

Elsevier L^AT_EX template[☆]

Elsevier¹

Radarweg 29, Amsterdam

Elsevier Inc^{a,b}, Global Customer Service^{b,}*

^a1600 John F Kennedy Boulevard, Philadelphia

^b360 Park Avenue South, New York

Abstract

This template helps you to create a properly formatted L^AT_EX manuscript.

Keywords: `elsarticle.cls`, L^AT_EX, Elsevier, template

2010 MSC: 00-01, 99-00

1. Introduction

Consider i.i.d. random sample $X_1, \dots, X_n \in \mathbb{R}^p$ which has means $\mu = (\mu_1, \dots, \mu_p)^T$ and covariance matrix Σ . The one sample testing problem

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

has been extensively studied by many researchers. A classical test statistic is Hotelling's T^2 ,

$$T_1 = n\bar{X}^T S^{-1} \bar{X}$$

where \bar{X} 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. However, the high-dimensional case $d > n$ invalidates the Hotelling's test. [1] is the first paper modifying Hotelling's

[☆]Fully documented templates are available in the `elsarticle` package on CTAN.

*Corresponding author

Email address: `support@elsevier.com` (Global Customer Service)

URL: `www.elsevier.com` (Elsevier Inc)

¹Since 1880.

T^2 statistic for high dimensional situations. This seminal paper removed S^{-1} from T_1 and proposed a test statistic based on

$$T_2 = \bar{X}^T \bar{X}.$$

Many subsequent papers relaxed the assumptions and generalized the idea of [1]. For example, [2] proposed a test based on

$$T_3 = \bar{X}^T [\text{diag}(S)]^{-1} \bar{X},$$

where $\text{diag}(S)$ is a matrix with diagonal elements equal to that of S and off-diagonal elements equal to 0. [3] proposed a test based on

$$T_4 = \sum_{i \neq j} X_i^T X_j.$$

Feng long proposed a test based on.

Note that these works are mostly can be written as a generalized quadratic form of data, see. Their theoretical prove mostly rely on the application of
5 martingale central limit theorem (MCLT) which is suitable for quadratic form.

The critical value of existing high dimensional tests are often determined by asymptotical distribution. Hence the test level can not be guaranteed in finite sample case.

The idea of randomization test dates back to [4], which is a tool to determine
10 the critical value for a given test statistic. It's strength is in that the resulting test procedure has exact level under mild condition. [5] described a general construction of the randomization test. In this paper, we focus on a special randomization method for one sample mean testing problem, which we will introduce shortly. If randomization method is robust and has good power in
15 high dimension, it will be a better choice compared to asymptotic method. As far as we know, there's no existing theoretical work investigating the behavior of randomization test in high dimensional setting.

Suppose the distribution of X_i is symmetric about 0 under the null hypothesis. T is a certain test statistic. Let $\epsilon_1, \dots, \epsilon_n$ be i.i.d. Rademacher variables

($\Pr(\epsilon_i = 1) = \Pr(\epsilon_i = -1) = 1/2$) which are independent of the data. The conditional distribution

$$\mathcal{L}(T(\epsilon_1 X_1, \dots, \epsilon_i X_i, \dots, \epsilon_n X_n) | X_1, \dots, X_n) \quad (2)$$

is the uniform distribution on 2^n values. The critical value of the randomization test is defined as the $1 - \alpha$ quantile of the above distribution. More specifically, the test function equals to 1 or 0 if $T(X_1, \dots, X_n)$ is greater or not greater than the critical value. The resulting test is a level α test. By some refinement of the test function when $T(X_1, \dots, X_n)$ equals to the critical value, the test is exact, see [5]. Since such extreme case occurs with little probability 2^{-n} , such refinement is often dropped in practice, which losses only a little power.

Equivalently, the randomization test can be implemented by p -value. Define

$$p(X_1, \dots, X_n) = \Pr(T(\epsilon_1 X_1, \dots, \epsilon_n X_n) \geq T(X_1, \dots, X_n) | X_1, \dots, X_n).$$

The randomization test rejects the null hypothesis if $p(X_1, \dots, X_n) \leq \alpha$.

