

a Randomization Test for Mean Vector In High Dimension

Elsevier¹

Radarweg 29, Amsterdam

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

^a*1600 John F Kennedy Boulevard, Philadelphia*

^b*360 Park Avenue South, New York*

Abstract

Testing mean vector is a fundamental problem in statistics. Many test procedures for high dimensional mean vector have been proposed in the literature.

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.

5 However, the high-dimensional case $d > n$ invalidates the Hotelling's test. [1] is the first paper modifying Hotelling's T^2 statistic for testing (1) in high dimensional setting. This seminal paper removed S^{-1} from T_1 and proposed a test

^{*}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.

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
10 $T_3 = \bar{X}^T [\text{diag}(S)]^{-1} \bar{X}$, 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
 $T_4 = \sum_{i \neq j} X_i^T X_j$. [4] proposed a test based on $T_5 = \sum_{i \neq j} Y_i^T Y_j$, where Y_i is defined as $Y_i = X_i / \|X_i\|$ if $X_i \neq 0$; and $Y_i = 0$ if $X_i = 0$.

Note that the test statistics T_2 – T_5 all can be written as a generalized quadratic
15 form of data, see [5] and their theoretical proves mostly rely on an application of martingale central limit theorem (MCLT).

The critical value of existing high dimensional tests are often determined by asymptotical distribution. We will call it asymptotic method. Asymptotic method can guarantee the test level asymptotically. However, in many real
20 world problems, e.g., gene testing (see [6]), sample size n may be very small. Hence the Type I error rates of Asymptotic method may be far away from nominal levels in this case.

The idea of randomization test dates back to [7], which is a tool to determine the critical value for a given test statistic. It's strength is in that the resulting
25 test procedure has exact level under mild condition. [8] described a general construction of the randomization test. In high dimension setting, randomization test is widely used in applied statistics, see [6]. However, theoretical analysis mainly focus on fixed p case. To the best of our knowledge, there's no existing theoretical work investigating the behavior of randomization test in high
30 dimensional setting. In this paper, we focus on a special randomization method for one sample mean testing problem, which we will introduce shortly. If the randomization method is robust and has good power in high dimension, it may be a better choice compared to asymptotic method.

Suppose the distribution of X_i is symmetric about 0 under the null hypothesis. T is certain test statistic for testing (1). 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
 35 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 if the distribution
 of X_i is symmetric under null. By some refinement of the test function when
 $T(X_1, \dots, X_n)$ equals to the critical value, the test is exact, see [8]. Since such
 40 extreme case occurs with little probability 2^{-n} , the refinement is often dropped
 in practice, which only losses 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 [1]'s test and [3]'s test are
 both equivalent to the randomization test based on

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

In this paper, we will study this special statistic for illustration. From now on,
 $T(X_1, \dots, X_n)$ will refer to (3). Other quadratic based statistic can be studied
 45 by similar method.

Our results show that even if the null distribution is not symmetric, the
 randomization test is still asymptotically exact under mild assumptions. We
 also give the local asymptotic power.

In fixed p setting, it's well known that randomization test has much higher
 50 time complexity than asymptotic method, which historically hampered it's use.
 Surprisingly, in high dimension setting the randomization test can be imple-
 mented as efficiently as asymptotic method.

Maybe the most widely used randomization method is the two sample per-
 mutation test. As [8] pointed out, the asymptotic property of randomization

55 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.

2. Randomization test

We assume, like [3] and [1], 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 $l_1 \neq l_2 \neq \dots \neq l_q$.

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

An important assumption in [3] is $\text{tr}(\Sigma^4) = o\{\text{tr}^2(\Sigma^2)\}$, which is equivalent 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

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

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 $\phi(X_1, \dots, X_n)$ of randomization test 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. 65

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 first result is:

Theorem 3. Under (4), (5), (6) and

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

we have that

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

Under Theorem 3's conditions, the randomization distribution is asymptotically a normal distribution. It's also interesting to understand the behavior of randomization distribution when condition (11) is not valid. We have the following asymptotic result: 70

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

$$\sqrt{\text{tr}\Sigma^2} = o(\mu^T \mu), \quad (12)$$

we have

$$\rho\left(\mathcal{L}\left(\frac{T(\epsilon_1 X_1, \dots, \epsilon_i X_i, \dots, \epsilon_n X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} \middle| X_1, \dots, X_n\right), \frac{\sqrt{2}}{2}(\chi_1^2 - 1)\right) \xrightarrow{P} 0. \quad (13)$$

Once the asymptotic distribution 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, $\xi_\alpha^* \xrightarrow{P} \Phi(1 - \alpha)$.

