A randomization test for mean vector in high dimension

Rui Wang[a], Xingzhong Xu[a,b,*]

[a] *School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081,China*
[b] *Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing 100081,China*

Abstract

The strength of randomization tests is that the size of the test is exact under certain symmetry assumption for distribution. 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 time 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 investigated 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: Randomization test, High dimension, Symmetry assumption, Asymptotic power function

---

[*] Corresponding author
Email address: xuxz@bit.edu.cn (Xingzhong Xu)

1. Introduction

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

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

A classical test statistic for hypothesis (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) removed S^{-1} from Hotelling's T^2 statistic and proposed a test statistic based on $\bar{X}^T \bar{X}$ which can be applied in high dimensional setting. Many subsequent papers relaxed the assumptions and generalized the idea of Bai and Saranadasa (1996). For example, Srivastava and Du (2008) proposed a test based on $\bar{X}^T [\mathrm{diag}(S)]^{-1} \bar{X}$, where $\mathrm{diag}(S)$ is a matrix with diagonal elements equal to that of S and off-diagonal elements equal to 0. Chen and Qin (2010) proposed a test based on U-statistic $\sum_{i \neq j} X_i^T X_j$. Wang et al. (2015) proposed a test based on $\sum_{i \neq j} \|X_i\|^{-1} \|X_j\|^{-1} X_i^T X_j$. All these high dimensional statistics can be written as generalized quadratic forms of data, see Jong (1987). And the asymptotic properties of these statistics are mostly derived by martingale central limit theorem (MCLT). 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 (see Bradley Efron (2007)), sample size n may be very small. In this case, the Type I error rates 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. Romano (1990) described a general construction of the randomization test. In high dimension setting, randomization test is widely used in applied statistics. See, for example, Bradley Efron (2007). It's strength is in that the resulting test procedure has exact level under mild condition. There are many papers give theoretical analysis for fixed p case. See, for example, Romano (1990), Zhu and Neuhaus (2000) and Chung and Romano (2016).

In this paper, we consider the following randomization method. Suppose $T(X_1, \ldots, X_n)$ is certain test statistic for hypothesis (1). Let $\epsilon_1, \ldots, \epsilon_n$ be i.i.d. Rademacher variables ($\Pr(\epsilon_i = 1) = \Pr(\epsilon_i = -1) = 1/2$) which are independent of the data. Denote by

$$\mathcal{L}\big(T(\epsilon_1 X_1, \ldots, \epsilon_i X_i, \ldots, \epsilon_n X_n) | X_1, \ldots, X_n\big) \tag{2}$$

the conditional distribution of $T(\epsilon_1 X_1, \ldots, \epsilon_i X_i, \ldots, \epsilon_n X_n)$ conditioning on $X_1, \ldots, X_n$. The null hypothesis is rejected when $T(X_1, \ldots, X_n)$ is greater than the $1 - \alpha$ quantile of the conditional distribution (2) and is accepted otherwise, where α is the test level and the $1 - \alpha$ quantile of a distribution function $F(\cdot)$ is defined as $\inf\{y : F(y) \geq 1 - \alpha\}$. For illustration, this paper only considers the test statistic

$$T(X_1, \ldots, X_n) = \sum_{j<i} X_i^T X_j. \tag{3}$$

Nevertheless, other quadratic form statistics can also be studied by similar method. In fixed p setting, it's well known that randomization test has much higher time complexity than asymptotic method, which historically hampered it's use. We give a fast implementation of the randomization procedure, the time complexity of which does not depend on p. When p is large, our method is even more efficient than asymptotic method. We 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 efficient and also has good statistical properties.

The rest of the paper is organized in the following way. In Section.

