Stat Comput manuscript No.
(will be inserted by the editor)

A feasible high dimensional randomization test for mean vector

Rui Wang · **Xingzhong Xu**

Received: date / Accepted: date

Abstract The strength of randomization tests is that they are exact tests under certain symmetry assumption for distributions. In this paper, we study a randomization test for mean vector in high dimensional setting. In classical statistics, a major down-side to randomization tests is that they are computational intensive. Surprisingly, it is not the case in high dimensional setting. We give an implementation of the randomization procedure, the computational complexity of which does not rely on data dimension. The theoretical property of randomization test is another important issue. So far, the asymptotic behaviors of randomization tests have only been studied in fixed dimension case. We investigate the asymptotic behavior of the randomization test in high dimensional setting. It turns out that even if the symmetry assumption is violated, the randomization test has correct level asymptotically. The asymptotic power function is also given. With fast implementation and good theoretical properties, the randomization test can be recommended in practice.

Keywords Asymptotic power function · High dimension · Randomization test · Symmetry assumption

This work was supported by the National Natural Science Foundation of China under Grant No. 11471035, 11471030.

Rui Wang · Xingzhong Xu
School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081,China

Xingzhong Xu (✉)
Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing 100081,China
E-mail: xuxz@bit.edu.cn
Tel.: +86-13681402299

1 Introduction

Suppose $X_1,\dots,X_n$ are independent and identically distributed (iid) p-dimensional random vectors with mean vector $\mu = (\mu_1,\dots,\mu_p)^T$ and covariance matrix Σ. In this paper, we consider the randomization test procedure for testing the hypotheses

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

A classical test statistic for hypotheses (1) is Hotelling's T^2, defined as $n\bar{X}^T S^{-1}\bar{X}$, where $\bar{X} = n^{-1}\sum_{i=1}^n X_i$ and $S = (n-1)^{-1}\sum_{i=1}^n (X_i-\bar{X})(X_i-\bar{X})^T$ are the sample mean vector and sample covariance matrix, respectively. Under normal distribution, Hotelling's T^2 is the likelihood ratio test and enjoys desirable properties in fixed p case. See, for example, Anderson (2003). However, Hotelling's test can not be defined when $p > n-1$ due to the singularity of S. In a seminal paper, Bai and Saranadasa (1996) considered two sample testing problem and proposed a statistic by removing S^{-1} from Hotelling's T^2 statistic. They studied the asymptotic properties of their test statistic when p/n tends to a positive constant. Many subsequent papers generalized the idea of Bai and Saranadasa (1996) to more general models (Srivastava and Du 2008; Chen and Qin 2010; Wang et al. 2015). The critical value of existing high dimensional tests are mostly determined by asymptotic distribution. We call it asymptotic method. In many real world problems, e.g., gene testing (Bradley Efron 2007), sample size n may be very small. In this case, the Type I error rate of asymptotic method may be far away from nominal level.

The idea of randomization test dates back to Fisher (1936), which is a tool to determine the critical value for a given test statistic. See Romano (1990) for a general construction of randomization test. Its strength is in that the resulting test procedure has exact level under mild condition. There are many papers concerning the theoretical properties of randomization tests for fixed p case. See, for example, Romano (1990), Zhu and Neuhaus (2000) and Chung and Romano (2016). In high dimensional setting, randomization tests are widely used in applied statistics (Subramanian et al. 2005; Bradley Efron 2007; Ko et al. 2016). However, little is known about the theoretical properties of the randomization test in high dimensional setting.

In this paper, we consider the following randomization method. Suppose $T(X_1,\ldots,X_n)$ is certain test statistic for hypotheses (1). Let $\varepsilon_1,\ldots,\varepsilon_n$ be iid Rademacher variables ($\Pr(\varepsilon_i=1)=\Pr(\varepsilon_i=-1)=1/2$) which are independent of data. Denote by $\mathscr{L}\big(T(\varepsilon_1X_1,\ldots,\varepsilon_iX_i,\ldots,\varepsilon_nX_n)|X_1,\ldots,X_n\big)$ the distribution of $T(\varepsilon_1X_1,\ldots,\varepsilon_iX_i,\ldots,\varepsilon_nX_n)$ conditioning on $X_1,\ldots,X_n$. The randomization test rejects the null hypothesis when $T(X_1,\ldots,X_n)$ is greater than the $1-\alpha$ quantile of the conditional distribution and accepts the null hypothesis otherwise, where α is the significant level and the $1-\alpha$ quantile of a distribution function $F(\cdot)$ is defined as $\inf\{y:F(y)\geq 1-\alpha\}$. In fixed p setting, it's well known that randomization test consumes much more computing time than asymptotic method, which historically hampered it's use. The goal of this paper is to show that randomization is feasible in high dimension while still have the statistic properties and asymptotic power. Inspired by the work of Bai and Saranadasa (1996) and Chen and Qin (2010), we propose a randomization test for hypotheses (1). We give a fast implementation of the randomization test, the computational complexity of which does not depend on p. When p is large, our method even consumes less computing time than asymptotic method. We also investigate the asymptotic behavior of test procedure. Our results show that even if the null distribution of X_1 is not symmetric, the randomization test is still asymptotically exact under mild assumptions. Hence the test procedure is robust. The local asymptotic power function is also given. To the best of our knowledge, this is the first work which gives the asymptotic behavior of randomization test in high dimensional setting. Our work shows that the randomization test is very suitable for high dimensional testing problem since it is not only easy to compute but also has good statistical properties.

The rest of the paper is organized in the following way. In Section 2, we propose a randomization test and give a fast implementation. In Section 3, we investigate the asymptotic behavior of the proposed test. The simulation results are reported in Section 4. The technical proofs are presented in Appendix.

2 Test Procedure

Consider testing the hypotheses (1) in high dimensional setting. It is known that Hotelling's T^2 can not be defined when $p>n-1$. Bai and Saranadasa (1996) removed the S^{-1} from Hotelling's T^2 statistic and proposed a statistic which has good power behavior in high dimensional setting. Their idea can also be used for testing hypotheses (1) and the statistic becomes $\bar{X}^T\bar{X}$. The asymptotic properties of the statistic requires p/n tends to a positive constant. Chen and Qin (2010) found that the restriction on p and n can be considerably relaxed by removing the diagonal elements in the statistic of Bai and Saranadasa (1996). For hypotheses (1), their statistic is $\sum_{i\neq j}X_i^TX_j$. Inspired by the statistic of Bai and Saranadasa (1996) and Chen and Qin (2010), we consider the statistic

$$T(X_1,\ldots,X_n)=\sum_{j<i}X_i^TX_j. \tag{2}$$

Let $\varepsilon_1,\ldots,\varepsilon_n$ be iid Rademacher variables which are independent of data. Denote by

$$\mathscr{L}\big(T(\varepsilon_1X_1,\ldots,\varepsilon_iX_i,\ldots,\varepsilon_nX_n)|X_1,\ldots,X_n\big) \tag{3}$$

the distribution of $T(\varepsilon_1X_1,\ldots,\varepsilon_iX_i,\ldots,\varepsilon_nX_n)$ conditioning on $X_1,\ldots,X_n$. We propose a test procedure with test function $\phi(X_1,\ldots,X_n)$ to be equal to 1 if $T(X_1,\ldots,X_n)$ is greater than the $1-\alpha$ quantile of the conditional distribution (3) and equal to 0 otherwise. Since $T(X_1,\ldots,X_n)$ equals to half of the Chen and Qin (2010)'s statistic $\sum_{i\neq j}X_i^TX_j$, the test procedure $\phi(X_1,\ldots,X_n)$ is the randomization version of Chen and Qin (2010)'s test procedure. On the other hand, note that Bai and Saranadasa (1996)'s statistic $\bar{X}^T\bar{X}$ can be written as $n^{-2}\sum_{i=1}^n\sum_{j=1}^nX_i^TX_j$. Since $\sum_{i=1}^nX_i^TX_i$ is invariance under randomization, the test procedure $\phi(X_1,\ldots,X_n)$ is also the randomization version of Bai and Saranadasa (1996)'s test.

Under certain symmetric assumption, randomization test is exact test, which is a desirable property. See, for example, E. L. Lehmann (2005, Chapter 15). In our problem, the Type I error of $\phi(X_1,\ldots,X_n)$ is not larger than α provided X_1 and $-X_1$ have the same distribution under null hypothesis. By refined definition of $\phi(X_1,\ldots,X_n)$ on the boundary of rejection region, one can obtain a test procedure with exact level. Such refinement only has minor effect on the test procedure and won't be considered in this paper.

