Journal of Multivariate Analysis 196 (2023) 105183

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Finite sample t-tests for high-dimensional means

Jun Li

Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA

ARTICLE INFO

Article history:
Received 22 April 2022
Received in revised form 22 March 2023
Accepted 24 March 2023
Available online 28 March 2023

AMS 2020 subject classifications:
primary 62H15
secondary 62G10

Keywords:
High-dimensional data
Nonparametric methods
Robust procedures

ABSTRACT

When sample sizes are small, it becomes challenging for an asymptotic test requiring diverging sample sizes to maintain an accurate Type I error rate. In this paper, we consider one-sample, two-sample and ANOVA tests for mean vectors when data are high-dimensional but sample sizes are very small. We establish asymptotic t-distributions of the proposed U-statistics, which only require data dimensionality to diverge but sample sizes to be fixed and no less than 3. The proposed tests maintain accurate Type I error rates for a wide range of sample sizes and data dimensionality. Moreover, the tests are nonparametric and can be applied to data which are normally distributed or heavy-tailed. Simulation studies confirm the theoretical results for the tests. We also apply the proposed tests to an fMRI dataset to demonstrate the practical implementation of the methods.

1. Introduction

Testing for population means is a classical problem in statistics, which has a wide range of applications in clinical trails, case-control and financial studies. In the univariate case, the Student's t-test can be applied to test whether a population mean equals a claimed value when the sample mean follows a normal distribution, sample variance follows a χ^2 distribution and the sample mean and sample variance are independent. In the traditional multivariate setting, Hotelling's T^2 test [9] can be applied to test a population mean vector when dimension p is fixed and dimension p and sample size n satisfy the relation $p \leq n-1$. The Student's t-test and Hotelling's T^2 test are exact tests for finite sample sizes if data are normally distributed. If data are non-normally distributed, they are asymptotic tests as relatively large sample sizes are required to approximate sample means by a normal distribution.

With the development of high-throughput technologies, high-dimensional data characterized by the "large p and small n" situation have been widely observed in functional magnetic resonance imaging (fMRI), microarray, next-generation sequencing (RNA-Seq), and genome-wide association (GWA) studies. When $p > n-1$, the Hotelling's T^2 test becomes infeasible due to singularity of the sample covariance matrix. Even when p is close to $n-1$, the Hotelling's T^2 test loses its power as revealed by [2]. Many approaches have been proposed to modify the Hotelling's T^2 test for high-dimensional data. Some were constructed to discard or stabilize the inverse of sample covariance matrix. Examples include [2,5–7,11,16,20]. Some were proposed to reduce the noise contributed by non-signal bearing components for sparse signal detection. Examples include the maximum type test proposed in [3] and the thresholding tests in [4,22]. Some were developed to utilize advantages of both maximum type and sum-of-squares type tests to achieve better power against both sparse and dense alternatives. Examples include [8,21]. Others were proposed to project the classical Hotelling's T^2 statistic to a low-dimensional space. Examples include [13,17,19]. Except the aforementioned methods, there are many other contributions on testing high-dimensional means. We refer the readers to [10] for a recent review.

E-mail address: jli49@kent.edu.

https://doi.org/10.1016/j.jmva.2023.105183

Most proposed tests for high-dimensional means are asymptotic procedures, in the sense that they require both dimensionality and sample sizes to diverge to infinity even though dimensionality can be much larger than sample sizes. In many biological and financial studies, high dimensional data with very small sample sizes often occur due to ethical and cost reasons. In hypothesis testing, a primary requirement for a proposed test is to maintain its accurate type I error rate. As demonstrated by the simulation studies in Section 4, when sample sizes are very small and the null hypothesis is true, tests established by requiring diverging sample sizes tend to reject the null hypothesis with a probability higher or lower than a preselected nominal significance level. They thus cannot control type I error rate accurately.

To propose a robust test which maintains type I error rate accurately especially for small sample sizes, we establish the asymptotic normality of a one-sample U-statistic standardized by its standard deviation only requiring data dimensionality to diverge to infinity but sample sizes to be fixed. In practice the standard deviation of the U-statistic is unknown and cannot be consistently estimated when sample size is small. By analogy with the univariate Student's t-statistic, we propose an estimator for the variance of the U-statistic, which is not needed to be consistent but shown to be asymptotically χ^2 distributed and independent to the U-statistic under the null hypothesis. The result enables us to establish the asymptotic t-distribution of the U-statistic standardized by the sample standard deviation. The test can be applied to high-dimensional data with any finite sample size no less than 3, and maintains accurate Type I error rate. Moreover, it is nonparametric and can be applied to normally distributed or heavy-tailed data. We further extend the test to the two-sample and ANOVA testing problems. It is worth mentioning that the current work is not the only one to propose a finite sample t-test for high-dimensional testing problems. In [18], the authors developed a modified distance correlation statistic for the problem of testing the independence of high-dimensional random vectors. As the dimension diverges, the test statistic was shown to converge to a t-distribution for any sample size greater than 3 and an approximate standard normal when the sample size is greater than 9.

The rest of the paper is organized as follows. Section 2 introduces the one-sample U-statistic and establishes its asymptotic t-distribution. Extension to the two-sample and ANOVA problems is provided in Section 3. Simulation and case studies are presented in Sections 4 and 5. Section 6 concludes the paper with discussion. Technical proofs of theorems are relegated to the Appendix.

2. One-sample test

Let $\{X_1, \ldots, X_n\}$ be independent and identically distributed p-dimensional random vectors with mean $\mu = \mathrm{E}X_i$ and covariance matrix $\Sigma = \mathrm{Var}\,X_i$. The hypotheses we are interested in are

$$H_0 : \mu = 0 \quad \text{versus} \quad H_1 : \mu \neq 0. \tag{1}$$

The hypotheses are general. If one wants to test whether $\mu = \mu_0$ with a non-zero vector μ_0, we can convert the problem into the above hypotheses after subtracting each X_i by μ_0.

To test the hypotheses, we consider the following U-statistic

$$U_n = \frac{2}{n(n-1)} \sum_{i<j}^{n} X_i^\top X_j. \tag{2}$$

A two-sample version of the U-statistic was considered in [6], where the authors established its asymptotic normality by requiring both dimensionality and sample size to diverge to infinity. In this paper we establish the asymptotic t-distribution of the standardized U_n via Theorems 1–3 which only require dimensionality to diverge. The technical details can be seen in the Appendix. We outline some main ideas of the proofs as follows. To start with, we notice that the hypotheses (1) are equivalent to

$$H_0^* : \mu^\top \mu = 0 \quad \text{versus} \quad H_1^* : \mu^\top \mu > 0,$$

which allows us to obtain an equivalent univariate sample $\{X_i^\top X_j, i < j\}_{i,j=1}^n$ of size $n(n-1)/2$ from the original sample $\{X_i, 1 \le i \le n\}$. Each univariate random variable $X_i^\top X_j$ has the expectation $\mu^\top \mu$, which is zero under H_0^* but positive under H_1^*. The U-statistic is therefore the sample mean of $\{X_i^\top X_j, i < j\}_{i,j=1}^n$, or an unbiased estimator of the population mean $\mu^\top \mu$. In Theorem 1, we derive an asymptotic normal distribution of U_n by applying the martingale central limit theorem where we only require dimensionality to diverge but sample size n to be finite. Most importantly, we show that the random variables $\{X_i^\top X_j, i < j\}_{i,j=1}^n$ are asymptotically mutually independent and normally distributed under the null hypothesis. In Theorem 2, we further establish an asymptotic χ^2 distribution for the scaled sample variance of the asymptotically mutually independent and normally distributed random sample $\{X_i^\top X_j, i < j\}_{i,j=1}^n$ under the null hypothesis. By analogy with the Student's t-statistic, we also show that the sample variance is independent of U_n which is the sample mean of the random sample $\{X_i^\top X_j, i < j\}_{i,j=1}^n$. Combining the results in Theorems 1 and 2, we finally establish an asymptotic t-distribution of U_n standardized by its sample standard deviation under the null hypothesis in Theorem 3.

To make the proposed test nonparametric, we follow the idea in [2,6] to model the sequence of p-dimensional random vectors $\{X_i, 1 \le i \le n\}$ by a linear high-dimensional time series

$$X_i = \mu + \Gamma Z_i, \quad i \in \{1, \ldots, n\}, \tag{3}$$

where μ is the p-dimensional population mean, Γ is a $p \times q$ matrix with $q \geq p$ satisfying $\Gamma\Gamma^\top = \Sigma$, and $Z_i = (z_{i1}, \ldots, z_{iq})^\top$ so that $\{z_{i\ell}\}_{\ell=1}^q$ are mutually independent and satisfy $\mathrm{E}\, z_{i\ell} = 0$, $\mathrm{Var}\, z_{i\ell} = 1$ and $\mathrm{E}\, z_{i\ell}^4 = 3 + \eta$ for some finite constant η.

As discussed in [6], the model (3) includes normally distributed $z_{i\ell}$ if $\eta = 0$ and other distributions such as heavy-tailed or skewed $z_{i\ell}$ if $\eta \neq 0$. It is worth mentioning that different from [6], we assume all the components of Z_i are mutually independent rather than pseudo-independent. As shown in the proof of Theorem 1, such an assumption allows us to apply the martingale central limit theorem to establish the asymptotic normal distribution of the U-statistic, when dimension p diverges to infinity but sample size n is finite. To implement the martingale central limit theorem, we assume the following condition for the covariance matrix Σ.

(C1). As $p \to \infty$, $\mathrm{tr}(\Sigma^4) = o\{\mathrm{tr}^2(\Sigma^2)\}$.

The condition is automatically satisfied if all the eigenvalues of Σ are bounded, but it also allows partial eigenvalues to be unbounded (see [6] for detailed discussion). We establish the asymptotic normality of the U-statistic as follows.

Theorem 1. *Assume the model (3) and the condition (C1). For any finite sample size $n \geq 2$,*

$$\frac{U_n - \mu^\top \mu}{\sigma_n} \xrightarrow{d} N(0,1), \qquad p \to \infty,$$

where

$$\sigma_n^2 = \frac{2}{n(n-1)} tr(\Sigma^2) + \frac{4}{n} \mu^\top \Sigma \mu.$$

Especially, under H_0 of (1),

$$\frac{U_n}{\sigma_{n,0}} \xrightarrow{d} N(0,1), \quad p \to \infty,$$

where

$$\sigma_{n,0}^2 = \frac{2}{n(n-1)} tr(\Sigma^2).$$

While dropping the requirement that $n \to \infty$ sheds some light on proposing a test workable for small sample sizes, it becomes technically challenging to estimate the unknown $\mathrm{tr}(\Sigma^2)$ in order to implement a testing procedure. If we allow the sample size n to diverge to infinity, $\mathrm{tr}(\Sigma^2)$ can be estimated consistently by an unbiased estimator in [6] or a U-statistic in [12]. Slutsky's theorem then shows that the asymptotic normality of U_n with the estimated $\sigma_{n,0}$ still holds under the null hypothesis. However, we consider the sample size to be fixed and a consistent estimator of $\mathrm{tr}(\Sigma^2)$ is thus not achievable.