2. Randomization Test

Throughout the paper, $T(X_1,\dots,X_n)$ refers to (3). We consider the test function $\phi(X_1,\dots,X_n)$ which equals to 1 if $T(X_1,\dots,X_n)$ is greater than the $1-\alpha$ quantile of the conditional distribution (2) and $\phi(X_1,\dots,X_n)=0$ otherwise. The test procedure $\phi(X_1,\dots,X_n)$ can be equivalently implemented by p-value. Define

$$p(X_1,\dots,X_n)=\Pr(T(\epsilon_1X_1,\dots,\epsilon_nX_n)\geq T(X_1,\dots,X_n)|X_1,\dots,X_n). \quad (4)$$

Then our test procedure rejects the null hypothesis if $p(X_1,\dots,X_n)\leq\alpha$.

Since $T(X_1,\dots,X_n)$ equals to half of the Chen and Qin (2010)'s U-statistic $\sum_{i\neq j}X_i^TX_j$, the test procedure $\phi(X_1,\dots,X_n)$ is equivalent to the randomization version of Chen and Qin (2010)'s test procedure. Note that Bai and Saranadasa (1996)'s statistic $\bar{X}^T\bar{X}$ can be written as $n^{-2}\sum_{i=1}^n\sum_{j=1}^nX_i^TX_j$ and $\sum_{i=1}^nX_i^TX_i$ is invariance under randomization. Hence the test procedure $\phi(X_1,\dots,X_n)$ is also equivalent to the randomization version of Bai and Saranadasa (1996)'s test.

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

Conditioning on $X_1,\dots,X_n$, the randomized test statistic $T(\epsilon_1X_1,\dots,\epsilon_nX_n)$ is uniformly distributed on 2^n values. To compute the exact quantile of (2), one need 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 $\epsilon_1^*, \ldots, \epsilon_n^*$ and compute $T^* = T(\epsilon_1^* X_1, \ldots, \epsilon_n^* X_n)$. Repeat B times for a large B and we obtain $T_1^*, \ldots, T_B^*$. Let

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

The test is rejected if $\tilde{p}(X_1, \ldots, X_n) \leq \alpha$. This procedure can also control the test level. In fact, we have $\Pr(\tilde{p}(X_1, \ldots, X_n) \leq u) \leq u$ for all $0 \leq u \leq 1$. See E. L. Lehmann (2005), Page 636. Moreover, as $B \to \infty$, we have $\tilde{p}(X_1, \ldots, X_n) \xrightarrow{P} p(X_1, \ldots, X_n)$ by law of large numbers. Here we emphasis that the convergence rate of $\tilde{p}(X_1, \ldots, X_n)$ only relies on α. Hence the choice of B does not rely on n and p.

We next consider the computation issue of the randomization procedure. Note that

$$T(\epsilon_1 X_1, \ldots, \epsilon_n X_n) = \sum_{1 \leq j < i \leq n} X_i^T X_j \epsilon_i \epsilon_j.$$

The computation of $T(X_1, \ldots, X_n)$ has complexity $O(n^2 p)$. This itself is not negligible when n and p are large and we can't afford to naively repeat computing it for B times. Note that $X_i^T X_j$ $(1 \leq j < i \leq n)$ can be computed beforehand. Once we have obtained $X_i^T X_j$, the computation of T_i^* has only complexity $O(n^2)$. The total complexity is thus $O(n^2 p + n^2 B)$. When p is large compared with n, the randomization procedure only adds minor complexity. The randomization method doesn't need an estimator of variance, which is a must need in asymptotic method. A good variance estimator is complicated in form, see Chen and Qin (2010), which add much time complexity to asymptotic method. Hence the randomization method is very competitive compared to asymptotic method. This is different from low dimensional setting where randomization test is a lot slower than asymptotic method.

If we only care about the decision (reject or accept) and the p-value is not needed, the computation efficiency of randomization method can be further improved. In fact, the rejection region is $\tilde{p}(X_1, \ldots, X_n) \leq \alpha$ or

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

Note that the left hand side is a sum of non-negative vaues. We can reject the null once $\sum_{i=1}^{B_0}(1 - \mathbf{1}_{\{T_i^* \geq T_0\}}) \geq B - (B+1)\alpha + 1$ for some B_0. Similarly, the acception region is

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

we can accept the null once the sum of left hand side exceeds the right hand side. The algorithm 1 summarizes our previous argument.