It's easy to see that the randomization version of Bai's test and Chen's test are both equivalent to the randomization test based on

$$T(X_1, \dots, X_n) = \sum_{j < i} X_i^T X_j, \quad (3)$$

the test statistic this paper adopts. Other quadratic based statistic can be studied by similar method.

Our results show that even if the null distribution is not symmetric, the randomization test is still asymptotically exact. The local asymptotic power is also given.

In low dimensional setting, it's well known that randomization test has much higher time complexity than asymptotic method, which historically hampered it's use. Surprisingly, in high dimension setting the randomization test can be implemented as efficiently as asymptotic method.

Maybe the most used randomization method is the two sample permutation test. As [5] 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 deal with permutation test.

40 2. Randomization test

We assume, like chen qin and bai, the following multivariate model:

$$X_i = \mu + \Gamma Z_i \quad \text{for } i = 1, \dots, n, \quad (4)$$

where Γ is a $p \times m$ matrix for some $m \geq p$ such that $\Gamma_i \Gamma_i^T = \Sigma_i$ and $\{Z_i\}_{i=1}^n$ are m -variate i.i.d. random vectors satisfying $E(Z_i) = 0$ and $\text{Var}(Z_i) = I_m$, the $m \times m$ identity matrix. Write $Z_i = (z_{i1}, \dots, z_{im})^T$, we assume $E(z_{ij}^4) = 3 + \Delta < \infty$ and

$$E(z_{il_1}^{\alpha_1} z_{il_2}^{\alpha_2} \dots z_{il_q}^{\alpha_q}) = E(z_{il_1}^{\alpha_1}) E(z_{il_2}^{\alpha_2}) \dots E(z_{il_q}^{\alpha_q}) \quad (5)$$

for a positive integer q such that $\sum_{l=1}^q \alpha_l \leq 8$ and $l_1 \neq l_2 \neq \dots \neq l_q$.

In the following, $T(X_1, \dots, X_n)$ will be specialized to (3).

An important assumption in Chen and Qin's paper is $\text{tr}(\Sigma^4) = o\{\text{tr}^2(\Sigma^2)\}$, which is equavelent to

$$\frac{\lambda_{\max}(\Sigma)}{\sqrt{\text{tr}\Sigma^2}} \rightarrow 0. \quad (6)$$

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

Theorem 1. *Under (4), (5), (6) and local alternatives*

$$\mu^T \Sigma \mu = o(n^{-1} \text{tr}(\Sigma)^2), \quad (7)$$

we have

$$\frac{T(X_1, \dots, X_n) - \frac{n(n-1)}{2} \mu^T \mu}{\sqrt{\frac{n(n-1)}{2} \text{tr}\Sigma^2}} \xrightarrow{\mathcal{L}} N(0, 1). \quad (8)$$

45 And

Theorem 2. *Under (4), (5), (6) and*

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

we have

$$\frac{T(X_1, \dots, X_n) - \frac{n(n-1)}{2} \mu^T \mu}{\sqrt{(n-1)^2 n \mu^T \Sigma \mu}} \xrightarrow{\mathcal{L}} N(0, 1). \quad (10)$$

We will call the conditional distribution

$$\mathcal{L}\left(\frac{T(\epsilon_1 X_1, \dots, \epsilon_i X_i, \dots, \epsilon_n X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} \middle| X_1, \dots, X_n\right)$$

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

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

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

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

It's well known that $\rho(F_n, F) \rightarrow 0$ if and only if $F_n \xrightarrow{\mathcal{L}} F$.

Our asymptotic results need a condition stronger than (9).

Theorem 3. *Under (4), (5), (6) and local alternatives*

$$\mu^T \mu = o(\sqrt{\text{tr}(\Sigma)^2}), \quad (11)$$

we have that

$$\rho\left(\mathcal{L}\left(\frac{T_2(\epsilon_1 X_1, \dots, \epsilon_i X_i, \dots, \epsilon_n X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} \middle| X_1, \dots, X_n\right), N(0, 1)\right) \xrightarrow{P} 0$$

⁵⁰ **Corollary 4.** *Under the condition of Theorem 3, we have $\xi_\alpha^* \xrightarrow{P} \Phi(1 - \alpha)$, where $\Phi(\cdot)$ is the CDF of standard normal distribution.*

Remark 1. The proof of this theorem relies on Proposition 1, a central limit theorem (CLT).

