Testing high-dimensional mean vector with applications

A normal reference approach

Jin-Ting Zhang[1] · Bu Zhou[2,3] (ID) **· Jia Guo[4]**

Abstract

A centered L^2-norm based test statistic is used for testing if a high-dimensional mean vector equals zero where the data dimension may be much larger than the sample size. Inspired by the fact that under some regularity conditions the asymptotic null distributions of the proposed test are the same as the limiting distributions of a chi-square-mixture, a three-cumulant matched chi-square-approximation is suggested to approximate this null distribution. The asymptotic power of the proposed test under a local alternative is established and the effect of data non-normality is discussed. A simulation study under various settings demonstrates that in terms of size control, the proposed test performs significantly better than some existing competitors. Several real data examples are presented to illustrate the wide applicability of the proposed test to a variety of high-dimensional data analysis problems, including the one-sample

✉ Bu Zhou
  bu.zhou@u.nus.edu

  Jin-Ting Zhang
  stazjt@nus.edu.sg

  Jia Guo
  jia.guo@u.nus.edu

[1] Department of Statistics and Data Science, National University of Singapore, Singapore, Singapore

[2] School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou, China

[3] Collaborative Innovation Center of Statistical Data Engineering, Technology & Application, Zhejiang Gongshang University, Hangzhou, China

[4] School of Management, Zhejiang University of Technology, Hangzhou, China

problem, paired two-sample problem, and MANOVA for correlated samples or independent samples.

Keywords High-dimensional data · Matrix variate data · One-sample problem · Two-sample problem · MANOVA · Linear hypothesis · Chi-square-type mixtures · Three-cumulant matched chi-square-approximation

1 Introduction

In this paper, we are interested in the one-sample problem for high-dimensional data. Here "high dimension" means "the data dimension is close to or even much larger than the sample size". High-dimensional data are encountered when many measurements are taken on only a few subjects. For example, in DNA microarray data, thousands of gene expression levels are often measured on a relatively few subjects. With rapid development of data collecting technologies, high-dimensional data become rather common and attract many research efforts nowadays. Many new methods are proposed for high-dimensional hypothesis testing problems about mean vectors or covariance matrices in recent years, see, for example, Li et al. (2020), Bai et al. (2021), Zhang et al. (2021), and Silva et al. (2021) among others. The canonical one-sample problem aims to test if the population mean vector of a sample is a zero vector, and many interesting and more complicated hypotheses can be converted to it by some simple transformations, such as in the one group repeat measurement designs (Ahmad et al. 2008), in the mean matrix structure of transposable data (Touloumis et al. 2015), the two-sample problem (Chen and Qin 2010), and the multi-sample problem (Schott 2007).

The classical solution to the multivariate one-sample problem is Hotelling's T^2 test. However, Hotelling's T^2 test does not apply to high-dimensional data when the data dimension is larger than the sample size because in this case the sample covariance matrix is not invertible. To overcome this problem, many alternative tests are then proposed to test the one-sample hypothesis in high-dimensional settings. Srivastava and Du (2008) proposed a scale-invariant test. Park and Ayyala (2013) proposed a leaving-one-out scale-invariant test. Wang et al. (2015) proposed a nonparametric one-sample test based on the multivariate spatial sign transformation for elliptically distributed data. Feng and Sun (2016) proposed a scale-invariant nonparametric test based on spatial ranks and inner standardization which can also take the scale difference of variables into account. Some other tests include the random permutation based test proposed by Shen and Lin (2015), the randomization test proposed by Wang and Xu (2019), block diagonal test by Zhao (2017), the diagonal likelihood ratio test by Hu et al. (2019), the sign test by Paindaveine and Verdebout (2016), the composite T^2 test by Feng et al. (2017), shrinkage-based regularization tests by Chen et al. (2011), Shen et al. (2011) and Dong et al. (2016), and the empirical likelihood test by Peng et al. (2014) among others.

Many existing tests, such as the tests by Srivastava and Du (2008) and Wang et al. (2015), use normal approximation to approximate their null distributions. However, for most tests, normal approximation is only valid under very strong conditions on

the underlying covariance matrix as noted by Katayama et al. (2013). One of the key conditions requires that the high-dimensional data are less or nearly not correlated. To relax the assumptions on the underlying covariance matrix, Zhang and Xu (2009) proposed an L^2-norm one-sample test for normal data based on the two-cumulant (2-c) matched Welch–Satterthwaite χ^2-approximation. For one-group normally distributed repeated measures designs, Ahmad et al. (2008) proposed a test with the 2-c matched χ^2-approximation, and Pauly et al. (2015) proposed a test with the three-cumulant (3-c) matched χ^2-approximation of Zhang (2005).

In this paper, we propose and study a normal reference test with the 3-c matched χ^2-approximation for a general one-sample problem with non-normal high-dimensional data. We show that under some regularity conditions, when the null hypothesis is true, the proposed test statistic and a χ^2-type mixture have the same normal or non-normal limit distributions. It is then justifiable to approximate the null distribution of the test statistic using that of the χ^2-type mixture. The distribution of the χ^2-type mixture which has both positive and negative unknown coefficients can be well approximated by a 3-c matched χ^2-approximation with the approximation parameters consistently estimated from the data. Since the χ^2-type mixture is obtained from the test statistic when the null hypothesis holds and when the data are normally distributed, the resulting test is termed as a normal reference test with 3-c matched χ^2-approximation.

The proposed test has a close relationship with the test proposed by Pauly et al. (2015) but the two tests have several different aspects as listed below. First of all, our test is investigated for general non-normal data and one-sample tests with other types of data can be reduced to our one-sample test via some simple transformations while their test is studied only for normally distributed repeated measure designs. Second, the test proposed by Pauly et al. (2015) is based on a nonnegative squared L^2-norm statistic and their approximation essentially follows Hall (1983) by matching the third moment of normalized variables. On the other hand, our statistic is a centered squared L^2-norm statistic and our approximation is formulated as in Zhang (2005) for a χ^2-mixture with both positive and negative coefficients. Third, our approximation parameter estimators are constructed directly without using U-statistics and are ratio-consistent under the null or any alternative hypotheses while their approximation parameter estimators are constructed using U-statistics which are often time and space consuming and are ratio-consistent only under the null hypothesis. In practice, one does not know if the null hypothesis holds. Fourth, the asymptotic power of our test is established, the effect of the data non-normality on our test is discussed, and a sufficient and necessary condition is found for the asymptotic normality of our test. These are not discussed in Pauly et al. (2015).

The rest of the paper is organized as follows. Our main results are presented in Sect. 2. A simulation study is presented in Sect. 3. Applications of our test to one-sample problems with other types of data are presented in Sect. 4. Some concluding remarks are given in Sect. 5. The technical proofs of the main results are outlined in the Appendix.

2 Main results

Our study is motivated by a multivariate analysis of variance (MANOVA) problem for dependent samples. Suppose we have n independent, identically distributed (i.i.d.) $q \times k$ matrix variate observations $X_i = (x_{i1}, \ldots, x_{ik})$, $i = 1, \ldots, n$. The k columns of the observation matrix X_i correspond to matched multivariate observations from k different samples. Unlike the usual MANOVA problem for independent samples, we assume the observations of the k samples are matched, and allow possible dependence between matched observations from different samples. Besides, as frequently encountered in many practical problems, such as in the time profiles analysis (Ahmad et al. 2008; Pauly et al. 2015), we allow k (or q) to be large, even be proportional to the sample size n. The interested problem is whether the mean vectors of the k samples are the same, i.e., to test

$$H_0 : \mathrm{E}(x_{11}) = \cdots = \mathrm{E}(x_{1k}) \text{ versus } H_1 : H_0 \text{ is not true.} \tag{1}$$

In this paper, instead of trying to solve above specific problem directly, we treat it as a special case of the following one-sample problem. Suppose we have one high-dimensional sample:

$$y_1, \ldots, y_n \text{ are i.i.d. } p \text{-dimensional random vectors,} \tag{2}$$

with $\mathrm{E}(y_1) = \mu$ and $\mathrm{Cov}(y_1) = \Sigma$ where the dimension p is big, and may be much larger than the sample size n. Consider the following hypotheses:

$$H_0 : \ \mu = 0 \ \text{ versus } \ H_1 : \ \mu \neq 0. \tag{3}$$

In many situations, one may be interested in testing the hypotheses: $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ for some known constant vector μ_0. This general one-sample problem can be reduced to the one-sample problem (3) based on the induced sample $y_i - \mu_0$, $i = 1, \ldots, n$ and with μ replaced by $\mu - \mu_0$. To see the connection between hypotheses (3) and (1), let $P = I_k - k^{-1}J_k$, where I_k is a $k \times k$ identity matrix and J_k is a $k \times k$ matrix of ones. The hypothesis H_0 in (1) is equivalent to $\mathrm{vec}[\mathrm{E}(X_1)P] = \mathrm{E}[\mathrm{vec}(X_1 P)] = 0$, where vec denotes the matrix vectorization by column operator, so to test the hypothesis H_0 in (1) for the original sample X_i, $i = 1, \ldots, n$, we can just test the hypothesis H_0 in (3) for the induced sample $y_i = \mathrm{vec}(X_i P), i = 1, \ldots, n$.

2.1 Asymptotic null distribution

Let

$$\bar{y} = n^{-1} \sum_{i=1}^{n} y_i, \ \text{ and } \ \hat{\Sigma} = (n-1)^{-1} \sum_{i=1}^{n} (y_i - \bar{y})(y_i - \bar{y})^\top \tag{4}$$

denote the sample mean vector and covariance matrix, respectively. Inspired by the two-sample test of Bai and Saranadasa (1996), the test statistic for testing the one-

sample problem (3) can be constructed as

$$T_{n,p} = n\|\bar{\mathbf{y}}\|^2 - \operatorname{tr}(\hat{\mathbf{\Sigma}}), \tag{5}$$

where $\|\cdot\|$ denotes the usual L^2-norm of a vector. We can write

$$T_{n,p} = T_{n,p,0} + 2S_{n,p} + n\|\boldsymbol{\mu}\|^2, \tag{6}$$

where

$$T_{n,p,0} = n\|\bar{\mathbf{y}} - \boldsymbol{\mu}\|^2 - \operatorname{tr}(\hat{\mathbf{\Sigma}}), \quad S_{n,p} = n\boldsymbol{\mu}^\top(\bar{\mathbf{y}} - \boldsymbol{\mu}). \tag{7}$$

Note that $T_{n,p,0}$ has the same distribution as $T_{n,p}$ under the null hypothesis.

When the sample (2) is normally distributed, it is easy to see that for any given n and p, the distribution of $T_{n,p,0}$ has the same distribution as that of the following χ^2-type mixture

$$T_{n,p,0}^* = \sum_{r=1}^p \lambda_{p,r}[A_r - B_r/(n-1)], \quad A_r \stackrel{\text{i.i.d.}}{\sim} \chi_1^2, \quad B_r \stackrel{\text{i.i.d.}}{\sim} \chi_{n-1}^2, \tag{8}$$

where χ_v^2 denotes a central chi-square distribution with v degrees of freedom, $\lambda_{p,r}$, $r = 1, \ldots, p$ are the eigenvalues of the covariance matrix $\mathbf{\Sigma}$. The first three cumulants of $T_{n,p,0}^*$ are given as $\mathrm{E}(T_{n,p,0}^*) = 0$,

$$\operatorname{Var}(T_{n,p,0}^*) = \frac{2n}{n-1}\operatorname{tr}(\mathbf{\Sigma}^2), \quad \text{and} \quad \mathrm{E}(T_{n,p,0}^{*3}) = \frac{8n(n-2)}{(n-1)^2}\operatorname{tr}(\mathbf{\Sigma}^3). \tag{9}$$

Now we study the asymptotic property of T_n when both n and p tend to infinity. Although the situation described by this kind of high-dimensional asymptotics never happens in reality, the high-dimensional property of T_n gives a hint how it behaves in the practical scenario that when both sample size and data dimension are large, or when the data dimension is comparable to the sample size. More importantly, the limiting behavior of T_n provides a guidance for properly approximating its null distribution and the p value of the corresponding test when both n and p are large.

Set $\rho_{p,r} = \lambda_{p,r}/\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}$, $r = 1, \ldots, p$. The following conditions are convenient for the theoretical study:

C1 We have $\mathbf{y}_i = \boldsymbol{\mu} + \mathbf{\Gamma}\mathbf{z}_i$, $i = 1, \ldots, n$, where $\mathbf{\Gamma}$ is a $p \times p$ matrix such that $\mathbf{\Gamma}\mathbf{\Gamma}^\top = \mathbf{\Sigma}$ and $\mathbf{z}_i$'s are i.i.d. p-vectors with $\mathrm{E}(\mathbf{z}_i) = \mathbf{0}$ and $\operatorname{Cov}(\mathbf{z}_i) = \mathbf{I}_p$, the $p \times p$ identity matrix.

C2 We have $\mathrm{E}(z_{ir}^4) = 3 + \Delta < \infty$ where z_{ir} is the r-th component of $\mathbf{z}_i$, Δ is some constant, and $\mathrm{E}(z_{ir_1}^{\alpha_1} \cdots z_{ir_q}^{\alpha_q}) = \mathrm{E}(z_{ir_1}^{\alpha_1}) \cdots \mathrm{E}(z_{ir_q}^{\alpha_q})$ for a positive integer q such that $\sum_{r=1}^q \alpha_r \leq 8$ and $r_1 \neq \cdots \neq r_q$.

C3 We have $\lim_{p \to \infty} \rho_{p,r} = \rho_r$, $r = 1, 2, \ldots$, uniformly and $\lim_{p \to \infty} \sum_{r=1}^p \rho_{p,r} = \sum_{r=1}^\infty \rho_r < \infty$.

C4 As $n, p \to \infty$, we have $p/n^2 \longrightarrow 0$.

C5 As $p \to \infty$, we have $\rho_{p,\max} \to 0$ where $\rho_{p,\max} = \max_{r=1}^{p} \rho_{p,r}$.