Algorithm 1 Randomization Algorithm

Require: α, B

- **for** $1 \leq j < i \leq n$ **do**
  - $D_{ij} \leftarrow X_i^T X_j$
- **end for**
- $T_0 \leftarrow \sum_{1 \leq j < i \leq n} D_{ij}$
- Set $A \leftarrow 0$.
- **for** $i = 1$ to B **do**
  - Generate $\epsilon_1, \ldots, \epsilon_n$ according to $\Pr(\epsilon_i = 1) = \Pr(\epsilon_i = -1) = \frac{1}{2}$.
  - **if** $\sum_{1 \leq j < i \leq n} D_{ij}\epsilon_i\epsilon_j \geq T_0$ **then**
    - $A \leftarrow A + 1$
    - **if** $A > (B+1)\alpha - 1$ **then**
      - **return** Accept
    - **end if**
  - **else**
    - **if** $i - A \geq B - (B+1)\alpha + 1$ **then**
      - **return** Reject
    - **end if**
  - **end if**
- **end for**

---

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 $\Gamma_i \Gamma_i^T = \Sigma_i$ and $Z_1, \dots, Z_n$ are m-variate i.i.d. 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.

An important assumption in Chen and Qin (2010) is $\mathrm{tr}(\Sigma^4) = o\{\mathrm{tr}^2(\Sigma^2)\}$, which is equivalent to

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

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

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

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

we have

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

And

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

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

we have

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

We shall call the conditional distribution

$$\mathcal{L}\Big(\frac{T(\epsilon_1 X_1,\ldots,\epsilon_i X_i,\ldots,\epsilon_n X_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^T X_j)^2}}\Big| X_1,\ldots,X_n\Big)$$

the randomization distribution. Let ξ_α^* be the $1-\alpha$ quantile of the randomization distribution. Then the test function $\phi(X_1,\ldots,X_n)$ equals to 1 when

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

and equals to 0 otherwise. Since ξ_α^* relies on data, the rejection region is determined by not only $T(X_1,\ldots,X_n)$ but also randomization distribution.

To study the asymptotic property of ξ_α^*, we need to derive the asymptotic behavior of randomization distribution. Since the randomization distribution itself is random, we need to define in what sense the convergence is. Let F and G be two distribution function on $\mathbb{R}$ and the Levy metric ρ of F and G is defined as

$$\rho(F,G)=\inf\{\epsilon : F(x-\epsilon)-\epsilon \le G(x) \le F(x+\epsilon)+\epsilon \quad \text{for all} \quad x\}.$$

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

Theorem 3. *Under* (5), (6), (7) *and*

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

we have that

$$\rho\Big(\mathcal{L}\Big(\frac{T_2(\epsilon_1 X_1,\ldots,\epsilon_i X_i,\ldots,\epsilon_n X_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^T X_j)^2}}\Big| X_1,\ldots,X_n\Big), N(0,1)\Big) \xrightarrow{P} 0.$$

It's also a theoretical interest to understand the behavior of randomization distribution when condition (10) is not valid. We have the following asymptotic result:

Theorem 4. *Under* (5), (6), (7) *and*

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

we have

$$\rho\Big(\mathcal{L}\Big(\frac{T(\epsilon_1 X_1, \ldots, \epsilon_i X_i, \ldots, \epsilon_n X_n)}{\sqrt{\sum_{1\le j<i\le n}\big(X_i^T X_j\big)^2}}\Big| X_1, \ldots, X_n\Big), \frac{\sqrt{2}}{2}\big(\chi_1^2 - 1\big)\Big) \xrightarrow{P} 0.$$

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 3, we have*

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

Corollary 2. *Under the conditions of Theorem 4,*

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

Now we are ready to derive the asymptotic power of randomization test. Since the limit property of T is different under (8) and (9). The following two theorems give the power under (8) and (9), separately.

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

If (10) *holds,*

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

If (11) *holds,*

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

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

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

Remark 1. These theorems don't assume that the distribution of X_1 is symmetric under null. Hence Theorem 5 shows that the level of randomization test is robust against asymmetry.