The test procedure $\phi(X_1,\ldots,X_n)$ can be equivalently implemented by p-value. Define

$$\begin{aligned}&p(X_1,\ldots,X_n)\\&=\Pr(T(\varepsilon_1X_1,\ldots,\varepsilon_nX_n)\geq T(X_1,\ldots,X_n)|X_1,\ldots,X_n).\end{aligned} \tag{4}$$

Then the test procedure rejects the null hypothesis when $p(X_1,\ldots,X_n)\leq\alpha$.

The randomized statistic $T(\varepsilon_1X_1,\ldots,\varepsilon_nX_n)$ is uniformly distributed on 2^n values conditioning on $X_1,\ldots,X_n$. To compute the exact quantile of (3) or the p-value (4), one needs to calculate at least 2^n values, which is not feasible even when n is moderate. In practice, randomization test is often realized through an approximation of p-value (4). More specifically, we sample $\varepsilon_1^*, \dots, \varepsilon_n^*$ and compute the randomized statistic $T^* = T(\varepsilon_1^* X_1, \dots, \varepsilon_n^* X_n)$. Repeat B times for a large B and we obtain $T_1^*, \dots, T_B^*$. Let $T_0 = T(X_1, \dots, X_n)$ be the original statistic and define

$$\tilde{p}(X_1, \dots, X_n) = \frac{1}{B+1}\big(1 + \sum_{i=1}^{B} \mathbf{1}_{\{T_i^* \geq T_0\}}\big).$$

The test is rejected when $\tilde{p}(X_1, \dots, X_n) \leq \alpha$. This procedure can also control the significant level. In fact, we have $\Pr(\tilde{p}(X_1, \dots, X_n) \leq u) \leq u$ for all $0 \leq u \leq 1$. See E. L. Lehmann (2005, Page 636). Moreover, by Bernoulli's law of large numbers, we have $\tilde{p}(X_1, \dots, X_n) \xrightarrow{P} p(X_1, \dots, X_n)$ as $B \to \infty$. Here we emphasis that the convergence rate of $\tilde{p}(X_1, \dots, X_n)$ to $p(X_1, \dots, X_n)$ only relies on $p(X_1, \dots, X_n)$. Hence the choice of B can be independent of the sample size n and the dimension of data p.

Now we consider the implementation of the randomization test procedure. The computation of T_0 costs $O(n^2 p)$ operations. To obtain T_i^*, $i = 1, \dots, B$, we need to generate $\varepsilon_1, \dots, \varepsilon_n$ and compute

$$T(\varepsilon_1 X_1, \dots, \varepsilon_n X_n) = \sum_{1 \leq j < i \leq n} X_i^T X_j \varepsilon_i \varepsilon_j.$$

Note that $X_i^T X_j$ $(1 \leq j < i \leq n)$ can be computed beforehand. Once we obtain $X_i^T X_j$, the computation of T_i^* cost $O(n^2)$ operations. Thus, the randomization test costs $O(n^2 p + n^2 B)$ operations in total. When p is large compared with n, the computation of T_0 consumes almost the whole computing time and the computing time of the randomization procedure is relatively low. The randomization method doesn't need an estimator of variance, which is necessary in asymptotic method. A good variance estimator is complicated, see Chen and Qin (2010), and consumes much computing time. Hence the randomization test method is very competitive compared with asymptotic method. This is different from low dimensional setting where randomization tests consume much more computing time than asymptotic method.

If we only care about the decision (reject or accept) and the p-value is not needed, the computing time of the randomization test can be further reduced. In fact, the rejection region $\tilde{p}(X_1, \dots, X_n) \leq \alpha$ can be written as

$$\sum_{i=1}^{B} (1 - \mathbf{1}_{\{T_i^* \geq T_0\}}) \geq B + 1 - (B+1)\alpha.$$

Since the left hand side is a sum of non-negative values, we can reject the null hypothesis once $\sum_{i=1}^{B_0} (1 - \mathbf{1}_{\{T_i^* \geq T_0\}}) \geq B + 1 - (B+1)\alpha$ for some B_0. Similarly, the acceptance region can be written as

$$\sum_{i=1}^{B} \mathbf{1}_{\{T_i^* \geq T_0\}} > (B+1)\alpha - 1.$$

we can accept the null hypothesis once $\sum_{i=1}^{B_0} \mathbf{1}_{\{T_i^* \geq T_0\}} > (B+1)\alpha - 1$ for some B_0. The Algorithm 1 summarizes our computing method.

Algorithm 1: Randomization Algorithm

Data: Data $X_1, \dots, X_n$
Result: Reject or accept the null hypothesis
for $i \leftarrow 2$ **to** n **do**
  for $j \leftarrow 1$ **to** $i-1$ **do**
    $D_{ij} \leftarrow X_i^T X_j$;
  end
end
Compute $T_0 \leftarrow \sum_{1 \leq j < i \leq n} D_{ij}$;
Set $A \leftarrow 0$;
for $i = 1$ *to* B **do**
  Generate $\varepsilon_1, \dots, \varepsilon_n$ according to $\Pr(\varepsilon_i = 1) = \Pr(\varepsilon_i = -1) = \frac{1}{2}$;
  if $\sum_{1 \leq j < i \leq n} D_{ij} \varepsilon_i \varepsilon_j \geq T_0$ **then**
    $A \leftarrow A + 1$;
    if $A > (B+1)\alpha - 1$ **then return** *Accept*;
  else
    if $i - A \geq B + 1 - (B+1)\alpha$ **then return** *Reject*;
  end
end

3 Asymptotic properties

In this section, we investigate the asymptotic properties of the test procedure $\phi(X_1, \dots, X_n)$. We assume, like Chen and Qin (2010) and Bai and Saranadasa (1996), the following multivariate model:

$$X_i = \mu + \Gamma Z_i \text{ for } i = 1, \dots, n, \tag{5}$$

where Γ is a $p \times m$ matrix for some $m \geq p$ such that $\Gamma \Gamma^T = \Sigma$ and $Z_1, \dots, Z_n$ are m-variate iid random vectors satisfying $\mathrm{E}(Z_i) = 0$ and $\mathrm{Var}(Z_i) = I_m$, the $m \times m$ identity matrix. Write $Z_i = (z_{i1}, \dots, z_{im})^T$. We assume $\mathrm{E}(z_{ij}^4) = 3 + \Delta < \infty$ and

$$\mathrm{E}(z_{il_1}^{\alpha_1} z_{il_2}^{\alpha_2} \cdots z_{il_q}^{\alpha_q}) = \mathrm{E}(z_{il_1}^{\alpha_1}) \mathrm{E}(z_{il_2}^{\alpha_2}) \cdots \mathrm{E}(z_{il_q}^{\alpha_q}) \tag{6}$$

for a positive integer q such that $\sum_{l=1}^{q} \alpha_l \leq 8$ and $l_1 \neq l_2 \neq \cdots \neq l_q$. Note that here X_1 and $-X_1$ don't need to have the same distribution under null hypothesis.

A key assumption in Chen and Qin (2010) is $\mathrm{tr}(\Sigma^4) = o\big(\mathrm{tr}^2(\Sigma^2)\big)$. Let $\lambda_i(\Sigma)$ be the ith largest eigenvalue of Σ. From

$$\frac{\lambda_1(\Sigma)^4}{(\sum_{i=1}^{p} \lambda_i(\Sigma)^2)^2} \leq \frac{\sum_{i=1}^{p} \lambda_i(\Sigma)^4}{(\sum_{i=1}^{p} \lambda_i(\Sigma)^2)^2} \leq \frac{\lambda_1(\Sigma)^2 \sum_{i=1}^{p} \lambda_i(\Sigma)^2}{(\sum_{i=1}^{p} \lambda_i(\Sigma)^2)^2},$$

we can see that $\mathrm{tr}(\Sigma^4) = o\big(\mathrm{tr}^2(\Sigma^2)\big)$ is equivalent to

$$\frac{\lambda_1(\Sigma)}{\sqrt{\mathrm{tr}(\Sigma^2)}} \to 0. \tag{7}$$

Although Chen and Qin (2010)'s results are for two sample case, their results can be proved similarly for one sample case. The following two lemmas restate their theorems.

Lemma 1 *Under* (5)*,* (6)*,* (7) *and local alternatives*

$$\mu^T \Sigma \mu = o(n^{-1}\operatorname{tr}(\Sigma^2)), \tag{8}$$

we have

$$\frac{T(X_1,\ldots,X_n) - \frac{n(n-1)}{2}\mu^T\mu}{\sqrt{\frac{n(n-1)}{2}\operatorname{tr}(\Sigma^2)}} \xrightarrow{\mathscr{L}} N(0,1),$$

where "$\xrightarrow{\mathscr{L}}$" means convergence in law.