Conditions C1 and C2 are also imposed by Bai and Saranadasa (1996) and Chen and Qin (2010), respectively. They specify a factor model for high-dimensional data analysis. Condition C3 is also imposed by Zhang et al. (2020), it ensures the existence of the limits of $\lambda_{p,r}$ as $p \to \infty$ and the exchangeability of the limit and summation operations in the expression $\lim_{p \to \infty} \sum_{r=1}^{p} \rho_{p,r}$. Condition C3 implies that $\sum_{r=q+1}^{p} \rho_{p,r} \longrightarrow \sum_{r=q+1}^{\infty} \rho_r$ as $p \to \infty$ for any fixed $q < p$, and $\sum_{r=q+1}^{\infty} \rho_r \longrightarrow 0$ as $q \to \infty$. It is used to ensure that the limiting distributions of the normalized versions of $T_{n,p,0}$ and $T_{n,p,0}^*$, namely,

$$\tilde{T}_{n,p,0} = \frac{T_{n,p,0}}{\sqrt{\frac{2n}{(n-1)} \operatorname{tr}(\boldsymbol{\Sigma}^2)}}, \quad \text{and} \quad \tilde{T}_{n,p,0}^* = \frac{T_{n,p,0}^*}{\sqrt{\frac{2n}{(n-1)} \operatorname{tr}(\boldsymbol{\Sigma}^2)}}, \tag{10}$$

are non-normal. Condition C4 is needed by Lemma 1 presented in the Appendix which proves the ratio-consistency of the estimator (20) of $\operatorname{tr}(\boldsymbol{\Sigma}^3)$. It is also needed by Theorems 4 and 5 . This condition is weaker than the condition "$p/n \longrightarrow c \in (0, \infty)$ as $n, p \to \infty$" imposed by Bai and Saranadasa (1996). It allows $p/n \longrightarrow \infty$ as $n, p \to \infty$ but only allows p to diverge in a slower rate than n^2. Condition C5 is also imposed by Bai and Saranadasa (1996) and it is used to ensure that the limiting distributions of $\tilde{T}_{n,p,0}$ and $\tilde{T}_{n,p,0}^*$ are normal. Conditions C3 and C5 impose two exclusive constraints on the eigenvalues of the covariance matrix $\boldsymbol{\Sigma}$ so that the limiting distributions of $\tilde{T}_{n,p,0}$ and $\tilde{T}_{n,p,0}^*$ are non-normal and normal, respectively. Theoretically speaking, when the eigenvalues of $\boldsymbol{\Sigma}$ are in the same order (e.g., under a non-spiked covariance model where no eigenvalues of $\boldsymbol{\Sigma}$ can dominate the other eigenvalues), Condition C5 is satisfied so that $\tilde{T}_{n,p,0}$ and $\tilde{T}_{n,p,0}^*$ will be asymptotically normally distributed and when the sequence of decreasingly ordered eigenvalues of $\boldsymbol{\Sigma}$ tends to 0 quickly (e.g., under a spiked covariance model where a finite number of eigenvalues dominate the remaining eigenvalues asymptotically) such that $\operatorname{tr}^2(\boldsymbol{\Sigma})/\operatorname{tr}(\boldsymbol{\Sigma}^2)$ tends to a finite limit, Condition C3 is satisfied. In real data analysis, largely speaking, when the p-components of an observation are nearly uncorrelated, Condition C5 is approximately satisfied and when they are moderately or highly correlated, Condition C3 is approximately satisfied. Let $\stackrel{d}{=}$ denote equality in distribution and $\stackrel{\mathcal{L}}{\longrightarrow}$ denote convergence in distribution. We have the following useful theorem.

Theorem 1 (a) *Under Conditions C1, C2 and C3, as $n, p \to \infty$, we have*

$$\tilde{T}_{n,p,0} \stackrel{\mathcal{L}}{\longrightarrow} \zeta, \quad \text{and} \quad \tilde{T}_{n,p,0}^* \stackrel{\mathcal{L}}{\longrightarrow} \zeta, \tag{11}$$

where $\zeta \stackrel{d}{=} \sum_{r=1}^{\infty} \rho_r (A_r - 1)/\sqrt{2}$, $A_r \stackrel{i.i.d.}{\sim} \chi_1^2$.
(b) *Under Conditions C1, C2 and C5, as $n, p \to \infty$, we have*

$$\tilde{T}_{n,p,0} \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}(0, 1), \quad \text{and} \quad \tilde{T}_{n,p,0}^* \stackrel{\mathcal{L}}{\longrightarrow} \mathcal{N}(0, 1). \tag{12}$$

Then under the conditions of (a) or (b), we always have

$$\sup_x |\Pr(T_{n,p,0} \le x) - \Pr(T^*_{n,p,0} \le x)| \longrightarrow 0. \tag{13}$$

In one-sample test for normally distributed repeated measures designs, a theorem comparable with Theorem 1 has been proved by Pauly et al. (2015). However, the authors failed to extend it to non-normal repeated measures designs in their paper. Theorem 1 provides a theoretical justification for us to use the distribution of $T^*_{n,p,0}$ to approximate the distribution of $T_{n,p,0}$. Notice that $T^*_{n,p,0}$ is obtained when the data (2) are normally distributed. Thus, we term the distribution of $T^*_{n,p,0}$ as the normal-reference distribution of $T_{n,p,0}$.

2.2 Implementation

To implement the proposed test, we approximate the null distribution of $T_{n,p}$ using that of $T^*_{n,p,0}$. Different from the L^2-norm test studied in Zhang and Xu (2009), whose null distribution is the same as a χ^2-type mixture with only positive coefficients, the distribution of $T^*_{n,p,0}$ is the same as a χ^2-type mixture with both positive and negative coefficients. For such a χ^2-type mixture, Zhang (2013) showed, with some simulation studies, that the 2-c matched χ^2-approximation method (Welch 1947; Satterthwaite 1946; Box 1954) adopted by Zhang and Xu (2009) should not be used to approximate the distribution of $T^*_{n,p,0}$. Rather, the 3-c matched χ^2-approximation method of Zhang (2005) should be used.

One obvious advantage of the 3-c χ^2-approximation method for approximating the distribution of $T^*_{n,p,0}$ over the normal approximation suggested by Bai and Saranadasa (1996), and the 2-c matched χ^2-approximation method used by Zhang and Xu (2009), is that the former matches the first three cumulants while the latter two only matches the first two cumulants. So it is expected that in terms of size control the 3-c χ^2-approximation should be more accurate than the normal approximation and the 2-c matched χ^2-approximation. In fact, Zhang (2005) showed, theoretically in terms of upper density approximation error bound and via simulation studies, that the 3-c matched χ^2-approximation has a much better accuracy than the normal approximation even when the normal approximation is adequate.

By the 3-c matched χ^2-approximation method of Zhang (2005), we approximate the distribution of $T^*_{n,p,0}$ using the distribution of the random variable

$$R = \beta_0 + \beta_1 \chi^2_d, \tag{14}$$

where the parameters β_0, β_1 and d are determined via matching the first three cumulants of $T^*_{n,p,0}$ and R. The first three cumulants of $T^*_{n,p,0}$ are given in (9) while by (14), the first three cumulants of R are given by $\beta_0 + \beta_1 d$, $2\beta_1^2 d$, and $8\beta_1^3 d$, respectively.

Matching the first three cumulants of $T^*_{n,p,0}$ and R then leads to

$$\beta_0 = -\frac{n\,\mathrm{tr}^2(\boldsymbol{\Sigma}^2)}{(n-2)\,\mathrm{tr}(\boldsymbol{\Sigma}^3)}, \quad \beta_1 = \frac{(n-2)\,\mathrm{tr}(\boldsymbol{\Sigma}^3)}{(n-1)\,\mathrm{tr}(\boldsymbol{\Sigma}^2)}, \quad d = \frac{n(n-1)}{(n-2)^2}\frac{\mathrm{tr}^3(\boldsymbol{\Sigma}^2)}{\mathrm{tr}^2(\boldsymbol{\Sigma}^3)}. \quad (15)$$

The parameter d is usually called as the approximate degrees of freedom of the 3-c matched χ^2-approximation to $T^*_{n,p,0}$. Note that since $\boldsymbol{\Sigma}$ is always nonnegative, we always have $\beta_0 < 0$, $\beta_1 > 0$, and $d > 0$. This is reasonable since $T^*_{n,p,0}$ is a χ^2-type mixture with both positive and negative coefficients. Using d defined above, the skewness of $T^*_{n,p,0}$ is given by

$$\mathrm{E}(T^{*3}_{n,p,0})/\,\mathrm{Var}^{3/2}(T^*_{n,p,0}) = (8/d)^{1/2}. \quad (16)$$

To implement the proposed test in real data analysis, we need estimate $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ and $\mathrm{tr}(\boldsymbol{\Sigma}^3)$ consistently. Let their ratio-consistent estimators be denoted respectively as $\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}$ and $\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^3)}$. Then the ratio-consistent estimators of β_0, β_1 and d are respectively given by

$$\hat{\beta}_0 = -\frac{n[\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}]^2}{(n-2)\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^3)}}, \quad \hat{\beta}_1 = \frac{(n-2)\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^3)}}{(n-1)\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}}, \quad \hat{d} = \frac{n(n-1)}{(n-2)^2}\frac{[\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}]^3}{[\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^3)}]^2}. \quad (17)$$

For any nominal significance level $\alpha > 0$, let $\chi^2_v(\alpha)$ denote the upper 100α percentile of χ^2_v. Then by (17), the proposed test for the one-sample problem (3) using $T_{n,p}$ with the 3-c matched χ^2-approximation is then conducted via using the approximate critical value $\hat{\beta}_0 + \hat{\beta}_1\chi^2_{\hat{d}}(\alpha)$ or the approximate p value $\Pr\left[\chi^2_{\hat{d}} \geq (T_{n,p} - \hat{\beta}_0)/\hat{\beta}_1\right]$.

In practice, one often uses the following normalized version of $T_{n,p}$:

$$\tilde{T}_{n,p} = \frac{T_{n,p}}{\sqrt{\frac{2n}{n-1}\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)}}}. \quad (18)$$

Then to approximate the distribution of $T_{n,p}$ using that of $\hat{\beta}_0 + \hat{\beta}_1\chi^2_{\hat{d}}$ is equivalent to approximate the distribution of $\tilde{T}_{n,p}$ using that of $(\chi^2_{\hat{d}} - \hat{d})/\sqrt{2\hat{d}}$. In this case, the proposed test for the one-sample problem (3) using $\tilde{T}_{n,p}$ with the 3-c matched χ^2-approximation can also be conducted via using the approximate critical value $[\chi^2_{\hat{d}}(\alpha) - \hat{d}]/\sqrt{2\hat{d}}$ or the approximate p value $\Pr\left(\chi^2_{\hat{d}} \geq \hat{d} + \sqrt{2\hat{d}}\tilde{T}_{n,p}\right)$.

We now consider the ratio-consistent estimators of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ and $\mathrm{tr}(\boldsymbol{\Sigma}^3)$. By Lemma S.3 of Zhang et al. (2020), a ratio-consistent estimator of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ is given by

$$\widehat{\mathrm{tr}(\boldsymbol{\Sigma}^2)} = \frac{(n-1)^2}{(n-2)(n+1)}\left[\mathrm{tr}(\hat{\boldsymbol{\Sigma}}^2) - \frac{\mathrm{tr}^2(\hat{\boldsymbol{\Sigma}})}{n-1}\right], \quad (19)$$

where $\hat{\boldsymbol{\Sigma}}$ is the sample covariance estimator of $\boldsymbol{\Sigma}$ as given in (4). When the data (2) are normally distributed, we have $\hat{\boldsymbol{\Sigma}} \sim W_p(n-1, \boldsymbol{\Sigma}/(n-1))$, a Wishart distribution with $n-1$ degrees of freedom and covariance matrix $\boldsymbol{\Sigma}/(n-1)$. Then under Condition C4, by Lemma 1 given in the Appendix, an unbiased and ratio-consistent estimator of $\text{tr}(\boldsymbol{\Sigma}^3)$ is given by

$$\widehat{\text{tr}(\boldsymbol{\Sigma}^3)} = \frac{(n-1)^4}{(n^2+n-6)(n^2-2n-3)}\left[\text{tr}(\hat{\boldsymbol{\Sigma}}^3) - \frac{3\,\text{tr}(\hat{\boldsymbol{\Sigma}})\,\text{tr}(\hat{\boldsymbol{\Sigma}}^2)}{(n-1)} + \frac{2\,\text{tr}^3(\hat{\boldsymbol{\Sigma}})}{(n-1)^2}\right]. \quad (20)$$

We conjecture that when Conditions C1, C2 and C4 are satisfied, $\widehat{\text{tr}(\boldsymbol{\Sigma}^3)}$ is also ratio-consistent for $\text{tr}(\boldsymbol{\Sigma}^3)$ for non-normal data. This is partially confirmed by the simulation results presented in Sect. 3 and in the Supplementary Material where the proposed test works well in terms of size control regardless of whether the data are nearly uncorrelated, moderately correlated or highly correlated and whether the data are normally or non-normally distributed. A theoretical justification of the ratio-consistency of $\widehat{\text{tr}(\boldsymbol{\Sigma}^3)}$ without the normality assumption, like the one given in Lemma 1 for normal data, is theoretically interesting and mathematically possible but expectedly rather laborious because the evaluation of the mean and variance of $\widehat{\text{tr}(\boldsymbol{\Sigma}^3)}$ for non-normal data will be much more complicated than those in the proof of Lemma 1 for normal data. Further research in this direction is interesting and warranted. It is worthwhile to mention that a U-statistic based estimator of $\text{tr}(\boldsymbol{\Sigma}^3)$ is given by Pauly et al. (2015) (Theorem 8.2). However, this estimator is often time-consuming, especially when both n and p are large. Further, its ratio-consistency is proved under the null hypothesis and the normality assumption as well.

