite matrix, depending on a nonnegative tuning parameter denoted as ρ. The tuning parameters δ and $\boldsymbol{h}$ are used to control the mean vector so that the power of a test will increase with increasing the value of δ and the tuning parameter ρ is used so that the data correlation will increase with increasing the value of ρ. For simplicity, without loss of generality, we set $\boldsymbol{h} = \boldsymbol{u}/\|\boldsymbol{u}\|$ with $\boldsymbol{u} = (1, \ldots, p)^\top$. To compare the performance of the tests under consideration with small, moderate, and large tuning parameters, we consider three cases of dimension with $p = 50, 500, 1000$, three cases of sample sizes with $n = 30, 60, 120$, and three cases of data correlation with $\rho = 0.1, 0.5, 0.9$. Other cases are considered in Figures 1–3.

The empirical sizes and powers are calculated based on $N = 10{,}000$ simulation runs with the nominal size $\alpha = 5\%$. In each run, the test statistics are computed and the associated p-values are calculated. When the p-value of a test is smaller than $\alpha = 5\%$, the null hypothesis is rejected. The empirical size or power of a test is then calculated as the proportions of the number of rejections out of N runs based on the calculated p-values. To assess the performance of a test in maintaining the nominal size (type I error), we use the following so-called average relative error $\mathrm{ARE} = 100M^{-1}\sum_{j=1}^{M}|\hat{\alpha}_j - \alpha|/\alpha$, where $\hat{\alpha}_j,\ j = 1, \ldots, M$ denote the empirical sizes under consideration. A smaller ARE value indicates a better overall performance of the associated test in terms of size control.

3.1 Simulation 1

In this simulation study, the matrix $\boldsymbol{\Sigma}$ is specified as $\boldsymbol{\Sigma} = (1-\rho)\boldsymbol{I}_p + \rho\boldsymbol{J}_p$, with $\boldsymbol{I}_p$ a $p \times p$ identity matrix and $\boldsymbol{J}_p$ a $p \times p$ matrix of ones, and we generate the i.i.d. random vectors $\boldsymbol{v}_i,\ i = 1, \ldots, n$ from the following two models:

- Model 1. $v_{ij},\ j = 1, \ldots, p$, i.i.d. follow the normal mixture $0.9\mathcal{N}(0,1) + 0.1\mathcal{N}(0,9)$.
- Model 2. $\boldsymbol{v}_i = \boldsymbol{w}_i/\sqrt{3}$, with $\boldsymbol{w}_i$ following a multivariate t-distribution with mean $\boldsymbol{0}$, correlation matrix $\boldsymbol{I}_p$, and 3 degrees of freedom.

The generated data are then with symmetric and heavy-tailed distributions. It is expected that they will favor the NEW and WPL tests against the BS and CQ tests in terms of power.

Table 1 displays the empirical sizes of the NEW, WPL, BS and CQ tests under various settings with their ARE values listed at the last row. It is seen that the NEW test performs well with most of its empirical sizes around 6% while the other three tests are rather liberal with most of their empirical sizes more than 7% except when $\rho = 0.1$, the BS test is very conservative with most of its empirical sizes less than 2%. Therefore, in terms of size control, the NEW test outperforms the other three tests. This is also indicated by their ARE values. The ARE values of the four tests are respectively $26.16, 36.77, 61.54, 41.82$ when $\rho = 0.1$, $23.32, 46.57, 29.93, 53.60$ when $\rho = 0.5$, and $14.04, 48.39, 45.36, 55.52$ when $\rho = 0.9$, showing that the ARE values of the NEW test are always smaller than those of the other three tests.

Table 2 displays the associated empirical powers of the four tests. It is seen that in terms of power, the NEW and WPL tests are generally comparable and they both outperform the BS and CQ tests. This shows that the spatial-sign-transformation indeed helps to improve the power of a one-sample test when the data are heavy-tailed. Note that compared with the NEW test, the slightly higher empirical powers of the WPL test are mainly due to the fact that it also has larger associated empirical sizes than the NEW test; see Figures 2 and 3 below for more details.

Table 3 presents the estimated approximate degrees of freedom, $\hat{d}$, of the NEW test. It is seen that the value of $\hat{d}$ has strong relationship with n, p and ρ. When n and p are fixed, $\hat{d}$ decreases with ρ increasing while when ρ is fixed, $\hat{d}$ increases with p/n increasing. In particular, when $\rho = 0.5$ and 0.9, $\hat{d}$ is generally small (≤ 10), showing that in these cases, the underlying null distribution of the NEW test is skewed. Since the underlying null distribution of the WPL test has a similar shape as that of the underlying null distribution of the NEW test (see (1.12)), the normal approximation to the null distribution of the WPL test is impossibly adequate when $\rho = 0.5$ and 0.9. This partially explains why the NEW test always outperforms the WPL test in terms of size control.

To further study the effect of ρ on the performance of the four tests in terms of size control, Figure 1 displays the empirical size curves of the four tests (NEW — solid, WPL — dashed, BS — dotted, and CQ — dot-dashed) and the

Table 2. Empirical powers (%) of the four tests in Simulation 1

Model	p	n	δ	$\rho = 0.1$ NEW	WPL	BS	CQ	$\rho = 0.5$ NEW	WPL	BS	CQ	$\rho = 0.9$ NEW	WPL	BS	CQ
1	50	30	1.5	85.56	86.62	52.55	71.76	37.42	40.55	28.33	30.85	23.98	29.90	21.28	22.11
		60	1.1	88.51	89.81	60.88	73.13	38.95	42.70	28.95	30.23	25.68	30.22	19.79	20.32
		120	0.8	90.95	91.47	68.74	75.31	40.82	44.98	29.37	30.31	26.27	31.88	19.71	20.08
	500	30	4.0	88.86	90.01	40.69	73.69	28.64	31.81	22.61	25.11	18.27	22.23	16.65	17.44
		60	2.8	89.22	89.80	48.45	70.54	27.33	29.47	20.43	21.67	17.75	20.57	14.91	15.43
		120	2.0	89.80	90.90	56.64	70.10	27.29	30.63	21.23	21.85	17.04	20.64	14.72	14.98
	1000	30	5.5	88.67	90.18	39.34	73.26	26.74	30.06	21.86	24.22	18.18	21.47	16.90	17.65
		60	4.0	91.49	92.02	50.20	73.33	28.14	31.60	21.81	23.14	16.72	21.56	15.56	16.04
		120	2.5	81.34	82.93	44.62	60.46	22.01	24.50	17.24	17.82	14.81	17.39	12.28	12.56
2	50	30	1.1	96.29	96.86	55.09	76.73	54.87	59.08	31.77	35.90	37.67	43.81	22.96	24.38
		60	0.8	97.87	98.18	59.77	77.12	57.59	62.17	30.16	33.40	39.17	46.62	20.85	21.93
		120	0.5	93.82	94.40	49.05	65.10	45.52	49.95	23.09	25.34	30.56	36.65	16.86	17.57
	500	30	2.7	95.60	96.12	38.06	72.41	35.59	39.19	21.12	24.89	23.11	27.36	16.20	17.44
		60	1.9	96.32	96.79	39.05	68.58	35.25	39.07	19.78	22.67	22.72	27.20	15.19	16.18
		120	1.3	95.83	96.15	39.73	63.93	33.02	36.72	17.92	19.82	20.56	25.19	13.49	14.08
	1000	30	3.5	92.76	93.34	31.70	67.03	31.36	34.57	19.22	22.63	20.10	24.09	14.89	16.03
		60	2.5	94.60	95.06	34.26	65.17	31.28	34.40	18.38	20.94	20.00	24.27	13.94	14.98
		120	1.8	96.27	96.77	38.00	63.67	31.47	34.85	17.25	19.21	19.96	24.68	13.83	14.47

Table 3. Estimated approximate degrees of freedom of the NEW test in Simulation 1.

p	n	Model 1 $\rho = 0.1$	$\rho = 0.5$	$\rho = 0.9$	Model 2 $\rho = 0.1$	$\rho = 0.5$	$\rho = 0.9$
50	30	39.29	8.30	2.32	39.23	8.32	2.33
	60	38.18	7.85	2.24	38.21	7.84	2.24
	120	37.69	7.65	2.20	37.62	7.65	2.20
500	30	124.00	9.44	2.37	124.50	9.44	2.37
	60	115.50	8.85	2.27	115.94	8.81	2.27
	120	111.18	8.58	2.24	110.57	8.56	2.24
1000	30	143.93	9.52	2.37	142.67	9.54	2.38
	60	129.64	8.90	2.28	130.57	8.92	2.28
	120	124.30	8.61	2.23	124.04	8.62	2.23

estimated approximate degrees of freedom ($\hat{d}$) of the NEW test for a set of ρ's values, simulated under Model 1 of Simulation 1 with $n = 120$ and $p = 1000$. From the empirical size curves of the four tests, it is seen that in terms of size control, the NEW test outperforms other three tests generally, the WPL and CQ tests are rather liberal when $\rho \geq 0.10$, and the BS test is very conservative when $\rho \leq 0.10$ and it becomes very liberal when $\rho \geq 0.40$. These conclusions are consistent with those observed from Table 1. From the estimated approximate degrees of freedom curve of the NEW test, it is seen that with the value of ρ increasing, $\hat{d}$ generally becomes smaller, showing that the normal approximation used in the WPL, BS and CQ tests becomes less adequate. This again partially explains why the NEW test generally outperforms the WPL, BS and CQ tests, especially when the value of ρ is large.

