



12-2013

# Optimal Hypothesis Testing for High Dimensional Covariance Matrices

Tony Cai  
*University of Pennsylvania*

Zongming Ma  
*University of Pennsylvania*

Follow this and additional works at: [http://repository.upenn.edu/statistics\\_papers](http://repository.upenn.edu/statistics_papers)

 Part of the [Physical Sciences and Mathematics Commons](#)

## Recommended Citation

Cai, T., & Ma, Z. (2013). Optimal Hypothesis Testing for High Dimensional Covariance Matrices. *Bernoulli*, 19 (5B), 2359-2388.  
<http://dx.doi.org/10.3150/12-BEJ455>

This paper is posted at ScholarlyCommons. [http://repository.upenn.edu/statistics\\_papers/45](http://repository.upenn.edu/statistics_papers/45)  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

# Optimal Hypothesis Testing for High Dimensional Covariance Matrices

## Abstract

This paper considers testing a covariance matrix  $\Sigma$  in the high dimensional setting where the dimension  $p$  can be comparable or much larger than the sample size  $n$ . The problem of testing the hypothesis  $H_0: \Sigma = \Sigma_0$  for a given covariance matrix  $\Sigma_0$  is studied from a minimax point of view. We first characterize the boundary that separates the testable region from the non-testable region by the Frobenius norm when the ratio between the dimension  $p$  over the sample size  $n$  is bounded. A test based on a  $U$ -statistic is introduced and is shown to be rate optimal over this asymptotic regime. Furthermore, it is shown that the power of this test uniformly dominates that of the corrected likelihood ratio test (CLRT) over the entire asymptotic regime under which the CLRT is applicable. The power of the  $U$ -statistic based test is also analyzed when  $p/n$  is unbounded.

## Keywords

correlation matrix, covariance matrix, high-dimensional data, likelihood ratio test, minimax hypothesis testing, power, testing covariance structure

## Disciplines

Physical Sciences and Mathematics

# Optimal hypothesis testing for high dimensional covariance matrices

T. TONY CAI\* and ZONGMING MA\*\*

Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA.  
E-mail: \*cai@wharton.upenn.edu; \*\*zongming@wharton.upenn.edu

This paper considers testing a covariance matrix  $\Sigma$  in the high dimensional setting where the dimension  $p$  can be comparable or much larger than the sample size  $n$ . The problem of testing the hypothesis  $H_0: \Sigma = \Sigma_0$  for a given covariance matrix  $\Sigma_0$  is studied from a minimax point of view. We first characterize the boundary that separates the testable region from the non-testable region by the Frobenius norm when the ratio between the dimension  $p$  over the sample size  $n$  is bounded. A test based on a  $U$ -statistic is introduced and is shown to be rate optimal over this asymptotic regime. Furthermore, it is shown that the power of this test uniformly dominates that of the corrected likelihood ratio test (CLRT) over the entire asymptotic regime under which the CLRT is applicable. The power of the  $U$ -statistic based test is also analyzed when  $p/n$  is unbounded.

*Keywords:* correlation matrix; covariance matrix; high-dimensional data; likelihood ratio test; minimax hypothesis testing; power; testing covariance structure

## 1. Introduction

Covariance structure plays a fundamental role in multivariate analysis and testing the covariance matrix is an important problem. Let  $X_1, \dots, X_n$  be  $n$  independent and identically distributed  $p$ -vectors following a multivariate normal distribution  $N_p(0, \Sigma)$ . A hypothesis testing problem of significant interest is testing

$$H_0: \Sigma = I. \tag{1}$$

Note that any null hypothesis  $H_0: \Sigma = \Sigma_0$  with a given positive definite covariance matrix  $\Sigma_0$  is equivalent to (1), since one can always transform  $X_i$  to  $\tilde{X}_i = \Sigma_0^{-1/2} X_i$  and then test (1) based on the transformed data.

This testing problem has been well studied in the classical setting of small  $p$  and large  $n$ . See, for example, Anderson [1] and Muirhead [13]. In particular, the likelihood ratio test (LRT) is commonly used. Driven by a wide range of contemporary scientific applications, analysis of high dimensional data is of significant current interest. In the high dimensional setting, where the dimension can be comparable to or even much larger than the sample size, the conventional testing procedures such as the LRT perform poorly or are not even well defined. Several testing procedures designed for the high-dimensional setting have been proposed. Let  $S = \frac{1}{n} \sum_{i=1}^n X_i X_i'$  be the sample covariance matrix. The existing tests for (1) in the literature can be categorized as the following according to the asymptotic regime under which they are suitable:

- $p$  fixed and  $n \rightarrow \infty$ . In this classical asymptotic regime, conventional tests for (1) include the likelihood ratio test (LRT) [1], Roy's largest root test [16], and Nagao's test [14]. In

particular, the LRT statistic is  $LR_n = nL_n$ , where

$$L_n = \text{tr } S - \log \det(S) - p.$$

The asymptotic distribution of  $LR_n$  under  $H_0$  is  $\chi_{p(p+1)/2}^2$ .

- Both  $n, p \rightarrow \infty$  and  $p/n \rightarrow c \in (0, \infty)$ . Investigation in this asymptotic regime has been very active in the past decade. For example, Johnstone [11] revisited Roy’s largest root test and derived the Tracy–Widom limit of its null distribution. Ledoit and Wolf [12] proposed a new test based on Nagao’s proposal. See also Srivastava [17]. When  $p$  grows, the chi-squared limiting null distribution of the LRT statistic  $LR_n$  is no longer valid. Recently, Bai *et al.* [2] proposed a corrected LRT when  $c < 1$ , and Jiang *et al.* [10] extended it to the case when  $p < n$  and  $c = 1$ . Here, for  $c_n = p/n$ , the test statistic of the corrected LRT is

$$CLR_n = \frac{L_n - p[1 - (1 - c_n^{-1}) \log(1 - c_n)] - 1/2 \log(1 - c_n)}{\sqrt{-2 \log(1 - c_n) - 2c_n}}, \tag{2}$$

whose asymptotic null distribution is  $N(0, 1)$ . Note that no test based on the likelihood ratio can be defined when  $p > n$  or  $c > 1$ .

- Both  $n, p \rightarrow \infty$  and  $p/n \rightarrow \infty$ . This is the ultra high-dimensional setting and both the LRT and corrected LRT are not well defined in this case. The testing problem in this asymptotic regime is not as well studied as in the previous categories. Birke and Dette [3] derived the asymptotic null distribution of the Ledoit–Wolf test under the current asymptotic regime. More recently, Chen *et al.* [7] proposed a new test statistic and derived its asymptotic null distribution when both  $n, p \rightarrow \infty$ , regardless of the limiting behavior of  $p/n$ .

When the dimension  $p$  grows together with the sample size  $n$ , the focus of most of the aforementioned papers is mainly on finding the asymptotic null distribution of the proposed test statistic, so the significance level of the test can be controlled. The few exceptions include Srivastava [17] and Chen *et al.* [7], where the asymptotic pointwise power of the proposed tests is also studied. Recently, Onatski *et al.* [15] established the regime of mutual contiguity of the joint distributions of the sample eigenvalues under the null and under the special alternative of rank one perturbation to the identity matrix, and then applied Le Cam’s third lemma to study the pointwise power of a collection of eigenvalue based tests for (1) against this special class of alternative.

In the present paper, we investigate this testing problem in the high-dimensional settings from a minimax point of view. Consider testing (1) against a composite alternative hypothesis

$$H_1: \Sigma \in \Theta, \quad \text{where } \Theta = \Theta_n = \{ \Sigma: \|\Sigma - I\|_F \geq \varepsilon_n \}. \tag{3}$$

Here,  $\|A\|_F = (\sum_{ij} a_{ij}^2)^{1/2}$  denotes the Frobenius norm of a matrix  $A = (a_{ij})$ . It is clear that the difficulty of testing between  $H_0$  and  $H_1$  depends on the value of  $\varepsilon_n$ : the smaller  $\varepsilon_n$  is, the harder it is to distinguish between the two hypotheses. An interesting question is: What is the boundary that separates the testable region, where it is possible to reliably detect the alternative based on the observations, from the untestable region, where it is impossible to do so? This problem is connected to the classical contiguity theory. It is also important to construct a test that can optimally distinguish between the two hypotheses in the testable region. The high-dimensional

settings here include all the cases where the dimension  $p = p_n \rightarrow \infty$  as the sample size  $n \rightarrow \infty$ , and there is no restriction on the limit of  $p/n$  unless otherwise stated.

For a given the significance level  $0 < \alpha < 1$ , our first goal is to identify the separation rate  $\varepsilon_n$  at which there exists a test  $\phi$  based on the random sample  $\{X_1, \dots, X_n\}$  such that

$$\inf_{\Sigma \in \Theta} P_{\Sigma}(\phi \text{ rejects } H_0) \geq \beta > \alpha.$$

Hence, the test is able to detect any alternative that is separated away from the null by a certain distance  $\varepsilon_n$  with a guaranteed power  $\beta > \alpha$ . Our second goal is to construct such a testing procedure  $\phi$ .

The major contribution of the current paper is threefold. First, we show that if  $p/n$  is bounded, then the rate  $\varepsilon_n$  needs to be no less than  $b\sqrt{p/n}$  for some constant  $b$ . In addition, it is shown that if  $\varepsilon_n = b\sqrt{p/n}$ , there exists a test  $\psi$  of significance level  $\alpha$ , such that  $\lim_{n \rightarrow \infty} \inf_{\Theta} P_{\Sigma}(\psi \text{ rejects } H_0) > \alpha$ , and the power tends to 1 if  $b = b_n \rightarrow \infty$ . The test is motivated by the proposal in Chen *et al.* [7]. (We use  $\psi$  to denote the specific test that we construct, while  $\phi$  is used to denote a generic test.) Here, we no longer require  $p/n$  to be bounded, and the explicit expression for the asymptotic power of  $\psi$  is also given. Moreover, we show that the asymptotic power of  $\psi$  on  $\Theta_n$  uniformly dominates that of the corrected LRT by Bai *et al.* [2] and Jiang *et al.* [10] over the entire asymptotic regime under which the corrected LRT is defined, that is,  $p < n$  and  $p/n \rightarrow c \in (0, 1]$ .

The rest of the paper is organized as the following. In Section 2, after introducing basic notation and definitions, we establish a lower bound of the separation rate  $\varepsilon_n$ . Section 3 introduces the test based on a  $U$ -statistic and provides a Berry–Essen bound for its weak convergence to the normal limit under both the null and the alternative hypotheses, which leads to the establishment of its guaranteed power over  $\Theta$  when  $\varepsilon_n = b\sqrt{p/n}$ . Furthermore, we also show that the power of this test uniformly dominates that of the corrected LRT. The theoretical results are supported by the numerical experiments in Section 4. Further discussions on the connections of our results and those of related testing problems are given in Section 5. The main results are proved in Section 6.

## 2. Lower bound

In this section, we establish a lower bound for the separation rate  $\varepsilon_n$  in (3). The result in Section 3 will show that this lower bound is rate-optimal. The lower and upper bounds together characterize the separation boundary between the testable and non-testable regions when the ratio of the dimension  $p$  over the sample size  $n$  is bounded. This separation boundary can then be used as a minimax benchmark for the evaluation of the performance of a test in this asymptotic regime.

We begin with basic notation and definitions. Throughout the paper, a test  $\phi = \phi_n(X_1, \dots, X_n)$  refers to a measurable function which maps  $X_1, \dots, X_n$  to the closed interval  $[0, 1]$ , where the value stands for the probability of rejecting  $H_0$ . So, the significance level of  $\phi$  is  $P_I(\phi \text{ rejects } H_0) = E_I \phi$ , and its power at a certain alternative  $\Sigma$  is  $P_{\Sigma}(\phi \text{ rejects } H_0) = E_{\Sigma} \phi$ . Here and after,  $P_{\Sigma}$ ,  $E_{\Sigma}$ ,  $\text{Var}_{\Sigma}$  and  $\text{Cov}_{\Sigma}$  denote the induced probability measure, expectation, variance and covariance when  $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N_p(0, \Sigma)$ . The subscript is shown only when clarity dictates.

To state the lower bound result, let  $\varepsilon_n = b\sqrt{p/n}$  for some constant  $b$ , and define

$$\Theta(b) = \left\{ \Sigma : \|\Sigma - I\|_F \geq b\sqrt{p/n} \right\}. \tag{4}$$

**Theorem 1 (Lower bound).** *Let  $0 < \alpha < \beta < 1$ . Suppose that as  $n \rightarrow \infty$ ,  $p \rightarrow \infty$  and that  $p/n \leq \kappa$  for some constant  $\kappa < \infty$  and all  $n$ . Then there exists a constant  $b = b(\kappa, \beta - \alpha) < 1$ , such that for any test  $\phi$  with significance level  $\alpha$  for testing  $H_0: \Sigma = I$ ,*

$$\limsup_{n \rightarrow \infty} \inf_{\Sigma \in \Theta(b)} \mathbb{E}_\Sigma \phi < \beta.$$

Theorem 1 shows that no level  $\alpha$  test for (1) can distinguish between the two hypotheses with power tending to 1 as  $n$  and  $p$  grow, when the separation rate  $\varepsilon_n$  is of order  $\sqrt{p/n}$ . Hence, it provides a lower bound for the separation rate.