Remark 2. 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).

The randomization test rejects the null when

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

The asymptotic method rejects the null when

$$T(X_1, \ldots, X_n) > \sqrt{\frac{n(n-1)}{2}\mathrm{tr}\Sigma^2}\,\Phi^{-1}(1-\alpha).$$

Note that the larger the reject region, the more powerful the test is. Thus we compare $\sqrt{\sum_{1\le j<i\le n} (X_i^T X_j)^2}\,\xi_\alpha^*$ and $\sqrt{\frac{n(n-1)}{2}\mathrm{tr}\hat{\Sigma}^2}\,\Phi^{-1}(1-\alpha)$. Suppose $\mathrm{tr}\hat{\Sigma}^2$ is a ratio consistent estimator of $\mathrm{tr}\Sigma^2$. By Lemma 5 in appendix,

$$\frac{\sqrt{\sum_{1\le j<i\le n} (X_i^T X_j)^2}\,\xi_\alpha^*}{\sqrt{\frac{n(n-1)}{2}\mathrm{tr}\hat{\Sigma}^2}\,\Phi^{-1}(1-\alpha)} = (1+o_P(1))\frac{\sqrt{\mathrm{tr}(\Sigma+\mu\mu^T)^2}\,\xi_\alpha^*}{\sqrt{\mathrm{tr}\Sigma^2}\,\Phi^{-1}(1-\alpha)},$$

which tends to 1 when (10) holds, and tends to ∞ when (11) holds. Hence randomization may loss some power.

4. Simulation Studies

In this section, we report the simulation performance of the randomization test in various setting. We compare randomization method with asymptotic method. Throughout, we set $B = 200$. Asymptotic method needs a ratio consistent estimator of $\mathrm{tr}\Sigma^2$. However, a good estimator is often complicated and thus time consuming. For simplicity, throughout the simulation the asymptotic method is implemented by using true $\mathrm{tr}\Sigma^2$, which will in principle perform better. The empirical power and size are computed based on 1000 simulations.

Let $c = \sqrt{n(n-1)}\mu^T\mu/\sqrt{2\text{tr}\Sigma^2}$ be the signal to noise ratio (SNR). The theoretic asymptotic power is an increasing function of SNR. Throughout the simulation, we will 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 consider two innovation structure: the moving average model and the factor model.

We first conduct the moving average model. In model (5), we set $m = p + k$ and $(\Gamma)_{ij} = \rho_{j-i}$ for $j = i, \ldots, i+k$ and $(\Gamma)_{ij} = 0$ otherwise. More precisely,

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

for $i = 1, \ldots, n$ and $j = 1, \ldots, p$. Here Z_{ij}'s are i.i.d. random variables with distribution F for $i = 1, \ldots, n$ and $j = 1, \ldots, p+k$. Like Chen and Qin (2010), we consider two different F. One is $N(0,1)$, and the other is $(\text{Gamma}(4,1) - 4)/2$. We set $k = 5$ and $k = p$. The ρ_i's are generated independently from $U(2,3)$ and kept fixed throughout the simulation. 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 justify 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 also shares similar power with asymptotic method.

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

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

		Dense means				Sparse means			
		$k = 3$		$k = 500$		$k = 3$		$k = 500$	
SNR	TP	RM	AM	RM	AM	RM	AM	RM	AM
0.0	0.050	0.040	0.006	0.042	0.065	0.048	0.015	0.047	0.059
0.5	0.126	0.190	0.069	0.125	0.169	0.146	0.063	0.081	0.111
1.0	0.260	0.429	0.235	0.218	0.262	0.361	0.175	0.145	0.193
1.5	0.442	0.629	0.425	0.314	0.384	0.633	0.413	0.229	0.313
2.0	0.639	0.795	0.632	0.399	0.458	0.817	0.652	0.415	0.524
2.5	0.804	0.923	0.828	0.469	0.529	0.960	0.861	0.645	0.839
3.0	0.912	0.966	0.919	0.582	0.647	0.998	0.975	0.819	0.999

generation mechanism is adopted in our next simulation study. We denote by $\{\epsilon_{ij}\}_{1\le i\le n,1\le j\le p}$ 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

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

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 $j = 1, \ldots, p$ and in case II a_i and b_i are generated independently from $U(0, 0.4)$ and $U(0, 0.2)$. Table 3 list the simulation results. Although the distribution is not symmetric, the results show that the level of randomization method is robust while asymptotic method suffers from level inflation.