To further study the effect of ρ on the performance of the four tests in terms of power, Figures 2 and 3 display the empirical power curves of the four tests (NEW — solid, WPL — dashed, BS — dotted, and CQ — dot-dashed) for a set of δ's values with $\rho = 0.1$ and $\rho = 0.9$ respectively, simulated under Model 1 of Simulation 1 with $n = 120$ and $p = 1000$. From these two figures, it is seen that in terms of power, both the NEW and WPL tests outperform the BS and CQ tests and when $\rho = 0.1$, the NEW and WPL tests are generally comparable since their empirical sizes are comparable and when $\rho = 0.9$, the WPL test has slightly larger empirical powers than the NEW test since the former also has a larger empirical sizes than the latter. These conclusions are consistent with those drawn from Table 2.

3.2 Simulation 2

In this simulation study, we use the same settings as Simulation 1 except the matrix $\boldsymbol{\Sigma}$ is specified as $\boldsymbol{\Sigma} = \boldsymbol{DRD}$, where $\boldsymbol{D} = \text{diag}(\boldsymbol{h})$, $\boldsymbol{h} = \boldsymbol{u}/\|\boldsymbol{u}\|$ with $\boldsymbol{u} = (1, \ldots, p)^\top$, and $\boldsymbol{R}$ is a $p \times p$ matrix with the (i, j)th element being $\rho^{|i-j|}$. The simulation results are presented in Tables 4–6. The conclusions drawn from these tables are similar to those drawn from Tables 1–3 of Simulation 1, i.e., in terms of size control, the NEW test outperforms the other three tests and in terms of power, the NEW and WPL tests are generally comparable and they outperform the BS and CQ tests except now most of the empirical sizes of the NEW test are around 5%, most of the empirical sizes of the WPL and CQ tests are around 6% and the BS test becomes very conservative with most of its empirical sizes less than 2%.

4. APPLICATIONS

An important application of a one-sample test is to test if two paired samples have the same mean vectors. Suppose

Table 4. Empirical sizes (%) of the four tests in Simulation 2

			$\rho = 0.1$				$\rho = 0.5$				$\rho = 0.9$			
Model	p	n	NEW	WPL	BS	CQ	NEW	WPL	BS	CQ	NEW	WPL	BS	CQ
1	50	30	5.25	6.58	1.57	6.59	5.44	6.65	3.05	6.90	5.58	7.04	6.39	7.66
		60	5.21	6.27	2.73	6.42	5.19	6.49	3.99	6.35	5.99	7.56	6.85	7.77
		120	4.85	6.00	3.96	6.11	5.45	6.68	5.11	6.79	5.16	6.73	6.32	6.79
	500	30	5.40	5.84	0.21	5.92	5.75	6.35	0.30	6.28	5.48	6.69	1.64	6.99
		60	4.95	5.28	0.01	4.94	5.14	5.65	0.09	5.71	5.79	6.84	2.99	6.83
		120	4.89	5.35	0.03	5.28	5.42	5.98	0.30	5.78	5.57	6.51	4.42	7.08
	1000	30	4.67	5.01	0.24	5.58	5.65	6.05	0.22	6.13	5.93	6.69	0.58	6.57
		60	5.55	5.94	0.02	5.60	5.06	5.47	0.01	5.46	5.55	6.16	1.11	6.47
		120	4.93	5.27	0.00	5.19	5.14	5.59	0.00	5.46	5.45	6.13	2.21	5.97
2	50	30	5.30	6.51	1.32	6.15	5.63	7.11	2.76	7.14	6.17	7.66	5.89	7.98
		60	5.25	6.54	1.97	6.20	5.56	6.76	3.27	6.86	5.94	7.54	6.14	7.52
		120	5.11	6.28	2.23	5.78	5.52	6.70	3.79	6.64	5.71	7.59	6.15	7.24
	500	30	5.29	5.72	0.01	5.53	5.55	6.19	0.01	6.14	5.63	6.63	1.55	6.75
		60	5.14	5.62	0.01	5.49	5.30	5.84	0.12	5.84	5.37	6.49	1.91	6.35
		120	5.29	5.72	0.08	5.69	5.21	5.75	0.11	6.16	5.22	6.08	2.57	6.23
	1000	30	5.40	5.77	0.00	5.58	5.37	5.77	0.00	6.04	5.74	6.41	0.50	6.42
		60	4.95	5.29	0.00	5.07	5.61	6.04	0.02	5.63	5.62	6.41	0.75	6.17
		120	5.20	5.41	0.00	5.43	5.48	5.91	0.00	5.55	5.84	6.73	1.46	6.49
ARE			4.61	16.00	84.01	14.08	8.30	23.31	74.52	23.18	13.04	35.43	51.17	36.98

Table 5. Empirical powers (%) of the four tests in Simulation 2

				$\rho = 0.1$				$\rho = 0.5$				$\rho = 0.9$			
Model	p	n	δ	NEW	WPL	BS	CQ	NEW	WPL	BS	CQ	NEW	WPL	BS	CQ
1	50	30	0.250	98.43	98.71	73.75	88.97	87.86	89.55	61.65	74.33	40.86	45.06	36.31	39.31
		60	0.170	97.53	98.13	75.95	85.41	84.75	86.85	61.69	69.27	38.81	43.37	34.26	35.94
		120	0.120	97.57	98.05	80.88	85.87	84.92	86.89	64.83	69.08	38.60	43.10	33.51	34.60
	500	30	0.120	97.50	97.74	5.55	82.26	86.18	87.09	7.56	66.28	35.98	38.98	13.78	29.33
		60	0.080	95.23	95.61	6.07	75.05	81.59	82.65	9.78	58.81	31.89	34.59	15.50	25.49
		120	0.060	98.10	98.24	25.77	82.35	86.77	87.90	26.49	64.95	36.20	38.96	21.13	27.26
	1000	30	0.090	92.21	92.55	4.11	70.87	77.24	78.20	3.55	55.10	29.65	31.83	5.36	23.85
		60	0.065	94.51	94.88	0.51	72.01	80.17	81.20	1.25	55.41	30.64	32.83	8.13	23.19
		120	0.050	98.48	98.60	5.29	82.85	88.83	89.47	9.06	65.04	35.91	38.10	14.96	26.16
2	50	30	0.130	92.19	93.40	39.79	64.18	76.71	79.01	35.19	50.22	33.72	37.41	23.78	27.67
		60	0.100	97.26	97.87	50.12	70.01	84.60	86.73	40.73	53.65	38.52	42.82	25.73	28.57
		120	0.080	99.61	99.68	64.68	80.12	93.81	95.01	52.66	63.39	47.06	51.87	29.32	31.74
	500	30	0.070	96.31	96.57	3.95	64.00	85.03	85.89	5.05	50.74	35.12	37.88	9.36	22.53
		60	0.050	97.49	97.63	4.94	63.26	87.03	88.06	6.52	47.50	36.69	39.47	11.05	22.36
		120	0.033	94.91	95.26	4.65	50.99	80.48	81.88	6.12	39.11	31.53	34.19	10.10	18.07
	1000	30	0.056	95.14	95.40	0.58	60.45	82.28	83.15	0.85	46.02	32.68	35.20	4.13	20.53
		60	0.040	96.51	96.81	0.53	57.26	84.93	85.74	1.17	43.76	33.79	36.08	4.95	19.07
		120	0.028	96.46	96.68	0.92	53.28	83.76	84.71	1.58	40.48	32.81	35.30	5.71	17.30

we have two paired samples: $(\boldsymbol{y}_{i1}, \boldsymbol{y}_{i2})$, $i = 1, \ldots, n$ which are i.i.d. but for each i, $\boldsymbol{y}_{i1}$ and $\boldsymbol{y}_{i2}$ may be correlated. It is often of interest to test if the two paired samples have the same mean vectors:

$$(4.1) \quad H_0: \ \mathrm{E}(\boldsymbol{y}_{11}) = \mathrm{E}(\boldsymbol{y}_{12}) \text{ vs } H_1: \ \mathrm{E}(\boldsymbol{y}_{11}) \neq \mathrm{E}(\boldsymbol{y}_{12}).$$

It is well known that the above paired two-sample problem can be easily transformed into a one-sample problem. In fact, set $\boldsymbol{x}_i = \boldsymbol{y}_{i1} - \boldsymbol{y}_{i2}$, $i = 1, \ldots, n$. Then testing (4.1) is equivalent to testing the one-sample problem (1.1) with $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{y}_{11}) - \mathrm{E}(\boldsymbol{y}_{12})$ being the mean vector of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. Thus, the proposed NEW test, together with the WPL, BS and CQ tests, can be used to test (4.1).