We now give an outline of the proof for Theorem 1, while the complete proof is provided in Section 6.1. Consider the following ‘‘least favorable’’ subset of  $\Theta(b)$ :

$$\Theta^*(b) = \left\{ \Sigma_v = \left[ 1 - \frac{b}{\sqrt{n(p-1)}} \right] I_{p \times p} + \frac{b}{\sqrt{n(p-1)}} vv' : v \in \{\pm 1\}^p \right\}. \tag{5}$$

With slight abuse of notation, let  $P_0$  be the probability measure when  $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N_p(0, I)$  and  $P_v$  the probability measure when  $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N_p(0, \Sigma_v)$ . In addition, let  $P_1 = \frac{1}{2^p} \sum_{v \in \{\pm 1\}^p} P_v$  be the average measure of the  $P_v$ 's. Then for any test  $\phi$ , the sum of probabilities of its two types of errors satisfies

$$\begin{aligned} \sup_v \mathbb{E}_0 \phi + \mathbb{E}_v(1 - \phi) &\geq \inf_\psi \sup_v \mathbb{E}_0 \psi + \mathbb{E}_v(1 - \psi) \\ &\geq \inf_\psi \frac{1}{2^p} \sum_v \mathbb{E}_0 \psi + \mathbb{E}_v(1 - \psi) \\ &= \inf_\psi \mathbb{E}_0 \psi + \mathbb{E}_1(1 - \psi) \\ &= 1 - \frac{1}{2} \|P_1 - P_0\|_1. \end{aligned}$$

Here,  $\mathbb{E}_0, \mathbb{E}_v$  and  $\mathbb{E}_1$  denote the expectation under  $P_0, P_v$  and  $P_1$  respectively, and  $\|P_1 - P_0\|_1$  is the  $L_1$  distance between  $P_0$  and  $P_1$ . Thus, we obtain

$$\inf_{\Sigma \in \Theta(b)} \mathbb{E}_\Sigma \phi \leq \inf_v \mathbb{E}_v \phi \leq \mathbb{E}_0 \phi + \frac{1}{2} \|P_1 - P_0\|_1 = \alpha + \frac{1}{2} \|P_1 - P_0\|_1.$$

To control the rightmost side, we bound the  $L_1$  distance by the chi-square divergence as

$$\|P_1 - P_0\|_1^2 \leq \mathbb{E}_0 \left| \frac{dP_1}{dP_0} - 1 \right|^2 = \mathbb{E}_0 \left| \frac{dP_1}{dP_0} \right|^2 - 1 = \int \frac{f_1^2}{f_0} - 1,$$

where  $f_i$  is the density function of  $P_i$  for  $i = 0, 1$ . So, the proof can be completed by showing that for an appropriate choice of the constant  $b$ , one obtains  $\int \frac{f_1^2}{f_0} - 1 \leq 4(\beta - \alpha)^2$ .

**Remark 1.** (a). Note that all the covariance matrices in the least favorable configuration  $\Theta^*(b)$  defined in (5) have diagonal elements all equal to 1. Thus, they are also correlation matrices. So the proof of Theorem 1 readily establishes an analogous lower bound result on testing  $H_0: R = I$  with  $R$  the population correlation matrix.

(b). The lower bound argument here does not extend to the case when  $p/n$  is unbounded, because the chi-square divergence becomes unbounded.

### 3. Upper bound

In this section, we show that there exists a level  $\alpha$  test whose power over  $\Theta_n$  is uniformly larger than a prescribed value  $\beta > \alpha$ , if  $\varepsilon_n = b\sqrt{p/n}$  for a large enough constant  $b$ . This matches the lower bound result in Theorem 1 when  $p/n$  is bounded. In addition, the results in the current section remain valid even when  $p/n$  is unbounded.

We first introduce the test statistic in Section 3.1, followed by a study on the rate of convergence of its distribution to the normal limit under both the null and the alternative hypotheses in Section 3.2. Section 3.3 then uses the rate of convergence result to study the asymptotic power of the proposed test. Finally, Section 3.4 shows that the test dominates the corrected LRT in (2) when  $p/n \rightarrow c \in (0, 1]$ .

#### 3.1. Test statistic

Given a random sample  $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N_p(0, \Sigma)$ , a natural approach to test between (1) and (3) is to first estimate the squared Frobenius norm  $\|\Sigma - I\|_F^2 = \text{tr}(\Sigma - I)^2$  by some statistic  $T_n = T_n(X_1, \dots, X_n)$ , and then reject the null hypothesis if  $T_n$  is too large. To estimate  $\|\Sigma - I\|_F^2 = \text{tr}(\Sigma - I)^2$ , note that  $\mathbb{E}_\Sigma h(X_1, X_2) = \text{tr}(\Sigma - I)^2$  where

$$h(X_1, X_2) = (X_1' X_2)^2 - (X_1' X_1 + X_2' X_2) + p. \tag{6}$$

Therefore,  $\text{tr}(\Sigma - I)^2$  can be estimated by the following  $U$ -statistic

$$T_n = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h(X_i, X_j), \tag{7}$$

for which we have

$$\mu_n(\Sigma) = \mathbb{E}_\Sigma(T_n) = \text{tr}(\Sigma - I)^2, \tag{8}$$

$$\sigma_n^2(\Sigma) = \text{Var}_\Sigma(T_n) = \frac{4}{n(n-1)} [\text{tr}^2(\Sigma^2) + \text{tr}(\Sigma^4)] + \frac{8}{n} \text{tr}(\Sigma^2(\Sigma - I)^2). \tag{9}$$

Here, verifying (8) is straightforward, and (9) is proved in Appendix A.2. For the  $U$ -statistic  $T_n$ , the proof for Theorem 2 of Chen *et al.* [7] essentially established the following.

**Proposition 2 (Theorem 2 of [7]).** *Suppose that  $p \rightarrow \infty$  as  $n \rightarrow \infty$ . If a sequence of covariance matrices satisfy  $\text{tr}(\Sigma^2) \rightarrow \infty$  and  $\text{tr}(\Sigma^4)/\text{tr}^2(\Sigma^2) \rightarrow 0$  as  $n \rightarrow \infty$ , then under  $\mathbb{P}_\Sigma$ , we have*

$$\frac{T_n - \mu_n(\Sigma)}{\sigma_n(\Sigma)} \Rightarrow N(0, 1).$$

Note that as  $p \rightarrow \infty$ , the identity matrix  $I_{p \times p}$  satisfies the condition of the above proposition. Also note that  $\mu_n(I) = 0$  and  $\sigma_n^2(I) = \frac{4p(p+1)}{n(n-1)}$ . Thus, Proposition 2 quantifies the behavior of  $T_n$  under  $H_0$ , and we could define the test as the following: For any  $\alpha \in (0, 1)$ , an asymptotic level  $\alpha$  test based on  $T_n$  is given by

$$\psi = I\left(T_n > z_{1-\alpha} \cdot 2\sqrt{\frac{p(p+1)}{n(n-1)}}\right). \tag{10}$$

Here,  $I(\cdot)$  is the indicator function, and  $z_{1-\alpha}$  denotes the  $100 \times (1 - \alpha)$ th percentile of the standard normal distribution. This test is motivated by the test introduced in Chen *et al.* [7], while the original proposal in [7] involves higher order symmetric functions of the  $X_i$ 's.

In addition to specifying the rejection region in (10), Proposition 2 can also be used to study the asymptotic power of  $\psi$  over a sequence of simple alternatives. However, to understand the power of  $\psi$  over the composite alternative  $\Theta$  in (3), it is necessary to understand the rate of convergence of  $[T_n - \mu_n(\Sigma)]/\sigma_n(\Sigma)$  to the normal limit, which is the central topic of the next subsection.

### 3.2. Rate of convergence

We now study the rate of convergence for the distribution of  $[T_n - \mu_n(\Sigma)]/\sigma_n(\Sigma)$  to its normal limit in Kolmogorov distance. Let  $\Phi(\cdot)$  be the cumulative distribution function of the standard normal distribution. We have the following Berry–Essen type bound.

**Proposition 3.** *Under the condition of Proposition 2, there exists a numeric constant  $C$  such that*

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}_\Sigma\left(\frac{T_n - \mu_n(\Sigma)}{\sigma_n(\Sigma)} \leq x\right) - \Phi(x) \right| \leq C \left[ \frac{1}{n} + \frac{\text{tr}(\Sigma^4)}{\text{tr}^2(\Sigma^2)} \right]^{1/5}.$$

We outline the proof of Proposition 3 below, while the complete proof is deferred to Section 6.2. The primary tool used in the proof is a Berry–Esseen type bound for martingale central limit theorem by Heyde and Brown [8].

We begin by giving a martingale representation of  $T_n - \mu_n(\Sigma)$ . Let  $X_i \stackrel{\text{i.i.d.}}{\sim} N_p(0, \Sigma)$ . Define filtration

$$\mathcal{F}_0 = \sigma(\emptyset), \quad \mathcal{F}_k = \sigma(X_1, \dots, X_k), \quad k = 1, \dots, n.$$



Also introducing the notation  $E_k[\cdot] = E_\Sigma[\cdot|\mathcal{F}_k]$ . Then

$$T_n - \mu_n(\Sigma) = \sum_{k=1}^n E_k[T_n] - E_{k-1}[T_n] = \sum_{k=1}^n D_{nk}. \tag{11}$$

Here,  $\{D_{nk}: k = 1, \dots, n\}$  is a martingale difference sequence. The explicit expression for  $D_{nk}$  is

$$D_{nk} = \frac{2}{n(n-1)} [X'_k Q_{k-1} X_k - \text{tr}(Q_{k-1} \Sigma)] + \frac{2}{n} [X'_k \Sigma X_k - \text{tr}(\Sigma^2)] - \frac{2}{n} [X'_k X_k - \text{tr}(\Sigma)], \tag{12}$$

with  $Q_{k-1} = \sum_{i=1}^{k-1} (X_i X'_i - \Sigma)$ . Let  $\sigma_{nk}^2 = E_{k-1}[D_{nk}^2]$ , and we have  $\sigma_n^2(\Sigma) = \sum_{k=1}^n E_\Sigma[\sigma_{nk}^2]$ . Under the current setup, the main theorem in [8] specializes to the following lemma.

**Lemma 1.** *There exist a numeric constant  $C$ , such that*

$$\begin{aligned} & \sup_{x \in \mathbb{R}} \left| \mathbb{P}_\Sigma \left( \frac{T_n - \mu_n(\Sigma)}{\sigma_n(\Sigma)} \leq x \right) - \Phi(x) \right| \\ & \leq C \left[ \frac{1}{\sigma_n^4(\Sigma)} \left( \sum_{k=1}^n E_\Sigma[D_{nk}^4] + E_\Sigma \left[ \sum_{k=1}^n \sigma_{nk}^2 - \sigma_n^2(\Sigma) \right]^2 \right) \right]^{1/5}. \end{aligned} \tag{13}$$

Define

$$E_1 = \sum_{k=1}^n E_\Sigma[D_{nk}^4] \quad \text{and} \quad E_2 = E_\Sigma \left[ \sum_{k=1}^n \sigma_{nk}^2 - \sigma_n^2(\Sigma) \right]^2. \tag{14}$$

The proof of Proposition 3 could then be completed by showing that  $E_1/\sigma_n^4(\Sigma) = O(1/n)$  and  $E_2/\sigma_n^4(\Sigma) = O(\text{tr}(\Sigma^4)/\text{tr}^2(\Sigma^2))$ . See Section 6.2 for details.

### 3.3. Power of the test

Equipped with Proposition 3, we now investigate the power of the test  $\psi$  in (10) over the composite alternative  $H_1: \Sigma \in \Theta(b)$ , with  $b < 1$ , where  $\Theta(b)$  is defined in (4). In particular, we have the following result.

**Theorem 4 (Upper bound).** *Suppose that  $p \rightarrow \infty$  as  $n \rightarrow \infty$ . For any significance level  $\alpha \in (0, 1)$  and  $\Theta(b)$  in (4), the power of the test  $\psi$  in (10) satisfies*

$$\lim_{n \rightarrow \infty} \inf_{\Theta(b)} E_\Sigma \psi = 1 - \Phi \left( z_{1-\alpha} - \frac{b^2}{2} \right) > \alpha.$$

Moreover, for  $b_n \rightarrow \infty$ ,  $\lim_{n \rightarrow \infty} \inf_{\Theta(b_n)} E_\Sigma \psi = 1$ .

Theorem 4 shows that the test  $\psi$  can distinguish between the null (1) and the alternative (3) with power tending to 1 when  $b = b_n \rightarrow \infty$ . Comparing with the lower bound given in Theorem 1, the test  $\psi$  is rate-optimal when  $p/n$  is bounded. When  $\|\Sigma - I\|_F \asymp \sqrt{p/n}$ , the proof of Theorem 4 essentially shows that the power of  $\psi$  is also monotone increasing in  $\|\Sigma - I\|_F$ .

To prove Theorem 4, we first notice that the second claim is a direct consequence of the first one. Indeed, if the first claim is true, then for any fixed constant  $b > 0$ ,

$$\liminf_{n \rightarrow \infty} \inf_{\Theta(b_n)} E_\Sigma \psi \geq \lim_{n \rightarrow \infty} \inf_{\Theta(b)} E_\Sigma \psi = 1 - \Phi\left(z_{1-\alpha} - \frac{b^2}{2}\right).$$

Because the above inequality holds for any  $b$ , we obtain  $\liminf_{n \rightarrow \infty} \inf_{\Theta(b_n)} E_\Sigma \psi \geq 1$ . On the other hand,  $\psi \leq 1$  and so  $\limsup_{n \rightarrow \infty} \inf_{\Theta(b_n)} E_\Sigma \psi \leq 1$ . This leads to the second claim.