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

		Dense means				Sparse means			
		$k = 3$		$k = 500$		$k = 3$		$k = 500$	
SNR	TP	RM	AM	RM	AM	RM	AM	RM	AM
0.0	0.050	0.050	0.016	0.043	0.060	0.033	0.005	0.050	0.069
0.5	0.126	0.211	0.077	0.128	0.160	0.159	0.062	0.065	0.098
1.0	0.260	0.410	0.228	0.255	0.308	0.386	0.159	0.151	0.188
1.5	0.442	0.609	0.418	0.318	0.375	0.643	0.410	0.231	0.321
2.0	0.639	0.799	0.634	0.392	0.449	0.875	0.685	0.373	0.501
2.5	0.804	0.912	0.809	0.496	0.550	0.961	0.858	0.589	0.828
3.0	0.912	0.963	0.908	0.543	0.596	0.985	0.967	0.843	1.000

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 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. It is interesting to investigate the robustness of randomization test in finite sample case.

In classical statistics, randomization procedure is time consuming. Nevertheless, our algorithm shows that the time complexity of randomization 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

Table 3: Empirical power and size of factor model innovation. $p = 600$, $n = 100$. 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		Case II	
SNR	TP	RM	AM	RM	AM	RM	AM	RM	AM
0.0	0.050	0.040	0.050	0.046	0.057	0.052	0.061	0.047	0.060
0.5	0.126	0.141	0.166	0.136	0.154	0.104	0.126	0.105	0.129
1.0	0.260	0.222	0.256	0.233	0.266	0.207	0.239	0.215	0.240
1.5	0.442	0.363	0.411	0.369	0.392	0.367	0.411	0.373	0.415
2.0	0.639	0.460	0.501	0.470	0.509	0.517	0.581	0.532	0.589
2.5	0.804	0.582	0.613	0.565	0.607	0.718	0.757	0.728	0.780
3.0	0.912	0.646	0.678	0.673	0.701	0.834	0.889	0.870	0.900

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.

6. Appendix

A CLT for quadratic form of Rademacher variables. The proof of the Theorem 3 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 $\epsilon_1, \ldots, \epsilon_n$ be independent Rademacher variables. Consider quadratic form $W_n = \sum_{1\leq j<i\leq n} a_{ij}\epsilon_i\epsilon_j$, where $\{a_{ij}\}$ are nonrandom numbers. Here $\{\epsilon_i\}$ and $\{a_{ij}\}$ may depend on n, a parameter we suppress. Obviously, $\mathrm{E}(W_n) = 0$ and $\mathrm{Var}(W_n) = \sum_{1\leq j<i\leq n} a_{ij}^2$.

Proposition 1. *A sufficient condition for*

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

is that

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

Proof. Define $U_{in} = \epsilon_i \sum_{j=1}^{i-1} a_{ij}\epsilon_j$, $i = 2,\dots,n$, and $\mathcal{F}_{in} = \sigma\{\epsilon_1,\dots,\epsilon_i\}$, $i = 1,\dots,n$. Now $W_n = \sum_{i=2}^n U_{in}$, $\{U_{in}\}$ is a martingale difference array with respect to $\{\mathcal{F}_{in}\}$. Then MCLT can be used. See, for example, Pollard (1984), Chapter VIII Theorem 1. To prove the proposition, we shall verify two conditions:

$$\frac{\sum_{i=2}^n \mathrm{E}(U_{in}^2|\mathcal{F}_{i-1,n})}{\sum_{1\le j<i\le n} a_{ij}^2} \xrightarrow{P} 1, \tag{12}$$

and

$$\frac{\sum_{i=2}^n \mathrm{E}\big(U_{in}^2\{U_{in}^2 > \epsilon\sum_{1\le j<i\le n} a_{ij}^2\}\big|\mathcal{F}_{i-1,n}\big)}{\sum_{1\le j<i\le n} a_{ij}^2} \xrightarrow{P} 0, \tag{13}$$

for every $\epsilon > 0$.