Lemma 2 *Under* (5)*,* (6)*,* (7) *and*

$$n^{-1}\operatorname{tr}(\Sigma)^2 = o(\mu^T\Sigma\mu), \tag{9}$$

we have

$$\frac{T(X_1,\ldots,X_n) - \frac{n(n-1)}{2}\mu^T\mu}{\sqrt{(n-1)^2 n\mu^T\Sigma\mu}} \xrightarrow{\mathscr{L}} N(0,1).$$

Now we study the asymptotic properties of the randomization test. The conditional distribution

$$\mathscr{L}\left(\frac{T(\varepsilon_1X_1,\ldots,\varepsilon_iX_i,\ldots,\varepsilon_nX_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}}\middle|X_1,\ldots,X_n\right)$$

plays an important role in our analysis, we shall call it randomization distribution. Let ξ_α^* be the $1-\alpha$ quantile of the randomization distribution. Then it can be seen that the test function $\phi(X_1,\ldots,X_n)$ equals to 1 if

$$\frac{T(X_1,\ldots,X_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}} > \xi_\alpha^*$$

and equals to 0 otherwise.

Since the randomization distribution itself is random, to study its asymptotic distribution, we need to define in what sense the convergence is. Let F and G be two distribution functions on $\mathbb{R}$, Levy metric ρ of F and G is defined as

$$\begin{aligned}&\rho(F,G)\\ =&\inf\{\varepsilon : F(x-\varepsilon)-\varepsilon \le G(x) \le F(x+\varepsilon)+\varepsilon \text{ for all } x\}.\end{aligned}$$

It's well known that $\rho(F_n,F)\to 0$ if and only if $F_n \xrightarrow{\mathscr{L}} F$. The following theorem shows that in high dimensional setting, the randomization distribution tends to a standard normal distribution.

Theorem 1 *Under* (5)*,* (6)*,* (7) *and*

$$\mu^T\mu = o\big(\sqrt{\operatorname{tr}(\Sigma^2)}\big), \tag{10}$$

we have that

$$\rho\left(\mathscr{L}\left(\frac{T(\varepsilon_1X_1,\ldots,\varepsilon_nX_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}}\middle|X_1,\ldots,X_n\right),N(0,1)\right) \xrightarrow{P} 0.$$

It can be proved that

$$\frac{\sum_{j<i}(X_i^TX_j)^2}{\frac{n(n-1)}{2}\operatorname{tr}(\Sigma^2)} \xrightarrow{P} 1.$$

By Lemma 1, under null hypothesis, we have that

$$\frac{T(X_1,\ldots,X_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}} \xrightarrow{\mathscr{L}} N(0,1).$$

Compare this with Theorem 1, we can see that if $\mu^T\mu = o(\sqrt{\operatorname{tr}(\Sigma^2)})$, the randomization distribution mimics the actual null distribution. However, the behavior of randomization distribution is different when condition (10) is not valid. In fact, we have the following result.

Theorem 2 *Under* (5)*,* (6)*,* (7) *and*

$$\sqrt{\operatorname{tr}\Sigma^2} = o\big(\mu^T\mu\big), \tag{11}$$

we have that

$$\rho\left(\mathscr{L}\left(\frac{T(\varepsilon_1X_1,\ldots,\varepsilon_nX_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}}\middle|X_1,\ldots,X_n\right),\frac{\sqrt{2}}{2}\big(\chi_1^2-1\big)\right) \xrightarrow{P} 0,$$

where χ_1^2 is the chi-squared distribution with freedom 1*.*

Once the limit of the randomization distribution is obtained, the asymptotic behavior of ξ_α^* can be derived immediately. Let $\Phi(\cdot)$ be the cumulative distribution function (CDF) of standard normal distribution, we have

Corollary 1 *Under the conditions of Theorem 1, we have*

$$\xi_\alpha^* \xrightarrow{P} \Phi^{-1}(1-\alpha).$$

Corollary 2 *Under the conditions of Theorem 2,*

$$\xi_\alpha^* \xrightarrow{P} \frac{\sqrt{2}}{2}\Big(\big(\Phi^{-1}(1-\tfrac{\alpha}{2})\big)^2-1\Big).$$

Now we are ready to derive the asymptotic power of randomization test. The following two theorems give the power under (8) and (9), respectively.

Theorem 3 *Suppose conditions* (5)*,* (6)*,* (7) *and* (8) *holds. Then*

1. *If* (10) *holds,*

$$\begin{aligned}&\Pr\left(\frac{T(X_1,\ldots,X_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}} > \xi_\alpha^*\right)\\ =&\Phi\Big(-\Phi^{-1}(1-\alpha)+\frac{\sqrt{n(n-1)}\mu^T\mu}{\sqrt{2\operatorname{tr}(\Sigma^2)}}\Big)+o(1).\end{aligned} \tag{12}$$

2. *If* (11) *holds,*

$$\Pr\Big(\frac{T(X_1,\dots,X_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}}>\xi_\alpha^*\Big)\to 1.$$

Theorem 4 *Under* (5), (6), (7) (9) *and either* (10) *or* (11),

$$\Pr\Big(\frac{T(X_1,\dots,X_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}}>\xi_\alpha^*\Big)=\Phi(\frac{\sqrt{n}\mu^T\mu}{2\sqrt{\mu^T\Sigma\mu}})+o(1).$$

Remark 1 Neither (8) or (10) implies the other one. For example, suppose $\Sigma=I_p$, then (8) is equivalent to $\mu^T\mu=o(p/n)$ and (10) is equivalent to $\mu^T\mu=o(\sqrt{p})$. In this case, if $\sqrt{p}/n\to 0$, then (8) implies (10); conversely, if $\sqrt{p}/n\to\infty$, then (10) implies (8).

Theorem 3 implies that under (5), (6) and (7), the Type I error rate of the randomization test tends to the nominal level. Note that this result doesn't assume that the distribution of X_1 is symmetric under null hypothesis. This implies that our test procedure is robust when the symmetry assumption is break down. This property is not held by all randomization tests. See, for example, Romano (1990).

Use the method of Chen and Qin (2010), the asymptotic method rejects the null hypothesis when

$$\frac{\sum_{i\neq j}X_i^TX_j}{\sqrt{2n(n-1)\widehat{\mathrm{tr}(\Sigma^2)}}}>\Phi^{-1}(1-\alpha),$$

where

$$\widehat{\mathrm{tr}(\Sigma^2)}=\frac{1}{n(n-1)}\mathrm{tr}\Big(\sum_{i\neq j}(X_i-\bar{X}_{(i,j)})X_i^T(X_j-\bar{X}_{(i,j)})X_j^T\Big)$$

is a ratio consistent estimator of $\mathrm{tr}(\Sigma^2)$ and $\bar{X}_{(i,j)}$ is the sample mean after excluding X_i and X_j. Lemma 1 implies that the Type I error rate of the asymptotic method tends to the nominal level under assumptions (5), (6) and (7). However, when these assumptions are not satisfied, Lemma 1 may not be valid. For example, we consider the model $X_i=u_i\mathbf{v}$, $i=1,\dots,n$, where $u_1,\dots,u_n$ are iid random variables with $\mathrm{E}u_1=0$, $\mathrm{E}u_1^2=1$ and $\mathbf{v}\in\mathbb{R}^p$ is a vector. In this case,

$$T(X_1,\dots,X_n)=\mathbf{v}^T\mathbf{v}\sum_{j<i}u_iu_j=\frac{1}{2}\mathbf{v}^T\mathbf{v}\big((\sum_{i=1}^n u_i)^2-\sum_{i=1}^n u_i^2\big).$$

By law of large number, we have $\sum_{i=1}^n u_i^2/n\xrightarrow{P}1$. By central limit theorem, we have $\sum_{i=1}^n u_i/\sqrt{n}\xrightarrow{\mathscr{L}}N(0,1)$. Then we have

$$\frac{2T(X_1,\dots,X_n)}{n\mathbf{v}^T\mathbf{v}}+1\xrightarrow{\mathscr{L}}\chi_1^2.$$

Since the asymptotic distribution of $T(X_1,\dots,X_n)$ is not normal distribution, the asymptotic method does not have the correct level asymptotically. For the randomization method, we have

$$\begin{aligned}T(\varepsilon_1X_1,\dots,\varepsilon_nX_n)&=\mathbf{v}^T\mathbf{v}\sum_{j<i}u_iu_j\varepsilon_i\varepsilon_j\\&=\frac{1}{2}\mathbf{v}^T\mathbf{v}\big((\sum_{i=1}^n u_i\varepsilon_i)^2-\sum_{i=1}^n u_i^2\big).\end{aligned}$$

4 Simulation Studies

In this section, we report the simulation performance of the randomization test in various setting. For comparison, we also conduct the simulation of the asymptotic method of Chen and Qin (2010).