Turn to the proof of the first claim, we divide  $\Theta(b)$  into two disjoint subsets  $\Theta(b) = \Theta(b, B) \cup \Theta(B)$ , where

$$\begin{aligned} \Theta(b, B) &= \{\Sigma: b\sqrt{p/n} \leq \|\Sigma - I\|_F < B\sqrt{p/n}\}, \\ \Theta(B) &= \{\Sigma: \|\Sigma - I\|_F \geq B\sqrt{p/n}\}. \end{aligned} \tag{15}$$

Here,  $B$  is a sufficiently large constant, the choice of which depends only on  $\alpha$  and  $b$ , but not on  $n$  or  $p$ . We employ different proof strategies on the two subsets. On  $\Theta(B)$ , Chebyshev’s inequality readily shows that

$$\inf_{\Theta(B)} E_\Sigma \psi > 1 - \Phi\left(z_{1-\alpha} - \frac{b^2}{2}\right).$$

Turn to  $\Theta(b, B)$ . On this subset, Proposition 3 then plays the key role in obtaining a uniform approximation to the power function  $E_\Sigma \psi$  by the normal distribution function  $\Phi(z_{1-\alpha} - \frac{\|\Sigma - I\|_F^2}{2p/n})$ , which in turn leads to the final claim. For a detailed proof, see Section 6.3.

**Remark 2.** (a). When  $p/n$  is bounded, the conclusion of the theorem matches the lower bound in Theorem 1. However, the result here holds even when  $p/n$  is unbounded.

(b). It can be seen from the proof of Theorem 4 that the simple expression

$$\Phi\left(z_{1-\alpha} - \frac{\|\Sigma - I\|_F^2}{2p/n}\right)$$

gives good approximation to the power of the test  $\psi$  defined in (10) at any  $\Sigma$  of interest in practice, because the approximation works well until the power of the test is extremely close to  $\alpha$  or 1.

### 3.4. Power comparison with the corrected LRT

In the classical asymptotic regime where  $p$  is fixed and  $n \rightarrow \infty$ , the likelihood ratio test (LRT) is one of the most commonly used tests. In the high-dimensional setting where both  $n$  and  $p$

are large and  $p < n$ , Bai *et al.* [2] showed that the LRT is not well behaved as the chi-squared limiting distribution under  $H_0$  no longer holds.

For testing (1), when  $p < n$  and  $p/n \rightarrow c \in (0, 1)$ , Bai *et al.* [2] proposed a corrected likelihood ratio test (CLRT) with the test statistic  $CLR_n$  given in (2). It was shown that the test statistic  $CLR_n \Rightarrow N(0, 1)$  under  $H_0$  and this leads to an asymptotically level  $\alpha$  test by rejecting  $H_0$  when  $CLR_n > z_{1-\alpha}$ . It was shown that the CLRT significantly outperforms the LRT when both  $n$  and  $p$  are large and  $p < n$ . Recently, Jiang *et al.* [10] also considered the CLRT and showed that the above limit holds even when  $p/n \rightarrow 1$ .

It is interesting to compare the power of the CLRT with that of the test defined in (10). Note that the test given in (10) is always well defined, but the CLRT is only properly defined in the asymptotic regime where  $p < n$  and  $p/n \rightarrow c \in (0, 1]$ . The following result shows that the power of the CLRT is uniformly dominated by that of  $\psi$  given in (10) over the entire asymptotic regime under which the CLRT is applicable.

**Proposition 5.** *Suppose that as  $n \rightarrow \infty$ ,  $p \rightarrow \infty$  with  $p < n$  and  $p/n \rightarrow c \in (0, 1]$ . Let  $\mathbb{C}(\tau) = \{\Sigma: \|\Sigma - I\|_F = \tau\sqrt{p/n}\}$ . Then for  $\psi$  in (10) and the corrected LRT  $\phi_{\text{CLR}}$ , we have*

$$\lim_{n \rightarrow \infty} \inf_{\mathbb{C}(\tau)} E_{\Sigma} \psi > \limsup_{n \rightarrow \infty} \inf_{\mathbb{C}(\tau)} E_{\Sigma} \phi_{\text{CLR}} \quad \text{for all } \tau \in (0, 1).$$

Moreover, for  $\Theta(b)$  in (4) with  $b \in (0, 1)$ ,

$$\lim_{n \rightarrow \infty} \inf_{\Theta(b)} E_{\Sigma} \psi > \limsup_{n \rightarrow \infty} \inf_{\Theta(b)} E_{\Sigma} \phi_{\text{CLR}}.$$

Hence, the CLRT is sub-optimal whenever it is properly defined. A proof of Proposition 5 is given in Section 6.4.

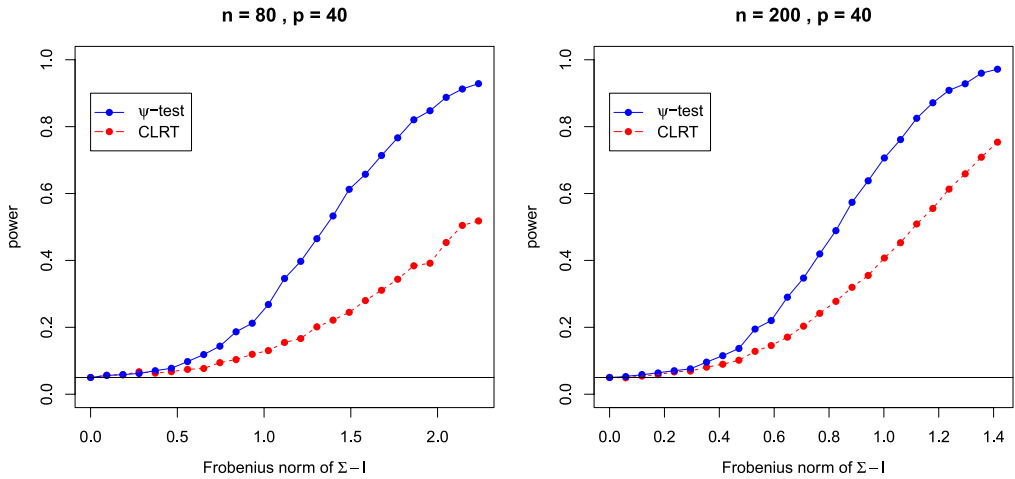
### 4. Numerical experiments

In this section, a small simulation study is carried out to compare the power of the test  $\psi$  defined in (10) with that of the CLRT under two specific alternatives.

The first alternative is the equi-correlation matrix  $\Sigma = (\sigma_{ij})$ , where for  $\rho \in (0, 1)$ ,

$$\sigma_{ij} = \begin{cases} 1, & i = j, \\ \rho, & i \neq j. \end{cases}$$

Figure 1 shows how the power functions of the  $\psi$  test and the CLRT grow with  $\|\Sigma - I\|_F$  when  $p = 40$  and  $n = 80$  or  $200$ . For both tests, the significance levels are fixed at  $\alpha = 0.05$ . To make a fair comparison, the 95th percentiles of the null distributions of both test statistics are obtained via simulation instead of using those of the asymptotic normal distributions. From Figure 1, it is clear that the  $\psi$  test is more powerful than the CLRT for both  $(n, p)$  configurations. The difference between the powers is smaller when  $n/p$  is larger. This is not surprising, because the LRT is a powerful test in the “large  $n$ , small  $p$ ” regime.



**Figure 1.** Power curves of the  $\psi$  test (solid) and the CLRT (dashed) under the equi-correlation alternative. Each dot is obtained from 5000 repetitions, and the curves are then obtained via linear interpolation.

The second alternative is the tridiagonal matrix  $\Sigma = (\sigma_{ij})$ , where for  $\rho \in (0, 1)$ ,

$$\sigma_{ij} = \begin{cases} 1, & i = j, \\ \rho, & |i - j| = 1, \\ 0, & |i - j| > 1. \end{cases}$$

Figure 2 shows how the power functions of the  $\psi$  test and the CLRT grow with  $\|\Sigma - I\|_F$  for the tridiagonal alternative. All the other setups remain unchanged. Here, the power of the  $\psi$  test still dominates, while the difference in power between the two tests is smaller.

### 5. Discussion

We have focused in the present paper on testing the hypotheses under the Frobenius norm. The technical arguments developed in this paper can also be used for testing under other matrix norms. Consider, for example, testing (1) against the following composite alternative hypothesis

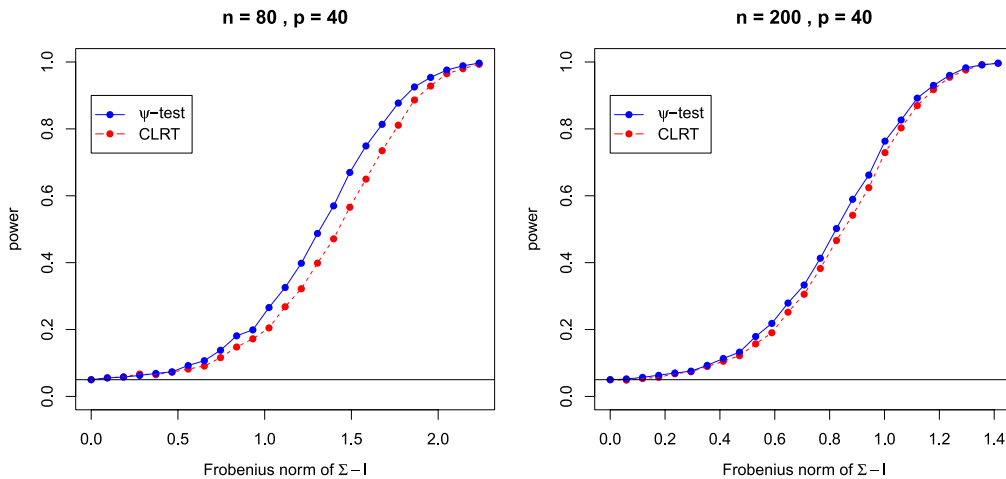
$$H_1: \Sigma \in \Theta, \quad \text{where } \Theta = \Theta_n = \{ \|\Sigma - I\|_s \geq \varepsilon_n \}.$$

Here  $\|A\|_s$  is the spectral norm defined by  $\|A\|_s = \max_{\|x\|_2=1} \|Ax\|_2$ . Define

$$\Theta_s(b) = \{ \Sigma: \|\Sigma - I\|_s \geq b\sqrt{p/n} \}. \tag{16}$$

Then the same lower bound holds for  $\Theta_s(b)$ . To be more precise, we have the following result.

**Theorem 6.** *Let  $0 < \alpha < \beta < 1$ . Suppose that as  $n \rightarrow \infty$ ,  $p \rightarrow \infty$  and  $p/n \leq \kappa$  for some constant  $\kappa < \infty$  and all  $n$ . Then there exist a constant  $b = b(\kappa, \beta - \alpha) < 1$ , such that for any test*



**Figure 2.** Power curves of the  $\psi$  test (solid) and the CLRT (dashed) against the tridiagonal alternative. Each dot is obtained from 5000 repetitions, and the curves are then obtained via linear interpolation.

$\phi$  with significance level  $0 < \alpha < 1$  for testing  $H_0: \Sigma = I$ ,

$$\limsup_{n \rightarrow \infty} \inf_{\Sigma \in \Theta_s(b)} \mathbb{E}_\Sigma \phi < \beta.$$

The proof of Theorem 6 is analogous to that of Theorem 1. We believe that the rate of  $\sqrt{p/n}$  in the lower bound is sharp. It is however unclear which test is optimal against the alternative (16) under the spectral norm. Obtaining a matching upper bound for a practically useful test is an interesting project for future research.

The results in the current paper also shed light on the problem of testing for independence, that is,  $H_0: R = I$ , where  $R$  is the population correlation matrix. Following Remark 1, the proof of Theorems 1 and 6 can be used directly to establish the same lower bound results on testing the correlation matrix.

Onatski *et al.* [15] also studied the hypothesis testing problem (1), but their attention is restricted to testing against alternatives that are rank one perturbations to the identity matrix. That is, under the alternative  $H_1$  the covariance matrix belongs to the set  $\Theta_h = \{I + hvv': \|v\|_2 = 1\}$ . The asymptotic regime is restricted to  $p/n \rightarrow c \in (0, \infty)$ . In this asymptotic regime, Theorem 7 in [15] gives a lower bound result analogous to Theorem 1. However, it does not cover the case when  $p/n \rightarrow 0$ , nor can it be extended to the case of testing correlation matrices. In addition, we notice that though the result in [15] enables one to study the asymptotic power of all the eigenvalue-based tests on each  $\Theta_h$  when  $p/n \rightarrow c \in (0, \infty)$ , it does not give a minimax claim as we did in Theorem 4.

The results in this paper also raised a number of interesting questions for future research. One example is the testing of equality of two covariance matrices based on the independent random samples  $X_1, \dots, X_{n_1} \stackrel{\text{i.i.d.}}{\sim} N_p(\mu_1, \Sigma_1)$  and  $Y_1, \dots, Y_{n_2} \stackrel{\text{i.i.d.}}{\sim} N_p(\mu_2, \Sigma_2)$ . The validity of many

commonly used statistical procedures including the classical Fisher’s linear discriminant analysis requires the assumption of equal covariance matrices. So it is of interest to test  $H_0: \Sigma_1 = \Sigma_2$ . Motivated by an unbiased estimator of the Frobenius norm of  $\Sigma_1 - \Sigma_2$ , Chen and Li [6] proposed a test using a linear combination of  $U$ -statistics and studied its power. Cai *et al.* [5] introduced a test based on the maximum of the standardized differences between the entries of the two sample covariance matrices. The test is shown to be powerful against sparse alternatives and robust with respect to the population distributions. However, the optimality of the two-sample tests has not been well studied. This is an important topic for future research that is of both theoretical and practical interest.