Proof of (12). Since $\mathrm{E}(U_{in}^2|\mathcal{F}_{i-1,n}) = (\sum_{j=1}^{i-1} a_{ij}\epsilon_j)^2$, we have

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

But

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

where the last equality holds by assumption. Hence (12) holds.

Proof of (13). By Markov's inequality, we only need to prove

$$\frac{\sum_{i=2}^{n} \mathrm{E}\big(U_{in}^4 \big| \mathcal{F}_{i-1,n}\big)}{\big(\sum_{1\le j<i\le n} a_{ij}^2\big)^2} \xrightarrow{P} 0. \tag{14}$$

Since the relevant random variables are all positive, we only need to prove (14) converges to 0 in mean. But

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

where the last equality holds by assumption. Hence (13) holds. □

The rest of Appendix is devoted to the proof of the theorems in the paper.

Lemma 1. *Suppose $\{\eta_n\}$ is a sequence of 1-dimensional random variables, weakly converges to η, a random variable with continuous distribution function. Then we have*

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

Lemma 2. *Suppose $A = (a_{ij})$ is an $m \times m$ positive semi-definite matrix. Under* (6)*, we have*

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

Proof. Notice that

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

Hence

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

By Cauchy inequality, $0 \le \mathrm{tr} A^2 \le (\mathrm{tr} A)^2$. The conclusion holds.

□

Lemma 3. *Under* (5), (6), *for* $i \neq j$ *we have*

$$
\mathrm{E}(X_i^T X_j)^4 = O(1)\Big(\mathrm{tr}\big(\Sigma + \mu\mu^T\big)^2\Big)^2. \tag{15}
$$

Proof.

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

We deal with the first term by applying Lemma 2 twice.

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

Similarly, we have

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

Thus, the conclusion holds. □

Lemma 4. *Under* (5), (6), *suppose* $i \neq j$, $i \neq k$, $j \neq k$, *we have*

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

Proof. Note that

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

By Lemma (2),

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

And

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

But for $i = 1, 2, \ldots$, we have

$$\mu^T\Sigma^i\mu \leq \lambda_{\max}^i(\Sigma)\mu^T\mu \leq (\mathrm{tr}(\Sigma^2))^{i/2}\mu^T\mu.$$

Thus,

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

□

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

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

Proof. Since

$$\begin{aligned}
\mathrm{E}(X_i^T X_j)^2 =& \mathrm{E}(X_i^T X_j X_j^T X_i) = \mathrm{E}(X_i^T(\Sigma + \mu\mu^T)X_i) \\
=& \mathrm{E}\mathrm{tr}((\Sigma + \mu\mu^T)X_i X_i^T) = \mathrm{tr}(\Sigma + \mu\mu^T)^2,
\end{aligned}$$

we have

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

So we only need to consider the variance. According to card$(\{i,j\}\cap\{k,l\}) = 0,1,2$, we have

$$\begin{aligned}\Big(\sum_{j<i}\left(X_i^T X_j\right)^2\Big)^2 &= \sum_{j<i}\left(X_i^T X_j\right)^4 + \sum_{j<i,k<l:\{i,j\}\cap\{k,l\}=\phi}\left(X_i^T X_j\right)^2\left(X_k^T X_l\right)^2 \\ &+ 2\sum_{j<i<k}\Big(\left(X_i^T X_j\right)^2\left(X_k^T X_i\right)^2 + \left(X_i^T X_j\right)^2\left(X_k^T X_j\right)^2 + \left(X_k^T X_j\right)^2\left(X_k^T X_i\right)^2\Big).\end{aligned} \tag{17}$$