11111111111111111111111111

According to Chen and Qin (2010), the asymptotic method rejects the null hypothesis when

$$\frac{\sum_{i\neq j}X_i^TX_j}{\sqrt{2n(n-1)\widehat{\mathrm{tr}(\Sigma^2)}}}>\Phi^{-1}(1-\alpha),$$

where

$$\widehat{\mathrm{tr}(\Sigma^2)}=\frac{1}{n(n-1)}\mathrm{tr}\Big(\sum_{i\neq j}(X_i-\bar{X}_{(i,j)})X_i^T(X_j-\bar{X}_{(i,j)})X_j^T\Big).$$

The associated p-value is

$$p_{CQ}(X_1,\dots,X_n)=1-\Phi\left(\frac{\sum_{i\neq j}X_i^TX_j}{\sqrt{2n(n-1)\widehat{\mathrm{tr}(\Sigma^2)}}}\right).$$

We consider two innovation structure: the moving average model and the factor model. The moving average model has the following structure:

$$X_{ij}=\sum_{l=0}^k\rho_lZ_{i,j+l}$$

for $i=1,\dots,n$ and $j=1,\dots,p$, where Z_{ij}'s are iid random variables with distribution F for $i=1,\dots,n$ and $j=1,\dots,p+k$. Like Chen and Qin (2010), we consider two different F. One is $N(0,1)$, and the other is $(\mathrm{Gamma}(4,1)-4)/2$. We also consider different k. The ρ_i's are generated independently from $U(2,3)$ and kept fixed throughout the simulation. The second model we consider is the factor model in Fan et al. (2007). In the simulation study of Fan et al. (2007), data are generated from a factor model to reflect aspects of gene expression data. The model involves three group factor and one common factor among all p variables. Their data generation mechanism is adopted in our next simulation study. We denote by $\{\xi_{ij}\}_{1\le i\le n,1\le j\le p}$ a sequence of independent $N(0,1)$ and by $\{\chi_{ij}\}_{1\le i\le n,1\le j\le 4}$ a sequence of independent random variables with distribution $(\chi_6^2-6)/\sqrt{12}$.

Note that χ_{ij} has mean 0, variance 1 and skewness $\sqrt{12}/3$. The data is generated by model

$$X_{ij} = \frac{a_{j1}\chi_{i1} + a_{j2}\chi_{i2} + a_{j3}\chi_{i3} + b_j\chi_{i4} + \xi_{ij}}{(1 + a_{j1}^2 + a_{j2}^2 + a_{j3}^2 + b_j^2)^{1/2}} \quad i = 1,\ldots,n \text{ and } j = 1,\ldots,p,$$

where $a_{jk} = 0$ except that $a_{j1} = a_j$ for $j = 1,\ldots,\frac{1}{3}p$, $a_{j2} = a_j$ for $\frac{1}{3}p+1,\ldots,\frac{2}{3}p$ and $a_{j3} = a_j$ for $\frac{2}{3}p+1,\ldots,p$. As in Fan et al. (2007), we consider two configurations of factor loadings. In case I we set $a_j = 0.25$ and $b_j = 0.1$ for $j = 1,\ldots,p$. In case II, a_i and b_i are generated independently from $U(0,0.4)$ and $U(0,0.2)$.

To control the significant level, the null distribution of a p-value should be close to $U(0,1)$, the uniform distribution on $(0,1)$. We simulate the p-value $\tilde{p}(X_1,\ldots,X_n)$ ($B = 1000$) and $p_{CQ}(X_1,\ldots,X_n)$ for 2000 times and Figure 1 plot the empirical distribution function (ECDF) of p-values. As shown by the plots, the p-values of randomization test method is uniform in all cases. As for asymptotic method, the uniformity of p-values depends on model. It performs well for moving average model with $k = 3$. In factor model, the lack of uniformity is obvious. The distribution of p-values is far away from uniform distribution for moving average model with $k = 500$.

In Theorem 1, we proved that the randomization distribution tends to a standard normal distribution under certain conditions. In Figure 2, we plot the histograms of randomization distribution under null hypothesis. For comparison, we also plot the standard normal density. From the plots, we can see that the randomization distribution is very similar to the standard normal distribution in factor model and moving average model with $k = 3$. This verifies our theorem **??**. However, under moving average model with $k = 500$, the randomization distribution is far from standard normal distribution. This implies the accuracy of normal approximation depends on the innovation model.

Now we simulate the empirical power and size. Let $\mathrm{SNR} = \sqrt{n(n-1)}\mu^T\mu/\sqrt{2\operatorname{tr}\Sigma^2}$ be the signal to noise ratio (SNR). The theoretic asymptotic power is an increasing function of SNR. We scale μ to reach different level of SNR. Our simulation consider two mean structure: dense mean and sparse mean. In the dense mean setting, each coordinate of μ is independently generated from $U(2,3)$ and then μ is scaled to reach a given SNR. In the sparse mean setting, we randomly select 5% of μ p coordinates to be non-zero. Each non-zero coordinate is again independently generated from $U(2,3)$ and then scaled to reach a given SNR. We set $B = 1000$ for randomization test method. The empirical power and size are computed based on 2000 simulations.

Table 1 and Table 2 list the empirical power and size for the moving average model. It's not surprising that randomization test can control level well in normal case since it can be proved in theory. The results also show that the randomization method can control level well even in Gamma case, which is not symmetric under null. It justifies the robustness of randomization method. On the other hand, asymptotic method has small size when dependence is weak and has inflated size when dependence is strong. Randomization method has similar power behavior with asymptotic method. The empirical power of both methods is similar to theoretical asymptotic power (12) while it is lower than theoretical asymptotic power when k is large. Table 3 lists the empirical power and size under factor model. Although the distribution is not symmetric, the results show that the level of randomization method is close to 0.05 while asymptotic method suffers from level inflation. In summary, the simulation results show that the randomization method is robust and has similar power with asymptotic method.

Fig. 1 The ECDF of p-values for asymptotic method (AM) and randomization method (RM). $p = 600$, $n = 100$.

5 Conclusion Remark

In this paper, we considered a randomization test for mean vector in high dimensional setting. A fast implementation was provided. We also derived some asymptotic properties of the test procedure. For illustration, we only considered a normal3.png

normal500.png

gamma3.png

gamma500.png

Table 1 Empirical power and size of moving average model with normal innovation. $p = 600$, $n = 100$, $\alpha = 0.05$. TP is the theoretical asymptotic power, RM represents the randomization method, AM represents the asymptotic method.

		Dense means				Sparse means		
		$k = 3$		$k = 500$		$k = 3$		k
SNR	TP	RM	AM	RM	AM	RM	AM	RM
0.0	0.050	0.0445	0.0515	0.0530	0.0745	0.0500	0.0585	0.0515
0.5	0.126	0.1815	0.1940	0.1330	0.1685	0.1735	0.1875	0.0780
1.0	0.260	0.4075	0.4295	0.2250	0.2785	0.4060	0.4325	0.1505
1.5	0.442	0.6295	0.6535	0.3435	0.3920	0.6520	0.6755	0.2480
2.0	0.639	0.7895	0.8055	0.3935	0.4665	0.8575	0.8765	0.3850
2.5	0.804	0.9165	0.9215	0.4775	0.5425	0.9655	0.9695	0.6355
3.0	0.912	0.9640	0.9695	0.5445	0.6090	0.9910	0.9935	0.8720

Table 2 Empirical power and size of moving average model with Gamma innovation. $p = 600$, $n = 100$, $\alpha = 0.05$. TP is the theoretical asymptotic power, RM represents the randomization method, AM represents the asymptotic method.

		Dense means				Sparse means		
		$k = 3$		$k = 500$		$k = 3$		k
SNR	TP	RM	AM	RM	AM	RM	AM	RM
0.0	0.050	0.0450	0.0550	0.0475	0.0660	0.0405	0.0465	0.0505
0.5	0.126	0.1815	0.1975	0.1365	0.1750	0.1765	0.1870	0.0985
1.0	0.260	0.3825	0.4050	0.2375	0.2765	0.4130	0.4335	0.1550
1.5	0.442	0.6210	0.6465	0.2975	0.3490	0.6580	0.6745	0.2225
2.0	0.639	0.8180	0.8325	0.3920	0.4450	0.8645	0.8800	0.3890
2.5	0.804	0.9115	0.9260	0.4900	0.5465	0.9635	0.9665	0.6280
3.0	0.912	0.9710	0.9765	0.5505	0.6085	0.9940	0.9945	0.8600

Table 3 Empirical power and size of factor model innovation. $p = 600$, $n = 100$, $\alpha = 0.05$. TP is the theoretical asymptotic power, RM represents the randomization method, AM represents the asymptotic method.