In the present paper, no structural assumption is imposed on the alternative class of the covariance matrices such as sparsity or bandedness. An optimal test against a structured alternative is potentially very different from the test (10) considered here. Recently, Cai and Jiang [4] considered testing the null hypothesis that  $\Sigma$  is a banded matrix and introduced a test based on the coherence of a random matrix. Xiao and Wu [18] proposed a test for testing  $H_0: \Sigma = I$  against sparse alternatives. Their test is based on the maximum of the standardized entries of the sample covariance matrix. The limiting null distribution is shown to be a type I extreme value distribution, the power of the test is not analyzed. It is interesting to investigate the optimality of these testing problems with structured alternatives.

## 6. Proofs

In this section, we prove Theorems 1, 4 and Propositions 3 and 5.

### 6.1. Proof of Theorem 1

Recall that  $P_0$  is the probability measure when  $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N_p(0, I)$  and  $P_v$  is the probability measure when  $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N_p(0, \Sigma_v)$ . In addition,  $P_1 = \frac{1}{2^p} \sum_{v \in \{\pm 1\}^p} P_v$  is the average measure of the  $P_v$ ’s. Let  $f_0$  and  $f_1$  be the density functions of  $P_0$  and  $P_1$ , respectively. By the discussion following Theorem 1, we could prove Theorem 1 by showing that  $\int f_1^2/f_0 - 1 \leq 4(\beta - \alpha)^2$ .

After some basic calculation (see Appendix A.1 for details), we obtain that if  $b < b_0(\kappa)$  such that

$$b < 1 \quad \text{and} \quad \frac{bp}{\sqrt{n(p-1)}} < \frac{1}{\sqrt{2}}, \tag{17}$$

then

$$\int \frac{f_1^2}{f_0} = \frac{(1-a^2)^{n-np/2}}{[1+(p-1)a^2]^n} \mathbb{E} \left[ 1 - \left( \frac{pa}{1+(p-1)a^2} \right)^2 \left( \frac{\mathbf{1}'V}{p} \right)^2 \right]^{-n/2}. \tag{18}$$

Here, the expectation is taken w.r.t.  $V = (V_1, \dots, V_p)'$  where the  $V_j$ ’s are i.i.d. Rademacher random variables which take values  $\pm 1$  with equal probability.

Note that (17) and  $p/n \leq \kappa$  implies

$$\left(\frac{pa}{1+(p-1)a^2}\right)^2 \leq \frac{1}{2}.$$

Also note that  $(\mathbf{1}'V/p)^2 \in [0, 1]$ . Thus, let  $\tilde{b}_{np} = (\frac{pa}{1+(p-1)a^2})^2$ , and  $(\mathbf{1}'V/p)^2 = \xi_p$ , we have

$$\begin{aligned} \mathbb{E}(1 - \tilde{b}_{np}\xi_p)^{-n/2} &= \mathbb{E}[(1 - \tilde{b}_{np}\xi_p)^{-1/(\tilde{b}_{np}\xi_p)}]^{n\tilde{b}_{np}\xi_p/2} \\ &\leq \mathbb{E} \exp\left(\frac{\log 4}{2} n\tilde{b}_{np}\xi_p\right) \quad (0 \leq (1-x)^{1/x} \leq 4, \text{ for all } x \in [0, 1/2]) \\ &\leq \mathbb{E} \exp\left(\frac{\log 4}{2} \frac{b^2 p^2}{p-1} \xi_p\right) \quad (\tilde{b}_{np} \leq p^2 b^2 / [n(p-1)]). \end{aligned}$$

For  $\xi_p$ , Hoeffding’s inequality [9], applied to Rademacher variables, yields

$$\mathbb{P}(\xi_p \geq \lambda) \leq 2e^{-2p\lambda} \quad \text{for all } \lambda > 0.$$

Thus, we obtain

$$\begin{aligned} \mathbb{E} \exp\left(\frac{\log 4}{2} \frac{b^2 p^2}{p-1} \xi_p\right) &= \int_0^\infty \mathbb{P}\left(\exp\left(\frac{\log 4}{2} \frac{b^2 p^2}{p-1} \xi_p\right) \geq u\right) du \\ &= 1 + \int_1^\infty \mathbb{P}\left(\xi_p \geq \frac{2 \log u}{\log 4} \frac{p-1}{b^2 p^2}\right) du \\ &\leq 1 + \int_1^\infty 2 \exp\left(-\frac{4(p-1) \log u}{b^2 p \log 4}\right) du \\ &= 1 + \frac{2b^2 p \log 2}{2(p-1) - b^2 p \log 2}. \end{aligned}$$

Here, the last equality holds if  $2(p-1) > b^2 p \log 2$ , which is always true for large  $p$  since  $b < 1$ .

In addition, with  $b$  satisfying (17), when  $n \rightarrow \infty$ ,

$$(1 - a^2)^{n-np/2} \rightarrow e^{b^2}, \quad [1 + (p-1)a^2]^n \rightarrow e^{b^2}.$$

Therefore, for large enough  $n \geq n_0(\kappa)$ ,

$$\int \frac{f_1^2}{f_0} - 1 \leq \frac{8b^2 p \log 2}{2(p-1) - b^2 p \log 2},$$

which, for sufficiently small  $b \leq b_0(\kappa, \beta - \alpha)$ , is no larger than  $4(\beta - \alpha)^2$ . This completes the proof.

### 6.2. Proof of Proposition 3

Following the outline of proof after Proposition 3, for  $E_1$  and  $E_2$  defined in (14), we complete the proof below by showing that  $E_1/\sigma_n^4(\Sigma) = O(1/n)$  and  $E_2/\sigma_n^4(\Sigma) = O(\text{tr}(\Sigma^4)/\text{tr}^2(\Sigma^2))$ .

To this end, we start with some preliminaries. Throughout the proof,  $E$  and  $\text{Var}$  are used as abbreviations for  $E_\Sigma$  and  $\text{Var}_\Sigma$ , respectively. Recall the martingale representation (11), where each martingale difference  $D_{nk}$  has the explicit expression (12). For  $D_{nk}$ , its conditional variance is

$$\begin{aligned} \sigma_{nk}^2 &= E_{k-1}[D_{nk}^2] \\ &= \frac{8}{n^2(n-1)^2} \text{tr}(Q_{k-1}\Sigma Q_{k-1}\Sigma) + \frac{16}{n^2(n-1)} \text{tr}(Q_{k-1}\Sigma^3) \\ &\quad - \frac{16}{n^2(n-1)} \text{tr}(Q_{k-1}\Sigma^2) + \frac{8}{n^2} \text{tr}(\Sigma^2(\Sigma - I)^2). \end{aligned} \tag{19}$$

Detailed derivation of (12) and (19) can be found in Appendix A.3. With (19), it is not difficult to verify that

$$E[\sigma_{nk}^2] = \frac{8(k-1)}{n^2(n-1)^2} [\text{tr}^2(\Sigma^2) + \text{tr}(\Sigma^4)] + \frac{8}{n^2} \text{tr}(\Sigma^2(\Sigma - I)^2),$$

and that  $\sigma_n^2 = \text{Var}(T_n) = \sum_{k=1}^n E[\sigma_{nk}^2]$ . Last but not least, we have for any  $j > k$ ,

$$E_{k-1}[\sigma_{nj}^2 - E\sigma_{nj}^2] = \sigma_{nk}^2 - E\sigma_{nk}^2. \tag{20}$$

Now, we turn to the studies of  $E_1$  and  $E_2$ .

*Term  $E_1$ .* We begin with the first term. Decompose the covariance matrix (as in [7]) as  $\Sigma = \Gamma\Gamma'$ , with  $\Gamma \in \mathbb{R}^{p \times p}$ . Then, we have the representation

$$X_i = \Gamma Z_i, \quad Z_i \stackrel{\text{i.i.d.}}{\sim} N_p(0, I), \quad i = 1, \dots, n. \tag{21}$$

We further define

$$A = \Gamma'\Gamma, \quad M_{k-1} = \Gamma' \sum_{i=1}^{k-1} (X_i X_i' - \Sigma) \Gamma = A \sum_{i=1}^{k-1} (Z_i Z_i' - I) A.$$

With the above definition, (12) can be rewritten as

$$D_{nk} = \frac{2}{n(n-1)} [Z_k' M_{k-1} Z_k - \text{tr}(M_{k-1})] + \frac{2}{n} [Z_k' (A^2 - A) Z_k - \text{tr}(A^2 - A)].$$

Therefore, we obtain from the Cauchy–Schwarz inequality and Lemma 3 that

$$\begin{aligned} E[D_{nk}^4] &\leq \frac{C}{n^4} E[Z_k' (A^2 - A) Z_k - \text{tr}(A^2 - A)]^4 + \frac{C}{n^4(n-1)^4} E[Z_k' M_{k-1} Z_k - \text{tr}(M_{k-1})]^4 \\ &\leq \frac{C}{n^4} \text{tr}^2(\Sigma^2(\Sigma - I)^2) + \frac{C}{n^4(n-1)^4} E[\text{tr}^2(M_{k-1}^2)]. \end{aligned}$$



For  $\text{tr}(M_{k-1}^2)$ , we use the following lemma, the proof of which is given in Appendix A.3.

**Lemma 2.** For  $\text{tr}(M_{k-1}^2)$ , we have

$$\begin{aligned} \mathbb{E}[\text{tr}(M_{k-1}^2)] &= (k-1)[\text{tr}^2(\Sigma^2) + \text{tr}(\Sigma^4)], \\ \text{Var}[\text{tr}(M_{k-1}^2)] &= (k-1)[24\text{tr}(\Sigma^8) + 16\text{tr}(\Sigma^6)\text{tr}(\Sigma^2) + 8\text{tr}^2(\Sigma^4) + 8\text{tr}(\Sigma^4)\text{tr}^2(\Sigma^2)] \\ &\quad + 2(k-1)(k-2)[6\text{tr}(\Sigma^8) + 2\text{tr}^2(\Sigma^4)]. \end{aligned}$$

For any sequences  $\{a_n\}$  and  $\{b_n\}$  of positive numbers, write  $a_n \lesssim b_n$  if  $\limsup_{n \rightarrow \infty} a_n/b_n < \infty$ . Note that  $\text{tr}(\Sigma^6) \leq \text{tr}(\Sigma^4)\text{tr}(\Sigma^2)$  and  $\text{tr}(\Sigma^8) \leq \text{tr}^2(\Sigma^4)$ . Since  $\text{tr}(\Sigma^4) = o(\text{tr}^2(\Sigma^2))$ , Lemma 2 implies that

$$\mathbb{E}[D_{nk}^4] \lesssim \frac{1}{n^4} \text{tr}^2(\Sigma^2(\Sigma - I)^2) + \frac{k^2}{n^8} \text{tr}^4(\Sigma^2),$$

and hence

$$E_1 \lesssim \frac{1}{n^3} \text{tr}^2(\Sigma^2(\Sigma - I)^2) + \frac{1}{n^5} \text{tr}^4(\Sigma^2). \tag{22}$$

*Term  $E_2$ .* For  $E_2$ , we can simplify it as

$$\begin{aligned} E_2 &= \mathbb{E} \left[ \sum_{k=1}^n (\sigma_{nk}^2 - \mathbb{E}\sigma_{nk}^2) \right]^2 \\ &= \mathbb{E} \left[ \sum_{k=1}^n (\sigma_{nk}^2 - \mathbb{E}\sigma_{nk}^2)^2 + 2 \sum_{k=1}^{n-1} \sum_{l=k+1}^n (\sigma_{nk}^2 - \mathbb{E}\sigma_{nk}^2)(\sigma_{nl}^2 - \mathbb{E}\sigma_{nl}^2) \right] \\ &= \sum_{k=1}^n \text{Var}(\sigma_{nk}^2) + 2 \sum_{k=1}^{n-1} (n-k) \text{Var}(\sigma_{nk}^2) = \sum_{k=1}^n (2n-2k+1) \text{Var}(\sigma_{nk}^2). \end{aligned}$$

Here, the second equality comes from (20).