2.3 Asymptotic power

In this subsection, we investigate the asymptotic power of $T_{n,p}$. By (6), we have the expansion $T_{n,p} \stackrel{d}{=} T_{n,p,0} + 2S_{n,p} + n\|\boldsymbol{\mu}\|^2$ where $T_{n,p,0}$ has the same distribution as $T_{n,p}$ under the null hypothesis and $\text{Var}(S_{n,p}) = n\boldsymbol{\mu}^\top \boldsymbol{\Sigma}\boldsymbol{\mu}$. Following Bai and Saranadasa (1996), let's consider the power of $T_{n,p}$ under the following local alternative:

$$\text{as } n, p \to \infty, \quad n\boldsymbol{\mu}^\top \boldsymbol{\Sigma}\boldsymbol{\mu} = o[\text{tr}(\boldsymbol{\Sigma}^2)]. \quad (21)$$

This is the case when $\text{Var}(S_{n,p}) = o[\text{Var}(T_{n,p,0})]$ so that $T_{n,p} = T_{n,p,0} + n\|\boldsymbol{\mu}\|^2 + o_p\left[\sqrt{\text{Var}(T_{n,p,0})}\right]$ since $\text{E}(S_{n,p}) = 0$.

Theorem 2 *Assume that $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{d}$ are the ratio-consistent estimators of β_0, β_1 and d as $n, p \to \infty$, respectively. Then, (a) Under Conditions C1, C2, C3, and the local alternative (21), as $n, p \to \infty$, we have*

$$\Pr\left[T_{n,p} > \hat{\beta}_0 + \hat{\beta}_1 \chi_{\hat{d}}^2(\alpha)\right] = \Pr\left[\zeta \geq \frac{\chi_d^2(\alpha) - d}{\sqrt{2d}} - \frac{n\|\boldsymbol{\mu}\|^2}{\sqrt{2\,\text{tr}(\boldsymbol{\Sigma}^2)}}\right][1 + o(1)],$$

where ζ is defined in Theorem 1(a).

(b) Under Conditions C1, C2, C4, C5 and the local alternative (21), as $n, p \to \infty$, we have

$$\Pr\left[T_{n,p} > \hat{\beta}_0 + \hat{\beta}_1 \chi_{\hat{d}}^2(\alpha)\right] = \Phi\left[-z_\alpha + \frac{n\|\boldsymbol{\mu}\|^2}{\sqrt{2\operatorname{tr}(\boldsymbol{\Sigma}^2)}}\right][1 + o(1)],$$

where z_α denotes the upper 100α-percentile of $\mathcal{N}(0, 1)$ and $\Phi(\cdot)$ denotes the cumulative distribution function of $\mathcal{N}(0, 1)$.

For any $d \geq 1$ and a small α, it is easy to check that we always have $z_\alpha < [\chi_d^2(\alpha) - d]/\sqrt{2d}$. This shows that under Conditions C1–C3 and the local alternative (21), the asymptotic size and power of the proposed test with the normal approximation are expected to be "artificially" larger than those of the proposed test with the 3-c χ^2-approximation. This is consistent with what we observe from the simulation results presented in Sect. 3.

2.4 Effect of data non-normality

The validness of the proposed normal reference test is guaranteed by Theorem 1. In this subsection, we aim to further investigate the effect of the data non-normality onto the proposed test. That is, how does the data non-normality affect the performance of the proposed test? To answer this question, we study how to approximate the distribution of $T_{n,p,0}$ directly using the 3-c matched χ^2-approximation. To this end, we compute the first three cumulants of $T_{n,p,0}$ as in the following theorem.

Theorem 3 *The first three cumulants of $T_{n,p,0}$ are given by $\mathrm{E}(T_{n,p,0}) = 0$,*

$$\operatorname{Var}(T_{n,p,0}) = \frac{2n}{n-1}\operatorname{tr}(\boldsymbol{\Sigma}^2), \ \text{and} \ \mathrm{E}(T_{n,p,0}^3) = \frac{8n(n-2)}{(n-1)^2}\operatorname{tr}(\boldsymbol{\Sigma}^3) + \frac{4n\Upsilon}{(n-1)^2},$$

where $\Upsilon = \mathrm{E}[(\boldsymbol{y}_1 - \boldsymbol{\mu})^\top(\boldsymbol{y}_2 - \boldsymbol{\mu})]^3$.

It is seen from Theorem 3 that the data non-normality affects the third moment of $T_{n,p,0}$ only. To approximate the distribution of $T_{n,p,0}$ directly using that of $W = b_0 + b_1\chi_f^2$ via matching the first three cumulants of $T_{n,p,0}$ and W, the parameters b_0, b_1 and f are obtained as

$$b_0 = \beta_0/\delta, \ \ b_1 = \beta_1\delta, \ \ \text{and} \ f = d/\delta^2, \ \ \text{where} \ \delta = 1 + \Upsilon/[2(n-2)\operatorname{tr}(\boldsymbol{\Sigma}^3)], \tag{22}$$

and β_0, β_1 and d are given in (15). Note that the skewness of $T_{n,p,0}$ is given by

$$\mathrm{E}(T_{n,p,0}^3)/\operatorname{Var}^{3/2}(T_{n,p,0}) = (8/f)^{1/2}. \tag{23}$$

The quantity Υ can be seen as a non-invariant measure of multivariate normality based on skewness (See, e.g., Sect. 3.1 of Henze 2002). When the data (2) are normal, it

is easy to show that $\varUpsilon = 0$ so that $\delta = 1$, $b_0 = \beta_0$, $b_1 = \beta_1$, $f = d$ and the skewness (23) of $T_{n,p,0}$ reduces to the skewness (16) of $T_{n,p,0}^*$ as expected. However, when the data (2) are non-normal, we may not have $\varUpsilon = 0$ and hence the approximation parameters b_0, b_1, f, and the skewness of $T_{n,p,0}$ are all affected by the data non-normality. Fortunately, we can show the following result.

Theorem 4 *(a) Under Conditions C1 and C2, we have* $\varUpsilon \leq (\Delta^2 + 6\Delta + 9)^{3/4} \operatorname{tr}^{3/2}(\boldsymbol{\Sigma}^2)$ *where Δ is given in Condition C2; and (b) Under either Conditions C1, C2 and C3 or Conditions C1, C2 and C4, we have* $\delta = 1 + o(1)$ *as $n, p \to \infty$.*

Theorem 4 says that under Conditions C1, C2 and C3 or Conditions C1, C2 and C4, the data non-normality on the proposed normal reference test can be ignorable asymptotically so that we have $b_0 = \beta_0[1 + o(1)]$, $b_1 = \beta_1[1 + o(1)]$, $f = d[1 + o(1)]$ and the skewness of $T_{n,p,0}$ and that of $T_{n,p,0}^*$ are also asymptotically equal. The following theorem gives a sufficient and necessary condition for the asymptotic normality of $\tilde{T}_{n,p,0}$.

Theorem 5 *Under Conditions C1, C2 and C4, as $n, p \to \infty$, $\tilde{T}_{n,p,0} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$ if and only if $d \longrightarrow \infty$ where d is given in* (15).

Theorem 5 indicates that when d is small, the normal approximation to the distribution of $\tilde{T}_{n,p,0}$ is unlikely to be adequate.

3 Simulation study

In this section, we conduct a simulation study to compare the proposed normal reference test with the 3-c matched χ^2-approximation (denoted as T_{new}), against the L^2-norm based test with the 2-c matched χ^2-approximation proposed by Zhang and Xu (2009) (denoted as T_{ZX}), and the tests proposed by Bai and Saranadasa (1996), Chen and Qin (2010) and Srivastava and Du (2008) (dented as T_{BS}, T_{CQ} and T_{SD}, respectively). The original T_{BS} and T_{CQ} are two-sample tests and the corresponding one-sample tests adopted here are respectively given by (1.2) and (1.5) of Zhou et al. (2019). Note that the null distributions of T_{BS}, T_{CQ} and T_{SD} are all computed using the normal approximation.

In each run, we generate the high-dimensional data (2) using $\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{z}_i$, $i = 1, \ldots, n$ where $\boldsymbol{\mu} = \delta\boldsymbol{h}$ with the components of $\boldsymbol{z}_i$ i.i.d. generated from the following three models:

- Model 1: $z_{ir}, \ r = 1, \ldots, p \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$.
- Model 2: $z_{ir} = w_{ir}/\sqrt{2}, \ r = 1, \ldots, p$ with $w_{ir}, \ r = 1, \ldots, p \overset{\text{i.i.d.}}{\sim} t_4$.
- Model 3: $z_{ik} = (w_{ir} - 1)/\sqrt{2}, \ r = 1, \ldots, p$ with $w_{ir}, \ r = 1, \ldots, p \overset{\text{i.i.d.}}{\sim} \chi_1^2$.

Based on the above three models, the resulting data are normal, symmetric but non-normal, and skewed and non-normal, respectively. The covariance matrix is specified as $\boldsymbol{\Sigma} = \sigma^2 \left[(1 - \rho)\boldsymbol{I}_p + \rho\boldsymbol{J}_p\right]$. Some additional simulation results with different covariance structures are presented in the Supplementary Material and the conclusions are similar to those presented in this section.

Note that the tuning parameters δ, $\boldsymbol{h}$ and ρ are used to control the mean vector and the data correlation, respectively. Note also that the power of a test will increase with increasing the value of δ and 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 $\boldsymbol{u} = (1, \dots, p)^\top$ and set $\sigma^2 = 1$. To compare the performance of the tests under consideration with various settings, 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 $\rho = 0.1, 0.5$ and 0.9.

In the simulations, empirical size and power of a test are calculated as the proportions of the number of rejections (i.e., number of runs when calculated p values of the associated test is smaller than nominal level $\alpha = 5\%$) out of $10,000$ runs. The empirical sizes are calculated with $\delta = 0$ so that the null hypothesis H_0 in (3) is true, and the empirical powers are calculated with $\delta > 0$. Different values of δ (see Table 2) are carefully selected for different combinations of n and p so that all the tests largely have non-trivial powers when $\rho = 0.1, 0.5$ and 0.9, respectively. To assess the performance of a test in maintaining the type I error, we define the average relative error as ARE $= 100 M^{-1} \sum_{j=1}^{M} |\hat{\alpha}_j - \alpha|/\alpha$, where α is the nominal size (5% here) and $\hat{\alpha}_j, \ j = 1, \dots, M$ denote the empirical sizes under consideration. A smaller ARE value indicates an overall better performance of the associated test in terms of maintaining the nominal size.

Table 1 displays the empirical sizes of the tests under various settings with the last row presenting the ARE values of the tests for three values of ρ. It is seen that under each setting, the empirical size of T_{new} is generally much closer to 5% than those of other tests. This shows that in terms of size control, our new test significantly outperforms other tests. This conclusion is also seen from the ARE values of the tests. In fact, from the last row of the table, it is seen that the ARE values of T_{new} are much smaller than those of other tests for $\rho = 0.1, 0.5$ and 0.9, respectively. From Table 1, we also see that in terms of size control, (a) T_{ZX} generally outperforms T_{BS}, T_{CQ} and T_{SD}; (b) T_{BS} and T_{SD} are generally comparable and they are generally very liberal with most of their empirical sizes close to 7%; and (c) T_{SD} performs quite well for Models 1 and 2 for $\rho = 0.1$ but it is very conservative for $\rho = 0.5$ and 0.9 with most of its empirical sizes much smaller than 5%. This implies that T_{SD} cannot work well for highly skewed or correlated high-dimensional data.

Table 2 displays the empirical powers of the tests under various settings. First of all, it is seen that T_{new} and T_{ZX} have comparable empirical powers, with T_{ZX}'s empirical powers slightly bigger than those of T_{new}'s. This is possibly due to the fact that as shown in Table 1, the empirical sizes of T_{ZX} are generally bigger than those of T_{new}. This observation is consistent with the conclusion drawn from Theorem 2. Second, T_{BS} and T_{CQ} have comparable empirical powers which are slightly bigger than those of T_{new} and T_{ZX}. This is also because the former tests generally have bigger empirical sizes than the latter tests. Third, T_{SD} has comparable empirical powers with other tests when $\rho = 0.1$ and under Models 1 and 2. It has lower empirical powers than other tests when $\rho = 0.5, 0.9$ or under Model 3. This again shows that T_{SD} does not work well for highly skewed or highly correlated high-dimensional data. Finally, we can see that under various settings, the empirical powers of all the tests are getting smaller

Table 1 Empirical sizes (in %) of the tests under various settings

Model	p	n	$\rho = 0.1$					$\rho = 0.5$					$\rho = 0.9$				
			T_{new}	T_{ZX}	T_{BS}	T_{CQ}	T_{SD}	T_{new}	T_{ZX}	T_{BS}	T_{CQ}	T_{SD}	T_{new}	T_{ZX}	T_{BS}	T_{CQ}	T_{SD}
1	50	30	4.99	5.69	6.50	6.47	5.88	5.76	6.39	7.72	7.68	3.89	6.06	6.70	7.97	7.98	1.48
		60	4.90	5.47	6.13	6.14	5.43	5.15	5.66	6.86	6.88	3.08	5.58	5.85	7.38	7.38	1.03
		120	5.06	5.69	6.44	6.45	5.53	5.22	5.51	6.94	6.92	3.06	5.20	5.34	7.19	7.19	0.89
	500	30	5.56	7.29	7.55	7.54	5.32	6.88	7.57	8.74	8.80	2.16	6.29	6.79	8.14	8.04	0.36
		60	4.99	6.77	7.18	7.19	5.26	5.44	5.86	7.23	7.24	1.24	5.22	5.48	6.98	6.94	0.19
		120	4.77	6.31	6.84	6.83	5.21	5.36	5.76	7.21	7.22	1.09	5.66	5.82	7.60	7.61	0.12
	1000	30	6.14	7.72	8.02	8.04	4.94	6.50	7.15	8.31	8.25	1.61	5.82	6.43	7.64	7.59	0.10
		60	5.39	7.07	7.41	7.38	4.99	5.31	5.89	7.40	7.35	0.96	5.70	6.02	7.70	7.68	0.10
		120	4.86	6.45	6.92	6.91	4.86	5.55	5.84	7.08	7.08	0.72	5.09	5.22	6.98	6.97	0.06
2	50	30	4.06	4.65	5.31	6.32	5.19	5.62	6.33	7.49	7.60	3.66	6.24	6.83	8.01	8.03	1.37
		60	4.63	5.43	6.14	6.86	5.62	5.31	5.77	7.20	7.35	3.17	5.67	5.96	7.26	7.28	1.04
		120	4.43	5.13	5.92	6.39	5.21	5.48	5.83	7.36	7.41	2.96	5.56	5.69	7.58	7.58	1.07
	500	30	5.36	6.82	7.18	7.51	5.17	6.64	7.32	8.46	8.45	1.83	6.19	6.73	7.85	7.90	0.32
		60	4.63	6.13	6.53	6.76	4.84	5.49	6.06	7.22	7.21	1.30	5.53	5.78	7.33	7.36	0.15
		120	4.47	5.83	6.37	6.56	5.04	5.30	5.73	7.13	7.14	1.07	5.26	5.45	7.17	7.18	0.14
	1000	30	5.64	7.16	7.36	7.53	4.65	6.27	7.09	8.10	8.09	1.19	6.52	7.03	8.40	8.41	0.12
		60	5.30	6.85	7.16	7.30	4.80	5.73	6.28	7.74	7.75	0.83	5.15	5.42	6.99	7.00	0.03
		120	4.70	6.11	6.61	6.72	4.73	5.32	5.71	6.97	6.97	0.65	5.43	5.57	7.29	7.29	0.03