As real data examples, we consider applications of the proposed NEW test, together with the WPL, BS and CQ tests, to three datasets. The first two datasets were provided by Chowdary et al. (2006), available at the Gene Expression Omnibus (GEO) (Barrett and Edgar 2006) with accession number GSE3726. The first dataset contains 31

Table 6. Estimated approximate degrees of freedom of the NEW test in Simulation 2

		Model 1			Model 2		
p	n	$\rho = 0.1$	$\rho = 0.5$	$\rho = 0.9$	$\rho = 0.1$	$\rho = 0.5$	$\rho = 0.9$
50	30	29.47	20.02	5.83	29.47	20.06	5.85
	60	28.85	19.54	5.66	28.87	19.53	5.66
	120	28.58	19.32	5.58	28.59	19.31	5.57
500	30	284.50	176.36	34.18	284.50	176.42	34.17
	60	278.40	172.69	33.41	278.40	172.87	33.39
	120	275.89	171.06	33.00	275.89	171.04	33.01
1000	30	567.68	349.68	64.72	567.78	349.95	64.79
	60	555.78	342.31	63.16	555.72	342.37	63.25
	120	550.52	339.10	62.56	550.57	339.26	62.55

Table 7. Results for the first two datasets, provided by Chowdary et al. (2006)

	1st dataset			2nd dataset		
Method	Statistic	p-value	$\hat{d}$	Statistic	p-value	$\hat{d}$
NEW	1.35	0.20	9.93	1.45	0.16	8.65
WPL	0.78	0.22	-	0.93	0.18	-
BS	0.03	0.49	-	0.19	0.42	-
CQ	0.03	0.49	-	0.49	0.31	-

pairs of matched lymph node-negative either "fresh frozen" or "stored in RNAlater preservative" breast tumor tissues. The second dataset contains 21 pairs of matched Dukes' B either "fresh frozen" or "stored in RNAlater preservative" colon tumor tissues. In both the datasets, each tumor tissue has $p = 22,283$ gene expression levels. Chowdary et al. (2006) used correlation analysis to demonstrate that tissues stored in RNAlater preservative can generate expression profiles similar to those produced by tissues that were snap-frozen and they suggested that prognostic signatures can be obtained from RNAlater preservative-suspended tissues. It is then of interest to check if the paired tissues of "fresh frozen" and "stored in RNAlater preservative" for these two datasets have the same mean gene expression levels, respectively.

Table 7 presents the test results for the first two datasets. It is seen that all the four tests indicate there is no strong evidence to reject the null hypothesis that the paired breast and colon tumor tissues have the same mean gene expression levels, respectively. These results are consistent with the findings of Chowdary et al. (2006). Note that both the estimated approximate degrees of freedom of the NEW test are less than 10. They indicate that the normal approximation used in the WPL, BS and CQ tests is less adequate so that their p-values are less trustful.

The third dataset was provided by Badea et al. (2008), available at GEO with accession number GSE15471. This dataset contains paired samples of pancreatic ductal adenocarcinoma tumors and matching normal pancreatic tissue from $n = 36$ pancreatic cancer patients, with each tissue

Table 8. Results for the third dataset, provided by Badea et al. (2008)

Method	Statistic	p-value	$\hat{d}$
NEW	11.98	9.57×10^{-49}	24.94
WPL	38.78	0.00	-
BS	35.83	2.24×10^{-281}	-
CQ	35.92	8.09×10^{-283}	-

having $p = 54,675$ gene expression levels. It is of interest if the paired tumor and normal tissues have same mean gene expression levels.

Table 8 presents the test results of the proposed NEW test, together with the WPL, BS and CQ tests. It is seen that all the four tests strongly reject the null hypothesis that the paired tumor and normal tissues have the same mean gene expression levels. Again, the number of the estimated approximate degrees of freedom of the NEW test is only 24.94, indicating that the normal approximation used in the WPL, BS and CQ tests is also less adequate so that the p-values of these three tests are also less reliable than that of the NEW test.

APPENDIX: TECHNICAL PROOFS

Proof of Theorem 1. Set $\boldsymbol{w}_n = \sqrt{n}\bar{\boldsymbol{z}}$. For any fixed finite p, by the central limit theorem, as $n \to \infty$, we have $\boldsymbol{w}_n \xrightarrow{\mathcal{L}} \boldsymbol{w}$ where $\boldsymbol{w} \sim \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{V}_p)$. By the continuous mapping theorem, we have $T_{n,p,0} = \|\boldsymbol{w}_n\|^2 \xrightarrow{\mathcal{L}} T_{p,0}$ where $T_{p,0} = \|\boldsymbol{w}\|^2 \stackrel{d}{=} \sum_{r=1}^p \lambda_{p,r} A_r$ with $A_1, \dots, A_r, \dots$ being i.i.d. χ_1^2 random variables.

We now prove the case when $n, p \to \infty$ via the characteristic function method. Let $\boldsymbol{u}_{p,1}, \dots, \boldsymbol{u}_{p,p}$ be the eigenvectors associated with the decreasing-ordered eigenvalues $\lambda_{p,1}, \dots, \lambda_{p,p}$ of $\boldsymbol{V}_p$. We have $\boldsymbol{w}_n = \sum_{r=1}^p \xi_{p,r} \boldsymbol{u}_{p,r}$ where $\xi_{p,r} = \boldsymbol{w}_n^\top \boldsymbol{u}_{p,r}$. It is known that $\xi_{p,r},\ r = 1, \dots, p$ are uncorrelated and $\mathrm{E}(\xi_{p,r}) = 0$ and $\mathrm{Var}(\xi_{p,r}) = \lambda_{p,r},\ r = 1, 2, \dots$. Note that $T_{n,p,0} = \sum_{r=1}^p \xi_{p,r}^2$. Set $T_{n,q,0} = \sum_{r=1}^q \xi_{p,r}^2$. Then we have

$$
\begin{aligned}
& |\psi_{T_{n,p,0}}(t) - \psi_{T_{n,q,0}}(t)| \\
\le\ & |t|\, \mathrm{E}\, |T_{n,p,0} - T_{n,q,0}| = |t| \textstyle\sum_{r=q+1}^p \mathrm{E}(\xi_{p,r}^2) \\
=\ & |t| \textstyle\sum_{r=q+1}^p \lambda_{p,r},
\end{aligned}
$$

which is valid for all large p. As $p \to \infty$, we have $T_{n,p,0} \to \sum_{r=1}^\infty \xi_{\infty,r}^2$ and the above result still holds with the upper bound $|t| \sum_{r=q+1}^p \lambda_{p,r}$ replaced by $|t| \sum_{r=q+1}^\infty \lambda_{\infty,r}$. Let t be fixed. Since $\sum_{r=1}^\infty \lambda_{\infty,r} = \mathrm{tr}(\boldsymbol{V}_\infty) = 1$, for any $\epsilon > 0$, there exist Q and N_1, both depending on t and ϵ, such that as $n \ge N_1$, we have $|\psi_{T_{n,p,0}}(t) - \psi_{T_{n,Q,0}}(t)| \le |t| \sum_{r=q+1}^\infty \lambda_{\infty,r} \le \epsilon$. For the fixed Q, by the central limit theorem we have $T_{n,Q,0} \xrightarrow{\mathcal{L}} T_{0,Q}$ where $T_{0,Q} \stackrel{d}{=} \sum_{r=1}^Q \lambda_{\infty,r} A_r$ since as $n, p \to \infty$, $\xi_{p,r} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \lambda_{\infty,r})$ and $\xi_{p,r}$'s are asymptotically independent. That is, there exists N_2, depending on t and ϵ, such that as $n > N_2$, we have $|\psi_{T_{n,Q,0}}(t) - \psi_{T_{0,Q}}(t)| \le \epsilon$. Note that $T_{\infty,0} = \sum_{r=1}^{\infty} \lambda_{\infty,r} A_r$. We have

$$\begin{aligned} & |\psi_{T_{0,Q}}(t) - \psi_{T_{\infty,0}}(t)| \\ \le \quad & |t| \sum_{r=Q+1}^{\infty} \lambda_{\infty,r} \operatorname{E}(A_r) \le |t| \sum_{r=Q+1}^{\infty} \lambda_{\infty,r} \le \epsilon. \end{aligned}$$

It follows that as $n \ge \min(N_1, N_2)$, we have $|\psi_{T_{n,p,0}}(t) - \psi_{T_{\infty,0}}(t)| \le 3\epsilon$. The theorem follows as we can let $\epsilon \to 0$. □

Proof of Theorem 2. If $T_{p,0}$ tends to normal as $p \to \infty$, its skewness $\sqrt{8/d^*}$ will tend to 0, showing that $d^* \to \infty$. On the other hand, when $d^* \to \infty$, by (2.4), we have $\Delta \to 0$. Then by (2.7), we have that $T_{p,0}$ tends normal. The first claim then follows. To show the second claim, note that when $d^* \to \infty$, by (2.7) again, we have $(T_{p,0} - 1)/\sqrt{2/d} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ and by (2.5), we have $d \to \infty$ so that $(R-1)/\sqrt{2/d} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$. The proof is completed. □