It turns out that constructing a consistent estimator is not necessary. Note that U_n is the sample mean of the sample $\{X_i^\top X_j, i < j\}_{i,j=1}^n$. From the proof of Theorem 1, we show that the random variables $\{X_i^\top X_j, i < j\}_{i,j=1}^n$ are asymptotically mutually independent and each component has the variance $\mathrm{tr}(\Sigma^2)$ under the null hypothesis. Based on $\{X_i^\top X_j, i < j\}_{i,j=1}^n$, we therefore estimate the unknown variance $\mathrm{tr}(\Sigma^2)$ by the sample variance

$$\widehat{\mathrm{tr}(\Sigma^2)} = \frac{1}{n(n-1)/2 - 1} \sum_{i<j}^{n} \left\{ X_i^\top X_j - \frac{2}{n(n-1)} \sum_{i<j}^{n} X_i^\top X_j \right\}^2. \tag{4}$$

By analogy with the Student's t-statistic, the scaled $\widehat{\mathrm{tr}(\Sigma^2)}$ is χ^2 distributed and asymptotically independent of the sample mean U_n.

Theorem 2. *Let the degrees of freedom $k = n(n-1)/2 - 1$. Assume the model (3) and the condition (C1). For any finite sample size $n \geq 3$ and under H_0 of (1),*

$$\frac{k\, \widehat{tr(\Sigma^2)}}{tr(\Sigma^2)} \xrightarrow{d} \chi^2(k), \quad p \to \infty.$$

After replacing $\mathrm{tr}(\Sigma^2)$ by the estimator $\widehat{\mathrm{tr}(\Sigma^2)}$, we estimate $\sigma_{n,0}^2$ by

$$\hat{\sigma}_{n,0}^2 = \frac{2}{n(n-1)} \widehat{\mathrm{tr}(\Sigma^2)}.$$

We establish the asymptotic t-distribution of $U_n / \hat{\sigma}_{n,0}$ as follows.

Theorem 3. *Assume the same conditions in Theorem 2. For any finite sample size $n \geq 3$ and under H_0 of (1), as $p \to \infty$,*

$$\frac{U_n}{\hat{\sigma}_{n,0}} \xrightarrow{d} t, \qquad k = n(n-1)/2 - 1 \quad \text{degrees of freedom}.$$

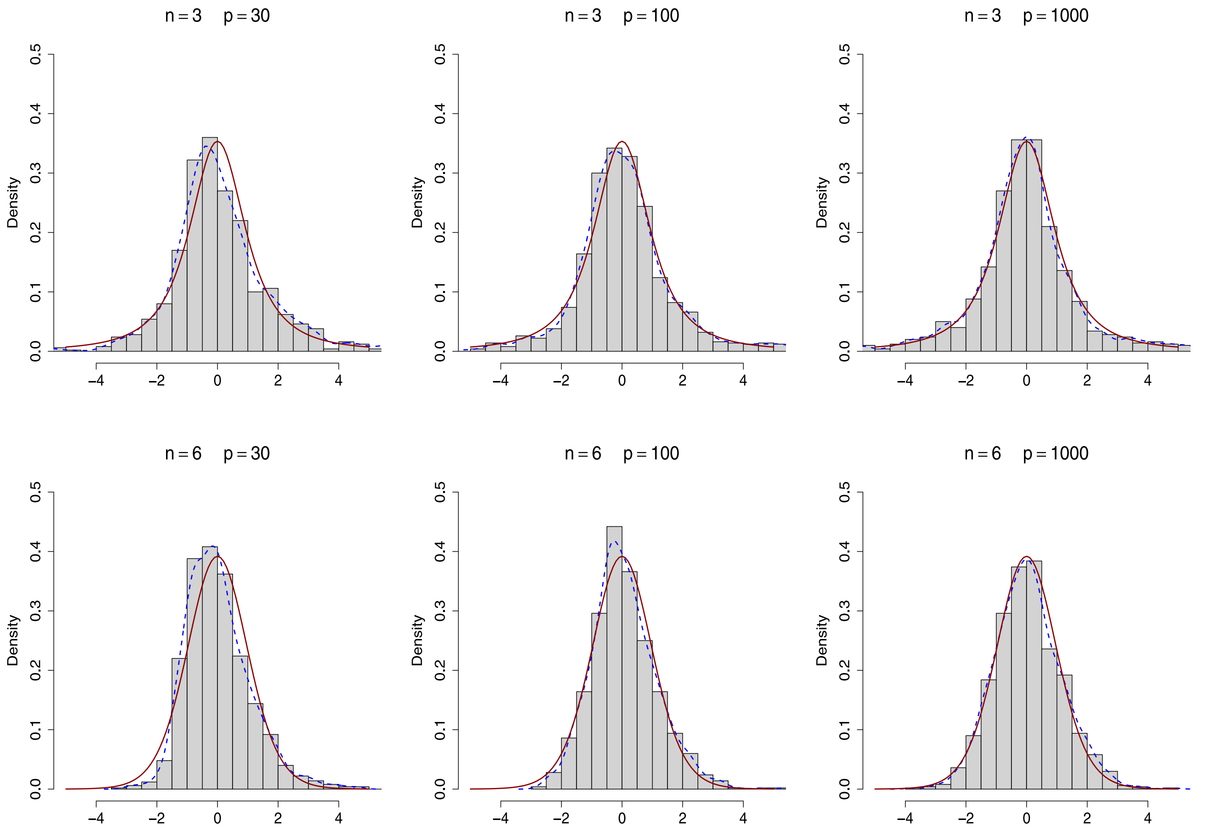

Fig. 1. Histogram of $U_n/\hat{\sigma}_{n,0}$ plus the fitted density curve (blue dashed line) versus the theoretical t-curve (red solid line). In the upper row, the sample size $n = 3$, and data dimension $p \in \{30, 100, 1000\}$. The theoretical t-curve has the degrees of freedom equal to 2; In the lower row, the sample size $n = 6$, and data dimension $p \in \{30, 100, 1000\}$. The theoretical t-curve has the degrees of freedom equal to 14.

To put the above result into a visual inspection, we simulate data from $N(0, \Sigma)$ where $\Sigma = (\sigma_{ij}) = (0.6^{|i-j|})$. Fig. 1 demonstrates the histogram of $U_n/\hat{\sigma}_{n,0}$ plus the corresponding fitted density curve (blue dashed line) versus the theoretical t-curve (red solid line), based on 1000 iterations with different sample sizes n and data dimensionality p. In the upper row, the sample size n is fixed to be 3 and the data dimension is increased from 30 to 1000. According to Theorem 3, the theoretical t-curve has 2 degrees of freedom. Since the asymptotic t-distribution is established as $p \to \infty$, the fitted density curve (blue dashed line) is skewed when $p = 30$ but closer to the t-curve (red solid line) when p becomes larger. Especially when $p = 1000$, the two density curves are nearly identical. Similar results can be observed in the lower row where the theoretical t-curve has 14 degrees of freedom. The visual inspection demonstrates the beneficial impact of the increasing dimension on the null distribution of the proposed test statistic.

Based on Theorem 3, the proposed test with a nominal α significance level rejects H_0 if $U_n/\hat{\sigma}_{n,0} \geq t_\alpha(k)$, where $t_\alpha(k)$ is the upper α quantile of t-distribution with $k = n(n-1)/2 - 1$ degrees of freedom. Moreover, the power function of the test when $\mu = \mu_0 \neq 0$ is

$$\begin{aligned}
B_1(\|\mu_0\|^2) &= \Pr\left\{\frac{U_n}{\hat{\sigma}_{n,0}} \geq t_\alpha(k) | \mu = \mu_0\right\} = 1 - \Pr\left\{\frac{U_n - \|\mu_0\|^2}{\sigma_n} < \frac{\hat{\sigma}_{n,0}}{\sigma_n} t_\alpha(k) - \frac{\|\mu_0\|^2}{\sigma_n} \quad | \mu = \mu_0\right\} \\
&= 1 - \Phi\left\{\frac{\hat{\sigma}_{n,0}}{\sigma_n} t_\alpha(k) - \frac{\sqrt{n(n-1)}\|\mu_0\|^2}{\sqrt{2\mathrm{tr}(\Sigma^2) + 4(n-1)\mu_0^\top \Sigma \mu_0}}\right\}, \quad p \to \infty,
\end{aligned}$$

where $\|\mu_0\|^2 = \mu_0^\top \mu_0$, σ_n is given in Theorem 1 and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal.

To see how the power $B_1(\|\mu_0\|^2)$ evolves with the sum-of-squares signal strength $\|\mu_0\|^2$, we first derive

$$\mathrm{E}\, \hat{\sigma}_{n,0}^2 = \frac{2}{n(n-1)} \mathrm{tr}(\Sigma^2) + \frac{4}{n(n+1)} \mu_0^\top \Sigma \mu_0.$$

Then from the Markov inequality, we obtain $\hat{\sigma}_{n,0}/\sigma_n = O_p(1)$ as $p \to \infty$. This indicates that the power $B_1(\|\mu_0\|^2)$ is largely determined by the signal-to-noise ratio

$$\mathrm{SNR}_1 = \frac{\sqrt{n(n-1)}\|\mu_0\|^2}{\sqrt{2\mathrm{tr}(\Sigma^2) + 4(n-1)\mu_0^\top \Sigma \mu_0}}.$$

A direct observation shows that $B_1(\|\mu_0\|^2) \to 1$ if $\mathrm{SNR}_1 \to \infty$ as $p \to \infty$. However, the test may lose its power when μ_0 is sparse. To appreciate this, we consider $\Sigma = I_p$ which is the $p \times p$ identity matrix, and let $p^{1-\beta}$ be the number of non-zero components in μ_0 and δ be the value of each non-zero component. Clearly β controls the sparsity of signals in the sense that a larger value of β leads to smaller numbers of nonzero signals. Based on the setup, we observe that when the sample size n and signal strength δ are fixed, $\mathrm{SNR}_1 = O(p^{1/2-\beta}) = o(1)$ if $\beta > 1/2$ representing the case of sparser signals. Even if δ^2 grows with p at the rate that $o(p^{\beta-1/2})$, we still obtain $\mathrm{SNR}_1 = O(\delta^2 p^{1/2-\beta}) = o(1)$. On the other hand, when the sample size n and signal strength δ are fixed, $\mathrm{SNR}_1 = O(p^{1/2-\beta}) \to \infty$ if $\beta < 1/2$ representing the case of denser signals.