		Dense means				Sparse means		
		Case I		Case II		Case I		C
SNR	TP	RM	AM	RM	AM	RM	AM	RM
0.0	0.050	0.0465	0.0610	0.0455	0.0590	0.0475	0.0625	0.0505
0.5	0.126	0.1315	0.1555	0.1465	0.1650	0.1200	0.1380	0.1080
1.0	0.260	0.2420	0.2780	0.2550	0.2780	0.1940	0.2250	0.2075
1.5	0.442	0.3635	0.3975	0.3555	0.3870	0.3670	0.4110	0.3740
2.0	0.639	0.4825	0.5165	0.4720	0.4975	0.5340	0.5930	0.5615
2.5	0.804	0.5860	0.6190	0.5825	0.6165	0.7040	0.7610	0.7120
3.0	0.912	0.6730	0.7060	0.6975	0.7210	0.8525	0.8815	0.8680

special statistic. In fact, the algorithm and the proof method can be applied to other quadratic based statistics.

We showed that even if the symmetric assumption is violated, the randomization test also has correct level asymptotically. Hence the test procedure is robust.

In classical statistics, randomization test procedure is time consuming. Nevertheless, our algorithm shows that the computational complexity of randomization test procedure is not affected by dimension. The randomization test can be even more efficient than asymptotic method while holding desirable statistic properties. Hence we have reason to believe that randomization tests may be generally suitable for high dimensional problems.

Maybe the most widely used randomization method is the two sample permutation test. As Romano (1990) pointed out, the asymptotic property of randomization tests depends heavily on the particular problem and the two sample case is quite distinct from the one sample case. The method used in this paper can not be applied to permutation test. We leave it for possible future work.

Appendix

A CLT for quadratic form of Rademacher variables The proof of the Theorem 1 is based on a CLT of the quadratic form of Rademacher variables. Such a CLT can be also used to study the asymptotic behavior of many other randomization test. Let $\varepsilon_1,\ldots,\varepsilon_n$ be independent Rademacher variables. Consider the quadratic form $W_n=\sum_{1\le j<i\le n}a_{ij}\varepsilon_i\varepsilon_j$, where $\{a_{ij}\}$ are nonrandom numbers. Here $\{\varepsilon_i\}$ and $\{a_{ij}\}$ may depend on n, a parameter we suppress. By direct calculation, we have $\mathrm{E}(W_n)=0$ and $\mathrm{Var}(W_n)=\sum_{1\le j<i\le n}a_{ij}^2$.

Proposition 1 *A sufficient condition for*

$$\frac{W_n}{\sqrt{\sum_{1\le j<i\le n}a_{ij}^2}}\xrightarrow{\mathscr{L}}N(0,1)$$

is that

$$\sum_{j<k}\big(\sum_{i:i>k}a_{ij}a_{ik}\big)^2+\sum_{j<i}a_{ij}^4+\sum_{j<k<i}a_{ij}^2a_{ik}^2=o\big((\sum_{j<i}a_{ij}^2)^2\big).$$

Proof Note that we have the decomposition $W_n=\sum_{i=2}^n U_{in}$, where $U_{in}=\varepsilon_i\sum_{j=1}^{i-1}a_{ij}\varepsilon_j$, $i=2,\ldots,n$. Let $\mathscr{F}_{in}$ be the σ-field generated by $\varepsilon_1,\ldots,\varepsilon_i$, $i=1,\ldots,n$. Then it can be seen that $\{U_{in}\}_{i=1}^n$ is a martingale difference array with respect to $\{\mathscr{F}_{in}\}_{i=1}^n$. Hence the martingale central limit theorem can be used. See, for example, Pollard (1984, Theorem 1 of Chapter VIII). By martingale central limit theorem, our conclusion holds if the following two conditions are satisfied:

$$\frac{\sum_{i=2}^n\mathrm{E}(U_{in}^2|\mathscr{F}_{i-1,n})}{\sum_{1\le j<i\le n}a_{ij}^2}\xrightarrow{P}1, \tag{13}$$

and

$$\frac{\sum_{i=2}^n\mathrm{E}\big(U_{in}^2\{U_{in}^2>\varepsilon\sum_{1\le j<i\le n}a_{ij}^2\}\big|\mathscr{F}_{i-1,n}\big)}{\sum_{1\le j<i\le n}a_{ij}^2}\xrightarrow{P}0, \tag{14}$$

for every $\varepsilon>0$.

Proof of (13) Since $\mathrm{E}(U_{in}^2|\mathscr{F}_{i-1,n})=\big(\sum_{j=1}^{i-1}a_{ij}\varepsilon_j\big)^2$, we have that

$$\begin{aligned}&\sum_{i=2}^n\mathrm{E}(U_{in}^2|\mathscr{F}_{i-1,n})=\sum_{i=2}^n\big(\sum_{j=1}^{i-1}a_{ij}\varepsilon_j\big)^2\\&=\sum_{i=2}^n\big(\sum_{j=1}^{i-1}a_{ij}^2+2\sum_{j,k:j<k<i}a_{ij}a_{ik}\varepsilon_j\varepsilon_k\big)\\&=\sum_{i=2}^n\sum_{j=1}^{i-1}a_{ij}^2+2\sum_{j<k<i}a_{ij}a_{ik}\varepsilon_j\varepsilon_k.\end{aligned}$$

For the second term, we have that

$$\begin{aligned}&\mathrm{E}\big(\sum_{j<k<i}a_{ij}a_{ik}\varepsilon_j\varepsilon_k\big)^2=\mathrm{E}\big(\sum_{j<k}\big(\sum_{i:i>k}a_{ij}a_{ik}\big)\varepsilon_j\varepsilon_k\big)^2\\&=\sum_{j<k}\big(\sum_{i:i>k}a_{ij}a_{ik}\big)^2=o\big((\sum_{j<i}a_{ij}^2)^2\big),\end{aligned}$$

where the last equality holds by assumption. Then it follows that

$$\frac{\sum_{j<k<i}a_{ij}a_{ik}\varepsilon_j\varepsilon_k}{\sum_{j<i}a_{ij}^2}\xrightarrow{P}0.$$

Hence (13) holds.

Proof of (14) By Markov inequality, it's sufficient to prove

$$\frac{\sum_{i=2}^n\mathrm{E}\big(U_{in}^4|\mathscr{F}_{i-1,n}\big)}{\big(\sum_{1\le j<i\le n}a_{ij}^2\big)^2}\xrightarrow{P}0. \tag{15}$$

Since the relevant random variables are all positive, we only need to prove (15) converges to 0 in mean. But

$$\begin{aligned}&\sum_{i=2}^n\mathrm{E}U_{in}^4=\sum_{i=2}^n\mathrm{E}\big(\sum_{j:j<i}a_{ij}\varepsilon_j\big)^4\\&=\sum_{i=2}^n\mathrm{E}\big(\sum_{j:j<i}a_{ij}^2+2\sum_{j,k:j<k<i}a_{ij}a_{ik}\varepsilon_j\varepsilon_k\big)^2\\&=\sum_{i=2}^n\Big(\big(\sum_{j:j<i}a_{ij}^2\big)^2+4\mathrm{E}\big(\sum_{j,k:j<k<i}a_{ij}a_{ik}\varepsilon_j\varepsilon_k\big)^2\Big)\\&=\sum_{i=2}^n\big(\sum_{j:j<i}a_{ij}^4+6\sum_{j,k:j<k<i}a_{ij}^2a_{ik}^2\big)\\&=\sum_{j<i}a_{ij}^4+6\sum_{j<k<i}a_{ij}^2a_{ik}^2=o\big((\sum_{j<i}a_{ij}^2)^2\big),\end{aligned}$$

where the last equality holds by assumption. Hence (14) holds.

The rest of Appendix is devoted to the proof of our main results.

Lemma 3 *Suppose $\{\eta_n\}_{n=1}^\infty$ is a sequence of random variables, weakly converges to η, a random variable with continuous distribution function. Then we have*

$$\sup_x|\Pr(\eta_n\le x)-\Pr(\eta\le x)|\to 0.$$

For two non-negative sequences $\{x_n\}_{n=1}^{\infty}$ and $\{y_n\}_{n=1}^{\infty}$, we write $a_n \asymp b_n$ to denote

$$cb_n \le a_n \le Cb_n$$

for some absolute constants $c>0, C>0$ and all $n=1,2,\ldots$.