Proof of Lemma 1. First of all, we have $T_{n,p} = n^{-1}\sum_{i=1}^n \boldsymbol{z}_i^\top \boldsymbol{z}_i + n^{-1}\sum_{i\neq j}^n \boldsymbol{z}_i^\top \boldsymbol{z}_j$. Since $\boldsymbol{z}_i^\top \boldsymbol{z}_i = 1$ and under H_0, we have $\operatorname{E}(\boldsymbol{z}_i) = \operatorname{E}\{U(\boldsymbol{\epsilon}_i)\} = \boldsymbol{0}$, it is easy to see that $\operatorname{E}(T_{n,p,0}) = 1$. By Lemma 3 of Zhang et al. (2015), under H_0, we have $\operatorname{Var}(n\|\bar{\boldsymbol{z}}\|^2) = n^{-2}\operatorname{Var}\left[\{\sum_{i=1}^n U(\boldsymbol{\epsilon}_i)^\top\}\{\sum_{i=1}^n U(\boldsymbol{\epsilon}_i)\}\right] = n^{-1}\kappa_{\boldsymbol{z},11} + 2\operatorname{tr}\left[\operatorname{E}\{U(\boldsymbol{\epsilon}_1)U(\boldsymbol{\epsilon}_1)^\top\}\right]^2 = n^{-1}\kappa_{\boldsymbol{z},11} + 2\operatorname{tr}(\boldsymbol{V}_p^2)$, where $\kappa_{\boldsymbol{z},11} = \operatorname{E}\|U(\boldsymbol{\epsilon}_1)\|^4 - \operatorname{tr}^2\left[\operatorname{E}\{U(\boldsymbol{\epsilon}_1)U(\boldsymbol{\epsilon}_1)^\top\}\right] - 2\operatorname{tr}\left[\operatorname{E}\{U(\boldsymbol{\epsilon}_1)U(\boldsymbol{\epsilon}_1)^\top\}\right]^2 = 1 - 1 - 2\operatorname{tr}(\boldsymbol{V}_p^2) = -2\operatorname{tr}(\boldsymbol{V}_p^2)$, since $\|U(\boldsymbol{\epsilon}_1)\|^2 = 1$. Thus under H_0 we have $\operatorname{Var}(n\|\bar{\boldsymbol{z}}\|^2) = 2(1 - \frac{1}{n})\operatorname{tr}(\boldsymbol{V}_p^2) = \frac{2(n-1)}{n}\operatorname{tr}(\boldsymbol{V}_p^2)$ as desired. □

Throughout this section, we define $P_{n,k} = n(n-1)\dots(n-k+1)$. The following lemma with its proof omitted holds obviously. It will be used in the proofs of Theorem 3 and Lemmas 3 and 4 stated later.

Lemma 2. *Denote $\sum_{i_1,\dots,i_m}^* = \sum_{i_1,\dots,i_m} 1\{i_1 \neq \dots \neq i_m\}$ as summation over mutually different indices $\{i_1,\dots,i_m\}$. For any function $f(j,k,j_1,k_1)$ of indices $\{j,k,j_1,k_1\}$, we have*

$$\begin{aligned} \sum_{j\neq k}\sum_{j_1\neq k_1} f(j,k,j_1,k_1) &= \sum_{j,k,j_1,k_1}^* f(j,k,j_1,k_1) \\ &+ \sum_{j,k,k_1}^* f(j,k,j,k_1) + \sum_{j,k,j_1}^* f(j,k,j_1,j) \\ &+ \sum_{j,k,k_1}^* f(j,k,k,k_1) + \sum_{j,k,j_1}^* f(j,k,j_1,k) \\ &+ \sum_{j,k}^* f(j,k,j,k) + \sum_{j,k}^* f(j,k,k,j). \end{aligned}$$

In addition, if $f(j,k,j_1,k_1) = f(k,j,j_1,k_1)$ and $f(j,k,j_1,k_1) = f(j,k,k_1,j_1)$, we have

$$\begin{aligned} \sum_{j_1\neq k_1}\sum_{j\neq k} f(j_1,k_1,j,k) &= \sum_{j,k,j_1,k_1}^* f(j,k,j_1,k_1) \\ &+ 4\sum_{j,k,k_1}^* f(j,k,j,k_1) + 2\sum_{j,k}^* f(j,k,j,k). \end{aligned}$$

Furthermore, if $f(j,k,j_1,k_1) = f(j_1,k_1,j,k)$, we have

$$\begin{aligned} \sum_{j_1\neq k_1}\sum_{j\neq k} f(j_1,k_1,j,k) &= \sum_{j,k,j_1,k_1}^* f(j,k,j_1,k_1) \\ &+ \sum_{j,k,k_1}^* f(j,k,j,k_1) + 2\sum_{j,k,j_1}^* f(j,k,j_1,j) + \sum_{j,k,j_1}^* f(j,k,j_1,k) \\ &+ \sum_{j,k}^* f(j,k,j,k) + \sum_{j,k}^* f(j,k,k,j). \end{aligned}$$

Proof of Theorem 3. By (1.10), we have

$$\begin{aligned} \widehat{\operatorname{tr}(\boldsymbol{V}_p^2)} =& P_{n,2}^{-1}\sum_{j\neq k}(\boldsymbol{z}_j^\top \boldsymbol{z}_k \boldsymbol{z}_k^\top \boldsymbol{z}_j - 2\boldsymbol{z}_j^\top \bar{\boldsymbol{z}}_{(j,k)} \boldsymbol{z}_k^\top \boldsymbol{z}_j \\ &+ \boldsymbol{z}_j^\top \bar{\boldsymbol{z}}_{(j,k)} \boldsymbol{z}_k^\top \bar{\boldsymbol{z}}_{(j,k)}) \\ :=& B_1 - 2B_2 + B_3, \end{aligned}$$

where

$$\begin{aligned} B_1 =& P_{n,2}^{-1}\sum_{j\neq k}(\boldsymbol{z}_j^\top \boldsymbol{z}_k \boldsymbol{z}_k^\top \boldsymbol{z}_j), \\ B_2 =& P_{n,2}^{-1}\sum_{j\neq k}(\boldsymbol{z}_j^\top \bar{\boldsymbol{z}}_{(j,k)} \boldsymbol{z}_k^\top \boldsymbol{z}_j), \\ B_3 =& P_{n,2}^{-1}\sum_{j\neq k}(\boldsymbol{z}_j^\top \bar{\boldsymbol{z}}_{(j,k)} \boldsymbol{z}_k^\top \bar{\boldsymbol{z}}_{(j,k)}). \end{aligned}$$

We firstly show the first claim of Theorem 3. Note that under H_0, by independence of $\boldsymbol{z}_j$, $\boldsymbol{z}_k$ and $\bar{\boldsymbol{z}}_{(j,k)}$, we have $\operatorname{E}(\boldsymbol{z}_j) = \operatorname{E}(\bar{\boldsymbol{z}}_{(j,k)}) = \boldsymbol{0}$. It follows that

$$\text{(A.1)} \qquad \begin{aligned} \operatorname{E}(B_1) &= P_{n,2}^{-1}\operatorname{tr}\left(\sum_{j\neq k} \boldsymbol{V}_p \boldsymbol{V}_p\right) = \operatorname{tr}(\boldsymbol{V}_p^2), \\ \operatorname{E}(B_2) &= \operatorname{E}(B_3) = 0. \end{aligned}$$

The first claim of Theorem 3 then follows immediately.

We now show the second claim of the theorem. By Lemmas 3 and 4 (stated and proved later), we only need to show $B_1/\operatorname{tr}(\boldsymbol{V}_p^2) \xrightarrow{P} 1$. By (A.1), the above expression holds if we can show

$$\text{(A.2)} \qquad \operatorname{Var}(B_1) = o\left\{\operatorname{tr}^2(\boldsymbol{V}_p^2)\right\}.$$

Now, we have

$$\begin{aligned} \operatorname{E}(B_1^2) =& P_{n,2}^{-2}\operatorname{E}\sum_{j\neq k}\sum_{j_1\neq k_1}(\boldsymbol{z}_j^\top \boldsymbol{z}_k \boldsymbol{z}_k^\top \boldsymbol{z}_j)(\boldsymbol{z}_{j_1}^\top \boldsymbol{z}_{k_1} \boldsymbol{z}_{k_1}^\top \boldsymbol{z}_{j_1}) \\ =& P_{n,2}^{-2}\operatorname{E}\sum_{j,k,j_1,k_1}^*(\boldsymbol{z}_j^\top \boldsymbol{z}_k \boldsymbol{z}_k^\top \boldsymbol{z}_j \boldsymbol{z}_{j_1}^\top \boldsymbol{z}_{k_1} \boldsymbol{z}_{k_1}^\top \boldsymbol{z}_{j_1}) \end{aligned}$$

$$
\begin{aligned}
&+4P_{n,2}^{-2}\,\mathrm{E}\sum_{j,k,k_1}^{*}(\boldsymbol{z}_j^\top\boldsymbol{z}_k\boldsymbol{z}_k^\top\boldsymbol{z}_j\boldsymbol{z}_j^\top\boldsymbol{z}_{k_1}\boldsymbol{z}_{k_1}^\top\boldsymbol{z}_j)\\
&+2P_{n,2}^{-2}\,\mathrm{E}\sum_{j,k}^{*}(\boldsymbol{z}_j^\top\boldsymbol{z}_k\boldsymbol{z}_k^\top\boldsymbol{z}_j\boldsymbol{z}_j^\top\boldsymbol{z}_k\boldsymbol{z}_k^\top\boldsymbol{z}_j)\\
=&\frac{(n-2)(n-3)}{n(n-1)}\,\mathrm{tr}^2(\boldsymbol{V}_p^2)+\frac{4(n-2)}{n(n-1)}\,\mathrm{E}(\boldsymbol{z}_1^\top\boldsymbol{V}_p\boldsymbol{z}_1)^2\\
&+\frac{2}{n(n-1)}\,\mathrm{E}(\boldsymbol{z}_1^\top\boldsymbol{z}_2)^4.
\end{aligned}
$$