Remark 2. Condition (11) is stronger than (9), which suggests the power function of randomization test may be a little bit smaller than non-randomization test.

We can derive the asymptotic power of randomization test.

Theorem 5. *Under the condition of Theorem 3 and Theorem 1, we have*

$$\Pr\left(\frac{T(X_1, \dots, X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} > \xi_\alpha^*\right) = \Phi(-\Phi(1 - \alpha) + \frac{\sqrt{n(n-1)}\mu^T \mu}{\sqrt{2\text{tr}\Sigma^2}}) + o(1),$$

Remark 3. The theorem doesn't assume the distribution of X_i is symmetric under null. Hence randomization test is robust in this sense.

Theorem 6. *Under the condition of Theorem 3 and Theorem 2, we have*

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

Remark 4. Chen's test power is

$$\Pr\left(\frac{T(X_1, \dots, X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} > \xi_\alpha^*\right) = \Phi(-\Phi(1 - \alpha) + \frac{\sqrt{n}\mu^T \mu}{2\sqrt{\mu^T \Sigma \mu}}) + o(1),$$

Hence our method is a bit more powerful.

3. Algorithm

The quantile ξ_α^* might not be computationally feasible, since the randomized test statistic is (conditionally) uniformly distributed on 2^n different values. In practice, randomization test is often realized through an approximation of p -value. More specifically, we can sample from randomization distribution by generate $\epsilon_1, \dots, \epsilon_n$ and compute $T(\epsilon_1 X_1, \dots, \epsilon_n X_n)$. Repeat M times for a large M and we obtain T_i^* , $i = 1, \dots, M$. Then $\sum_{i=1}^M \mathbf{1}_{\{T_i^* \geq T_0\}}/M$ is an approximation of the p -value

$$p(X_1, \dots, X_n) = \Pr(T_i^* \geq T_0 | X_1, \dots, X_n).$$

Hence we reject the null if $\sum_{i=1}^M \mathbf{1}_{\{T_i^* \geq T_0\}}/M \leq \alpha$.

To be more efficient, we can accept the null once the sum of ξ_i exceeds $M\alpha$ and reject the null once the sum of $1 - \xi_i$ reaches $M(1 - \alpha)$.

65 Note that 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 $O(n^2(p + M))$. Hence randomization test is very efficient in high dimension setting. For example, when $M = p$, time spent by randomization is at most twice of that of asymptotic method. This is different from low dimension setting where randomization is a
70 lot slower than asymptotic method.

Algorithm 1 Randomization Algorithm

Require: α, M

Set $A \leftarrow 0, T_0 \leftarrow T(X_1, \dots, X_n)$.

Compute $X_i^T X_j$ for $1 \leq j < i \leq n$

for $i = 1$ to M **do**

Generate $\epsilon_1, \dots, \epsilon_n$ and compute $T(\epsilon_1 X_1, \dots, \epsilon_n X_n) = \sum_{j < i} X_i^T X_j \epsilon_i \epsilon_j$.

if $T(\epsilon_1 X_1, \dots, \epsilon_n X_n) > T_0$ **then**

$A \leftarrow A + 1$

end if

if $A > M\alpha$ **then**

return Accept

end if

if $i - A \geq M(1 - \alpha)$ **then**

return Reject

end if

end for

4. Appendix

CLT for quadratic form of Rademacher variables. Our proof of the main results 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
 75 test. Let $\epsilon_1, \dots, \epsilon_n$ be indepent 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, $E(W_n) = 0$ and
 $\text{Var}(W_n) = \sum_{1 \leq j < i \leq n} a_{ij}^2$.