Lemma 4 *Under* (6)*, suppose $A=(a_{ij})$ is an $m\times m$ positive semi-definite matrix, we have*

$$\mathrm{E}\left(Z_i^T A Z_i\right)^2 \asymp (\mathrm{tr} A)^2.$$

Proof *Notice that*

$$\begin{aligned}
&\left(Z_i^T A Z_i\right)^2 = (\sum_{j=1}^m a_{jj} z_{ij}^2 + 2\sum_{k<j} a_{jk} z_{ij} z_{ik})^2 \\
=&(\sum_{j=1}^m a_{jj} z_{ij}^2)^2 + 4(\sum_{j=1}^m a_{jj} z_{ij}^2)(\sum_{k<j} a_{jk} z_{ij} z_{ik}) \\
&+4(\sum_{k<j} a_{jk} z_{ij} z_{ik})^2 \\
=&\sum_{j=1}^m a_{jj}^2 z_{ij}^4 + 2\sum_{k<j} a_{jj} a_{kk} z_{ij}^2 z_{ik}^2 + 4(\sum_{j=1}^m a_{jj} z_{ij}^2)(\sum_{k<j} a_{jk} z_{ij} z_{ik}) \\
&+4\Big(\sum_{k<j} a_{jk}^2 z_{ij}^2 z_{ik}^2 \\
&+\sum_{k<j, l<\alpha : \mathrm{card}(\{k,j\}\cap\{l,\alpha\})<2} a_{jk} a_{\alpha l} z_{ij} z_{ik} z_{i\alpha} z_{il}\Big),
\end{aligned}$$

where $\mathrm{card}(\cdot)$ *is the cardinality of a set. By the assumption* (6)*, we have*

$$\begin{aligned}
&\mathrm{E}\left(Z_i^T A Z_i\right)^2 \\
=&\sum_{j=1}^n a_{jj}^2 \mathrm{E} z_{ij}^4 + 2\sum_{k<j} a_{jj} a_{kk} \mathrm{E}(z_{ij}^2 z_{ik}^2) + 4\sum_{k<j} a_{jk}^2 \mathrm{E}(z_{ij}^2 z_{ik}^2) \\
\asymp&\sum_{j=1}^n \sum_{k=1}^n a_{jj} a_{kk} + \sum_{j=1}^n \sum_{k=1}^n a_{jk}^2 = (\mathrm{tr}(A))^2 + \mathrm{tr}(A^2).
\end{aligned}$$

Then the conclusion holds from inequality

$$\mathrm{tr}(A^2) \le \lambda_1(A)\,\mathrm{tr}(A) \le (\mathrm{tr} A)^2.$$

Lemma 5 *Under* (5) *and* (6)*, for $i\ne j$ we have*

$$\mathrm{E}\left(X_i^T X_j\right)^4 = O(1)\Big(\mathrm{tr}\left(\Sigma+\mu\mu^T\right)^2\Big)^2. \tag{16}$$

Proof Under (5) and (6), we have

$$\begin{aligned}
&\left(X_i^T X_j\right)^4 = \left(Z_i^T \Gamma^T \Gamma Z_j + \mu^T \Gamma Z_i + \mu^T \Gamma Z_j + \mu^T \mu\right)^4 \\
\le&64\left((Z_i^T \Gamma^T \Gamma Z_j)^4 + (\mu^T \Gamma Z_i)^4 + (\mu^T \Gamma Z_j)^4 + (\mu^T \mu)^4\right)
\end{aligned}$$

We can deal with the first term by applying Lemma 4 twice:

$$\begin{aligned}
&\mathrm{E}(Z_i^T \Gamma^T \Gamma Z_j)^4 = \mathrm{E}(Z_i^T \Gamma^T \Gamma Z_j Z_j^T \Gamma^T \Gamma Z_i)^2 \\
=&\mathrm{E}\,\mathrm{E}\left((Z_i^T \Gamma^T \Gamma Z_j Z_j^T \Gamma^T \Gamma Z_i)^2 | Z_j\right) \\
\asymp&\mathrm{E}\left(Z_j^T \Gamma^T \Sigma \Gamma Z_j\right)^2 \asymp \left(\mathrm{tr}(\Sigma^2)\right)^2.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&\mathrm{E}(\mu^T \Gamma Z_i)^4 = \mathrm{E}(Z_i^T \Gamma^T \mu\mu^T \Gamma Z_i)^2 \asymp \left(\mu^T \Sigma \mu\right)^2 \\
\le&\lambda_1^2(\Sigma)(\mu^T\mu)^2 \le \mathrm{tr}(\Sigma^2)(\mu^T\mu)^2 \le \left(\mathrm{tr}(\Sigma^2)\right)^2 + (\mu^T\mu)^4.
\end{aligned}$$

Hence

$$\mathrm{E}(X_i^T X_j)^4 = O(1)\Big(\left(\mathrm{tr}(\Sigma^2)\right)^2 + (\mu^T\mu)^4\Big).$$

Then the theorem follows by noting that

$$\Big(\mathrm{tr}\left(\Sigma+\mu\mu^T\right)^2\Big)^2 \asymp \left(\mathrm{tr}(\Sigma^2)\right)^2 + (\mu^T\mu)^4.$$

Lemma 6 *Under* (5), (6)*, suppose $i\ne j$, $i\ne k$, $j\ne k$, we have*

$$\mathrm{E}\left(X_i^T X_j\right)^2 \left(X_k^T X_i\right)^2 = O(1)\Big(\mathrm{tr}\left(\Sigma+\mu\mu^T\right)^2\Big)^2. \tag{17}$$

Proof Note that

$$\begin{aligned}
&\mathrm{E}\left(X_i^T X_j\right)^2 \left(X_k^T X_i\right)^2 = \mathrm{E}\,\mathrm{E}\left(\left(X_i^T X_j\right)^2 \left(X_k^T X_i\right)^2 | X_i\right) \\
=&\mathrm{E}\left(X_i^T(\Sigma+\mu\mu^T)X_i\right)^2 \\
=&\mathrm{E}\Big(Z_i^T \Gamma^T(\Sigma+\mu\mu^T)\Gamma Z_i + 2\mu^T(\Sigma+\mu\mu^T)\Gamma Z_i \\
&+\mu^T\Sigma\mu + (\mu^T\mu)^2\Big)^2 \\
\le&4\mathrm{E}(Z_i^T \Gamma^T(\Sigma+\mu\mu^T)\Gamma Z_i)^2 + 16\mathrm{E}(\mu^T(\Sigma+\mu\mu^T)\Gamma Z_i)^2 \\
&+4(\mu\Sigma\mu)^2 + 4(\mu^T\mu)^4.
\end{aligned}$$

By Lemma (4), we have

$$\begin{aligned}
&\mathrm{E}(Z_i^T \Gamma^T(\Sigma+\mu\mu^T)\Gamma Z_i)^2 \\
\asymp&\left(\mathrm{tr}(\Gamma^T(\Sigma+\mu\mu^T)\Gamma)\right)^2 \\
=&\left(\mathrm{tr}\,\Sigma^2 + \mu^T\Sigma\mu\right)^2 \\
\le&2\left(\mathrm{tr}\,\Sigma^2\right)^2 + 2\left(\mu^T\Sigma\mu\right)^2.
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
&\mathrm{E}(\mu^T(\Sigma+\mu\mu^T)\Gamma Z_i)^2 \\
=&\mu^T(\Sigma+\mu\mu^T)\Sigma(\Sigma+\mu\mu^T)\mu \\
=&\mu^T\Sigma^3\mu + 2(\mu^T\mu)(\mu^T\Sigma^2\mu) + (\mu^T\mu)^2(\mu^T\Sigma\mu).
\end{aligned}$$

For $i=1,2,3$, we have

$$\mu^T\Sigma^i\mu \le \lambda_1^i(\Sigma)\mu^T\mu \le \left(\mathrm{tr}(\Sigma^2)\right)^{i/2}\mu^T\mu.$$

Combining these yields

$$\begin{aligned}
&\mathrm{E}\left(X_i^T X_j\right)^2 \left(X_k^T X_i\right)^2 \\
=&O(1)\Big(\left(\mathrm{tr}(\Sigma^2)\right)^2 + \left(\mathrm{tr}(\Sigma^2)\right)^{3/2}\mu^T\mu + \left(\mathrm{tr}(\Sigma^2)\right)(\mu^T\mu)^2 \\
&+\left(\mathrm{tr}(\Sigma^2)\right)^{1/2}(\mu^T\mu)^3 + (\mu^T\mu)^4\Big) \\
=&O(1)\Big(\left(\mathrm{tr}(\Sigma^2)\right)^2 + (\mu^T\mu)^4\Big) \\
=&O(1)\Big(\mathrm{tr}\left(\Sigma+\mu\mu^T\right)^2\Big)^2.
\end{aligned}$$

Lemma 7 *Under* (5) *and* (6)*, we have*

$$\frac{\sum_{j<i}(X_i^TX_j)^2}{\frac{n(n-1)}{2}\operatorname{tr}(\Sigma+\mu\mu^T)^2}\xrightarrow{P}1.$$