Note that  $\text{tr}(Q_{k-1}\Sigma Q_{k-1}\Sigma) = \text{tr}(M_{k-1}^2)$  and  $\text{tr}(Q_{k-1}(\Sigma^3 - \Sigma^2)) = \text{tr}(M_{k-1}(A^2 - A))$ . So, by (19), there exist numeric constants  $C$  and  $C'$ , such that

$$\text{Var}(\sigma_{nk}^2) \leq \frac{C}{n^4(n-1)^4} \text{Var}[\text{tr}(M_{k-1}^2)] + \frac{C'}{n^4(n-1)^2} \text{Var}[\text{tr}(M_{k-1}(A^2 - A))].$$

We have studied  $\text{Var}[\text{tr}(M_{k-1}^2)]$  in Lemma 2. On the other hand, we have from Lemma 3 that

$$\begin{aligned} \text{Var}[\text{tr}(M_{k-1}(A^2 - A))] &= (k-1) \text{Var}[\text{tr}(AZZ'A(A^2 - A))] = (k-1) \text{Var}[Z'(A^4 - A^3)Z] \\ &= (k-1) \{ \mathbb{E}[(Z'(A^4 - A^3)Z)^2] - [\mathbb{E}[Z'(A^4 - A^3)Z]]^2 \} \\ &= (k-1) [2\text{tr}((A^4 - A^3)^2) + \text{tr}^2(A^4 - A^3) - \text{tr}^2(A^4 - A^3)] \end{aligned}$$

$$\begin{aligned}
 &= 2(k-1) \operatorname{tr}(\Sigma^6(\Sigma - I)^2) \\
 &\leq 2(k-1) \operatorname{tr}(\Sigma^4) \operatorname{tr}(\Sigma^2(\Sigma - I)^2).
 \end{aligned}$$

Since  $\operatorname{tr}(\Sigma^6) \leq \operatorname{tr}(\Sigma^4) \operatorname{tr}(\Sigma^2)$ ,  $\operatorname{tr}(\Sigma^8) \leq \operatorname{tr}^2(\Sigma^4)$  and  $\operatorname{tr}(\Sigma^4) = o(\operatorname{tr}^2(\Sigma^2))$ , we obtain that

$$\operatorname{Var}(\sigma_{nk}^2) \lesssim \frac{k}{n^8} \operatorname{tr}(\Sigma^4) \operatorname{tr}^2(\Sigma^2) + \frac{k^2}{n^2} \operatorname{tr}^2(\Sigma^4) + \frac{k}{n^6} \operatorname{tr}(\Sigma^4) \operatorname{tr}(\Sigma^2(\Sigma - I)^2),$$

which leads to the bound

$$E_2 \lesssim \frac{1}{n^3} \operatorname{tr}(\Sigma^4) \operatorname{tr}(\Sigma^2(\Sigma - I)^2) + \frac{1}{n^4} \operatorname{tr}^2(\Sigma^4) + \frac{1}{n^5} \operatorname{tr}(\Sigma^4) \operatorname{tr}^2(\Sigma^2). \tag{23}$$

*Summing up.* By (9), we have

$$\sigma_n^4 \asymp \frac{1}{n^2} \operatorname{tr}^2(\Sigma^2(\Sigma - I)^2) + \frac{1}{n^3} \operatorname{tr}^2(\Sigma^2) \operatorname{tr}(\Sigma^2(\Sigma - I)^2) + \frac{1}{n^4} \operatorname{tr}^4(\Sigma^2). \tag{24}$$

Here and after, for any sequences  $\{a_n\}$  and  $\{b_n\}$  of positive numbers, we write  $a_n \asymp b_n$  if  $a_n/b_n$  is bounded away from both 0 and  $\infty$ . Thus, we obtain that

$$\sigma_n^{-4} E_1 = O(n^{-1}), \quad \sigma_n^{-4} E_2 = O(\operatorname{tr}(\Sigma^4)/\operatorname{tr}^2(\Sigma^2)).$$

Plugging these estimates in Lemma 1, we complete the proof.

### 6.3. Proof of Theorem 4

Following the discussion after Theorem 4, we give below the detailed proof of the first claim in the theorem. In particular, we bound the power of the test separately on  $\Theta(B)$  and  $\Theta(b, B)$ , which are defined in (15).

*Case 1:  $\Theta(B)$ .* Here, we shall proceed heavy-handedly by using Chebyshev’s inequality, because the alternative class is sufficiently far away from  $H_0$ .

For any  $\Sigma \in \Theta(B)$ , there exists  $\tau \geq B$ , s.t.  $\|\Sigma - I\|_F = \tau \sqrt{p/n}$ . Suppose  $B$  is large enough s.t.  $\tau^2 \geq B^2 \geq 3z_{1-\alpha}$ . Note that  $\sigma_n(I) = (2p/n)(1 + o(1))$ , and so

$$E_\Sigma T_n = \|\Sigma - I\|_F^2 = \frac{\tau^2}{2} \sigma_n(I)(1 + o(1)) > z_{1-\alpha} \sigma_n(I).$$

Thus, we can use Chebyshev’s inequality to bound the type II error of  $\psi$  at  $\Sigma$  as the following:

$$\begin{aligned}
 1 - E_\Sigma \psi &= \mathbb{P}_\Sigma(T_n \leq z_{1-\alpha} \sigma_n(I)) = \mathbb{P}_\Sigma(T_n - E_\Sigma T_n \leq z_{1-\alpha} \sigma_n(I) - E_\Sigma T_n) \\
 &\leq \mathbb{P}_\Sigma(|T_n - E_\Sigma T_n| \geq |z_{1-\alpha} \sigma_n(I) - E_\Sigma T_n|) \\
 &\leq \frac{\operatorname{Var}_\Sigma(T_n)}{[z_{1-\alpha} \sigma_n(I) - E_\Sigma T_n]^2}.
 \end{aligned} \tag{25}$$

For  $\text{Var}_\Sigma(T_n) = \sigma_n^2(\Sigma)$ , we have its explicit expression given in (9). Let  $\lambda_{\max}(\Sigma)$  denote the largest eigenvalue of  $\Sigma$ . When  $\|\Sigma - I\|_F = \tau\sqrt{p/n}$ , we have  $\lambda_{\max}(\Sigma) \leq 1 + \tau\sqrt{p/n}$ , and so

$$\text{tr}(\Sigma^2) \leq p \left(1 + \frac{\tau}{\sqrt{n}}\right)^2, \quad \text{tr}(\Sigma^2(\Sigma - I)^2) \leq \lambda_{\max}^2(\Sigma) \|\Sigma - I\|_F^2 \leq \frac{\tau^2 p}{n} \left(1 + \tau\sqrt{\frac{p}{n}}\right)^2.$$

Since  $\text{tr}(\Sigma^4) \leq \text{tr}^2(\Sigma^2)$  and  $\sigma_n^2(I) = (4p^2/n^2)(1 + o(1))$ , the above inequalities, together with (9), lead to

$$\frac{\sigma_n^2(\Sigma)}{\sigma_n^2(I)} \leq \left[2 \left(1 + \frac{\tau}{\sqrt{n}}\right)^2 + \frac{2\tau^2}{p} \left(1 + \tau\sqrt{\frac{p}{n}}\right)^2\right] (1 + o(1)).$$

Since  $\tau^2/2 - z_{1-\alpha} \geq \tau^2/6$ , there exists some constant  $C_\alpha$  depending only on  $\alpha$ , such that

$$\begin{aligned} \frac{\text{Var}_\Sigma(T_n)}{[z_{1-\alpha}\sigma_n(I) - \mathbb{E}_\Sigma T_n]^2} &\leq \frac{2(1 + \tau/\sqrt{n})^2 + (2\tau^2/p)(1 + \tau\sqrt{p/n})^2}{(\tau^2/2 - z_{1-\alpha})^2} (1 + o(1)) \\ &\leq C_\alpha \left[ \left(\frac{1}{\sqrt{\tau}} + \frac{1}{\sqrt{n}}\right)^4 + \frac{1}{\tau^2 p} + \frac{1}{n} \right]. \end{aligned}$$

Note that all the  $o(1)$  terms in the above derivation are uniform over  $\Theta(B)$ . Therefore, given  $\alpha$  and  $b$ , there exist a constant  $B = B(\alpha, b)$ , such that

$$\begin{aligned} \liminf_{n \rightarrow \infty} \inf_{\Theta(B)} \mathbb{E}_\Sigma \psi &\geq 1 - \frac{\text{Var}_\Sigma(T_n)}{[z_{1-\alpha}\sigma_n(I) - \mathbb{E}_\Sigma T_n]^2} \\ &\geq 1 - C_\alpha \left[ \left(\frac{1}{\sqrt{B}} + \frac{1}{\sqrt{n}}\right)^4 + \frac{1}{B^2 p} + \frac{1}{n} \right] \\ &\geq 1 - \Phi\left(z_{1-\alpha} - \frac{b^2}{2}\right) > \alpha. \end{aligned} \tag{26}$$

*Case 2:  $\Theta(b, B)$ .* On this subset, we use Proposition 3 to obtain the following uniform approximation to the power function by the normal distribution function

$$\sup_{\Theta(b, B)} \left| \mathbb{E}_\Sigma \psi - \Phi\left(z_{1-\alpha} - \frac{\|\Sigma - I\|_F^2}{2p/n}\right) \right| \rightarrow 0. \tag{27}$$

If (27) is true, then we obtain

$$\lim_{n \rightarrow \infty} \inf_{\Theta(b, B)} \mathbb{E}_\Sigma \psi = \lim_{n \rightarrow \infty} \inf_{\Theta(b, B)} \Phi\left(z_{1-\alpha} - \frac{\|\Sigma - I\|_F^2}{2p/n}\right) = 1 - \Phi\left(z_{1-\alpha} - \frac{b^2}{2}\right) > \alpha.$$

Together with (26), this leads to the desired claim.

Turn to the proof of (27). First, note that uniformly on  $\Theta(b, B)$ , we have

$$p(1 - B/\sqrt{n})^2 \leq \text{tr}(\Sigma^2) \leq p(1 + B/\sqrt{n})^2, \tag{28}$$

and  $\text{tr}(\Sigma^4) \leq \lambda_{\max}^2(\Sigma) \text{tr}(\Sigma^2) \leq B^2(p/n)p(1 + B/\sqrt{n})^2$ . Therefore, as  $n \rightarrow \infty$ ,

$$\sup_{\Theta(b, B)} |\text{tr}(\Sigma^2)/p - 1| \rightarrow 0, \quad \sup_{\Theta(b, B)} \text{tr}(\Sigma^4)/\text{tr}^2(\Sigma^2) \rightarrow 0. \tag{29}$$

So the condition of Proposition 3 is satisfied. Next, we observe that

$$\begin{aligned} \mathbb{E}_\Sigma \psi &= \mathbb{P}_\Sigma(T_n > z_{1-\alpha} \sigma_n(I)) = \mathbb{P}_\Sigma\left(\frac{T_n - \|\Sigma - I\|_F^2}{\sigma_n(\Sigma)} \geq \frac{\sigma_n(I)}{\sigma_n(\Sigma)} z_{1-\alpha} - \frac{\|\Sigma - I\|_F^2}{\sigma_n(\Sigma)}\right) \\ &= \mathbb{P}_\Sigma\left(\frac{T_n - \mu_n(\Sigma)}{\sigma_n(\Sigma)} \geq \frac{\sigma_n(I)}{\sigma_n(\Sigma)} z_{1-\alpha} - \frac{\|\Sigma - I\|_F^2}{\sigma_n(\Sigma)}\right). \end{aligned}$$

Thus, Proposition 3 and (29) together imply that

$$\sup_{\Theta(b, B)} \left| \mathbb{E}_\Sigma \psi - \Phi\left(\frac{\sigma_n(I)}{\sigma_n(\Sigma)} z_{1-\alpha} - \frac{\|\Sigma - I\|_F^2}{\sigma_n(\Sigma)}\right) \right| \rightarrow 0.$$

To complete the proof, what is left to be verified is that

$$\sup_{\Theta(b, B)} \left| \frac{\sigma_n^2(\Sigma)}{4p^2/n^2} - 1 \right| \rightarrow 0, \tag{30}$$

because it implies  $\sup_{\Theta(b, B)} |\Phi(z_{1-\alpha} - \frac{\|\Sigma - I\|_F^2}{2p/n}) - \Phi(\frac{\sigma_n(I)}{\sigma_n(\Sigma)} z_{1-\alpha} - \frac{\|\Sigma - I\|_F^2}{\sigma_n(\Sigma)})| \rightarrow 0$ , which together with the last display before (30), leads to (27). To show (30), first recall the expression of  $\sigma_n^2(\Sigma)$  in (9). By (29), we obtain that the first term in (9) is  $4(p^2/n^2)(1 + o(1))$  where  $o(1)$  is uniform on  $\Theta(b, B)$ . On the other hand,

$$\begin{aligned} \text{tr}(\Sigma^2(\Sigma - I)^2) &\leq \lambda_{\max}^2(\Sigma) \|\Sigma - I\|_F^2 \\ &\leq \left(1 + B\sqrt{\frac{p}{n}}\right)^2 \frac{Bp}{n} \leq C(B) \max\left(1, \frac{p}{n}\right) \cdot \frac{p}{n}. \end{aligned}$$

Here,  $C(B)$  is a constant depending only on  $B$ . Therefore, we have that the second term in (9) is of order  $o(p^2/n^2)$  uniformly over  $\Theta(b, B)$ . Putting the two parts together leads to (30). This completes the proof.

### 6.4. Proof of Proposition 5

Fix any  $\tau \in (0, 1)$ . At each dimension  $p$ , consider a single point in  $\mathbb{C}(\tau)$ :

$$\Sigma^* = I_{p \times p} + \tau \sqrt{\frac{p}{n}} uu',$$

where  $u$  is an arbitrarily fixed unit vector in  $\mathbb{R}^p$ . Since  $\tau < 1$ , Proposition 10 in [15] leads to

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\Sigma^*} \phi_{\text{CLR}} = 1 - \Phi(z_{1-\alpha} - h(\tau, c)) \quad \text{for } h(\tau, c) = \frac{\tau\sqrt{c} - \log(1 + \tau\sqrt{c})}{\sqrt{-2\log(1-c) - 2c}}.$$

Note that for all  $\tau > 0$  and  $c \in (0, 1)$ ,  $\tau^2/2 > h(\tau, c) > 0$ . Therefore,

$$\begin{aligned} \lim_{n \rightarrow \infty} \inf_{\mathbb{C}(\tau)} \mathbb{E}_{\Sigma} \psi &= 1 - \Phi\left(z_{1-\alpha} - \frac{\tau^2}{2}\right) \\ &> 1 - \Phi(z_{1-\alpha} - h(\tau, c)) \\ &= \lim_{n \rightarrow \infty} \mathbb{E}_{\Sigma^*} \phi_{\text{CLR}} \\ &\geq \limsup_{n \rightarrow \infty} \inf_{\mathbb{C}(\tau)} \mathbb{E}_{\Sigma} \phi_{\text{CLR}}. \end{aligned}$$

The proof of the second claim is obtained by replacing  $\mathbb{C}(\tau)$  with  $\Theta(b)$  and  $\tau$  with  $b$  in the above arguments.