Proposition 1. *A sufficient condition for*

$$\frac{W_n}{\sqrt{\sum_{1 \leq j < i \leq n} a_{ij}^2}} \xrightarrow{\mathcal{L}} N(0, 1) \quad (12)$$

is that

$$\sum_{j < k} \left(\sum_{i: i > k} a_{ij} a_{ik} \right)^2 + \sum_{j < i} a_{ij}^4 + \sum_{j < k < i} a_{ij}^2 a_{ik}^2 = o\left(\left(\sum_{j < i} a_{ij}^2\right)^2\right). \quad (13)$$

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 $\{\mathcal{F}_{in}\}$. To prove the proposition, we shall verify two conditions (See David Pollard's book):

$$\frac{\sum_{i=2}^n E(U_{in}^2 | \mathcal{F}_{i-1, n})}{\sum_{1 \leq j < i \leq n} a_{ij}^2} \xrightarrow{P} 1, \quad (14)$$

and

$$\frac{\sum_{i=2}^n E(U_{in}^2 \{U_{in}^2 > \epsilon \sum_{1 \leq j < i \leq n} a_{ij}^2\} | \mathcal{F}_{i-1, n})}{\sum_{1 \leq j < i \leq n} a_{ij}^2} \xrightarrow{P} 0, \quad (15)$$

for every $\epsilon > 0$.

Proof of (14). Since $E(U_{in}^2 | \mathcal{F}_{i-1, n}) = \left(\sum_{j=1}^{i-1} a_{ij} \epsilon_j\right)^2$, we have

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

But

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

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

Proof of (15). By Markov's inequality, we only need to prove

$$\frac{\sum_{i=2}^n \mathbb{E}(U_{in}^4 | \mathcal{F}_{i-1, n})}{\left(\sum_{1 \leq j < i \leq n} a_{ij}^2\right)^2} \xrightarrow{P} 0. \quad (16)$$

Since the relevant random variables are all positive, we only need to prove (16)

converges to 0 in mean. But

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

where the last equality holds by assumption. Hence (15) holds. \square

5. Asymptotic normality

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

$$\sup_x |\Pr(\eta_n \leq x) - \Pr(\eta \leq x)| \rightarrow 0.$$

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

$$\mathbb{E}(Z_i^T A Z_i)^2 \asymp (\text{tr}(A))^2 + \text{tr} A^2 \quad (17)$$

Proof. Notice that

$$\begin{aligned} (Z_i^T A Z_i)^2 &= \left(\sum_{j=1}^m a_{jj} z_{ij}^2 + 2 \sum_{k < j} a_{jk} z_{ij} z_{ik} \right)^2 \\ &= \left(\sum_{j=1}^m a_{jj} z_{ij}^2 \right)^2 + 4 \left(\sum_{j=1}^m a_{jj} z_{ij}^2 \right) \left(\sum_{k < j} a_{jk} z_{ij} z_{ik} \right) + 4 \left(\sum_{k < j} a_{jk} z_{ij} z_{ik} \right)^2 \\ &= \sum_{j=1}^m a_{jj}^2 z_{ij}^4 + 2 \sum_{k < j} a_{jj} a_{kk} z_{ij}^2 z_{ik}^2 + 4 \left(\sum_{j=1}^m a_{jj} z_{ij}^2 \right) \left(\sum_{k < j} a_{jk} z_{ij} z_{ik} \right) \\ &\quad + 4 \left(\sum_{k < j} a_{jk}^2 z_{ij}^2 z_{ik}^2 + \sum_{k < j, l < \alpha: \text{card}(\{i, j\} \cap \{l, \alpha\}) < 2} a_{jk} a_{\alpha l} z_{ij} z_{ik} z_{i\alpha} z_{il} \right) \end{aligned} \quad (18)$$

Hence

$$\begin{aligned} \mathbb{E}(Z_i^T A Z_i)^2 &= \sum_{j=1}^n a_{jj}^2 \mathbb{E} z_{ij}^4 + 2 \sum_{k < j} a_{jj} a_{kk} \mathbb{E}(z_{ij}^2 z_{ik}^2) + 4 \sum_{k < j} a_{jk}^2 \mathbb{E}(z_{ij}^2 z_{ik}^2) \\ &\asymp \sum_{j=1}^n \sum_{k=1}^n a_{jj} a_{kk} + \sum_{j=1}^n \sum_{k=1}^n a_{jk}^2 = (\text{tr}(A))^2 + \text{tr} A^2 \end{aligned} \quad (19)$$

□

Lemma 3. Under (4), (5) and (6), we have

$$\frac{\sum_{j < i} (X_i^T X_j)^2}{\frac{n(n-1)}{2} \text{tr}(\Sigma + \mu \mu^T)^2} \xrightarrow{P} 1. \quad (20)$$

Proof.