Proof Since

$$\begin{aligned}\mathrm{E}(X_i^TX_j)^2=&\mathrm{E}(X_i^TX_jX_j^TX_i)=\mathrm{E}(X_i^T(\Sigma+\mu\mu^T)X_i)\\=&\mathrm{E}\operatorname{tr}((\Sigma+\mu\mu^T)X_iX_i^T)=\operatorname{tr}(\Sigma+\mu\mu^T)^2,\end{aligned}$$

we have

$$\mathrm{E}\sum_{j<i}(X_i^TX_j)^2=\frac{n(n-1)}{2}\operatorname{tr}(\Sigma+\mu\mu^T)^2.$$

Next we need to deal with $\mathrm{E}\big(\sum_{j<i}(X_i^TX_j)^2\big)^2$. Write

$$\big(\sum_{j<i}(X_i^TX_j)^2\big)^2=\big(\sum_{j<i}(X_i^TX_j)^2\big)\big(\sum_{k<l}(X_l^TX_k)^2\big)$$

According to $\operatorname{card}(\{i,j\}\cap\{k,l\})=0,1,2$, we have

$$\begin{aligned}\big(\sum_{j<i}(X_i^TX_j)^2\big)^2=&\sum_{j<i}(X_i^TX_j)^4\\&+\sum_{j<i,k<l:\{i,j\}\cap\{k,l\}=\phi}(X_i^TX_j)^2(X_l^TX_k)^2\\+2\sum_{j<i<k}&\Big((X_i^TX_j)^2(X_k^TX_i)^2+(X_i^TX_j)^2(X_k^TX_j)^2\\&+(X_k^TX_j)^2(X_k^TX_i)^2\Big).\end{aligned}$$

There are $n(n-1)/2$, $n(n-1)(n-2)(n-3)/4$ and $n(n-1)(n-2)/6$ terms in each sum, respectively. This, combined with Lemma 5 and Lemma 6, yields

$$\begin{aligned}&\mathrm{E}\big(\sum_{j<i}(X_i^TX_j)^2\big)^2\\=&\frac{n(n-1)(n-2)(n-3)}{4}\big(\operatorname{tr}(\Sigma+\mu\mu^T)^2\big)^2\\&+O(1)(\frac{n(n-1)}{2}+n(n-1)(n-2))\big(\operatorname{tr}(\Sigma+\mu\mu^T)^2\big)^2.\end{aligned}$$

Hence we have

$$\begin{aligned}&\frac{\operatorname{Var}(\sum_{j<i}(X_i^TX_j)^2)}{\big(\mathrm{E}\sum_{j<i}(X_i^TX_j)^2\big)^2}\\=&\frac{\mathrm{E}\big(\sum_{j<i}(X_i^TX_j)^2\big)^2-\big(\mathrm{E}\sum_{j<i}(X_i^TX_j)^2\big)^2}{\big(\mathrm{E}\sum_{j<i}(X_i^TX_j)^2\big)^2}=O(\frac{1}{n}).\end{aligned}$$

This implies

$$\frac{\sum_{j<i}(X_i^TX_j)^2}{\mathrm{E}\sum_{j<i}(X_i^TX_j)^2}\xrightarrow{P}1.$$

The proof is complete.

Lemma 8 *Under* (5), (6), (7) *and* (10)*, we have*

$$\sum_{j<k}\big(\sum_{i:i>k}X_i^TX_jX_i^TX_k\big)^2=o_P\Big(\big(\frac{n(n-1)}{2}\operatorname{tr}(\Sigma+\mu\mu^T)^2\big)^2\Big).\tag{18}$$

$$\sum_{j<k}(X_i^TX_j)^4=o_P\Big(\big(\frac{n(n-1)}{2}\operatorname{tr}(\Sigma+\mu\mu^T)^2\big)^2\Big)\tag{19}$$

$$\sum_{j<k<i}(X_i^TX_j)^2(X_i^TX_k)^2=o_P\Big(\big(\frac{n(n-1)}{2}\operatorname{tr}(\Sigma+\mu\mu^T)^2\big)^2\Big)\tag{20}$$

Proof We have

$$\begin{aligned}&\mathrm{E}\sum_{j<k}\big(\sum_{i:i>k}X_i^TX_jX_i^TX_k\big)^2\\=&\mathrm{E}\sum_{j<k}\Big(\sum_{i:i>k}(X_i^TX_j)^2(X_i^TX_k)^2\\&+2\sum_{i_1,i_2:i_1>i_2>k}X_{i_1}^TX_jX_{i_1}^TX_kX_{i_2}^TX_jX_{i_2}^TX_k\Big)\\=&\mathrm{E}\sum_{j<k<i}(X_i^TX_j)^2(X_i^TX_k)^2\\&+2\mathrm{E}\sum_{j<k<i_2<i_1}X_{i_1}^TX_jX_{i_1}^TX_kX_{i_2}^TX_jX_{i_2}^TX_k.\end{aligned}$$

By Lemma 5, we have

$$\mathrm{E}\sum_{j<k<i}(X_i^TX_j)^2(X_i^TX_k)^2=O(n^3)\big(\operatorname{tr}(\Sigma+\mu\mu^T)^2\big)^2.$$

And

$$\begin{aligned}&\mathrm{E}\sum_{j<k<i_2<i_1}X_{i_1}^TX_jX_{i_1}^TX_kX_{i_2}^TX_jX_{i_2}^TX_k\\=&\frac{n(n-1)(n-2)(n-3)}{6}\operatorname{tr}(\Sigma+\mu\mu^T)^4\\\leq&\frac{n(n-1)(n-2)(n-3)}{6}8(\operatorname{tr}(\Sigma^4)+(\mu^T\mu)^4)\\\leq&O(n^4)(\lambda_1^2(\Sigma)\operatorname{tr}(\Sigma^2)+(\mu^T\mu)^4)\\=&o\Big(n^4\big(\operatorname{tr}(\Sigma^2)\big)^2\Big),\end{aligned}$$

where the last line follows by assumption (7) and (10). This proves (18). And (19) and (20) follow by Lemma 5 and Lemma 6, respectively.

Proof of Theorem 1

Proof By a standard subsequence argument, we only need to prove

$$\rho\Big(\mathscr{L}\Big(\frac{T_2(\varepsilon_1X_1,\ldots,\varepsilon_iX_i,\ldots,\varepsilon_nX_n)}{\sqrt{\sum_{1\leq j<i\leq n}(X_i^TX_j)^2}}\Big|X_1,\ldots,X_n\Big),N(0,1)\Big)\xrightarrow{a.s.}0$$

$$(21)$$

along a subsequence. But there exists a subsequence $\{n(k)\}$ along which (18), (19) and (20) holds almost surely. By Proposition 1, we have

$$\mathscr{L}\Big(\frac{T_2(\varepsilon_1 X_1,\ldots,\varepsilon_i X_i,\ldots,\varepsilon_n X_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^T X_j)^2}}\Big|X_1,\ldots,X_n\Big)\xrightarrow{\mathscr{L}} N(0,1)$$

almost surely along $\{n(k)\}$, which means that (21) holds along $\{n(k)\}$.

Proof of Theorem 2

Proof

$$\begin{aligned}
&\sum_{j<i} X_i^T X_j \varepsilon_i \varepsilon_j\\
=&\sum_{j<i} Z_i^T \Gamma^T \Gamma Z_j \varepsilon_i\varepsilon_j\\
&+\sum_{j<i}\mu^T\Gamma Z_i\varepsilon_i\varepsilon_j+\sum_{j<i}\mu^T\Gamma Z_j\varepsilon_i\varepsilon_j+\mu^T\mu\sum_{j<i}\varepsilon_i\varepsilon_j\\
\stackrel{def}{=}&C_1+C_2+C_3+C_4.
\end{aligned}$$

Term C_4 plays a major role. Note that

$$C_4=\frac{n}{2}\mu^T\mu\Big(\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n\varepsilon_i\Big)^2-1\Big).$$

By central limit theorem, we have

$$\rho\Big(\mathscr{L}\Big(\frac{C_4}{\frac{n}{2}\mu^T\mu}\Big|X_1,\ldots,X_n\Big),\chi_1^2-1\Big)\xrightarrow{a.s.}0.$$