3. Two-sample and ANOVA tests

Let $i \in \{1, 2\}$ and $\{X_{i1}, \ldots, X_{in_i}\}$ be two independent and identically distributed p-dimensional random samples with mean μ_i and covariance matrix Σ_i. The two-sample testing problem considers the hypotheses

$$H_0 : \mu_1 = \mu_2 \quad \text{versus} \quad H_1 : \mu_1 \neq \mu_2. \tag{5}$$

If the two samples have the same sample size n, we can directly extend the one-sample statistic (2) to the two-sample case by replacing X_i with $X_{1i} - X_{2i}$, $i \in \{1, \ldots, n\}$. However, two sample sizes are different in many cases. Without loss of generality, we assume $n_1 \leq n_2$ and consider

$$Y_i = X_{1i} - \sqrt{\frac{n_1}{n_2}} X_{2i} + \frac{1}{\sqrt{n_1 n_2}} \sum_{j=1}^{n_1} X_{2j} - \frac{1}{n_2} \sum_{j=1}^{n_2} X_{2j}, \quad i \in \{1, \ldots, n_1\}. \tag{6}$$

This procedure for the difference of two samples was first suggested by [15] in the univariate case to construct the confidence intervals by using the t-distribution. It was extended by [1] in the multivariate case to obtain the generalized Hotelling's T^2 statistic. We adopt the same procedure to propose the two-sample statistic

$$V_{n_1 n_2} = \frac{2}{n_1(n_1-1)} \sum_{i<j}^{n_1} Y_i^\top Y_j, \tag{7}$$

which is similar to (2) but we replace X_i with Y_i.

Note that $\bar{Y} = n_1^{-1} \sum_{i=1}^{n_1} Y_i = n_1^{-1} \sum_{i=1}^{n_1} X_{1i} - n_2^{-1} \sum_{i=1}^{n_2} X_{2i}$. We can write

$$V_{n_1 n_2} = \frac{n_1}{n_1 - 1} \bar{Y}^\top \bar{Y} - \frac{1}{n_1(n_1-1)} \sum_{i=1}^{n_1} Y_i^\top Y_i,$$

where the first term on the right hand side uses the difference of the two sample means $n_1^{-1} \sum_{i=1}^{n_1} X_{1i} - n_2^{-1} \sum_{i=1}^{n_2} X_{2i}$, which is the most relevant to $\mu_1 - \mu_2$. We subtract the second term from the first term so that $\mathrm{E}\, V_{n_1 n_2} = (\mu_1 - \mu_2)^\top (\mu_1 - \mu_2)$.

Similar to (3), we model the two independent and identically distributed p-dimensional random samples by the linear high-dimensional time series

$$X_{ij} = \mu_i + \Gamma_i Z_{ij}, \qquad i \in \{1, 2\}, \quad j \in \{1, \ldots, n_i\}, \tag{8}$$

where Γ_i is a $p \times q_i$ matrix with $q_i \geq p$ satisfying $\Gamma_i \Gamma_i^\top = \Sigma_i$, and $Z_{ij} = (z_{ij1}, \ldots, z_{ijq_i})^\top$ so that $\{z_{ij\ell}\}_{\ell=1}^{q_i}$ are mutually independent and satisfy $\mathrm{E}\, z_{ij\ell} = 0$, $\mathrm{Var}\, z_{ij\ell} = 1$ and $\mathrm{E}\, z_{ij\ell}^4 = 3 + \eta$ for some finite constant η.

By analogy with (C1), we consider the following condition for the two covariance matrices Σ_1 and Σ_2.

(C2). As $p \to \infty$, $\mathrm{tr}(\Sigma_i \Sigma_j \Sigma_k \Sigma_\ell) = o[\mathrm{tr}^2\{(\Sigma_1 + \Sigma_2)^2\}]$, $i, j, k, \ell \in \{1, 2\}$.

Under H_0 of (5), the variance of V_{n_1, n_2} is

$$\sigma_{n_1 n_2, 0}^2 = \frac{2}{n_1(n_1-1)} \sigma_{Y^\top Y, 0}^2, \qquad \sigma_{Y^\top Y, 0}^2 = \mathrm{tr}(\Sigma_1^2) + \frac{n_1^2}{n_2^2} \mathrm{tr}(\Sigma_2^2) + \frac{2n_1}{n_2} \mathrm{tr}(\Sigma_1 \Sigma_2)$$

is the variance of $Y_i^\top Y_j$ for $i < j$. Similar to the proof of Theorem 1, $\{Y_i^\top Y_j, i < j\}_{i,j=1}^{n_1}$ can be shown to be a sequence of $n_1(n_1-1)/2$ independent random variables under H_0 of (5). We therefore estimate $\sigma_{Y^\top Y, 0}^2$ by

$$\hat{\sigma}_{Y^\top Y, 0}^2 = \frac{1}{n_1(n_1-1)/2 - 1} \sum_{i<j}^{n_1} \left\{ Y_i^\top Y_j - \frac{2}{n_1(n_1-1)} \sum_{i<j}^{n_1} Y_i^\top Y_j \right\}^2, \tag{9}$$

which is similar to (4) but we replace n and X_i by n_1 and Y_i, respectively.

As a result, the unbiased estimator of $\sigma^2_{n_1 n_2,0}$ is

$$\hat{\sigma}^2_{n_1 n_2,0} = \frac{2}{n_1(n_1-1)} \hat{\sigma}^2_{Y^\top Y,0}. \tag{10}$$

The following theorem establishes the asymptotic t-distribution of $V_{n_1 n_2}/\hat{\sigma}_{n_1 n_2,0}$, which is a direct extension of Theorem 3.

Theorem 4. *Assume the model (8) and the condition (C2). For any finite sample sizes $n_2 \geq n_1 \geq 3$ and under H_0 of (5), as $p \to \infty$,*

$$\frac{V_{n_1 n_2}}{\hat{\sigma}_{n_1 n_2,0}} \xrightarrow{d} t, \quad k_2 = n_1(n_1-1)/2 - 1 \quad \textit{degrees of freedom}.$$

Based on Theorem 4, the proposed test with a nominal α level of significance rejects H_0 if $V_{n_1 n_2}/\hat{\sigma}_{n_1 n_2,0} \geq t_\alpha(k_2)$, where $t_\alpha(k_2)$ is the upper α quantile of t-distribution with $k_2 = n_1(n_1-1)/2 - 1$ degrees of freedom. Moreover, as $p \to \infty$, the power of the two-sample test is

$$B_2(\|\mu_1 - \mu_2\|^2) = 1 - \Phi\left\{ \frac{\hat{\sigma}_{n_1 n_2,0}}{\sigma_{n_1 n_2}} t_\alpha(k_2) - \mathrm{SNR}_2 \right\},$$

where the signal-to-noise ratio

$$\mathrm{SNR}_2 = \frac{\sqrt{n_1(n_1-1)}\|\mu_1 - \mu_2\|^2}{\sqrt{2\mathrm{tr}\{(\Sigma_1 + \frac{n_1}{n_2}\Sigma_2)^2\} + 4(n_1-1)(\mu_1-\mu_2)^\top(\Sigma_1 + \frac{n_1}{n_2}\Sigma_2)(\mu_1-\mu_2)}}. \tag{11}$$

Similar to the one-sample test, the power of the two-sample test is largely determined by SNR_2, the analysis of which demonstrates that the proposed test is powerful in detecting dense and strong differences between μ_1 and μ_2, but encounters a power loss when differences between μ_1 and μ_2 are sparse and weak.

When there are more than two populations, the ANOVA problem is to test whether $\mu_1 = \cdots = \mu_m$, where μ_ℓ and Σ_ℓ are the population mean and covariance matrix of the ℓth population for $\ell \in \{1, \ldots, m\}$. It is equivalent to testing whether $\|\mu_2 - \mu_1\|^2 + \|\mu_3 - \mu_1\|^2 + \cdots + \|\mu_m - \mu_1\|^2 = 0$, where $\|\mu_\ell - \mu_1\|^2$ can be estimated by the two-sample statistic (7) based on the ℓth and first samples.

Let $\ell \in \{1, \ldots, m\}$ and $\{X_{\ell 1}, \ldots, X_{\ell n_\ell}\}$ be an independent and identically distributed random sample from the ℓth population. Without loss of generality, we assume the sample sizes $n_1 \leq n_2 \leq \cdots \leq n_m$. The proposed multiple-sample statistic is

$$W_m = \frac{2}{n_1(n_1-1)} \sum_{i<j}^{n_1} \sum_{\ell=2}^{m} Y_{\ell i}^\top Y_{\ell j},$$

where $Y_{\ell i}$ is the difference of ℓth and first samples given by (6) with X_{2i} replaced by $X_{\ell i}$. Similar to the two-sample statistic (7), W_m can be shown to be unbiased to $\|\mu_2 - \mu_1\|^2 + \|\mu_3 - \mu_1\|^2 + \cdots + \|\mu_m - \mu_1\|^2$. Moreover, similar to (9) and (10), the variance of W_m under the null can be estimated by

$$\hat{\sigma}^2_{m,0} = \frac{2}{n_1^2(n_1-1)^2/2 - n_1(n_1-1)} \sum_{i<j}^{n_1} \left\{ \sum_{\ell=2}^{m} Y_{\ell i}^\top Y_{\ell j} - \frac{2}{n_1(n_1-1)} \sum_{i<j}^{n_1} \sum_{\ell=2}^{m} Y_{\ell i}^\top Y_{\ell j} \right\}^2.$$

To establish the asymptotic t-distribution of $W_m/\hat{\sigma}_{m,0}$, we can directly extend the model (8) to

$$X_{\ell j} = \mu_\ell + \Gamma_\ell Z_{\ell j}, \qquad \ell \in \{1, \ldots, m\}, \quad j \in \{1, \ldots, n_\ell\}, \tag{12}$$

where Γ_ℓ is a $p \times q_\ell$ matrix with $q_\ell \geq p$ satisfying $\Gamma_\ell \Gamma_\ell^\top = \Sigma_\ell$, and $Z_{\ell j} = (z_{\ell j1}, \ldots, z_{\ell j q_\ell})^\top$ so that $\{z_{\ell jh}\}_{h=1}^{q_\ell}$ are mutually independent and satisfy $\mathrm{E}\, z_{\ell jh} = 0$, $\mathrm{Var}\, z_{\ell jh} = 1$ and $\mathrm{E}\, z^4_{\ell jh} = 3 + \eta$ for some finite constant η.