$$\begin{aligned} \mathbb{E}(X_i^T X_j)^2 &= \mathbb{E}(X_i^T X_j X_j^T X_i) = \mathbb{E}(X_i^T (\Sigma + \mu \mu^T) X_i) \\ &= \mathbb{E} \text{tr}((\Sigma + \mu \mu^T) X_i X_i^T) = \text{tr}(\Sigma + \mu \mu^T)^2 \end{aligned} \quad (21)$$

Hence

$$\mathbb{E} \sum_{j < i} (X_i^T X_j)^2 = \frac{n(n-1)}{2} \text{tr}(\Sigma + \mu \mu^T)^2 \quad (22)$$

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

$$\begin{aligned} \left(\sum_{j < i} (X_i^T X_j)^2 \right)^2 &= \sum_{j < i} (X_i^T X_j)^4 + \sum_{j < i, k < l: \{i, j\} \cap \{k, l\} = \emptyset} (X_i^T X_j)^2 (X_k^T X_l)^2 \\ &+ 2 \sum_{j < i < k} ((X_i^T X_j)^2 (X_k^T X_i)^2 + (X_i^T X_j)^2 (X_k^T X_j)^2 + (X_k^T X_j)^2 (X_k^T X_i)^2) \end{aligned} \quad (23)$$

$$\begin{aligned} (X_i^T X_j)^4 &= (Z_i^T \Gamma^T \Gamma Z_j + \mu^T \Gamma Z_i + \mu^T \Gamma Z_j + \mu^T \mu)^4 \\ &\leq 64((Z_i^T \Gamma^T \Gamma Z_j)^4 + (\mu^T \Gamma Z_i)^4 + (\mu^T \Gamma Z_j)^4 + (\mu^T \mu)^4) \end{aligned} \quad (24)$$

Note that by lemma 2, we have

$$\mathbb{E}(\mu^T \Gamma Z_i)^4 = \mathbb{E}(Z_i^T \Gamma^T \mu \mu^T \Gamma Z_i)^2 \asymp (\mu^T \Sigma \mu)^2 \leq \lambda_{\max}^2(\Sigma) (\mu^T \mu)^2. \quad (25)$$

and apply lemma 2, we get

$$\begin{aligned} \mathbb{E}(Z_i^T \Gamma^T \Gamma Z_j)^4 &= \mathbb{E}(Z_i^T \Gamma^T \Gamma Z_j Z_j^T \Gamma^T \Gamma Z_i)^2 \\ &= \mathbb{E} \mathbb{E}((Z_i^T \Gamma^T \Gamma Z_j Z_j^T \Gamma^T \Gamma Z_i)^2 | Z_j) \\ &\asymp \mathbb{E}(Z_j^T \Gamma^T \Sigma \Gamma Z_j)^2 \\ &\asymp (\text{tr} \Sigma^2)^2 + \text{tr} \Sigma^4 \\ &\leq (\text{tr} \Sigma^2)^2 + \lambda_{\max}^2(\Sigma) \text{tr} \Sigma^2. \end{aligned} \quad (26)$$

By condition (??), we have

$$\mathbb{E}(X_i^T X_j)^4 = O\left((\text{tr}(\Sigma + \mu \mu^T))^2\right) \quad (27)$$