By Condition A, we have (A.2). The theorem is then proved. □

Lemma 3. *Under H_0 and Condition A, we have $B_2 = o\left\{\mathrm{tr}(\boldsymbol{V}_p^2)\right\}$.*

Proof of Lemma 3. By (A.1), we only need to show $\mathrm{E}(B_2^2) = o\left\{\mathrm{tr}^2(\boldsymbol{V}_p^2)\right\}$. Note that

$$
\begin{aligned}
\mathrm{E}(B_2^2)=&P_{n,2}^{-2}\,\mathrm{E}\sum_{j\neq k}\sum_{j_1\neq k_1}(\boldsymbol{z}_j^\top\bar{\boldsymbol{z}}_{(j,k)}\boldsymbol{z}_k^\top\boldsymbol{z}_j)(\boldsymbol{z}_{j_1}^\top\bar{\boldsymbol{z}}_{(j_1,k_1)}\boldsymbol{z}_{k_1}^\top\boldsymbol{z}_{j_1})\\
=&P_{n,2}^{-2}\,\mathrm{E}\sum_{j,k,j_1,k_1}^{*}(\boldsymbol{z}_j^\top\bar{\boldsymbol{z}}_{(j,k)}\boldsymbol{z}_k^\top\boldsymbol{z}_j)(\boldsymbol{z}_{j_1}^\top\bar{\boldsymbol{z}}_{(j_1,k_1)}\boldsymbol{z}_{k_1}^\top\boldsymbol{z}_{j_1})\\
&+P_{n,2}^{-2}\,\mathrm{E}\sum_{j,k,k_1}^{*}(\boldsymbol{z}_j^\top\bar{\boldsymbol{z}}_{(j,k)}\boldsymbol{z}_k^\top\boldsymbol{z}_j)(\boldsymbol{z}_j^\top\bar{\boldsymbol{z}}_{(j,k_1)}\boldsymbol{z}_{k_1}^\top\boldsymbol{z}_j)\\
&+2P_{n,2}^{-2}\,\mathrm{E}\sum_{j,k,j_1}^{*}(\boldsymbol{z}_j^\top\bar{\boldsymbol{z}}_{(j,k)}\boldsymbol{z}_k^\top\boldsymbol{z}_j)(\boldsymbol{z}_{j_1}^\top\bar{\boldsymbol{z}}_{(j_1,j)}\boldsymbol{z}_j^\top\boldsymbol{z}_{j_1})\\
&+P_{n,2}^{-2}\sum_{j,k,j_1}^{*}(\boldsymbol{z}_j^\top\bar{\boldsymbol{z}}_{(j,k)}\boldsymbol{z}_k^\top\boldsymbol{z}_j)(\boldsymbol{z}_{j_1}^\top\bar{\boldsymbol{z}}_{(j_1,k)}\boldsymbol{z}_k^\top\boldsymbol{z}_{j_1})\\
&+P_{n,2}^{-2}\,\mathrm{E}\sum_{j,k}^{*}(\boldsymbol{z}_j^\top\bar{\boldsymbol{z}}_{(j,k)}\boldsymbol{z}_k^\top\boldsymbol{z}_j)(\boldsymbol{z}_j^\top\bar{\boldsymbol{z}}_{(j,k)}\boldsymbol{z}_k^\top\boldsymbol{z}_j)\\
&+P_{n,2}^{-2}\,\mathrm{E}\sum_{j,k}^{*}(\boldsymbol{z}_j^\top\bar{\boldsymbol{z}}_{(j,k)}\boldsymbol{z}_k^\top\boldsymbol{z}_j)(\boldsymbol{z}_k^\top\bar{\boldsymbol{z}}_{(k,j)}\boldsymbol{z}_j^\top\boldsymbol{z}_k)\\
:=&B_{21}+B_{22}+2B_{23}+B_{24}+B_{25}+B_{26},
\end{aligned}
$$

where $B_{21}=B_{22}=0$. We now show B_{23}, B_{24}, B_{25} and B_{26} are $o\left\{\mathrm{tr}^2(\boldsymbol{V}_p^2)\right\}$. Notice that $\bar{\boldsymbol{z}}_{(j,k)}=\{(n-3)\bar{\boldsymbol{z}}_{(j,k,j_1)}+\boldsymbol{z}_{j_1}\}/(n-2)$, $\bar{\boldsymbol{z}}_{(j_1,j)}=\{(n-3)\bar{\boldsymbol{z}}_{(j,k,j_1)}+\boldsymbol{z}_k\}/(n-2)$, and $\bar{\boldsymbol{z}}_{(j_1,k)}=\{(n-3)\bar{\boldsymbol{z}}_{(j,k,j_1)}+\boldsymbol{z}_j\}/(n-2)$. First of all, we have

$$
\begin{aligned}
B_{23}=&P_{n,2}^{-2}\,\mathrm{E}\sum_{j,k,j_1}^{*}(\boldsymbol{z}_j^\top\bar{\boldsymbol{z}}_{(j,k)}\boldsymbol{z}_k^\top\boldsymbol{z}_j)(\boldsymbol{z}_{j_1}^\top\bar{\boldsymbol{z}}_{(j_1,j)}\boldsymbol{z}_j^\top\boldsymbol{z}_{j_1})\\
=&P_{n,3}^{-2}\,\mathrm{E}\sum_{j,k,j_1}^{*}\left[\boldsymbol{z}_j^\top\{(n-3)\bar{\boldsymbol{z}}_{(j,k,j_1)}+\boldsymbol{z}_{j_1}\}\boldsymbol{z}_k^\top\boldsymbol{z}_j\right]\\
&\left[\boldsymbol{z}_{j_1}^\top\{(n-3)\bar{\boldsymbol{z}}_{(j,k,j_1)}+\boldsymbol{z}_k\}\boldsymbol{z}_j^\top\boldsymbol{z}_{j_1}\right]\\
:=&B_{231}+B_{232}+B_{233}+B_{234},
\end{aligned}
$$

with $B_{231}=B_{232}=B_{233}=0$, and

$$
\begin{aligned}
B_{234}=&P_{n,3}^{-2}\,\mathrm{E}\sum_{j,k,j_1}^{*}(\boldsymbol{z}_j^\top\boldsymbol{z}_{j_1}\boldsymbol{z}_k^\top\boldsymbol{z}_j\boldsymbol{z}_{j_1}^\top\boldsymbol{z}_k\boldsymbol{z}_j^\top\boldsymbol{z}_{j_1})\\
=&P_{n,3}^{-1}\,\mathrm{E}\left\{(\boldsymbol{z}_1^\top\boldsymbol{z}_2)^2(\boldsymbol{z}_1^\top\boldsymbol{V}_p\boldsymbol{z}_2)\right\}\\
\leq&P_{n,3}^{-1}\sqrt{\mathrm{E}(\boldsymbol{z}_1^\top\boldsymbol{z}_2)^4\,\mathrm{E}(\boldsymbol{z}_1^\top\boldsymbol{V}_p\boldsymbol{z}_2)^2}\\
=&o\left\{\mathrm{tr}^2(\boldsymbol{V}_p^2)\right\},
\end{aligned}
$$

where note $\mathrm{E}(\boldsymbol{z}_1^\top\boldsymbol{V}_p\boldsymbol{z}_2)^2=\mathrm{tr}\left\{\mathrm{E}(\boldsymbol{z}_2^\top\boldsymbol{V}_p\boldsymbol{z}_1\boldsymbol{z}_1^\top\boldsymbol{V}_p\boldsymbol{z}_2)\right\}=\mathrm{tr}(\boldsymbol{V}_p^4)$, and the last equality is due to the fact $\mathrm{tr}(\boldsymbol{V}_p^4)\leq\mathrm{tr}^2(\boldsymbol{V}_p^2)$ and Condition A. Secondly, we have

$$
\begin{aligned}
B_{24}=&P_{n,2}^{-2}\,\mathrm{E}\sum_{j,k,j_1}^{*}(\boldsymbol{z}_j^\top\bar{\boldsymbol{z}}_{(j,k)}\boldsymbol{z}_k^\top\boldsymbol{z}_j)(\boldsymbol{z}_{j_1}^\top\bar{\boldsymbol{z}}_{(j_1,k)}\boldsymbol{z}_k^\top\boldsymbol{z}_{j_1})\\
=&P_{n,3}^{-2}\,\mathrm{E}\sum_{j,k,j_1}^{*}\left[\boldsymbol{z}_j^\top\{(n-3)\bar{\boldsymbol{z}}_{(j,k,j_1)}+\boldsymbol{z}_{j_1}\}\boldsymbol{z}_k^\top\boldsymbol{z}_j\right]\\
&\left[\boldsymbol{z}_{j_1}^\top\{(n-3)\bar{\boldsymbol{z}}_{(j,k,j_1)}+\boldsymbol{z}_j\}\boldsymbol{z}_k^\top\boldsymbol{z}_{j_1}\right]\\
:=&B_{241}+B_{242}+B_{243}+B_{244},
\end{aligned}
$$