## Appendix: Technical details

### A.1. Proof details for Theorem 1

Here we give the calculation leading to (18) in the proof of Theorem 1.

Consider  $\Theta^*(b)$  in (5). For any  $v \in \{\pm 1\}^p$ , we have  $\|\Sigma_v - I\|_F = b\sqrt{p/n}$ ,  $\text{diag}(\Sigma_v) = (1, \dots, 1)$ , and for  $a = b/\sqrt{n(p-1)}$ ,

$$\begin{aligned} \Sigma_v^{-1} &= \frac{1}{1-a} I_{p \times p} - \left[ \frac{1}{1-a} - \frac{1}{1+(p-1)a} \right] \frac{1}{p} vv', \\ \det \Sigma_v &= (1-a)^{p-1} [1+(p-1)a]. \end{aligned} \tag{31}$$

Therefore, we have

$$\begin{aligned} f_0(x_1, \dots, x_n) &= \frac{1}{(2\pi)^{np/2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^n x'_i x_i\right\}, \\ f_1(x_1, \dots, x_n) &= \frac{1}{2^p} \sum_v \frac{1}{(2\pi)^{np/2} (\det \Sigma_v)^{n/2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^n x'_i \Sigma_v^{-1} x_i\right\} \\ &= \frac{1}{(2\pi)^{np/2}} \exp\left\{-\frac{1}{2(1-a)} \sum_{i=1}^n x'_i x_i\right\} \frac{1}{(1-a)^{n(p-1)/2} [1+(p-1)a]^{n/2}} \\ &\quad \times \frac{1}{2^p} \sum_v \exp\left\{\frac{1}{2p} \left[ \frac{1}{1-a} - \frac{1}{1+(p-1)a} \right] \sum_{i=1}^n (v'_i x_i)^2\right\}. \end{aligned}$$

And so

$$\begin{aligned} \frac{f_1^2}{f_0} &= \frac{1}{(1-a)^n(p-1)[1+a(p-1)]^n} \times \frac{1}{(2\pi)^{np/2}} \exp\left\{-\frac{1}{2}\left(\frac{1+a}{1-a}\right) \sum_{i=1}^n x'_i x_i\right\} \\ &\quad \times \frac{1}{2^{2p}} \left\{ \sum_v \exp\left[\frac{1}{p}\left(\frac{1}{1-a} - \frac{1}{1+(p-1)a}\right) \sum_{i=1}^n (v'x_i)^2\right] \right. \\ &\quad \left. + \sum_{v \neq u} \exp\left[\frac{1}{2p}\left(\frac{1}{1-a} - \frac{1}{1+(p-1)a}\right) \left(\sum_{i=1}^n (v'x_i)^2 + \sum_1^n (u'x_i)^2\right)\right] \right\}. \end{aligned}$$

Now we compute the integral. Fix any  $v$ , we have

$$\begin{aligned} &\int \frac{1}{(2\pi)^{np/2}} \exp\left\{-\frac{1}{2}\left(\frac{1+a}{1-a}\right) \sum_{i=1}^n x'_i x_i\right\} \exp\left\{\frac{1}{p}\left(\frac{1}{1-a} - \frac{1}{1+(p-1)a}\right) \sum_{i=1}^n (v'x_i)^2\right\} dx \\ &= \left[ \int \frac{1}{(2\pi)^{p/2}} \exp\left\{-\frac{1}{2}\left(\frac{1+a}{1-a}\right) x'x\right\} \exp\left\{\frac{1}{p}\left(\frac{1}{1-a} - \frac{1}{1+(p-1)a}\right) (v'x)^2\right\} dx \right]^n \\ &= \left(\frac{1+a}{1-a}\right)^{-np/2} [\mathbb{E} \exp(tY)]^n \\ &= \left(\frac{1+a}{1-a}\right)^{-np/2} (1-2t)^{-n/2} \quad (\text{for } t \leq 1/2), \end{aligned}$$

where  $Y \sim \chi_{(1)}^2$ , and

$$t = \left(\frac{1}{1-a} - \frac{1}{1+(p-1)a}\right) \left(\frac{1+a}{1-a}\right)^{-1}. \tag{32}$$

In addition, fix any  $v \neq u$ , we have

$$\begin{aligned} &\int \frac{1}{(2\pi)^{np/2}} \exp\left\{-\frac{1}{2}\left(\frac{1+a}{1-a}\right) \sum_{i=1}^n x'_i x_i\right\} \\ &\quad \times \exp\left\{\frac{1}{2p}\left(\frac{1}{1-a} - \frac{1}{1+(p-1)a}\right) \left(\sum_{i=1}^n (v'x_i)^2 + (u'x_i)^2\right)\right\} dx \\ &= \left[ \int \frac{1}{(2\pi)^{p/2}} \exp\left\{-\frac{1}{2}\left(\frac{1+a}{1-a}\right) x'x\right\} \right. \\ &\quad \left. \times \exp\left\{\frac{1}{2p}\left(\frac{1}{1-a} - \frac{1}{1+(p-1)a}\right) [(v'x)^2 + (u'x)^2]\right\} dx \right]^n. \end{aligned}$$

Let  $X \sim N_p(0, I)$ ,  $Z_1 = v'X/\sqrt{p}$ , and  $Z_2 = u'X/\sqrt{p}$ . Then

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & v'u/p \\ v'u/p & 1 \end{bmatrix}\right),$$

and so  $Z_1^2 + Z_2^2 \stackrel{d}{=} (1 + v'u/p)Y_1 + (1 - v'u/p)Y_2$ , with  $Y_i \stackrel{\text{i.i.d.}}{\sim} \chi_{(1)}^2$ . Therefore, the second last display equals

$$\begin{aligned} & \left(\frac{1+a}{1-a}\right)^{-np/2} [\mathbb{E} \exp(t_1 Y_1 + t_2 Y_2)]^n \\ &= \left(\frac{1+a}{1-a}\right)^{-np/2} (1-2t_1)^{-n/2} (1-2t_2)^{-n/2}, \end{aligned}$$

where

$$t_1 = \frac{t}{2} \left(1 + \frac{1}{p} v'u\right), \quad t_2 = \frac{t}{2} \left(1 - \frac{1}{p} v'u\right).$$

Collecting terms, we obtain after some linear algebra that

$$\begin{aligned} \int \frac{f_1^2}{f_0} &= \frac{(1-a^2)^{n-np/2}}{[1+(p-1)a^2]^n} \mathbb{E}_{V,U} \left[ 1 - \left(\frac{pa}{1+(p-1)a^2}\right)^2 \left(\frac{V'U}{p}\right)^2 \right]^{-n/2} \\ &= \frac{(1-a^2)^{n-np/2}}{[1+(p-1)a^2]^n} \mathbb{E}_V \left[ 1 - \left(\frac{pa}{1+(p-1)a^2}\right)^2 \left(\frac{\mathbf{1}'V}{p}\right)^2 \right]^{-n/2}. \end{aligned}$$

Here, both  $V$  and  $U$  have i.i.d. Rademacher entries which take values  $\pm 1$  with equal probability, and in the first expectation,  $V$  and  $U$  are independent.

### A.2. Variance of $T_n$

In this part, we establish the variance of  $T_n$  given in (9). We begin with a technical lemma, which is closely connected to [7], Proposition A.1.

**Lemma 3.** Let  $Z_1, Z_2 \stackrel{\text{i.i.d.}}{\sim} N_p(0, I)$ , and  $M, N$  be two  $p \times p$  p.s.d. matrices, then

$$\mathbb{E}[(Z_1' M Z_1)(Z_1' N Z_1)] = \text{tr}(M) \text{tr}(N) + 2 \text{tr}(MN); \tag{33}$$

$$\mathbb{E}[(Z_1' M Z_2)^4] = 3 \text{tr}^2(M^2) + 6 \text{tr}(M^4); \tag{34}$$

$$\mathbb{E}[(Z_1' M Z_1 - \text{tr}(M))^4] = 48 \text{tr}(M^4) + 12 \text{tr}^2(M^2). \tag{35}$$

**Proof.** Denote the ordered eigenvalues of  $M$  by  $\lambda_1 \geq \dots \geq \lambda_p$ , and those of  $N$  by  $\mu_1 \geq \dots \geq \mu_p$ . Let  $U_j \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ ,  $j = 1, \dots, p$ . For (33), we have

$$\begin{aligned} \mathbb{E}[(Z_1' M Z_1)(Z_1' N Z_1)] &= \mathbb{E}\left[\left(\sum_{j=1}^p \lambda_j U_j^2\right)\left(\sum_{j=1}^p \mu_j U_j^2\right)\right] \\ &= \sum_{j=1}^p \lambda_j \mu_j \mathbb{E}[U_j^4] + \sum_{j \neq l}^p \lambda_j \mu_l \mathbb{E}[U_j^2] \mathbb{E}[U_l^2] \\ &= 3 \sum_{j=1}^p \lambda_j \mu_j + \sum_{j \neq l}^p \lambda_j \mu_l = \text{tr}(M) \text{tr}(N) + 2 \text{tr}(MN). \end{aligned}$$

For (34), we define  $V_j \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ ,  $j = 1, \dots, p$ , which are independent from the  $U_j$ 's. Then

$$\begin{aligned} \mathbb{E}[(Z_1' M Z_2)^4] &= \mathbb{E}\left[\left(\sum_{j=1}^p \lambda_j U_j V_j\right)^4\right] \\ &= \sum_{j=1}^p \lambda_j^4 \mathbb{E}[U_j^4] \mathbb{E}[V_j^4] + \binom{4}{2} \sum_{j \neq l}^p \lambda_j^2 \lambda_l^2 \mathbb{E}[U_j^2] \mathbb{E}[U_l^2] \mathbb{E}[V_j^2] \mathbb{E}[V_l^2] \\ &= 9 \sum_{j=1}^p \lambda_j^4 + 6 \sum_{j \neq l}^p \lambda_j^2 \lambda_l^2 = 3 \text{tr}^2(M^2) + 6 \text{tr}(M^4). \end{aligned}$$

Finally, for (35), we have

$$\begin{aligned} \mathbb{E}[(Z_1' M Z_1 - \text{tr}(M))^4] &= \mathbb{E}\left[\left(\sum_{j=1}^p \lambda_j (U_j^2 - 1)\right)^4\right] \\ &= \sum_{j=1}^p \lambda_j^4 \mathbb{E}[(U_j^2 - 1)^4] + 6 \sum_{j \neq l}^p \lambda_j^2 \lambda_l^2 \mathbb{E}[(U_j^2 - 1)^2] \mathbb{E}[(U_l^2 - 1)^2] \\ &= 60 \sum_{j=1}^p \lambda_j^4 + 24 \sum_{j \neq l}^p \lambda_j^2 \lambda_l^2 = 48 \text{tr}(M^4) + 12 \text{tr}^2(M^2). \end{aligned}$$

This completes the proof of the lemma. □

In order to understand the variance of  $T_n$ , we need the following lemma.



**Lemma 4.** For  $X_1, X_2, X_3 \stackrel{\text{i.i.d.}}{\sim} N_p(0, \Sigma)$ , we have

$$\begin{aligned}\text{Var}(h(X_1, X_2)) &= 2[\text{tr}^2(\Sigma^2) + \text{tr}(\Sigma^4)] + 4\text{tr}(\Sigma^2(\Sigma - I)^2), \\ \text{Cov}(h(X_1, X_2), h(X_1, X_3)) &= 2\text{tr}(\Sigma^2(\Sigma - I)^2).\end{aligned}$$

**Proof.** For the variance, we first decompose it as

$$\text{Var}(h(X_1, X_2)) = \text{Var}(X_1'X_2)^2 + 2\text{Var}(X_1'X_1) - 4\text{Cov}((X_1'X_2)^2, (X_1'X_1)).$$

For  $\text{Var}(X_1'X_2)^2 = \mathbb{E}[(X_1'X_2)^4] - [\mathbb{E}(X_1'X_2)^2]^2$ , we have from (34) that

$$\mathbb{E}[(X_1'X_2)^4] = \mathbb{E}[(Z_1'AZ_2)^4] = 3\text{tr}^2(A^2) + 6\text{tr}(A^4) = 3\text{tr}^2(\Sigma^2) + 6\text{tr}(\Sigma^4).$$

On the other hand, we have

$$\mathbb{E}[(X_1'X_2)^2] = \mathbb{E}[Z_1'AZ_2Z_2'AZ_1] = \mathbb{E}[\text{tr}(AZ_2Z_2'A)] = \mathbb{E}[Z_2'A^2Z_2] = \text{tr}(A^2) = \text{tr}(\Sigma^2).$$

Thus, we obtain  $\text{Var}(X_1'X_2)^2 = 2\text{tr}^2(\Sigma^2) + 6\text{tr}(\Sigma^4)$ . Similar type of calculation yields that

$$\text{Var}(X_1'X_1) = 2\text{tr}(\Sigma^2), \quad \text{Cov}((X_1'X_2)^2, (X_1'X_1)) = 2\text{tr}(\Sigma^3).$$

Assembling the pieces, we prove the variance formula.