Next we show that C_1, C_2 and C_3 are negligible under the assumptions of the theorem. By a standard subsequence argument and Slutsky's theorem, we can obtain

$$\rho\Big(\mathscr{L}\Big(\frac{T(\varepsilon_1 X_1,\ldots,\varepsilon_i X_i,\ldots,\varepsilon_n X_n)}{\frac{n}{2}\mu^T\mu}\Big|X_1,\ldots,X_n\Big),\chi_1^2-1\Big)\xrightarrow{P}0 \qquad (22)$$

by showing that

$$\mathrm{E}\Big(\Big(\frac{C_i}{\frac{n}{2}\mu^T\mu}\Big)^2\Big|X_1,\ldots,X_n\Big)\xrightarrow{P}0,\quad i=1,2,3.$$

It in turn suffices to show

$$\mathrm{E}\Big(\frac{C_i}{\frac{n}{2}\mu^T\mu}\Big)^2\to 0,\quad i=1,2,3. \qquad (23)$$

By direct calculation, we have

$$\mathrm{E}(C_1^2)=\mathrm{E}\,\mathrm{E}(C_1^2|X_1,\ldots,X_n)=\sum_{j<i}\mathrm{E}\,(Z_i^T\Gamma^T\Gamma Z_j)^2=\frac{n(n-1)}{2}\operatorname{tr}\Sigma^2,$$

and

$$\mathrm{E}(C_2^2)=\mathrm{E}(C_3^2)=\frac{n(n-1)}{2}\mu^T\Sigma\mu\le\frac{n(n-1)}{2}\sqrt{\operatorname{tr}\Sigma^2}\mu^T\mu.$$

Thus (23) follows by Assumption (11). Having (22) holds, the theorem follows by Slutsky's theorem, Lemma 7 and Assumption (11).

Proof (***Proof of Corollaries 1 and 2***) For every subsequence, there is a further subsequence along which

$$\rho\Big(\mathscr{L}\Big(\frac{T_2(\varepsilon_1 X_1,\ldots,\varepsilon_i X_i,\ldots,\varepsilon_n X_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^T X_j)^2}}\Big|X_1,\ldots,X_n\Big),N(0,1)\Big)\to 0$$

almost surely. By the property of convergence in law, $\xi_\alpha^*\to\Phi^{-1}(1-\alpha)$ almost surely along this subsequence. That is, For every subsequence, there is a further subsequence along which $\xi_\alpha^*\to\Phi^{-1}(1-\alpha)$ almost surely. This is equivalent to $\xi_\alpha^*\xrightarrow{P}\Phi^{-1}(1-\alpha)$. The proof of Corollary 2 is similar.

Proof (*Proof of Theorem 3*) Note that

$$\begin{aligned}
&\Pr\Big(\frac{T(X_1,\ldots,X_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^T X_j)^2}}>\xi_\alpha^*\Big)\\
=&\Pr\Big(\frac{T(X_1,\ldots,X_n)-\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{\sum_{1\le j<i\le n}(X_i^T X_j)^2}}>\xi_\alpha^*-\frac{\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{\sum_{1\le j<i\le n}(X_i^T X_j)^2}}\Big)
\end{aligned} \qquad (24)$$

If (10) holds, by Lemma 7, we have

$$\frac{\sum_{1\le j<i\le n}(X_i^T X_j)^2}{\frac{n(n-1)}{2}\operatorname{tr}\Sigma^2}\xrightarrow{P}1.$$

By Corollary 1, we have $\xi_\alpha^*\xrightarrow{P}\Phi(1-\alpha)$. Thus,

$$\begin{aligned}
(24)=&\Pr\Big(\frac{T(X_1,\ldots,X_n)-\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{\frac{n(n-1)}{2}\operatorname{tr}\Sigma^2}}-\frac{\sqrt{\sum_{1\le j<i\le n}(X_i^T X_j)^2}}{\sqrt{\frac{n(n-1)}{2}\operatorname{tr}\Sigma^2}}\xi_\alpha^*>-\sqrt{n}\\
=&\Pr\Big(N(0,1)-\Phi(1-\alpha)>-\frac{\sqrt{n(n-1)}\mu^T\mu}{\sqrt{2\operatorname{tr}\Sigma^2}}\Big)+o(1)\\
=&\Phi(-\Phi(1-\alpha)+\frac{\sqrt{n(n-1)}\mu^T\mu}{\sqrt{2\operatorname{tr}\Sigma^2}})+o(1),
\end{aligned}$$

where the last two equality holds by Lemma 1, Slutsky's theorem and Lemma 3.

If (11) holds, by Lemma 7, we have

$$\frac{\sum_{1\le j<i\le n}(X_i^T X_j)^2}{\frac{n(n-1)}{2}(\mu^T\mu)^2}\xrightarrow{P}1.$$

Thus

$$
\begin{aligned}
&\frac{T(X_1,\dots,X_n)-\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}}\\
=&\frac{T(X_1,\dots,X_n)-\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{\frac{n(n-1)}{2}\operatorname{tr}\Sigma^2}}\frac{\sqrt{\frac{n(n-1)}{2}\operatorname{tr}\Sigma^2}}{\sqrt{\frac{n(n-1)}{2}(\mu^T\mu)^2}}\frac{\sqrt{\frac{n(n-1)}{2}(\mu^T\mu)^2}}{\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}}\xrightarrow{P}0.
\end{aligned}
$$

By Corollary 2, $\xi_\alpha^* \xrightarrow{P} \frac{\sqrt{2}}{2}\Big(\big(\Phi^{-1}(1-\frac{\alpha}{2})\big)^2-1\Big)$. And

$$
\frac{\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}}=\sqrt{\frac{n(n-1)}{2}}\frac{\sqrt{\frac{n(n-1)}{2}(\mu^T\mu)^2}}{\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}}\xrightarrow{P}+\infty.
$$

Then (24) $\to 1$.

Proof (Proof of Theorem 4) Note that

$$
\begin{aligned}
&\Pr\Big(\frac{T(X_1,\dots,X_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}}>\xi_\alpha^*\Big)\\
=&\Pr\Big(\frac{T(X_1,\dots,X_n)-\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{(n-1)^2n\mu^T\Sigma\mu}}>\frac{\sqrt{\sum_{1\le j<i\le n}(X_i^TX_j)^2}}{\sqrt{(n-1)^2n\mu^T\Sigma\mu}}\xi_\alpha^*-\frac{\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{(n-1)^2n\mu^T\Sigma\mu}}\Big).
\end{aligned}
\tag{25}
$$

If (10) holds, the theorem follows by Lemma 2 and the fact that if (10) holds, the coefficient of ξ_α^* in (25) tends to 0.

If (11) holds, the theorem follows by noting that

$$
(25)=\Pr\Big(\frac{T(X_1,\dots,X_n)-\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{(n-1)^2n\mu^T\Sigma\mu}}>-(1+o_P(1))\frac{\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{(n-1)^2n\mu^T\Sigma\mu}}\Big).
$$

References

T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, 3rd edition edition, 2003.

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

Muni S. Srivastava and Meng Du. A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis*, 99(3):386–402, mar 2008. doi: 10.1016/j.jmva.2006.11.002.

Song Xi Chen and Ying Li Qin. A two-sample test for high-dimensional data with applications to gene-set testing. *Annals of Statistics*, 38(2):808–835, 2010.

L. Wang, B. Peng, and R. Li. A high-dimensional nonparametric multivariate test for mean vector. *Journal of the American Statistical Association*, 110(512):00–00, 2015.

Robert JTibshirani Bradley Efron. On testing the significance of sets of genes. *The Annals of Applied Statistics*, 1(1):107–129, 2007.

R A Fisher. The design of experiments. 1936.

Joseph P. Romano. On the behavior of randomization tests without a group invariance assumption. *Journal of the American Statistical Association*, 85(411):686–692, 1990.

Li-Xing Zhu and Georg Neuhaus. Nonparametric monte carlo tests for multivariate distributions. *Biometrika*, 87(4):919–928, 2000.

Eun Yi Chung and Joseph P. Romano. Multivariate and multiple permutation tests. *Journal of Econometrics*, 193(1):76–91, 2016.

A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550, sep 2005. doi: 10.1073/pnas.0506580102. URL https://doi.org/10.1073/pnas.0506580102.

Hyoseok [illegible]o, Kipoong Kim, and Hokeun Sun. Multiple group testing procedures for analysis of high-dimensional genomic data. *Genomics & Informatics*, 14(4):187, 2016. doi: 10.5808/gi.2016.14.4.187. URL https://doi.org/10.5808/gi.2016.14.4.187.

Joseph P. Romano E. L. Lehmann. *Testing Statistical Hypotheses*. Springer New York, 2005. doi: 10.1007/0-387-27605-X.

Jianqing Fan, Peter Hall, and Qiwei Yao. To how many simultaneous hypothesis tests can normal, student's t or bootstrap calibration be applied? *Journal of the American Statistical Association*, 102 (480):1282–1288, 2007.

David Pollard. *Convergence of stochastic processes*. 1984.