We also directly modify (C2) to the following condition.

(C3). As $p \to \infty$, $\mathrm{tr}(\Sigma_i \Sigma_j \Sigma_k \Sigma_\ell) = o[\mathrm{tr}^2\{(\sum_{h=1}^m \Sigma_h)^2\}]$, $i, j, k, \ell \in \{1, \ldots, m\}$.

The asymptotic t-distribution of $W_m/\hat{\sigma}_{m,0}$ can be established as follows.

Theorem 5. *Assume the model (12) and the condition (C3). For any finite sample sizes $n_m \geq \cdots \geq n_2 \geq n_1 \geq 3$ and under the null hypothesis $\mu_1 = \cdots = \mu_m$, as $p \to \infty$,*

$$\frac{W_m}{\hat{\sigma}_{m,0}} \xrightarrow{d} t, \quad k_2 = n_1(n_1-1)/2 - 1 \quad \textit{degrees of freedom}.$$

Based on Theorem 5, the proposed test with a nominal α level of significance rejects the null hypothesis $\mu_1 = \cdots = \mu_m$ if $W_m/\hat{\sigma}_{m,0} \geq t_\alpha(k_2)$, where $t_\alpha(k_2)$ is the upper α quantile of t-distribution with $k_2 = n_1(n_1-1)/2 - 1$ degrees of freedom.

Table 1
Empirical sizes of Chen and Qin's test (CQ), Bai and Sarandasa's test (BS), Srivastava and Du's test (SD), and the proposed test (New), based on 1000 replications with normally distributed and t-distributed Z_i in (3) under models (a) and (b) for the covariance matrix Σ, respectively. The nominal significance level $\alpha = 0.05$. The sample size $n \in \{4, 6, 15, 30\}$ and the dimension $p \in \{200, 400, 1000\}$.

	$p = 200$				$p = 400$				$p = 1000$			
	$n = 4$	6	15	30	$n = 4$	6	15	30	$n = 4$	6	15	30
Normally distributed Z_i												
Model (a)												
CQ	0.129	0.088	0.055	0.058	0.122	0.086	0.062	0.061	0.134	0.078	0.062	0.052
BS	0.088	0.080	0.056	0.056	0.090	0.084	0.061	0.064	0.095	0.059	0.059	0.053
SD	0.217	0.149	0.055	0.047	0.205	0.118	0.046	0.038	0.213	0.072	0.018	0.024
New	0.056	0.062	0.051	0.055	0.059	0.059	0.058	0.058	0.058	0.052	0.059	0.050
Model (b)												
CQ	0.138	0.079	0.044	0.057	0.112	0.080	0.067	0.059	0.119	0.070	0.050	0.057
BS	0.099	0.069	0.047	0.057	0.064	0.079	0.062	0.060	0.090	0.061	0.049	0.058
SD	0.218	0.157	0.043	0.035	0.192	0.114	0.042	0.036	0.227	0.101	0.014	0.033
New	0.067	0.049	0.042	0.057	0.040	0.063	0.064	0.059	0.046	0.053	0.048	0.055
t-distributed Z_i												
Model (a)												
CQ	0.145	0.076	0.066	0.062	0.130	0.074	0.054	0.060	0.116	0.089	0.054	0.056
BS	0.072	0.043	0.047	0.042	0.053	0.034	0.032	0.044	0.040	0.038	0.028	0.037
SD	0.193	0.089	0.033	0.032	0.187	0.070	0.018	0.031	0.175	0.047	0.007	0.014
New	0.064	0.053	0.061	0.059	0.048	0.049	0.051	0.058	0.049	0.065	0.051	0.054
Model (b)												
CQ	0.143	0.085	0.067	0.052	0.141	0.069	0.076	0.044	0.128	0.064	0.050	0.069
BS	0.051	0.039	0.038	0.039	0.062	0.033	0.048	0.025	0.049	0.023	0.021	0.049
SD	0.157	0.106	0.038	0.028	0.169	0.080	0.030	0.021	0.148	0.034	0.006	0.017
New	0.056	0.058	0.061	0.052	0.068	0.049	0.068	0.043	0.059	0.046	0.048	0.069

4. Simulation studies

4.1. One-sample test

We compare the proposed one-sample t-test with the one-sample version CQ test in [6], the BS test in [2], and the SD test in [16]. To generate random samples, we considered two types of innovations in (3): the Gaussian $Z_i \sim N(0, I_p)$ and the standardized t-distribution with 4 degrees of freedom for each component of Z_i, where the latter has heavier tails than the former used to demonstrate nonparametric performance of the proposed test. Under H_0, we simply assumed $\mu = 0$. Under H_1, we considered μ to have $[p^{1-\beta}]$ non-zero entires which were randomly selected from $\{1, \ldots, p\}$. Here $[a]$ denotes the integer part of a. The value of each non-zero entry was r. From the simulation setup, the two parameters $\beta > 0$ and $r > 0$ were chosen to control the sparsity and strength of signals, respectively. We also considered the following two structures for the covariance Σ, where model (a) specifies a bandable structure of Σ and Model (b) leads to a sparse Σ.

(a) AR(1) model: $\sigma_{j_1 j_2} = 0.6^{|j_1 - j_2|}$ for $1 \le j_1, j_2 \le p$.
(b) Random sparse matrix model: first generate a $p \times p$ matrix Γ each row of which has only four non-zero element that is randomly chosen from $\{1, \ldots, p\}$ with magnitude generated from Unif(1, 2) multiplied by a random sign. Then $\Sigma = \Gamma\Gamma^\top + I_p$ where I_p is the $p \times p$ identity matrix.

All the simulation results were based on 1000 replications with the nominal significance level $\alpha = 0.05$.

Table 1 displays the empirical sizes of the four tests with normally distributed and t-distributed Z_i in (3) under models (a) and (b) for the covariance matrix Σ, respectively. The sample size and dimension were chosen to be $n \in \{4, 6, 15, 30\}$ and $p \in \{200, 400, 1000\}$. While the CQ, BS and SD tests were able to maintain the empirical sizes close to the nominal significance level $\alpha = 0.05$ when sample sizes were relatively large, they encountered size distortion especially when sample size was very small ($n = 4$). Unlike the competitors, the proposed test always had the empirical sizes close to the nominal significance level for both normally distributed and t-distributed Z_i. The results confirm Theorem 3 that the proposed testing procedure was established without requiring the diverging sample size and without assuming Gaussian distribution of data.

Due to the size distortion of the CQ, BS and SD tests when sample sizes are small, we compared the power performance of the four tests with relatively large sample sizes $n \in \{15, 30\}$. Table 2 demonstrates the empirical powers of the four tests with respect to different signal strength r when the sparsity of signal $\beta = 0.4$. As we can see, the powers of the four tests were increased as the signal strength r increased. The proposed test performed similarly to the other three tests. This is not surprising as the four tests were all proposed based on similar sum-of-squares type statistics. The powers of the proposed test were less than the CQ when $n = 15$ and only slightly less than the CQ when $n = 30$. The power loss can be explained by the higher value of $\hat{\sigma}_{n,0}$ in the proposed test statistic $U_n/\hat{\sigma}_{n,0}$ than the corresponding $\hat{\sigma}^*_{n,0}$ in the

Table 2
Empirical powers of Chen and Qin's test (CQ), Bai and Sarandasa's test (BS), Srivastava and Du's test (SD), and the proposed test (New), based on 1000 replications with normally distributed and t-distributed Z_i in (3) under models (a) and (b) for the covariance matrix Σ, respectively. The nominal significance level $\alpha = 0.05$. The sample size $n \in \{15, 30\}$ and the dimension $p = 400$. The signal strength $r \in \{0.1, 0.2, 0.3, 0.4\}$ and the signal sparsity $\beta = 0.4$.

	Model (a)				Model (b)			
	$r = 0.1$	0.2	0.3	0.4	$r = 0.1$	0.2	0.3	0.4
Normally distributed Z_i with $p = 400$, $\beta = 0.4$, $n = 15$								
CQ	0.067	0.141	0.307	0.634	0.084	0.170	0.347	0.695
BS	0.070	0.142	0.307	0.627	0.083	0.174	0.351	0.688
SD	0.038	0.098	0.215	0.470	0.054	0.124	0.255	0.569
New	0.067	0.137	0.295	0.627	0.077	0.168	0.343	0.684
Normally distributed Z_i with $p = 400$, $\beta = 0.4$, $n = 30$								
CQ	0.099	0.292	0.709	0.990	0.095	0.310	0.796	0.999
BS	0.098	0.290	0.711	0.990	0.094	0.311	0.795	0.999
SD	0.062	0.211	0.596	0.972	0.058	0.237	0.726	0.991
New	0.097	0.288	0.705	0.991	0.094	0.307	0.794	0.999
t distributed Z_i with $p = 400$, $\beta = 0.4$, $n = 15$								
CQ	0.082	0.158	0.328	0.630	0.063	0.158	0.377	0.684
BS	0.041	0.098	0.239	0.510	0.043	0.104	0.272	0.571
SD	0.025	0.051	0.155	0.370	0.026	0.076	0.189	0.495
New	0.074	0.152	0.318	0.624	0.065	0.154	0.366	0.684
t distributed Z_i with $p = 400$, $\beta = 0.4$, $n = 30$								
CQ	0.097	0.290	0.728	0.985	0.093	0.296	0.792	0.996
BS	0.068	0.241	0.650	0.953	0.060	0.233	0.728	0.978
SD	0.039	0.170	0.535	0.906	0.033	0.175	0.672	0.986
New	0.096	0.287	0.724	0.985	0.090	0.290	0.792	0.996

CQ test statistic $U_n/\hat{\sigma}^*_{n,0}$ under the alternative, although both test statistics employ the same U_n. From [6], the variance estimator $\hat{\sigma}^{*2}_{n,0}$ is ratio consistent to $2n^{-1}(n-1)^{-1}\mathrm{tr}(\Sigma^2)$ under both the null and alternative hypotheses. On the other hand, according to Theorem 2, the proposed variance estimator $\hat{\sigma}^2_{n,0}$ is unbiased to $2n^{-1}(n-1)^{-1}\mathrm{tr}(\Sigma^2)$ under the null hypothesis but the expectation becomes $2n^{-1}(n-1)^{-1}\mathrm{tr}(\Sigma^2) + 4n^{-1}(n+1)^{-1}\mu_0^\top \Sigma \mu_0$ under the alternative hypothesis. The extra term $4n^{-1}(n+1)^{-1}\mu_0^\top \Sigma \mu_0$ reduces the power of the proposed test but it diminishes as sample size n increases.