Similarly, we have

$$\begin{aligned}
& \mathbb{E}(X_i^T X_j)^2 (X_k^T X_i)^2 \\
&= \mathbb{E} \mathbb{E}((X_i^T X_j)^2 (X_k^T X_i)^2 | X_i) \\
&= \mathbb{E}(X_i^T (\Sigma + \mu \mu^T) X_i)^2 \\
&= \mathbb{E}(Z_i^T \Gamma^T (\Sigma + \mu \mu^T) \Gamma Z_i + 2\mu^T (\Sigma + \mu \mu^T) \Gamma Z_i + \mu \Sigma \mu + (\mu^T \mu)^2)^2 \\
&\leq 4\mathbb{E}(Z_i^T \Gamma^T (\Sigma + \mu \mu^T) \Gamma Z_i)^2 + 16\mathbb{E}(\mu^T (\Sigma + \mu \mu^T) \Gamma Z_i)^2 + 4(\mu \Sigma \mu)^2 + 4(\mu^T \mu)^4 \\
&\asymp (\text{tr}(\Gamma^T (\Sigma + \mu \mu^T) \Gamma))^2 + \text{tr}(\Gamma^T (\Sigma + \mu \mu^T) \Sigma (\Sigma + \mu \mu^T) \Gamma) \\
&\quad + \mu^T (\Sigma + \mu \mu^T) \Sigma (\Sigma + \mu \mu^T) \mu + (\mu \Sigma \mu)^2 + (\mu^T \mu)^4 \\
&= \text{tr} \Sigma^4 + (\mu^T \mu)^4 + (\text{tr} \Sigma^2)^2 + 2\text{tr} \Sigma^2 (\mu^T \Sigma \mu) + 3(\mu^T \Sigma \mu)^2 \\
&\quad + 3\mu^T \Sigma^3 \mu + 2(\mu^T \Sigma^2 \mu)(\mu^T \mu) + (\mu^T \Sigma \mu)(\mu^T \mu)^2 \\
&= O(1)((\text{tr} \Sigma^2)^2 + (\mu^T \mu)^4) \\
&= O\left((\text{tr}(\Sigma + \mu \mu^T)^2)^2\right)
\end{aligned} \tag{28}$$

The last 2 equality follows by the fact $\mu^T \Sigma^i \mu \leq \lambda_{\max}^i(\Sigma) \mu^T \mu$ ($i = 1, 2, \dots$), the
85 condition (6) and the inequality $ab \leq a^p/p + b^q/p$ for $1/p + 1/q = 1$ and $p, q > 0$.

In (23), 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. Hence

$$\begin{aligned}
\mathbb{E}\left(\sum_{j < i} (X_i^T X_j)^2\right)^2 &= \frac{n(n-1)(n-2)(n-3)}{4} (\text{tr}(\Sigma + \mu \mu^T)^2)^2 \\
&\quad + O(1)\left(\frac{n(n-1)}{2} + n(n-1)(n-2)\right) (\text{tr}(\Sigma + \mu \mu^T)^2)^2
\end{aligned} \tag{29}$$

Hence $\text{Var}(\sum_{j < i} (X_i^T X_j)^2) = o\left((\mathbb{E}(\sum_{j < i} (X_i^T X_j)^2))^2\right)$, which prove the lemma. \square

Lemma 4. Under (4), (5), (6) and local alternatives

$$\mu^T \mu = o(\sqrt{\text{tr} \Sigma^2}), \tag{30}$$

we have

$$\sum_{j < k} \left(\sum_{i: i > k} X_i^T X_j X_i^T X_k \right)^2 = o_P\left(\left(\frac{n(n-1)}{2} \text{tr}(\Sigma + \mu \mu^T)^2\right)^2\right) \tag{31}$$

$$\sum_{j < k} (X_i^T X_j)^4 = o_P \left(\left(\frac{n(n-1)}{2} \text{tr}(\Sigma + \mu\mu^T)^2 \right)^2 \right) \quad (32)$$

$$\sum_{j < k < i} (X_i^T X_j)^2 (X_i^T X_k)^2 = o_P \left(\left(\frac{n(n-1)}{2} \text{tr}(\Sigma + \mu\mu^T)^2 \right)^2 \right) \quad (33)$$

Proof.

$$\begin{aligned} & \mathbb{E} \sum_{j < k} \left(\sum_{i: i > k} X_i^T X_j X_i^T X_k \right)^2 \\ &= \mathbb{E} \sum_{j < k} \left(\sum_{i: i > k} (X_i^T X_j)^2 (X_i^T X_k)^2 + 2 \sum_{i_1, i_2: i_1 > i_2 > k} X_{i_1}^T X_j X_{i_1}^T X_k X_{i_2}^T X_j X_{i_2}^T X_k \right). \end{aligned} \quad (34)$$

By (28), we have

$$\mathbb{E} \sum_{j < k < i} (X_i^T X_j)^2 (X_i^T X_k)^2 = O(n^3) (\text{tr}(\Sigma + \mu\mu^T)^2)^2 \quad (35)$$