Table 1 continued

Model	p	n	$\rho = 0.1$					$\rho = 0.5$					$\rho = 0.9$				
			T_{new}	T_{ZX}	T_{BS}	T_{CQ}	T_{SD}	T_{new}	T_{ZX}	T_{BS}	T_{CQ}	T_{SD}	T_{new}	T_{ZX}	T_{BS}	T_{CQ}	T_{SD}
3	50	30	4.14	4.90	5.34	6.58	38.80	5.75	6.48	7.35	7.43	4.87	6.13	6.68	7.96	7.99	1.48
		60	4.58	5.35	6.17	6.84	22.60	5.50	6.01	7.49	7.52	4.16	5.22	5.46	7.15	7.16	0.92
		120	4.02	4.67	5.54	5.96	13.57	5.19	5.66	7.31	7.33	2.98	5.66	5.84	7.62	7.62	1.00
	500	30	5.50	7.04	7.33	7.63	38.04	5.86	6.73	7.75	7.72	2.03	5.98	6.48	7.55	7.51	0.35
		60	4.96	6.52	6.97	7.15	22.92	5.50	5.95	6.97	7.01	1.34	5.46	5.74	7.19	7.19	0.25
		120	5.08	6.46	6.89	6.99	14.11	5.07	5.40	6.88	6.88	1.16	5.21	5.39	7.09	7.09	0.17
	1000	30	5.53	7.26	7.58	7.78	34.66	6.54	7.09	8.37	8.28	1.60	5.93	6.55	8.08	8.12	0.17
		60	5.14	6.58	6.91	6.99	20.56	5.10	5.62	6.90	6.91	0.73	5.44	5.76	7.57	7.53	0.09
		120	5.13	6.46	6.82	6.84	13.20	5.56	5.92	7.22	7.21	0.78	5.27	5.40	7.26	7.27	0.02
	ARE		7.93	25.46	34.16	39.01	132.23	12.89	23.41	49.19	49.39	59.91	12.94	19.56	50.32	50.29	90.33

Table 2 Empirical powers (in %) of the tests under various settings

Model	p	n	δ	ρ = 0.1					ρ = 0.5					ρ = 0.9				
				T_{new}	T_{ZX}	T_{BS}	T_{CQ}	T_{SD}	T_{new}	T_{ZX}	T_{BS}	T_{CQ}	T_{SD}	T_{new}	T_{ZX}	T_{BS}	T_{CQ}	T_{SD}
1	50	30	1.5	90.55	91.65	92.33	92.29	90.82	37.27	39.06	42.21	42.18	29.50	22.95	24.45	27.38	27.34	8.27
		60	1.1	92.65	93.55	94.32	94.28	93.10	39.13	40.52	45.00	45.03	29.98	23.50	24.37	28.11	28.15	8.31
		120	0.8	94.15	94.86	95.30	95.31	94.40	40.45	42.06	46.95	46.94	30.44	24.17	24.63	29.55	29.56	8.19
	500	30	4.0	90.66	93.28	93.73	93.74	89.62	27.69	29.85	32.82	32.83	11.89	17.98	19.18	21.49	21.46	2.05
		60	2.8	89.99	92.64	93.20	93.19	90.02	26.74	28.03	31.46	31.48	11.16	16.69	17.45	20.63	20.61	1.17
		120	2.0	90.74	93.13	93.76	93.78	91.32	27.27	28.37	32.38	32.34	10.12	16.96	17.28	20.82	20.84	0.72
	1000	30	5.5	90.09	92.75	93.14	93.16	86.95	26.40	28.19	31.26	31.19	9.69	16.96	18.09	20.77	20.79	0.82
		60	4.0	92.20	94.46	94.92	94.94	90.94	27.78	29.25	33.01	33.05	8.37	17.05	17.69	21.15	21.13	0.61
		120	2.5	81.34	84.78	85.73	85.73	80.96	21.88	22.99	26.15	26.16	5.20	13.85	14.04	17.42	17.40	0.36
2	50	30	1.5	88.69	90.04	91.01	92.46	93.48	37.27	39.49	42.81	43.06	29.82	23.97	25.56	28.52	28.57	8.47
		60	1.1	91.37	92.51	93.29	94.11	94.45	39.43	41.04	45.41	45.62	29.70	24.37	25.09	28.87	28.92	8.63
		120	0.8	93.21	93.91	94.61	95.11	95.23	40.12	41.36	46.21	46.33	30.31	24.31	24.59	29.04	29.12	8.24
	500	30	4.0	89.54	92.12	92.62	93.19	90.18	28.20	30.09	33.03	33.09	12.49	17.62	18.73	21.47	21.45	1.99
		60	2.8	90.09	92.85	93.40	93.74	91.32	26.62	28.11	31.39	31.37	10.11	17.09	17.67	20.71	20.72	1.20
		120	2.0	90.25	92.75	93.47	93.63	91.58	27.20	28.37	32.20	32.18	9.94	16.55	16.74	20.25	20.23	1.03
	1000	30	5.5	89.60	92.29	92.84	93.24	87.54	27.34	29.08	32.01	31.93	8.48	16.65	17.79	20.39	20.29	0.99
		60	4.0	91.51	94.02	94.43	94.69	90.85	27.89	29.26	32.71	32.70	8.10	17.59	18.14	21.61	21.62	0.49
		120	2.5	80.47	84.54	85.37	85.55	80.55	21.22	22.27	26.18	26.19	4.82	13.81	14.10	17.29	17.29	0.34

Table 2 continued

Model	p	n	δ	ρ = 0.1					ρ = 0.5					ρ = 0.9				
				T_{new}	T_{ZX}	T_{BS}	T_{CQ}	T_{SD}	T_{new}	T_{ZX}	T_{BS}	T_{CQ}	T_{SD}	T_{new}	T_{ZX}	T_{BS}	T_{CQ}	T_{SD}
3	50	30	1.5	90.84	91.86	92.79	94.34	85.99	36.10	38.31	41.89	42.30	19.74	20.18	21.89	25.09	25.19	6.15
		60	1.1	93.80	94.43	95.04	95.61	89.44	37.28	38.93	43.18	43.33	21.83	22.99	23.71	27.70	27.76	6.64
		120	0.8	94.96	95.61	96.10	96.43	91.70	38.69	40.06	44.90	44.98	23.98	22.87	23.31	27.72	27.79	6.97
	500	30	4.0	91.74	94.07	94.42	94.77	77.26	27.98	29.91	32.88	32.88	7.14	17.68	19.01	21.82	21.78	1.53
		60	2.8	90.27	92.74	93.36	93.56	78.85	26.13	27.56	31.27	31.26	6.80	16.53	17.09	20.53	20.57	1.06
		120	2.0	90.71	93.10	93.63	93.70	82.54	26.67	27.68	31.58	31.59	7.74	16.09	16.33	19.67	19.68	0.71
	1000	30	5.5	90.28	92.93	93.32	93.57	68.26	26.51	28.48	31.50	31.53	4.90	17.02	18.20	20.58	20.62	0.87
		60	4.0	91.78	94.18	94.73	94.79	78.11	28.04	29.47	32.81	32.78	5.20	16.13	16.81	20.15	20.19	0.67
		120	2.5	82.19	85.90	86.91	87.01	65.62	21.47	22.72	26.30	26.31	3.74	13.95	14.22	17.75	17.76	0.30

Table 3 Estimated degrees of freedoms of T_{new}, T_{ZX} under various settings

Model	p	n	$\rho = 0.1$		$\rho = 0.5$		$\rho = 0.9$	
			T_{new}	T_{ZX}	T_{new}	T_{ZX}	T_{new}	T_{ZX}
1	50	30	13.47	35.11	1.45	4.30	1.35	1.26
		60	9.78	34.33	1.24	4.01	1.16	1.25
		120	8.28	33.86	1.14	3.89	1.08	1.24
	500	30	2.45	99.62	1.36	4.62	1.34	1.27
		60	1.91	90.85	1.17	4.26	1.16	1.25
		120	1.71	87.18	1.08	4.11	1.08	1.24
	1000	30	1.83	112.09	1.35	4.60	1.34	1.27
		60	1.51	100.47	1.16	4.26	1.16	1.25
		120	1.38	95.73	1.08	4.13	1.08	1.24
2	50	30	11.29	30.03	1.47	4.20	1.35	1.26
		60	8.96	30.97	1.25	3.98	1.16	1.24
		120	8.02	31.88	1.15	3.87	1.08	1.24
	500	30	2.79	90.98	1.36	4.56	1.34	1.27
		60	2.07	87.20	1.17	4.25	1.16	1.25
		120	1.79	84.89	1.09	4.11	1.08	1.24
	1000	30	2.04	105.61	1.35	4.58	1.34	1.27
		60	1.60	97.70	1.17	4.27	1.16	1.25
		120	1.42	94.58	1.08	4.11	1.08	1.24
3	50	30	12.05	28.73	1.47	4.22	1.35	1.26
		60	9.53	30.64	1.25	3.99	1.16	1.24
		120	8.33	31.90	1.15	3.87	1.08	1.24
	500	30	2.83	92.82	1.36	4.59	1.34	1.27
		60	2.06	88.31	1.17	4.26	1.16	1.25
		120	1.78	86.00	1.09	4.12	1.08	1.24
	1000	30	2.01	107.86	1.35	4.62	1.34	1.27
		60	1.58	99.41	1.17	4.28	1.16	1.25
		120	1.41	94.95	1.08	4.12	1.08	1.24

with increasing the value of ρ. This is reasonable since with increasing the value of ρ, the data variations are also increasing.

Table 3 displays the estimated approximate degrees of freedom of T_{new} and T_{ZX} under various settings. First of all, it is seen that under the same setting, the estimated approximate degrees of freedom of T_{new} is smaller than T_{ZX} in most cases. Secondly, it is seen that with increasing the values of ρ, the estimated approximate degrees of freedom of T_{new} and T_{ZX} become smaller. This shows that with increasing the data correlation, the normal approximation becomes less adequate. This explains why in terms of size control, T_{BS} and T_{CQ} perform worse with increasing the data correlation.

In summary, the simulation results presented in this section show that in terms of size control, T_{new} outperforms other tests significantly; T_{ZX} outperforms T_{BS}, T_{CQ}

Table 4 Results for testing if the mean gene expression levels of the normal colon issues and the tumor colon tissues are the same

Method	Statistic	p value	$\hat{d}$
T_{new}	1.51×10^9	1.34×10^{-5}	4.69
T_{ZX}	1.93×10^9	6.15×10^{-7}	10.87
T_{BS}	8.10	2.78×10^{-16}	–
T_{CQ}	8.60	3.83×10^{-18}	–
T_{SD}	1.30	0.096	–

and T_{SD}; T_{BS} and T_{SD} are generally comparable and are generally liberal; and T_{SD} performs well for symmetric and less correlated high-dimensional data but it is very conservative when the high-dimensional data are highly skewed or highly correlated.

4 Some interesting applications

4.1 Paired two-sample problem

One important application of the one-sample test considered in this paper is testing the mean difference for two paired samples. Suppose we have n paired observations $(\boldsymbol{x}_{11}, \boldsymbol{x}_{12}), \ldots, (\boldsymbol{x}_{n1}, \boldsymbol{x}_{n2})$ which are i.i.d., we are interested in testing the following hypotheses

$$H_0 : \mathrm{E}(\boldsymbol{x}_{11}) = \mathrm{E}(\boldsymbol{x}_{12}), \text{ versus } H_1 : \mathrm{E}(\boldsymbol{x}_{11}) \neq \mathrm{E}(\boldsymbol{x}_{12}). \tag{24}$$

Then testing (24) is equivalent to testing (3) based on the induced i.i.d. sample $\boldsymbol{y}_i = \boldsymbol{x}_{i1} - \boldsymbol{x}_{i2}, i = 1, \ldots, n$ and with $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{y}_1)$. Therefore, the one-sample test discussed previously can be used to test the hypothesis (24).

As a real data example, we consider the colon dataset provided by Alon et al. (1999). The colon dataset contains 22 normal colon tissues and 40 tumor colon tissues from 40 colon-cancer patients, with each observation consisting of 2000 gene expressions. It is of interest to check whether the mean gene expression levels of the normal and tumor colon tissues are the same. For simplicity, we remove the unpaired colon tissues and keep $n = 22$ paired colon tissues only.

As an application, we apply the tests T_{new}, T_{ZX}, T_{BS}, T_{CQ} and T_{SD} to the colon dataset to test whether the normal colon tissues and the tumor colon tissues have significantly different mean gene expression levels.

Table 4 presents the results based on the 22 paired colon issues only. It is seen that all the tests except T_{SD} strongly reject the null hypothesis. The estimated degrees of freedom of T_{new} and T_{ZX} are small, showing that the normal approximation used in T_{BS}, T_{CQ} and T_{SD} is not adequate to the respective null distributions. Therefore, the p values of T_{BS}, T_{CQ} and T_{SD} are less liable. The p value of T_{SD} indicates that T_{SD} failed to detect the difference between the gene expression levels of the normal colon tissues and the tumor colon tissues at the 5% significance level, showing that T_{SD} is

conservative in this example. This result is consistent with what we observed from the simulation results presented in Sect. 3.