For the covariance formula, the basic quantity to compute is

$$\begin{aligned}\mathbb{E}[(X_1'X_2)^2 - (X_1'X_1) - (X_2'X_2)][(X_1'X_3)^2 - (X_1'X_1) - (X_3'X_3)] \\ &= \mathbb{E}(X_1'X_2)^2(X_1'X_3)^2 - \mathbb{E}(X_1'X_1)(X_1'X_3)^2 - \mathbb{E}(X_2'X_2)\mathbb{E}(X_1'X_3)^2 \\ &\quad - \mathbb{E}(X_1'X_2)^2(X_1'X_1) + \mathbb{E}(X_1'X_1)^2 + \mathbb{E}(X_1'X_1)\mathbb{E}(X_2'X_2) \\ &\quad - \mathbb{E}(X_3'X_3)\mathbb{E}[(X_1'X_2)^2 - (X_1'X_1) - (X_2'X_2)] \\ &= \mathbb{E}(X_1'X_2)^2(X_1'X_3)^2 - 2\mathbb{E}(X_1'X_1)(X_1'X_3)^2 \\ &\quad - 2\mathbb{E}(X_1'X_1)\mathbb{E}(X_1'X_2)^2 + \mathbb{E}(X_1'X_1)^2 + 3[\mathbb{E}(X_1'X_1)]^2.\end{aligned}$$

First, we compute  $\mathbb{E}(X_1'X_2)^2(X_1'X_3)^2$ , for which we have

$$\begin{aligned}\mathbb{E}(X_1'X_2)^2(X_1'X_3)^2 &= \mathbb{E}[(Z_1'AZ_2)^2(Z_1'AZ_3)^2] \\ &= \mathbb{E}[\mathbb{E}[Z_2'AZ_1Z_1'AZ_2|Z_1]\mathbb{E}[Z_3'AZ_1Z_1'AZ_3|Z_1]] \\ &= \mathbb{E}[\text{tr}^2(AZ_1Z_1'A)] \\ &= \mathbb{E}[(Z_1'A^2Z_1)^2] \\ &= \text{tr}^2(A^2) + 2\text{tr}(A^4) = \text{tr}^2(\Sigma^2) + 2\text{tr}(\Sigma^4).\end{aligned}$$

Next, we compute  $E(X'_1 X_1)(X'_1 X_3)^2$ , for which we have

$$\begin{aligned} E(X'_1 X_1)(X'_1 X_3)^2 &= E[(Z'_1 A Z_1)(Z'_1 A Z_3)^2] \\ &= E[(Z'_1 A Z_1)E[Z'_3 A Z_1 Z'_1 A Z_3 | Z_1]] \\ &= E[(Z'_1 A Z_1)(Z'_1 A^2 Z_1)] \\ &= 2 \operatorname{tr}(A^3) + \operatorname{tr}(A^2) \operatorname{tr}(A) = 2 \operatorname{tr}(\Sigma^3) + \operatorname{tr}(\Sigma^2) \operatorname{tr}(\Sigma). \end{aligned}$$

We further note that  $E(X'_1 X_1)^2 = E(Z'_1 A Z_1)^2 = 2 \operatorname{tr}(\Sigma^2) + \operatorname{tr}^2(\Sigma)$ , that  $E(X'_1 X_2)^2 = \operatorname{tr}(\Sigma^2)$ , and that  $E(X'_1 X_1) = \operatorname{tr}(\Sigma)$ . Thus, we obtain that

$$\begin{aligned} &E[(X'_1 X_2)^2 - (X'_1 X_1) - (X'_2 X_2)][(X'_1 X_3)^2 - (X'_1 X_1) - (X'_3 X_3)] \\ &= \operatorname{tr}^2(\Sigma^2) + 2 \operatorname{tr}(\Sigma^4) - 4 \operatorname{tr}(\Sigma^3) - 4 \operatorname{tr}(\Sigma^2) \operatorname{tr}(\Sigma) + 2 \operatorname{tr}(\Sigma^2) + 4 \operatorname{tr}^2(\Sigma). \end{aligned}$$

Noting that  $E[(X'_1 X_2)^2 - (X'_1 X_1) - (X'_2 X_2)] = \operatorname{tr}(\Sigma^2) - 2 \operatorname{tr}(\Sigma)$ , we obtain the claim. □

**Proof of (9).** With Lemma 4, we have

$$\begin{aligned} &\operatorname{Var}\left(\sum_{1 \leq i < j \leq n} h(X_i, X_j)\right) \\ &= \sum_{1 \leq i < j \leq n} \operatorname{Var}(h(X_i, X_j)) + 2 \sum_{\substack{1 \leq i < j, i' < j' \leq n \\ i=i' \text{ or } j=j'}} \operatorname{Cov}(h(X_i, X_j), h(X_{i'}, X_{j'})) \\ &= \frac{n(n-1)}{2} \operatorname{Var}(h(X_1, X_2)) + 2 \frac{n(n-1)}{2} (n-2) \operatorname{Cov}(h(X_1, X_2), h(X_1, X_3)) \\ &= n(n-1)[\operatorname{tr}^2(\Sigma^2) + \operatorname{tr}(\Sigma^4)] + 2n(n-1)^2 \operatorname{tr}(\Sigma^2(\Sigma - I)^2). \end{aligned}$$

Multiplying both sides with  $4n^{-2}(n-1)^{-2}$ , we obtain (9). □

### A.3. Proof details for Proposition 3

#### A.3.1. Proof of (12)

First of all, we give a formal proof of the representation (12).

**Proof of (12).** The computations made in [7], Appendix, are handy for the proof here. Indeed, we have

$$u_{nk} = (E_k - E_{k-1}) \left[ \frac{1}{n} X'_k X_k \right] = \frac{1}{n} [X'_k X_k - \operatorname{tr}(\Sigma)], \tag{36}$$

$$\begin{aligned} v_{nk} &= (\mathbf{E}_k - \mathbf{E}_{k-1}) \left[ \frac{2}{n(n-1)} \sum_{i \neq k} (X'_i X_k)^2 \right] \\ &= \frac{2}{n(n-1)} [X'_k Q_{k-1} X_k - \text{tr}(Q_{k-1} \Sigma)] + \frac{2}{n} [X'_k \Sigma X_k - \text{tr}(\Sigma^2)], \end{aligned}$$

where  $Q_{k-1} = \sum_{i=1}^{k-1} (X_i X'_i - \Sigma)$ . Noting that  $D_{nk} = v_{nk} - 2u_{nk}$ , we obtain (12). □

A.3.2. Proof of (19)

To calculate  $\sigma_{nk}^2$ , we note that

$$\sigma_{nk}^2 = \mathbf{E}_{k-1} [D_{nk}^2] = 4\mathbf{E}_{k-1} [u_{nk}^2] - 4\mathbf{E}_{k-1} [u_{nk} v_{nk}] + \mathbf{E}_{k-1} [v_{nk}^2].$$

Thus, (19) is immediate with the following lemma.

**Lemma 5.** For  $u_{nk}, v_{nk}$  defined as in (36), we have

$$\begin{aligned} \mathbf{E}_{k-1} [u_{nk}^2] &= \frac{2}{n^2} \text{tr}(\Sigma^2), \\ \mathbf{E}_{k-1} [u_{nk} v_{nk}] &= \frac{4}{n^2(n-1)} \text{tr}(Q_{k-1} \Sigma^2) + \frac{4}{n^2} \text{tr}(\Sigma^3), \\ \mathbf{E}_{k-1} [v_{nk}^2] &= \frac{8}{n^2(n-1)^2} \text{tr}(Q_{k-1} \Sigma Q_{k-1} \Sigma) \\ &\quad + \frac{16}{n^2(n-1)} \text{tr}(Q_{k-1} \Sigma^3) + \frac{8}{n^2} \text{tr}(\Sigma^4). \end{aligned}$$

**Proof.** First, we have

$$\mathbf{E}_{k-1} [u_{nk}^2] = \frac{1}{n^2} \text{Var}(X'_k X_k) = \frac{2}{n^2} \text{tr}(\Sigma^2).$$

Next, we have from (33) that

$$\begin{aligned} \mathbf{E}_{k-1} [u_{nk} v_{nk}] &= \frac{2}{n^2(n-1)} \mathbf{E}_{k-1} [X'_k X_k - \text{tr}(\Sigma)] X'_k Q_{k-1} X_k + \frac{2}{n^2} [X'_k X_k - \text{tr}(\Sigma)] X'_k \Sigma X_k \\ &= \frac{2}{n^2(n-1)} [\text{tr}(\Sigma) \text{tr}(Q_{k-1} \Sigma) + 2 \text{tr}(Q_{k-1} \Sigma^2) - \text{tr}(\Sigma) \text{tr}(Q_{k-1} \Sigma)] \\ &\quad + \frac{2}{n^2} [\text{tr}(\Sigma) \text{tr}(\Sigma^2) + 2 \text{tr}(\Sigma^3) - \text{tr}(\Sigma) \text{tr}(\Sigma^2)] \\ &= \frac{4}{n^2(n-1)} \text{tr}(Q_{k-1} \Sigma^2) + \frac{4}{n^2} \text{tr}(\Sigma^3). \end{aligned}$$

Finally, we have

$$\begin{aligned} E_{k-1}[v_{nk}^2] &= \frac{4}{n^2(n-1)^2} E_{k-1}[X'_k Q_{k-1} X_k - \text{tr}(Q_{k-1} \Sigma)] X'_k Q_{k-1} X_k \\ &\quad + \frac{8}{n^2(n-1)} E_{k-1}[X'_k \Sigma X_k - \text{tr}(\Sigma^2)] X'_k Q_{k-1} X_k \\ &\quad + \frac{4}{n^2} E_{k-1}[X'_k \Sigma X_k - \text{tr}(\Sigma^2)] X'_k \Sigma X_k. \end{aligned}$$

Note that

$$\begin{aligned} E_{k-1}[X'_k Q_{k-1} X_k X'_k Q_{k-1} X_k] &= E_{k-1}[Z'_k \Gamma' Q_{k-1} \Gamma Z_k Z'_k \Gamma' Q_{k-1} \Gamma Z_k] \\ &= \text{tr}^2(\Gamma' Q_{k-1} \Gamma) + 2 \text{tr}(\Gamma' Q_{k-1} \Gamma \Gamma' Q_{k-1} \Gamma) \\ &= 2 \text{tr}(Q_{k-1} \Sigma Q_{k-1} \Sigma) + \text{tr}^2(Q_{k-1} \Sigma), \\ E[X'_k Q_{k-1} X_k X'_k \Sigma X_k] &= E_{k-1}[Z'_k \Gamma' Q_{k-1} \Gamma Z_k Z'_k \Gamma' \Sigma \Gamma Z_k] \\ &= \text{tr}(\Gamma' Q_{k-1} \Gamma) \text{tr}(\Gamma' \Sigma \Gamma) + 2 \text{tr}(\Gamma' Q_{k-1} \Gamma \Gamma' \Sigma \Gamma) \\ &= 2 \text{tr}(Q_{k-1} \Sigma^3) + \text{tr}(Q_{k-1} \Sigma) \text{tr}(\Sigma^2), \\ E_{k-1}[X'_k \Sigma X_k X'_k \Sigma X_k] &= E_{k-1}[Z'_k \Gamma' \Sigma \Gamma Z_k Z'_k \Gamma' \Sigma \Gamma Z_k] \\ &= 2 \text{tr}(\Gamma' \Sigma \Gamma \Gamma' \Sigma \Gamma) + \text{tr}^2(\Gamma' \Sigma \Gamma) \\ &= 2 \text{tr}(\Sigma^4) + \text{tr}^2(\Sigma^2). \end{aligned}$$

Collecting terms, we complete the proof. □

### A.3.3. Proof of Lemma 2

Finally, we shall complete the proof of Lemma 2.

Recall that  $M_{k-1} = A \sum_{i=1}^{k-1} (Z_i Z'_i - I) A$ , and so

$$\begin{aligned} \text{tr}(M_{k-1}^2) &= \sum_{i=1}^{k-1} \text{tr}(A(Z_i Z'_i - I) A^2 (Z_i Z'_i - I) A) \\ &\quad + 2 \sum_{i=1}^{k-2} \sum_{j=i+1}^{k-1} \text{tr}(A(Z_i Z'_i - I) A^2 (Z_j Z'_j - I) A). \end{aligned} \tag{37}$$

For any fixed  $i$ , we have

$$\begin{aligned} E[\text{tr}(A(Z_i Z'_i - I) A^2 (Z_i Z'_i - I) A)] &= E[\text{tr}(A Z_i Z'_i A^2 Z_i Z'_i A)] - 2E[\text{tr}(A^2 Z_i Z'_i A^2)] + \text{tr}(A^4) \\ &= E[(Z'_i A^2 Z_i)^2] - 2E[Z'_i A^4 Z_i] + \text{tr}(A^4) \end{aligned}$$

$$\begin{aligned}
 &= 2 \operatorname{tr}(A^4) + \operatorname{tr}^2(A^2) - 2 \operatorname{tr}(A^4) + \operatorname{tr}(A^4) \\
 &= \operatorname{tr}^2(\Sigma^2) + \operatorname{tr}(\Sigma^4).
 \end{aligned}$$

On the other hand, for any  $i \neq j$ , we have

$$\mathbb{E}[\operatorname{tr}(A(Z_i Z_i' - I)A^2(Z_j Z_j' - I)A)] = 0.$$

In addition, we note that the terms in (37) are all uncorrelated. Therefore, we obtain that

$$\mathbb{E}[\operatorname{tr}(M_{k-1}^2)] = (k - 1)[\operatorname{tr}^2(\Sigma^2) + \operatorname{tr}(\Sigma^4)].$$

Moreover, we have

$$\begin{aligned}
 \operatorname{Var}[\operatorname{tr}(M_{k-1}^2)] &= \sum_{i=1}^{k-1} \operatorname{Var}[\operatorname{tr}(A(Z_i Z_i' - I)A^2(Z_i Z_i' - I)A)] \\
 &\quad + 2 \sum_{i=1}^{k-2} \sum_{j=i+1}^{k-1} \operatorname{Var}[\operatorname{tr}(A(Z_i Z_i' - I)A^2(Z_j Z_j' - I)A)] \quad (38) \\
 &= (k - 1)V_1 + 2(k - 1)(k - 2)V_2.
 \end{aligned}$$

Here,  $V_1 = \operatorname{Var}[\operatorname{tr}(A(Z_1 Z_1' - I)A^2(Z_1 Z_1' - I)A)]$  and  $V_2 = \operatorname{Var}[\operatorname{tr}(A(Z_1 Z_1' - I)A^2(Z_2 Z_2' - I)A)]$ .