But

$$\begin{aligned} & \mathbb{E} \sum_{j < k < i_2 < i_1} X_{i_1}^T X_j X_{i_1}^T X_k X_{i_2}^T X_j X_{i_2}^T X_k \\ &= \frac{n(n-1)(n-2)(n-3)}{6} \text{tr}(\Sigma + \mu\mu^T)^4 \\ &\leq \frac{n(n-1)(n-2)(n-3)}{6} 8(\text{tr}(\Sigma)^4 + (\mu^T \mu)^4) \\ &\leq O(n^4) (\lambda_{\max}^2(\Sigma) \text{tr}(\Sigma)^2 + (\mu^T \mu)^4) \\ &= o \left(n^4 (\text{tr} \Sigma^2)^2 \right). \end{aligned} \quad (36)$$

It follows that (31) holds. (32) and (33) can be proved similarly by (28) and (27). \square

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

$$\rho \left(\mathcal{L} \left(\frac{T_2(\epsilon_1 X_1, \dots, \epsilon_i X_i, \dots, \epsilon_n X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} \middle| X_1, \dots, X_n \right), N(0, 1) \right) \xrightarrow{a.s.} 0 \quad (37)$$

along a subsequence. But there is a subsequence $\{n(k)\}$ along which (31), (32) and (33) holds almost surely. By Proposition 1, we have

$$\mathcal{L} \left(\frac{T_2(\epsilon_1 X_1, \dots, \epsilon_i X_i, \dots, \epsilon_n X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} \middle| X_1, \dots, X_n \right) \xrightarrow{\mathcal{L}} N(0, 1) \quad (38)$$

almost surely along $\{n(k)\}$, which means (37) holds along $\{n(k)\}$. □

Proof of Theorem 5. Similar to the proof of Chen's theorem and by Lemma 3, we have

$$\frac{T_2(X_1, \dots, X_n) - \frac{n(n-1)}{2} \mu^T \mu}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} \xrightarrow{\mathcal{L}} N(0, 1). \quad (39)$$

By Corollary 4, we have $\xi_\alpha^* \xrightarrow{P} \Phi(1 - \alpha)$. Now the asymptotic power function can be derived by Slutsky's theorem

$$\begin{aligned} & \Pr \left(\frac{T_2(X_1, \dots, X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} > \xi_\alpha^* \right) \\ &= \Pr \left(\frac{T_2(X_1, \dots, X_n) - \frac{n(n-1)}{2} \mu^T \mu}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} > \xi_\alpha^* - \frac{\frac{n(n-1)}{2} \mu^T \mu}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} \right) \\ &= \Pr \left(\frac{T_2(X_1, \dots, X_n) - \frac{n(n-1)}{2} \mu^T \mu}{\sqrt{\frac{n(n-1)}{2} \text{tr} \Sigma^2}} - \frac{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}}{\sqrt{\frac{n(n-1)}{2} \text{tr} \Sigma^2}} \xi_\alpha^* > -\frac{\sqrt{n(n-1)} \mu^T \mu}{\sqrt{2 \text{tr} \Sigma^2}} \right) \\ &= \Pr \left(N(0, 1) - \Phi(1 - \alpha) > -\frac{\sqrt{n(n-1)} \mu^T \mu}{\sqrt{2 \text{tr} \Sigma^2}} \right) + o(1) \\ &= \Phi(-\Phi(1 - \alpha) + \frac{\sqrt{n(n-1)} \mu^T \mu}{\sqrt{2 \text{tr} \Sigma^2}}) + o(1), \end{aligned} \quad (40)$$

where the last two equality holds by Lemma 1. □

References

- [1] Z. Bai, H. Saranadasa, Effect of high dimension: by an example of a two sample problem, *Statistica Sinica* 6 (2) (1996) 311–329.
- [2] M. S. Srivastava, A test for the mean vector with fewer observations than the dimension under non-normality, *Journal of Multivariate Analysis* 100 (3) (2009) 518–532. As the access to this document is restricted, you may want to look for a different version under "Related research" (further below) or for a different version of it.

- [3] S. X. Chen, Y. L. Qin, A two-sample test for high-dimensional data with applications to gene-set testing, *Annals of Statistics* 38 (2) (2010) 808–835.
- [4] R. A. Fisher, *The design of experiments*.
- 105 [5] J. P. Romano, On the behavior of randomization tests without a group invariance assumption, *Journal of the American Statistical Association* 85 (411) (1990) 686–692.