4.2 One-sample problem for transposable data

In many applications, measurements of a subject can be naturally organized in a matrix, especially when the rows and columns correspond to two different sets of variables. Such a kind of data is called transposable data in Allen and Tibshirani (2010). Given n i.i.d. transposable $q \times k$ random matrices $X_1, \ldots, X_n$, Touloumis et al. (2015) considered the following testing problem on the structure of the mean matrix:

$$H_0 : \ M = \left(\mu_1 \mathbf{1}_{k_1}^\top, \ldots, \mu_g \mathbf{1}_{k_g}^\top \right), \ \text{versus } H_1 : \ M \neq \left(\mu_1 \mathbf{1}_{k_1}^\top, \ldots, \mu_g \mathbf{1}_{k_g}^\top \right), \ (25)$$

where $M = \mathrm{E}(X_1)$, $k_1, \ldots, k_g$ are positive integers such that $\sum_{i=1}^{g} k_i = k$ with at least one $k_i \geq 2$, $\mu_1, \ldots, \mu_g$ are g unknown $q \times 1$ vectors. For each $i = 1, \ldots, g$, set $P_{k_i} = I_{k_i} - J_{k_i}/k_i$ as a centering matrix of size $k_i \times k_i$. Note that the MANOVA hypothesis (1) for dependent samples can be seen as a special case of (25). Set $P = \mathrm{diag}(P_{k_1}, \ldots, P_{k_g})$, a $k \times k$ block diagonal matrix. Then testing the null hypothesis in (25) is equivalent to testing $\mathrm{vec}(MP) = \mathbf{0}$. Set

$$y_i = \mathrm{vec}(X_i P), \ i = 1, \ldots, n, \tag{26}$$

which are i.i.d. $(qk) \times 1$ random vectors. Then testing (25) based on the i.i.d. random matrices $X_1, \ldots, X_n$ is equivalent to testing (3) with the induced i.i.d. random vectors (26) and with $\mu = \mathrm{E}(y_1) = \mathrm{vec}(MP)$. Therefore, our normal reference one-sample test described in Sect. 2 can then be applied to test (25) via applying it to the induced i.i.d. random vectors (26). Similar structural hypotheses on the rows of the mean matrix can also be tested accordingly. Besides, the technical Conditions C1–C5 can be easily adapted to the original transposable data as in Touloumis et al. (2015), so asymptotic results derived in Sect. 2 also apply here. To test (25), Touloumis et al. (2015) constructed a test using U-statistics as in T_{CQ} of Chen and Qin (2010). Like T_{CQ}, their test requires some strong assumptions so that a normal approximation to the null distribution of the test statistic is valid.

As a real data example, we consider the following mean matrix structure hypothesis studied by Touloumis et al. (2015) on the glioblastoma (GB) transposable dataset provided by Sottoriva et al. (2013):

$$H_0 : \ M = (\mu_1, \mu_2, \mu_3 \mathbf{1}_5^\top), \ \text{versus } H_1 : H_0 \text{ is not true}, \tag{27}$$

where the columns of M represent the mean gene expression patterns of different brain compartments, with μ_1 corresponding to the tumor margin (MA), μ_2 corresponding to the sub-ventricular zone (SVZ, normal brain tissue that surrounds the tumor mass), and $\mu_3, \ldots, \mu_7$ corresponding to 5 different fragments in the tumor mass such that earlier fragments are closer to MA and later fragments closer to SVZ. The null hypothesis in (27) corresponds to the biological hypothesis of the conservation of the mean vectors

Table 5 Testing the null hypotheses in (27) and (28) for mean gene expression levels of the glioblastoma data

Hypothesis	(27)			(28)		
Method	Statistic	p value	$\hat{d}$	Statistic	p value	$\hat{d}$
T_{new}	-149.96	0.5774	33.00	8.74×10^3	2.38×10^{-26}	136.82
T_{ZX}	3447	0.5893	96.80	1.34×10^4	1.08×10^{-29}	151.98
T_{BS}	-0.2717	0.6071	–	15.38	1.11×10^{-53}	–
T_{CQ}	-0.2782	0.6096	–	14.03	4.79×10^{-45}	–
T_{TTM}	-0.2818	0.6110	–	15.24	9.22×10^{-53}	–
T_{SD}	0.0392	0.4843	–	4.429	4.74×10^{-6}	–

of gene expression levels across the tumor mass. The GB dataset consists of $n = 8$ patients for $k = 7$ mRNA samples (column variables), with each sample having $q = 16,810$ (row variables) gene expression levels measured. We apply the test T_{TTM} proposed by Touloumis et al. (2015), and T_{new}, T_{ZX}, T_{BS}, T_{CQ} and T_{SD} to the transformed data (26) to test the null hypothesis in (27). The associated p values are given in the left panel of Table 5. It is seen that all the p values are comparable and they suggest that there is not enough evidence to reject the null hypothesis in (27). Because we do not reject the null hypothesis in (27), it is of interest to further test the following hypotheses:

$$H_0: \ \boldsymbol{M} = (\boldsymbol{\mu}_1 \mathbf{1}_2^\top, \boldsymbol{\mu}_2 \mathbf{1}_5^\top), \ \text{versus } H_1: \ H_0 \text{ is not true}, \tag{28}$$

where the null hypothesis corresponds to the biological hypothesis that MA and SVZ have a common mean gene expression pattern and the 5 different fragments in the tumor mass also have a common mean gene expression pattern. The testing results are given in the right panel of Table 5. It is seen that all the tests reject the null hypothesis in (28). From Table 5, it is seen that the estimated degrees of freedom's of T_{new} and T_{ZX} are quite large, showing that the normal approximation to the respective null distributions of T_{BS}, T_{CQ}, T_{TTM} and T_{SD} may be adequate.

4.3 Two-sample problem and MANOVA

In this subsection, we show how to use the proposed one-sample test to solve problems with two or more independent samples, e.g., the two-sample problem and MANOVA, by transforming them into a one-sample problem. There is abundant literature in the high-dimensional two-sample problem and MANOVA, see Dempster (1958), Bai and Saranadasa (1996), Srivastava and Du (2008), Chen and Qin (2010), Schott (2007), Yamada and Himeno (2015), Hu et al. (2017) and references therein. One of the advantages of using the transformation method to solve k-sample problems as a one-sample problem is that heteroscedasticity can be automatically overcome so there is no need to assume a common covariance matrix for different samples (Zhang and Xu 2009, Nishiyama et al. 2013).

Given k independent normal samples $\boldsymbol{x}_{ij}$, $i = 1, \ldots, n_j$ $\overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$; $j = 1, \ldots, k$, where suppose $n_1 \leq \cdots \leq n_k$, we firstly consider testing the simple linear hypotheses

$$H_0 : \sum_{j=1}^{k} c_j \boldsymbol{\mu}_j = \boldsymbol{0}, \text{ versus } H_1 : \sum_{j=1}^{k} c_j \boldsymbol{\mu}_j \neq \boldsymbol{0}, \tag{29}$$

where $c_1, \ldots, c_k$ are some given scalars. To apply the proposed one-sample test to the above problem, we can transform the k samples into one sample by the following transformation (Anderson 2003, Sect. 5.5): for $i = 1, \ldots, n_1$,

$$\boldsymbol{y}_i = c_1 \boldsymbol{x}_{i1} + \sum_{j=2}^{k} c_j (n_1/n_j)^{1/2} \left[\boldsymbol{x}_{ij} - n_1^{-1} \sum_{\ell=1}^{n_1} \boldsymbol{x}_{\ell j} + (n_1 n_j)^{-1/2} \sum_{\ell=1}^{n_j} \boldsymbol{x}_{\ell j} \right]. \tag{30}$$

Then we have $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{n_1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\sum_{i=1}^{k} c_i \boldsymbol{\mu}_i, \sum_{i=1}^{k} c_i^2 n_1 n_i^{-1} \boldsymbol{\Sigma}_i)$. Applying the proposed one-sample test to the induced sample we can test the hypotheses (29). In particular, let $k = 2$ and $c_1 = 1$, $c_2 = -1$, hypotheses (29) reduce to the two-sample problem studied in Chen and Qin (2010), and the transformation (30) reduces to the multivariate Scheffé (1943)'s transformation, also known as Bennett (1950)'s transformation.

For the MANOVA problem, i.e., testing

$$H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k, \text{ versus } H_1 : H_0 \text{ is not true}, \tag{31}$$

where $k \geq 3$, we can use the "dimension stacking" trick described by Anderson (1963). By applying the transformation (30) $k - 1$ times, where in the j-th time, set $c_1 = 1$, $c_j = -1$ other coefficients zero, we get $k - 1$ samples $\boldsymbol{y}_{ij}$, $i = 1, \ldots, n_1$; $j = 1, \ldots, (k-1)$. By stacking $(k-1)$ observations from different samples into a single observation, i.e., define $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}^{\top}, \ldots, \boldsymbol{y}_{i(k-1)}^{\top})^{\top}$, $i = 1, \ldots, n_1$, the original k samples are transformed into one sample with mean vector $(\boldsymbol{\mu}_1^{\top} - \boldsymbol{\mu}_2^{\top}, \ldots, \boldsymbol{\mu}_1^{\top} - \boldsymbol{\mu}_k^{\top})^{\top}$, and the MANOVA problem (31) for the original k samples is also converted to the one-sample problem for the induced sample. See also Zhang and Xu (2009) for more details of this approach.

As a real data example, we consider the peripheral blood mononuclear cells (PBMC) data provided by Burczynski et al. (2006), which is a microarray data contains 22,283 gene expression levels of 42 normal, 26 ulcerative colitis (UC), and 59 Crohn's disease (CD) tissues. We apply different one-sample tests based on the transformation method to check whether the three PBMC tissues have the same mean expression levels. The testing results are given in Table 6, where all tests reject the null hypothesis that the mean gene expression levels of the three PBMC tissues are the same. This result is consistent with the result reported in Table 7 of Zhang et al. (2017), and the testing result given by the one-sample test T_{ZX} is very similar to the result given by the MANOVA test proposed by Zhang et al. (2017). It is seen that the estimated degrees

Table 6 Testing if the mean gene expression levels of the three PBMC tissues are the same

Method	Statistic	p value	$\hat{d}$
T_{new}	5.62×10^9	1.77×10^{-7}	3.59
T_{ZX}	6.74×10^9	1.06×10^{-10}	11.99
T_{BS}	12.07	7.95×10^{-34}	–
T_{CQ}	12.37	2.01×10^{-35}	–
T_{SD}	3.99	3.30×10^{-5}	–

of freedom's of T_{new} and T_{ZX} are quite small, indicating that the normal approximation to the respective null distributions of T_{BS}, T_{CQ} and T_{SD} is not adequate.

4.4 One-sample problem for heavy tailed data

Direct applications of T_{new}, T_{ZX}, T_{BS}, T_{CQ} and T_{SD} to one-sample problem for heavy tailed high-dimensional data are often less powerful. To overcome this difficulty, one may apply these tests to the induced one sample yielded from the following multivariate spatial sign transformation:

$$\boldsymbol{u}_i = U(\boldsymbol{y}_i) = \begin{cases} \frac{\boldsymbol{y}_i}{||\boldsymbol{y}_i||}, & \boldsymbol{y}_i \neq \boldsymbol{0}, \\ 0, & \boldsymbol{y}_i = \boldsymbol{0}. \end{cases}, \ i = 1, \ldots, n. \tag{32}$$

For example, Wang et al. (2015) and Zhou et al. (2019) successfully apply T_{CQ} and T_{ZX} to the induced sample (32) for elliptically distributed high-dimensional data, respectively.

To compare the performance of T_{new}, T_{ZX}, T_{BS}, T_{CQ} and T_{SD} on the induced one sample (32) for heavy tailed high-dimensional data, we conduct the following simulation study. We generate a heavy tailed high-dimensional sample using $\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{z}_i$, $i = 1, \ldots, n$ where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are specified as in Sect. 3 and $\boldsymbol{z}_i$, $i = 1, \ldots, n$ are generated using the following two models:

– Model 4: z_{ir}, $r = 1, \ldots, p$ i.i.d. follow a Gaussian mixture $0.9\mathcal{N}(0, 1) + 0.1\mathcal{N}(0, 9)$.
– Model 5: $\boldsymbol{z}_i = \boldsymbol{w}_i/\sqrt{0.3}$, with $\boldsymbol{w}_i$ following a p-dimensional multivariate t-distribution with 3 degrees of freedom, mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{I}_p$.

Tables 7, 8 and 9 present the empirical sizes, powers of the tests and estimated degrees of freedoms of T_{new} and T_{ZX}, respectively. As expected, the conclusions drawn from these three tables are similar to those drawn from Tables 1, 2 and 3 in Sect. 3. In particular, in terms of size control, T_{new} again outperforms other tests significantly.