To further investigate how the power of the proposed test varies with different sample sizes and data dimensionalities, we chose a range of sample sizes $n \in \{3, 6, 12\}$ and data dimensionalities $p \in \{30, 100, 400, 1000\}$ with normally distributed Z_i under model (a). For each of the sample sizes, the empirical powers of the proposed test were obtained with respect to a range of signal strength r from 0.2 to 1 and a range of dimensionalities p from 30 to 1000. As illustrated in the first row of Fig. 2, when signals were denser ($\beta = 0.4$), the powers of the proposed test increased as the signal strength r increased for each sample size n, and as the sample size increased for each signal strength r. Moreover, for each fixed sample size n, the powers increased as data dimensionality p increased especially with relatively large signal strength r. For example, when $\beta = 0.4$, $n = 6$ and $r = 0.8$, the powers were 0.674, 0.755, 0.848 and 0.928 for $p \in \{30, 100, 400, 1000\}$, respectively. When signals were sparser ($\beta = 0.6$), we still observed that the powers increased as the signal strength r increased for each sample size n, and as the sample size increased for each signal strength r. However, for each fixed sample size n, the powers decreased as data dimensionality p increased. For example, when $\beta = 0.6$, $n = 6$ and $r = 0.8$, the powers were 0.324, 0.255, 0.234 and 0.220 for $p \in \{30, 100, 400, 1000\}$, respectively. The simulation studies confirm the theoretical results in the end of Section 2 that for each fixed sample size, the proposed test can enhance its power as dimensionality p increases if the signal sparsity $\beta < 0.5$, but loses its power if $\beta > 0.5$.

4.2. Two-sample and ANOVA tests

Under H_0 of the two-sample testing problem, we compared the size performance of the proposed test with the two-sample version CQ test in [6], the maximum type CLX test in [3], the multi-level thresholding CLZ test in [4] and a permutation test based on the proposed test statistic. The permutation test is obtained by first calculating the observed two-sample test statistic $V_{n_1n_2}/\hat{\sigma}_{n_1n_2,0}$ from the two samples. Next, the observations of the two samples are pooled. For each permutation of the pooled sample, the test statistic $V_{n_1n_2}/\hat{\sigma}_{n_1n_2,0}$ is calculated from the two samples of first n_1 observations and remaining n_2 observations. At last, the p-value is calculated as the proportion of the test statistics from the permuted two samples greater than the observed test statistic from the original two samples. To reduce the computational cost, we considered 1000 random permutations. The random samples were generated from two types of innovations in (8). The Gaussian $Z_{1j} \sim N(0, I_p)$ and $Z_{2j} \sim N(0, I_p)$, and the standardized t-distribution with 4 degrees of freedom for each component of Z_{1j} and Z_{2j}. For simplicity, we assigned $\mu_1 = \mu_2 = 0$, and considered $\Sigma_1 = \Sigma_2$ modeled by the AR(1) structure (a) in Section 4.1.

Table 3 displays the empirical sizes of the five tests. The dimensions of random vector were $p \in \{200, 400, 1000\}$. The sample size $n_2 = 30$, and the sample size $n_1 \in \{4, 6, 15, 30\}$. The sample sizes are unbalanced when $n_1 \in \{4, 6, 15\}$.

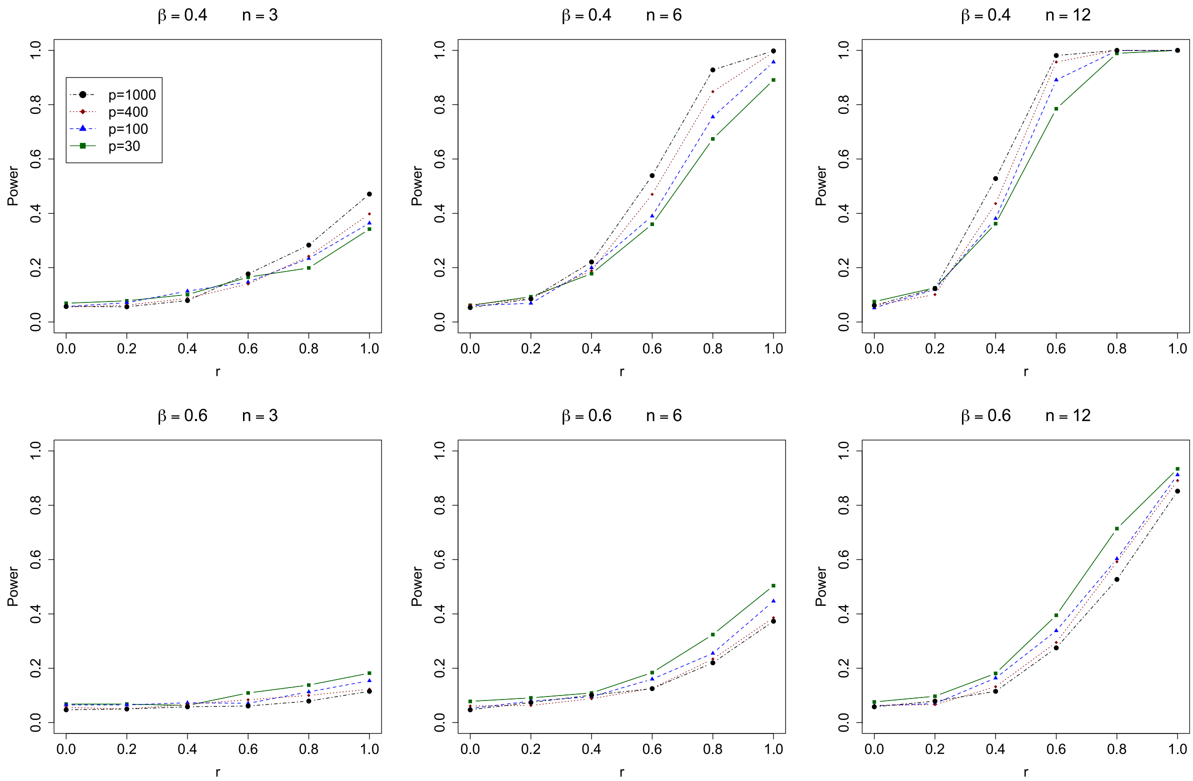

Fig. 2. Empirical powers of the proposed test with respect to signal strength r and signal sparsity β by choosing different sample sizes and dimensionalities. In the upper row, the signal sparsity $\beta = 0.4$ (denser signals). In the lower row, the signal sparsity $\beta = 0.6$ (sparser signals).

Table 3
Empirical sizes of Chen and Qin's test (CQ), Cai, Liu and Xia's test (CLX), Chen, Li and Zhong's test (CLZ), the proposed test (New), and the proposed permutation test (New-perm), based on 1000 replications with normally distributed and t-distributed Z_{1j} and Z_{2j} in (8) under model (a) for the covariance matrices $\Sigma_1 = \Sigma_2$. The nominal significance level $\alpha = 0.05$. The sample size $n_2 = 30$, and the sample size $n_1 \in \{4, 6, 15, 30\}$. The dimension $p \in \{200, 400, 1000\}$.

$n_2 = 30$	$p = 200$				$p = 400$				$p = 1000$			
	$n_1 = 4$	6	15	30	$n_1 = 4$	6	15	30	$n_1 = 4$	6	15	30
					Normally distributed Z_{1j} and Z_{2j}							
CQ	0.108	0.084	0.064	0.068	0.085	0.069	0.060	0.060	0.106	0.067	0.059	0.054
CLX	0.971	0.599	0.046	0.014	0.994	0.763	0.062	0.024	1	0.934	0.076	0.003
CLZ	0.998	0.869	0.113	0.036	1	0.980	0.168	0.025	1	1	0.236	0.015
New	0.054	0.062	0.055	0.066	0.058	0.053	0.059	0.059	0.057	0.060	0.055	0.061
New-perm	0.044	0.041	0.048	0.053	0.048	0.051	0.052	0.050	0.057	0.052	0.054	0.058
					t-distributed Z_{1j} and Z_{2j}							
CQ	0.095	0.054	0.052	0.060	0.096	0.060	0.063	0.062	0.089	0.066	0.054	0.051
CLX	0.968	0.582	0.060	0.006	0.996	0.745	0.058	0.008	1	0.929	0.081	0.001
CLZ	0.999	0.849	0.137	0.034	1	0.973	0.166	0.019	1	1	0.266	0.011
New	0.055	0.050	0.054	0.061	0.056	0.048	0.065	0.061	0.049	0.059	0.063	0.050
New-perm	0.047	0.053	0.046	0.056	0.057	0.051	0.044	0.053	0.055	0.049	0.056	0.055

From Theorem 4, the asymptotic t-distribution of the proposed test statistic holds under the null hypothesis as long as the smaller sample size is no less than 3. For all the cases including the unbalanced sample sizes, the proposed test and permutation test maintained the empirical sizes close to the nominal significance level $\alpha = 0.05$. However, the CQ, CLX and CLZ tests had inflated sizes when the sample size n_1 was extremely small at 4.

Due to the size distortion of the CLX and CLZ tests, it is not reasonable to compare their powers with the proposed test when sample sizes are small. We therefore compared the proposed test with the CQ and the permutation test. Under H_1, $\mu_1 = 0$ but μ_2 had $[p^{1-\beta}]$ non-zero entires which were randomly selected from $\{1, \ldots, p\}$. The value of each non-zero entry was r. Similarly to the setup under H_0, the random samples were generated from two types of innovations. The Gaussian $Z_{1j} \sim N(0, I_p)$ and $Z_{2j} \sim N(0, I_p)$, and the standardized t-distribution with 4 degrees of freedom for each

Table 4
Empirical powers of Chen and Qin's test (CQ), the proposed test (New) and the proposed permutation test (New-perm), based on 1000 replications with normally distributed and t-distributed Z_{1j} and Z_{2j} in (8) under model (a) for the covariance matrices $\Sigma_1 = \Sigma_2$. The nominal significance level $\alpha = 0.05$. The dimension $p = 400$, and the sample sizes $(n_1, n_2) = (6, 15), (15, 15), (6, 30), (15, 30), (6, 40)$ and $(15, 40)$. The signal strength $r \in \{0.4, 0.6, 0.8, 1\}$.