Corollary 2. Under the conditions of Theorem 4, $\xi_\alpha^* \xrightarrow{P} \frac{\sqrt{2}}{2}(\Phi^{-1}(1 - \frac{\alpha}{2}) - 1)$.

Now we are ready to derive the asymptotic power of randomization test. Since the limit property of T is different under (7) and (9). The following two theorems give the power under (7) and (9), separately.

Theorem 5. Suppose conditions (4), (5), (6) and (7) holds.

If (11) holds,

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

If (12) holds,

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

Theorem 6. Under (4), (5), (6) (9) and either (11) or (12),

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

Remark 1. The theorem doesn't assume that the distribution of X_i is symmetric under null. Hence Theorem 5 shows that the level of randomization test is robust against asymmetry.

Remark 2. Neither (7) or (11) implies the other one. For example, suppose $\Sigma = I_p$, then (7) is equivalent to $\mu^T \mu = o(p/n)$ and (11) is equivalent to $\mu^T \mu = o(\sqrt{p})$. In this case, if $\sqrt{p}/n \rightarrow 0$, then (7) implies (11); conversely, if $\sqrt{p}/n \rightarrow \infty$, then (11) implies (7).

The randomization test rejects the null when

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

The asymptotic method rejects the null when

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

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

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

which tends to 1 when (11) holds, and tends to ∞ when (12) holds. Hence

randomization may loss some power.

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 $M(1 - \alpha)$.

95 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
100 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. 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

105 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*

$$\frac{W_n}{\sqrt{\sum_{1 \leq j < i \leq n} a_{ij}^2}} \xrightarrow{\mathcal{L}} N(0, 1) \quad (14)$$

is that

$$\sum_{j < k} \left(\sum_{i: i > k} a_{ij} a_{ik} \right)^2 + \sum_{j < i} a_{ij}^4 + \sum_{j < k < i} a_{ij}^2 a_{ik}^2 = o\left(\left(\sum_{j < i} a_{ij}^2\right)^2\right). \quad (15)$$

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 [9]):

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

and

$$\frac{\sum_{i=2}^n E(U_{in}^2 \{U_{in}^2 > \epsilon \sum_{1 \leq j < i \leq n} a_{ij}^2\} | \mathcal{F}_{i-1, n})}{\sum_{1 \leq j < i \leq n} a_{ij}^2} \xrightarrow{P} 0, \quad (17)$$

for every $\epsilon > 0$.

Proof of (16). 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

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

110 where the last equality holds by assumption. Hence (16) holds.

Proof of (17). By Markov's inequality, we only need to prove

$$\frac{\sum_{i=2}^n \mathbb{E}(U_{in}^4 | \mathcal{F}_{i-1,n})}{(\sum_{1 \leq j < i \leq n} a_{ij}^2)^2} \xrightarrow{P} 0. \quad (18)$$

Since the relevant random variables are all positive, we only need to prove (18) 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 (17) 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. Under (5), we have

$$\mathbb{E}(Z_i^T A Z_i)^2 \asymp (\text{tr} A)^2 \quad (19)$$

Proof. Notice that

$$\begin{aligned}
(Z_i^T A Z_i)^2 &= \left(\sum_{j=1}^m a_{jj} z_{ij}^2 + 2 \sum_{k < j} a_{jk} z_{ij} z_{ik} \right)^2 \\
&= \left(\sum_{j=1}^m a_{jj} z_{ij}^2 \right)^2 + 4 \left(\sum_{j=1}^m a_{jj} z_{ij}^2 \right) \left(\sum_{k < j} a_{jk} z_{ij} z_{ik} \right) + 4 \left(\sum_{k < j} a_{jk} z_{ij} z_{ik} \right)^2 \\
&= \sum_{j=1}^m a_{jj}^2 z_{ij}^4 + 2 \sum_{k < j} a_{jj} a_{kk} z_{ij}^2 z_{ik}^2 + 4 \left(\sum_{j=1}^m a_{jj} z_{ij}^2 \right) \left(\sum_{k < j} a_{jk} z_{ij} z_{ik} \right) \\
&\quad + 4 \left(\sum_{k < j} a_{jk}^2 z_{ij}^2 z_{ik}^2 + \sum_{k < j, l < \alpha: \text{card}(\{k, j\} \cap \{l, \alpha\}) < 2} a_{jk} a_{\alpha l} z_{ij} z_{ik} z_{i\alpha} z_{il} \right)
\end{aligned} \tag{20}$$