In (17), there are $n(n-1)/2$, $n(n-1)(n-2)(n-3)/4$ and $n(n-1)(n-2)/6$ terms in each summation. By Lemma 3 and Lemma 4, we have

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

Hence

$$\frac{\mathrm{Var}\big(\sum_{j<i}\left(X_i^T X_j\right)^2\big)}{\big(\mathrm{E}\sum_{j<i}\left(X_i^T X_j\right)^2\big)^2} = \frac{\mathrm{E}\big(\sum_{j<i}\left(X_i^T X_j\right)^2\big)^2 - \big(\mathrm{E}\sum_{j<i}\left(X_i^T X_j\right)^2\big)^2}{\big(\mathrm{E}\sum_{j<i}\left(X_i^T X_j\right)^2\big)^2} = O(\frac{1}{n}).$$

Thus the conclusion holds. $\square$

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

$$\mu^T\mu = o(\sqrt{\mathrm{tr}\Sigma^2}), \tag{18}$$

we have

$$\sum_{j<k}\Big(\sum_{i:i>k} X_i^T X_j X_i^T X_k\Big)^2 = o_P\Big(\big(\frac{n(n-1)}{2}\mathrm{tr}(\Sigma+\mu\mu^T)^2\big)^2\Big). \tag{19}$$

$$\sum_{j<k}\left(X_i^T X_j\right)^4 = o_P\Big(\big(\frac{n(n-1)}{2}\mathrm{tr}(\Sigma+\mu\mu^T)^2\big)^2\Big) \tag{20}$$

$$\sum_{j<k<i}\left(X_i^T X_j\right)^2\left(X_i^T X_k\right)^2 = o_P\Big(\big(\frac{n(n-1)}{2}\mathrm{tr}(\Sigma+\mu\mu^T)^2\big)^2\Big) \tag{21}$$

Proof.

$$\begin{aligned}
&\mathrm{E}\sum_{j<k}\Big(\sum_{i:i>k} X_i^T X_j X_i^T X_k\Big)^2 \\
=&\mathrm{E}\sum_{j<k}\Big(\sum_{i:i>k}(X_i^T X_j)^2(X_i^T X_k)^2 + 2\sum_{i_1,i_2:i_1>i_2>k} X_{i_1}^T X_j X_{i_1}^T X_k X_{i_2}^T X_j X_{i_2}^T X_k\Big).
\end{aligned}$$

By Lemma 3, we have

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

And

$$\begin{aligned}
&\mathrm{E}\sum_{j<k<i_2<i_1} X_{i_1}^T X_j X_{i_1}^T X_k X_{i_2}^T X_j X_{i_2}^T X_k \\
=&\frac{n(n-1)(n-2)(n-3)}{6}\mathrm{tr}(\Sigma+\mu\mu^T)^4 \\
\le&\frac{n(n-1)(n-2)(n-3)}{6}8(\mathrm{tr}(\Sigma)^4+(\mu^T\mu)^4) \\
\le&O(n^4)(\lambda_{\max}^2(\Sigma)\mathrm{tr}(\Sigma)^2+(\mu^T\mu)^4) \\
=&o\Big(n^4\big(\mathrm{tr}\Sigma^2\big)^2\Big),
\end{aligned}$$

where the last line follows by assumption (7) and (18). Thus (19) holds. (20) and (21) follow by Lemma 3 and Lemma 4.

□

Proof of Theorem 3. By a standard subsequence argument, we only need to prove

$$\rho\Big(\mathcal{L}\Big(\frac{T_2(\epsilon_1 X_1,\ldots,\epsilon_i X_i,\ldots,\epsilon_n X_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^T X_j)^2}}\Big| X_1,\ldots,X_n\Big), N(0,1)\Big) \xrightarrow{a.s.} 0 \qquad (22)$$

along a subsequence. But there is a subsequence $\{n(k)\}$ along which (19), (20) and (21) holds almost surely. By Proposition 1, we have

$$\mathcal{L}\Big(\frac{T_2(\epsilon_1 X_1,\ldots,\epsilon_i X_i,\ldots,\epsilon_n X_n)}{\sqrt{\sum_{1\le j<i\le n}(X_i^T X_j)^2}}\Big| X_1,\ldots,X_n\Big) \xrightarrow{\mathcal{L}} N(0,1)$$

almost surely along $\{n(k)\}$, which means (22) holds along $\{n(k)\}$.