	$r = 0.4$	0.6	0.8	1	$r = 0.4$	0.6	0.8	1
				Normally distributed Z_{1j} and Z_{2j}				
		$(n_1 = 6, n_2 = 15)$				$(n_1 = 15, n_2 = 15)$		
CQ	0.170	0.367	0.713	0.951	0.245	0.695	0.982	1.000
New	0.137	0.319	0.648	0.913	0.247	0.681	0.976	1.000
New-perm	0.115	0.294	0.625	0.900	0.217	0.655	0.974	1.000
		$(n_1 = 6, n_2 = 30)$				$(n_1 = 15, n_2 = 30)$		
CQ	0.192	0.450	0.795	0.979	0.398	0.902	0.999	1.000
New	0.169	0.392	0.724	0.966	0.379	0.889	0.999	1.000
New-perm	0.153	0.363	0.702	0.952	0.354	0.870	0.999	1.000
		$(n_1 = 6, n_2 = 40)$				$(n_1 = 15, n_2 = 40)$		
CQ	0.208	0.484	0.817	0.981	0.424	0.934	1.000	1.000
New	0.175	0.409	0.763	0.971	0.397	0.921	1.000	1.000
New-perm	0.155	0.381	0.731	0.968	0.378	0.913	1.000	1.000
				t-distributed Z_{1j} and Z_{2j}				
		$(n_1 = 6, n_2 = 15)$				$(n_1 = 15, n_2 = 15)$		
CQ	0.161	0.382	0.718	0.950	0.281	0.701	0.985	0.999
New	0.152	0.335	0.624	0.906	0.272	0.696	0.984	0.999
New-perm	0.124	0.300	0.596	0.893	0.247	0.675	0.977	0.999
		$(n_1 = 6, n_2 = 30)$				$(n_1 = 15, n_2 = 30)$		
CQ	0.196	0.410	0.795	0.972	0.383	0.890	0.999	1.000
New	0.163	0.363	0.736	0.958	0.375	0.875	0.999	1.000
New-perm	0.147	0.332	0.709	0.944	0.344	0.854	0.998	1.000
		$(n_1 = 6, n_2 = 40)$				$(n_1 = 15, n_2 = 40)$		
CQ	0.178	0.479	0.822	0.979	0.433	0.928	1.000	1.000
New	0.152	0.417	0.781	0.963	0.418	0.916	1.000	1.000
New-perm	0.131	0.379	0.756	0.956	0.384	0.908	1.000	1.000

component of Z_{1j} and Z_{2j}. Again, we considered $\Sigma_1 = \Sigma_2$ modeled by the AR(1) structure (a) in Section 4.1. All the simulation results were based on 1000 replications with nominal significance level $\alpha = 0.05$.

As we can see in Table 4, the powers of the CQ test, the proposed test and permutation test increased as the signal strength r increased for each combination of n_1 and n_2. From Theorem 4, the power of the proposed two-sample test is largely determined by the signal-to-noise ratio (11), which can also be written as

$$\mathrm{SNR}_2 = \frac{\|\mu_1 - \mu_2\|^2}{\sqrt{\frac{2n_1}{n_1-1}\mathrm{tr}\{(\frac{\Sigma_1}{n_1} + \frac{\Sigma_2}{n_2})^2\} + 4(\mu_1 - \mu_2)^\top(\frac{\Sigma_1}{n_1} + \frac{\Sigma_2}{n_2})(\mu_1 - \mu_2)}}.$$

The expression shows that the power of the proposed two-sample test increases as either n_1 or n_2 increases even though the two sample sizes could be unbalanced. This is confirmed by the results in Table 4 where n_1 was fixed and n_2 was increased from 15 to 40 (rows 1 to 3), and n_2 was fixed and n_1 was increased from 6 to 15 (columns 1 to 2) for both normally and t distributed Z_{1j} and Z_{2j}. However, the unbalanced sample sizes could affect the power of the proposed test as the degrees of freedom rely on the smaller sample size n_1 rather than the combined sample size $n_1 + n_2$. When n_1 is extremely small, say $n_1 = 6$, the degrees of freedom equal $n_1(n_1 - 1)/2 - 1 = 14$ and the corresponding quantile $t_{0.05}(14) = 1.761$. It is therefore harder for the proposed test to reject the null hypothesis than any other testing rule based on the combined sample size or the asymptotic normal approximation. This is why the power of the proposed test is less than the CQ test when $n_1 = 6$. Similar to the one-sample case, another reason for the power loss of the proposed two-sample test is that it involves the higher value of $\hat{\sigma}_{n_1n_2,0}$ in the proposed test statistic $V_{n_1n_2}/\hat{\sigma}_{n_1n_2,0}$ than the corresponding variance estimator in the CQ test statistic under the alternative. From [6], the variance estimator of the CQ statistic is ratio consistent to $2n_1^{-1}(n_1 - 1)^{-1}\mathrm{tr}(\Sigma_1^2) + 2n_2^{-1}(n_2 - 1)^{-1}\mathrm{tr}(\Sigma_2^2) + 4n_1^{-1}n_2^{-1}\mathrm{tr}(\Sigma_1\Sigma_2)$ under both the null and alternative hypotheses. However, the proposed variance estimator $\hat{\sigma}^2_{n_1n_2,0}$ is only unbiased under the null hypothesis but has an additional term $4n_1^{-1}n_2^{-1}(\mu_1 - \mu_2)^\top(\Sigma_1 + \Sigma_2)(\mu_1 - \mu_2)$ under the alternative hypothesis. The extra term reduces the power of the proposed test but it diminishes as either n_1 or n_2 increases.

It is worth mentioning that when the two high dimensional mean vectors differ only in sparse coordinates and the differences are faint, the CLX and CLZ tests were proposed to improve power performance of the CQ test. It is therefore not surprising that they perform better than the proposed test for sparse and faint signal detection. But we need to emphasize that such a superior performance relies on the requirement of relatively large sample sizes.

At last, we conducted simulation studies to demonstrate the empirical performance of the proposed ANOVA test. We considered testing the equality of three mean vectors. Three random samples were generated from the model (12) with two types of innovations: one is Gaussian $Z_{\ell j} \sim N(0, I_p)$ and another is the standardized t-distribution with 4 degrees of freedom for each component of $Z_{\ell j}$. For simplicity, we chose $\mu_1 = \mu_2 = \mu_3 = 0$ under H_0. Under H_1, $\mu_1 = \mu_2 = 0$

Table 5
Empirical sizes and powers of the proposed ANOVA test, based on 1000 replications of three random samples with normally distributed and t-distributed Z_{1j}, Z_{2j} and Z_{3j} in (12). The covariance matrices Σ_1 and Σ_2 were the model (a), and Σ_3 was the model (b). The nominal significance level $\alpha = 0.05$. Empirical sizes were obtained with the signal strength $r = 0$, and empirical powers were obtained with $r \in \{0.6, 0.8, 1.0\}$ for different combinations of (p, n_1, n_2, n_3).

(p, n_1, n_2, n_3)	Size	Power		
		$r = 0.6$	$r = 0.8$	$r = 1.0$
Normally distributed Z_{1j}, Z_{2j} and Z_{3j}				
(200,3,15,30)	0.046	0.089	0.121	0.176
(200,15,15,30)	0.049	0.453	0.842	0.989
(200,30,30,30)	0.066	0.864	1	1
(400,3,15,30)	0.049	0.104	0.108	0.187
(400,15,15,30)	0.066	0.479	0.882	0.998
(400,30,30,30)	0.062	0.900	0.999	1
(1000,3,15,30)	0.041	0.095	0.134	0.199
(1000,15,15,30)	0.058	0.556	0.941	1
(1000,30,30,30)	0.053	0.957	1	1
t-distributed Z_{1j}, Z_{2j} and Z_{3j}				
(200,3,15,30)	0.063	0.089	0.135	0.172
(200,15,15,30)	0.059	0.445	0.839	0.982
(200,30,30,30)	0.060	0.878	1	1
(400,3,15,30)	0.056	0.104	0.119	0.175
(400,15,15,30)	0.049	0.490	0.878	0.995
(400,30,30,30)	0.068	0.904	0.997	1
(1000,3,15,30)	0.057	0.100	0.130	0.202
(1000,15,15,30)	0.048	0.551	0.938	1
(1000,30,30,30)	0.051	0.964	1	1

and μ_3 had $[p^{0.6}]$ non-zero entries of equal value r, which were uniformly allocated among the p components of μ_3. The covariance matrices Σ_1 and Σ_2 were chosen to be the model (a), and Σ_3 was the model (b) specified in Section 4.1.

Table 5 displays the empirical sizes and powers of the proposed ANOVA test subject to different values of p, n_1, n_2, n_3 and r. For all the cases, the empirical sizes were quite close to the nominal significance level of 0.05. Moreover, for each specific (p, n_1, n_2, n_3), the powers of the ANOVA test were increased as the signal strength r was increased. For each specific r and p, the powers were increased as sample sizes were increased.

5. Application to fMRI dataset

To demonstrate the practical use of the proposed tests, we consider the StarPlus fMRI data in [14], which is publicly available from Carnegie Mellon University's Center for Cognitive Brain Imaging. The original data consist of different trials and we use a subset in which each of six human subjects was provided a sentence first for four seconds, followed by a blank screen for four seconds. The subject was then provided a picture for four seconds, followed by answering whether the sentence correctly described the picture. At last, the subject was given a rest for fifteen seconds. There are in total 55 images collected every 0.5 s. At each time point, the image is marked with 25–30 anatomically defined regions called regions of interest (ROIs). In fMRI, ROI analysis is a useful method of selecting a cluster of voxels for exploring patterns of activation across stimuli.

Our interest is to identify the ROIs which react differently to a sentence and a picture. To accommodate high dimensionality, we consider the ROIs with the number of voxels greater than 60. The names of these ROIs are described in Table 6. We let μ_{1i} and μ_{2i} be the population means of the ith ROI with respect to a picture and a sentence, respectively. The null hypotheses of interest are $H_{0i} : \mu_{1i} = \mu_{2i}, i \in \{1, \ldots, 15\}$, where 15 is the number of ROIs. Since the dataset has a very small sample size 6, it is more appropriate to apply the proposed test rather than other competitors requiring diverging sample sizes. For each ROI, we computed the difference between the fMRI image at 29 s and that at 9 s. The two time points are the ends of the time intervals during which the picture and the sentence were presented, respectively. We applied the proposed test to obtain the p-values of the 15 ROIs displayed by Fig. 3. By further applying the Bonferroni correction to control the family-wise error rate at 0.05, the two ROIs named LIPS and LT were identified to be significant as their p-values were less than 0.05/15. The functions of LIPS are related to directing eye movements and reaching and visual attention. On the other hand, LT, which is the most dominant in people, is associated with understanding language and remembering verbal information.