Hence

$$\begin{aligned}
\mathbb{E}(Z_i^T A Z_i)^2 &= \sum_{j=1}^n a_{jj}^2 \mathbb{E} z_{ij}^4 + 2 \sum_{k < j} a_{jj} a_{kk} \mathbb{E}(z_{ij}^2 z_{ik}^2) + 4 \sum_{k < j} a_{jk}^2 \mathbb{E}(z_{ij}^2 z_{ik}^2) \\
&\asymp \sum_{j=1}^n \sum_{k=1}^n a_{jj} a_{kk} + \sum_{j=1}^n \sum_{k=1}^n a_{jk}^2 = (\text{tr}(A))^2 + \text{tr} A^2.
\end{aligned} \tag{21}$$

By Cauchy inequality, $0 \leq \text{tr} A^2 \leq (\text{tr} A)^2$. The conclusion holds. \square

Lemma 3. Under (4), (5), for $i \neq j$ we have

$$\mathbb{E}(X_i^T X_j)^4 = O(1) \left(\text{tr}(\Sigma + \mu \mu^T)^2 \right)^2. \tag{22}$$

Proof.

$$\begin{aligned}
(X_i^T X_j)^4 &= (Z_i^T \Gamma^T \Gamma Z_j + \mu^T \Gamma Z_i + \mu^T \Gamma Z_j + \mu^T \mu)^4 \\
&\leq 64 \left((Z_i^T \Gamma^T \Gamma Z_j)^4 + (\mu^T \Gamma Z_i)^4 + (\mu^T \Gamma Z_j)^4 + (\mu^T \mu)^4 \right)
\end{aligned} \tag{23}$$

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

$$\begin{aligned}
\mathbb{E}(Z_i^T \Gamma^T \Gamma Z_j)^4 &= \mathbb{E}(Z_i^T \Gamma^T \Gamma Z_j Z_j^T \Gamma^T \Gamma Z_i)^2 = \mathbb{E} \mathbb{E}((Z_i^T \Gamma^T \Gamma Z_j Z_j^T \Gamma^T \Gamma Z_i)^2 | Z_j) \\
&\asymp \mathbb{E}(Z_j^T \Gamma^T \Sigma \Gamma Z_j)^2 \asymp (\text{tr} \Sigma^2)^2,
\end{aligned} \tag{24}$$

Similarly, we have

$$\begin{aligned}
\mathbb{E}(\mu^T \Gamma Z_i)^4 &= \mathbb{E}(Z_i^T \Gamma^T \mu \mu^T \Gamma Z_i)^2 \asymp (\mu^T \Sigma \mu)^2 \\
&\leq \lambda_{\max}^2(\Sigma) (\mu^T \mu)^2 \leq \text{tr}(\Sigma^2) (\mu^T \mu)^2 \leq (\text{tr}(\Sigma^2))^2 + (\mu^T \mu)^4.
\end{aligned} \tag{25}$$

115 Thus, the conclusion holds. \square

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

$$\mathbb{E}(X_i^T X_j)^2 (X_k^T X_i)^2 = O(1) \left(\text{tr}(\Sigma + \mu\mu^T)^2 \right)^2. \quad (26)$$

Proof. Note that

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

By Lemma (2),

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

And

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

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

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

Thus,

$$\begin{aligned} &\mathbb{E}(X_i^T X_j)^2 (X_k^T X_i)^2 \\ &= O(1) \left((\text{tr} \Sigma^2)^2 + (\text{tr} \Sigma^2)^{3/2} (\mu^T \mu) + (\text{tr} \Sigma^2) (\mu^T \mu)^2 + (\text{tr} \Sigma^2)^{1/2} (\mu^T \mu)^3 + (\mu^T \mu)^4 \right) \\ &= O(1) \left(\text{tr}(\Sigma + \mu\mu^T)^2 \right)^2. \end{aligned} \quad (27)$$