5 Concluding remarks

In this paper, we propose and study a normal reference test with three-cumulant matched χ^2-approximation for the one-sample problem for high-dimensional data. A

Table 7 Empirical sizes (in %) of the tests for heavy tailed distributions

Model	p	n	$\rho=0.1$					$\rho=0.5$					$\rho=0.9$				
			T_{new}	T_{ZX}	T_{BS}	T_{CQ}	T_{SD}	T_{new}	T_{ZX}	T_{BS}	T_{CQ}	T_{SD}	T_{new}	T_{ZX}	T_{BS}	T_{CQ}	T_{SD}
4	50	30	5.83	6.59	7.32	6.89	6.67	6.06	6.87	8.02	7.82	5.45	6.01	6.69	7.83	7.58	2.70
		60	5.43	6.10	6.84	6.61	6.08	5.83	6.71	7.77	7.68	4.98	5.83	6.15	7.78	7.61	2.08
		120	4.76	5.33	6.28	6.16	5.13	5.17	5.76	6.86	6.81	4.22	5.14	5.32	6.83	6.74	1.77
	500	30	5.89	7.44	7.60	7.34	5.55	6.37	7.29	8.12	7.90	3.52	6.12	6.61	7.95	7.71	1.04
		60	4.87	6.35	6.57	6.45	5.12	5.77	6.50	7.56	7.46	2.52	5.36	5.70	7.14	6.99	0.42
		120	4.77	6.38	6.71	6.65	5.34	5.49	6.22	7.20	7.15	2.35	5.46	5.68	7.28	7.19	0.33
	1000	30	5.76	7.43	7.59	7.35	5.01	6.54	7.65	8.49	8.28	2.70	5.92	6.53	7.90	7.59	0.57
		60	4.93	6.63	6.99	6.80	4.67	5.81	6.49	7.56	7.39	2.13	5.47	5.77	7.35	7.20	0.38
		120	5.29	6.87	7.23	7.15	5.49	5.32	6.09	7.10	6.99	1.60	5.04	5.25	6.92	6.85	0.21
5	50	30	5.18	5.87	6.62	6.08	6.16	6.54	7.37	8.38	8.06	6.00	5.69	6.23	7.31	7.15	2.51
		60	4.87	5.55	6.33	6.08	5.49	5.28	6.04	7.12	7.00	4.38	5.69	5.97	7.54	7.33	2.27
		120	5.31	5.82	6.72	6.61	5.72	5.43	6.02	7.07	6.98	4.26	5.12	5.28	6.78	6.70	1.87
	500	30	6.24	7.52	7.73	7.48	5.89	6.26	7.30	8.24	8.03	3.58	6.28	6.80	8.21	7.89	0.92
		60	5.18	6.64	6.91	6.78	5.37	5.72	6.48	7.59	7.45	2.80	5.33	5.62	7.08	6.93	0.55
		120	4.90	6.43	6.83	6.74	5.49	5.07	6.02	7.13	7.05	2.29	4.79	5.00	6.63	6.56	0.52
	1000	30	6.37	8.05	8.31	7.93	5.44	6.37	7.49	8.43	8.15	2.78	6.02	6.49	7.63	7.36	0.67
		60	5.01	6.58	6.83	6.73	4.75	5.51	6.30	7.28	7.11	1.80	5.46	5.82	7.28	7.19	0.44
		120	5.17	6.83	7.14	7.04	5.39	5.29	5.94	7.16	7.06	1.68	5.47	5.66	7.10	6.99	0.24
ARE			8.40	31.57	40.61	36.52	11.02	15.37	31.71	52.31	49.30	37.62	11.80	18.41	47.27	43.96	78.34

Table 8 Empirical powers (in %) of the tests for heavy tailed distributions

Model	p	n	δ	ρ = 0.1					ρ = 0.5					ρ = 0.9				
				T_{new}	T_{ZX}	T_{BS}	T_{CQ}	T_{SD}	T_{new}	T_{ZX}	T_{BS}	T_{CQ}	T_{SD}	T_{new}	T_{ZX}	T_{BS}	T_{CQ}	T_{SD}
4	50	30	1.2	64.54	66.47	68.02	67.01	65.42	24.09	26.40	28.91	28.12	22.57	16.75	17.97	20.81	20.29	9.20
		60	0.8	58.02	59.96	61.94	61.49	58.98	20.80	22.53	25.23	24.93	18.52	14.37	14.98	17.87	17.63	6.52
		120	0.6	64.04	65.92	67.91	67.69	65.31	22.52	24.48	27.54	27.44	19.35	15.47	15.98	19.39	19.16	6.89
	500	30	3.2	66.36	71.38	72.09	71.19	65.29	17.96	20.31	22.24	21.60	10.82	13.28	14.18	16.55	16.13	2.96
		60	2.0	52.54	58.02	59.20	58.62	53.41	14.88	16.41	18.52	18.25	8.19	11.06	11.57	13.80	13.58	1.62
		120	1.4	51.57	56.81	58.03	57.85	53.52	13.66	15.03	17.22	17.11	6.98	10.47	10.84	13.51	13.37	1.25
	1000	30	4.0	55.84	60.99	61.58	60.67	51.79	15.75	17.52	19.09	18.57	7.90	11.73	12.88	15.09	14.60	1.70
		60	3.2	67.88	72.91	73.86	73.41	66.57	17.78	19.48	21.67	21.32	8.30	12.03	12.59	15.29	15.01	1.07
		120	1.8	44.36	49.80	50.84	50.61	44.83	12.82	14.47	16.39	16.26	5.04	9.27	9.66	12.24	12.13	0.53
5	50	30	1.2	98.50	98.76	98.92	98.84	98.56	62.28	65.35	68.62	67.60	58.18	44.65	47.05	52.34	51.09	27.91
		60	0.8	97.67	97.99	98.21	98.16	97.80	56.66	59.76	63.88	63.31	52.10	39.00	40.38	46.46	45.86	20.89
		120	0.6	99.02	99.12	99.27	99.25	99.07	62.79	65.89	70.58	70.24	57.34	44.65	45.57	53.01	52.64	24.15
	500	30	3.2	98.98	99.51	99.57	99.51	98.77	47.69	50.91	54.07	53.18	33.71	31.54	33.60	38.24	37.41	8.54
		60	2.0	97.38	98.41	98.58	98.47	97.36	36.77	39.56	42.97	42.57	23.50	24.33	25.18	29.73	29.23	4.68
		120	1.4	97.45	98.41	98.66	98.63	97.72	36.00	38.67	42.74	42.39	22.14	23.88	24.49	29.44	29.21	3.74
	1000	30	4.0	97.04	98.15	98.27	98.12	95.54	37.29	40.17	43.10	42.17	21.93	25.83	27.35	30.67	30.13	4.98
		60	3.2	99.69	99.82	99.88	99.86	99.55	47.52	50.50	54.33	53.85	26.67	32.03	33.14	38.29	37.68	4.48
		120	1.8	93.98	96.32	96.67	96.63	94.22	29.89	32.27	35.80	35.58	14.33	19.41	19.97	24.18	24.01	1.78

Table 9 Estimated approximate degrees of freedom when $\delta = 0$ for heavy tailed distributions

Model	p	n	$\rho = 0.1$		$\rho = 0.5$		$\rho = 0.9$	
			T_{new}	T_{ZX}	T_{new}	T_{ZX}	T_{new}	T_{ZX}
4	50	30	20.08	40.43	1.79	8.47	1.40	2.31
		60	14.26	38.70	1.47	7.95	1.19	2.24
		120	12.16	37.93	1.33	7.72	1.10	2.19
	500	30	3.28	127.71	1.47	9.67	1.38	2.35
		60	2.37	116.35	1.22	8.93	1.17	2.28
		120	2.06	111.84	1.12	8.61	1.08	2.23
	1000	30	2.32	148.12	1.45	9.70	1.38	2.35
		60	1.76	132.86	1.21	9.00	1.17	2.27
		120	1.55	125.52	1.11	8.67	1.08	2.24
5	50	30	19.93	40.29	1.80	8.53	1.40	2.31
		60	14.25	38.69	1.47	7.97	1.19	2.23
		120	12.16	37.94	1.33	7.72	1.10	2.20
	500	30	3.30	128.35	1.47	9.68	1.38	2.36
		60	2.37	116.61	1.22	8.93	1.17	2.27
		120	2.05	111.42	1.12	8.63	1.08	2.23
	1000	30	2.31	147.56	1.45	9.71	1.38	2.36
		60	1.76	131.70	1.21	8.99	1.17	2.28
		120	1.55	125.54	1.11	8.67	1.08	2.24

simulation study shows that in terms of size control, the proposed test outperforms several existing competitors. The proposed test can also be applied for testing one-sample problems with other types of data via some simple transformations. When the data are normally distributed, it is known that the estimated approximation parameters are ratio-consistent. However, whether they are also ratio-consistent for non-normal high-dimensional data is interesting and warranted. Since the normal reference test with the 3-c matched χ^2-approximation for one-sample problems for high-dimensional data has much better size control than several existing tests, it is also interesting and warranted to extend this normal reference test to other high-dimensional hypothesis testing problems.

Appendix: Technical proofs

We first prove the following useful lemma.

Lemma 1 *Let $W \sim W_p(v, \Sigma/v)$ denote a Wishart distribution with v degrees of freedom and a covariance matrix Σ/v. Assume that $p/v^2 \longrightarrow 0$ as $v, p \to \infty$. Then the unbiased and ratio-consistent estimator of $\mathrm{tr}(\Sigma^3)$ is given by*

$$\widehat{\mathrm{tr}(\Sigma^3)} = \frac{v^4}{(v-1)(v+4)(v^2-4)} \left[\mathrm{tr}(W^3) - \frac{3}{v} \mathrm{tr}(W) \mathrm{tr}(W^2) + \frac{2}{v^2} \mathrm{tr}^3(W) \right]. \quad \text{(A.33)}$$

Proof of Lemma 1. Let $V = vW$. Then $V \sim W_p(v, \Sigma)$. To show (A.33) is equivalent to show

$$\widehat{\mathrm{tr}(\Sigma^3)} = \frac{v}{(v-1)(v+4)(v^2-4)} \left[\mathrm{tr}(V^3) - \frac{3}{v} \mathrm{tr}(V) \mathrm{tr}(V^2) + \frac{2}{v^2} \mathrm{tr}^3(V) \right] \quad \text{(A.34)}$$

is unbiased and ratio consistent for $\mathrm{tr}(\Sigma^3)$. By the proof of Theorem 2.1 of Srivastava and Yanagihara (2010), we can express the above expression as $\widehat{\mathrm{tr}(\Sigma^3)} = I_1 + I_2 + I_3 + I_4$, where

$$I_1 = \frac{1}{v(v+2)(v+4)} \sum_{i=1}^p \lambda_i^3 w_{ii}^3,$$
$$I_2 = \frac{3}{(v-1)(v+2)(v+4)} \sum_{i \neq j} \lambda_i^2 \lambda_j \left(w_{ii} w_{ij}^2 - v^{-1} w_{ii}^2 w_{jj} \right),$$
$$I_3 = \frac{v}{(v-1)(v+4)(v^2-4)} \sum_{i \neq j \neq k} \lambda_i \lambda_j \lambda_k \left(w_{ij} w_{jk} w_{ki} - v^{-2} w_{ii} w_{jj} w_{kk} \right),$$
$$I_4 = \frac{-3}{(v-1)(v+4)(v^2-4)} \sum_{i \neq j \neq k} \lambda_i \lambda_j \lambda_k \left(w_{ij} w_{jk}^2 - v^{-1} w_{ii} w_{jj} w_{kk} \right),$$

where $w_{ij} = \boldsymbol{u}_i^\top \boldsymbol{u}_j$, $i, j = 1, \ldots, p$ and $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p \overset{\text{i.i.d.}}{\sim} \mathcal{N}_v(\boldsymbol{0}, \boldsymbol{I}_v)$. By Lemma 6.2 (a) and (e) of Srivastava and Yanagihara (2010), we have $\mathrm{E}(w_{ii}^3) = v(v+2)(v+4)$ and $\mathrm{Var}(w_{ii}^3) = 6v(v+2)(v+4)(3v^2+30v+80)$. Therefore,

$$\mathrm{E}(I_1) = \frac{1}{v(v+2)(v+4)} \sum_{i=1}^p \lambda_i^3 \mathrm{E}(w_{ii}^3) = \mathrm{tr}(\Sigma^3),$$
$$\mathrm{Var}(I_1) = [\frac{1}{v(v+2)(v+4)}]^2 \sum_{i=1}^p \lambda_i^6 \mathrm{Var}(w_{ii}^3) = v^{-1} \mathrm{tr}(\Sigma^6)[1 + o(1)].$$

Noting that

$$\mathrm{tr}(A^r)/\mathrm{tr}^r(A) \leq 1, \ r = 1, 2, \ldots \quad \text{(A.35)}$$

hold for any nonnegative matrix A, we have $\mathrm{Var}[I_1/\mathrm{tr}(\Sigma^3)] < v^{-1}[1 + o(1)]$. It follows that $I_1/\mathrm{tr}(\Sigma^3) \longrightarrow 1$ as $v \to \infty$.

Following the proof of Theorem 2.1 of Srivastava and Yanagihara (2010) closely, let $r_{ij} = w_{ii} w_{ij}^2 - v^{-1} w_{ii}^2 w_{jj}$. Then we have $\mathrm{E}(r_{ij}) = 0$ and $\mathrm{Cov}(r_{ij}, r_{kl}) = 0$ where $(i \neq j)$, $(k \neq l)$ and $(i, j) \neq (k, l)$. In addition, by Lemma 6.2 (f), (g) and (h) of Srivastava and Yanagihara (2010), we have

$$\mathrm{Var}(r_{ij}) = \mathrm{E}\left(w_{ii} w_{ij}^2 - v^{-1} w_{ii}^2 w_{jj} \right)^2 = \mathrm{E}\left(w_{ii}^2 w_{ij}^4 - 2v^{-1} w_{ii}^3 w_{ij}^2 w_{jj} + v^{-2} w_{ii}^4 w_{jj}^2 \right)$$
$$= O(v^4) - O(v^{-1} v^5) + O(v^{-2} v^6) = O(v^4).$$

It follows that $E(I_2) = 0$ and

$$\text{Var}(I_2) = 9\left[(v-1)(v+2)(v+4)\right]^{-2} \sum_{i \neq j} \lambda_i^4 \lambda_j^2 \text{Var}(r_{ij})$$
$$\leq 9v^{-2} \text{tr}(\boldsymbol{\Sigma}^4) \text{tr}(\boldsymbol{\Sigma}^2)[1+o(1)].$$

It follows from (A.35) that $\text{Var}\left[I_2/\text{tr}(\boldsymbol{\Sigma}^3)\right] \leq [\text{tr}(\boldsymbol{\Sigma}^4)(9\kappa)]/[\text{tr}^2(\boldsymbol{\Sigma}^2)v^2][1+o(1)] \leq 9\kappa/v^2[1+o(1)]$, where

$$\kappa = \frac{\text{tr}^3(\boldsymbol{\Sigma}^2)}{\text{tr}^2(\boldsymbol{\Sigma}^3)}. \tag{A.36}$$

By Theorem 5 of Zhang et al. (2020), we have $\kappa/p \leq 1$. Since $p/v^2 \longrightarrow 0$ as $v, p \to \infty$, we have $\kappa/v^2 \longrightarrow 0$ as $v, p \to \infty$. Therefore, we have that $I_2 = o_p[\text{tr}(\boldsymbol{\Sigma}^3)]$.
  Similarly, we have $E(I_3) = 0$, $E(I_4) = 0$ and

$$\text{Var}[I_3/\text{tr}(\boldsymbol{\Sigma}^3)] \leq \kappa v^{-3}[1+o(1)] \longrightarrow 0, \quad \text{Var}[I_4/\text{tr}(\boldsymbol{\Sigma}^3)] \leq \kappa v^{-4}[1+o(1)] \longrightarrow 0,$$

showing that $I_3 = o_p[\text{tr}(\boldsymbol{\Sigma}^3)]$ and $I_4 = o_p[\text{tr}(\boldsymbol{\Sigma}^3)]$. It follows that $\widehat{\text{tr}(\boldsymbol{\Sigma}^3)}$ is an unbiased and ratio-consistent estimator of $\text{tr}(\boldsymbol{\Sigma}^3)$. The lemma is then proved. □

  We now prove the main results.