6. Conclusions

Many existing asymptotic tests for high-dimensional data require both dimensionality and sample sizes to diverge. Type I error rate may not be accurately controlled when sample sizes are very small. We addressed the one-sample,

Table 6
The 15 ROIs (left column) and the corresponding full names (right column) from the StarPlus dataset. Each ROI consists of at least 60 voxels in order to accommodate high dimensionality. The proposed two-sample test is applied to select the ROIs which react differently to a sentence and a picture.

ROI	Full name
CALC	Calcarine Sulcus
LDLPFC/RDLPFC	Left/Right Dorsolateral Prefrontal Cortex
LIPL	Left Inferior Parietal Lobule
LIPS/RIPS	Left/Right Intraparietal Sulcus
LIT/RIT	Left/Right Inferior Temporal Lobe
LOPER/ROBER	Left/Right Opercularis
LSPL/RSPL	Left Superior Parietal Lobe
LT/RT	Left/Right Temporal Lobe
LTRIA/RTRIA	Left/Right Triangularis
SMA	Supplementary Motor Area

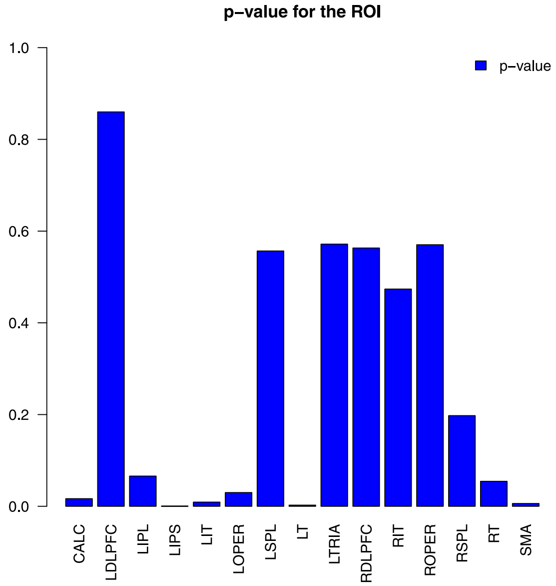

Fig. 3. Bar plot of the p-values for the 15 ROIs. The p-value for each ROI is the probability of the t random variable with 14 degrees of freedom greater than the two-sample test statistic $V_{n_1 n_2}/\hat{\sigma}_{n_1 n_2,0}$ calculated from the fMRI images at 29 s and at 9 s.

two-sample and ANOVA testing problems for high-dimensional means with new proposed tests, which only require dimensionality to diverge but sample sizes to be fixed no less than 3. Without requiring relatively large sample sizes, the tests can be applied to the high-dimensional data such as fMRI, microarray, RNA-Seq, and GWA data with very small sample sizes due to ethical and cost reasons. Except maintaining accurate type I error rate, the tests are nonparametric without assuming normal distributions of data.

The proposed tests are of sum-of-squares type. Similar to other sum-of-squares type tests, the tests are powerful when mean difference of all data components is dense and sum-of-squares signal strength is strong, but not very powerful if mean difference is sparse and signal strength is weak. Some sophisticated technicals such as thresholding and random projection may be applied to the proposed approach for power improvement. Note that in hypothesis testing, maintaining an accurate type I error is the primary requirement before seeking a better power performance. The proposed tests made an attempt on this direction and extending the current approach for power improvement provides an interesting direction for future work.

The proposed tests are computationally inexpensive. For the proposed one-sample test, the test statistic can be quickly computed as the ratio of the sample mean to the sample standard deviation, once the equivalent univariate sample $\{X_i^\top X_j, i < j\}_{i,j=1}^n$ of size $n(n-1)/2$ is obtained from the original sample $\{X_i, 1 \le i \le n\}$. Similar ideas are applied to the proposed two-sample and ANOVA tests. To provide a quantitative evaluation, we compared the computational time of the proposed one-sample test with the one-sample CQ test for the samples with different sample sizes and data dimensionalities. The computations were executed on a 12-core 2.1 GHz Intel processor with 32 GB RAM. In general, the proposed one-sample test was faster than the CQ test. For example, the average execution time of the proposed one-sample test was 3.07×10^{-5}, 8.42×10^{-5}, 7.34×10^{-5} and 5.06×10^{-4} seconds, respectively when the corresponding

$(n, p) = (3, 100), (3, 1000), (30, 100)$ and $(30, 1000)$. On the other hand, the average execution time of the one-sample CQ test was 4.83×10^{-5}, 1.07×10^{-4}, 3.40×10^{-4} and 7.05×10^{-4} seconds, respectively for the same combinations of n and p.

CRediT authorship contribution statement

Jun Li: Conceptualization, Methodology, Software, Writing, Editing.

Acknowledgments

We thank the Editor, Associate Editor and two referees for their constructive comments which have significantly improved the quality of the paper. The author is partially supported by NSF grant DMS-1916239, and NIH grant R15GM139115.

Appendix. Technical details

Proof of Theorem 1. From $\{X_i, 1 \le i \le n\}$, we construct a sequence of $n(n-1)/2$ random variables $\{X_i^\top X_j, i < j\}_{i,j=1}^n$. To establish the asymptotic normality of U_n, we need to show that the sequence $\{X_i^\top X_j/\sqrt{\mathrm{tr}(\Sigma^2) + 2\mu^\top \Sigma \mu}, i < j\}_{i,j=1}^n$ converges to a joint multivariate normal distribution as $p \to \infty$. According to the Cramér–Wold device, we only need to show that $\sum_{i<j}^n c_{ij} X_i^\top X_j/\sqrt{\mathrm{tr}(\Sigma^2) + 2\mu^\top \Sigma \mu}$ is asymptotically normal as $p \to \infty$, where $\{c_{ij}, i < j\}_{i,j=1}^n$ are some constants and at least one of them is nonzero.

We first establish the asymptotic normality of $\sum_{i<j}^n c_{ij} X_i^\top X_j/\sqrt{\mathrm{tr}(\Sigma^2)}$ under the null hypothesis. From (3), we see that $X_i = \Gamma Z_i$. Based on that,

$$\sum_{i<j}^n \frac{c_{ij} X_i^\top X_j}{\sqrt{\mathrm{tr}(\Sigma^2)}} = \sum_{i<j}^n \frac{c_{ij}}{\sqrt{\mathrm{tr}(\Sigma^2)}} Z_i^\top \Gamma^\top \Gamma Z_j = \sum_{k=1}^p \sum_{\ell=1}^p \sum_{i<j}^n \frac{c_{ij}}{\sqrt{\mathrm{tr}(\Sigma^2)}} (\Gamma^\top \Gamma)_{k\ell} z_{ik} z_{j\ell}.$$

To simplify notation, we define

$$A_{ij,k\ell} = \frac{c_{ij}}{\sqrt{\mathrm{tr}(\Sigma^2)}} (\Gamma^\top \Gamma)_{k\ell},$$

if $k > \ell$, and if $k = \ell$,

$$A_{ij,kk} = \frac{c_{ij}}{2\sqrt{\mathrm{tr}(\Sigma^2)}} (\Gamma^\top \Gamma)_{kk}.$$

Then, using the symmetry, we can write

$$\sum_{i<j}^n \frac{c_{ij} X_i^\top X_j}{\sqrt{\mathrm{tr}(\Sigma^2)}} = \sum_{k=1}^p \sum_{\ell=1}^k \sum_{i<j}^n A_{ij,k\ell}(z_{ik} z_{j\ell} + z_{i\ell} z_{jk}) = \sum_{k=1}^p V_k,$$

where $V_k = \sum_{\ell=1}^k \sum_{i<j}^n A_{ij,k\ell}(z_{ik} z_{j\ell} + z_{i\ell} z_{jk})$. Let $S_h = \sum_{k=1}^h V_k$. Since we have $\mathrm{E}(S_q|S_h) = S_h$ for any $q > h$, we see that S_h is a martingale. We therefore use the martingale central limit theorem to establish the asymptotic normality of $\sum_{i<j}^n c_{ij} X_i^\top X_j/\sqrt{\mathrm{tr}(\Sigma^2)}$. According to the martingale central limit theorem, it is equivalent to proving the following two results:

$$\sum_{k=1}^p \mathrm{E}(V_k^2|\mathcal{F}_{k-1}) \to \sum_{i<j}^n c_{ij}^2 \quad \text{in probability}, \tag{13}$$

$$\sum_{k=1}^p \mathrm{E}\{V_k^2 \mathrm{I}(|V_k| > \epsilon)|\mathcal{F}_{k-1}\} \to 0 \quad \text{in probability}, \tag{14}$$

where $\mathcal{F}_{k-1}$ is the σ algebra generated by $\{z_{i1}, \ldots, z_{ik-1}\}$ for $i \in \{1, \ldots, n\}$, and ϵ is any small positive number.

To prove (13), we show $\sum_{k=1}^p \mathrm{E}\, V_k^2 \to \sum_{i<j}^n c_{ij}^2$ and $\mathrm{Var}\{\sum_k \mathrm{E}(V_k^2|\mathcal{F}_{k-1})\} \to 0$, respectively. To this end, we notice

$$V_k^2 = \sum_{\ell_1=1}^k \sum_{\ell_2=1}^k \sum_{i_1<j_1}^n \sum_{i_2<j_2}^n A_{i_1 j_1, k\ell_1} A_{i_2 j_2, k\ell_2} (z_{i_1 k} z_{j_1 \ell_1} + z_{i_1 \ell_1} z_{j_1 k})(z_{i_2 k} z_{j_2 \ell_2} + z_{i_2 \ell_2} z_{j_2 k}).$$

Then,

$$\sum_{k=1}^p \mathrm{E}\, V_k^2 = 2 \sum_{k=1}^p \sum_{\ell=1}^k \sum_{i<j}^n A_{ij,k\ell}^2 + 2 \sum_{k=1}^p \sum_{i<j}^n A_{ij,kk}^2 = \sum_{i<j}^n c_{ij}^2.$$