\square

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

$$\frac{\sum_{j < i} (X_i^T X_j)^2}{\frac{n(n-1)}{2} \text{tr}(\Sigma + \mu\mu^T)^2} \xrightarrow{P} 1. \quad (28)$$

Proof. Since

$$\begin{aligned} \mathbb{E}(X_i^T X_j)^2 &= \mathbb{E}(X_i^T X_j X_j^T X_i) = \mathbb{E}(X_i^T (\Sigma + \mu\mu^T) X_i) \\ &= \mathbb{E} \text{tr}((\Sigma + \mu\mu^T) X_i X_i^T) = \text{tr}(\Sigma + \mu\mu^T)^2, \end{aligned} \quad (29)$$

we have

$$\mathbb{E} \sum_{j < i} (X_i^T X_j)^2 = \frac{n(n-1)}{2} \text{tr}(\Sigma + \mu\mu^T)^2 \quad (30)$$

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

$$\begin{aligned} \left(\sum_{j < i} (X_i^T X_j)^2 \right)^2 &= \sum_{j < i} (X_i^T X_j)^4 + \sum_{j < i, k < l: \{i, j\} \cap \{k, l\} = \emptyset} (X_i^T X_j)^2 (X_k^T X_l)^2 \\ &+ 2 \sum_{j < i < k} ((X_i^T X_j)^2 (X_k^T X_i)^2 + (X_i^T X_j)^2 (X_k^T X_j)^2 + (X_k^T X_j)^2 (X_k^T X_i)^2). \end{aligned} \quad (31)$$

In (31), 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

$$\begin{aligned} \mathbb{E} \left(\sum_{j < i} (X_i^T X_j)^2 \right)^2 &= \frac{n(n-1)(n-2)(n-3)}{4} (\text{tr}(\Sigma + \mu\mu^T)^2)^2 \\ &+ O(1) \left(\frac{n(n-1)}{2} + n(n-1)(n-2) \right) (\text{tr}(\Sigma + \mu\mu^T)^2)^2. \end{aligned} \quad (32)$$

Hence

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

Thus the conclusion holds. \square

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

$$\mu^T \mu = o(\sqrt{\text{tr} \Sigma^2}), \quad (33)$$

we have

$$\sum_{j < k} \left(\sum_{i: i > k} X_i^T X_j X_i^T X_k \right)^2 = o_P \left(\left(\frac{n(n-1)}{2} \text{tr}(\Sigma + \mu\mu^T)^2 \right)^2 \right). \quad (34)$$

$$\sum_{j < k} (X_i^T X_j)^4 = o_P \left(\left(\frac{n(n-1)}{2} \text{tr}(\Sigma + \mu\mu^T)^2 \right)^2 \right) \quad (35)$$

$$\sum_{j < k < i} (X_i^T X_j)^2 (X_i^T X_k)^2 = o_P \left(\left(\frac{n(n-1)}{2} \text{tr}(\Sigma + \mu\mu^T)^2 \right)^2 \right) \quad (36)$$

Proof.

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

By Lemma 3, we have

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

And

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

where the last line follows by assumption (6) and (33). Thus (34) holds. (35)

and (36) follow by Lemma 3 and Lemma 4. \square

120

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 (34), (35)

and (36) holds almost surely. By Proposition 1, we have

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

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

□

Proof of Theorem 5. Note that

$$\begin{aligned} & \Pr\left(\frac{T(X_1, \dots, X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} > \xi_\alpha^*\right) \\ &= \Pr\left(\frac{T(X_1, \dots, X_n) - \frac{n(n-1)}{2} \mu^T \mu}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} > \xi_\alpha^* - \frac{\frac{n(n-1)}{2} \mu^T \mu}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}}\right) \end{aligned} \quad (38)$$