Consider  $V_1$  first, for which we have the decomposition

$$\begin{aligned}
 V_1 &= \operatorname{Var}[(Z_1' A^2 Z_1)^2 - 2Z_1' A^4 Z_1] \\
 &= \operatorname{Var}[(Z_1' A^2 Z_1)^2] + 4 \operatorname{Var}(Z_1' A^4 Z_1) - 4 \operatorname{Cov}[(Z_1' A^2 Z_1)^2, Z_1' A^4 Z_1].
 \end{aligned}$$

To calculate  $\operatorname{Var}[(Z_1' A^2 Z_1)^2]$ , we note that the eigenvalues of  $A^2$  are  $\lambda_1^2 \geq \dots \geq \lambda_p^2$ , where  $\lambda_1 \geq \dots \geq \lambda_p$  are the eigenvalues of  $\Sigma$ . Let  $U_j \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ , for  $j = 1, \dots, p$ . By the moment generating function of  $\chi_{(1)}^2$  distribution, we have  $\mathbb{E}[U_j^2] = 1$ ,  $\mathbb{E}[U_j^4] = 3$ ,  $\mathbb{E}[U_j^6] = 15$ , and  $\mathbb{E}[U_j^8] = 105$ . Then, we obtain that

$$\begin{aligned}
 \mathbb{E}(Z_1' A^2 Z_1)^4 &= \mathbb{E}\left[\left(\sum_{j=1}^p \lambda_j^2 U_j^2\right)^4\right] \\
 &= \mathbb{E}\left[\sum_{j=1}^p \lambda_j^8 U_j^8 + 4 \sum_{j \neq l}^p \lambda_j^6 \lambda_l^2 U_j^6 U_l^2 + \frac{6}{2!} \sum_{j \neq l}^p \lambda_j^4 \lambda_l^4 U_j^4 U_l^4 \right. \\
 &\quad \left. + \frac{12}{2!} \sum_{j \neq l \neq m}^p \lambda_j^4 \lambda_l^2 \lambda_m^2 U_j^4 U_l^2 U_m^2 + \frac{24}{4!} \sum_{j \neq l \neq m \neq r}^p \lambda_j^2 \lambda_l^2 \lambda_m^2 \lambda_r^2 U_j^2 U_l^2 U_m^2 U_r^2\right]
 \end{aligned}$$

$$\begin{aligned}
 &= 105 \sum_{j=1}^p \lambda_j^8 + 15 \cdot 4 \sum_{j \neq l}^p \lambda_j^6 \lambda_l^2 + 9 \cdot \frac{6}{2!} \sum_{j \neq l}^p \lambda_j^4 \lambda_l^4 \\
 &\quad + 3 \cdot \frac{12}{2!} \sum_{j \neq l \neq m}^p \lambda_j^4 \lambda_l^2 \lambda_m^2 + \frac{24}{4!} \sum_{j \neq l \neq m \neq r}^p \lambda_j^2 \lambda_l^2 \lambda_m^2 \lambda_r^2 \\
 &= \text{tr}^4(\Sigma^2) + 12 \text{tr}(\Sigma^4) \text{tr}^2(\Sigma^2) + 12 \text{tr}^2(\Sigma^4) + 32 \text{tr}(\Sigma^6) \text{tr}(\Sigma^2) + 48 \text{tr}(\Sigma^8).
 \end{aligned}$$

Observing that  $E[(Z_1' A^2 Z_1)^2] = 2 \text{tr}(\Sigma^4) + \text{tr}^2(\Sigma^2)$ , we get

$$\text{Var}[(Z_1' A^2 Z_1)^2] = 8 \text{tr}(\Sigma^4) \text{tr}^2(\Sigma^2) + 8 \text{tr}^2(\Sigma^4) + 32 \text{tr}(\Sigma^6) \text{tr}(\Sigma^2) + 48 \text{tr}(\Sigma^8).$$

Next, we compute  $\text{Var}(Z_1' A^4 Z_1)$ , for which we have

$$\begin{aligned}
 E[(Z_1' A^4 Z_1)^2] &= 2 \text{tr}(\Sigma^8) + \text{tr}^2(\Sigma^4), \\
 E[Z_1' A^4 Z_1] &= \text{tr}(\Sigma^4).
 \end{aligned}$$

Therefore, we get  $\text{Var}(Z_1' A^4 Z_1) = 2 \text{tr}(\Sigma^8)$ .

Now switch to  $\text{Cov}[(Z_1' A^2 Z_1)^2, Z_1' A^4 Z_1]$ . We note that

$$\begin{aligned}
 &E[(Z_1' A^2 Z_1)^2 Z_1' A^4 Z_1] \\
 &= E \left[ \left( \sum_{j=1}^p \lambda_j^4 U_j^4 + \sum_{j \neq l}^p \lambda_j^2 \lambda_l^2 U_j^2 U_l^2 \right) \sum_{j=1}^p \lambda_j^4 U_j^2 \right] \\
 &= E \left[ \sum_{j=1}^p \lambda_j^8 U_j^6 + \sum_{j \neq l}^p \lambda_j^4 \lambda_l^4 U_j^4 U_l^2 + 2 \sum_{j \neq l}^p \lambda_j^6 \lambda_l^2 U_j^4 U_l^2 + \sum_{j \neq l \neq m}^p \lambda_j^4 \lambda_l^2 \lambda_m^2 U_j^2 U_l^2 U_m^2 \right] \\
 &= 15 \sum_{j=1}^p \lambda_j^8 + 3 \sum_{j \neq l}^p \lambda_j^4 \lambda_l^4 + 6 \sum_{j \neq l}^p \lambda_j^6 \lambda_l^2 + \sum_{j \neq l \neq m}^p \lambda_j^4 \lambda_l^2 \lambda_m^2 \\
 &= \text{tr}(\Sigma^4) \text{tr}^2(\Sigma^2) + 4 \text{tr}(\Sigma^6) \text{tr}(\Sigma^2) + 2 \text{tr}^2(\Sigma^4) + 8 \text{tr}(\Sigma^8).
 \end{aligned}$$

By previous expression for  $E[(Z_1' A^2 Z_1)^2]$  and  $E[Z_1' A^4 Z_1]$ , we obtain that

$$\text{Cov}[(Z_1' A^2 Z_1)^2, Z_1' A^4 Z_1] = 4 \text{tr}(\Sigma^6) \text{tr}(\Sigma^2) + 8 \text{tr}(\Sigma^8).$$

Finally, we obtain that

$$V_1 = 8 \text{tr}(\Sigma^4) \text{tr}^2(\Sigma^2) + 8 \text{tr}^2(\Sigma^4) + 16 \text{tr}(\Sigma^6) \text{tr}(\Sigma^2) + 24 \text{tr}(\Sigma^8). \tag{39}$$

Switch to the calculation of  $V_2$ . We first note that

$$\begin{aligned} V_2 &= \text{Var}[\text{tr}(A(Z_1 Z_1' - I)A^2(Z_2 Z_2' - I)A)] \\ &= \text{Var}[\text{tr}(AZ_1 Z_1' A^2 Z_1 Z_1' A) - \text{tr}(A^2 Z_1 Z_1' A^2) - \text{tr}(A^2 Z_2 Z_2' A^2)] \\ &= \text{Var}[(Z_1' A^2 Z_2)^2] + 2 \text{Var}(Z_1' A^4 Z_1) - 2 \text{Cov}[(Z_1' A^2 Z_2)^2, Z_1' A^4 Z_1]. \end{aligned}$$

Note that  $E[(Z_1' A^2 Z_2)^4] = 3 \text{tr}^2(\Sigma^4) + 6 \text{tr}(\Sigma^8)$ , and  $E[(Z_1' A^2 Z_2)^2] = \text{tr}(\Sigma^4)$ . We then get

$$\text{Var}[(Z_1' A^2 Z_2)^2] = 2 \text{tr}^2(\Sigma^4) + 6 \text{tr}(\Sigma^8).$$

In addition, previous calculation gives  $\text{Var}(Z_1' A^4 Z_1) = 2 \text{tr}(\Sigma^8)$ . Then for  $\text{Cov}[(Z_1' A^2 Z_2)^2, Z_1' A^4 Z_1]$ , we have

$$\begin{aligned} E[(Z_1' A^2 Z_1)^2 Z_1' A^4 Z_1] &= E\left[\left(\sum_{j=1}^p \lambda_j^2 U_j V_j\right)^2 \sum_{j=1}^p \lambda_j^4 U_j^2\right] \\ &= E\left[\left(\sum_{j=1}^p \lambda_j^4 U_j^2 V_j^2\right)\left(\sum_{j=1}^p \lambda_j^4 U_j^2\right)\right] \\ &= E\left[\sum_{j=1}^p \lambda_j^8 U_j^4 V_j^2 + \sum_{j \neq l}^p \lambda_j^4 \lambda_l^4 U_j^2 V_j^2 U_l^2\right] \\ &= 3 \sum_{j=1}^p \lambda_j^8 + \sum_{j \neq l}^p \lambda_j^4 \lambda_l^4 \\ &= \text{tr}^2(\Sigma^4) + 2 \text{tr}(\Sigma^8). \end{aligned}$$

This leads to the conclusion that  $\text{Cov}[(Z_1' A^2 Z_2)^2, Z_1' A^4 Z_1] = 2 \text{tr}(\Sigma^8)$ , and so

$$V_2 = 2 \text{tr}^2(\Sigma^4) + 6 \text{tr}(\Sigma^8). \tag{40}$$

Replacing  $V_1$  and  $V_2$  in (38) by (39) and (40), we obtain the claimed formula for  $\text{Var}[\text{tr}(M_{k-1}^2)]$ .

### Acknowledgements

The research of Tony Cai was supported in part by NSF FRG Grant DMS-08-54973. The research of Zongming Ma was supported in part by the Dean’s Research Fund of The Wharton School.

### References

[1] Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley-Interscience [John Wiley & Sons]. MR1990662

- [2] Bai, Z., Jiang, D., Yao, J.F. and Zheng, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840. [MR2572444](#)
- [3] Birke, M. and Dette, H. (2005). A note on testing the covariance matrix for large dimension. *Statist. Probab. Lett.* **74** 281–289. [MR2189467](#)
- [4] Cai, T.T. and Jiang, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525. [MR2850210](#)
- [5] Cai, T.T., Liu, W. and Xia, Y. (2011). Two-sample covariance matrix testing and support recovery. Technical report.
- [6] Chen, S.X. and Li, J. (2012). Two sample tests for high dimensional covariance matrices. *Ann. Statist.* **40** 908–940. [MR2985938](#)
- [7] Chen, S.X., Zhang, L.X. and Zhong, P.S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. [MR2724863](#)
- [8] Heyde, C.C. and Brown, B.M. (1970). On the departure from normality of a certain class of martingales. *Ann. Math. Statist.* **41** 2161–2165. [MR0293702](#)
- [9] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. [MR0144363](#)
- [10] Jiang, D., Jiang, T. and Yang, F. (2012). Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. *J. Statist. Plann. Inference* **142** 2241–2256. [MR291184](#)
- [11] Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. [MR1863961](#)
- [12] Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. [MR1926169](#)
- [13] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley Series in Probability and Mathematical Statistics. New York: Wiley. [MR0652932](#)
- [14] Nagao, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.* **1** 700–709. [MR0339405](#)
- [15] Onatski, A., Moreira, M.J. and Hallin, M. (2011). Asymptotic power of sphericity tests for high-dimensional data. Available at <http://www.econ.cam.ac.uk/people/faculty/ao319/pubs/WPOnatskiMoreira.pdf>.
- [16] Roy, S.N. (1957). *Some Aspects of Multivariate Analysis*. New York: Wiley. [MR0092296](#)
- [17] Srivastava, M.S. (2005). Some tests concerning the covariance matrix in high dimensional data. *J. Japan Statist. Soc.* **35** 251–272. [MR2328427](#)
- [18] Xiao, H. and Wu, W.B. (2011). Simultaneous inference on sample covariances. Available at [arXiv:1109.0524v1](http://arxiv.org/abs/1109.0524v1).

*Received October 2011 and revised May 2012*