Next, it can be seen that

$$\sum_k \mathrm{E}(V_k^2|\mathcal{F}_{k-1}) = 4\sum_{k=1}^{p}\sum_{i<j} A_{ij,kk}^2 + \sum_{k=1}^{p}\sum_{\ell_1=1}^{k-1}\sum_{\ell_2=1}^{k-1}\sum_{i<j_1}\sum_{i<j_2} A_{ij_1,k\ell_1}A_{ij_2,k\ell_2}z_{j_1\ell_1}z_{j_2\ell_2}$$
$$+\sum_{k=1}^{p}\sum_{\ell_1=1}^{k-1}\sum_{\ell_2=1}^{k-1}\sum_{i_1<j_1}\sum_{i_2<i_1} A_{i_1j_1,k\ell_1}A_{i_2i_1,k\ell_2}z_{j_1\ell_1}z_{i_2e\ell_2} + \sum_{k=1}^{p}\sum_{\ell_1=1}^{k-1}\sum_{\ell_2=1}^{k-1}\sum_{i_1<i_2}\sum_{i_2<j_2} A_{i_1i_2,k\ell_1}A_{i_2j_2,k\ell_2}z_{i_1\ell_1}z_{j_2\ell_2}$$
$$+\sum_{k=1}^{p}\sum_{\ell_1=1}^{k-1}\sum_{\ell_2=1}^{k-1}\sum_{i_1<j_1}\sum_{i_2<j_1} A_{i_1j_1,k\ell_1}A_{i_2j_1,k\ell_2}z_{i_1\ell_1}z_{i_2\ell_2}.$$

Taking the expectation of the above, we can show that

$$\mathrm{E}[\{\sum_k \mathrm{E}(V_k^2|\mathcal{F}_{k-1})\}]^2 = (\sum_{i<j}^{n} c_{ij}^2)^2 + O\left\{(\sum_{i<j}^{n} c_{ij}^2)^2\frac{\mathrm{tr}(\Sigma^4)}{\mathrm{tr}^2(\Sigma^2)}\right\}.$$

As a result, when $p\to\infty$,

$$\mathrm{Var}\{\sum_k \mathrm{E}(V_k^2|\mathcal{F}_{k-1})\} = O\left\{(\sum_{i<j}^{n} c_{ij}^2)^2\frac{\mathrm{tr}(\Sigma^4)}{\mathrm{tr}^2(\Sigma^2)}\right\}\to 0,$$

because $\mathrm{tr}(\Sigma^4)=o\{\mathrm{tr}^2(\Sigma^2)\}$ according to the condition (C1).

By Chebyshev Inequality, to prove (14), we only need to show that

$$\sum_{k=1}^{p}\mathrm{E}\,V_k^4\to 0.$$

Using $V_k=\sum_{\ell=1}^{k}\sum_{i<j}^{n} A_{ij,k\ell}(z_{ik}z_{j\ell}+z_{i\ell}z_{jk})$, we can show that for some constant C,

$$\sum_{k=1}^{p}\mathrm{E}\,V_k^4 \le \frac{C}{\mathrm{tr}^2(\Sigma^2)}(\sum_{i<j}c_{ij})^4\sum_k\sum_{\ell_1\ell_2}(\Gamma^\top\Gamma)_{k\ell_1}(\Gamma^\top\Gamma)_{k\ell_1}(\Gamma^\top\Gamma)_{k\ell_2}(\Gamma^\top\Gamma)_{k\ell_2}$$
$$\le \frac{C}{\mathrm{tr}^2(\Sigma^2)}(\sum_{i<j}c_{ij})^4\mathrm{tr}(\Sigma^4)\to 0,$$

because $\mathrm{tr}(\Sigma^4)=o\{\mathrm{tr}^2(\Sigma^2)\}$ according to the condition (C1).

Based on (13) and (14), we apply the martingale central limit theorem to establish the asymptotic normality of $\sum_{i<j}^{n}c_{ij}X_i^\top X_j/\sqrt{\mathrm{tr}(\Sigma^2)}$ under the null hypothesis.

To prove asymptotic normality of $\sum_{i<j}^{n}c_{ij}X_i^\top X_j$ under the alternative, we use (3) to write

$$\sum_{i<j}^{n}c_{ij}X_i^\top X_j=\sum_{i<j}^{n}c_{ij}\mu^\top\mu+\sum_{i<j}^{n}c_{ij}\mu^\top\Gamma Z_j+\sum_{i<j}^{n}c_{ij}\mu^\top\Gamma Z_i+\sum_{i<j}^{n}c_{ij}Z_i^\top\Gamma^\top\Gamma Z_j,$$

where the last term remains under the null hypothesis and its asymptotic normality has been established. Next, we need to establish the asymptotic normality of $\sum_{i<j}^{n}c_{ij}\mu^\top\Gamma Z_j+\sum_{i<j}^{n}c_{ij}\mu^\top\Gamma Z_i$. By observing that $Z_i=(z_{i1},\ldots,z_{ip})^\top$ and $\{z_k\}_{k=1}^{p}$ are mutually independent, the asymptotic normality of $\sum_{i<j}^{n}c_{ij}\mu^\top\Gamma Z_j+\sum_{i<j}^{n}c_{ij}\mu^\top\Gamma Z_i$ can be established from the Lyapunov's condition. At last, we can show that $\sum_{i<j}^{n}c_{ij}\mu^\top\Gamma Z_j+\sum_{i<j}^{n}c_{ij}\mu^\top\Gamma Z_i$ and $\sum_{i<j}^{n}c_{ij}Z_i^\top\Gamma^\top\Gamma Z_j$ are asymptotically independent. We thus establish the asymptotic normality of $\sum_{i<j}^{n}c_{ij}X_i^\top X_j$ under the alternative hypothesis. This completes the proof of Theorem 1. □

Proof of Theorem 2. In Theorem 1, we show that under the null hypothesis, for $1\le i<j\le n$, $(X_1^\top X_2,\ldots,X_i^\top X_j,\ldots,X_{n-1}^\top X_n)^\top$ follows an asymptotic $n(n-1)/2$-variate multivariate normal distribution with mean equal to zero and covariance matrix equal to $\mathrm{tr}(\Sigma^2)I_{n(n-1)/2}$, where $I_{n(n-1)/2}$ is the $n(n-1)/2\times n(n-1)/2$ identity matrix. Based on $\{X_i^\top X_j, i<j\}_{i,j=1}^{n}$, we estimate the unknown $\mathrm{tr}(\Sigma^2)$ by the sample variance (4). Since $\{X_i^\top X_j, i<j\}_{i,j=1}^{n}$ are asymptotically independent and normally distributed random variables, $k\,\widehat{\mathrm{tr}(\Sigma^2)}/\mathrm{tr}(\Sigma^2)$ converges to $\chi^2(k)$ as $p\to\infty$, where $k=n(n-1)/2-1$. This completes the proof of Theorem 2. □

Proof of Theorem 3. The statistic U_n is the sample mean of $\{X_i^\top X_j, i<j\}_{i,j=1}^{n}$. From the proof of Theorem 2, U_n is asymptotically independent of the sample variance (4). As a result, $U_n/\hat{\sigma}_{n,0}$ converges to a t-distribution with $k=n(n-1)/2-1$ degrees of freedom. This completes the proof of Theorem 3. □

Proof of Theorems 4 and 5. Theorems 4 and 5 can be shown by replacing X_i with Y_i in the proofs of Theorems 1–3. We therefore omit it. □

References

[1] T.W. Anderson, An Introduction To Multivariate Statistical Analysis, Wiley, Hoboken, NJ, 2003.
[2] Z. Bai, H. Saranadasa, Effect of high dimension: by an example of a two sample problem, Statist. Sinica 6 (1996) 311–329.
[3] T. Cai, W. Liu, Y. Xia, Two-sample test of high dimensional means under dependence, J. R. Stat. Soc. Ser. B Stat. Methodol. 76 (2014) 349–372.
[4] S.X. Chen, J. Li, P. Zhong, Two-sample and ANOVA tests for high dimensional means, Ann. Statist. 47 (2019) 1443–1474.
[5] L. Chen, D. Paul, R. Prentice, P. Wang, A regularized Hotelling's T^2 test for pathway analysis in proteomic studies, J. Amer. Statist. Assoc. 106 (2011) 1345–1360.
[6] S.X. Chen, Y. Qin, A two sample test for high dimensional data with applications to gene-set testing, Ann. Statist. 38 (2010) 808–835.
[7] K. Gregory, R. Carroll, V. Baladandayuthapani, S. Lahiri, A two-sample test for equality of means in high dimension, J. Amer. Statist. Assoc. 110 (2015) 837–849.
[8] Y. He, G. Xu, C. Wu, W. Pan, Asymptotically independent U-statistics in high-dimensional testing, Ann. Statist. 49 (2021) 154–181.
[9] H. Hotelling, The generalization of student's ratio, Ann. Math. Stat. 2 (1931) 360–378.
[10] Y. Huang, C. Li, R. Li, S. Yang, An overview of tests on high-dimensional means, J. Multivariate Anal. 188 (2022).
[11] H. Li, A. Aue, D. Paul, J. Peng, P. Wang, An adaptable generalization of Hotelling's T^2 test in high dimension, Ann. Statist. 48 (2020) 1815–1847.
[12] J. Li, S.X. Chen, Two sample tests for high dimensional covariance matrices, Ann. Statist. 40 (2012) 908–940.
[13] C. Li, R. Li, Linear hypothesis testing in linear models with high dimensional responses, J. Amer. Statist. Assoc. 117 (2022) 1738–1750.
[14] T. Mitchell, R. Hutchinson, M. Just, R. Niculescu, F. Pereira, X. Wang, Classifying instantaneous cognitive states from fMRI data, in: The American Medical Informatics Association 2003 Annual Symposium, American Medical Informatics Association, Washington, DC, 2003, pp. 8–12.
[15] H. Scheffé, On solutions of the behrens-Fisher problem, based on the t-distribution, Ann. Math. Stat. 14 (1943) 35–44.
[16] M. Srivastava, M. Du, A test for the mean vector with fewer observations than the dimension, J. Multivariate Anal. 99 (2008) 386–402.
[17] R. Srivastava, P. Li, D. Ruppert, RAPTT: an exact two-sample test in high dimensions using random projections, J. Comput. Graph. Stat. 25 (2016) 954–970.
[18] G. Székely, M. Rizzo, The distance correlation t-test of independence in high dimension, J. Multivariate Anal. 117 (2013) 193–213.
[19] M. Thulin, A high-dimensional two-sample test for the mean using random subspaces, Comput. Statist. Data Anal. 74 (2014) 26–38.
[20] L. Wang, B. Peng, R. Li, A high-dimensional nonparametric multivariate test for mean vector, J. Amer. Statist. Assoc. 110 (2015) 1658–1669.
[21] G. Xu, L. Lin, P. Wei, W. Pan, An adaptive two-sample test for high-dimensional means, Biometrika 103 (2016) 609–624.
[22] P. Zhong, S.X. Chen, M. Xu, Tests alternative to higher criticism for high dimensional means under sparsity and column-wise dependence, Ann. Statist. 41 (2013) 2820–2851.