with

$$
\begin{aligned}
B_{241}=&P_{n,3}^{-2}\,\mathrm{E}\sum_{j,k,j_1}^{*}\left\{(n-3)^2\boldsymbol{z}_j^\top\bar{\boldsymbol{z}}_{(j,k,j_1)}\right.\\
&\left.\times\boldsymbol{z}_k^\top\boldsymbol{z}_j\boldsymbol{z}_{j_1}^\top\bar{\boldsymbol{z}}_{(j,k,j_1)}\boldsymbol{z}_k^\top\boldsymbol{z}_{j_1}\right\}\\
=&(n-3)P_{n,3}^{-1}\,\mathrm{tr}(\boldsymbol{V}_p^4)=o\left\{\mathrm{tr}^2(\boldsymbol{V}_p^2)\right\},\\
B_{242}=&B_{243}=0,\\
B_{244}=&P_{n,3}^{-2}\,\mathrm{E}\sum_{j,k,j_1}^{*}(\boldsymbol{z}_j^\top\boldsymbol{z}_{j_1}\boldsymbol{z}_k^\top\boldsymbol{z}_j\boldsymbol{z}_{j_1}^\top\boldsymbol{z}_j\boldsymbol{z}_k^\top\boldsymbol{z}_{j_1})\\
=&P_{n,3}^{-1}\,\mathrm{E}\left\{(\boldsymbol{z}_1^\top\boldsymbol{z}_2)^2(\boldsymbol{z}_1^\top\boldsymbol{V}_p\boldsymbol{z}_2)\right\}\\
=&o\left\{\mathrm{tr}^2(\boldsymbol{V}_p^2)\right\}.
\end{aligned}
$$

Thirdly, we have

$$
\begin{aligned}
B_{25}=&P_{n,2}^{-2}\,\mathrm{E}\sum_{j,k}^{*}(\boldsymbol{z}_j^\top\bar{\boldsymbol{z}}_{(j,k)}\boldsymbol{z}_k^\top\boldsymbol{z}_j)(\boldsymbol{z}_j^\top\bar{\boldsymbol{z}}_{(j,k)}\boldsymbol{z}_k^\top\boldsymbol{z}_j)\\
=&P_{n,2}^{-1}(n-3)^{-1}\,\mathrm{E}(\boldsymbol{z}_1^\top\boldsymbol{V}_p\boldsymbol{z}_1)^2=o\left\{\mathrm{tr}^2(\boldsymbol{V}_p^2)\right\}.
\end{aligned}
$$

And finally, we have

$$
\begin{aligned}
B_{26}=&P_{n,2}^{-2}\,\mathrm{E}\sum_{j,k}^{*}(\boldsymbol{z}_j^\top\bar{\boldsymbol{z}}_{(j,k)}\boldsymbol{z}_k^\top\boldsymbol{z}_j)(\boldsymbol{z}_k^\top\bar{\boldsymbol{z}}_{(k,j)}\boldsymbol{z}_j^\top\boldsymbol{z}_k)\\
=&P_{n,3}^{-2}\,\mathrm{E}\sum_{j,k}^{*}\left[\boldsymbol{z}_j^\top\{(n-3)\bar{\boldsymbol{z}}_{(j,k,j_1)}+\boldsymbol{z}_{j_1}\}\boldsymbol{z}_k^\top\boldsymbol{z}_j\right]\\
&\left[\boldsymbol{z}_k^\top\{(n-3)\bar{\boldsymbol{z}}_{(j,k,j_1)}+\boldsymbol{z}_{j_1}\}\boldsymbol{z}_j^\top\boldsymbol{z}_k\right]\\
:=&B_{261}+B_{262}+B_{263}+B_{264},
\end{aligned}
$$

with $B_{262}=B_{263}=0$,

$$
\begin{aligned}
B_{261} =& P_{n,3}^{-2} \mathrm{E} \sum_{j,k}^{*} \left[\boldsymbol{z}_j^\top \{(n-3)\bar{\boldsymbol{z}}_{(j,k,j_1)}\} \boldsymbol{z}_k^\top \boldsymbol{z}_j \right] \\
& \left[\boldsymbol{z}_k^\top \{(n-3)\bar{\boldsymbol{z}}_{(j,k,j_1)}\} \boldsymbol{z}_j^\top \boldsymbol{z}_k \right] \\
=& P_{n,2}(n-3) P_{n,3}^{-2} \mathrm{E} \left\{ (\boldsymbol{z}_1^\top \boldsymbol{V}_p \boldsymbol{z}_2)(\boldsymbol{z}_1^\top \boldsymbol{z}_2)^2 \right\} \\
=& o\left\{ \mathrm{tr}^2(\boldsymbol{V}_p^2) \right\}, \\
B_{264} =& P_{n,3}^{-2} \mathrm{E} \sum_{j,k}^{*} (\boldsymbol{z}_k^\top \boldsymbol{z}_j \boldsymbol{z}_k^\top \boldsymbol{z}_{j_1} \boldsymbol{z}_{j_1}^\top \boldsymbol{z}_j \boldsymbol{z}_j^\top \boldsymbol{z}_k) \\
=& P_{n,3}^{-2} P_{n,2} \mathrm{E} \left\{ (\boldsymbol{z}_1^\top \boldsymbol{z}_2)^2 \boldsymbol{z}_1^\top \boldsymbol{V}_p \boldsymbol{z}_2) \right\} \\
=& o\left\{ \mathrm{tr}^2(\boldsymbol{V}_p^2) \right\}.
\end{aligned}
$$

The proof is then completed. □

Lemma 4. *Under H_0 and Condition A, we have $B_3 = o\left\{\mathrm{tr}(\boldsymbol{V}_p^2)\right\}$.*

Proof of Lemma 4.. By (A.1), we only need to show $\mathrm{E}(B_3^2) = o\left\{\mathrm{tr}^2(\boldsymbol{V}_p^2)\right\}$. Now

$$
\begin{aligned}
\mathrm{E}(B_3^2) =& P_{n,2}^{-2} \mathrm{E} \sum_{j \neq k} (\boldsymbol{z}_j^\top \bar{\boldsymbol{z}}_{(j,k)} \boldsymbol{z}_k^\top \bar{\boldsymbol{z}}_{(j,k)})(\boldsymbol{z}_{j_1}^\top \bar{\boldsymbol{z}}_{(j_1,k_1)} \boldsymbol{z}_{k_1}^\top \bar{\boldsymbol{z}}_{(j_1,k_1)}) \\
=& P_{n,2}^{-2} \mathrm{E} \sum_{j,k,j_1,k_1}^{*} (\boldsymbol{z}_j^\top \bar{\boldsymbol{z}}_{(j,k)} \boldsymbol{z}_k^\top \bar{\boldsymbol{z}}_{(j,k)}) \\
& (\boldsymbol{z}_{j_1}^\top \bar{\boldsymbol{z}}_{(j_1,k_1)} \boldsymbol{z}_{k_1}^\top \bar{\boldsymbol{z}}_{(j_1,k_1)}) \\
& + 4P_{n,2}^{-2} \mathrm{E} \sum_{j,k,k_1}^{*} (\boldsymbol{z}_j^\top \bar{\boldsymbol{z}}_{(j,k)} \boldsymbol{z}_k^\top \bar{\boldsymbol{z}}_{(j,k)}) \\
& (\boldsymbol{z}_j^\top \bar{\boldsymbol{z}}_{(j,k_1)} \boldsymbol{z}_{k_1}^\top \bar{\boldsymbol{z}}_{(j,k_1)}) \\
& + 2P_{n,2}^{-2} \mathrm{E} \sum_{j,k}^{*} (\boldsymbol{z}_j^\top \bar{\boldsymbol{z}}_{(j,k)} \boldsymbol{z}_k^\top \bar{\boldsymbol{z}}_{(j,k)})(\boldsymbol{z}_j^\top \bar{\boldsymbol{z}}_{(j,k)} \boldsymbol{z}_k^\top \bar{\boldsymbol{z}}_{(j,k)}) \\
:=& B_{31} + 4B_{32} + 2B_{33}.
\end{aligned}
$$

Note that $\bar{\boldsymbol{z}}_{(j,k)} = \{(n-4)\bar{\boldsymbol{z}}_{(j,k,j_1,k_1)} + \boldsymbol{z}_{j_1} + \boldsymbol{z}_{k_1}\}/(n-2)$ and $\bar{\boldsymbol{z}}_{(j_1,k_1)} = \{(n-4)\bar{\boldsymbol{z}}_{(j,k,j_1,k_1)} + \boldsymbol{z}_j + \boldsymbol{z}_k\}/(n-2)$, we have