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

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

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

$$\begin{aligned} (38) &= \Pr\left(\frac{T(X_1, \dots, X_n) - \frac{n(n-1)}{2} \mu^T \mu}{\sqrt{\frac{n(n-1)}{2} \text{tr} \Sigma^2}} - \frac{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}}{\sqrt{\frac{n(n-1)}{2} \text{tr} \Sigma^2}} \xi_\alpha^* > -\frac{\sqrt{n(n-1)} \mu^T \mu}{\sqrt{2 \text{tr} \Sigma^2}}\right) \\ &= \Pr\left(N(0, 1) - \Phi(1 - \alpha) > -\frac{\sqrt{n(n-1)} \mu^T \mu}{\sqrt{2 \text{tr} \Sigma^2}}\right) + o(1) \\ &= \Phi(-\Phi(1 - \alpha) + \frac{\sqrt{n(n-1)} \mu^T \mu}{\sqrt{2 \text{tr} \Sigma^2}}) + o(1), \end{aligned}$$

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

If (12) holds, by Lemma 5, we have

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

Thus

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

125 Since $\xi_\alpha^* \xrightarrow{P} \frac{\sqrt{2}}{2} (\Phi^{-1}(1 - \frac{\alpha}{2}) - 1)$ by Corollary 2, we have that (38) $\rightarrow 1$.

□

Proof of Theorem 6. Note that

$$\begin{aligned}
& \Pr \left(\frac{T(X_1, \dots, X_n)}{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}} > \xi_\alpha^* \right) \\
&= \Pr \left(\frac{T(X_1, \dots, X_n) - \frac{n(n-1)}{2} \mu^T \mu}{\sqrt{(n-1)^2 n \mu^T \Sigma \mu}} > \frac{\sqrt{\sum_{1 \leq j < i \leq n} (X_i^T X_j)^2}}{\sqrt{(n-1)^2 n \mu^T \Sigma \mu}} \xi_\alpha^* - \frac{\frac{n(n-1)}{2} \mu^T \mu}{\sqrt{(n-1)^2 n \mu^T \Sigma \mu}} \right).
\end{aligned} \tag{39}$$

If (11) holds, the theorem follows by Theorem 2 and the fact that if (11) holds, the coefficient of ξ_α^* in (39) tends to 0.

If (12) holds, the theorem follows by noting that

$$(39) = \Pr \left(\frac{T(X_1, \dots, X_n) - \frac{n(n-1)}{2} \mu^T \mu}{\sqrt{(n-1)^2 n \mu^T \Sigma \mu}} > -(1 + o_P(1)) \frac{\frac{n(n-1)}{2} \mu^T \mu}{\sqrt{(n-1)^2 n \mu^T \Sigma \mu}} \right).$$

□

Proof of Theorem 4.

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

Term C_4 plays a major role. Note that

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

By central limit theorem, we have

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

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

$$\rho \left(\mathcal{L} \left(\frac{T(\epsilon_1 X_1, \dots, \epsilon_i X_i, \dots, \epsilon_n X_n)}{\frac{n}{2} \mu^T \mu} \middle| X_1, \dots, X_n \right), \chi_1^2 - 1 \right) \xrightarrow{P} 0 \tag{40}$$

by showing that

$$E\left(\left(\frac{C_i}{\frac{n}{2}\mu^T\mu}\right)^2 \middle| X_1, \dots, X_n\right) \xrightarrow{P} 0, \quad i = 1, 2, 3.$$

130 It in turn suffices to show

$$E\left(\frac{C_i}{\frac{n}{2}\mu^T\mu}\right)^2 \rightarrow 0, \quad i = 1, 2, 3. \quad (41)$$

By direct calculation, we have

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

and

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

Thus (41) follows by Assumption (12). Having (40) holds, the theorem follows by Slutsky's theorem, Lemma 5 and Assumption (12). \square

References

- [1] Z. Bai, H. Saranadasa, Effect of high dimension: by an example of a two
135 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) 518–532As the access to this document is restricted, you may want
to look for a different version under "Related research" (further below) orfor
140 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] L. Wang, B. Peng, R. Li, A high-dimensional nonparametric multivari-
ate test for mean vector., *Journal of the American Statistical Association*
145 110 (512) (2015) 00–00.

- [5] P. D. Jong, A central limit theorem for generalized quadratic forms, *Probability Theory & Related Fields* 75 (2) (1987) 261–277.
- [6] R. J. Bradley Efron, On testing the significance of sets of genes, *The Annals of Applied Statistics* 1 (1) (2007) 107–129.
- 150 [7] R. A. Fisher, The design of experiments.
- [8] 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.
- [9] D. Pollard, *Convergence of stochastic processes*, 1984.