Proof of Theorem 1 We first prove (a). We shall use the characteristic function $(\psi_X(t) = E(e^{itX})$ for a random variable $X)$ method. Set $\boldsymbol{x}_i = \boldsymbol{y}_i - \boldsymbol{\mu}$, $i = 1, \ldots, n$ and hence $\bar{\boldsymbol{x}} = \bar{\boldsymbol{y}} - \boldsymbol{\mu}$. Set $\boldsymbol{w}_{n,p} = \sqrt{n}\bar{\boldsymbol{x}}$. By (10), we have

$$\tilde{T}_{n,p,0} = \left[\|\boldsymbol{w}_{n,p}\|^2 - \text{tr}(\boldsymbol{\Sigma})\right]/\sqrt{2\text{tr}(\boldsymbol{\Sigma}^2)}[1+o(1)], \tag{A.37}$$

since $\text{tr}(\hat{\boldsymbol{\Sigma}})/\text{tr}(\boldsymbol{\Sigma}) \longrightarrow 1$ as $n, p \to \infty$ (See Proof of Theorem 9 in Zhang et al. 2020). Further, we have $E(\boldsymbol{w}_{n,p}) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{w}_{n,p}) = \boldsymbol{\Sigma}$. Write $\boldsymbol{w}_{n,p} = \sum_{r=1}^p \xi_{n,p,r} \boldsymbol{u}_{p,r}$, where $\xi_{n,p,r} = \boldsymbol{w}_{n,p}^\top \boldsymbol{u}_{p,r}$, and $\boldsymbol{u}_{p,1}, \ldots, \boldsymbol{u}_{p,p}$ denote the eigenvectors associated with the eigenvalues $\lambda_{p,1}, \ldots, \lambda_{p,p}$ of $\boldsymbol{\Sigma}$ in the descending order. We have $E(\xi_{n,p,r}) = 0$, and $\text{Var}(\xi_{n,p,r}) = \lambda_{p,r}$, $r = 1, 2, \ldots$, and $\xi_{n,p,r}$, $r = 1, \ldots, p$, are uncorrelated.
  By Lemma S.4 of Zhang et al. (2020), we have

$$\text{Var}(\xi_{n,p,r}^2) = 2\lambda_{p,r}^2 + [E(\boldsymbol{x}_1^\top \boldsymbol{u}_{p,r})^4 - 3\lambda_{p,r}^2]/n.$$

Under Conditions C1 and C2, by some simple algebra, we have $E(\boldsymbol{x}_1^\top \boldsymbol{u}_{p,r})^4 \leq (3+\Delta)\lambda_{p,r}^2$. Thus, we have

$$\text{Var}(\xi_{n,p,r}^2) \leq (2+\Delta/n)\lambda_{p,r}^2, \quad r = 1, 2, \ldots. \tag{A.38}$$

It follows from (A.37) that

$$\tilde{T}_{n,p,0} = \sum_{r=1}^p (\xi_{n,p,r}^2 - \lambda_{p,r})/\sqrt{2\text{tr}(\boldsymbol{\Sigma}^2)}[1+o(1)].$$

Set $\tilde{T}_{n,p,0}^{(q)} = \sum_{r=1}^{q} (\xi_{n,p,r}^2 - \lambda_{p,r})/\sqrt{2\operatorname{tr}(\boldsymbol{\Sigma}^2)}$ where $q < p$. Then

$$
\begin{aligned}
\mathrm{E}\left(\tilde{T}_{n,p,0} - \tilde{T}_{n,p,0}^{(q)}\right)^2 &= \mathrm{E}\left[\sum_{r=q+1}^{p} (\xi_{n,p,r}^2 - \lambda_{p,r})/\sqrt{2\operatorname{tr}(\boldsymbol{\Sigma}^2)}\right]^2 [1 + o(1)] \\
&= \operatorname{Var}\left(\sum_{r=q+1}^{p} \xi_{n,p,r}^2\right) / \left[2\operatorname{tr}(\boldsymbol{\Sigma}^2)\right] [1 + o(1)] \\
&\leq \left[\sum_{r=q+1}^{p} \sqrt{\operatorname{Var}(\xi_{n,p,r}^2)}\right]^2 / \left[2\operatorname{tr}(\boldsymbol{\Sigma}^2)\right] [1 + o(1)].
\end{aligned}
$$

By (A.38), we have

$$
\left[\sum_{r=q+1}^{p} \sqrt{\operatorname{Var}(\xi_{n,p,r}^2)}\right]^2 / \left[2\operatorname{tr}(\boldsymbol{\Sigma}^2)\right] \leq (1 + \Delta/n)\left(\sum_{r=q+1}^{p} \rho_{p,r}\right)^2.
$$

It follows that

$$
\begin{aligned}
|\psi_{\tilde{T}_{n,p,0}}(t) - \psi_{\tilde{T}_{n,p,0}^{(q)}}(t)| &\leq |t|\left[\mathrm{E}(\tilde{T}_{n,p,0} - \tilde{T}_{n,p,0}^{(q)})^2\right]^{1/2} \\
&\leq |t|(1 + \Delta/n)^{1/2}\left(\sum_{r=q+1}^{p} \rho_{p,r}\right)[1 + o(1)].
\end{aligned} \tag{A.39}
$$

Set $\zeta^{(q)} = \sum_{r=1}^{q} \rho_r (A_r - 1)/\sqrt{2}$, we have

$$
\begin{aligned}
\left|\psi_{\tilde{T}_{n,p,0}}(t) - \psi_\zeta(t)\right| &\leq \left|\psi_{\tilde{T}_{n,p,0}}(t) - \psi_{\tilde{T}_{n,p,0}^{(q)}}(t)\right| + \left|\psi_{\tilde{T}_{n,p,0}^{(q)}}(t) - \psi_{\tilde{T}_{p,0}^{(q)}}(t)\right| \\
&\quad + \left|\psi_{\tilde{T}_{p,0}^{(q)}}(t) - \psi_{\zeta^{(q)}}(t)\right| + \left|\psi_{\zeta^{(q)}}(t) - \psi_\zeta(t)\right|.
\end{aligned}
$$

By similar arguments in Proof of Theorem 2 of Zhang et al. (2020), we can show under Conditions C1–C3 all the four terms on the right hand side of the previous inequality converge to zero as $n, p \to \infty$, so the first expression of (11) follows. In particular, the convergence of the first term can be derived from (A.39) and Condition C3, the convergence of the second term is ensured by the standard central limit theorem, and the convergence of the last two terms is due to Condition C3.

Notice that when the data (2) are normally distributed, Conditions C1–C2 are automatically satisfied so that under Condition C3, the second expression of (11) follows immediately since under the normality assumption, we have $T_{n,p,0} \overset{d}{=} T_{n,p,0}^*$.

We now prove (b). By Lemma 8.1 of Pauly et al. (2015), Condition C5 is equivalent to the condition "$\operatorname{tr}(\boldsymbol{\Sigma}^4)/\operatorname{tr}^2(\boldsymbol{\Sigma}^2) = o(1)$" imposed by Chen and Qin (2010). Therefore, under Conditions C1, C2 and C5, the proofs of the asymptotic normality of $\tilde{T}_{n,p,0}$ and $\tilde{T}_{n,p,0}^*$ are along the same lines as the one given by Chen and Qin (2010) for the asymptotic normality of their test statistic.

Finally, we prove (13). We have $\sup_x \left| \Pr(T_{n,p,0} \leq x) - \Pr(T^*_{n,p,0} \leq x) \right| \leq$
$\sup_x \left| \Pr(\tilde{T}_{n,p,0} \leq \tilde{x}) - \Pr(\zeta \leq \tilde{x}) \right| + \sup_x \left| \Pr(\tilde{T}^*_{n,p,0} \leq \tilde{x}) - \Pr(\zeta \leq \tilde{x}) \right|$, where
$\tilde{x} = [x - \text{tr}(\hat{\Sigma})]/\sqrt{\frac{2n}{(n-1)} \text{tr}(\Sigma^2)}$. Under the conditions of (a), the two terms on the
right hand side of previous inequality both converge to zero because both $\tilde{T}_{n,p,0}$ and
$\tilde{T}^*_{n,p,0}$ converge to ζ in distribution. Under the conditions of (b), the proof of (13) are
along the same lines as the above and hence are omitted for space saving.   □

Proof of Theorem 2 Under the local alternative (21), $T_{n,p} = \left(T_{n,p,0} + n\|\boldsymbol{\mu}\|^2 \right) [1 + o_p(1)]$. In addition, under the given conditions, we have $\hat{\beta}_0/\beta_0 \xrightarrow{P} 1$, $\hat{\beta}_1/\beta_1 \xrightarrow{P}$
1 and $\hat{d}/d \xrightarrow{P} 1$ as $n, p \to \infty$. We first prove (a). Under Conditions C1,
C2 and C3, Theorem 1(a) indicates that as $n, p \to \infty$, we have $\tilde{T}_{n,p,0} = T_{n,p,0}/\sqrt{\frac{2n}{(n-1)} \text{tr}(\Sigma^2)} \xrightarrow{\mathcal{L}} \zeta$. It follows that as $n, p \to \infty$, we have

$$\Pr\left[T_{n,p} \geq \hat{\beta}_0 + \hat{\beta}_1 \chi_{\hat{d}}^2(\alpha) \right] = \Pr\left[\tilde{T}_{n,p,0} \geq \frac{\beta_0 + \beta_1 \chi_{\hat{d}}^2(\alpha)}{\sqrt{\frac{2n}{(n-1)} \text{tr}(\Sigma^2)}} - \frac{n\|\boldsymbol{\mu}\|^2}{\sqrt{\frac{2n}{(n-1)} \text{tr}(\Sigma^2)}} \right] [1 + o(1)] \tag{A.40}$$
$$= \Pr\left[\zeta \geq \frac{\chi_{\hat{d}}^2(\alpha) - d}{\sqrt{2d}} - \frac{n\|\boldsymbol{\mu}\|^2}{\sqrt{2\text{tr}(\Sigma^2)}} \right] [1 + o(1)].$$

We now prove (b). Under the given conditions, Theorem 1(b) indicates that as
$n \to \infty$, we have $\tilde{T}_{n,p,0} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ and by Theorem 5, as $n, p \to \infty$, we have
$d \to \infty$ and $[\chi_{\hat{d}}^2(\alpha) - d]/\sqrt{2d} \to z_\alpha$ where z_α denotes the upper 100α-percentile
of $\mathcal{N}(0,1)$. Then by (A.40), as $n, p \to \infty$, we have

$$\Pr\left[T_{n,p} \geq \hat{\beta}_0 + \hat{\beta}_1 \chi_{\hat{d}}^2(\alpha) \right] = \Pr\left[\tilde{T}_{n,p,0} \geq \frac{\beta_0 + \beta_1 \chi_{\hat{d}}^2(\alpha)}{\sqrt{\frac{2n}{(n-1)} \text{tr}(\Sigma^2)}} - \frac{n\|\boldsymbol{\mu}\|^2}{\sqrt{\frac{2n}{(n-1)} \text{tr}(\Sigma^2)}} \right] [1 + o(1)]$$
$$= \Phi\left[-z_\alpha + \frac{n\|\boldsymbol{\mu}\|^2}{\sqrt{2\text{tr}(\Sigma^2)}} \right] [1 + o(1)],$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of $\mathcal{N}(0,1)$. The proof is
complete.   □

Proof of Theorem 3 Let $\boldsymbol{x}_i = \boldsymbol{y}_i - \boldsymbol{\mu}$, $i = 1, \ldots, n$. Then $\boldsymbol{x}_i$, $i = 1, \ldots, n$ are
i.i.d. with $\text{E}(\boldsymbol{x}_i) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{x}_i) = \Sigma$. It is easy to verify that $T_{n,p,0} = 2(n - 1)^{-1} \sum_{i<j} \boldsymbol{x}_i^\top \boldsymbol{x}_j$. It follows that $\text{E}(T_{n,p,0}) = 0$ and

$$\text{Var}(T_{n,p,0}) = \text{E}(T_{n,p,0}^2) = 4(n-1)^{-2} \sum_{i<j} \text{E}(\boldsymbol{x}_i^\top \boldsymbol{x}_j)^2 = \frac{2n}{(n-1)} \text{tr}(\Sigma^2).$$

Furthermore, we have

$$
\begin{aligned}
E(T_{n,p,0}^3) &= 8(n-1)^{-3} \, E\left[\sum_{i<j} (x_i^\top x_j) \right]^3 \\
&= 8(n-1)^{-3} \, E\left[\sum_{i<j} (x_i^\top x_j)^3 + 3 \sum_{*} (x_i^\top x_j)^2 (x_u^\top x_v) \right. \\
&\quad \left. + 6 \sum_{**} (x_i^\top x_j)(x_u^\top x_v)(x_\alpha^\top x_\beta) \right] \\
&= 8(n-1)^{-3} \left\{ \frac{n(n-1)}{2} E(x_1^\top x_2)^3 + n(n-1)(n-2) \, E[(x_1^\top x_2)(x_2^\top x_3)(x_3^\top x_1)] \right\} \\
&= \frac{8n(n-2)}{(n-1)^2} \operatorname{tr}(\boldsymbol{\Sigma}^3) + \frac{4n\Upsilon}{(n-1)^2},
\end{aligned}
$$

where $\Upsilon = E(x_1^\top x_2)^3 = E[(y_1 - \mu)^\top (y_2 - \mu)]^3$, $*$ means "$i < j$, $u < v$" and "$(i,j) \neq (u,v)$" while $**$ means "$i < j$, $u < v$, $\alpha < \beta$" and "(i,j), (u,v), (α, β) are not mutually equal to each other." The proof is complete. $\qquad\square$