$$
\begin{aligned}
B_{31} =& P_{n,2}^{-2} \mathrm{E} \sum_{j,k,j_1,k_1}^{*} (\boldsymbol{z}_j^\top \bar{\boldsymbol{z}}_{(j,k)} \boldsymbol{z}_k^\top \bar{\boldsymbol{z}}_{(j,k)}) \\
& (\boldsymbol{z}_{j_1}^\top \bar{\boldsymbol{z}}_{(j_1,k_1)} \boldsymbol{z}_{k_1}^\top \bar{\boldsymbol{z}}_{(j_1,k_1)}) \\
=& P_{n,2}^{-2}(n-2)^{-4} \mathrm{E} \sum_{j,k,j_1,k_1}^{*} \left[\boldsymbol{z}_j^\top \{(n-4)\bar{\boldsymbol{z}}_{(j,k,j_1,k_1)} \right. \\
& \left. + \boldsymbol{z}_{j_1} + \boldsymbol{z}_{k_1}\} \boldsymbol{z}_k^\top \{(n-4)\bar{\boldsymbol{z}}_{(j,k,j_1,k_1)} + \boldsymbol{z}_{j_1} + \boldsymbol{z}_{k_1}\} \right] \\
& \left[\boldsymbol{z}_{j_1}^\top \{(n-4)\bar{\boldsymbol{z}}_{(j,k,j_1,k_1)} + \boldsymbol{z}_j + \boldsymbol{z}_k\} \right. \\
& \left. \boldsymbol{z}_{k_1}^\top \{(n-4)\bar{\boldsymbol{z}}_{(j,k,j_1,k_1)} + \boldsymbol{z}_j + \boldsymbol{z}_k\} \right] \\
=& 2P_{n,2}^{-2}(n-2)^{-4} \mathrm{E} \sum_{j,k,j_1,k_1}^{*} (\boldsymbol{z}_j^\top \boldsymbol{z}_{j_1} \boldsymbol{z}_k^\top \boldsymbol{z}_{k_1} \boldsymbol{z}_{j_1}^\top \boldsymbol{z}_j \boldsymbol{z}_{k_1}^\top \boldsymbol{z}_k) \\
& + 2P_{n,2}^{-2}(n-2)^{-4} \\
& \mathrm{E} \sum_{j,k,j_1,k_1}^{*} (\boldsymbol{z}_j^\top \boldsymbol{z}_{k_1} \boldsymbol{z}_k^\top \boldsymbol{z}_{j_1} \boldsymbol{z}_{j_1}^\top \boldsymbol{z}_j \boldsymbol{z}_{k_1}^\top \boldsymbol{z}_k) \\
:=& 2B_{311} + 2B_{312},
\end{aligned}
$$

with

$$
\begin{aligned}
B_{311} =& P_{n,2}^{-2}(n-2)^{-4} \mathrm{E} \sum_{j,k,j_1,k_1}^{*} (\boldsymbol{z}_j^\top \boldsymbol{z}_{j_1} \boldsymbol{z}_{j_1}^\top \boldsymbol{z}_j \boldsymbol{z}_{k_1}^\top \boldsymbol{z}_k \boldsymbol{z}_k^\top \boldsymbol{z}_{k_1}) \\
=& P_{n,2}^{-2}(n-2)^{-4} P_{n,4} \mathrm{tr}^2(\boldsymbol{V}_p^2) = o\left\{ \mathrm{tr}^2(\boldsymbol{V}_p^2) \right\}, \\
B_{312} =& P_{n,2}^{-2}(n-2)^{-4} \mathrm{E} \sum_{j,k,j_1,k_1}^{*} (\boldsymbol{z}_j^\top \boldsymbol{z}_{k_1} \boldsymbol{z}_k^\top \boldsymbol{z}_{j_1} \boldsymbol{z}_{j_1}^\top \boldsymbol{z}_j \boldsymbol{z}_{k_1}^\top \boldsymbol{z}_k) \\
=& P_{n,2}^{-1}(n-2)^{-3}(n-3) \mathrm{tr}(\boldsymbol{V}_p^4) = o\left\{ \mathrm{tr}^2(\boldsymbol{V}_p^2) \right\}.
\end{aligned}
$$

Note that $\bar{\boldsymbol{z}}_{(j,k)} = \{(n-3)\bar{\boldsymbol{z}}_{(j,k,k_1)} + \boldsymbol{z}_{k_1}\}/(n-2)$ and $\bar{\boldsymbol{z}}_{(j,k_1)} = \{(n-3)\bar{\boldsymbol{z}}_{(j,k,k_1)} + \boldsymbol{z}_k\}/(n-2)$, we have

$$
\begin{aligned}
B_{32} =& P_{n,2}^{-2} \mathrm{E} \sum_{j,k,k_1}^{*} (\boldsymbol{z}_k^\top \bar{\boldsymbol{z}}_{(j,k)} \bar{\boldsymbol{z}}_{(j,k)}^\top \boldsymbol{V}_p \bar{\boldsymbol{z}}_{(j,k_1)} \bar{\boldsymbol{z}}_{(j,k_1)}^\top \boldsymbol{z}_{k_1}) \\
=& P_{n,2}^{-2}(n-2)^{-4} \mathrm{E} \sum_{j,k,k_1}^{*} \left[\boldsymbol{z}_k^\top \{(n-3)\bar{\boldsymbol{z}}_{(j,k,k_1)} + \boldsymbol{z}_{k_1}\} \right. \\
& \left. \{(n-3)\bar{\boldsymbol{z}}_{(j,k,k_1)} + \boldsymbol{z}_{k_1}\}^\top \right] \left[\boldsymbol{V}_p \{(n-3)\bar{\boldsymbol{z}}_{(j,k,k_1)} + \boldsymbol{z}_k\} \right. \\
& \left. \{(n-3)\bar{\boldsymbol{z}}_{(j,k,k_1)} + \boldsymbol{z}_k\}^\top \boldsymbol{z}_{k_1} \right] \\
=& P_{n,2}^{-2}(n-2)^{-4} \mathrm{E} \sum_{j,k,k_1}^{*} \left[\boldsymbol{z}_k^\top \{(n-3)\boldsymbol{z}_{k_1} \bar{\boldsymbol{z}}_{(j,k,k_1)}^\top \right. \\
& \left. + 2(n-3)\bar{\boldsymbol{z}}_{(j,k,k_1)} \boldsymbol{z}_{k_1}^\top + \boldsymbol{z}_{k_1} \boldsymbol{z}_{k_1}^\top \} \right] \\
& \times \left[\boldsymbol{V}_p \{(n-3)\bar{\boldsymbol{z}}_{(j,k,k_1)} \boldsymbol{z}_k^\top \right. \\
& \left. + 2(n-3)\boldsymbol{z}_k \bar{\boldsymbol{z}}_{(j,k,k_1)}^\top + \boldsymbol{z}_k \boldsymbol{z}_k^\top \} \boldsymbol{z}_{k_1} \right] \\
:=& B_{321} + 2B_{322} + B_{323} + 2B_{324} + 4B_{325} + 2B_{326} \\
& + B_{327} + 2B_{328} + B_{329},
\end{aligned}
$$

with

$$
\begin{aligned}
B_{321} =& P_{n,2}^{-2}(n-2)^{-4}(n-3)^2 \\
& \mathrm{E} \sum_{j,k,k_1}^{*} \left(\boldsymbol{z}_k^\top \boldsymbol{z}_{k_1} \bar{\boldsymbol{z}}_{(j,k,k_1)}^\top \boldsymbol{V}_p \bar{\boldsymbol{z}}_{(j,k,k_1)} \boldsymbol{z}_k^\top \boldsymbol{z}_{k_1} \right) \\
=& P_{n,2}^{-2}(n-2)^{-4} P_{n,4} \mathrm{tr}^2(\boldsymbol{V}_p^2) = o\left\{ \mathrm{tr}^2(\boldsymbol{V}_p^2) \right\}, \\
B_{322} =& P_{n,2}^{-2}(n-2)^{-4}(n-3)^2 \\
& \mathrm{E} \sum_{j,k,k_1}^{*} \left(\boldsymbol{z}_k^\top \boldsymbol{z}_{k_1} \bar{\boldsymbol{z}}_{(j,k,k_1)}^\top \boldsymbol{V}_p \boldsymbol{z}_k \bar{\boldsymbol{z}}_{(j,k,k_1)}^\top \boldsymbol{z}_{k_1} \right) \\
=& P_{n,2}^{-2}(n-2)^{-4} P_{n,4} \mathrm{tr}(\boldsymbol{V}_p^4) = o\left\{ \mathrm{tr}^2(\boldsymbol{V}_p^2) \right\}, \\
B_{323} =& P_{n,2}^{-2}(n-2)^{-4} \\
& \mathrm{E} \sum_{j,k,k_1}^{*} \left[\boldsymbol{z}_k^\top \{(n-3)\boldsymbol{z}_{k_1} \bar{\boldsymbol{z}}_{(j,k,k_1)}^\top\} (\boldsymbol{V}_p \boldsymbol{z}_k \boldsymbol{z}_k^\top \boldsymbol{z}_{k_1}) \right] = 0,
\end{aligned}
$$

$$
\begin{aligned}
B_{324} =& P_{n,2}^{-2}(n-2)^{-4}\,\mathrm{E}\sum_{j,k,k_1}^{*}\left(\left[\boldsymbol{z}_k^\top\{(n-3)\bar{\boldsymbol{z}}_{(j,k,k_1)}\boldsymbol{z}_{k_1}^\top\}\right]\right.\\
&\left.\left[\boldsymbol{V}_p\{(n-3)\bar{\boldsymbol{z}}_{(j,k,k_1)}\boldsymbol{z}_k^\top\}\boldsymbol{z}_{k_1}\right]\right)\\
=& P_{n,2}^{-2}(n-2)^{-4}P_{n,4}\operatorname{tr}(\boldsymbol{V}_p^4) = o\left\{\operatorname{tr}^2(\boldsymbol{V}_p^2)\right\},\\
B_{325} =& P_{n,2}^{-2}(n-2)^{-4}(n-3)^2\\
&\mathrm{E}\sum_{j,k,k_1}^{*}\left(\boldsymbol{z}_k^\top\bar{\boldsymbol{z}}_{(j,k,k_1)}\boldsymbol{z}_{k_1}^\top\boldsymbol{V}_p\boldsymbol{z}_k\bar{\boldsymbol{z}}_{(j,k,k_1)}^\top\boldsymbol{z}_{k_1}\right)\\
=& P_{n,2}^{-2}(n-2)^{-4}P_{n,4}\operatorname{tr}(\boldsymbol{V}_p^4) = o\left\{\operatorname{tr}^2(\boldsymbol{V}_p^2)\right\},
\end{aligned}
$$