□

Proof of Theorem 4.

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

Term C_4 plays a major role. Note that

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

By central limit theorem, we have

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

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

$$\rho \Big(\mathcal{L} \Big(\frac{T(\epsilon_1 X_1, \ldots, \epsilon_i X_i, \ldots, \epsilon_n X_n)}{\frac{n}{2} \mu^T \mu} \Big| X_1, \ldots, X_n \Big), \chi_1^2 - 1 \Big) \xrightarrow{P} 0 \tag{23}$$

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

$$\mathrm{E} \Big(\frac{C_i}{\frac{n}{2} \mu^T \mu} \Big)^2 \to 0, \quad i = 1, 2, 3. \tag{24}$$

By direct calculation, we have

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

and

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

Thus (24) follows by Assumption (11). Having (23) holds, the theorem follows by Slutsky's theorem, Lemma 5 and Assumption (11). $\square$

Proof of Theorem 5. Note that

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

If (10) holds, by Lemma 5, we have

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

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

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

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

If (11) holds, by Lemma 5, we have

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

Thus

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

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

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

Then (25) $\to 1$.

$\square$

Proof of Theorem 6. Note that

$$
\begin{aligned}
&\Pr\Big(\frac{T(X_1,\ldots,X_n)}{\sqrt{\sum_{1\le j<i\le n}\left(X_i^TX_j\right)^2}}>\xi_\alpha^*\Big)\\
=&\Pr\Big(\frac{T(X_1,\ldots,X_n)-\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{(n-1)^2n\mu^T\Sigma\mu}}>\frac{\sqrt{\sum_{1\le j<i\le n}\left(X_i^TX_j\right)^2}}{\sqrt{(n-1)^2n\mu^T\Sigma\mu}}\xi_\alpha^*-\frac{\frac{n(n-1)}{2}\mu^T\mu}{\sqrt{(n-1)^2n\mu^T\Sigma\mu}}\Big).
\end{aligned}
\tag{26}
$$

If (10) holds, the theorem follows by Theorem 2 and the fact that if (10) holds, the coefficient of ξ_α^* in (26) tends to 0.

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

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

$\square$

References

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

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

Bradley Efron, R.J., 2007. On testing the significance of sets of genes. The Annals of Applied Statistics 1, 107–129.

Chen, S.X., Qin, Y.L., 2010. A two-sample test for high-dimensional data with applications to gene-set testing. Annals of Statistics 38, 808–835.

Chung, E.Y., Romano, J.P., 2016. Multivariate and multiple permutation tests. Journal of Econometrics 193, 76–91.

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

Fan, J., Hall, P., Yao, Q., 2007. To how many simultaneous hypothesis tests can normal, student's t or bootstrap calibration be applied? Journal of the American Statistical Association 102, 1282–1288.

Fisher, R.A., 1936. The design of experiments. .

Jong, P.D., 1987. A central limit theorem for generalized quadratic forms. Probability Theory & Related Fields 75, 261–277.

Pollard, D., 1984. Convergence of stochastic processes.

Romano, J.P., 1990. On the behavior of randomization tests without a group invariance assumption. Journal of the American Statistical Association 85, 686–692.

Srivastava, M.S., Du, M., 2008. A test for the mean vector with fewer observations than the dimension. Journal of Multivariate Analysis 99, 386–402. doi:10.1016/j.jmva.2006.11.002.

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

Zhu, L.X., Neuhaus, G., 2000. Nonparametric monte carlo tests for multivariate distributions. Biometrika 87, 919–928.