Proof of Theorem 4 We first show (a). Under Condition C1, we have $y_i = \mu + \boldsymbol{\Gamma} z_i$, $i = 1, \ldots, n$ where z_i, $i = 1, \ldots, n$ are i.i.d. with $E(z_i) = \mathbf{0}$ and $\operatorname{Cov}(z_i) = \boldsymbol{I}_p$ and $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top$. It follows that $\Upsilon = E[(y_1 - \mu)^\top (y_2 - \mu)]^3 = E(z_1^\top \boldsymbol{\Omega} z_2)^3$ where $\boldsymbol{\Omega} = \boldsymbol{\Gamma}^\top \boldsymbol{\Gamma}$. By Jensen's inequality, we have

$$
\Upsilon = E(z_1^\top \boldsymbol{\Omega} z_2)^3 \leq \left[E(z_1^\top \boldsymbol{\Omega} z_2)^4 \right]^{3/4}. \tag{A.41}
$$

Denote the (i,j)-th entry of $\boldsymbol{\Omega}$ as w_{ij}, $i, j, = 1, \ldots, p$. Under Conditions C1 and C2, from the proof of Lemma 6.2 of Srivastava and Kubokawa (2013) (p. 215), we have

$$
E\left(z_1^\top \boldsymbol{\Omega} z_2\right)^4 = \Delta^2 \sum_{i=1}^p \sum_{j=1}^p w_{ij}^4 + 6\Delta \sum_{i=1}^p \left(\sum_{j=1}^p w_{ij}^2 \right)^2 + 6\operatorname{tr}(\boldsymbol{\Omega}^4) + 3\operatorname{tr}^2(\boldsymbol{\Omega}^2), \tag{A.42}
$$

where Δ is given in Condition C2. Notice that we have

$$
\begin{aligned}
\sum_{i=1}^p \sum_{j=1}^p w_{ij}^4 &\leq \left(\sum_{i=1}^p \sum_{j=1}^p w_{ij}^2 \right)^2 = \operatorname{tr}^2(\boldsymbol{\Omega}^2), \\
\sum_{i=1}^p \left(\sum_{j=1}^p w_{ij}^2 \right)^2 &\leq \left(\sum_{i=1}^p \sum_{j=1}^p w_{ij}^2 \right)^2 = \operatorname{tr}^2(\boldsymbol{\Omega}^2), \\
\operatorname{tr}(\boldsymbol{\Omega}^4) &\leq \operatorname{tr}^2(\boldsymbol{\Omega}^2).
\end{aligned}
$$

These, together with (A.42), imply that $E\left(z_1^\top \boldsymbol{\Omega} z_2\right)^4 \leq (\Delta^2 + 6\Delta + 9) \operatorname{tr}^2(\boldsymbol{\Omega}^2)$. Then by (A.41), we have

$$
\Upsilon \leq (\Delta^2 + 6\Delta + 9)^{3/4} \operatorname{tr}^{3/2}(\boldsymbol{\Omega}^2) = (\Delta^2 + 6\Delta + 9)^{3/4} \operatorname{tr}^{3/2}(\boldsymbol{\Sigma}^2), \tag{A.43}
$$

where we use the fact that $\mathrm{tr}(\boldsymbol{\Omega}^2) = \mathrm{tr}(\boldsymbol{\Gamma}^\top \boldsymbol{\Gamma} \boldsymbol{\Gamma}^\top \boldsymbol{\Gamma}) = \mathrm{tr}(\boldsymbol{\Gamma} \boldsymbol{\Gamma}^\top \boldsymbol{\Gamma} \boldsymbol{\Gamma}^\top) = \mathrm{tr}(\boldsymbol{\Sigma}^2)$.

We now show (b). First of all, under Conditions C1 and C2, by (22) and (A.43), we have

$$\delta \leq 1 + \frac{(\Delta^2 + 6\Delta + 9)^{3/4}\, \mathrm{tr}^{3/2}(\boldsymbol{\Sigma}^2)}{2(n-2)\, \mathrm{tr}(\boldsymbol{\Sigma}^3)} = 1 + (\Delta^2 + 6\Delta + 9)^{3/4} \frac{\sqrt{\kappa}}{2(n-2)}, \tag{A.44}$$

where κ is given in (A.36). By Theorem 5 of Zhang et al. (2020), we have

$$1 \leq \kappa \leq \frac{\mathrm{tr}^2(\boldsymbol{\Sigma})}{\mathrm{tr}(\boldsymbol{\Sigma}^2)} \leq p. \tag{A.45}$$

Under Condition C3, as $p \to \infty$, we have $\mathrm{tr}^2(\boldsymbol{\Sigma})/\mathrm{tr}(\boldsymbol{\Sigma}^2) = (\sum_{r=1}^p \rho_{p,r})^2 \longrightarrow (\sum_{r=1}^\infty \rho_r)^2 < \infty$. It follows that under Conditions C1, C2 and C3, κ is bounded as $p \to \infty$. This, together with (A.44), implies that under Conditions C1, C2 and C3, as $n, p \to \infty$, we have $\delta = 1 + o(1)$ as $n, p \to \infty$.

Under Conditions C1, C2 and C4, by (A.45), as $n, p \to \infty$, we have $\sqrt{\kappa}/[2(n-2)] \leq \sqrt{p/[2(n-2)]^2} = o(1)$. This, together with (A.44), implies that under Conditions C1, C2 and C4, as $n, p \to \infty$, we always have $\delta = 1 + o(1)$. The proof is complete. □

Proof of Theorem 5 By (23), the skewness of $T_{n,p,0}$ is $\sqrt{8/f}$ where f is defined in (22) and under Conditions C1, C2 and C4, by Theorem 4, we have $f = d[1 + o(1)]$. On the one hand, when $\tilde{T}_{n,p,0} \overset{\mathcal{L}}{\longrightarrow} \mathcal{N}(0,1)$, the skewness of $T_{n,p,0}$ must tend to 0 as $n, p \to \infty$, it follows that we must have $d \longrightarrow \infty$ as $n, p \to \infty$. On the other hand, by (15), as $n \to \infty$, we have $d = \kappa[1 + o(1)]$ where κ is defined in (A.36). We have $\kappa = \mathrm{tr}^3(\boldsymbol{\Sigma}^2)/\mathrm{tr}^2(\boldsymbol{\Sigma}^3) \geq \mathrm{tr}(\boldsymbol{\Sigma}^2)/\lambda_{p,\max}^2$ where $\lambda_{p,\max}$ is the largest eigenvalue of $\boldsymbol{\Sigma}$. Therefore, as $d \to \infty$, we have $\kappa \longrightarrow \infty$ and $\lambda_{p,\max}^2/\mathrm{tr}(\boldsymbol{\Sigma}^2) \geq \kappa^{-1} \longrightarrow 0$ as $p \to \infty$. That is, Condition C5 holds. This, together with Conditions C1, C2 and C4, implies that by Theorem 1(b), we have $\tilde{T}_{n,p,0} \overset{\mathcal{L}}{\longrightarrow} \mathcal{N}(0,1)$. The proof is then complete. □

References

Ahmad MR, Werner C, Brunner E (2008) Analysis of high-dimensional repeated measures designs: the one sample case. Comput Stat Data Anal 53(2):416–427

Allen GI, Tibshirani R (2010) Transposable regularized covariance models with an application to missing data imputation. Ann Appl Stat 4(2):764–790

Alon U, Barkai N, Notterman D, Gish K, Ybarra S, Mack D, Levine A (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci 96(12):6745–6750

Anderson TW (1963) A test for equality of means when covariance matrices are unequal. Ann Math Stat 34(2):671–672

Anderson TW (2003) An introduction to multivariate statistical analysis. Wiley, New York

Bai ZD, Saranadasa H (1996) Effect of high dimension: by an example of a two sample problem. Stat Sin 6(2):311–329

Bai Z, Hu J, Wang C, Zhang C (2021) Test on the linear combinations of covariance matrices in high-dimensional data. Stat Pap 62:701–719

Bennett BM (1950) Note on a solution of the generalized Behrens-Fisher problem. Ann Inst Stat Math 2(1):87–90

Box GE (1954) Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. Ann Math Stat 25(2):290–302

Burczynski ME, Peterson RL, Twine NC, Zuberek KA, Brodeur BJ, Casciotti L, Maganti V, Reddy PS, Strahs A, Immermann F, Spinelli W, Schwertschlag U, Slager AM, Cotreau MM, Dorner AJ (2006) Molecular classification of Crohn's disease and ulcerative colitis patients using transcriptional profiles in peripheral blood mononuclear cells. J Mol Diagn 8(1):51–61

Chen SX, Qin YL (2010) A two-sample test for high-dimensional data with applications to gene-set testing. Ann Stat 38(2):808–835

Chen LS, Paul D, Prentice RL, Wang P (2011) A regularized Hotelling's T^2 test for pathway analysis in proteomic studies. J Am Stat Assoc 106(496):1345–1360

Dempster AP (1958) A high dimensional two sample significance test. Ann Math Stat 29(4):995–1010

Dong K, Pang H, Tong T, Genton MG (2016) Shrinkage-based diagonal Hotelling's tests for high-dimensional small sample size data. J Multivar Anal 143:127–142

Feng L, Sun F (2016) Spatial-sign based high-dimensional location test. Electron J Stat 10(2):2420–2434

Feng L, Zou C, Wang Z, Zhu L (2017) Composite T^2 test for high-dimensional data. Stat Sin 27:1419–1436

Hall P (1983) Chi squared approximations to the distribution of a sum of independent random variables. Ann Probab 11(4):1028–1036

Henze N (2002) Invariant tests for multivariate normality: a critical review. Stat Pap 43:467–506

Hu J, Bai Z, Wang C, Wang W (2017) On testing the equality of high dimensional mean vectors with unequal covariance matrices. Ann Inst Stat Math 69:365–387

Hu Z, Tong T, Genton MG (2019) Diagonal likelihood ratio test for equality of mean vectors in high-dimensional data. Biometrics 75:256–267

Katayama S, Kano Y, Srivastava MS (2013) Asymptotic distributions of some test criteria for the mean vector with fewer observations than the dimension. J Multivar Anal 116:410–421

Li H, Aue A, Paul D (2020) High-dimensional general linear hypothesis tests via non-linear spectral shrinkage. Bernoulli 26(4):2541–2571

Nishiyama T, Hyodo M, Seo T, Pavlenko T (2013) Testing linear hypotheses of mean vectors for high-dimension data with unequal covariance matrices. J Stat Plann Inference 143(11):1898–1911

Paindaveine D, Verdebout T (2016) On high-dimensional sign tests. Bernoulli 22(3):1745–1769

Park J, Ayyala DN (2013) A test for the mean vector in large dimension and small samples. J Stat Plann Inference 143(5):929–943

Pauly M, Ellenberger D, Brunner E (2015) Analysis of high-dimensional one group repeated measures designs. Statistics 49(6):1243–1261

Peng L, Qi Y, Wang F (2014) Test for a mean vector with fixed or divergent dimension. Stat Sci 29(1):113–127

Satterthwaite FE (1946) An approximate distribution of estimates of variance components. Biometrics Bull 2(6):110–114

Scheffé H (1943) On solutions of the Behrens-Fisher problem, based on the t-distribution. Ann Math Stat 14(1):35–44

Schott JR (2007) Some high-dimensional tests for a one-way MANOVA. J Multivar Anal 98(9):1825–1839

Shen Y, Lin Z (2015) An adaptive test for the mean vector in large-p-small-n problems. Comput Stat Data Anal 89:25–38

Shen Y, Lin Z, Zhu J (2011) Shrinkage-based regularization tests for high-dimensional data with application to gene set analysis. Comput Stat Data Anal 55(7):2221–2233

Silva IR, Zhuang Y, da Silva Junior JCA (2021) Kronecker delta method for testing independence between two vectors in high-dimension. Stat Pap (In Press)

Sottoriva A, Spiteri I, Piccirillo SGM, Touloumis A, Collins VP, Marioni JC, Curtis C, Watts C, Tavaré S (2013) Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proc Natl Acad Sci 110(10):4009–4014

Srivastava MS, Du M (2008) A test for the mean vector with fewer observations than the dimension. J Multivar Anal 99(3):386–402

Srivastava MS, Kubokawa T (2013) Tests for multivariate analysis of variance in high dimension under non-normality. J Multivar Anal 115:204–216

Srivastava MS, Yanagihara H (2010) Testing the equality of several covariance matrices with fewer observations than the dimension. J Multivar Anal 101(6):1319–1329

Touloumis A, Tavaré S, Marioni JC (2015) Testing the mean matrix in high-dimensional transposable data. Biometrics 71(1):157–166

Wang R, Xu X (2019) A feasible high dimensional randomization test for the mean vector. J Stat Plann Inference 199:160–178

Wang L, Peng B, Li R (2015) A high-dimensional nonparametric multivariate test for mean vector. J Am Stat Assoc 110(512):1658–1669

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

Yamada T, Himeno T (2015) Testing homogeneity of mean vectors under heteroscedasticity in high-dimension. J Multivar Anal 139:7–27

Zhang JT (2005) Approximate and asymptotic distributions of chi-squared-type mixtures with applications. J Am Stat Assoc 100(469):273–285

Zhang JT (2013) Analysis of variance for functional data. CRC Press, Boca Raton

Zhang JT, Xu J (2009) On the k-sample Behrens-Fisher problem for high-dimensional data. Sci China Ser A 52(6):1285–1304

Zhang JT, Guo J, Zhou B (2017) Linear hypothesis testing in high-dimensional one-way MANOVA. J Multivar Anal 155:200–216

Zhang JT, Guo J, Zhou B, Cheng MY (2020) A simple two-sample test in high dimensions based on L^2-norm. J Am Stat Assoc 115(530):1011–1027

Zhang T, Wang Z, Wan Y (2021) Functional test for high-dimensional covariance matrix, with application to mitochondrial calcium concentration. Stat Pap 62:1213–1230

Zhao J (2017) A new test for the mean vector in large dimension and small samples. Commun Stat-Simul Comput 46(8):6115–6128

Zhou B, Guo J, Chen J, Zhang JT (2019) An adaptive spatial-sign-based test for mean vectors of elliptically distributed high-dimensional data. Stat Interface 12:93–106