and

$$
\begin{aligned}
B_{326} =& P_{n,2}^{-2}(n-2)^{-4}\,\mathrm{E}\sum_{j,k,k_1}^{*}\left(\left[\boldsymbol{z}_k^\top\{(n-3)\bar{\boldsymbol{z}}_{(j,k,k_1)}\boldsymbol{z}_{k_1}^\top\}\right]\right.\\
&\left.\left\{\boldsymbol{V}_p(\boldsymbol{z}_k\boldsymbol{z}_k^\top)\boldsymbol{z}_{k_1}\right\}\right) = 0,\\
B_{327} =& P_{n,2}^{-2}(n-2)^{-4}\,\mathrm{E}\sum_{j,k,k_1}^{*}\left((\boldsymbol{z}_k^\top\boldsymbol{z}_{k_1}\boldsymbol{z}_{k_1}^\top)\right.\\
&\left.\left[\boldsymbol{V}_p\{(n-3)\bar{\boldsymbol{z}}_{(j,k,k_1)}\boldsymbol{z}_k^\top\}\boldsymbol{z}_{k_1}\right]\right) = 0,\\
B_{328} =& P_{n,2}^{-2}(n-2)^{-4}\,\mathrm{E}\sum_{j,k,k_1}^{*}\left((\boldsymbol{z}_k^\top\boldsymbol{z}_{k_1}\boldsymbol{z}_{k_1}^\top)\right.\\
&\left.\left[\boldsymbol{V}_p\{(n-3)\boldsymbol{z}_k\bar{\boldsymbol{z}}_{(j,k,k_1)}^\top\}\boldsymbol{z}_{k_1}\right]\right) = 0,\\
B_{329} =& P_{n,2}^{-2}(n-2)^{-4}\,\mathrm{E}\sum_{j,k,k_1}^{*}\left(\boldsymbol{z}_k^\top\boldsymbol{z}_{k_1}\boldsymbol{z}_{k_1}^\top\boldsymbol{V}_p\boldsymbol{z}_k\boldsymbol{z}_k^\top\boldsymbol{z}_{k_1}\right)\\
=& P_{n,2}^{-2}(n-2)^{-4}P_{n,3}\,\mathrm{E}\left\{(\boldsymbol{z}_1^\top\boldsymbol{z}_2)^2\left(\boldsymbol{z}_1^\top\boldsymbol{V}_p\boldsymbol{z}_2\right)\right\}\\
=& o\left\{\operatorname{tr}^2(\boldsymbol{V}_p^2)\right\}.
\end{aligned}
$$

Finally,

$$
\begin{aligned}
B_{33} =& P_{n,2}^{-2}\,\mathrm{E}\sum_{j,k}^{*}\left(\boldsymbol{z}_k^\top\bar{\boldsymbol{z}}_{(j,k)}\bar{\boldsymbol{z}}_{(j,k)}^\top\boldsymbol{z}_j\boldsymbol{z}_j^\top\bar{\boldsymbol{z}}_{(j,k)}\bar{\boldsymbol{z}}_{(j,k)}^\top\boldsymbol{z}_k\right)\\
=& P_{n,2}^{-2}\sum_{j,k}^{*}\mathrm{E}\left(\bar{\boldsymbol{z}}_{(j,k)}^\top\boldsymbol{V}_p\bar{\boldsymbol{z}}_{(j,k)}\bar{\boldsymbol{z}}_{(j,k)}^\top\boldsymbol{V}_p\bar{\boldsymbol{z}}_{(j,k)}\right)\\
=& P_{n,2}^{-2}(n-2)^{-4}\\
&\sum_{j,k}^{*}\sum_{\ell_1,\ell_2,\ell_3,\ell_4\neq j,k}\mathrm{E}\left(\boldsymbol{z}_{\ell_1}^\top\boldsymbol{V}_p\boldsymbol{z}_{\ell_2}\boldsymbol{z}_{\ell_3}^\top\boldsymbol{V}_p\boldsymbol{z}_{\ell_4}\right),
\end{aligned}
$$

where

$$
\begin{aligned}
&\sum_{\ell_1,\ell_2,\ell_3,\ell_4\neq j,k}\mathrm{E}\left(\boldsymbol{z}_{\ell_1}^\top\boldsymbol{V}_p\boldsymbol{z}_{\ell_2}\boldsymbol{z}_{\ell_3}^\top\boldsymbol{V}_p\boldsymbol{z}_{\ell_4}\right)\\
=& \sum_{\ell\neq j,k}\mathrm{E}\left(\boldsymbol{z}_\ell^\top\boldsymbol{V}_p\boldsymbol{z}_\ell\boldsymbol{z}_\ell^\top\boldsymbol{V}_p\boldsymbol{z}_\ell\right)\\
&+\sum_{\ell_1,\ell_2\neq j,k}^{*}\mathrm{E}\left(\boldsymbol{z}_{\ell_1}^\top\boldsymbol{V}_p\boldsymbol{z}_{\ell_1}\boldsymbol{z}_{\ell_2}^\top\boldsymbol{V}_p\boldsymbol{z}_{\ell_2}\right)\\
&+2\sum_{\ell_1,\ell_2\neq j,k}^{*}\mathrm{E}\left(\boldsymbol{z}_{\ell_1}^\top\boldsymbol{V}_p\boldsymbol{z}_{\ell_2}\boldsymbol{z}_{\ell_1}^\top\boldsymbol{V}_p\boldsymbol{z}_{\ell_2}\right)\\
=& (n-2)\,\mathrm{E}(\boldsymbol{z}_1^\top\boldsymbol{V}_p\boldsymbol{z}_1)^2 + (n-2)(n-3)\operatorname{tr}^2(\boldsymbol{V}_p^2)\\
&+2(n-2)(n-3)\operatorname{tr}(\boldsymbol{V}_p^4).
\end{aligned}
$$

Thus, $B_{33} = o\left\{\operatorname{tr}^2(\boldsymbol{V}_p^2)\right\}$. The proof is then completed. □

ACKNOWLEDGMENT

The work of Zhou was financially supported by First Class Discipline of Zhejiang – A (Zhejiang Gongshang University – Statistics). The work of Guo was financially supported by the Australian Research Council (ARC) Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS). The work of Zhang was financially supported by the National University of Singapore Academic Research Grant R-155-000-187-114. The authors thank the Editor, an AE and a reviewer for their insightful comments and suggestions which help improve the presentation of the paper substantially.

Received 5 March 2018

REFERENCES

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis*. Wiley New York, 3rd edition. MR1990662

Badea, L., Herlea, V., Dima, S., Dumitrascu, T., and Popescu, I. (2008). Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. *Hepatogastroenterology*, 55(88):2016–2027.

Bai, Z. D. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*, 6(2):311–329. MR1399305

Barrett, T. and Edgar, R. (2006). Gene expression omnibus (geo): Microarray data storage, submission, retrieval, and analysis. *Methods in enzymology*, 411(352–369).

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(2):808–835. MR2604697

Chowdary, D., Lathrop, J., Skelton, J., Curtin, K., Briggs, T., Zhang, Y., Yu, J., Wang, Y., and Mazumder, A. (2006). Prognostic gene expression signatures can be measured in tissues collected in rnalater preservative. *The Journal of Molecular Diagnostics*, 8(1):31–39.

Fang, K.-T., Kotz, S., and Ng, K. W. (1990). *Symmetric multivariate and related distributions*. Chapman and Hall. MR1071174

Oja, H. (2010). *Multivariate Nonparametric Methods with R*. Springer-Verlag New York. MR2598854

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2(6):110–114.

Wang, L., Peng, B., and Li, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *Journal of the American Statistical Association*, 110(512):1658–1669. MR3449062

Welch, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, 34(1/2):28–35. MR0019277

Zhang, J.-T. (2005). Approximate and asymptotic distributions of chi-squared-type mixtures with applications. *Journal of the American Statistical Association*, 100(469):273–285. MR2156837

Zhang, J.-T., Guo, J., Zhou, B., and Cheng, M.-Y. (2015). *A simple and adaptive two-sample test in high dimensions based on L^2 norm.* Manuscript.

Bu Zhou
School of Statistics and Mathematics
Zhejiang Gongshang University
China
E-mail address: bu.zhou@u.nus.edu

Jia Guo
College of Economics and Management
Zhejiang University of Technology
China
E-mail address: jia.guo@u.nus.edu

Jianwei Chen
Department of Mathematics and Statistics
San Diego State University
USA
and Center of Modern Applied Statistics and Big Data
School of Statistics
Huaqiao University
China
E-mail address: jianweichen@hqu.edu.cn

Jin-Ting Zhang
Department of Statistics and Applied Probability
National University of Singapore
Singapore
E-mail address: stazjt@nus.edu.sg